

# DISSERTATION

submitted to the  
Combined Faculties of the Natural Sciences and Mathematics  
of the Ruperto-Carola-University of Heidelberg, Germany  
for the degree of

**Doctor of Natural Sciences**

Put forward by

**Dipl.-Phys. Marco Meißner**

born in: Mannheim, Germany

Oral examination: 5th February 2015





---

**Abstract:** In this thesis, measurements of particle production and particle correlations using proton-proton ( $pp$ ) and proton-ion ( $p\text{Pb}$ ) collisions recorded with the LHCb experiment at the Large Hadron Collider are presented.

The charged particle multiplicity and the differential charged particle density are studied in  $pp$  collisions at a centre-of-mass energy of  $\sqrt{s} = 7$  TeV. The measurement is performed for events with at least one charged particle in the kinematic range of  $p > 2$  GeV/ $c$ ,  $p_T > 0.2$  GeV/ $c$  and  $2.0 < \eta < 4.8$ . The mean particle multiplicity,  $\mu$ , and the corresponding root-mean-square deviation,  $\sigma$ , are measured to be  $\mu = 11.30 \pm 0.01 \pm 0.09$  and  $\sigma = 9.50 \pm 0.01 \pm 0.02$ , where the uncertainties are statistical and systematic, respectively. Results for particle multiplicities and particle densities are presented as functions of  $\eta$  and  $p_T$  of the particles, and are compared to predictions from different Monte Carlo event generators.

Furthermore, a measurement of two-particle angular correlations is performed by using charged particles produced in  $p\text{Pb}$  collisions at a nucleon-nucleon centre-of-mass energy of  $\sqrt{s_{NN}} = 5.02$  TeV. Data with two different beam configurations are analysed, in which either the proton or the nucleus remnant is traversing the LHCb detector. The particle correlations are measured using the azimuthal angle separation,  $\Delta\phi$ , and pseudorapidity separation,  $\Delta\eta$ . The emergence of a correlation structure on the near side ( $\Delta\phi = 0$ ), elongated over a long range in  $\Delta\eta$  is observed for particles in a  $p_T$ -range of  $1 - 3$  GeV/ $c$  in events exhibiting a high particle multiplicity. This measurement is the first observation of the *near-side ridge* in  $p\text{Pb}$  collisions in the forward region and complements similar observations in  $pp$  and  $p\text{Pb}$  collisions in the central rapidity region.

**Kurzfassung:** In vorliegender Arbeit werden Messungen von Teilchenproduktion und Teilchenkorrelationen vorgestellt, die auf am LHCb-Experiment am Large Hadron Collider gewonnenen Daten aus Proton-Proton ( $pp$ ) und Proton-Ionen ( $p\text{Pb}$ ) Kollisionen basieren.

Die Multiplizität sowie die differenzielle Dichte geladener Teilchen werden in  $pp$  Kollisionen bei einer Schwerpunktsenergie von  $\sqrt{s} = 7$  TeV bestimmt. Die Messung basiert auf Ereignissen mit mindestens einem geladenen Teilchen im kinematischen Bereich von  $p > 2$  GeV/ $c$ ,  $p_T > 0.2$  GeV/ $c$  und  $2.0 < \eta < 4.8$ . Die mittlere Teilchenmultiplizität  $\mu$  und die Breite der Verteilung  $\sigma$  werden hierbei zu  $\mu = 11.30 \pm 0.01 \pm 0.09$  und  $\sigma = 9.50 \pm 0.01 \pm 0.02$  ermittelt, wobei die angegebenen Fehler jeweils statistischen bzw. systematischen Ursprungs sind. Die Messergebnisse für die Teilchenmultiplizität und die differenzielle Teilchendichte werden als Funktion von  $\eta$  und  $p_T$  der Teilchen präsentiert und mit Vorhersagen verschiedener Monte Carlo Ereignisgeneratoren verglichen.

Im Weiteren wird eine Messung von Zwei-Teilchen Winkelkorrelationen vorgestellt, welche für geladene Teilchen aus  $p\text{Pb}$  Kollisionen bei einer Nukleon-Nukleon Schwerpunktsenergie von  $\sqrt{s_{NN}} = 5.02$  TeV bestimmt wird. Hierfür werden Daten mit zwei verschiedenen Strahlkonfigurationen analysiert, in welchen entweder Proton- oder Bleikernfragmente den LHCb-Detektor durchqueren. Die Teilchenkorrelationen werden als Azimutalwinkeldifferenz  $\Delta\phi$  und Pseudorapiditätssdifferenz  $\Delta\eta$  zweier Teilchen gemessen. In Kollisionen mit einer hohen Aktivität wird für Teilchen mit einem Transversalimpuls im Bereich von  $1 - 3$  GeV/ $c$  die Entstehung einer Korrelationsstruktur bei  $\Delta\phi = 0$  nachgewiesen, die sich über einen großen  $\Delta\eta$ -Bereich erstreckt. Diese Messung ist der erste Nachweis des sogenannten *near-side ridge* in  $p\text{Pb}$  Kollisionen im kinematischen Vorwärtsbereich und ergänzt vergleichbare Beobachtungen in  $pp$  und  $p\text{Pb}$  Kollisionen im zentralen Rapiditätsbereich.

|  |           |
|--|-----------|
| <b>1. Introduction</b>   | <b>1</b>  |
| <b>2. Theory of Particle Production at the LHC</b>               | <b>5</b>  |
| 2.1. The Standard Model in a nutshell . . . . .                  | 5         |
| 2.2. Particle Production in $pp$ collisions at the LHC . . . . . | 8         |
| 2.2.1. Anatomy of a $pp$ collision . . . . .                     | 10        |
| 2.2.2. Production of $b$ hadrons at the LHC . . . . .            | 11        |
| 2.3. Monte Carlo event generators . . . . .                      | 12        |
| 2.3.1. The PYTHIA generator . . . . .                            | 12        |
| 2.3.2. PYTHIA tunes . . . . .                                    | 15        |
| 2.3.3. The PHOJET generator . . . . .                            | 17        |
| 2.3.4. The HERWIG++ generator . . . . .                          | 17        |
| <b>3. The Large Hadron Collider at CERN</b>                      | <b>21</b> |
| 3.1. The LHC experiments . . . . .                               | 22        |
| 3.2. A multi-purpose collider . . . . .                          | 23        |
| <b>4. The LHCb experiment at the LHC</b>                         | <b>25</b> |
| 4.1. Tracking system . . . . .                                   | 27        |
| 4.1.1. The magnet . . . . .                                      | 27        |
| 4.1.2. Vertex Locator . . . . .                                  | 28        |
| 4.1.3. Tracker Turicensis . . . . .                              | 30        |
| 4.1.4. Inner Tracker . . . . .                                   | 31        |
| 4.1.5. Outer Tracker . . . . .                                   | 31        |
| 4.2. Beam pipe . . . . .   | 32        |
| 4.3. Particle identification . . . . .                           | 33        |
| 4.3.1. Ring Imaging Cherenkov detectors . . . . .                | 33        |
| 4.3.2. Calorimeter system . . . . .                              | 35        |
| 4.3.3. Muon system . . . . .                                     | 37        |
| 4.4. Trigger system . . . . .                                    | 38        |

|            |   |           |
|------------|---|-----------|
| 4.5.       | Data taking . . . . .   | 40        |
| 4.5.1.     | Proton-proton period . . . . .  | 40        |
| 4.5.2.     | Proton-ion period . . . . .   | 41        |
| <b>5.</b>  | <b>From particles to tracks - Event and track reconstruction at LHCb</b>                  | <b>43</b> |
| 5.1.       | The LHCb software framework . . . . .   | 44        |
| 5.2.       | Track types in LHCb . . . . .   | 45        |
| 5.3.       | Track reconstruction LHCb . . . . .   | 46        |
| 5.4.       | Reconstruction artefacts and inefficiencies . . . . .                                     | 48        |
| 5.4.1.     | Definitions and nomenclatures . . . . .   | 49        |
| 5.4.2.     | Reconstruction efficiencies . . . . .   | 50        |
| 5.4.3.     | Reconstruction artefacts . . . . .  | 51        |
| <b>1.</b>  | <b>Particle production in <math>pp</math> collisions at <math>\sqrt{s} = 7</math> TeV</b> | <b>53</b> |
| <b>6.</b>  | <b>Introduction and analysis strategy</b>   | <b>55</b> |
| <b>7.</b>  | <b>Data samples</b>   | <b>59</b> |
| <b>8.</b>  | <b>Event definition and data selection</b>  | <b>61</b> |
| 8.1.       | Event definition and selection . . . . .  | 61        |
| 8.2.       | Track selection . . . . .   | 63        |
| <b>9.</b>  | <b>Data analysis and correction</b>   | <b>67</b> |
| 9.1.       | Correction for undetected events . . . . .  | 68        |
| 9.2.       | Correction for reconstruction artefacts and non-prompt particles . . . . .                | 70        |
| 9.2.1.     | Fake tracks . . . . .   | 71        |
| 9.2.2.     | Duplicate tracks . . . . .  | 73        |
| 9.2.3.     | Non-prompt tracks . . . . .   | 74        |
| 9.2.4.     | Background subtraction - mean particle densities . . . . .                                | 75        |
| 9.2.5.     | Background subtraction - multiplicity distributions . . . . .                             | 76        |
| 9.3.       | Pile-up correction . . . . .  | 78        |
| 9.4.       | Efficiency correction and unfolding procedure . . . . .                                   | 81        |
| 9.4.1.     | Detector acceptance . . . . .   | 81        |
| 9.4.2.     | Track finding efficiency . . . . .  | 82        |
| 9.4.3.     | Efficiency correction for particle densities . . . . .                                    | 83        |
| 9.4.4.     | Unfolding of multiplicity distributions . . . . .   | 84        |
| <b>10.</b> | <b>Systematic studies</b>   | <b>93</b> |
| 10.1.      | Fake track estimation on data and simulation . . . . .                                    | 93        |
| 10.2.      | Cross-check for duplicate tracks . . . . .  | 97        |
| 10.3.      | Non-prompt particles . . . . .  | 98        |
| 10.4.      | Uncertainties on track finding and detector acceptance . . . . .                          | 99        |
| 10.4.1.    | Detector acceptance . . . . .   | 99        |
| 10.4.2.    | Track finding . . . . .   | 100       |
| 10.4.3.    | Multiplicity unfolding with modified response matrix . . . . .                            | 100       |

|  |            |
|--|------------|
| 10.4.4. Efficiency uncertainties of particle densities . . . . .   | 101        |
| 10.5. Uncertainties to unobserved events . . . . .   | 101        |
| 10.6. Total systematics . . . . .  | 104        |
| 10.7. Check of unfolding procedure . . . . .   | 106        |
| <b>11. Results of charged particle densities and multiplicities</b>  | <b>109</b> |
| 11.1. Charged particle densities . . . . .   | 109        |
| 11.2. Charged particle multiplicities . . . . .  | 111        |
| <b>II. Two-particle correlations in</b>  |            |
| <b><math>p+\text{Pb}</math> and <math>\text{Pb}+p</math> collisions at <math>\sqrt{s_{NN}} = 5.02</math> TeV</b> | <b>117</b> |
| <b>12. An introduction to two-particle correlations</b>  | <b>119</b> |
| 12.1. Brief review of recent two-particle correlation measurements . . . . .                                     | 120        |
| 12.2. Properties of the two-dimensional correlation function . . . . .   | 122        |
| 12.3. Theoretical interpretations of the ridge . . . . .   | 126        |
| <b>13. Strategy of the analysis</b>  | <b>129</b> |
| <b>14. Preparation and selection of data and simulation</b>  | <b>131</b> |
| 14.1. The data samples . . . . .   | 133        |
| 14.2. Trigger configuration . . . . .  | 133        |
| 14.3. Event selection . . . . .  | 135        |
| 14.4. HIJING simulation . . . . .  | 144        |
| 14.5. Track selection . . . . .  | 145        |
| 14.5.1. Definition of the kinematic range . . . . .  | 145        |
| 14.5.2. Suppression of reconstruction artefacts . . . . .  | 146        |
| 14.5.3. Selection of prompt particles . . . . .  | 147        |
| 14.6. Selection summary and overview of data samples . . . . .   | 148        |
| <b>15. Measurement of angular correlations in proton-ion collisions</b>  | <b>151</b> |
| 15.1. Data correction for reconstruction artefacts and mis-selection . . . . .                                   | 151        |
| 15.1.1. Fake-track contamination . . . . .   | 152        |
| 15.1.2. Non-prompt contribution . . . . .  | 155        |
| 15.1.3. Detector acceptance . . . . .  | 157        |
| 15.1.4. Track finding efficiency . . . . .   | 158        |
| 15.1.5. Correction procedure . . . . .   | 160        |
| 15.2. Event Classification . . . . .   | 162        |
| 15.3. The correlation function . . . . .   | 167        |
| 15.4. One-dimensional long-range correlations in $\Delta\phi$ . . . . .  | 173        |
| 15.5. Correlations in HIJING simulation . . . . .  | 174        |
| 15.6. Cross-check for artificial correlation structures . . . . .  | 175        |
| <b>16. Results for angular correlations</b>  | <b>177</b> |
| 16.1. Results from $p+\text{Pb}$ data . . . . .  | 177        |
| 16.1.1. Overview of 2D correlations in $p+\text{Pb}$ . . . . .   | 177        |

|  |            |
|--|------------|
| 16.1.2. Long-range correlation in $p+\text{Pb}$ . . . . .                  | 181        |
| 16.1.3. 1D associated yield versus $\Delta\phi$ in $p+\text{Pb}$ . . . . . | 181        |
| 16.2. Results from $\text{Pb}+p$ data . . . . .                            | 184        |
| 16.2.1. Overview of 2D correlations in $\text{Pb}+p$ . . . . .             | 184        |
| 16.2.2. Long-range correlation in $\text{Pb}+p$ . . . . .                  | 186        |
| 16.2.3. 1D associated yield versus $\Delta\phi$ in $\text{Pb}+p$ . . . . . | 186        |
| 16.3. Qualitative comparison of the ridge among LHC experiments . . . . .  | 189        |
| <b>17. Systematic checks</b>   | <b>193</b> |
| 17.1. Consistency check - magnetic field polarity . . . . .                | 193        |
| 17.2. Alternative event-activity classification . . . . .                  | 196        |
| 17.3. Influence of correction factors . . . . .                            | 198        |
| 17.4. Uncertainties related to multiple PVs . . . . .                      | 201        |
| 17.5. Uncertainties related to track selection . . . . .                   | 204        |
| <b>18. Conclusion</b>  | <b>207</b> |
| <b>Bibliography</b>  | <b>212</b> |

# CHAPTER 1

---

## Introduction

---

Tapestries are made by many artisans working together. The contributions of separate workers cannot be discerned in the completed work, and the loose and false threads have been covered over. So it is in our picture of particle physics.

---

Nobel Lecture, 8 Dec 1979,  
Sheldon Lee Glashow

The field of particle physics comprises the science about the nature of the smallest particles which our world is build of and about the mechanisms which hold them together. Our knowledge of the fundamental particles and their interactions is described by a single theory, the Standard Model (SM). It is founded on theoretical work [1–3] which has been developed already in the 1960's. As the introductory quote from one of its founders says, there have been a variety of theories built and experiments performed which all contributed in some way to our current knowledge, although not all leading to discoveries. Since the time of this statement, the Standard Model has been tested extensively by experiments in order to prove or disprove its predictions and to search for new phenomena.

Many discoveries in fundamental particle physics are achieved in particle collisions. New, so far unknown, interactions or particles can only be accessed by probing higher energy regimes, performing high-precision measurements, or searching for very rare processes. These kind of discoveries rely on the technical evolution which is needed for building new machines and experiments that supersede previous ones in terms of the energy of the colliding particles and the rate of collisions.

With the start of the Large Hadron Collider (LHC) on November 23, 2009, the day on which the first proton-proton collisions were delivered, a new chapter began. The up to date most powerful particle accelerator enabled to find the last fundamental particle predicted by the SM, the Higgs Boson [4,5]. Besides this key discovery, the LHC and its four major experiments have been build to address open questions in particle physics.

The SM by all its achievements does not give a complete description of nature. Several observations imply that unknown physics beyond the SM must exist.

An unsolved puzzle of our universe is the fact that only a tiny fraction (4.9%) of the mass/energy inventory consists of ordinary matter which stars and galaxies are made of. According to recent measurements of the cosmic-microwave-background by the Planck collaboration [6], the remaining bulk is related to dark matter (26.8%) and dark energy (68.3%) which are both not described within the Standard Model. The large multi-purpose detectors, ATLAS and CMS, perform direct searches of new particles that can be related to this, yet, undiscovered form of matter or energy.

Another open question addresses the visible matter surrounding us. From our understanding, matter and anti-matter are always created in equal parts. However, the universe, including our earth, consists of matter only. That implies a mechanism being responsible for a slight excess of matter in the early stages of the universe which is still present today. Within the Standard Model an identical behaviour for particles and their anti-particles is claimed. The conversion between both counterparts is obtained by a Charge ( $C$ ) and Parity ( $P$ ) transformation. However, this symmetry is not fully conserved and commonly referred to as  $CP$ -violation. The size of the asymmetry is still several orders of magnitude smaller than the one needs to explain the observed imbalance of matter/anti-matter in our universe. One of the main goals of the LHCb experiment is to precisely measure  $CP$ -asymmetries predicted in the SM and search for evidences of new physics that can explain this mystery.

Coming back to the incipient quote, the direct and indirect discoveries in modern big-scale experiments require a precise understanding of the detectors and the bulk of surrounding known physics processes. The observation of a new process or a new particle, such as the Higgs boson discovery, requires a discrimination of the signal from background which can be a sophisticated task. Very often this critically depends on good simulations of the known physics processes which dominate all measurements as a background. Even if the nature of the underlying single processes is theoretically understood, the calculation of a complex system from first principles, such as the always present soft particle production, is impossible. Instead, this is approached by a phenomenological modelling of the processes. The applied models use parameters which have to be optimised depending on the colliding particle species and their energies. The first half of the presented thesis describes an analysis of charged particle multiplicities and densities in proton-proton collisions at  $\sqrt{s} = 7$  TeV, which gives fundamental input used for these tuning procedures. The different geometry of the LHCb detector compared to the other LHC experiments allows accessing a unique kinematic range and performing complementary measurements which will help improving future simulations.

In addition to the proton-proton program of the LHC, which successfully explores the high energy frontier, the accelerator also provides collisions of heavy lead nuclei. The fourth LHC experiment, ALICE, is dedicated to the heavy-ion physics program. It addresses investigations of a new state of matter which existed in the early universe, the quark-gluon-plasma (QGP). Previous experiments found strong evidences that a QGP can also be produced in heavy-ion collisions. Its existence should be confirmed at the LHC.

---

General studies of particle production in heavy-ion collisions show collective behaviour when analysing the angular distributions of the particles. A certain two-particle correlation pattern, which was previously only known from heavy-ion collisions, has recently been also discovered in proton-proton [7] and proton-ion [8] collisions. This unexpected angular correlation of particles over a long range, often referred to as *the ridge*, generated high interest among physicist, since its existence is in contrast to theoretical explanations which were thought of being only applicable to the heavy-ion system. So far, this specific correlation pattern has only been confirmed in the *central* collision region. The complementary *forward* region can only be accessed by the LHCb detector.

The second half of this thesis will be dedicated to an analysis of angular correlations for particles created in proton-ion collisions at a nucleon-nucleon centre-of-mass energy of  $\sqrt{s_{NN}} = 5.02$  TeV. The analysis is based on data collected with the LHCb detector in early 2013. Proving or disproving the existence of this unexpected correlation structure in these data will help in understanding its origin and gives insights to particle production mechanisms.

This thesis is based on two analyses for which the author is the main proponent and data from the LHCb collaboration is used. It is outlined as follows. The first chapters, 2-5, give a general discussion about the theoretical background of particle production and the experimental conditions with focus on the LHCb experiment and the reconstruction of proton-proton collisions using the LHCb software. The first analysis, chapters 6-11, about the *"Measurement of charged particle multiplicities and densities in pp collisions at  $\sqrt{s} = 7$  TeV in the forward region"* has been published in EUR. PHYS. J. C (2014) 74:2888 [9] and is also documented in the corresponding internal LHCb note, LHCb-ANA-2011-084 [10]. The LHCb measurement was solely performed by the author, which is also the contact author of the LHCb publication. The second analysis, chapters 12-17, about *"Two-particle angular correlations in proton-ion collisions in the forward region"* is documented in the internal note LHCb-ANA-2014-091 [11] and is currently prepared for publication. The last chapter, 18, concludes both analyses and gives an outlook on future related measurements.



---

## Theory of Particle Production at the LHC

---

This chapter gives a brief introduction to the Standard Model of particle physics (based on Refs. [12–14]) and discusses particle production in the context of high-energetic  $pp$  collisions at the LHC. Particle collisions, including the production of particles and their decays, can be simulated using Monte Carlo event generators. These are extensively used within the high-energy physics community. The concept of an event generator is explained by using PYTHIA [15] as an example. Two further generators are briefly introduced which implement different realisations of phenomenological models. Furthermore, a selection of generator specific tunes (parameter sets) is given. Generator predictions using these tunes are compared to the data results obtained in the analysis of charged particle production.

### 2.1. The Standard Model in a nutshell

The Standard Model of particle physics (SM) is a relativistic, renormalisable quantum field theory which represents our current knowledge of fundamental particles and their interactions. With the exception of gravitation, it describes the whole phenomenology of matter and radiation at microscopic level. The SM is a combined theory of three fundamental interactions: strong, electromagnetic and weak interaction. The electromagnetic force is responsible for the interaction between electrically charged particles. As an example, atoms, molecules and solid state bodies are held together by this force. The strong force is responsible for binding protons and neutrons to nuclei. At short distances, it is the strongest of all interactions. All fundamental particles can interact via the weak force. It describes the decays of leptons and quarks and is the origin of radioactivity as seen *e.g.* in the beta-decay of the neutron. The fourth force, gravity, could not yet be incorporated into the Standard Model. However, the impact of the gravitational force is negligible for fundamental particle processes that can be accessed by current experiments.

In the SM, the matter is described by quantum fields that represent particles carrying

| Fermions (spin 1/2) |          |                 |   |            |                 |                  |
|---------------------|----------|-----------------|---|------------|-----------------|------------------|
| Generation          | Quarks   |                 |   | Leptons    |                 |                  |
|                     | Flavour  | Electric charge | Mass                                    | Flavour    | Electric charge | Mass             |
| 1 <sup>st</sup>     | <i>u</i> | +2/3 <i>e</i>   | 2.3 <sup>+0.7</sup> <sub>-0.5</sub> MeV | $\nu_e$    | 0               | < 2 eV           |
|                     | <i>d</i> | -1/3 <i>e</i>   | 4.8 <sup>+0.5</sup> <sub>-0.3</sub> MeV | <i>e</i>   | -1 <i>e</i>     | 511.0 keV        |
| 2 <sup>nd</sup>     | <i>c</i> | +2/3 <i>e</i>   | 1.28 ± 0.03 GeV                         | $\nu_\mu$  | 0               | < 2 eV           |
|                     | <i>s</i> | -1/3 <i>e</i>   | 95 ± 5 GeV                              | $\mu$      | -1 <i>e</i>     | 105.7 MeV        |
| 3 <sup>rd</sup>     | <i>t</i> | +2/3 <i>e</i>   | 173.2 ± 1.2 GeV                         | $\nu_\tau$ | 0               | < 2 eV           |
|                     | <i>b</i> | -1/3 <i>e</i>   | 4.18 ± 0.03 GeV                         | $\tau$     | -1 <i>e</i>     | 1776.8 ± 0.2 MeV |

Table 2.1.: Properties of fermions in the SM. Quarks and leptons are grouped in three generations. Masses<sup>1</sup> are taken from Ref. [16].

a half-integer spin, called fermions. They are divided into quarks and leptons, where both are arranged in three generations of doublets. The six different quark flavours are grouped in *up-type* (up, charm and top) and *down-type* quarks (down, strange and bottom):

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

The quark masses within a doublet increase from the first to the third generation. All quarks carry a colour charge which allows them participating in the strong interaction. In addition, all quarks contain a weak charge and an electrical charge depending on their type. Up-type quarks have a positive charge of 2/3 of the elementary charge (*e*), down-type quarks have a negative charge of -1/3*e*. Thus, quarks can participate in all three fundamental interactions.

The three generations of leptons also appear in different flavours: electron, muon or tau flavour. Each doublet consists of a charged lepton and a corresponding neutral neutrino:

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

Due to missing colour and electrical charge, neutrinos can only interact via the weak force, the electrically charged leptons can also interact electromagnetically.

All fermions are summarised in Tab. 2.1. For each fermion, *f*, there is an anti-particle,  $\bar{f}$ , which carries the opposite quantum-numbers, but has the same mass.

The dynamics of particles are described by the SM Lagrangian, which is invariant under local gauge transformation of the combined symmetry group,  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ . This symmetry group has  $8 + 3 + 1 = 12$  generators, where each is associated to a gauge boson. Bosons are particles with an integer spin. Gauge bosons are the force carriers of the fundamental interactions. A conserved quantum number can be assigned to each of the three interactions, these are introduced in the following.

---

<sup>1</sup>throughout this document, natural units are used, *i.e.*  $c = 1$ .

| Bosons   |                                |                 |        |                             |            |
|----------|--------------------------------|-----------------|--------|-----------------------------|------------|
| Spin 1   | Mass                           | Interaction     | Spin 0 | Mass                        | couples to |
| $\gamma$ | 0                              | electromagnetic |        |                             |            |
| $Z^0$    | $91.188 \pm 0.002 \text{ GeV}$ | weak            | $H^0$  | $125.7 \pm 0.4 \text{ GeV}$ | mass       |
| $W^\pm$  | $80.385 \pm 0.015 \text{ GeV}$ |                 |        |                             |            |
| $g(8)$   | 0                              | strong          |        |                             |            |

Table 2.2.: Properties of bosons in the SM. The mediators of the strong and electroweak interaction are Spin 1 particles, the Higgs boson is a Spin 0 particle. Masses are taken from Ref. [16].

The theory of Quantum Chromo Dynamics (QCD) is based on the  $SU(3)_C$  symmetry group and describes the **strong interaction**. The generators of this group correspond to eight massless gluons,  $g$ . Within QCD, the conserved quantum number is the colour charge ( $C$ ). There are three different colour charges existing: red, green, blue (and the corresponding anti-colours). Only quarks and the mediators of the strong interaction, the gluons, carry a colour charge. As a result, not only quark-gluon but also gluon-gluon self-interactions are possible. The strength of the QCD coupling,  $\alpha_s$ , is not constant but has a strong dependence on the momentum transfer within an interaction. This leads to two important effects. At low energies  $\alpha_s$  increases dramatically and leads to an effect called *confinement*. This phenomenon describes the fact that colour charges cannot be isolated. Towards large energies,  $\alpha_s$  is very small and leads to an *asymptotic freedom* of the coloured particles. As a result, quarks are quasi-free while they keep only small distances. On the other hand, the confinement implies that only colour-neutral objects can exist. Hadrons, strongly interacting composite particles, are either made of three quarks (baryons) or quark-antiquark pairs (mesons). Recent measurements [17] also confirm the existence of colour-neutral four-quark states. However, it is debated if the four-quark state is a new type of hadron, a so-called tetra-quark, or a molecule-like two-meson system. Another consequence of the self-interacting nature of the strong force is that its effective range is very limited.

The  $SU(2)_L \otimes U(1)_Y$  symmetry group represents the unified electroweak sector of the SM. Gauge theory predicts that a triplet of massless bosons,  $W_\mu$ , is associated to the  $SU(2)_L$  group of the **weak interaction**. The conserved quantum number is the weak isospin  $T$ . Another massless boson appears as a singlet,  $B_\mu$ , which corresponds to the  $U(1)_Y$  group of the **electromagnetic interaction**. The conserved quantum number is the hypercharge  $Y$ . It is constructed by  $Y = Q + T_3$ , where  $Q$  is the electric charge and  $T_3$  is the third component of the weak isospin. The  $W_\mu$  bosons can only couple to left-handed particles (or right-handed anti-particle) with weak isospin. Thus, the theory is called a chiral theory and the symmetry group has a index  $L$  for left-handed. As all particles carry hypercharge, the  $B_\mu$  boson can couple to all particles.

The described symmetry of the electroweak interaction is spontaneously broken by the Higgs mechanism, as explained in the next paragraph. As a result, the physical force carriers of the weak and the electromagnetic interaction are linear combination of the previous  $W_\mu$  and  $B_\mu$  fields. In addition, the formerly massless bosons acquire a mass

through the coupling to the Higgs Boson. The weak interaction distinguishes between charged and neutral currents. The massive  $W^+$  and  $W^-$  bosons ( $\approx 80$  GeV) are the mediators of the charged currents. Both are a combination of the  $W_\mu^1$  and  $W_\mu^2$  bosons. Thus, also  $W^\pm$  couple only to left-handed particles. The mediator of the neutral current is the  $Z^0$  boson ( $\approx 91$  GeV), which is a combination of the  $W_\mu^3$  and  $B_\mu$ . It couples with different strength to left- and right-handed particles. Due to the high mass of  $W$  and  $Z$  bosons the weak interaction is only short-ranged.

The electromagnetic mediator is the massless photon,  $\gamma$ . It is also a combination of the  $W_\mu^3$  and  $B_\mu$  bosons but it is orthogonal to the  $Z^0$ . The photon couples to the electric charge of particle but does not carry charge itself and thus it does not self-interact. As the photon remains massless, the range of the electromagnetic interaction is infinite.

According to gauge symmetry, the fundamental particles described by the SM have to be massless. Mass terms, such as for massive  $W$  and  $Z$  bosons, would violate the local gauge invariance of the theory. The above quoted **Higgs mechanism** solves this conflict by introducing an additional doublet of complex scalar (spin 0) fields. A combination of these fields builds the *Higgs potential* which has a non-zero ground state and spontaneously breaks the electroweak symmetry. This leads to the massive bosons of the weak interaction but also predicts another massive spin-0 particle, the *Higgs boson*. It was the last missing particle of the SM until its discovery in July 2012 at the LHC. Further, also fermion masses (c.f. Tab. 2.1) are generated by Yukawa interactions which describe the coupling between fermion fields and the Higgs potential. The boson masses and their properties are listed in Tab. 2.2.

## 2.2. Particle Production in $pp$ collisions at the LHC

At the LHC, high-energy protons are collided at a centre-of-mass energy of up to  $\sqrt{s} = 8$  TeV. At these energies, the interaction between the protons becomes very complex. If the momentum transfer between the protons is small, the collision can be described by scattering of two compound objects. Towards large momentum transfer, the proton breaks up and the single quarks and gluons, collectively referred to as partons, participate in the collision.

The proton consists of three valence quarks ( $uud$ ), sea quarks and gluons. The momentum of the proton is shared among these partons and can be described by parton density functions (PDFs). In Fig. 2.1 a) the proton PDFs for an energy scale of  $Q^2 = 10$  GeV<sup>2</sup> is displayed as an example. The distributions show the probability density to find a certain parton with a momentum fraction  $x$  of the proton. At high energies, the interaction is not only sensitive to the valence quarks but also to gluons and to sea quarks that are created in quark/antiquark pairs from the vacuum. Each of the partons in the two protons can interact in a collision. This results either in a quark-quark, gluon-gluon or quark-gluon interaction, mediated by the strong interaction (QCD). In hadronic processes two different classes of interactions are distinguished, *hard* and *soft* processes.

**Hard scattering processes** involve either one or two partons which carry a large momentum fraction. The resulting large momentum transfer in the interaction is of

the order of the "hard" energy scale, which is typically  $\gtrsim 1 \text{ GeV}^2$ . Hard QCD processes can be calculated using perturbation theory. Production rates and event properties are accurately predicted. The large amount of energy which is available in a hard scattering process allows the creation of heavy particles. These can be heavy bosons, *e.g.*  $W^\pm$  or  $H^0$  bosons, leptons, *e.g.*  $\tau$ , or heavy quarks, such as bottom or top quarks, which are produced either alone or in pairs. The probability to produce heavy objects depends on several parameters. The most important are the mass of the produced particles, the available energy in the collision and the colliding particle species which defines the possible production channels.

For proton-(anti)proton collisions the production cross section for specific particles or processes is visualised in Fig. 2.1 b). The total  $pp$  cross section at the LHC ( $\sqrt{s} = 7 \text{ TeV}$ ) has been measured to be approximately 100 mb [20]. Cross sections for different processes span over many orders of magnitudes. As an example, the cross section to produce  $b\bar{b}$  pairs is only two orders of magnitude smaller than the total one. However, the cross section for Higgs production is another nine orders of magnitude smaller.

In experimental searches, very rare processes are contaminated by the second class of interactions that are called **soft processes**. These processes are characterised by the energy scale of the size of a hadron ( $\approx 1 \text{ fm}$ ). In contrast to parton scattering at large momentum transfer, soft processes typically have a small momentum transfer of the order of a few hundred  $\text{MeV}^2$ . This regime cannot be accessed by perturbative calculations due to the rising strong coupling constant. Soft processes also include

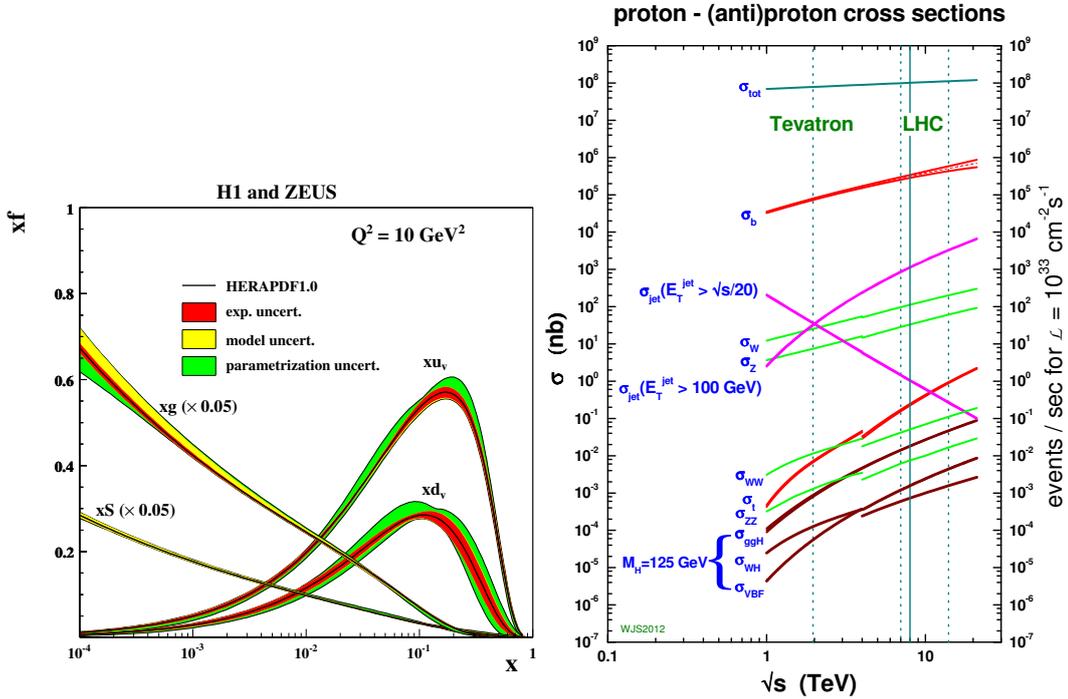


Figure 2.1.: (a) Proton PDFs from HERAPDF2.0 at  $Q^2 = 10 \text{ GeV}^2$ . The PDFs for valence quarks ( $u$ ,  $d$ ), sea quarks ( $s$ ) and gluons ( $g$ ) are displayed. Figure taken from Ref. [18]. (b) Proton-(anti)proton cross sections as function of  $\sqrt{s}$ . The kink represent the transition from  $p\bar{p}$  to  $pp$  collisions. Figure taken from Ref. [19]

the phenomenon of diffractive dissociation in hadron-hadron collisions. [21] Diffraction describes a high-energy collision, in which no quantum numbers are exchanged between the colliding particles. Classically, diffraction is considered as being a pure soft process. However, also diffractive processes that have both soft and hard components are observed. More details about diffraction is given in the subsequent section. The understanding of soft processes, so far, relies only on phenomenological descriptions. Predictions are less precise, in particular when going to higher, unexplored collider energies. However, the modelling of soft processes is important for many analyses. Signatures of exclusive hard processes are polluted by the background from soft particle production. The phenomenological approaches adopted in simulations are further discussed in Sect. 2.3.

### 2.2.1. Anatomy of a $pp$ collision

Proton-proton collisions are often classified by the incoming and outgoing particles in a collision. In *elastic scattering*, both incoming protons come out of the collision unscathed ( $p+p \rightarrow p+p$ ). The contrary *inelastic* interactions are further distinguished as follows. If only one of the incident protons remains intact but the second proton brakes up to create new particles ( $p+p \rightarrow p+X$ ) the collision is defined as *single-diffractive*. It is important that no quantum numbers are exchanged between the protons and thus the quantum numbers of the produced final-state particles are the same as the ones of the original proton. Accordingly, if both protons separately give rise to a bunch of new particles the system is called *double-diffractive* ( $p+p \rightarrow X+X$ ). Again, no exchange of quantum numbers between both incident particles is allowed. Furthermore, there are *non-diffractive* collisions ( $p+p \rightarrow X$ ) where non of the protons survive and quantum numbers are exchanged. Apart from the previous categories there are *central exclusive processes* (see *e.g.* Ref. [22]) of the type  $p+p \rightarrow p+X+p$ . In these diffractive processes, a colour-neutral object, *e.g.* a  $J/\psi$  or  $H^0$ , is created even though no exchange of quantum numbers between the initial protons occurred.<sup>2</sup>

From the experimental point of view, the identification of diffractive processes is very difficult. Certain signatures (*e.g.* rapidity gaps [23]) can be exploited to obtain enriched samples of (non-)diffractive events. However, in the course of this thesis all event samples are treated as inclusive inelastic interactions. Elastic interactions are not considered as they cannot be measured with the LHCb detector.

A typical inelastic  $pp$  collision at high energies is sketched in Fig. 2.2. There are two incoming protons, their partons interact with each other by strong interaction. A part of the original proton usually survives the collision, this so-called *beam remnant* is only slightly deflected and travels close along its original flight path. A hard parton interaction can produce a heavy resonance, such as the  $Z$  boson in this example. Hard processes are accompanied by **initial-state radiation** (ISR) of photons or soft gluons. In analogy there is also **final-state radiation** (FSR) where produced partons further emit bremsstrahlung or gluon-strahlung. In addition, there are soft interactions between the incoming partons which also produce quarks or gluons resulting in multiple-particle production of light hadrons.

---

<sup>2</sup>Since the protons can in turn dissociate or not, one has to distinguish between elastic and inelastic central exclusive processes.

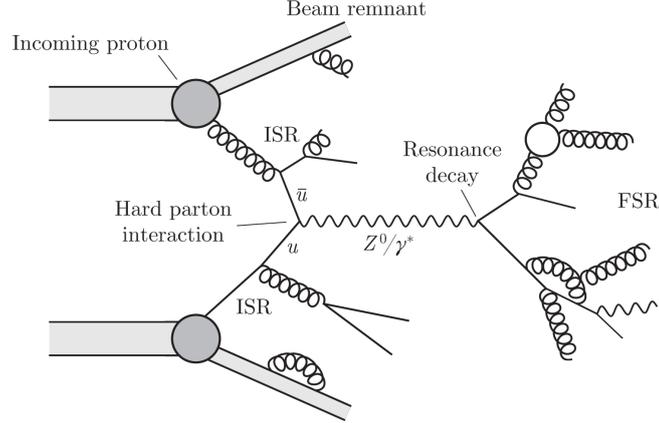


Figure 2.2.: Illustration of an inelastic  $pp$  collision at the LHC. The various interactions are discussed within the text. Figure taken from Ref. [24].

The partons produced by any of these processes must all form colour-neutral mesons or baryons. This procedure is called **hadronisation**. The formation of hadrons is driven by the confinement implied by the strong interaction. The low momentum transfer between the quarks and the resulting large coupling constant does not permit perturbative calculations. Thus, hadronisation in simulations is realised by phenomenological models.

The soft component of a collision is often referred to as the *underlying event* (UE). It comprises all created particles except the hard scatter products. The large number of final-state particles produced in a collision is primarily due to soft processes. Consequently, a reliable description of the underlying event is crucial for most of the data analyses.

### 2.2.2. Production of $b$ hadrons at the LHC

In this thesis, data recorded by the LHCb detector are used. The focus on hadrons containing  $b$  quarks is one of the key aspects that motivated the design of the detector, as it is later discussed in Chap. 4. The production of  $b$  quarks and the resulting kinematic implications are briefly discussed in the following.

At a  $pp$  collider, the leading-order production mechanisms for  $b\bar{b}$  pairs are quark-antiquark annihilation and gluon fusion. Two Feynman diagrams showing these processes are given as an example in Figs. 2.3 a,b). For centre-of-mass energies at the TeV scale, next-to-leading order (NLO) contributions, such as flavour excitations and gluon splitting, become more and more important. At  $\sqrt{s} = 7$  TeV, the dominant process [25] is flavour excitations, the corresponding Feynman diagrams are depicted in Figs. 2.3 c,d).

The mass of a  $b$  quark ( $\approx 4.2$  GeV) is relatively small compared to the energy which is available in  $pp$  collisions at the TeV scale. Thus, there is no need to have two high energetic partons to produce a light  $b\bar{b}$  pair. Looking at the parton density functions of the proton (c.f. Fig. 2.1 a) one finds that towards lower momentum-fractions (Bjorken- $x$ ) the gluon density is dominating. Statistically, it is unlikely to find two gluons (or  $q\bar{q}$  pairs) with the same momentum fraction. An imbalance of the two interacting partons is more likely, which results in a boost of the created  $b\bar{b}$  pair. The boost is directed along the flight direction of one of the colliding protons.

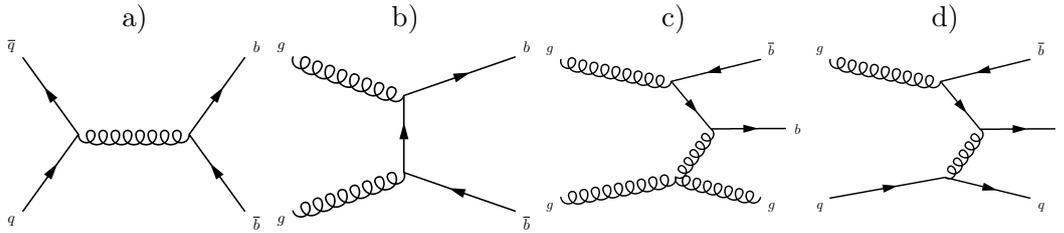


Figure 2.3.: Feynman diagrams for the production of  $b\bar{b}$  pairs in  $pp$  collisions at the LHC. Examples are given for leading-order: quark-antiquark annihilation (a) and gluon fusion (b); next-to-leading-order: flavour excitation (c-d). Figures taken from Ref. [26].

Theoretical calculations [27], including NLO contributions, of the  $b\bar{b}$  cross section at  $\sqrt{s} = 7 \text{ TeV}$  predict a value of  $\sigma_{b\bar{b}} \approx 300 \mu\text{b}$ . The total  $b\bar{b}$  cross section has also been measured by LHCb [28] to be

$$\sigma_{pp \rightarrow b\bar{b}X} = (284 \pm 20 \pm 49) \mu\text{b} , \quad (2.1)$$

being in good agreement with the prediction.

## 2.3. Monte Carlo event generators

This section is dedicated to the simulation of  $pp$  collisions using Monte Carlo<sup>3</sup> event generators. The basis of an event generator is the statistical sampling of physics processes using random numbers.

The simulation of a  $pp$  collision is a complex task which can be split into separate steps. By using the PYTHIA generator [15] as an example, the single simulation steps are introduced. Further generators are explained with a focus on different hadronisation models.

### 2.3.1. The PYTHIA generator

PYTHIA is a general purpose event generator to simulate  $pp$ ,  $p\bar{p}$  and  $ee$  collisions. It is extensively used in the high-energy physics community, and also by LHCb. PYTHIA combines both classes of interactions, hard scattering is calculated by using perturbative QCD, soft interactions are described phenomenologically. The transition between both regimes is defined by the momentum transfer in the parton interaction which is a tunable parameter ( $p_{Tmin}$ ). The PYTHIA generator exists in two different implementations which are both used. PYTHIA 6.4 [15] is written in FORTRAN and is no longer developed but still supported. The corresponding C++ implementation is PYTHIA 8.1 [29] and is the recommended version for LHC studies. PYTHIA 8.1 supports additional new features, however, the basic physics content is the same. The following discussion is based on the PYTHIA documentation [15].

---

<sup>3</sup>The name comes from the resemblance of the technique to the act of playing and recording results in a real gambling casino.' [wikipedia]

The total hadron-hadron cross section at an energy scale  $s$  (square of the centre-of-mass energy) is calculated in PYTHIA using a simple formula [30],

$$\sigma_{\text{tot}}^{pp}(s) = X^{pp} \cdot s^{\epsilon} + Y^{pp} \cdot s^{-\eta}. \quad (2.2)$$

The first term is related to a pomeron and the second to a reggeon exchange. Pomerons and reggeons are hypothetical particles postulated in the framework of Regge theory [31], which is a concept to describe proton-proton scattering. Both particles carry the quantum numbers of the vacuum. The pomeron can be interpreted as a colour-neutral combination of gluons. A review about Regge theory can be found *e.g.* in Ref [32].

In the phenomenological parametrisation of Eq. 2.2, the constants  $\epsilon = 0.0808$  and  $\eta = 0.4525$  are assumed to be universal, the parameters  $X^{pp} = 21.70$  and  $Y^{pp} = 56.08$  are obtained by fitting data, see Fig. 2.4.

The obtained total cross section is subdivided into four terms which are treated individually in PYTHIA:

$$\sigma_{\text{tot}} = \sigma_{\text{EL}} + \sigma_{\text{SD}} + \sigma_{\text{DD}} + \sigma_{\text{ND}}. \quad (2.3)$$

Following the definition of Sect. 2.2.1, the first term (EL) is related to elastic scattering, followed by single diffractive (SD), double diffractive (DD) and non-diffractive (ND) processes. The elastic component is approximated by the optical theorem,  $\sigma_{\text{EL}} = \sigma_{\text{tot}}^2/16\pi B_{\text{EL}}$ , where the parameter  $B_{\text{EL}}$  depends on the energy  $s$ . The single and double diffractive cross sections are calculated in the Regge framework. The remaining cross section is identified as the non-diffractive  $\sigma_{\text{ND}}$ . For  $pp$  collisions at  $\sqrt{s} = 7$  TeV, the single cross sections are summarised in Tab. 2.3. Inelastic interactions account for around 50% to 80% (including diffraction) of the total proton-proton cross section.

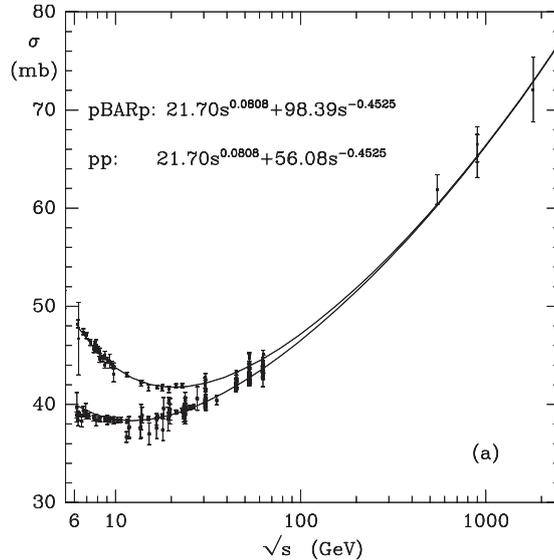


Figure 2.4.: Total cross section for  $pp$  and  $p\bar{p}$  scattering measured as function of centre-of-mass energy. The curves corresponds to a fit using Eq. 2.2. Figure taken from Ref. [30].

| Cross section $\sigma^{pp}$<br>( $\sqrt{s} = 7 \text{ TeV}$ ) | PYTHIA<br>6.426           | PYTHIA<br>8.180            | PHOJET<br>1.12 | HERWIG++ |
|---|---------------------------|----------------------------|----------------|----------|
| Elastic   | 19.4 mb                   |                            |                |          |
| Single diffractive  | $2 \times 6.8 \text{ mb}$ | $2 \times 6.19 \text{ mb}$ | 10.7 mb        |          |
| Double diffractive  | 9.3 mb                    | 8.11 mb                    | 3.9 mb         |          |
| Inelastic (tot-elastic-diff)                                  | 48.4 mb                   | 50.9 mb                    |                |          |
| Inelastic (tot-elastic)                                       | 71.5 mb                   | 71.5 mb                    | 75.3 mb        | 71.0 mb  |
| Total   | 91.1 mb                   |                            |                | 101.9 mb |

Table 2.3.: Cross sections implemented in different Monte Carlo event generators.

The single steps of the simulation of a  $pp$  collisions and the implemented processes are outlined in the following by using an event containing a hard interaction as example:

**Parton interaction** Protons with a defined energy are simulated. Their quark and gluon densities and the corresponding momentum fractions are defined by parton density functions, such as in Fig. 2.1 a). The used PDF sets can be selected in advance, by default CTEQ5L [33] PDFs are implemented in PYTHIA, the standard LHCb configuration relies on CTEQ6L [34]. Both PDF sets are based on various measurements of high-energy lepton-hadron and hadron-hadron interactions. The respective physics process of the partonic scattering is sampled according to cross sections and by using a random number generator.

**ISR** Before calculating the hard process, partons of both protons can radiate gluons and photons by initial-state-radiation.

**MPI** The two partons involved in the hard interaction produce additional partons according to the generated physics process. PYTHIA also includes the possibility of multi-parton-interactions (MPI) where additional hard parton interactions occur simultaneously in a single  $pp$  collision. The remaining partons that were not involved in the interaction build the beam remnant.

**UE** Along the hard interaction, additional soft and semi-hard parton interactions can take place and build the underlying event.

**FSR / parton shower** All final-state partons can branch and produce additional gluons and photons. Final-state-radiation in PYTHIA is implemented by different models of parton showers. The main difference is the ordering in which partons are allowed to shower. Also the interplay of parton showers and multi-parton-interactions is supported. Parton showering stops when a mass scale of 1 GeV is reached. Obviously, this arbitrary cut-off is sensitive to the number of final-state partons.

**Hadronisation** In the next step, the fragmentation and hadronisation of a partonic final state is approached. The QCD confinement forces single quarks and gluons to

form colourless bound states. The hadronisation process is not understood from first principles and requires the use of phenomenological models.

The PYTHIA generator, also known as *the Lund Monte Carlo*, is a famous representative for the Lund string fragmentation model. The idea of this model can best be explained by a simple  $q\bar{q}$  system where both colour-charged quarks move into different directions. The field between colour charges is responsible for the creation of a *Lund string*. Increasing the distance between the colour charges of both quarks increases the energy density described by the string tension  $\kappa = 1 \text{ GeV/fm}$ . This corresponds to the linear part of the effective quark-antiquark potential

$$V(r) = -\frac{4}{3} \frac{\alpha_s \hbar c}{r} + \kappa \cdot r, \quad (2.4)$$

with  $\alpha_s = \alpha_s(Q^2)$  being the energy dependent strong coupling "constant". The first term is similar to a Coulomb potential and is only important for very short distances ( $r < 0.2 \text{ fm}$ ) and is not considered in the string model. The energy of the stretched colour string can be picked up by vacuum fluctuations and produce another  $q'\bar{q}'$  pair. The original string is then cut into two shorter strings between the colour charges of the new quark pairs,  $q\bar{q}'$  and  $q'\bar{q}$ . The resulting total energy, including the created quark-antiquark pair, is much lower than that of the original long string. This fragmentation is further repeated until the energy is low enough and bound mesons are formed.

Baryons are produced in a related fashion, *e.g.* in the *diquark picture*. Instead of creating a quark-antiquark pair ( $q'\bar{q}'$ ), the string energy is used to produce an antiquark-diquark pair, ( $\bar{q}'q'q'$ ). The created antiquark-diquark pair allows an equivalent colour state to a quark-antiquark pair. The fragmentation process is then carried out identically as for single quark pairs. An original  $q\bar{q}$  pair can then produce two systems containing three quarks ( $qq'q' + \bar{q}'\bar{q}'\bar{q}$ ), where each can potentially be joint to a baryon.

**Particle decay** PYTHIA includes the decay of unstable particles according to detailed lists of particle properties (masses, widths and lifetimes) and decay properties (branching ratios). The decay description comprises decays of hadrons and leptons, as well as resonance decays which are handled directly in the simulation of the hard process. There are interfaces (*e.g.* EVTGEN [35]) which allow to update or modify the decay properties outside of PYTHIA.

### 2.3.2. PYTHIA tunes

The PYTHIA generator comprises a variety of adjustable parameters which define active physics processes and control phenomenological models which are adopted *e.g.* in hadronisation or soft particle production. These parameters are typically tested and optimised using data, the procedure is called *tuning*. Fundamental input to the tuning of these parameters are various soft-QCD measurements, such as charged particle multiplicities, cross section measurements, energy flow or production ratios of different particle species. A set of parameters is typically referred to as a *tune*. In this thesis, several tunes of PYTHIA 6.4 and PYTHIA 8.1 are compared to the measured charged particle multiplicities presented in Part I. The tunes are presented in the following.

Of particular interest is the **LHCb tune** [36]. It contains the parameters used for the default simulation of  $pp$  collisions in the LHCb collaboration. It can be applied to both PYTHIA implementations, PYTHIA 6.4 and 8.1, leading to almost identical results. The tune is based on central-rapidity charged particle multiplicity measurements by the UA5 [37, 38] and CDF [39] experiments in  $pp$  and  $p\bar{p}$  collisions at different centre-of-mass energies. The simulated charged particle multiplicity is sensitive to the previously mentioned  $p_{Tmin}$  cut-off. The evolution of this parameter as a function of the centre-of-mass energy is fitted to the data of the quoted references with the function:

$$p_{Tmin} = p_{Tmin}^{\text{LHC}} \left( \frac{\sqrt{s}}{\sqrt{s_0}} \right)^{2\epsilon} \quad (2.5)$$

where  $p_{Tmin}^{\text{LHC}}$  is the extrapolated  $p_{Tmin}$  at the LHC reference energy of  $\sqrt{s_0} = 14$  TeV. The obtained multiple interaction parameters are  $\text{PARP}(82) = p_{Tmin} = 4.28$  GeV and  $\text{PARP}(90) = 2\epsilon = 0.238$ . All non-default PYTHIA parameters of the LHCb tune are summarised in Tab. 2.4.

For the old PYTHIA version 6.4, two widely-used tunes are selected for the multiplicity comparison. These are the **Perugia 0** and **Perugia NOCR** tunes, both of which rely on the CTEQ5L PDF sets. The models were obtained by using data from the Tevatron and the SPS, comprising hadron-hadron collisions at centre-of-mass energies in the range of 200 – 1960 GeV. Besides several further model parameters, the tuning comprises parton shower and underlying-event models. A focus was set on the energy scaling of charged particle multiplicities. In order to provide a good estimates for LHC energies, the strategy is to obtain a smooth description of the measured multiplicities at different lower centre-of-mass energies.

The Perugia 0 tune can be interpreted as the central parameter set of all Perugia tunes. Perugia NOCR is a special variation where no *colour reconnection* [40] in the parton showering and hadronisation models is allowed. In a  $pp$  collision, many coloured partons are produced by the aforementioned processes, *e.g.* ISR, MPI or parton showers. Partons produced by a single process, *e.g.* a parton shower, are afterwards combined to form colour-neutral hadrons, according to their colour charges. This original colour-structure of a single process is referred to as the dominant structure. In

| Parameter | Value | Parameter | Value | Parameter | Value |
|-----------|-------|-----------|-------|-----------|-------|
| CKIN(41)  | 3.0   | PARP(82)  | 4.28  | PARP(150) | 0.085 |
| MSTP(2)   | 2     | PARP(85)  | 0.33  | PARJ(11)  | 0.5   |
| MSTP(33)  | 3     | MSTJ(26)  | 0     | PARJ(12)  | 0.4   |
| MSTP(81)  | 21    | PARJ(33)  | 0.4   | PARJ(13)  | 0.79  |
| MSTP(82)  | 3     | PARP(86)  | 0.66  | PARJ(14)  | 0.0   |
| MSTP(52)  | 2     | PARP(89)  | 14000 | PARJ(15)  | 0.018 |
| MSTP(51)  | 10042 | PARP(90)  | 0.238 | PARJ(16)  | 0.054 |
| MSTP(142) | 2     | PARP(91)  | 1.0   | PARJ(17)  | 0.131 |
| MSTP(67)  | 1     | PARP(149) | 0.02  |           |       |

Table 2.4.: Non-default PYTHIA parameters used in the LHCb tune [36].

colour-reconnection models, also the colour charges produced by independent processes can be reconnected. This results in a more realistic energy and momentum spectrum of the produced hadrons but is more complex to compute. The Perugia NOCR tune, in which this reconnection is not allowed, emphasises a large baryon-number-transport from the beam particles to the created final-state particles. The ratio of final-state antibaryons to baryons is sensitive to baryon number of the colliding particles and depends on the kinematic range. Details to the Perugia tunes can be found in Ref. [41].

For PYTHIA 8.1, only default configurations are selected when comparing charged particle multiplicity distributions of simulation and data. Two version are selected, namely PYTHIA 8.145 and 8.180. Each employs a different default parameter set which has strong influence on the underlying event scenario or the simulation of minimum-bias<sup>4</sup> event samples. Up to PYTHIA 8.145, there has been no tune available which was optimised on the basis of LHC measurements. However, an improved parton shower model [42] was introduced in this version. The more recent version 8.180 is selected to represent results based on **Tune 4C** [43]. This tune includes LHC measurements of charged particle multiplicities in  $pp$  collisions at  $\sqrt{s} = 0.9, 2.36$  and 7 TeV in the central rapidity region and is still the default choice to date.

### 2.3.3. The PHOJET generator

The PHOJET event generator [44, 45] is based on a two-component approach which also divides interactions into a soft and a hard component. The focus of PHOJET is on soft interactions which are described phenomenologically by an implementation of the dual-parton model [46], based on the Regge theory. Soft interactions, including diffraction, are modelled by Pomeron exchanges which also allow multi-parton interactions. A major difference to the PYTHIA 6.4 generator is the inclusion of central diffractive processes, which are described by a double pomeron exchange.<sup>5</sup>

The hard interaction is calculated in the same way as in PYTHIA using perturbative QCD calculations. A similar cut-off parameter defines the scale of hard interactions. In the dual-parton model the exchanged pomeron can be either soft or hard and thus allows a smooth transition of both regimes. Multiple soft and hard interactions can occur in the same event.

PHOJET includes initial- and final-state radiation and also uses the same string fragmentation model as PYTHIA. The generator provides simulations of  $pp$ ,  $p\gamma$  and  $\gamma\gamma$  interactions. In this thesis, PHOJET in the last version 1.12 is used and the default options are applied. The development of this generator is no longer continued.

### 2.3.4. The HERWIG++ generator

HERWIG++ [48] is a recent general purpose event generator to simulate hadron-hadron, lepton-lepton and hadron-lepton collisions. It is specialised to provide a variety of hard

<sup>4</sup> Ideally, a no-bias event sample represents an unbiased selection of events, recorded with an totally inclusive trigger. Experimentally, an event sample is always biased by the trigger. The term minimum-bias implies an event selection that introduces the smallest possible bias.

<sup>5</sup>The more recent PYTHIA 8.1 generator also includes central-diffraction, see Ref. [47].

scattering processes and the simulation of *Beyond Standard Model* physics processes. Recent development has been done to obtain a good description of the underlying event in hadron-hadron collisions. This involves the implementation of new models to describe multi-parton interactions [49] which have two tunable parameters,  $p_{Tmin}$  and  $\mu^2$ . The first parameter defines the minimum transverse momentum needed for additional hard scatters, the second parameter,  $\mu$ , is an effective inverse proton radius which describes the spatial overlap of the incoming hadrons. The MPI model has been tested to successfully reproduce Tevatron data.

The underlying event activity is further modelled within the parton shower approach to simulated initial- and final-state QCD radiation<sup>6</sup>. Parton showers are implemented differently compared to PYTHIA, a *coherent branching algorithm* [50] is adopted to account for soft gluon interference. The branching of partons is carried out until there is no phase space left for further emissions. At this point perturbative calculations are no longer possible and hadronisation effects start to dominate.

Another major difference of HERWIG++ compared to PYTHIA is a different phenomenological description of the hadronisation. The model implemented in HERWIG++ is the cluster hadronisation model [48], based on the idea of colour preconfinement [51]. The description of the formation of hadrons has a crucial impact on the multiplicity of final-state particles which are visible in the detector. The idea of cluster hadronisation is outlined in the following.

Starting point of the cluster model is the final evolution step of the parton shower. The information obtained from the perturbative phase is carried forward as far as possible. All quarks and gluons from the previous parton evolution are put onto their constituent mass-shells. With the preconfinement condition, the quarks tend to form colour singlets (colour neutral states) if they are close to each other in momentum space. The remaining gluons are split non-perturbatively to quark-antiquark ( $g \rightarrow q\bar{q}$ ) or diquark-antidiquark ( $g \rightarrow qq\bar{q}\bar{q}$ ) pairs. Depending on the probability given by the available phase space, the gluon can decay into any quark flavour. The colour singlet combinations of  $q\bar{q}$  pairs in the event are assumed to form so-called *clusters*, which can be interpreted as highly excited pre-hadronic states. In a simple picture, the formation of a cluster is shown in Fig. 2.5 a). Clusters that are too heavy lead to non-physical observations and thus can further split into lighter clusters and produce additional  $q\bar{q}$  pairs. In the last step, each cluster isotropically decays in its rest frame into a pair of hadrons. Assuming a cluster of a given flavour ( $q_1\bar{q}_2$ ), a quark-antiquark pair ( $q\bar{q}$ ) is extracted from the vacuum and a pair of mesons is formed ( $q_1\bar{q}, \bar{q}_2q$ ). In analogy, baryons are formed by creating a diquark-antidiquark pair from the vacuum.

The hadronisation model in combination with the implemented MPI model show good results in describing Tevatron data on the UE. However, the description of LHC measurements was not satisfying and lead to the implementation of colour reconnection [52] (CR) in the MPI model. The basic idea is to improve the colour structure by assigning colour connections to jets which are produced nearby in momentum space but originate from separate hard scatters. The formation of colour-singlet states was previously attempted by following the dominant colour structure of the event, defined

<sup>6</sup>According to the documentation, special emphasis is put on the correct description of radiation from heavy particles which could be of particular interest for future LHCb analysis.

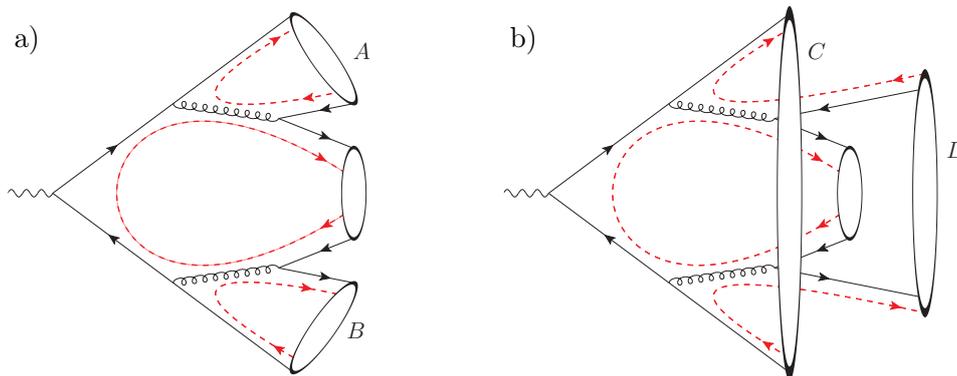


Figure 2.5.: Cluster formation implemented in HERWIG++. a) Clusters (indicated as ovals) are formed according to the dominant colour structure (colour lines shown as red dotted lines). As a result, clusters A and B are formed. b) Rearranged clusters C and D are formed by reconnecting the colour lines of partons from the original clusters A and B. Figures taken from Ref. [52].

by the separate processes (c.f. Fig. 2.5 a)). The CR model allows the reconnection of colour lines in a different cluster configuration, see Fig. 2.5 b), which results in a lighter invariant mass of the clusters. With the inclusion of CR in the MPI model, HERWIG++ allows a proper description of non-diffractive observables from LHC data measured in central rapidity. The forward region has not been tested so far.

Two versions of the HERWIG++ event generator, version 2.6.3 [53] and 2.7.0 [54], are used later for a comparison of predictions of charged particle multiplicities to the measured data. Both versions are operated in the minimum bias configuration which implements a respective underlying-event tune [55]. Correspondent to the name, these tunes are optimised to describe the underlying event in hadron collisions and accordingly minimum bias event samples. The tunes comprise a set of parameters used to configure the phenomenological MPI and CR models.

In HERWIG++ version 2.6.3 the default underlying-event tune is UE-EE-4-MRST (based on Ref. [52]), while version 2.7.0 relies on tune UE-EE-5-MRST [56]. Both tunes are based on Tevatron and LHC measurements on charged particle multiplicities in the central rapidity region. They are utilising the MRST LO\*\* [57] PDF set, in contrast to the CTEQ PDFs used by the selected PYTHIA tunes.



---

## The Large Hadron Collider at CERN

---

The *Large Hadron Collider* (LHC) is, at present, the worlds most powerful particle accelerator. It was planned and build as a common project of more than 85 nations all around the world and is located at CERN, the *European Organization for Nuclear Research*<sup>1</sup>, close to Geneva in Switzerland. It is installed in the 26.7km long tunnel that has originally been build for the LEP (*Large Electron Positron*) collider which was in operation between 1989 to 2000. Figure 3.1 shows the tunnel being located below the Swiss and French border in a depth of 50 to 175 m.

The LHC is designed as a high-luminosity proton-proton ( $pp$ ) collider but can also be operated to accelerate and collide heavy ions. It consists out of two beam lines surrounded by powerful superconducting magnets which keep charged particles on their orbit. The beams are brought to collision at four intersection points, where the main LHC

---

<sup>1</sup>The abbreviation originates from the French name, *Conseil Européen pour la Recherche Nucléaire*

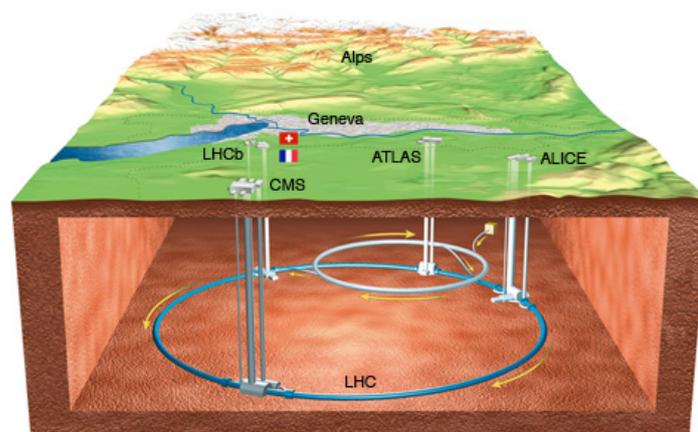


Figure 3.1.: Schematic overview of the Large Hadron Collider and the position of the four major LHC experiments. Figure taken from Ref. [58]

experiments are located. Further details about the LHC, a comprehensive description of the machine can be found in Ref. [59].

### 3.1. The LHC experiments



ATLAS



ALICE



In total there are seven experiments attached to the LHC, each of the four collision points accommodates one of the four big experiments ATLAS, CMS, ALICE, and LHCb.

ATLAS [60] and CMS [61], **A Toroidal LHC ApparatuS** and the **Compact Muon Solenoid**, are both multi-purpose detectors. Their main physics goals are to perform direct searches of new particles, such as candidates for dark matter or supersymmetry, and to look for the missing piece of the Standard Model, the Higgs boson. Both experiments successfully discovered the Higgs boson in July 2012 [4, 5].

The ALICE detector [62], **A Large Ion Colliding Experiment**, is specialised for heavy-ion collisions. The physics program focusses on the physics of strong interactions and the properties of a new phase of matter, the *quark-gluon-plasma* (QGP). This phase emerges at extremely high densities and temperatures, which can be achieved in high energetic collisions of heavy ions.

The **L**arge **H**adron **C**ollider **b**eauty experiment, LHCb [63], is a specialised experiment dedicated to study CP violation and rare decays in the system of *b*-hadrons. The detector has proven to be suitable to successfully extend the physics program to a variety of other fields up to also covering proton-ion physics, as it will be presented in the course of this thesis.

There are three additional smaller experiments attached to the LHC. The TOTEM [64] experiment is installed next to the CMS detector. The main goal is to precisely determine the size of protons and to perform cross-section and luminosity measurements in *pp* collisions. LHCf [65] is a very small detector located at the interaction point of ATLAS. It is designed to detect particles that are produced in the extreme forward region of a collision. The results of this experiment can be used to calibrate and simulate cosmic-ray detectors. MoEDAL [66] is the most recent LHC experiment, it shares the cavern of the LHCb experiment. The physics goal is to search for magnetic monopoles.



Figure 3.2.: Picture of the LHC with its superconducting magnets placed in the tunnel of the former LEP collider. Figure taken from Ref. [58]

### 3.2. A multi-purpose collider

For the LHC, three different operation modes are distinguished, which allow colliding either protons, lead-ions, or a mixed setup of protons and lead-ions. The three machine setups are briefly discussed below:

**Proton-proton mode** The design target of the LHC is to collide two proton beams at a maximum centre-of-mass energy of  $\sqrt{s} = 14$  TeV. The corresponding energy of 7 TeV per beam is defined by the peak dipole field of the superconducting magnets which keep the protons on a circular orbit along the collider. A picture of the LHC and its magnets is depicted in Fig. 3.2.

At nominal running conditions, bunches are equally distributed along the LHC, corresponding to a bunch spacing of 25 ns. This results in a maximum frequency of proton-proton collisions of 40 MHz. Due to restrictions of the injection procedure and reserved empty bunches used to operate the beam, 2808 bunches can be populated with protons (*proton bunches*). The nominal peak luminosity of  $L = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  is obtained by filling  $1.1 \times 10^{11}$  protons in each of proton bunches.

Due to an incident during the commissioning of the LHC in 2008, where a large number of the superconducting magnets had been damaged, the LHC started to provide first *pp* collisions only in 2009 and at a lower centre-of-mass energy of 900 GeV. In the years 2010 to 2012 the energy had been increased up to 7 and 8 TeV, leading to a maximum luminosity of  $0.76 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  at a total number of 1374 proton bunches per beam and a 50 ns bunch spacing. In addition, short periods of *pp* collisions at  $\sqrt{s} = 1.36$  TeV and 2.76 TeV have been provided.

After the current ongoing upgrade of the LHC magnets the energy will be further increase towards the nominal values of 13 or 14 TeV.

**Heavy-ion mode** In addition to protons, the design of the machine also allows to accelerate and collide heavy ions. For this purpose, fully stripped lead ions ( $^{208}\text{Pb}^{82+}$ ) are used. In nominal conditions, the 82 protons per nucleus can be accelerated to 7 TeV corresponding to an energy of 2.76 TeV per nucleon. The nucleon-nucleon centre-of-mass

energy of two colliding lead-lead (PbPb) beams then amounts to  $\sqrt{s_{NN}} = 5.52$  TeV. The target peak luminosity in PbPb runs is  $1.0 \times 10^{27} \text{ cm}^{-2} \text{ s}^{-1}$ , obtained with 592 bunches filled with  $7 \times 10^7$  lead ions.

Besides the  $pp$  running periods, also two dedicated heavy-ion runs took place in 2010 and 2011. Due to the mentioned problems with the superconducting magnets also the energy during the heavy-ion runs was decreased to half of the design value leading to centre-of-mass energy of  $\sqrt{s_{NN}} = 2.76$  TeV in the nucleon-nucleon system.

**Proton-ion mode** A third operation mode, which was not foreseen in the original design of the LHC, involves an asymmetric beam configuration of colliding protons and lead ions. The main motivation for adding this type of collision is to provide benchmark measurements for PbPb collisions. The energy of both beams in this configuration is different. While the proton beam had an energy of 4 TeV, the lead beam is accelerated to an energy of 1.58 TeV/nucleon, accounting for the ratio of protons to nucleons ( $Z/A = 82/208$ ) within the ion. The nucleon-nucleon centre-of-mass energy  $\sqrt{s_{NN}}$  of a proton ( $p$ ) in one beam and a nucleon ( $N$ ) in the other beam can be calculated in the lab system as follows:

$$\begin{aligned} s_{NN} &= (E_p + E_N)^2 - (p_{z,p} - p_{z,N})^2, \text{ with } E_N = (Z/A)E_p \\ &= E_p^2 \cdot (1 + Z/A)^2 - p_{z,p}^2 \cdot (1 - Z/A)^2, \quad p_{z,N} \approx p_{z,p} \approx E_p \\ \sqrt{s_{NN}} &= 2E_p \sqrt{Z/A} \end{aligned} \quad (3.1)$$

The resulting energy for proton-ion beams amounts to  $\sqrt{s_{NN}} = 5.02$  TeV.

Since the momenta of the nucleons in the ion- and proton-beam are different, the centre-of-mass system (cms) of the collision is boosted. It follows, that the measured rapidity of particles in the detector's rest frame is shifted by a factor  $\Delta y$  with respect to the rapidity in the cms. This boost acts along the direction of the proton and amount to

$$\Delta y = 1/2 \ln A/Z \approx +0.465. \quad (3.2)$$

The LHCb detector, designed to originally collect proton-proton data only, also participated in the proton-lead program of the LHC. The detector design and its key features are presented in the next chapter.

## The LHCb experiment at the LHC

LHCb is the dedicated experiment for heavy flavour physics at the LHC. Its design is driven by the properties of heavy quark production and their decay characteristics. The geometry of the LHCb detector is different compared to most of other particle detectors.

As discussed previously (c.f. Chap. 2.2.2), the production of boosted  $b\bar{b}$  pairs at the LHC predominantly takes place in the forward and backward direction. Therefore, the LHCb detector is a single-arm magnetic dipole spectrometer in the forward region, the layout is given in Fig. 4.1. The LHCb coordinate system is defined as a right-handed

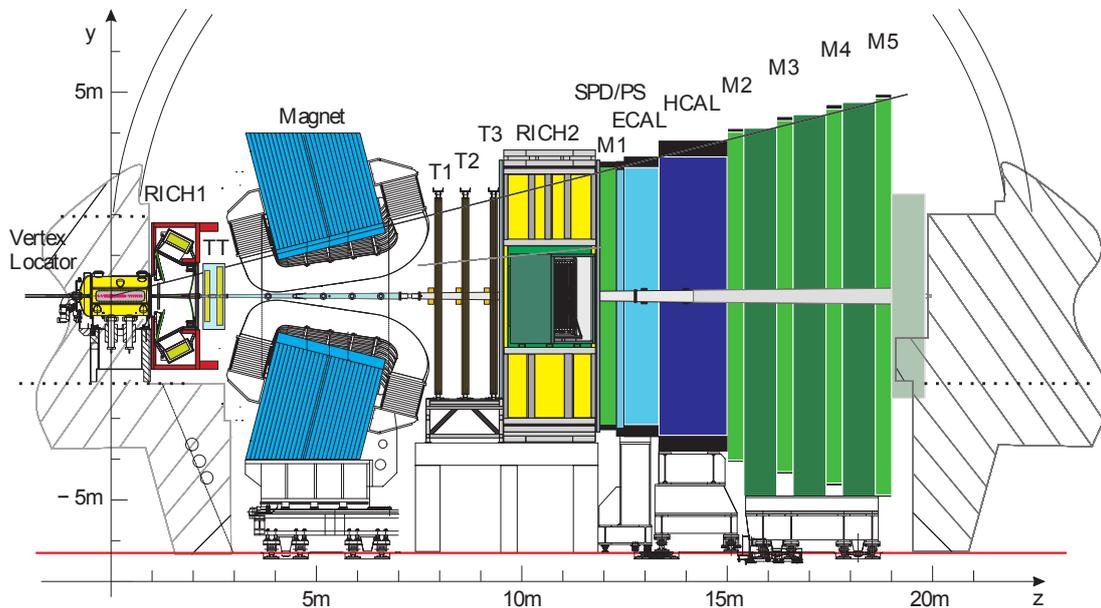


Figure 4.1.: Schematic overview of the LHCb detector in the  $y$ - $z$  plane. The detector's subsystems are described in the text. Figure is taken from Ref. [63].

system with its origin at the nominal interaction point of the colliding beams. The z-axis is aligned along the beam line and pointing towards the magnet, the y-axis is pointing upwards.

The detector acceptance covers angles of approximately 10 to 300 mrad in the bending plane (x-z-plane) and 10 to 250 mrad in the non-bending plane (y-z-plane). Even though the total coverage of the solid angle is small, the forward design allows a large fraction of  $b$ -hadrons to be reconstructed as indicated in Fig. 4.2. The histogram shows the angles between  $b$  and  $\bar{b}$  quarks w.r.t. the z-axis produced at the LHC. The red area indicates the coverage by the LHCb detector in the forward region.

The topology of  $b$ -hadron decays further motivates the key aspects of the detector layout. Even though  $b$ -hadrons have relatively short lifetimes the average boosted flight distances are of the order of centimetres before they decay. This is the result of the boost due to the asymmetric parton collision. In order to resolve the primary interaction point and the decay vertex a high precision tracking device, the Vertex Locator (VELO) surrounds the  $pp$  interaction region.

The VELO is only one component of the LHCb tracking system. It further comprises the Tracker Turicensis (TT) located in front of the dipole magnet, and the main tracking stations (T-stations) located behind. Each station, T1, T2 and T3 consist of two independent sub-systems. The central part around the beam axis is covered by the Inner Tracker (IT) while the outer region is covered by the Outer Tracker (OT).

Another group of sub-detectors is related to particle identification. LHCb has two Ring Imaging Cherenkov Detectors, RICH1 and RICH2, which allow the distinction between different hadron species. RICH1 is located behind the VELO, RICH2 is found behind the T-stations. RICH2 is followed by the calorimetry system. Its major constituents are the electromagnetic (ECAL) and the hadronic (HCAL) calorimeters

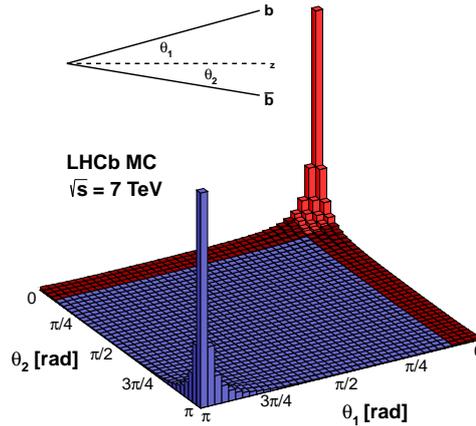


Figure 4.2.: The histogram shows the distribution of the angle between the beam axis ( $z$ ) and  $b$  and  $\bar{b}$  quarks produced in  $pp$  collisions at  $\sqrt{s} = 7$  TeV using a PYTHIA simulation. The red area indicates the LHCb acceptance. Figure is taken from Ref. [67].

which are used to determine the energy of particles by absorption. They are assisted by a Scintillation Pad Detector (SPD) and a Pre-Shower (PS) detector, both are primarily used for particle identification. The muon system consists of five stations, M1 is placed in front of the calorimeters, M2-M5 are placed behind.

The single detector components and the layout of the beam pipe within the range of the detector are presented in the following sections. A more detailed description can be found in Ref. [63,68]. Towards the end of this chapter, a brief overview of the trigger system and the performance of the proton-proton and proton-ion data taking periods are given.

## 4.1. Tracking system

The tracking system is designed to detect signatures of charged particles traversing the detector. The single detector *hits* are used to reconstruct the trajectories (*tracks*) of the original particles and to determine their momenta from the deflection by the magnetic field. Charged particles can be reconstructed in the VELO and TT, both located in front of the magnet, and in the T-stations behind the magnet. Combining these information gives access to the momentum of the respective particle.

### 4.1.1. The magnet

The LHCb experiment utilises a warm dipole magnet [69] to deflect charged particles traversing the magnetic field. The magnet is centred around 5 m downstream of the nominal interaction point and has an integrated magnetic field of  $\int B dl = 4 \text{ Tm}$ . With the bending power of the magnet, a momentum resolution  $\Delta p/p$  of 0.4% at 2 GeV and 0.6% at 100 GeV is achieved for particles traversing the entire tracking system.

The main magnetic field component,  $B_y$ , is pointing along the  $y$ -axis and thus, deflects

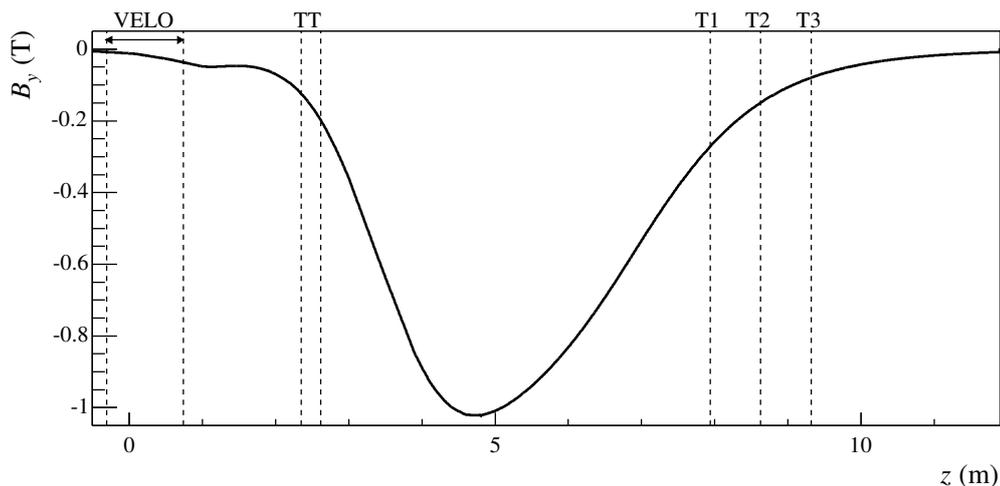


Figure 4.3.: The main component of the magnetic field strength as function of  $z$ . The position of the tracking detectors is indicated by dashed lines. Figure taken from Ref. [70].

charged particles in the horizontal  $x$ - $z$  plane. The shape and field strength along the  $z$ -axis are illustrated in Fig. 4.3. It should be noted that the tracking detectors are located at positions where the influence of the magnetic field is small.

As particles with a given charge are mostly deflected onto one side of the detector, the magnetic field polarity is reversed regularly. This allows studying systematic effects which are related to potential asymmetries of the detector.

#### 4.1.2. Vertex Locator

The Vertex Locator (VELO) [71, 72] is the sub-detector located closest to the primary interaction region. It provides the first measurements of charged particles' trajectories and can be used independently in the track reconstruction. The VELO is designed to precisely measure the position of the primary interaction vertex (PV) and to distinguish it from any displaced vertex. These can be secondary decay vertices or other primary vertices from additional  $pp$  collisions within the same event, so-called *pile-up* interactions.

As the Vertex Locator is surrounding the interaction region with the highest particles flux, radiation tolerant silicon sensors are adopted. These sensors employ a half-disc shape made of silicon microstrips. Two different types of sensors are installed which are designed to measure  $r$ - and  $\phi$ -coordinates of a particle traversing the sensor. The layout for both is given in Fig. 4.4. The  $r$ -type sensors use strips arranged in concentric semi-circles. In order to minimise the occupancy, each strip is subdivided into  $45^\circ$  segments. The strip pitch decreases linearly from  $102\ \mu\text{m}$  at the outer radius to  $38\ \mu\text{m}$  at the innermost radius. The  $\phi$ -type sensors employ an approximately radial strip design which is grouped in two regions of different pitch size. This prevents the sensors from too large occupancies and too large pitch sizes at the edge region.

The VELO consists of 21 stations where each has a detector module left and right of the beam axis. The modules itself comprise one  $r$ - and one  $\phi$ -sensor. On both sides the modules are slightly displaced in order to have a small overlap between the sensors. A schematic overview of the VELO layout is given in Fig. 4.5. Around the nominal interaction region, the modules are spaced more closely compared to the downstream

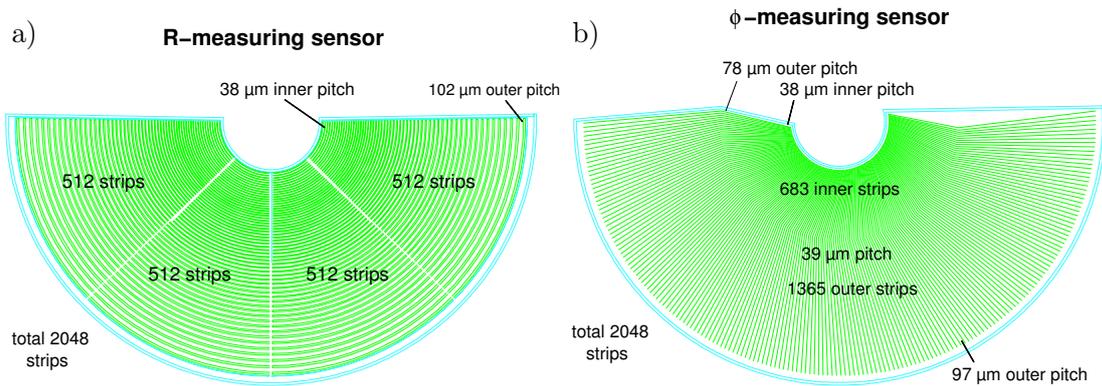


Figure 4.4.: Sketch of the VELO  $r$ - (a) and  $\phi$ - (b) sensors. Note the increasing pitch size when going towards greater radii, and the kink in the strips of the  $\phi$ -sensor. Figures are taken from Ref. [67].

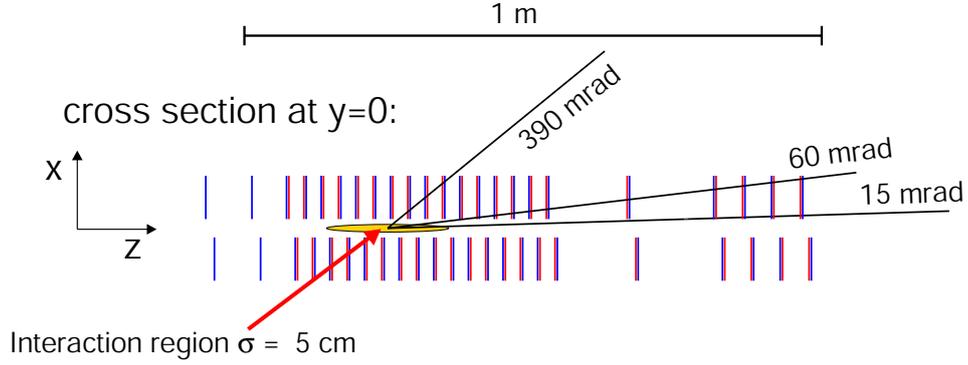


Figure 4.5.: Layout of VELO modules along the  $z$ -axis in top view.  $r$ -sensors are marked in blue,  $\phi$ -sensors in red. Figure is taken from Ref. [67].

edge area. In addition to the regular modules, there are two pile-up stations using  $r$ -sensors only. These are designed for vetoing pile-up events and are located upstream the interaction point.

In order to minimise the amount of material that is responsible for multiple scattering of particles, the beam pipe is removed within the VELO and all modules are mounted in a vacuum vessel. A thin RF foil, made of an aluminium alloy, is used to separate the beam and the sensor discs. This is important for two reasons. The ultra-high vacuum of the LHC beam is protected from out-gassing of the VELO sensors, and further, the sensors are shielded from electromagnetic effects induced by the LHC beam. The RF foil accounts for around 40% of the material budget of the VELO. Due to multiple scattering, the foil has a significant impact on the track reconstruction and also on measurements of prompt particle production. Figure 4.6 a) shows the average radiation length seen by particles passing through the VELO. Around azimuthal angles of  $|\phi| = [60, 120]^\circ$ , the material of the overlapping VELO discs and the RF foil are visible. Figure 4.6 b) shows the distribution of reconstructed vertices of hadronic interactions within the VELO.

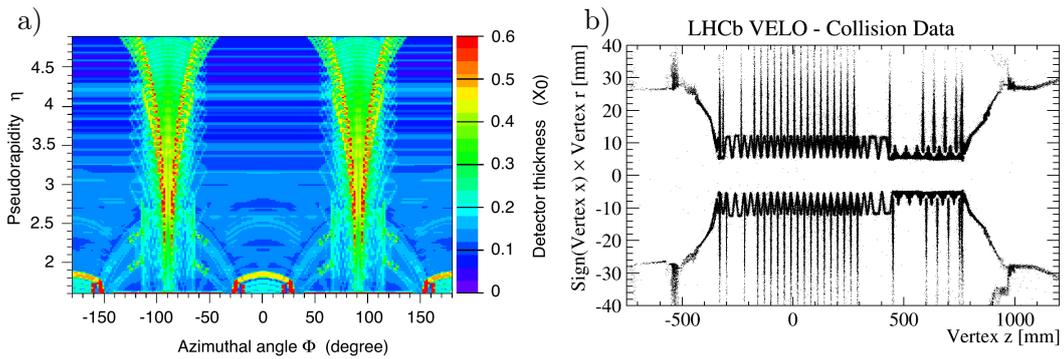


Figure 4.6.: a) Average radiation length seen by particles passing through the VELO as function of  $\eta$  and  $\phi$ . b) Reconstructed vertices of hadronic interactions in VELO material as function of  $z$  and radius  $\pm r$  (left/right of beam). Figures are taken from Refs [63, 72].

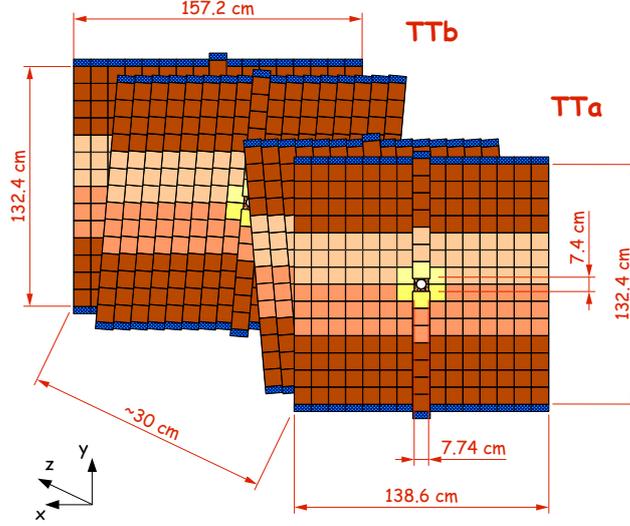


Figure 4.7.: Schematic view of the Tracker Turicensis. The x-layers are aligned straight, the stereo layers are tilted by  $5^\circ$ . Different read-out sectors are indicated by different shadings, the electronics are marked in blue. Figure taken from Ref. [74].

Besides the support structure including modules the RF foil can be clearly seen.

The two halves of the VELO are movable transverse to the beam line (along  $x$ -direction). During the injection of the LHC beams, the proton- or ion-beams are less stable in position. To protect the VELO sensors from severe damages during that phase, the half-discs are brought to a safety position around 3 cm away from the beam. When stable beam conditions are achieved, the VELO closes fully automated. The active sensor material then has a distance of only 8 mm to the beam.

#### 4.1.3. Tracker Turicensis

The Tracker Turicensis (TT)<sup>1</sup> [73] is situated behind RICH1 in front of the dipole magnet. The detector uses  $2 \times 2$  detection layers of silicon microstrip sensors. The strips have a pitch size of  $183 \mu\text{m}$  and a maximum length of 38 cm. Two different kinds of layers can be distinguished, so-called  $x$ -layers and stereo-layers ( $u$ ,  $v$ ). One layer of each type is arranged in a station, TTa has a  $(x, u)$  and TTb has a  $(v, x)$  configuration. The strips in  $x$ -layers are aligned vertically, while the strips in stereo layers are tilted by either  $-5^\circ$  ( $u$ ) or  $+5^\circ$  ( $v$ ) with respect to the vertical axis. This stereo view allows track parameters to be measured with a high resolution in the bending plane and with additional information in  $y$ -direction. Figure 4.7 shows the layout of both TT stations, which are separated by a distance of 27 cm. Each module consists of half modules which are read out from the top or the bottom, outside the acceptance of the detector. In total, the TT comprises an active silicon area of  $8.4 \text{ m}^2$  and 143 360 readout channels.

<sup>1</sup>The TT was formerly known as the Trigger Tracker.

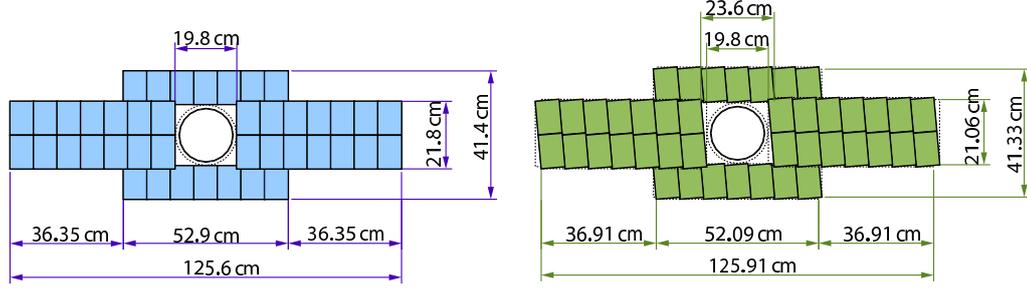


Figure 4.8.: Schematic view of  $x$ -layers (a) and stereo layers (b) in the Inner Tracker. Due to the deflection of charged particles by the magnetic field, there are more boxes in the bending plane compared to the non-bending plane. Figures taken from Ref. [74].

#### 4.1.4. Inner Tracker

Another silicon microstrip detector is located downstream of the magnet. The Inner Tracker (IT) [73] forms the inner part of the main tracking stations close the beam pipe. In this very forward region a much larger particle flux is reached compared to the outer region which is covered by a straw tube detector, the Outer Tracker.

The IT consists of three stations, where each station is built of four individual detector boxes that are arranged in a cross-shape around the beam pipe, see Fig. 4.8. Similar to the TT, each box comprises four detection layers in a  $(x, u, v, x)$ -configuration. The  $u$  and  $v$  layers are tilted by  $-5^\circ$  and  $+5^\circ$  with respect to the vertical  $x$ -layers. The layers have a height of 40 cm and width of 125 cm. In total this results in an active silicon area of around  $4 \text{ m}^2$ .

The silicon sensors used in the IT are very similar to the TT sensors. The strips have a pitch of  $198 \mu\text{m}$  and a length of 11 cm. Two different sensor sizes can be distinguished. The sensors left and right of the beam pipe have a thickness of  $410 \mu\text{m}$ , and  $320 \mu\text{m}$  for the sensors on the top and bottom. Similar to the TT, the obtained resolution for a single hit is  $50 \mu\text{m}$ . In total, the IT has 129 024 readout channels.

#### 4.1.5. Outer Tracker

The second component of the main tracking stations is called Outer Tracker (OT) [75–77]. It covers the large area outside of the IT acceptance. Figure 4.9 a) shows the dimensions of the OT, it has a size of around  $6 \times 5 \text{ m}$  in width and height. Due to the large size, the OT is not built of expensive silicon sensors. Instead, the Outer Tracker is a drift-time gas detector based on arrays of straw-tubes.

The straw-tubes have a cylindrical shape with a diameter of 4.9 mm and are filled with a gas mixture of  $\text{Ar}/\text{CO}_2/\text{O}_2$  (70.0%, 28.5%, 1.5%). This provides a spatial resolution of around  $200 \mu\text{m}$  and a fast drift time of below 50 ns, which is necessary to reduce the spill-over if the LHC is operated at 25 ns bunch spacing rate. In the centre of each tube there is an anode wire with a size of  $25 \mu\text{m}$  that collects the charge of the ionised gas.

The single straw tubes are placed next to each other in mono-layers where two staggered layers are combined to a double-layer in one module. A cross section of such a module and the placement of the straw-tubes is given in Fig. 4.9 b). The modules

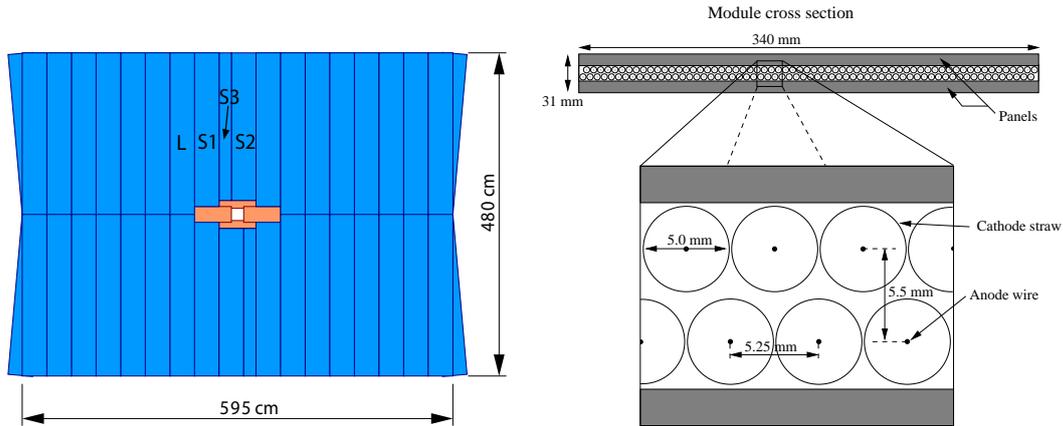


Figure 4.9.: a) Schematic view of an OT-station, the orange part indicates the IT. b) Cross section of an OT straw-tube module. Figures taken from Refs [63, 73].

are arranged in three stations with four layers, again in an alternating configuration of vertical  $x$ - and tilted stereo-layers. Due to the presence of the IT, the OT is made up of two different module sizes. There are 168 long modules with a length of 5 m containing 256 straw-tubes and 96 shorter modules containing half the number of tubes. This sums up to a total number of almost 53 760 readout channels.

The readout electronics are placed either on top or bottom of the modules, outside the active detection area. Inside the acceptance, the amount of material of one station (4 double-layers) sums up to 3.1% of a radiation length.

The acceptance of the OT is 300 mrad in the bending plane and 250 mrad in the non-bending plane. The efficiency to detect a hit in the central half of the straw is about 99.2%, and the position resolution is determined to be approximately  $200 \mu\text{m}$  [77].

## 4.2. Beam pipe

Since the LHCb experiment is focussed on the high rapidity region with a high particle density, the design of the beam pipe [63] is of particular importance. Any material seen by primary particles increases the probability to create secondary particles or accounts for multiple scattering. Both effects have a strong impact on the measurement of primary-produced charged particles. Therefore, the choice of material and the design of the beam pipe are crucial for the experiment and are briefly discussed in the following.

Within the LHCb detector a 19 m long beam pipe traverses the centre of all sub-systems, except for the VELO which is placed in a vacuum vessel. The first part of the beam pipe includes the exit window of the VELO which connects the vacuum vessel with the actual pipe containing the LHC beams. The spherically shaped exit window covers the entire LHCb acceptance. It is built of aluminium and has a diameter of 800 mm and a thickness of 2 mm.

The beam pipe itself is segmented into four conical sections. The first three sections, with a total length of 12 m, are made of beryllium. This material has a higher modulus of elasticity than steel and also has a small radiation length. For incident particles, beryllium

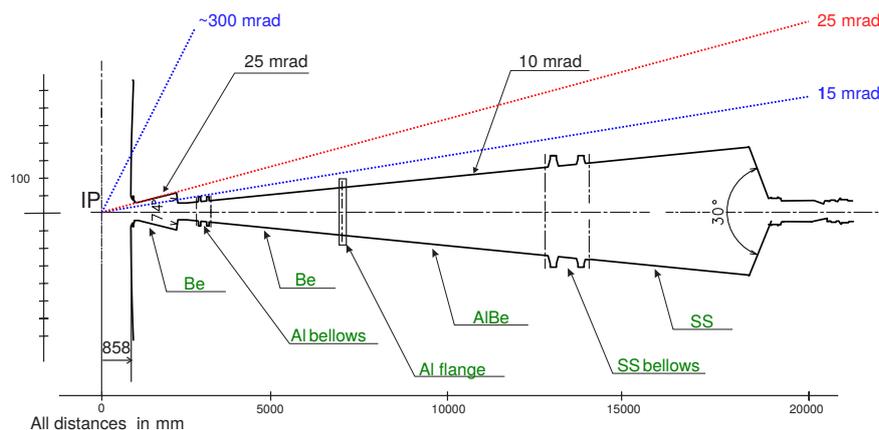


Figure 4.10.: Layout of the beam pipe in the LHCb coordinate system. The blue dotted lines indicate the borders of the main tracking system. The red dotted line marks the specific angle with which particles have to traverse a greater amount of material, related to the first section of the beam pipe. The material of the beam pipe and the connectors is given in green. Original figure is taken from Ref. [68].

is almost transparent. The first three segments are interconnected by aluminium bellows. Stainless steel is chosen for the last segment and the bellows used for the link. This part of the beam pipe is located behind of the calorimeter system, hence material interactions are not crucial any more.

The total layout of the beam pipe is given in Fig. 4.10. Beside its geometry also the angular acceptance of the main tracking system ( $\approx 15 - 300$  mrad) is indicated in blue. Two different opening angles of the conical pipes are visible. The rear part, which has an opening angle of 10 mrad, is outside of the acceptance and has no impact on the track reconstruction. In contrast to this, the front part of the beam pipe, with an opening angle of 25 mrad, will affect the reconstruction of particles in this specific range as the particles have to traverse a greater amount of material.

### 4.3. Particle identification

For many analyses, in particular in flavour physics, it is crucial to identify the final-state decay particles ( $\gamma$ ,  $e$ ,  $\mu$ ,  $\pi$ ,  $K$  and  $p$ ). The LHCb experiment provides an excellent particle identification (PID) by combining information from two Cherenkov detectors, the calorimeter system and the muon system.

Both analyses presented in this thesis only use inclusive particle samples without information about the particle species. However, information from PID-related detectors is also used to identify and suppress reconstruction artefacts.

In the following, the sub-detectors designed for particle identification will be explained briefly.

#### 4.3.1. Ring Imaging Cherenkov detectors

The LHCb experiment benefits from two Ring Imaging Cherenkov (RICH) [78, 79] detectors which are used for separating pions and kaons. Cherenkov detectors exploit

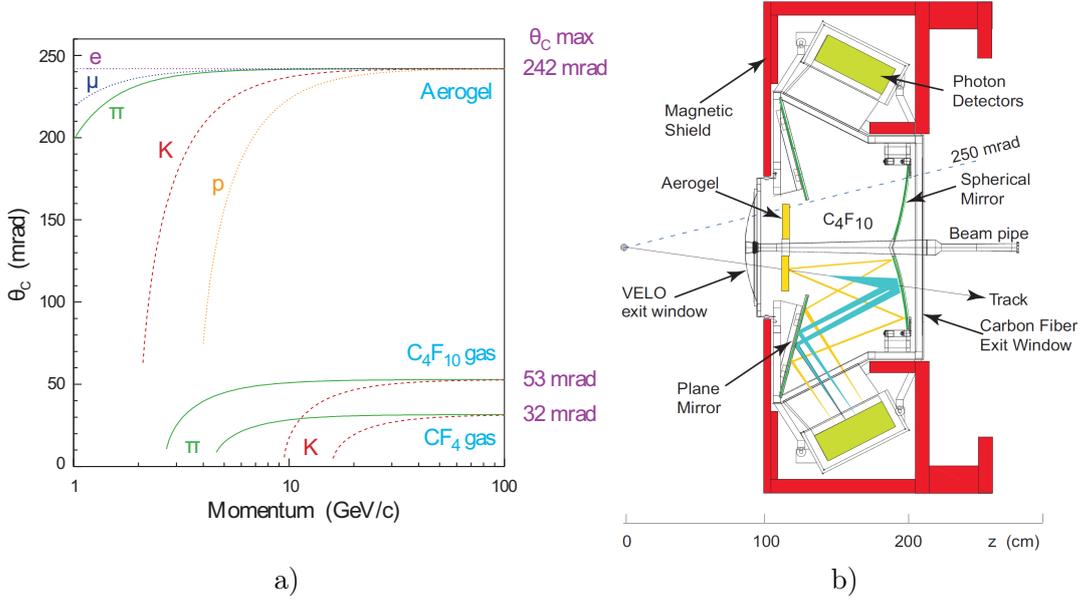


Figure 4.11.: Cherenkov angle as function of particle momentum for the RICH radiators (a). Schematic layout of RICH1 in side view (b). Figures taken from Ref. [63].

the Cherenkov effect [80] to distinguish different charged particle species. If a charged particle traverses a medium (called radiator) faster than the speed of light in this medium, electromagnetic radiation is emitted. These Cherenkov photons are produced in a cone around the flight direction of the particle with a specific opening angle  $\theta_C$ . Depending on the refractive index,  $n$ , of the radiator, this angle is given by

$$\cos \theta_C = \frac{1}{n\beta}, \quad (4.1)$$

where  $\beta = v/c$  is the relative speed of the particle with respect to the speed of light. By measuring the particle's momentum with the tracking system and the opening angle  $\theta_C$  with a RICH detector the invariant mass and thus, the particle type can be determined. There are two RICH detectors installed at LHCb, they use different radiators and cover a different momentum range. The relation of the Cherenkov angle with the particle's momentum is given in Fig. 4.11 a) for all radiators used in the RICH system.

RICH1 is located between VELO and TT and covers a wide acceptance range of  $\pm 15$  ( $\pm 25$ ) to 300 mrad (250 mrad) in the bending (non-bending) plane. A schematic view of the RICH1 is depicted in Fig. 4.11 b). It uses two types of radiators to cover a momentum range from a few GeV up to 60 GeV. Plates of aerogel with a refractive index of  $n = 1.03$  are directly placed behind the VELO exit window and allow a separation of particles with very low momenta. The inner volume of RICH1 is filled with  $C_4F_{10}$  gas ( $n = 1.0014$ ) which is a good radiator for separating pions and kaons up to 60 GeV. The Cherenkov light is focussed and reflected by a mirror system which maps the ring images onto a photon detector plane composed of Pixel Hybrid Photon Detectors (HPDs). These are located outside the LHCb acceptance and are placed in iron shielding boxes to attenuate the fringe field of the dipole. The photo-electrons created in a HPD are

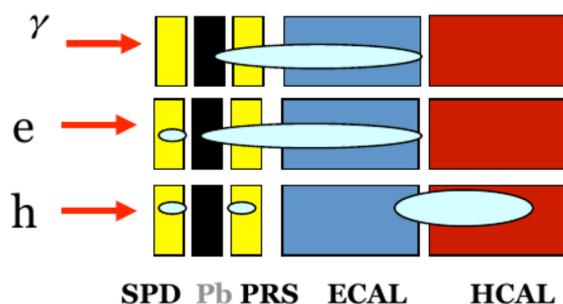


Figure 4.12.: Schematic view of the different signatures of photons, electrons and hadron in the components of the LHCb calorimeter system. Figures taken from Ref. [83].

detected by silicon pixel detectors which provide a position measurement. From this, the Cherenkov ring is calculated which is proportional to  $\theta_C$ .

The second RICH detector is situated behind the last tracking station, T3, and the first muon station, M1. RICH2 contains a  $\text{CF}_4$  gas radiator with a refractive index of  $n = 1.0005$ . It is designed to allow particle identification for high-momentum particles in the range of 15 – 100 GeV. Thus, in comparison to RICH1, RICH2 has a smaller angular acceptance of 15 to 120 mrad (100 mrad) in the bending (non-bending) plane. RICH2 is built similar to RICH1, except that the mirror system and the HPDs mounted left and right of the beam pipe.

#### 4.3.2. Calorimeter system

The calorimeter system [81, 82] of LHCb can be divided into four components, the two large electromagnetic (ECAL) and hadronic (HCAL) calorimeters and the two assisting detection layers, the Scintillating Pad Detector (SPD) and the Pre-Shower Detector (PS or PRS).

The calorimeter system is the only detector which is sensitive to neutral particles. It is used for three purposes. The first is to reconstruct photons and neutral pions. The second is to provide a total energy measurement of all particles except muons. The third is to identify photons, electrons and hadrons depending on their signatures in each of the calorimeter components. This is visualised in Fig. 4.12. While hadrons show the dominant signatures in the HCAL, photons and electrons both deposit their entire energy in the ECAL. By using two scintillating layers in combination with a lead plate, a discriminations between electrons and photons is possible. The fourth task is related to the hardware trigger. Calorimeter signal allow fast decisions, *e.g.* used to identify events with large transverse energy.

The single components of the calorimeter system are briefly discussed in the following.

**Scintillating Pad / Pre-Shower Detector** The first part of the calorimeter comprises the layers of the SPD/PS detectors which are located behind the first muon station. Each detector consists of rectangular scintillator pads which are placed in front of and behind a 15 mm lead absorber. Both detectors adopt a variable lateral segmentation

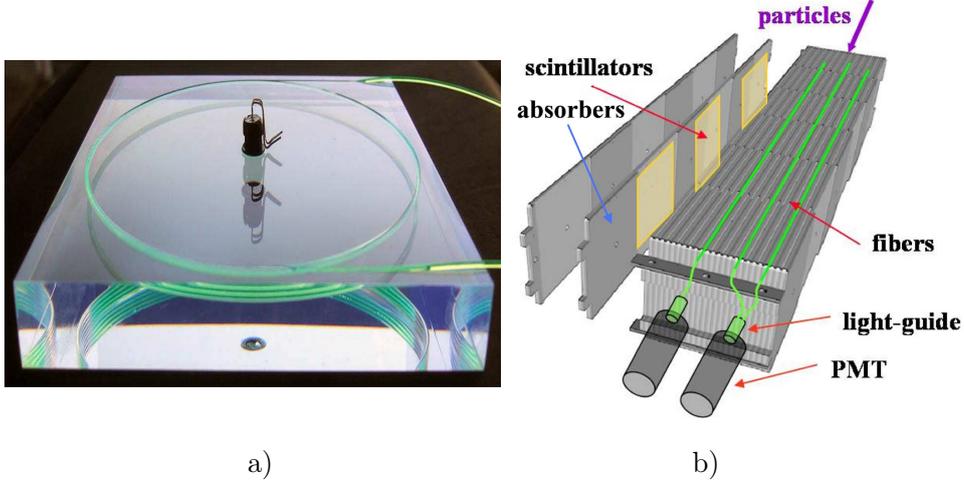


Figure 4.13.: Scintillator pad used in the SPD and PS (a). Schematic view of a HCAL cell showing the periodic absorber/scintillator design (b). Figures taken from Ref. [63].

since the particle density varies by two orders of magnitude comparing the central and peripheral region of the detector. The pads, an example is given in Fig. 4.13 a), have a size of approximately  $(4\text{ cm})^2$ ,  $(6\text{ cm})^2$  or  $(12\text{ cm})^2$ . A common segmentation for the SPD and PS is chosen such that it is in agreement with the subsequent ECAL segmentation. The resulting three regions are displayed in Fig. 4.14 a).

The hit-multiplicity in the SPD can be used as a measure of charged particle multiplicity as it is sensitive to all charged particles. In combination with the lead absorber photons can be identified, since they start showering in the PS but are not detected in the SPD. By exploiting the different shower lengths of electrons and hadrons, the PS helps to distinguish between electromagnetic and hadronic showers and is important for pion identification.

**Electromagnetic Calorimeter** The electromagnetic calorimeter is located next to the PS and is designed to measure particle showers of photons and electrons. The ECAL utilises a *shashlik* sampling calorimeter technology consisting of alternating absorber/detection layers. As absorbers, 2 mm thick lead layers are employed while the detection layers consist of scintillating polystyrene with a thickness of 4 mm. In total 66 absorber and detection layers are installed which results in 25 radiation lengths and 1.1 hadronic interaction lengths.

The obtained energy resolution for the ECAL is approximately

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\%, \quad (E \text{ in GeV}). \quad (4.2)$$

The first term of the resolution is related to statistical fluctuation, the second is due to the readout. Both terms have to be summed in quadrature.

**Hadronic Calorimeter** The hadronic calorimeter is placed between the ECAL and the muon system (M2-M5). Its purpose is to measure showers induced by hadrons. The

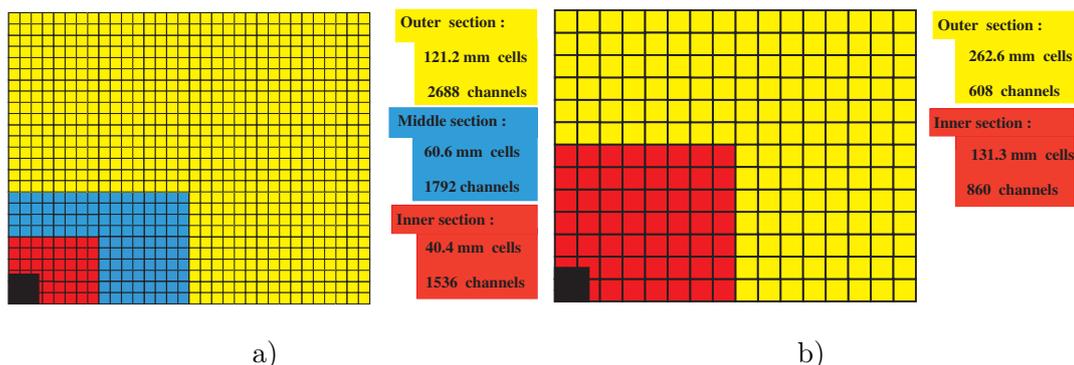


Figure 4.14.: Granularity of the calorimeter system showing one quarter of the detectors in front view. Three cell sizes are adopted for the SPD, PS and ECAL (a), the HCAL is divided into two regions. The position of the beam pipe is indicated in black. Figures are taken from Ref. [63].

HCAL is also a sampling device, but it uses iron as an absorber (16 mm) and scintillating pads (4 mm) as active material. In contrast to the ECAL, the iron absorbers and the scintillating tiles are oriented parallel to the beam line. The internal structure of a HCAL cell is shown in Fig. 4.13 b). A cell is composed of three absorbers and three scintillators connected to a photomultiplier by fibres. The length of the absorber and the scintillator corresponds to 5.6 hadronic interaction lengths. Also the HCAL is segmented in order to account for the higher particle flux towards the beam pipe. As hadronic showers are broader compared to electromagnetic ones, only two different cell sizes are employed as it is visualised in Fig. 4.14 b).

The performance of the HCAL gives an energy resolution of

$$\frac{\sigma_E}{E} = \frac{80\%}{\sqrt{E}} \oplus 10\%, \quad (E \text{ in GeV}). \quad (4.3)$$

The first term is related to statistical fluctuation, the second to readout. They have to be summed in quadrature.

### 4.3.3. Muon system

The muon system [84, 85] of LHCb consists of five rectangular shaped muon stations (M1-M5), as shown in Fig. 4.15. The first station, M1, is located in front of the calorimeter system while the stations M2-M5 are placed behind.

The purpose of M1 is to improve the transverse momentum resolution of muons in the trigger, as there is less multiple scattering before entering the calorimeter. The stations M2-M5 are interleaved with 80 cm thick iron plates acting as an absorber for all remaining detectable particles except muons. Including the preceding calorimeter system the absorber thickness corresponds to approximately 20 hadronic interaction lengths. A muon is required to have at least 6 GeV of momentum to cross all five muon stations.

The muon system has an angular acceptance of 20 (16) mrad to 306 (258) mrad in the bending (non-bending) plane, respectively. The total area covered by the five stations

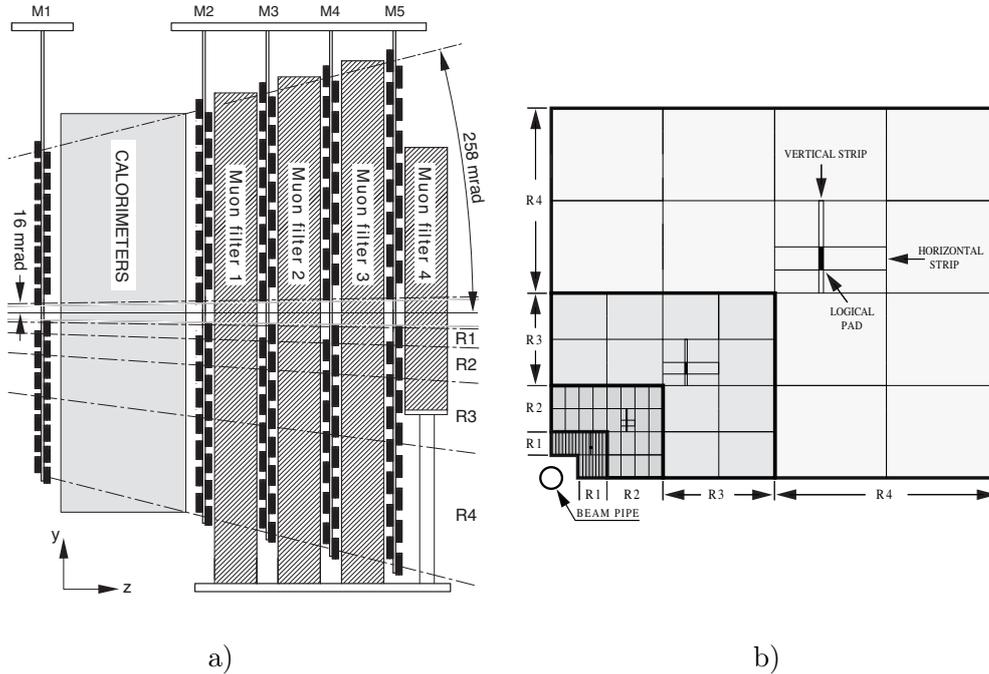


Figure 4.15.: Schematic view of the muon system. Layout of the five muons stations in side view (a). Front view of a quadrant showing the four regions of different granularity. Figures taken from Refs [63, 84].

sums up to  $435 \text{ m}^2$  filled by 1380 detection chambers. Each station is divided into four regions with increasing chamber size when increasing the distance to the beam line. The layout of the stations is given in Fig. 4.15 b).

Two different technologies are used in the construction of the muon stations. Except for the innermost region of station M1, multi-wire proportional chambers (MWPC) are used to build the muon system. Due to the high particle flux the remaining part of M1 uses Gas Electron Multipliers (GEM) which have a higher radiation tolerance.

The spatial resolution varies within the different muon stations. The purpose of the last two stations is mainly identification of muons while M1-M3 are also used to reconstruct muon tracks in the trigger. More details about the muon system and its usage in the trigger can be found in Refs. [84, 86].

#### 4.4. Trigger system

The nominal rate of collisions provided by the LHC is 40 MHz. The purpose of the trigger system is to reduce the event rate to approximately 5 kHz which corresponds to the maximum rate that can be written to storage. The trigger is designed to select only potentially interesting events that will be used for physics analyses. The crucial constraint for the trigger is to find a decision of rejecting or retaining the event within a very short time. In order to achieve a fast decision and to keep flexibility in defining events of interest, the trigger system is built of hardware and software components.

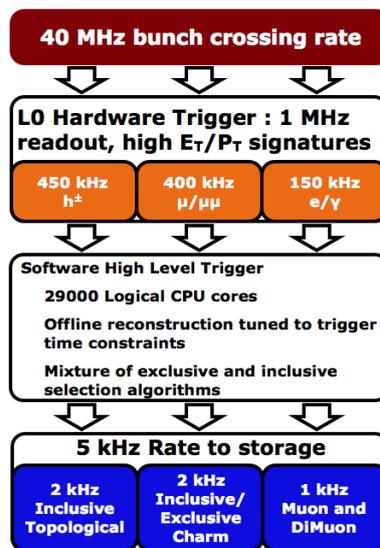


Figure 4.16.: Overview of the LHCb trigger scheme, by using the example of the 2012 configuration. Figure is taken from Ref. [58].

The analyses presented in this thesis uses data of special data taking periods. The interaction rates of these periods are low enough that essentially no trigger system is needed. Thus, only the basic idea of the LHCb trigger is explained in the following by using the configuration of the  $pp$  data taking in 2012 as an example (see Fig. 4.16).

The first stage of the trigger system is the hardware implemented **Level-0 (L0)** trigger. The incoming 40 MHz bunch crossing rate is reduced to around 1.1 MHz, at which the entire detector is read out. The L0 relies on two separate systems<sup>2</sup>, the calorimeter and the muon system. The calorimeter system is used to identify hadrons, electrons or photons with a large transverse momentum or large transverse energy. The muon trigger reconstructs the two highest  $p_T$  muons per event. These are typical signatures of decays of heavy B-mesons. An event is selected if certain threshold conditions on the calorimeter clusters or the muon tracks are fulfilled. The latency of the L0, the time between collision and decision, is fixed to  $4 \mu\text{s}$ . The output rate of the L0 consists of 450 kHz assigned to hadrons, followed by 400 kHz for muons and 150 kHz for electrons/photons.

The second trigger stage is dedicated to a software based **High Level Trigger (HLT)** which processes all events selected by the L0. The software trigger itself consists of components, HLT1 and HLT2, which are working subsequently. The HLT1 is designed to confirm the decision of the hardware trigger by performing a partial event reconstruction around the object that triggered the L0. In addition, also track based trigger configurations (*trigger lines*) are included which perform a partial event reconstruction independent of the L0 objects. The purpose is to look for topological signatures of B-decays, such as high momentum tracks with large impact parameters. The total output rate of the HLT1 is around 50 – 80 kHz. This allows the subsequent HLT2 to

<sup>2</sup>Originally, also the pile-up sensors were included into the L0 system. It was intended to reject events with more than one  $pp$  interaction, this strategy has been dropped during data taking.

perform a full event reconstruction close to the quality of the offline reconstruction. The HLT2 comprises a variety of specific exclusive and inclusive trigger lines that reduce the trigger rate to 5 kHz. These events are sent to storage and are used for the offline physics analyses.

More details about the design of the trigger system can be found in Refs [63,87], the performance of the trigger is evaluated in Ref. [86].

## 4.5. Data taking

Since the LHC has started to deliver  $pp$  collisions at a centre-of-mass energy of  $\sqrt{s} = 900$  GeV in 2009, the LHCb detector collected a huge amount of data in different beam configurations and at various energies. This section summarises the main data taking periods of the proton-proton runs in 2010-2012 and the proton-ion run in 2013.

### 4.5.1. Proton-proton period

The first proton-proton collisions with a centre-of-mass energy of  $\sqrt{s} = 7$  TeV were delivered and recorded in 2010. During the early data taking period after the commissioning, the LHC beams were filled with less than 10 proton-bunches per beam. The instantaneous luminosity at this time was only around  $\mathcal{L}_{inst} = 2 \cdot 10^{28} \text{ cm}^{-2}\text{s}^{-1}$  with an average number of interactions per bunch-crossing of less than 0.1. The recorded data samples of this period contain basically no pile-up interactions. These are ideal conditions to perform particle multiplicity measurements. Hence, data of this early period are used for the multiplicity analysis presented in the first part of this thesis.

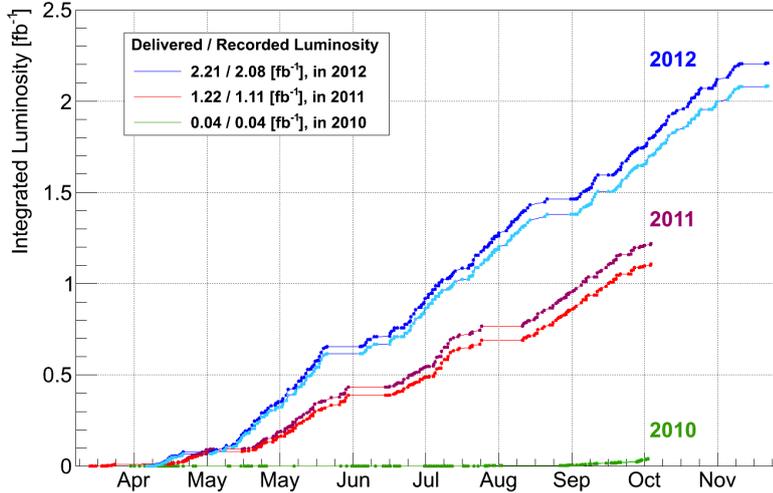


Figure 4.17.: Summary of the delivered (dark lines) and recorded (light lines) luminosity at the LHCb detector during the  $pp$  runs in the years 2010 – 2012. Figure is taken from Ref. [58].

With an increasing number of bunches (up to  $\approx 350$ ), also the interaction rate and luminosity increased up to  $\mathcal{L}_{inst} = 1.2 \cdot 10^{32} \text{ cm}^{-2}\text{s}^{-1}$  by the end of the year. The collected data for the entire year 2010 amounts to an integrated luminosity of  $\mathcal{L} = 0.04 \text{ fb}^{-1}$ .

In 2011, the delivered instantaneous luminosity went up to  $\mathcal{L}_{inst} = 3.8 \cdot 10^{32} \text{ cm}^{-2}\text{s}^{-1}$  with an average number of  $pp$  interactions of 1.5. The number of filled bunches in the LHC reached 1380 which is almost half of the design value of 2808 bunches. The size of the collected data sample in 2011 corresponds to  $\mathcal{L} = 1.11 \text{ fb}^{-1}$ .

In 2012, the energy of the proton beams could be increased, leading to a centre-of-mass energy of  $\sqrt{s} = 8 \text{ TeV}$ . The running conditions with around 1380 filled bunches and an average interaction rate of 1.7 are comparable to those of 2011. The stable conditions in 2012 were aiming at collecting as much data as possible. The recorded luminosity by LHCb in this year amounts to  $\mathcal{L} = 2.08 \text{ fb}^{-1}$ .

The size of the data samples for each year are summarised in Fig. 4.17.

#### 4.5.2. Proton-ion period

After a short pilot run of proton-ion collisions in 2012, where a small data sample with a luminosity of  $\mathcal{L} \approx 0.6 \mu\text{b}^{-1}$  was collected, a period of four weeks dedicated to proton-ion data taking followed in January/February of 2013. During that time, the LHCb detector collected  $\mathcal{L} \approx 1.7 \text{ nb}^{-1}$  of proton-ion collisions in different configurations. The instantaneous luminosity during that time was on average  $\mathcal{L}_{inst} = 3 \cdot 10^{27} \text{ cm}^{-2}\text{s}^{-1}$ .

In proton-ion collisions, two different beam configurations have to be distinguished. Depending on the orientation of the beams, either the proton or the lead remnant is travelling through the LHCb detector after the collision. The two configurations are visualised in Fig. 4.18. In the *forward* configuration the proton beam is pointing downstream the LHCb detector. Proton-lead collisions in this setup are further referred to as  $p+\text{Pb}$  collisions. In the opposite *backward* configuration the lead ion is pointing downstream the LHCb detector. This results in a larger particle density accompanied by larger detector occupancies. Collisions in this beam configuration are referred to as  $\text{Pb}+p$  collisions.

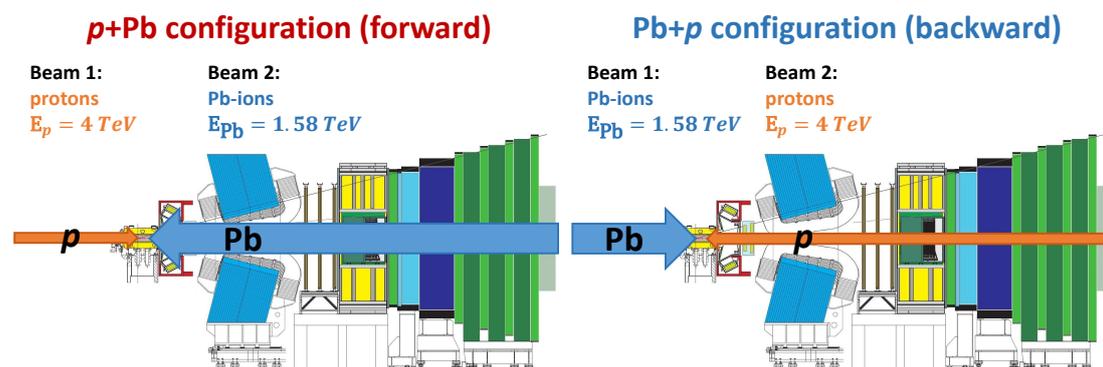


Figure 4.18.: Visualisation of the forward (left) and backward (right) beam configuration for proton-ion collisions. The nucleon-nucleon centre-of-mass energy in both configurations is  $\sqrt{s_{NN}} = 5.02 \text{ TeV}$ .

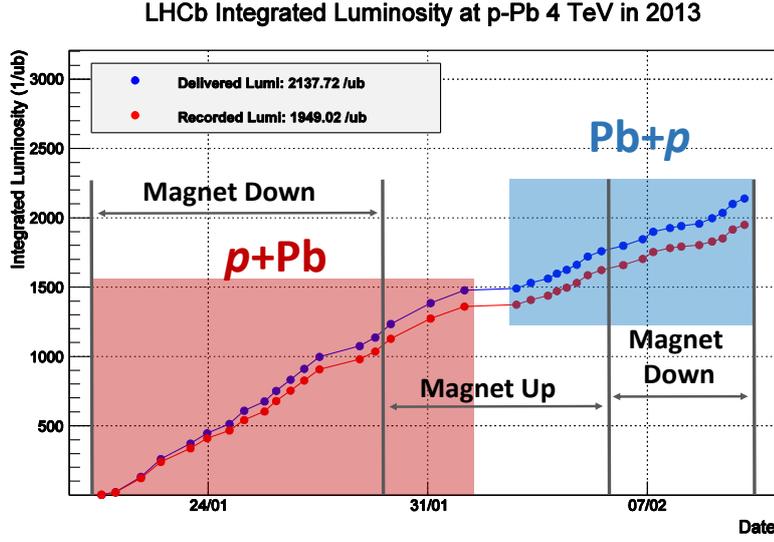


Figure 4.19.: Delivered and recorded integrated luminosity of the proton-ion run in 2013. Further indicated are the two different beam configurations and the reversal of the LHCb dipole. Original figure is taken from Ref. [58].

The asymmetric beam configuration of the proton and the lead beam results in a boost of the nucleon-nucleon centre-of-mass system in the direction of the proton, as shown in Chap. 3.2. As a result, two different (pseudo)rapidity regions in the nucleon-nucleon centre-of-mass system are probed with the LHCb detector. In the forward configuration the coverage is  $1.5 < \eta_{\text{cms}} < 4.5$ , in the backward configuration  $2.5 < \eta_{\text{cms}} < 5.5$ .

During the data taking, also the magnetic field has been reversed. The collected integrated luminosity of the entire proton-ion data sample is given in Fig. 4.19, indicating the different periods of  $p+Pb$  and  $Pb+p$  collisions together with the corresponding magnet polarities. The sizes of the data samples are listed in Tab. 4.1. Except for a larger amount of data collected in  $p+Pb$  configuration with magnet down polarity, the data samples have a comparable size. For the analysis of angular correlations in proton-ion collisions presented in the second part of this thesis, smaller equally sized sub-samples are used. Details are given in Chap. 14.

| beam configuration | magnet configuration | number of recorded events | data size (reconstructed) | integrated luminosity $\mathcal{L}$ |
|--------------------|----------------------|---------------------------|---------------------------|-------------------------------------|
| $p+Pb$             | down                 | $2248 \times 10^6$        | 120.0 Tbytes              | $769 \mu\text{b}^{-1}$              |
| $p+Pb$             | up                   | $533 \times 10^6$         | 27.9 Tbytes               | $298 \mu\text{b}^{-1}$              |
| $Pb+p$             | down                 | $675 \times 10^6$         | 51.2 Tbytes               | $303 \mu\text{b}^{-1}$              |
| $Pb+p$             | up                   | $591 \times 10^6$         | 47.2 Tbytes               | $263 \mu\text{b}^{-1}$              |

Table 4.1.: Overview of the proton-ion data samples recorded with the LHCb detector. The number events corresponds to all recorded events, the quoted integrated luminosity only considers data taking periods of good quality data, in which the detector was fully operational.

---

## From particles to tracks - Event and track reconstruction at LHCb

---

Particles produced in high-energetic  $pp$  or  $pPb$  collisions traverse the LHCb detector and leave signals in various components. These signals have to be first converted and prepared, before the properties of the original collision can be reconstructed in different software steps.

The focus of this thesis are charged particles whose trajectories can be reconstructed as *tracks* by exploiting the information of the tracking detectors. By knowing the position of a particle at a few different points in space the full flight path can be calculated. With the information of the traversed magnetic field also the momentum of a particle is determined. The specific particle species is not of interest for the presented analyses.

The performance of the track reconstruction is limited by experimental precision. The reconstruction is always affected by inefficiencies and artefacts related to the detector or the applied reconstruction software. These effects have an impact on the physics analyses and need to be accounted for.

The reconstructed tracks are further used to determine the primary interaction vertex (PV) of the collision. This allows separating primary produced particles that originate from the PV and secondary particles that are produced in decays.

This chapter focusses on the event reconstruction which is performed at LHCb. Firstly, the LHCb software framework is presented. It comprises software used to reconstruct the recorded data, to prepare event samples for the physics analyses, but also to simulate particle collisions within the detector. Secondly, an overview of the different track types is given. They are defined by the sub-detectors which are involved in the reconstruction process. This is followed by a discussion of the relevant reconstruction algorithms used in this thesis. The last part is dedicated to reconstruction artefacts and inefficiencies which have an impact on the physics analyses.

## 5.1. The LHCb software framework

The LHCb software comprises several packages used for the simulation and reconstruction of events, but also to perform the physics analysis. The whole software is based on the GAUDI [88] framework which provides the common interface. A brief overview of the relevant software packages is given in the following.

Particle physics analyses are typically based on two types of data. These are actual data that have been recorded by the detector but also simulated events which are used to build and test the analysis strategy. The simulation of particle collisions in the LHCb software is based on three steps before a common treatment of simulated and recorded data is performed. At first the software only used for **simulation** is discussed:

**Generator & simulation phase:** The GAUSS [89] framework represents the platform to run programs used to simulate entire physics events. The generator phase comprises different Monte Carlo event generators, *e.g.* PYTHIA, which are used for the simulation of the particle collision. The decay of heavy hadrons is simulated by using the EVTGEN [35] package in which radiative decays are governed by the PHOTOS [90] program. In the subsequent simulation phase of an event the propagation of particles through the detector is simulated using the GEANT4 [91,92] software. It describes electromagnetic and hadronic interactions with detector material.

**Digitization** For simulated interactions, the detector response is generated by the BOOLE [93] software. The resulting output mimics real data coming from the real detector and can be processed similarly.

**Trigger emulation** The MOORE [94] software provides the possibility to emulate the trigger response for simulated data using identical configurations as during data taking.

The common starting point of the reconstruction are *raw data* events that have been either simulated or triggered by the LHCb detector and subsequently written to a mass storage system. The raw data contain the full detector response which needs to be reconstructed and prepared first, before users can perform physics analyses:

**Offline reconstruction:** The raw data of an event are reconstructed by using the BRUNEL [95] software package. It includes algorithms to reconstruct properties of all particles that are visible to the detector. The trajectories of charged particles are reconstructed by using information of the tracking detectors. Depending on the involved sub-detectors, different track types are created, as discussed in the subsequent section. Neutral particles are reconstructed on the basis of the calorimeter system. The reconstructed charged and neutral objects are combined with information of the particle identification system and basic particle candidates are created.

**Analysis software:** The DAVINCI [96] project corresponds to the physics analysis software in LHCb. The particle candidates obtained from the track-reconstruction are

selected to form groups of *standard particles* according to predefined quality requirements. The groups comprise either specific particle species or inclusive selections of all quasi-stable particles. By using a specified decay chain, DAVINCI can subsequently combine standard particles in order to reconstruct unstable particles. The output of DAVINCI are typically smaller data samples which only contain a selection of events and particles used for the individual physics analysis.

## 5.2. Track types in LHCb

Charged particles travelling through the tracking system of LHCb can be reconstructed by tracking algorithms if a minimum number of detector hits has been recorded. In this context a hit is a well defined point or line in space which corresponds to a detector signal that has been induced by a particle, and not by noise. The important algorithms are discussed in the next section. Figure 5.1 illustrates different track types which are defined depending on the traversed sub-detectors, the VELO, the TT and the T-stations:

- **VELO tracks** are exclusively reconstructed in the VELO sub-detector with hits in the  $r$ - and  $\phi$ -sensors. They contain no momentum information as the VELO is outside the magnetic field. VELO tracks are predominately used in the primary vertex reconstruction but can be also used for physics analyses, such as a momentum inclusive measurement on charged particle multiplicities [98].
- **Upstream tracks** traverse the VELO and the TT-stations. The momentum of the corresponding particles is typically too small (below 2 GeV) to reach the T-stations and the particles are deflected out of the detector by the magnetic field. Upstream tracks contain momentum information, however, due to the small integrated magnetic field around the TT, the resolution  $\delta p/p$  is only  $\approx 15\%$ .

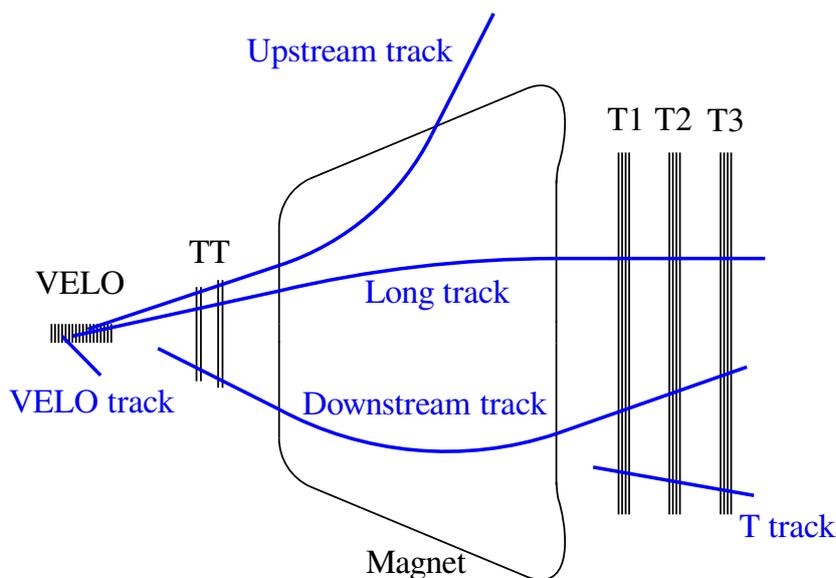


Figure 5.1.: Sketch of track types used in LHCb. Original figure taken from Ref. [97].

- **Long tracks** are the most important track type in LHCb. They traverse the entire tracking system and have hits in the VELO and in the T-stations. Optionally, also TT-hits are assigned. Long tracks also traverse the full magnetic field and thus, contain the most precise momentum information of all track types. The relative uncertainty varies from 0.4% at 2 GeV to 0.6% at 100 GeV momentum. Two separate tracking algorithms are used to reconstruct long tracks, the *forward tracking* and the *track matching*.
- **Downstream tracks** traverse the TT- and the T-stations but have no track segment in the VELO. Downstream tracks contain momentum information, too. However, the resolution is significantly worse compared to that of long tracks due to the missing VELO information and the resulting longer lever arm. This track type is important for the reconstruction of decay products of particles that typically decay outside the VELO, such as  $K_s^0$  mesons or  $\Lambda$  baryons.
- **T-tracks** are exclusively reconstructed in the T-stations. They are typically induced by particles produced in secondary decays or material interactions.

### 5.3. Track reconstruction LHCb

The different track types used in LHCb are based on several track finding algorithms. The tracking detectors are all located in the peripheral region of the magnetic field. The influence of the field is limited approximately to the area between VELO and T-stations (c.f. Fig. 4.3). While traversing the tracking detectors, charged particles are barely affected by the residual field so that their trajectories can be considered as almost straight lines. By exploiting this fact, the *pattern recognition* of the reconstruction software attempts to identify common hits from a single particle. The momentum information of a charged particle is obtained by combining information of its trajectory before and behind the magnet.

A particle's trajectory in the presence of a magnetic field can be described by five parameters, assuming that the magnetic field map is known. In LHCb the choice for these parameters are the positions in  $x, y$ , the corresponding slopes  $dx/dz$ ,  $dy/dz$  and the ratio of charge and momentum,  $q/p$ , where each of these parameter depends on the position in  $z$ .<sup>1</sup>

Track types relevant for the presented analyses are VELO tracks and long tracks, their implemented algorithms are briefly discussed in the following.

**VELO tracking:** VELO tracks are the first type of tracks that are reconstructed by the tracking software. They build the basis for the subsequently reconstructed long tracks. The standalone VELO reconstruction is performed by the *VELO pattern recognition* [99] which has been replaced by the *FastVelo* [100] algorithm in 2011.

The strength of the magnetic field in the VELO is negligible and particles are not deflected. Furthermore, all particles originate from the primary interaction along the beam line. These facts are exploited by the pattern recognition, where straight tracks

---

<sup>1</sup>In principle, also  $q/p$  depends on the  $z$ -position, since particles loose energy by interacting with the detector material.

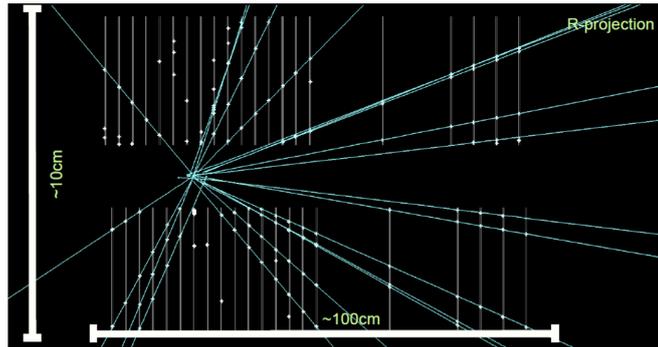


Figure 5.2.: The  $r$ - $z$ -projection of reconstructed tracks in the VELO. As there is no magnetic field, all tracks are reconstructed as straight lines.

are first reconstructed in the  $r$ - $z$ -projection. For this, only information of the  $r$ -sensors is considered. In the second step, the information of the  $\phi$ -sensors is added to form three dimensional tracks. An exemplary event display of reconstructed tracks in the VELO is given in Fig. 5.2.

**Long track algorithms:** Two independent algorithms are used to reconstruct long tracks. The *forward tracking* [101, 102] starts with a track *seed* in the VELO which is extrapolated through the magnetic field into the T-stations. The straight VELO track can only provide four of the five track parameters: the positions and the slopes. One additional measurement in or behind the magnetic field is sufficient to determine the remaining momentum information of the particle. Therefore, the forward tracking combines VELO tracks with hits in the T-station. The  $\pm y$  orientation of the magnetic field deflects particles, to first order, only in  $x$ -direction. Thus, the VELO-track seed points onto an  $x$ -plane of potential measurements in the T-stations. All hits within this plane are used to build an independent combination with the VELO track. For each hit-track combination, the T-station measurement is projected onto a reference plane,  $z_{\text{ref}}$ . Measurements corresponding to the original VELO track are supposed to cluster, while unrelated hits should be randomly distributed. The procedure, a *Hough transform*, is sketched in Fig. 5.3.

Hits that cluster in the reference plane are fitted in order to identify and remove outliers by using a  $\chi^2$  criterion. Typically, several track candidates with a sufficient number of hits are found for each VELO seed. The best candidate is identified by a quality variable,  $Q$ . The variable comprises information of the fit-quality, the momentum of the reconstructed track and the deviation from a straight-line extrapolation in  $y$  as the deflection acts only in  $x$ -direction. The track with the smallest value  $Q_{\text{min}}$  is retained, as well as tracks with  $Q \in [Q_{\text{min}}, Q_{\text{min}} + 1]$ , all others are rejected.

*Track matching* [103] is the second algorithm used to reconstruct long tracks. Starting point are two independent track segments reconstructed in the VELO and in the T-stations.<sup>2</sup> The standalone VELO and T-tracks are extrapolated onto common planes.

<sup>2</sup>The implementation of the matching procedure changed over the years. In 2010, the *track matching* algorithm based on *VELO pattern recognition* and *TsaSeeding & PatSeeding* were used, since 2011

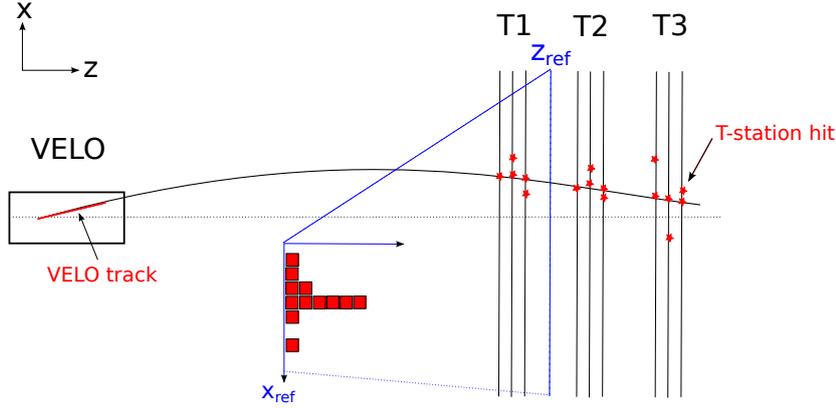


Figure 5.3.: Sketch visualising the forward tracking. A VELO track is extrapolated into the T-stations (T1-T3). Each of the found T-station hits is used to form a combined VELO-T-track. The T-hit is projected along the extrapolation onto a reference plane at  $z = z_{\text{ref}}$ . Common hits that belong to a track will cluster while unrelated hits are distributed randomly.

By considering the  $x$ - and  $y$ -distances between both extrapolated tracks a compatibility criterion is derived. Depending on the agreement between both track segments, they are combined and fitted to a long track or rejected.

The long tracks reconstructed by both algorithms can further have TT-hits. The TT-information is not essential and does not affect the track reconstruction itself. However, the quality of the tracks can be improved by adding the extra information of the TT.

In order to avoid multiple reconstructed tracks, an algorithm called *clone killer* [104] is run after the reconstruction sequence. The algorithm checks for shared hits among different tracks. If two tracks are found that have more than 70% of their hits in common, the track with the smaller number of hits is identified as a *clone track* and discarded.

## 5.4. Reconstruction artefacts and inefficiencies

The design of a particle detector and in particular a tracking system is a complex task and depends on the focus of the experiment. Two important effects must be considered in order to obtain a good performance.

On one hand, a high efficiency to reconstruct charged particle is desired. As explained for the forward tracking in LHCb, essentially two position measurements in front of a magnetic field and another measurement within or behind the field are sufficient to reconstruct a particle with momentum information.

On the other hand, a high purity of the reconstructed track sample is favoured. However, reconstruction artefacts always have an impact on the final reconstruction performance. The simplest example for such artefacts are combinatorial tracks. Increasing the number of detection planes reduces this particular kind of artefact but

---

the update *PatMatching* based on *FastVelo* and *PatSeeding* is applied.

causes new sources of problems: the more material is located within a detector, the larger is the probability of multiple scattering and material interaction. Both effects decrease the overall reconstruction efficiency, multiple scattering further affects the momentum resolution.

The design of the LHCb detector minimises the amount of material used for the tracking detectors. Except for the VELO and the three T-stations, the tracking system consists of air. This design allows a high track-reconstruction efficiency even for low-momentum particles ( $p \approx 2$  GeV). However, the limited number of tracking stations over a long distance results in a non-negligible amount of mis-reconstructed tracks.

In the following, definitions of relevant terms used in the context of charged particle reconstruction is given. Afterwards, sources for reconstruction artefacts and inefficiencies are discussed.

#### 5.4.1. Definitions and nomenclatures

The analyses on charged particle multiplicities and two-particle correlations are both based on charged particles that have been produced directly in the primary interaction:

**Charged particles:** All *quasi-stable* particles carrying electric charge are defined as charged particles. Namely, this class of particles comprises electrons, muons, charged pions, charged kaons, protons and their anti-particles, which are all visible to the detector.

**Prompt particles:** Within this thesis, prompt particles are defined as particles originating either directly from the primary interaction vertex or from a decay chain in which the sum of mean lifetimes does not exceed 10 ps. As a consequence, decay products of beauty and charm hadrons are treated as prompt particles.

To reconstruct prompt charged particles with momentum information, only long tracks can be used. Thus, the following discussion about inefficiencies and artefacts is limited to this track type. The following definitions apply to simulated particles but can be often analogously used for real particles.

**Kinematic requirements:** The first prerequisite for a (prompt) charged particle to be reconstructed is to fulfil the kinematic requirements defined by the tracking detectors. A long track requires measurements in the VELO and the T-stations, this constrains the acceptance in pseudorapidity to approximately  $2 < \eta < 4.9$ .<sup>3</sup> Requiring charged particles to stay within the geometric acceptance of the T-stations after deflection by the magnetic field, further restricts the phase space to a minimum momentum of approximately 2 GeV.

**Reconstructibility criterion:** A charged particle is considered as being *reconstructible* as a long track, if it leaves at least three  $r$ -sensor and three  $\phi$ -sensor hits in the VELO, and at least one  $x$ - and one *stereo*-hit in each of the T-stations (T1, T2 and T3).

---

<sup>3</sup>The  $\eta$ -acceptance is not a strict limit and may vary from analysis to analysis. Towards the edges of the acceptance ( $\eta \approx 2, 5$ ) the efficiency to reconstruct particles decreases drastically.

**Matched tracks:** In simulation, a particle is successfully reconstructed, if there is a track that shares at least 70% of its hits with the simulated particle. The track is then identified as *matched* to at least one simulated particle.

**Fake tracks:** The most important reconstruction artefacts are *fake* tracks, also referred to as *ghost* tracks. A track is considered as fake, if it does not correspond to the trajectory of a charged particle. In simulation, a track is identified as fake, if it cannot be matched to a simulated particle.

**Duplicate tracks:** Another reconstruction artefact are multiple reconstructed tracks. Two tracks are defined as *duplicate* or *clone* tracks, if they share at least 70% of their hits. In simulation, duplicate tracks are both matched to the same simulated particle.

#### 5.4.2. Reconstruction efficiencies

The prior condition for a charged particle to be reconstructed as a long track is to fulfil the kinematic requirements to traverse the tracking system. However, the reconstruction of these particles can be still prevented by several reasons which can be grouped into two categories:

- **Detector acceptance efficiency:** For particles fulfilling the kinematic requirements, the detector acceptance efficiency describes the fraction that reach the end of the T-stations and are unlikely to interact with material or to be deflected out of the detector by the magnetic field. Even though a particle meets the kinematic requirements its trajectory can still be bent out of the detector. This primarily affects particles that are produced close to the edges of the acceptance. Depending on its charge ( $\pm$ ), the resulting deflection is directed towards or away from the centre of the detector. As a consequence, particles produced with large pseudorapidities can be deflected away from the outer T-stations acceptance. Particles with small  $\eta$  can be deflected into the beam pipe. In both cases, a reconstruction is no longer possible. The probability for material interaction increases when a larger amount of material is traversed. Thus, the material of beam pipe has also a significant impact to the total acceptance efficiency. The average size of the acceptance efficiency depends on the kinematic range of the considered particles. As an example, in the pseudorapidity range of  $2.0 < \eta < 4.9$  and for particles with  $p_T > 150$  MeV, the average efficiency is  $\approx 70\%$ .<sup>4</sup>
- **Track-finding efficiency:** The efficiency to reconstruct a particle fulfilling the reconstructible criteria is quantified by the track-finding (tracking) efficiency. The efficiency depends on several external parameters, such as the implementation of the tracking algorithms, the alignment of the tracking detectors, etc. In addition, physics effects, such as multiple scattering have an impact on the efficiency. Particles with lower momentum are typically more difficult to reconstruct as their trajectories are more curved within the magnetic field and they have a higher

---

<sup>4</sup>This value is corresponds to minimum-bias  $p$ Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV.

probability for multiple scattering and material interaction. The average track-finding efficiency has been measured to be better than 95%, using muons from  $J/\psi$ -decays in a momentum range of  $5 < p < 200$  GeV [105]. However, towards the low boundary of the kinematic range of the detector, the efficiency decreases drastically.

### 5.4.3. Reconstruction artefacts

Reconstructed tracks can be grouped into 'good' tracks, which represent the properties of a genuine particle, and in reconstruction artefacts, which are fake and duplicate tracks. The origins for mis-reconstructed tracks are discussed in this section.

- **Fake tracks** are typically a result of high combinatorics of detector hits. The larger the number of hits in an event, the higher is the probability to find a random combination of hits that accidentally can form a track with a good quality.

With the LHCb detector design, the fraction of fake tracks is non-negligible since the extrapolation of tracks through the magnetic field is performed over a distance of several meters. Long tracks are prone to a mismatching of unrelated VELO and T-station segments, which then results in fake tracks. While the amount of fake tracks within the VELO is small, the T-stations are more sensitive to mis-reconstruction. This is primarily related to the smaller number of detection layers, but also due to hits induced by secondary particles and detector noise.

There are four important sources for the creation of fake long tracks [106]. The largest source are hadronic interactions of particles with detector material or the beam pipe. As an example, a successfully reconstructed VELO track is extrapolated to T-hits induced by a hadronic shower resulting in a fake long track. The second important category is the mismatching of track segments from two different particles. With a comparable probability, good VELO tracks are matched with a fake T-station parts formed by random hits. Vice versa, also fake VELO

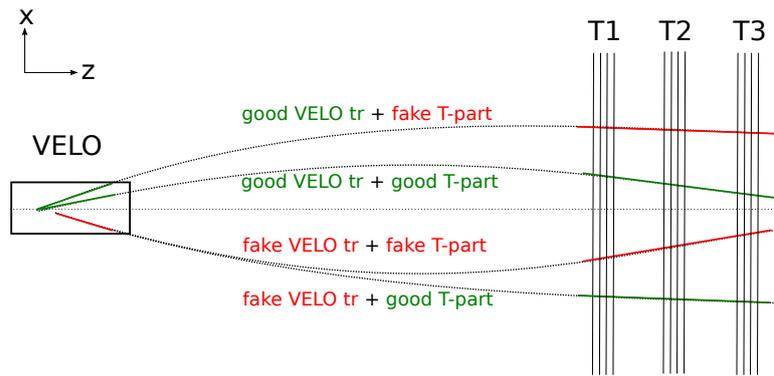


Figure 5.4.: Sketch visualising a good and fake long-tracks, consisting of different VELO and a T-station parts. Most of the fake tracks are made of good VELO track matched to a fake T-part that is built of random hits or hits induced by hadronic showers. Also combination based on fake VELO tracks are possible but less likely.

tracks can be matched to real T-tracks or T-hits. However, this is less likely by a factor of two. The combinations of VELO and T-station segments is visualised in Fig. 5.4.

- **Duplicate tracks** play a less important role in the reconstruction. In 2010 data, the fraction of duplicate tracks in the long track sample was of the order of a few percent. Clones originate from a splitting of a VELO track into two discrete tracks. Hits of a single particle are divided into two parts where each forms a similar but different VELO track. Both VELO segments then are combined with an identical T-station part and thus form a pair of duplicate tracks. In 2014 data, improvements of the reconstruction algorithms prevent this splitting. Further, an optimised clone killer suppresses duplicate tracks to a negligible amount.

The intrinsic difficulties with the track reconstruction, artefacts and inefficiencies have a major impact on the two following measurements. Therefore the quantification and correction are central aspects in these analyses and are discussed in more detail.

Part I.

Particle production in  
*pp* collisions at  $\sqrt{s} = 7 \text{ TeV}$



---

Introduction and analysis strategy

---

In this part of the thesis the measurement of charged particle multiplicities and densities in  $pp$  collisions is presented. The data has been collected with the LHCb detector at a centre-of-mass energy of  $\sqrt{s} = 7$  TeV in 2010. The presented analysis corresponds to the publication that can be found in Ref. [9].

As discussed in Chap. 2.2, soft-QCD processes cannot be predicted by perturbative calculations. In Monte Carlo event generators, soft processes, such as fragmentation, hadronisation and the underlying-event simulation, are described by different phenomenological models. These contain parameters that need to be tuned depending on the colliding particle species and the collision energy. The key motivation for multiplicity measurements is to provide new data for the parameter optimisation of phenomenological models.

Measurements of charged particle multiplicities performed with  $pp$  collisions at the LHC were reported by all four major experiments. The measurements from ALICE [107, 108], ATLAS [109, 110] and CMS [111] are all probing the central rapidity region. The forward region has been studied by the LHCb experiment in an early analysis, where an inclusive measurement without using momentum information was performed [98].

In the analysis presented here, information of the full LHCb tracking system is used, which enables the measurement of the momentum dependence of charged particle multiplicities. With the LHCb spectrometer,  $pp$  interactions are studied that produce at least one prompt charged particle (c.f. Chap. 5.4.1) in the pseudorapidity range of  $2 < \eta < 4.8$ , with a momentum of  $p > 2$  GeV and transverse momentum of  $p_T > 0.2$  GeV. Multiplicity distributions,  $P(n)$ , are reported for the total accessible phase space region as well as for  $\eta$  and  $p_T$  ranges. Further, mean particle densities are presented as functions of transverse momentum,  $dn/dp_T$ , and of pseudorapidity,  $dn/d\eta$ .

For illustration, an example of simulated mean particle densities as a function of  $\eta$  is shown in Fig. 6.1 a). It describes the average number of prompt charged particles that are produced in the selected phase space region, per  $pp$  collision and per unit of pseudorapidity. The two displayed distributions compare particle densities in  $pp$  collisions at  $\sqrt{s} = 7$  TeV for particles within the full phase space and for particles that fulfil the

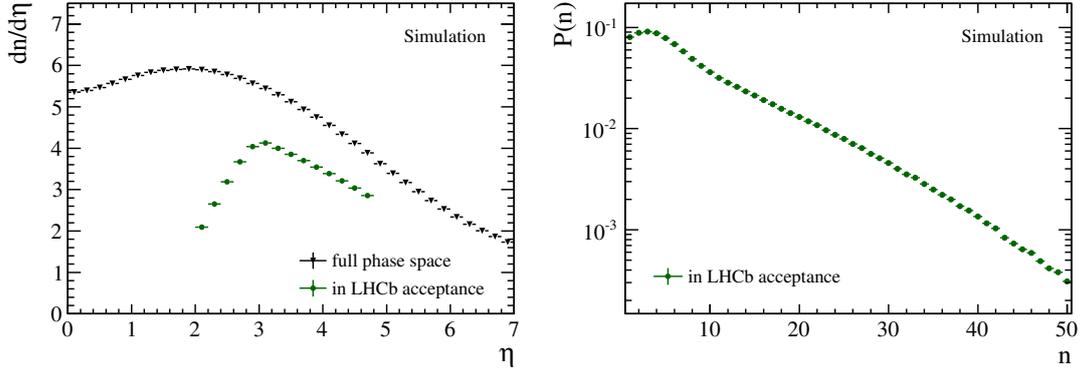


Figure 6.1.: Exemplary distributions for charged particle densities as function of pseudorapidity,  $dn/d\eta$ , and charged particle multiplicities,  $P(n)$ , in simulated  $pp$  collisions at  $\sqrt{s} = 7$  TeV. a) The  $dn/d\eta$  distributions shows results for the entire phase space (no kinematic constraints) and for particles within the accessible kinematic range of the LHCb detector ( $p > 2$  GeV,  $p_T > 0.2$  GeV,  $2.0 < \eta < 4.8$ ). b) Multiplicity distribution for charged particles within the full LHCb acceptance.

kinematic constrains defined by the acceptance of the LHCb detector. The resulting distribution corresponds to the accessible region of the this analysis. The minimum total and transverse momentum requirements are responsible for the falling edge in the particle density distribution towards small  $\eta$ . In the central region, typically a huge amount of particles with lower momenta are produced, except in hard interactions.

Figure 6.1 b) shows an example of the corresponding charged particle multiplicity distribution for particles fulfilling the kinematic range of the analysis. The distribution describes the probability that a certain number of prompt charged particles is produced in a single  $pp$  collision.

In the following, the structure of the subsequent chapters is outlined together with an explanation of the analysis strategy .

The analysis begins with Chap. 7, in which the recorded data samples and the employed Monte Carlo simulation are described. The simulation is used to test the analysis procedure and to determined correction factors. These are required to account for detector effects that affect the purity of the measured sample.

In order to provide a measurement that can be directly compared to predictions of event generators, a precise definition of a *visible event* for this measurement is necessary. Chapter 8 addresses the visibility definition and discusses the selection criteria used to define a fiducial region and to select prompt particles.

The analysis method is explained in Chap. 9. In order to have access to momentum information for the measurements, long tracks are used throughout the analysis. Reconstruction artefacts and limited efficiencies of the event and track reconstruction require the use of correction factors. These are determined on the basis of simulated event samples. A validation of the correction factor is done using data-driven methods. The corrections applied to measured particle multiplicities and mean particle densities are arranged in four steps:

- 
1. Reconstructed events are corrected on an event-by-event basis by weighting tracks according to purity factors. These account for reconstruction artefacts and for mis-selected non-prompt particles.
  2. The event sample is further corrected for unobserved events which fulfil the visibility criteria but in which no tracks are reconstructed.
  3. In order to obtain results for single  $pp$  collisions only, a correction to remove pile-up interactions is applied.
  4. The effects of detector and reconstruction inefficiencies are accounted for in the last correction step.

In Chap. 10, systematic uncertainties are studied. The precision of the measurements is not limited by statistical but systematic uncertainties. Most of these are related to potential differences between the simulation and data. Systematic uncertainties affect the precision of the correction factors and thus the measurements. Of particular interest are fake tracks in the reconstruction. The data used in this analysis correspond to an early data taking period which requires an evaluation of the reconstruction performance. A method has been developed in order to estimate the amount of fake tracks directly in data.

In Chap 11, the final results of the particle multiplicity and the particle density measurements are presented. The data are compared to a variety of predictions from different Monte Carlo event generators and a selection of their tunes. Besides the old PYTHIA 6 and PHOJET event generators also predictions of recent versions of PYTHIA 8 and HERWIG++ are included in the comparison.



The following analysis on charged particle multiplicities and densities is performed on a data sample of  $pp$  collisions at a centre-of-mass energy of  $\sqrt{s} = 7$  TeV. The data were recorded with the LHCb detector during a low-luminosity running period in May 2010. In that time, the average number of interactions in the detector acceptance per recorded bunch crossing ( $\mu$ ) was less than 0.1. The contribution from bunch crossings with more than one collision (pile-up events) is less than 4%. The data consists of 3 million events recorded in equal proportions of both magnetic field polarities.

As mentioned previously, the low luminosity during this early data taking period allowed the LHCb detector to be operated in a simplified trigger scheme. The hardware stage of the trigger system accepted all  $pp$  bunch-crossing events which were then directly reconstructed in the software trigger. An inclusive trigger selected all events with at least one reconstructed track segment in the VELO.<sup>1</sup> The resulting data sample corresponds to a *minimum-bias* event sample with no additional selection.

This analysis on charged particles is based on off-line reconstructed long tracks. By using a sample of *NoBias* events that are selected by a random trigger, it is tested that the minimum-bias trigger efficiency for events containing at least one long track is 100%.<sup>2</sup>

Despite the default configuration of the LHCb track reconstruction, only one of the two redundant long-track reconstruction-algorithms (forward tracking) is utilised in this analysis. This choice is motivated by a simpler estimation of systematic uncertainties related to reconstruction artefacts. Chapter 10 discusses this topic in more detail. The exact definition of a visible event in this analysis is given in the subsequent chapter.

The simulation sample comprises  $2 \times 10$  million minimum-bias events, produced in magnet up and down configuration. This large sample is used to determine correction factors for detector acceptance and resolution as well as to quantify background contributions and the reconstruction performance. To test the analysis procedure, a smaller

---

<sup>1</sup>The trigger line in LHCb is called `Hlt1MBMicroBiasRZVelo`.

<sup>2</sup>The long track reconstruction requires an offline VELO track first. The performance of the offline and online VELO reconstruction is almost identical. Thus, a 100% trigger efficiency is reasonable.

sub-sample of  $2 \times 3$  million events is used. That corresponds to double of the statistics which is considered in the data sample.

To obtain the fully simulated minimum-bias  $pp$  collisions the PYTHIA 6.4 event generator is used, operated in the LHCb configuration, c.f. Chap. 2.3.2. The simulation contains elastic and inelastic interactions, where the latter also includes single and double diffractive components. The generator was further adjusted in order to contain no pile-up interactions. At generator level, only events with exactly one  $pp$  interaction are selected.

In simulation, the colliding proton beams have been implemented with effectively no crossing angle. The crossing angle  $\theta$  is defined as the full angle between both proton beams. In real beam optics, a crossing angle is required to collide the two beams at a defined interaction point. However, the angle in data is expected to be very small, such that measurements remain unaffected. The crossing angle is determined from straight line fits through reconstructed primary vertices in so-called beam-gas events. These are composed of *beam-empty* and *empty-beam* events in which only one of the two crossing beams contains filled proton bunches at the time of the recorded collision. The position of primary vertices, thus, is distributed along the respective beam line. Figure 7.1 shows the fit results and the values obtained for  $\theta$  in the horizontal and vertical plane, separately for both magnetic field configuration. Depending on the magnetic field polarity slightly different beam-optics are required which results in different crossing angles. In the vertical plane the beams are aligned in parallel while in the horizontal plane a small angle of  $\theta = (0.64 \pm 0.04)$  mrad ( $\theta = 0.45 \pm 0.02$  mrad) is measured for the magnet-up (magnet-down) data sample. The obtained horizontal crossing angles are confirmed to be as small as expected. The impact to the measurement is negligible.

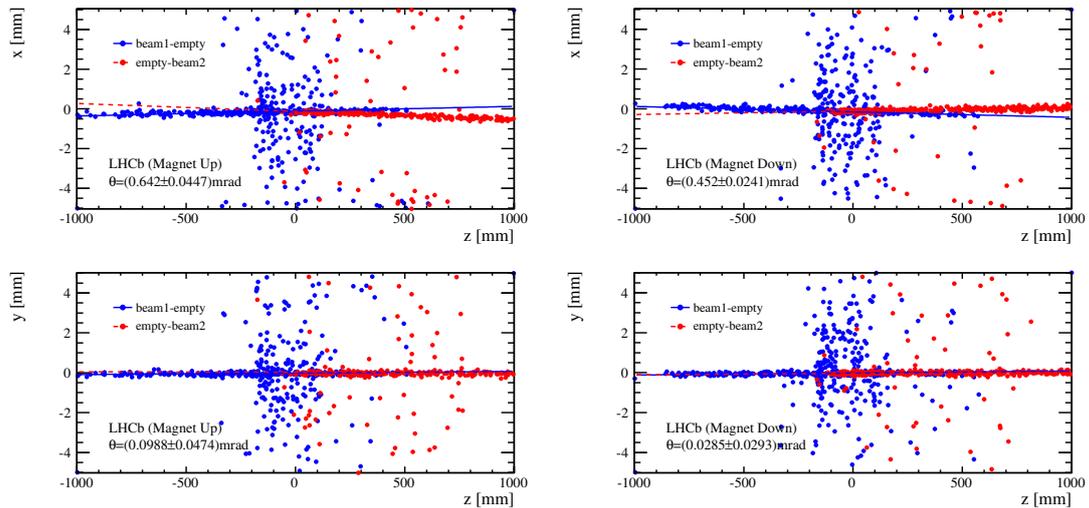


Figure 7.1.: Distribution of reconstructed primary vertices in the horizontal (top) and vertical (bottom) plane for the magnet up (left) and magnet down (right) data samples. Fitted lines represent the individual beam lines of both colliding beams using beam-empty and empty-beam events. The crossing angle  $\theta$  is given in the plots.

---

## Event definition and data selection

---

In this chapter, the basis for this analysis is set by first defining visible events for the following measurements. According to this definition, the experimental selection of events is specified. Afterwards, the track selection is discussed, which is applied to reconstructed tracks in order to obtain a sample of prompt charged particles. The event and track selections are identical for the measurements of charged particle multiplicities and densities.

### 8.1. Event definition and selection

The precise definition of a visible event is crucial for a particle multiplicity measurement. A slightly ambiguous event definition can already have a strong impact on the normalisation of the measurement. This has a critical impact when using data to optimise event generator predictions.

For this analysis a simple event definition is chosen that allows the measurement to be directly compared to simulation.

► An event is defined as *visible* if it contains at least one charged particle in the pseudorapidity range of  $2.0 < \eta < 4.8$  with  $p_T > 0.2 \text{ GeV}$  and  $p > 2 \text{ GeV}$ .

These criteria correspond to the typical kinematic requirements for particles traversing the magnetic field and reaching the main tracking stations. In order to compare the data directly to event generator predictions without having a full detector simulation available, the visibility definition is based on the actual presence of real charged particles, regardless of whether they are reconstructed as tracks or not.

In simulation, only events fulfilling the visibility definition are considered. The event selection in data is chosen in analogy, but is constrained by experimental observables. As a first pre-condition, only events are considered that have been recorded with a *beam-beam* crossing flag. This requires that the colliding proton-bunches of both

beams were actually filled when crossing at the interaction point. The events must be further selected by the minimum-bias trigger which requires at least one reconstructed VELO track. The only additional event selection criterion requires that at least one reconstructed long track is present, which fulfils the track selection given below. This condition corresponds to the minimum selection requirement for the analysis. However, this implies a fraction of events that remain undetected although they fulfil the visibility criterion. A correction dedicated to these undetected events is applied in the course of the analysis.

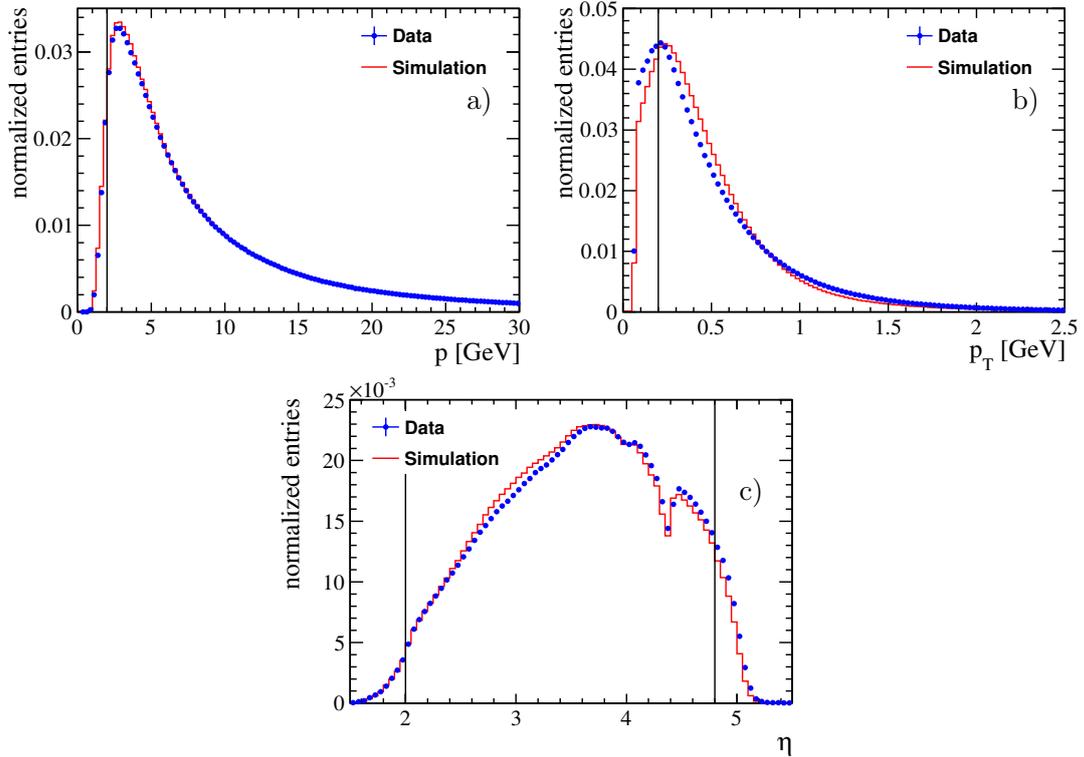


Figure 8.1.: Distributions of the track variables  $p$ ,  $p_T$  and  $\eta$  in data and simulation. The variables are used to define the kinematic range of the analysis, which is indicated by the vertical lines. No additional selection has been applied to the track sample before.

## 8.2. Track selection

The target of the track selection is to obtain a sample of charged particles that are directly produced in the  $pp$  collision. The selection criteria are chosen to be relaxed in order to retain a large reconstruction efficiency. Resulting contaminations of reconstruction artefacts or non-prompt particles are accounted for later in the analysis.

Long tracks which traverse the entire tracking system are considered for the measurements. The **kinematic phase space** which is covered by long tracks corresponds to:

- total momentum of  $p > 2 \text{ GeV}$
- transverse momentum of  $p_T > 0.2 \text{ GeV}$
- pseudorapidity range of  $2.0 < \eta < 4.8$

The kinematic criteria are explicitly applied to all tracks to restrict the measurement to a kinematic range in which the reconstruction efficiency is high. The distribution of the reconstructed kinematic variables in data and in simulation are presented in Fig. 8.1, together with the selection cuts. Data and simulation show good agreement, only the transverse momentum distribution towards lower  $p_T$  could be improved in simulation. The kinematic cuts are selected to be close to the intrinsic acceptance cut-off of the detector.

The three variables  $p$ ,  $p_T$  and  $\eta$  are strongly correlated with each other. By limiting the allowed range of two of these variables the third variable is affected indirectly. For example, the pseudorapidity distribution in Fig. 8.1 c) shows a decrease towards low  $\eta$ . This  $\eta$ -range is populated by many particles with very low total and transverse momenta. The minimum  $p$  and  $p_T$  requirements of the detector and the selection, prevent this particles from being reconstructed, which results in the observed decrease in the  $\eta$ -distribution.

The track reconstruction requires a minimum number of detector hits (in the VELO and in the T-stations), and a successful track fit. To ensure a high reconstruction efficiency no additional quality requirements for suppressing mis-reconstructed tracks are applied. Instead, the amount of reconstruction artefacts is precisely quantified and statistically corrected afterwards.

Tracks fulfilling the kinematic conditions are further required to originate from the primary interaction. Two conditions are set: The smallest distance  $d_{beam}$  of a track extrapolated to the beam line is required to be smaller than 2 mm. Furthermore, all tracks have to originate from the *luminous region*. The distance  $|z_0|$  of a track to the centre of this region has to fulfil  $|z_0| < 3\sigma_L$ . The beam line, the width  $\sigma_L$  and the mean position of the luminous region are determined from events with reconstructed primary vertices. In Fig. 8.2, the  $x$ -,  $y$ -, and  $z$ -position of these vertices are displayed. A Gaussian distribution is fitted to the central part of each distribution in which nominal  $pp$  collisions are peaking. From the longitudinal position of the vertices (e-f), the width  $\sigma_L$  is determined to be of the order of 40 mm in data. For simulation, a smaller spread of the longitudinal position of the primary interaction point was implemented, discrepancies to data are not of importance. Restricting tracks to originate from the luminous region

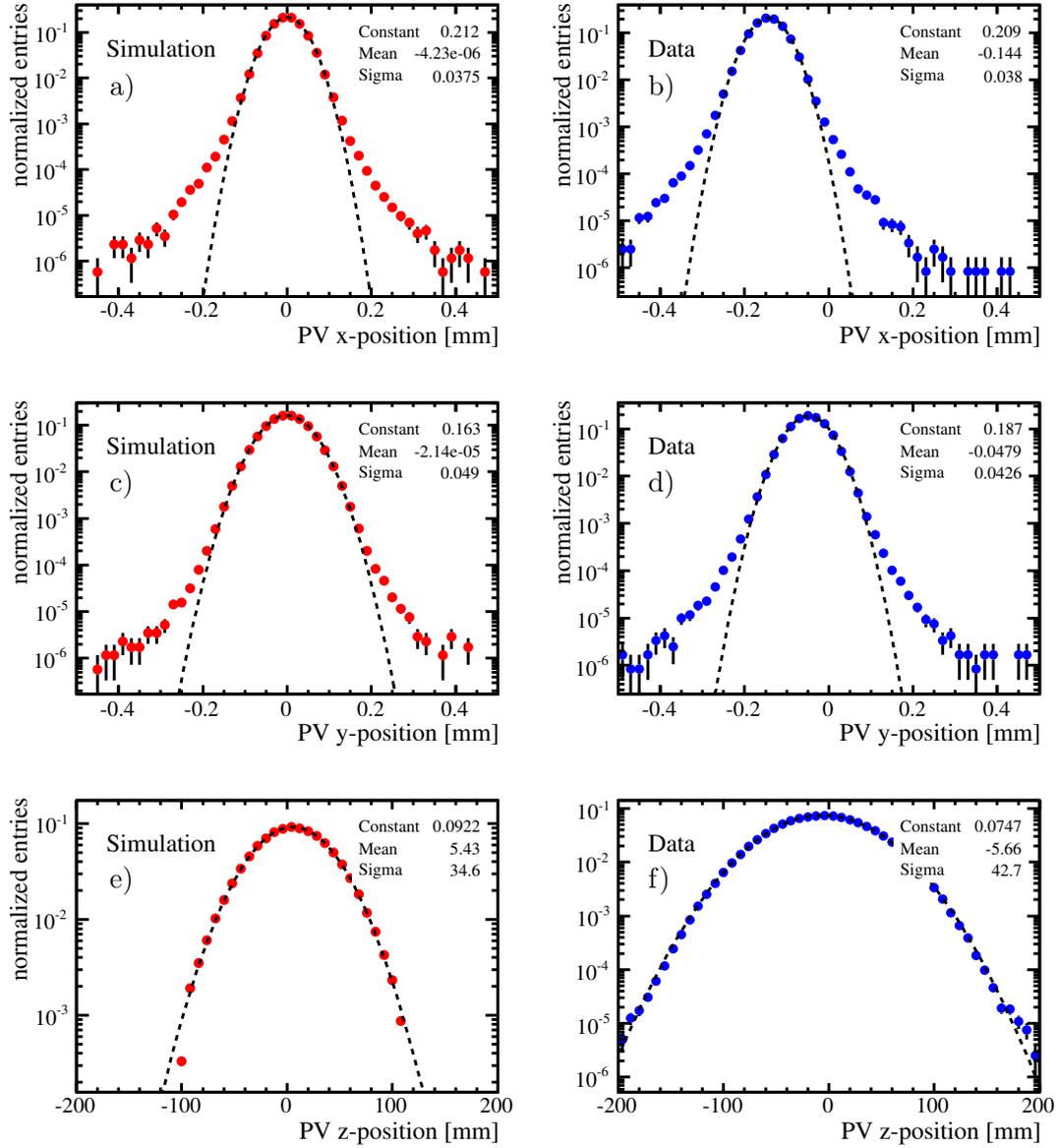


Figure 8.2.: Position of reconstructed primary vertices in the three space coordinates for simulation and data. The core region of the distributions is described by a Gaussian distribution. The data events correspond to only one LHC fill. However, the spread of the PV's  $z$ -position is slightly larger than implemented in the simulation.

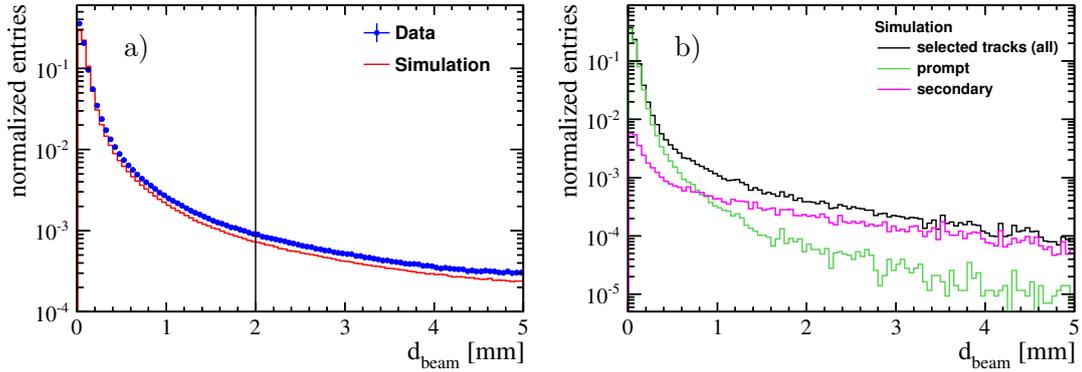


Figure 8.3.: a) Distribution of  $d_{\text{beam}}$  for reconstructed tracks in data and simulation, no prior selection is applied. The vertical line indicates the selection value. b) Separation of prompt and secondary particles as function of  $d_{\text{beam}}$  in simulation. Only selected track within the kinematic region of the analysis are considered.

also suppresses the contaminations from beam-gas interactions to a negligible amount, since beam-gas interactions are distributed flat along the beam line.

The distribution of the variable  $d_{\text{beam}}$  in data and simulation is shown in Fig. 8.3, together with the discrimination between prompt and secondary particles using generator information. The selected cut value of  $d_{\text{beam}} < 2$  mm is motivated to be consistent with other LHCb analyses (see *e.g.* Ref. [98]). The actual selection value is not of importance, since the resulting inefficiency of retaining prompt particles and the contamination by mis-selected secondary particles is determined and corrected afterwards.

The distribution of the variable  $z_0$ , for data and simulation, is depicted in Fig. 8.4. Prompt particles originating from the primary interaction are peaking around small values of  $|z_0|$ . The selection ranges defined for the luminous region ( $\pm 3\sigma_L$ ) is indicated in the respective distributions. Also for  $z_0$ , a smaller width of the core region in the

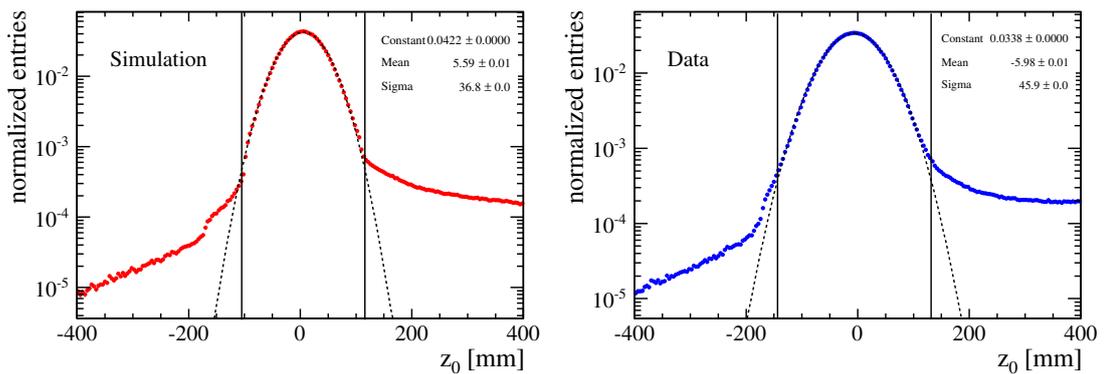


Figure 8.4.: Measured distances for extrapolated tracks to the centre of the luminous region. Tracks within a range of  $|z_0| < 3\sigma_L$  are selected, as indicated by the vertical lines. Outside of the core region, secondary particles and particles from beam-gas interactions dominate the distribution.

simulation sample is found. This is a direct result of the implemented smaller primary interaction region in simulation. Restricting the track samples to the defined  $3\sigma_L$  range acts as a loose criterion to suppresses apparent contaminations from secondary particles and particles originating from beam-gas background interactions.

It should be emphasised that there is no explicit requirement for a reconstructed primary vertex in this analysis. The reconstruction of a vertex requires a minimum number of tracks (originating from a common vertex), which would bias the minimum number of particles in an event. Without requiring a reconstructed PV and by using the chosen definition of a visible event, the measurement can be performed for events containing only a single particle.

All selection criteria are summarised again in Tab. 8.1.

|                           | Selection Parameter                 | Cut value           |
|---------------------------|-------------------------------------|---------------------|
| quality requirements      | long track reconstruction           | successful fit      |
|                           | explicit quality requirements       | none                |
| kinematic selection       | pseudorapidity $\eta$               | $\in [2.0, 4.8]$    |
|                           | total momentum $p$                  | $> 2.0 \text{ GeV}$ |
|                           | transv. momentum $p_T$              | $> 0.2 \text{ GeV}$ |
| prompt particle selection | distance to beam line $d_{beam}$    | $< 2.0 \text{ mm}$  |
|                           | distance to luminous region $ z_0 $ | $< 3\sigma_L$       |

Table 8.1.: Summary of the track selection applied to reconstructed tracks in data and in simulation.

---

## Data analysis and correction

---

The simple idea of counting charged particles in an experiment is complicated by several effects which bias the observed number of tracks towards larger and smaller multiplicities. The mis-reconstruction of non-physical tracks and the imperfect suppression of non-prompt particles both add additional tracks to the selected track sample. The individual identification of a 'good' track is impossible and thus implies a statistical correction to the number of selected tracks. While for determining a particle density, which represents an average value, applying naive multiplicative correction factors is sufficient, the correction of an integer multiplicity distribution requires a more sophisticated approach.

Limited reconstruction efficiencies have an opposite impact on the observed track multiplicity. Correcting the reduced number of particles is again more difficult for a multiplicity distribution than for an average density. Using average correction factors results in a non-integer track multiplicities per event.

As a consequence, the correction of particle multiplicities is treated in a different way than that of particle densities. However, the correction factors, which are determined from simulation, are the same for both observables.

To obtain fully corrected particle densities and particle multiplicity distributions the analysis is arranged in four separate correction steps, which are applied one after another. In the first step, the impact of unobserved events is considered. In the second step, the contamination of reconstruction artefacts and non-prompt particles is accounted for. In the third step, contributions from pile-up events are removed. In the last step, effects of various sources of inefficiencies are corrected.

The measurement of charged particle multiplicities comprises results in the full kinematic range of the analysis and in bins of pseudorapidity and transverse momentum. In order to limit the length of the analysis description, the procedure is only explained for the full kinematic range. The correction procedure in the kinematic sub-ranges is carried out analogously.

## 9.1. Correction for undetected events

Defining a visible event based on the presence of real particles in an event rather than on reconstructed tracks introduces a fraction of spuriously undetected events. An event that is visible according to the definition based on particles remains undetected if no tracks are reconstructed. These *unobserved* events are most likely to occur when only few charged particles are present within the acceptance of the detector. As discussed in Chap. 5.4, the track reconstruction can fail if particles undergo material interaction or multiple scattering, or due to inefficiencies of the detector and of the reconstruction algorithms. The fraction of unobserved events can be determined in simulation. However, using this fraction to correct data introduces a model dependence. The composition and amount of low-multiplicity events depends on the respective event generator and its settings. Instead, a data-driven approach is adopted in this analysis, which is described in the following.

By definition, the true multiplicity distribution for visible events,  $T(n)$ , where  $n$  denotes the number of charged particles, starts at  $n = 1$ . Since some of these events contain no reconstructed tracks, a different multiplicity distribution  $U(n)$  is observed, which starts from  $n = 0$ . The number of undetected events,  $U(0)$ , cannot be determined directly, as a reconstructed event must always contain at least one track. However, the number of undetected events can be estimated from the observed multiplicity distribution,  $U(n)$ , if the average survival probability,  $\mathcal{P}_{\text{sur}}$ , for a single particle in the kinematic acceptance is known.

The survival probability is determined to be  $\mathcal{P}_{\text{sur}} = 73.60\%$  by using  $N = 500k$  one-particle events in simulation. The statistical uncertainty ( $\sqrt{\mathcal{P}_{\text{sur}}(1 - \mathcal{P}_{\text{sur}})/N}$ ) amounts to  $\Delta_{\text{stat}} = 0.06\%$ . Possible reasons preventing a single particle in the kinematic acceptance from being reconstructed are listed in Tab. 9.1.

To exemplify this idea, events that contain exactly one charged particle in the acceptance,  $T(1)$ , are surveyed. The survival probability of the single track defines the number of events that are reconstructed with one track,  $U'(1) = \mathcal{P}_{\text{sur}} \cdot T(1)$ , and the number of events that remain undetected,  $U'(0) = (1 - \mathcal{P}_{\text{sur}}) \cdot T(1)$ .

Assuming that  $\mathcal{P}_{\text{sur}}$  is independent for two or more particles, the observed multiplicity distribution is approximated in terms of the still unknown actual multiplicity distribution  $T$ ,

$$U(n) = \sum_{l \geq n} \binom{l}{n} \mathcal{P}_{\text{sur}}^n (1 - \mathcal{P}_{\text{sur}})^{l-n} T(l). \quad (9.1)$$

| not reconstructed due to               | fraction      |
|--|---------------|
| any kind of interaction with material  | $\sim 57.9\%$ |
| not within VELO & T-station acceptance | $\sim 19.7\%$ |
| decay in flight                        | $\sim 15.7\%$ |
| inefficiency of track finding          | $\sim 6.7\%$  |

Table 9.1.: Potential sources preventing single particles from being reconstructed. Only visible one-particle events are considered.

All elements  $U(n)$  are expressed by the elements of the true distribution,  $T(l)$ , with  $l \geq n$ . This equation is only valid in the simplified picture that reconstruction artefacts, such as fake and duplicate tracks, which increase the number of observed tracks with respect to true tracks, can be ignored. In this approach, an event containing a certain number of particles is only reconstructed with the same number of tracks or fewer, but not with more tracks. The uncertainties due to these assumptions are evaluated using simulation and are accounted for as systematic uncertainties.

By extracting the first summand of Eq. 9.1, also the elements  $T(n)$  can be expressed by using the corresponding uncorrected measured bin  $U(n)$  and correction terms of  $T(l)$  at higher values of  $l > n$  :

$$T(n) = \frac{U(n)}{\mathcal{P}_{sur}^n} - \sum_{l>n} \binom{l}{n} (1 - \mathcal{P}_{sur})^{l-n} T(l). \quad (9.2)$$

Equation 9.1 allows estimating  $U(0)$  as a sum of the true distribution  $T$ , where the actual elements of  $T$  are recursively expressed in Eq. 9.2. This leads to a recursive formula to calculate  $U(0)$ ,

$$U(0) \approx \sum_{k=1}^r (1 - \mathcal{P}_{sur})^k T(k) \quad \text{with} \quad (9.3)$$

$$T(k) \approx \frac{U(k)}{\mathcal{P}_{sur}^k} - \sum_{l=k+1}^{k+r} \binom{l}{k} (1 - \mathcal{P}_{sur})^{l-k} T(l),$$

which can be calculated numerically up to a given order  $r$ . This expression is solved for different orders using the computer algebra software `mathematica`.

The result is depicted in Fig. 9.1, showing the ratio of undetected events to the total number of detected events. In simulation, the estimated fractions using this calculation,  $f_{unobs}^{est}$ , is compared to the true fraction of  $f_{unobs}^{true} = 3.07\%$ , obtained from using generator information. The estimated fractions show a strong variation for small values of  $r$ , where only few correction terms are considered in the calculation. For  $r \geq 4$  the estimate

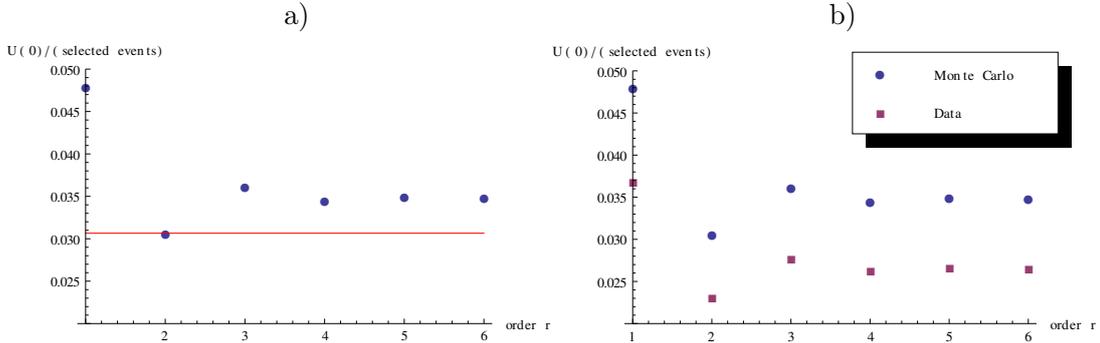


Figure 9.1.: Estimated fraction of undetected events calculated according to Eq. 9.3 for different orders  $r$ . After initial fluctuation, the estimated fraction is stable for  $r \geq 4$ . The true fraction of undetected events in simulation (a) is indicated by the red line. Data results compared to simulation are shown in (b).

| Data type  | Magnetic field | $f_{\text{unobs}}^{\text{est}}$ | $f_{\text{unobs}}^{\text{true}}$ | $f_{\text{unobs}}^{\text{data}}$ |
|------------|----------------|---------------------------------|----------------------------------|----------------------------------|
| Simulation | down           | 3.47%                           | 3.07%                            | –                                |
| Data       | down           | 2.67%                           | –                                | 2.36%                            |
| Simulation | up             | 3.49%                           | 3.07%                            | –                                |
| Data       | up             | 2.65%                           | –                                | 2.34%                            |

Table 9.2.: Fraction of undetected but visible events in simulation and data. The estimated fraction is compared to the true value in simulation and the corrected value in data. The magnet-up and down samples are independent but show very similar results.

becomes stable and results in a value of  $f_{\text{unobs}}^{\text{est}} = 3.47\%$  at order  $r = 6$ . This amounts to a relative systematic bias of

$$\Delta_{\text{method}} = (f_{\text{unobs}}^{\text{est}} - f_{\text{unobs}}^{\text{true}}) / f_{\text{unobs}}^{\text{true}} = 13.0\%, \quad (9.4)$$

which is related to the aforementioned assumptions made in the calculation.

Applying the method to data shows a similar behaviour compared to simulation, see Fig. 9.1 b). The estimated ratios show the same characteristics and become stable when including higher orders. For  $r = 6$  the estimated fraction of unobserved events is  $f_{\text{unobs}}^{\text{est,data}} = 2.67\%$ . The systematic overestimation of the calculation is assumed to be identical for data and simulation. The central value of the estimated fraction is corrected for this effect and then results in

$$f_{\text{unobs}}^{\text{data}} = f_{\text{unobs}}^{\text{est,data}} / (1 + \Delta_{\text{method}}) = 2.36\%. \quad (9.5)$$

The corrected overestimation  $\Delta_{\text{method}}$  of the calculation is considered later as a systematic uncertainty among others.

The so far discussed numbers are derived from the magnet-down data sample. However, the full calculation is carried out separately for data in both magnetic field configurations. The individual sub-samples show very similar results, as it is listed in Tab. 9.2. No uncertainties to  $f_{\text{unobs}}$  are given at this point as a separation of pure statistical and systematic uncertainties cannot be given. The full discussion of uncertainties to  $f_{\text{unobs}}$  is given later, in Chap. 10. The total combined relative uncertainty will be of the order of 15%, corresponding to  $\pm 0.35\%$  absolute uncertainty.

The obtained fractions of undetected events in data of 2.36% (magnet-down) and 2.34% (magnet-up) are added as empty events to the measured multiplicity distributions, and are also considered in the event normalisation of the mean particle density measurements.

## 9.2. Correction for reconstruction artefacts and non-prompt particles

The selected track sample includes three significant categories of impurities. The largest contribution with approximately 6.5% are fake tracks, duplicate tracks contribute with less than 1%. Non-prompt particles that are accidentally selected correspond to a fraction of around 4.5% in the track sample. Henceforth, all impurity categories are collectively

referred to as background tracks. The individual contributions are determined separately using simulation. Starting point for the following discussion is the raw track-sample which consists of all reconstructed and selected tracks. The total number of raw tracks is denoted as  $N_{\text{raw}}$  which still includes all background tracks.

### 9.2.1. Fake tracks

If a reconstructed track does not correspond to the trajectory of a genuine charged particle, the track is considered as a fake (ghost) track, c.f. Chap. 5.4.1. The probability of reconstructing a fake track,  $\mathcal{P}_{\text{fake}}$ , is studied and quantified using simulated events. By exploiting the possibility to check the origin of a reconstructed track in simulation, the probability to create a fake track can be determined as follows:

$$\mathcal{P}_{\text{fake}} = \frac{N_{\text{fake}}}{N_{\text{raw}}}, \quad (9.6)$$

where  $N_{\text{fake}}$  is the number of reconstructed tracks that cannot be matched to a generated particle in the simulation. The ratio is calculated *track weighted*, i.e.  $N_{\text{fake}}$  and  $N_{\text{raw}}$  are first summed over all events before the ratio is build.

The formation of fake tracks has been studied extensively. The probability of reconstructing a fake track is primarily depending on the occupancy of the involved tracking detectors which represent the global activity of an event. Further, track parameters, such as momentum or the flight direction, have a strong impact on the probability of creating an non-physical ghost track.

As discussed in Chap. 5.4, long tracks are formed by using information of the VELO and the T-stations. Thus, the combinatorial probability to find random suitable hits that can form a fake track increases with the hit-multiplicity in each of both detectors. The long track reconstruction algorithm uses already reconstructed VELO tracks as input which are then extrapolated into the T-stations. As a consequence,  $\mathcal{P}_{\text{fake}}$  is most sensitive to the track multiplicity in the VELO and the hit-multiplicity in the T-stations.

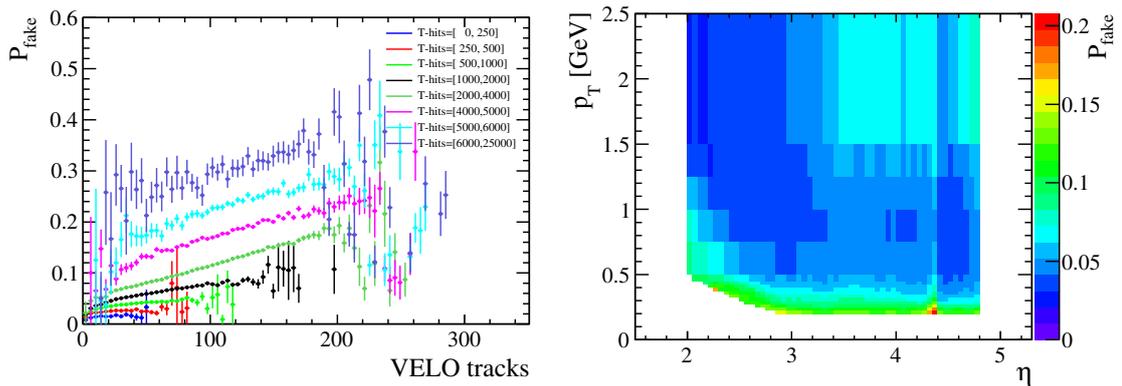


Figure 9.2.: The probability to reconstruct a fake track,  $\mathcal{P}_{\text{fake}}$ , studied as function of event quantities (a) and track parameters (b).  $\mathcal{P}_{\text{fake}}$  shows a strong dependence on global event parameters (hit and track multiplicity) but is also sensitive to individual track parameters such as  $p_T$  and  $\eta$ .

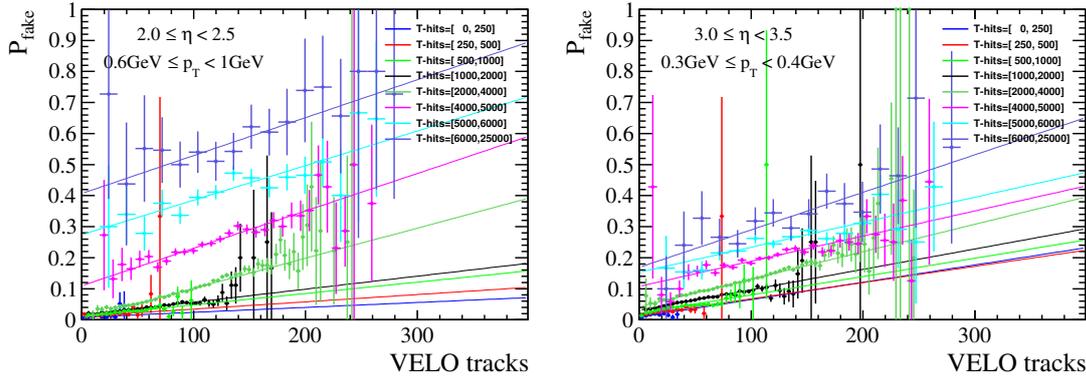


Figure 9.3.: Two exemplary  $(\eta, p_T)$ -bins showing the track and hit multiplicity dependence of  $\mathcal{P}_{\text{fake}}$ . The linear VELO-track multiplicity dependencies are fitted by straight lines.

This increasing probability of  $\mathcal{P}_{\text{fake}}$  with both variables is shown Fig. 9.2 a). A linear increase with VELO-track multiplicity is found in each of the hit-multiplicity bins. While  $\mathcal{P}_{\text{fake}}$  is of the order of a few percent at low hit multiplicities, a drastic increase of up to 40% is found in events with large hit and track multiplicities.

The probability to reconstruct fake tracks further depends on different track parameters. Naturally, a particle with low momentum is more sensitive to multiple scattering and material interaction. Thus, the efficiency to reconstruct particles decreases towards lower momenta. This in turn implies an increasing probability to reconstruct a fake track in the low-momentum region. If a particle scatters or decays before it fully traverses the main tracking stations, the well-reconstructed VELO track points into a T-station region in which no hits induced by the particle can be found. As a consequence, it is no more possible to reconstruct the original particle and the probability of creating a fake track is increased. This argument can be applied for  $p$ ,  $p_T$  and  $\eta$ , however, all three variables are correlated with each other. Hence,  $\mathcal{P}_{\text{fake}}$  is only studied as function of two parameters:  $\eta$  and  $p_T$ , both are presented in Fig. 9.2 b). The histogram shows the expected increase of  $\mathcal{P}_{\text{fake}}$  towards low- $p_T$  but also exhibits a small peaking structure around a pseudorapidity of 4.3 – 4.4. Particles in this  $\eta$ -range have to traverse a larger amount of material, which is related to the beam pipe going through the LHCb detector, c.f. Chap. 4.2. The opening angles that corresponds to this  $\eta$ -range point to a flange that connects two parts of the beam pipe. The larger amount of material locally increases  $\mathcal{P}_{\text{fake}}$  to a value of  $\approx 20\%$  on average.

To account for the contamination of fake tracks in the measurement, a four-dimensional

|                             | Variable                       | Number of bins | Bin range                                    |
|-----------------------------|--------------------------------|----------------|--|
| $\mathcal{P}_{\text{fake}}$ | $\eta$                         | 6              | [2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 4.8]          |
|                             | $p_T$                          | 6              | [0.2, 0.3, 0.4, 0.6, 1.0, 2.0, 100] GeV      |
|                             | $N_{\text{VELO}}^{\text{trk}}$ | 350            | [0, 349]                                     |
|                             | $N_{\text{T}}^{\text{hit}}$    | 7              | [0, 250, 500, 1000, 2000, 4000, 6000, 25000] |

Table 9.3.: Binning scheme used for the determination of fake tracks.

parametrisation of  $\mathcal{P}_{\text{fake}}$  is chosen. The variables are selected in order to provide the best result for the differential measurements of the multiplicity distributions (in bins of  $\eta$  and  $p_T$ ). Therefore, the individual track dependencies are also parametrised as a function of pseudorapidity and transverse momentum. The global event properties are best described by the discussed VELO-track multiplicity,  $N_{\text{VELO}}^{\text{trk}}$ , and the number of hits in the T-stations,  $N_{\text{T}}^{\text{hit}}$ :

$$\mathcal{P}_{\text{fake}} = \mathcal{P}_{\text{fake}}(\eta, p_T, N_{\text{VELO}}^{\text{trk}}, N_{\text{T}}^{\text{hit}}). \quad (9.7)$$

Due to the limited statistical power of the simulation sample, it is beneficial to parametrise one variable by an analytic function. Parametrising the dependence on the VELO-track multiplicity is most practical, since simple first-order polynomials describes the distributions. Examples for the VELO track parametrisation are depicted in Fig. 9.3. The full binning scheme selected for  $\mathcal{P}_{\text{fake}}$  is listed in Tab. 9.3.

### 9.2.2. Duplicate tracks

Another kind of reconstruction artefacts are multiple reconstructed tracks, which are denoted as pairs of duplicate tracks. The probability of reconstructing duplicate tracks,  $\mathcal{P}_{\text{dup}}$ , is estimated as follows:

$$\mathcal{P}_{\text{dup}} = \frac{N_{\text{dup}}}{N_{\text{raw}}}, \quad (9.8)$$

where  $N_{\text{dup}}$  is the number of duplicate tracks found in simulation and  $N_{\text{raw}}$  the corresponding total number of reconstructed tracks. In analogy to fake tracks,  $\mathcal{P}_{\text{dup}}$  is calculated track weighted.

For this analysis, the total contamination of duplicate tracks is small (less than 1%). However, a certain kinematic range is found to be particular sensitive to the creation of these kind of artefacts. The probability  $\mathcal{P}_{\text{dup}}$  has been studied as function of various kinematic variables. In Fig. 9.4 a) the  $(\eta, p_T)$ -plane shows a distinct region for creating duplicate tracks. In a pseudorapidity range of  $\eta = [3.4 - 4.1]$ , the fraction of duplicate tracks is significantly increased. Towards low transverse momenta,  $\mathcal{P}_{\text{dup}}$  reaches a maximum of  $\approx 3.5\%$ . It is tested that pairs of duplicate tracks in this range share identical T-hits but have a different VELO-track part. Both tracks show a small separation in pseudorapidity. The different VELO parts are identified to originate from a splitting of a single particle's trajectory, where half of the hits are used to form two very similar tracks.

The creation of duplicate tracks is also tested against global event properties, such as hit multiplicity in different sub-detectors or track multiplicity of different track types. In contrast to fake tracks, only a weak dependence is found. The average fraction of clone tracks varies from around 0.5% to 1.2% in events with a low and high number of tracks, respectively. To account for the light variation,  $\mathcal{P}_{\text{dup}}$  is determined in bins of VELO-track multiplicity. An additional dependence on hit multiplicity is found to be negligible and is thus not parametrised. In total, a three-dimensional parametrisation of  $\mathcal{P}_{\text{dup}}$  is adopted:

$$\mathcal{P}_{\text{dup}} = \mathcal{P}_{\text{dup}}(\eta, p_T, N_{\text{VELO}}^{\text{trk}}). \quad (9.9)$$

The full binning scheme is listed in Tab. 9.4.

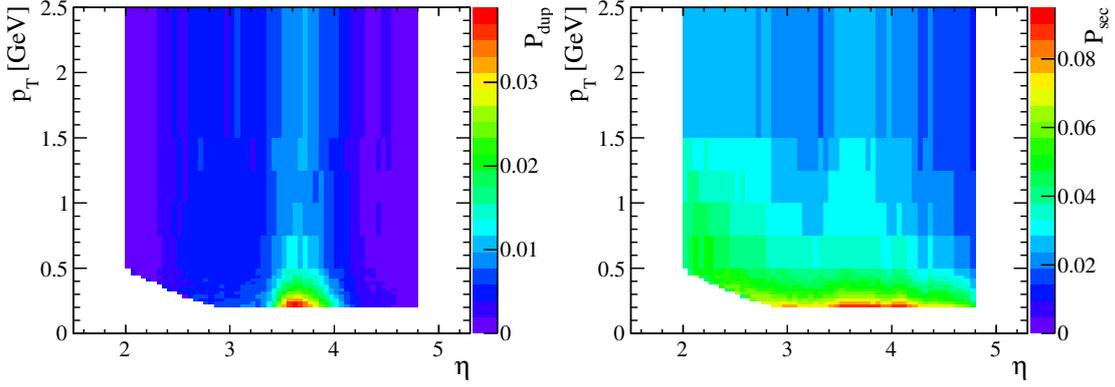


Figure 9.4.: The probability to reconstruct duplicate tracks (a), and to select non-prompt particles (b). Both probabilities are studied in simulation and are presented as function of  $\eta$  and  $p_T$ .

### 9.2.3. Non-prompt tracks

In addition to reconstruction artefacts also mis-selected secondary particles contaminate the track sample of prompt charged particles. The probability that a non-prompt particle is selected,  $\mathcal{P}_{\text{sec}}$ , is estimated in simulation:

$$\mathcal{P}_{\text{sec}} = \frac{N_{\text{sec}}}{N_{\text{raw}}}, \quad (9.10)$$

where  $N_{\text{sec}}$  is the number of selected secondary particles. The ratio is again determined as track weighted. The predominant contribution to  $\mathcal{P}_{\text{sec}}$  is due to material interaction, such as photon conversion, and depends on the amount of material traversed in the detector. Low- $p_T$  particles are affected more strongly. This can be seen in Fig. 9.4 b) where  $\mathcal{P}_{\text{sec}}$  is shown as a function of  $\eta$  and  $p_T$ .

The accidentally selected secondary particles can be classified in four categories, depending on their origin. As mentioned previously, the largest contribution of  $\approx 40\%$  is due to photon conversion and other material interaction. Another  $\approx 32\%$  are pions produced in  $K_s^0$  decays, followed by  $\approx 20\%$  due to  $\Lambda$  and hyperon decay products. The remaining 9% are secondary particles from various particle decays.

As for duplicate tracks, also the probability that a non-prompt particle passes the track selection is almost independent of the hit or track multiplicity of the events. However, in order to account for slight variations of  $\mathcal{P}_{\text{sec}}$  between low and high multiplicity events and to have a consistent binning scheme, different bins of VELO-track multiplicity are considered. In total,  $\mathcal{P}_{\text{sec}}$  is estimated as a function of the same variables as for duplicate tracks:

$$\mathcal{P}_{\text{sec}} = \mathcal{P}_{\text{sec}}(\eta, p_T, N_{\text{VELO}}^{\text{trk}}). \quad (9.11)$$

The binning scheme (c.f. Tab. 9.4) is identical to that of duplicate tracks.

Different methods are applied to correct the observed particle densities and particle multiplicities. However, the estimated probabilities of contaminating background in the track sample are the same. The two procedures are explained in the following.

### 9.2.4. Background subtraction - mean particle densities

For each track, a combined impurity probability,  $\mathcal{P}_{\text{bkg}}$ , is calculated, which is the sum of the three contamination types:

$$\mathcal{P}_{\text{bkg}} = \mathcal{P}_{\text{fake}} + \mathcal{P}_{\text{dup}} + \mathcal{P}_{\text{sec}}. \quad (9.12)$$

The combined probability then also depends on kinematic properties of the track,  $\eta$  and  $p_{\text{T}}$ , the VELO-track multiplicity, and the hit multiplicity of the T-stations.

When measuring particle densities, it is sufficient to assign a per-track weighting factor,  $\omega$ , that accounts for the impurities mentioned above:

$$\omega(\eta, p_{\text{T}}, N_{\text{VELO}}^{\text{trk}}, N_{\text{T}}^{\text{hit}}) = (1 - \mathcal{P}_{\text{bkg}}). \quad (9.13)$$

In Fig. 9.5 a,b), simulated  $dn/d\eta$  and  $dn/dp_{\text{T}}$  particle densities are displayed for reconstructed raw data and after applying the purity weighting factors. The  $dn/d\eta$  raw distribution shows a decreasing density towards small pseudorapidities which is related to the acceptance of the detector. The largest density is found at  $\eta \approx 3.5$  followed by a characteristic dip induced by the beam pipe flange. After applying the weighting factors, naturally the overall particle density is decreased but no additional patterns are induced. The weighted distribution is compared to a simulated distribution free of background tracks: By using information of the generator, reconstruction artefacts and non-prompt particles are identified and removed. A reasonable agreement between the weighted and the true distribution is obtained. The remaining differences that can be seen in the pull distribution are related to the binning scheme of the correction factors. Only statistical uncertainties are considered for the pull calculation, systematic uncertainties related to the parametrisation and the binning scheme are larger and are addressed later.

The corresponding  $dn/dp_{\text{T}}$  distribution does not indicate notable features. Comparing the raw distribution with the weighted distribution shows a stronger change of the density at low- $p_{\text{T}}$ , where a larger amount of background tracks is found. As for pseudorapidity, a reasonable agreement between the weighted and true distribution is obtained. Fluctuations in the pull distribution represent the bin sizes used for the determination of the correction factors, in particular that of fake tracks.

The correction is applied identically to measured data, the result is displayed in Fig. 9.5 c,d). The measured raw densities in data are more pronounced at low- $p_{\text{T}}$  compared to the simulated densities. As expected, the effect of the purity weighting is similar for data and simulation.

|  | Variable                       | Number of bins | Bin range  |
|--|--------------------------------|----------------|--|
| $\mathcal{P}_{\text{dup}}, \mathcal{P}_{\text{sec}}$ | $\eta$                         | 56             | [2.0, 4.8], $\Delta\eta = 0.05$  |
|  | $p_{\text{T}}$                 | 13             | [0.2, 0.5] with $\Delta p_{\text{T}} = 0.05$ GeV, and<br>[0.5, 0.75, 1.0, 1.25, 1.5, 2.5, 5.0, 15] GeV |
|  | $N_{\text{VELO}}^{\text{trk}}$ | 5              | [1, 7, 15, 30, 60, 350]  |

Table 9.4.: Combined binning scheme used for the determination of duplicate tracks and mis-selected secondary particles.

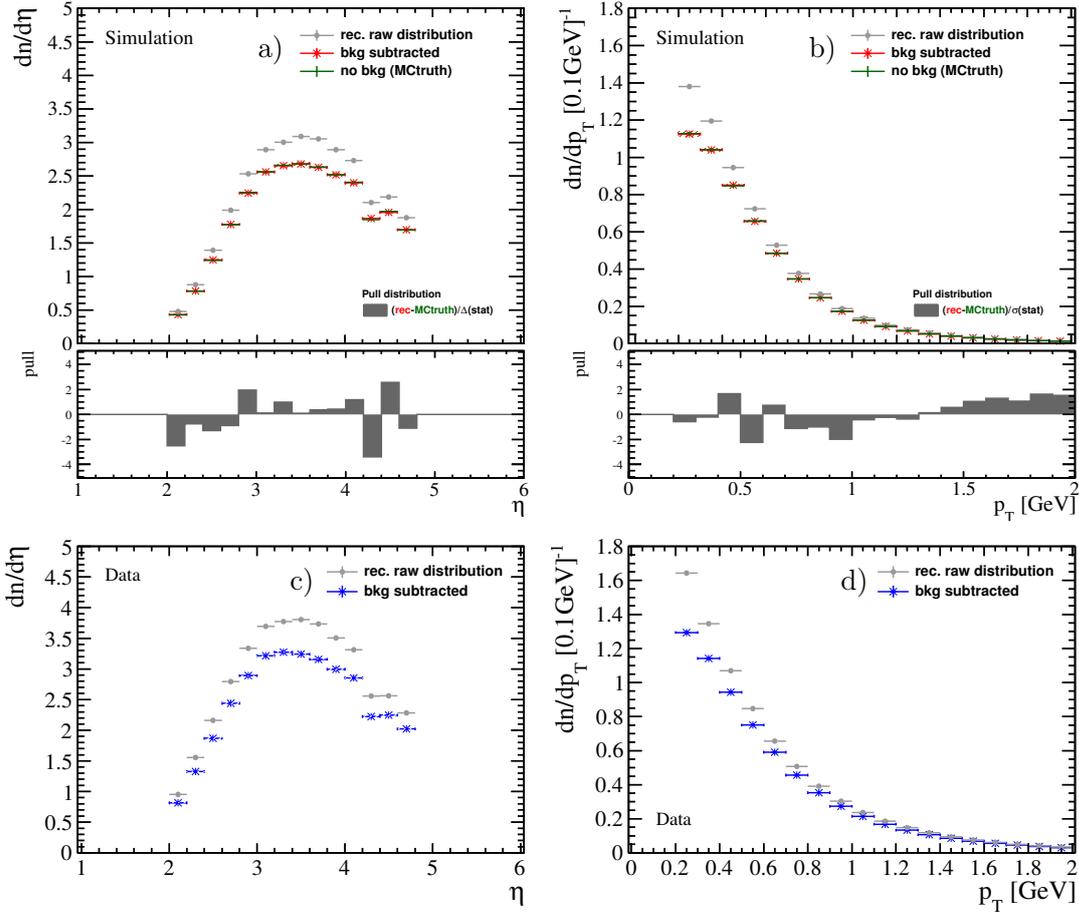


Figure 9.5.: Background correction applied to particle densities,  $dn/d\eta$  (left) and  $dn/dp_T$  (right). Simulated distributions (a,b) compare the raw track sample (grey), the same sample after applying purity weighting factors, and a sample free of background by using generator information. The corresponding pull distribution is given below. Data distributions (c,d) are compared w/ and w/o correction. All errors are statistical.

### 9.2.5. Background subtraction - multiplicity distributions

Correcting particle multiplicity distributions in the same way by applying per-track weighting factors would lead to non-physical fractional event multiplicities. The procedure to obtain background-subtracted multiplicity distributions is described below. To simplify the discussion, the description is explained for the full kinematic range, but the procedure is performed in each of the  $\eta$  and  $p_T$  sub-ranges separately.

The impurity probability,  $\mathcal{P}_{\text{bkg},i}$ , of each track  $i$ , is summed for all tracks in an event,  $n_{\text{ev}}$ , to obtain a total event impurity correction,  $\mu_{\text{ev}}$ :

$$\mu_{\text{ev}} = \sum_{i=1}^{n_{\text{ev}}} \mathcal{P}_{\text{bkg},i}, \quad \text{with} \quad \Delta\mu_{\text{ev}} = \sqrt{\sum_{i=1}^{n_{\text{ev}}} (\Delta\mathcal{P}_{\text{bkg},i})^2}, \quad (9.14)$$

where  $\Delta\mu_{\text{ev}}$  is the statistical uncertainty obtained from the uncertainties of the combined

impurity probability,  $\Delta\mathcal{P}_{\text{bkg},i}$ , of each track. The obtained mean number of background tracks,  $\mu_{\text{ev}}$ , permits to calculate the probability to reconstruct a certain number of background tracks in each event, assuming Poisson statistics. The number of observed background tracks,  $k$ , in an event then obeys the probability distribution

$$\mathcal{P}_{\text{bkg}}(\mu_{\text{ev}}, k) = \frac{\mu_{\text{ev}}^k}{k!} e^{-\mu_{\text{ev}}}. \quad (9.15)$$

From this relation a normalised probability distribution,  $\mathcal{P}_{\text{clean}}$ , is derived that an event contains a given number of prompt charged particles ( $n_{\text{ev}} - k$ ):

$$\mathcal{P}_{\text{clean}}(n_{\text{ev}} - k) = \frac{1}{I} \mathcal{P}_{\text{bkg}}(\mu_{\text{ev}}, k). \quad (9.16)$$

The normalisation factor  $I = \sum_{k=0}^{n_{\text{ev}}} \mathcal{P}_{\text{bkg}}(\mu_{\text{ev}}, k)$  ensures a total probability of one.

In Fig. 9.6, the idea of the background correction is visualised for a single event. Assuming a raw event with  $n_{\text{ev}} = 6$  reconstructed tracks contributes to the raw distribution with a single entry of weight one. Further assuming a mean background correction of  $\mu_{\text{ev}} = 1.1$  results in a probability distribution of background corrected events, which covers a multiplicity range of zero to six tracks. The total weight of the distribution remains one.

By summing the normalised probability distribution of all events a background corrected multiplicity distribution is obtained. The results of the method applied to events in simulation and in data are shown in Fig. 9.7. Simulated and measured multiplicities cover around six orders of magnitude. The uncorrected multiplicity distribution populates larger multiplicities compared to the background corrected distribution. For comparison, also the background free distribution, obtained by using generator information, is included in Fig. 9.7 a). As a result of the Poisson correction, a tail towards large multiplicities remains. Events with a large number of background tracks still have residual probabilities of being composed of good tracks only. Thus, the range beyond 50 tracks is ignored for further considerations.

The pull distribution in Fig. 9.7 a) shows that the correction is not perfect. The reason for the observed discrepancy is attributed to the used correction values and not to the Poisson method itself. Testing the method with ideal correction values using generator

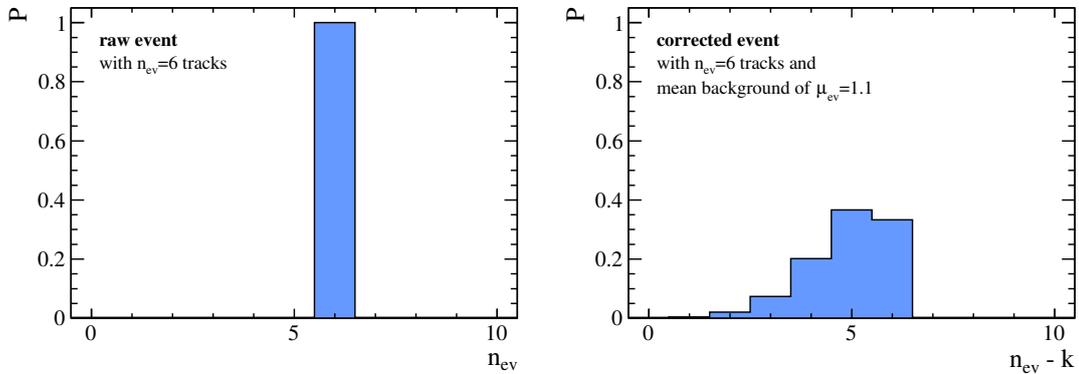


Figure 9.6.: Example showing the statistical correction of background tracks in a single event.

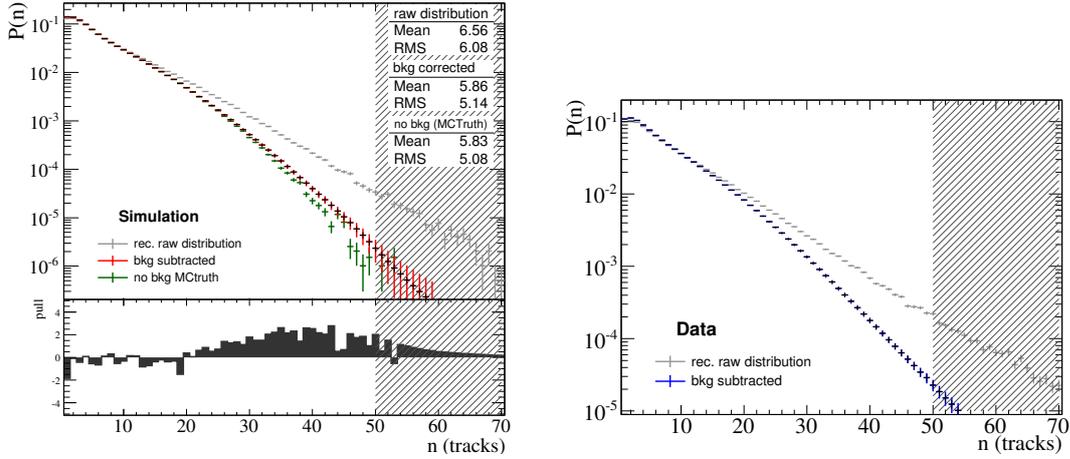


Figure 9.7.: Background correction applied to data and simulation.

information results in a perfect agreement between true and corrected distribution. The limitation of the four-dimensional parametrisation is reflected in the correction factors and propagates into the background corrected multiplicity distribution. However, the error due to this correction is not the limiting uncertainty in the final results.

The corrected data multiplicity distribution is depicted in Fig. 9.7 b). There are no apparent differences visible compared to simulation except for a larger population towards high multiplicities.

### 9.3. Pile-up correction

As mentioned in Chap. 7, the average number of  $pp$  interactions per bunch crossing in the selected data taking period is small and results in a limited bias from pile-up interactions. The probability,  $p_\mu(k)$ , to find a certain number of  $pp$  collisions,  $k$ , in a single bunch crossing event is described by a Poisson distribution,

$$p_\mu(k) = e^{-\mu} \frac{\mu^k}{k!} \quad \text{with} \quad \sum_{k=0}^{\infty} k p_\mu(k) = \mu. \quad (9.17)$$

The mean value,  $\mu$ , represents the average number of  $pp$  collisions in a single event. If suppressing events with zero interactions, this average number becomes

$$\mu_+ = \frac{\sum_{k=1}^{\infty} k p_\mu(k)}{\sum_{k=1}^{\infty} p_\mu(k)} = \frac{\mu}{1 - e^{-\mu}}. \quad (9.18)$$

For  $\mu \ll 1$ , Eq. 9.18 can be expanded and results in

$$\mu_+ = \frac{\mu}{1 - e^{-\mu}} \approx \frac{\mu}{1 - (1 - \mu + \mu^2/2)} \approx 1 + \frac{\mu}{2}, \quad (9.19)$$

and accordingly

$$(\mu_+)^{-1} \approx 1 - \frac{\mu}{2}. \quad (9.20)$$

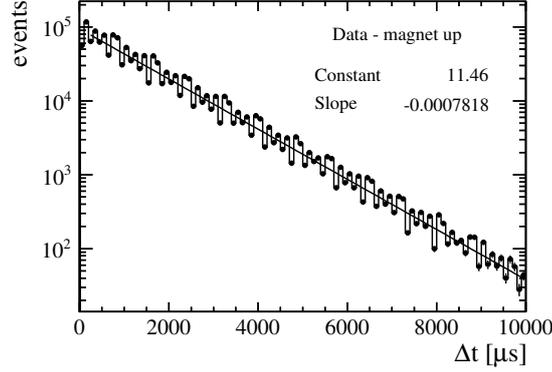


Figure 9.8.: Time difference  $\Delta t$  between consecutive recorded  $pp$  collisions with at least one reconstructed in the acceptance.

Experimentally,  $\mu$  can be deduced from the distribution of time differences  $\Delta t$ , between consecutive recorded events. Figure 9.8 shows the  $\Delta t$ -distribution for recorded events of the magnet-up data sample, which consists of only one LHC fill. The distribution follows an exponential function,

$$P \propto e^{-\mu \cdot b \cdot f_{\text{LHC}} \cdot \Delta t}, \quad (9.21)$$

where  $f_{\text{LHC}} = 11.246$  kHz is the revolution frequency of the LHC, and  $b$  is the number of colliding proton bunches in a specific LHC fill. By fitting an exponential function to the  $\Delta t$ -distribution, the average number of collisions per event can be estimated. In the magnet-up sample, which is shown in Fig. 9.8, only  $b = 1$  colliding bunch was present. This results in an average number of  $pp$  collisions of  $\mu = 0.0695$ . For the magnet-down sample, with  $b = 2$  colliding bunches<sup>1</sup>, a value of  $\mu = 0.0803$  is found, accordingly.

With the obtained values of  $\mu$  for each of the two data sets, the pile-up contribution in the measured Physics observables can be afterwards accounted for. Two different approaches [112] are adopted to correct the mean particle densities and the multiplicity distributions.

**Correction to particle densities:** The pile-up in the measured particle densities is corrected by re-normalising the measurement to the actual number of  $pp$  collisions instead of the number of recorded events. The number of particles for single  $pp$  collisions, free of pile-up, is henceforth denoted as  $N_{pp}$ . The observed (uncorrected) number of particles,  $N_{\text{obs}}$ , is described by

$$N_{\text{obs}} = \frac{1}{\sum_{k=1}^{\infty} p_{\mu}(k)} \sum_{k=1}^{\infty} k p_{\mu}(k) N_{pp} = \frac{\mu}{1 - e^{-\mu}} N_{pp}, \quad (9.22)$$

according to the probability to find events with  $k \geq 1$   $pp$  collisions. Rearranging Eq. 9.22 and using Eq. 9.20 leads to a simple relation between the number of particles

<sup>1</sup>The magnet-down data sample, which was recorded after the magnet-up data sample, has a larger number of colliding bunches, since the number of bunches in the LHC was steadily increasing with time.

in single-collision and recorded bunch-crossing events:

$$N_{pp} \approx \left(1 - \frac{\mu}{2}\right) N_{\text{obs}}. \quad (9.23)$$

This resulting factor is applied to the observed particle densities to normalise the measurement to the actual number of  $pp$  collisions.

**Correction to particle multiplicities:** In the presence of pile-up, the observed multiplicity distribution,  $M_{\text{obs}} = M_{\text{obs}}(n)$ , is composed of

$$M_{\text{obs}} = \frac{1}{\sum_{k=1}^{\infty} p_{\mu}(k)} \sum_{k=1}^{\infty} p_{\mu}(k) S_k \quad \text{with} \quad S_k = \underbrace{S_{pp} \otimes S_{pp} \dots \otimes S_{pp}}_{k \text{ times}}, \quad (9.24)$$

where  $S_k$  is the convolution of  $k$  single- $pp$ -collision multiplicity distributions,  $S_{pp}$ . In the limit of  $\mu \ll 1$ , it is sufficient to consider only the terms up to the order of  $k = 2$  in Eq. 9.24 and obtains:

$$M_{\text{obs}} \approx \frac{p_{\mu}(1)S_1 + p_{\mu}(2)S_2}{p_{\mu}(1) + p_{\mu}(2)} = \frac{S_1 + (\mu/2)S_2}{1 + (\mu/2)}, \quad (9.25)$$

with using  $p_{\mu}(2)/p_{\mu}(1) = \mu/2$ . Solving Eq. 9.25 for the single-collision distribution  $S_1$  results in an iterative calculation rule:

$$S_1^{(i+1)} = \left(1 + \frac{\mu}{2}\right) M_{\text{obs}} - \frac{\mu}{2} S_2^{(i)} \quad \text{with} \quad S_2^{(i)} = S_1^{(i)} \otimes S_1^{(i)}, \quad (9.26)$$

where the index  $i$  denotes the  $i$ -th iteration step. The distribution  $S_1^{(i+1)}$  is the improved iteration of the previous estimate  $S_1^{(i)}$ . Starting with the observed distribution as the  $i = 0$  estimate, this leads to

$$\begin{aligned} S_1^{(0)} &= M_1 = M_{\text{obs}}, \\ S_1^{(1)} &= M_1 + \frac{\mu}{2}(M_1 - M_2), \\ S_1^{(2)} &= M_1 + \frac{\mu}{2}(M_1 - M_2) - \frac{\mu^2}{2}(M_2 - M_3) + \frac{\mu^3}{8}(M_2 - 2M_3 + M_4), \end{aligned} \quad (9.27)$$

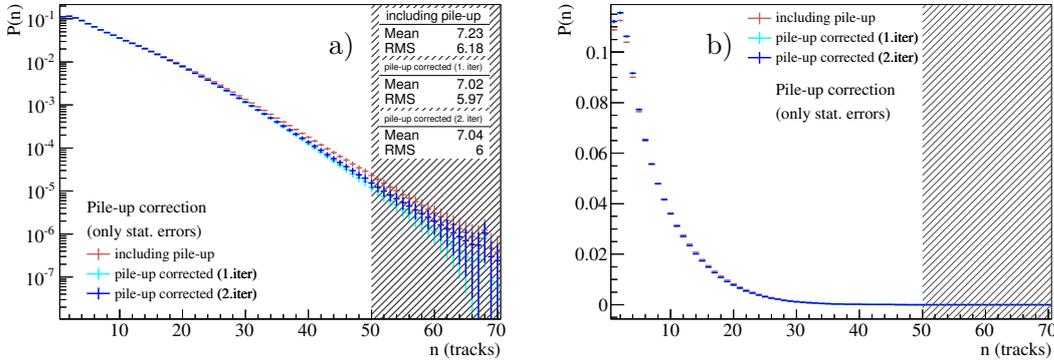


Figure 9.9.: Impact of the pile-up correction to the particle multiplicity distribution in the full kinematic range. Displayed are data results in logarithmic (a) and linear (b) scale.

where  $M_k$  is again identified as the convolution of  $k$  observed distributions,  $M_{\text{obs}}$ . The third iteration,  $S_1^{(3)}$ , would affect only the  $\mu^3$ -term. Since the average number of interactions in the used data sample is below 0.1, the correction due to the third iteration is already  $< 1\text{‰}$  and is not further considered.

The impact of the iterative pile-up correction to the multiplicity distribution in the full kinematic range is presented in Fig. 9.9. The correction is applied to the background corrected multiplicity distribution, described in the previous section. The impact of the pile-up correction is small, the largest effect is found towards low multiplicities. It is further visible that already the second iteration is almost negligible, since the mean value of the distribution is changed by less than a percent. The shaded multiplicity region,  $n > 50$ , identical to that of the background correction, is not considered in the following data unfolding. The charged particle multiplicity distributions in  $\eta$  and  $p_T$  sub-ranges are corrected by the same method as described for the full kinematic range.

## 9.4. Efficiency correction and unfolding procedure

In the fourth and last correction step, the background and pile-up corrected measurements are further corrected for inefficiencies that prevent charged particles from being reconstructed. As a reminder, only particles that fulfil the kinematic requirements of  $2.0 < \eta < 4.8, p > 2 \text{ GeV}$  and  $p_T > 0.2 \text{ GeV}$  are considered. The following correction accounts for limited efficiencies due to the detector acceptance ( $\epsilon_{\text{acc}}$ ) within this kinematic range and the track reconstruction ( $\epsilon_{\text{tr}}$ ). Both effects are evaluated separately, but only the product of both efficiencies,  $\epsilon_{\text{acc}} \cdot \epsilon_{\text{tr}}$ , represents the actual reconstruction efficiency in the kinematic range. Afterwards, the correction methods are discussed, again two different procedures must be applied to particle densities and to particle multiplicity distributions.

### 9.4.1. Detector acceptance

For particles fulfilling the kinematic requirements, the detector acceptance  $\epsilon_{\text{acc}}$  describes the fraction of particles that reach the end of the downstream (main) tracking stations and that neither interact with material nor are deflected out of the detector by the magnetic field. The efficiency  $\epsilon_{\text{acc}}$  is calculated in simulation by

$$\epsilon_{\text{acc}} = \frac{N_{\text{rb}}}{N_{\text{true}}}, \quad (9.28)$$

where  $N_{\text{true}}$  are all primary charged particles that fulfil the kinematic requirements of the analysis, and  $N_{\text{rb}}$  represents the number of particles that also fulfil the reconstructible requirements for long tracks, as defined in Chap. 5.4.1. In Fig. 9.10 a) the detector acceptance is studied as function of  $\eta$  and  $p_T$ . The averaged detector acceptance is approximately 73%. In the medium- and high- $p_T$  range the efficiency is  $\approx 80\%$ . At low- $p_T$ , particles are deflected more strongly by the magnetic field. As a result, there is a higher probability that the particles are bent out of the geometric acceptance of the detector. This probability is also increased for particles that are produced with pseudorapidities close to the detector edges.

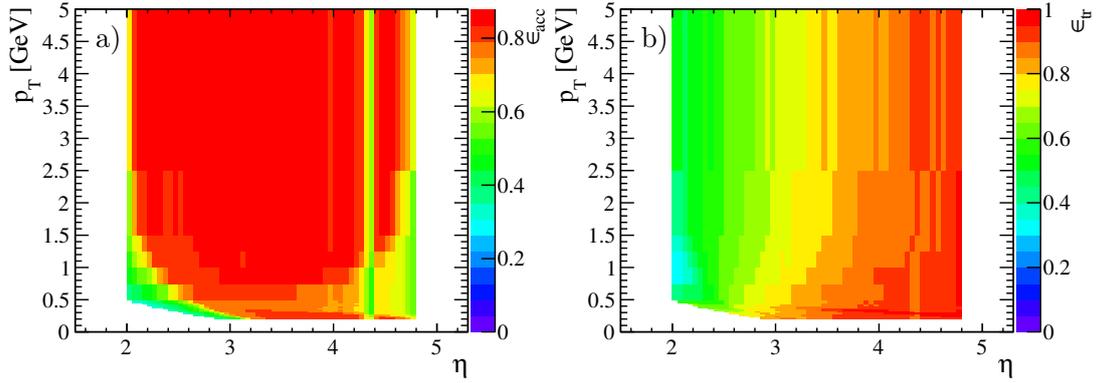


Figure 9.10.: Efficiency maps determined in simulation. The detector acceptance (a) and track finding efficiency (b) are presented as a function of  $\eta$  and  $p_T$ .

Furthermore, Fig. 9.10 a) shows a distinct drop in the acceptance map at  $\eta \approx 4.3 - 4.4$ . Particles in this  $\eta$ -range experience a larger probability to undergo multiple scattering and material interaction when traversing the larger amount of material that is related to the beam pipe flange.

The detector acceptance is also tested for dependencies on global event properties. No significant hit or track multiplicity dependence is found as the acceptance is predominantly defined by the design of the detector.

#### 9.4.2. Track finding efficiency

In addition to the given detector acceptance, the efficiency of the track finding algorithms is of importance for the total performance of the reconstruction. The track finding efficiency is defined for selected particles that fulfil the reconstructible criterion for long tracks, i.e. particles within the geometric acceptance of the VELO and the T-stations. The track finding efficiency,  $\epsilon_{\text{tr}}$ , is determined in simulation and calculated from the ratio of selected and reconstructed tracks<sup>2</sup>,  $N_{\text{rd}}$ , with respect to reconstructible particles:

$$\epsilon_{\text{tr}} = \frac{N_{\text{rd}}}{N_{\text{rb}}}. \quad (9.29)$$

The definition of  $\epsilon_{\text{tr}}$  is chosen such that the detector acceptance can be evaluated separately. Only the product of both efficiencies gives the total reconstruction efficiency of the selected primary charged particles.

The track finding efficiency is displayed as a function of  $\eta$  and  $p_T$  in Fig. 9.10 b). The efficiency shows a clear dependence as a function of pseudorapidity. The decrease towards lower  $\eta$  is identified as an indirect effect related to the increasing amount of low-momentum particles in this range, which are more prone to multiple scattering than particles with large momentum. The histogram also shows a kinematic cut-off that is visible at small values of  $\eta$  and  $p_T$ . The minimum- $p$  requirement in the selection forbids particles in this region.

<sup>2</sup>in contrast to the standard track reconstruction, only one of the two long track reconstruction algorithms is used in this analysis, the forward tracking.

The track finding efficiency is evaluated as function of  $\eta$  and  $p_T$  for the measurement of mean particle densities. For the measurement of the particle multiplicity distributions, an unfolding procedure is adopted which does not account for individual track properties but global event reconstruction performances. Both procedure are explained in the following.

### 9.4.3. Efficiency correction for particle densities

To correct average observables, such as the measured particle densities, it is adequate to use average correction factors. The limited efficiencies related to the detector acceptance and the track finding are corrected by assigning per-track weighting factors,  $\omega_{\text{eff}}$ , in analogy to the background correction in Sec. 9.2.4. The weighting factors are composed of the product of the detector acceptance and the track finding efficiency, which are both determined as function of  $\eta$  and  $p_T$ :

$$\omega_{\text{eff}}(\eta, p_T) = \frac{1}{\epsilon_{\text{acc}}(\eta, p_T)\epsilon_{\text{tr}}(\eta, p_T)}. \quad (9.30)$$

Each track is assigned an individual weighting factor which enters the particle density measurements.

The performance of the efficiency correction is depicted in Fig. 9.11 for the particle density measurements  $dn/d\eta$  and  $dn/dp_T$ , using simulated data. Both figures show a large correction due to the efficiency weighting. In particular the low- $\eta$  range is strongly affected. The corrected reconstructed distributions agree with the generated distributions within the combined statistical error of all applied corrections. The pull distribution of the  $dn/dp_T$  measurement shows a moderate discrepancy in the range of

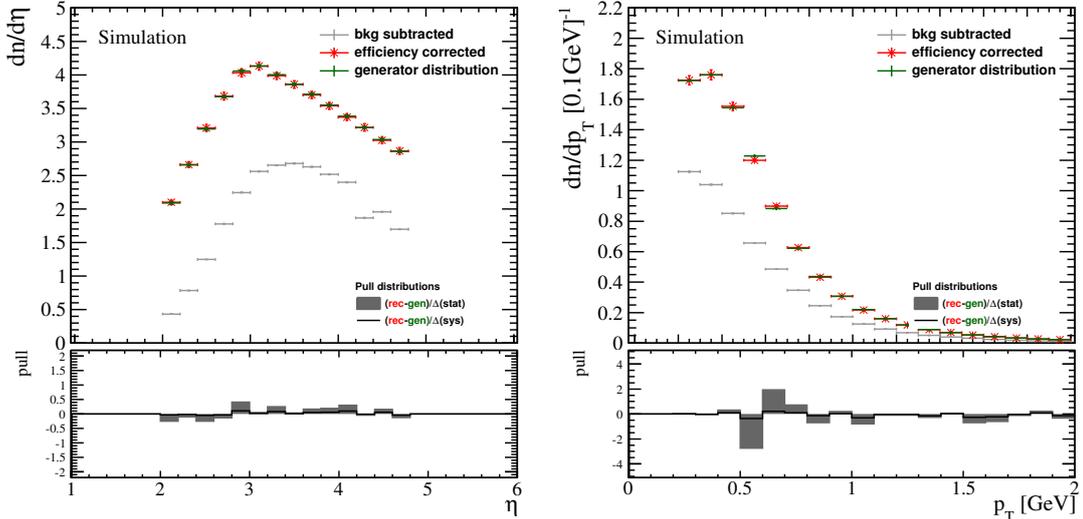


Figure 9.11.: Efficiency correction applied to background corrected simulated particle densities,  $dn/d\eta$  (a) and  $dn/dp_T$  (b). Efficiency weighted distributions (red) are compared to generated particle densities (green), the corresponding pull distributions are given below. The previous background corrected distributions are shown in grey.

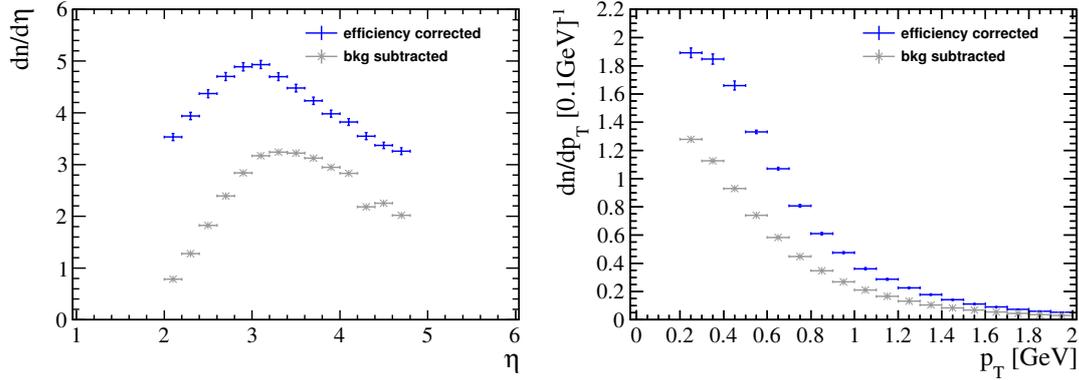


Figure 9.12.: Efficiency correction applied to measured particle densities,  $dn/d\eta$  (a) and  $dn/dp_T$  (b). The previous background corrected distributions are shown in grey, the efficiency weighted distributions are shown in blue. The corrected distributions also include the pile-up considered normalisation. Only statistical errors are given.

$p_T \approx 0.5 - 0.7$  GeV which reflects the chosen finite bin sizes of the correction factors. Apart from that, the generator distributions are well reproduced.

The effect of the efficiency correction to data is presented in Fig. 9.12, so far only statistical errors are given. Besides the generally larger particle densities, no noticeable differences to simulation are visible. The  $dn/d\eta$  distribution shows that the decrease induced by the beam pipe material ( $\eta \approx 4.3$ ) is sufficiently compensated.

#### 9.4.4. Unfolding of multiplicity distributions

The observed particle multiplicity distributions must be corrected for inefficiencies due to the detector acceptance and the track finding. Weighting the observed multiplicity of an event according to an efficiency factor, as adopted for the particle density measurement, is not possible, since this leads to non-integer multiplicities. In order to infer the true particle multiplicity distribution from the observed one, an unfolding technique based on a detector response matrix is employed.

The response matrix,  $R$ , is defined in such a way that it only accounts for inefficiencies due to the detector acceptance and the track finding, not for reconstruction artefacts and mis-selected tracks. The matrix is constructed from the relation between the distribution of true prompt charged particles,  $T(n)$ , and the distribution of measured tracks,  $M(m)$ , subtracted for background and pile-up:

$$M(m) = \sum_n R_{m,n} T(n). \quad (9.31)$$

Simulated events are used to determine the matrix. The generated number of charged particles per event,  $n \geq 1$ , is compared to the corresponding number of reconstructed and background subtracted number of tracks,  $m \geq 0$ . As a consequence, each possible value of the generated particle multiplicity ( $n$ ) is mapped to a distribution of reconstructed tracks ( $m$ ). The obtained detector response matrix for the full kinematic range is depicted in Fig. 9.13.

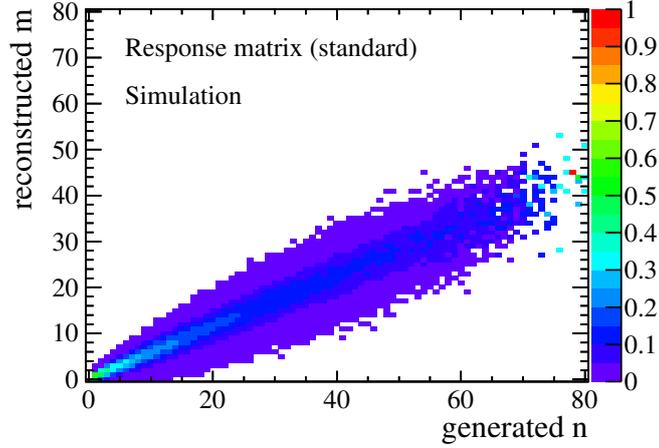


Figure 9.13.: The detector response matrix in the full kinematic range. Simulated events are used to map the reconstructed and background corrected track multiplicity distribution to the generated distribution of prompt charged particles.

The response matrix has been normalised such that the values of each column sum up to one. As a result, the entries of a given column represent the probabilities that an event with  $n$  prompt charged particles is reconstructed with a certain number of  $m$  (background subtracted) tracks.

The matrix is limited by the available event statistics of the simulation sample. For very high multiplicities, the number of events is not sufficient to populate the entire matrix. To avoid this statistical limitation, the matrix is parametrised in order to extrapolate the entries to higher multiplicities.

The parametrisation of the response matrix is explained on the basis of Fig. 9.14. The mapping of a certain true multiplicity  $n$  to a distribution of reconstructed tracks  $m$  is well described by a Gaussian distribution with mean value  $\bar{m}$  and standard deviation  $\sigma_m$ , four examples are presented in (a-d). The distributions of  $\bar{m}$  and  $\sigma_m$  as a function of  $n$  can be parametrised by combinations of polynomial and logarithmic functions, as displayed in (e-f). The model parameters are fitted in a range of high statistics,  $n = [4, 65]$ . In order to suppresses the effect of statistical fluctuations in the entries of the response matrix, the matrix is repopulated for multiplicities  $n > 40$ , according to the Gaussian distributions with the extrapolated set of parameters. This further allows the detector response matrix (g) to be extrapolated up to large values of  $n$ .

The true charged particle multiplicity distribution  $T(n)$  can be formally calculated from the measured distribution  $M(m)$ , by using the inverse of the response matrix,  $R^{-1}$ :

$$T(n) = \sum_m R_{n,m}^{-1} M(m). \quad (9.32)$$

However, it is in general not possible to determine the inverse matrix, since the matrix  $R$  can be singular (degenerated). By assuming that  $R$  is invertible and  $R^{-1}$  is computed, the true multiplicity distribution still cannot be extracted by using Eq. 9.32. As discussed in literature, see *e.g.* Refs [113, 114], the obtained solution typically has severe non-

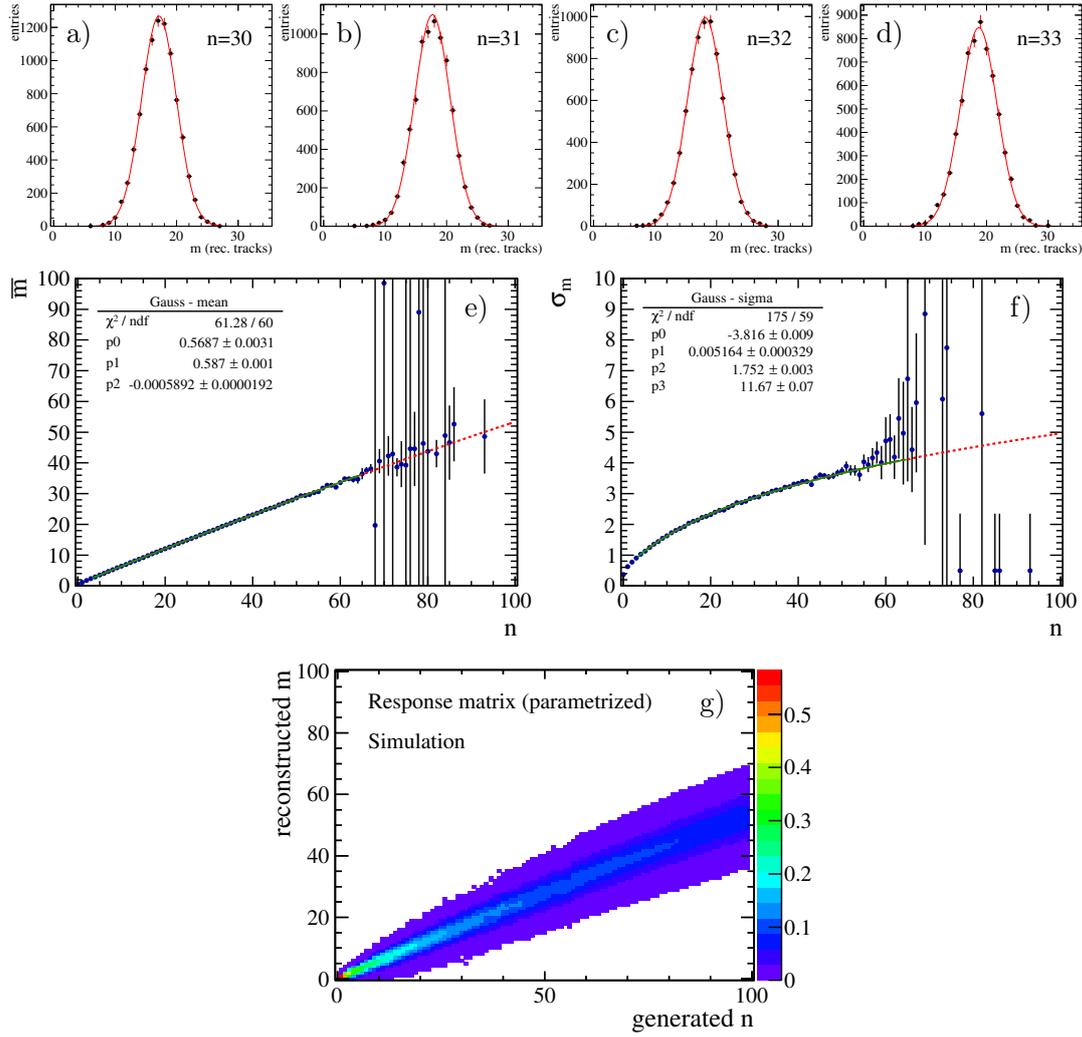


Figure 9.14.: Parametrising the detector response matrix. The distribution of reconstructed tracks  $m$  for a given true multiplicity  $n$  is described by a Gaussian distribution, as exemplified for  $n = [30, 33]$  (a-d). The obtained mean values (e) are well described by a second-order polynomial, the standard deviations (f) by a sum of a first-order polynomial and a logarithmic function. Fit ranges are represented by solid green lines, extrapolations by dotted red lines. By using these extrapolated parameters the response matrix is extended to large values of  $n$  (g).

physical oscillations. Resolution effects of the detector response and small statistical fluctuations of the measurement directly lead to strong non-physical fluctuations of the unfolded solution. Thus the matrix inversion cannot be applied to unfold the original true distribution.

One possible way to avoid fluctuations due to information lost during the detection is to parametrise the solution. In this analysis, an approach is chosen that further does not require an inversion of the response matrix. The true charged particle multiplicity distribution  $T(n)$  is extracted from the measured distribution  $M(m)$  by using an un-

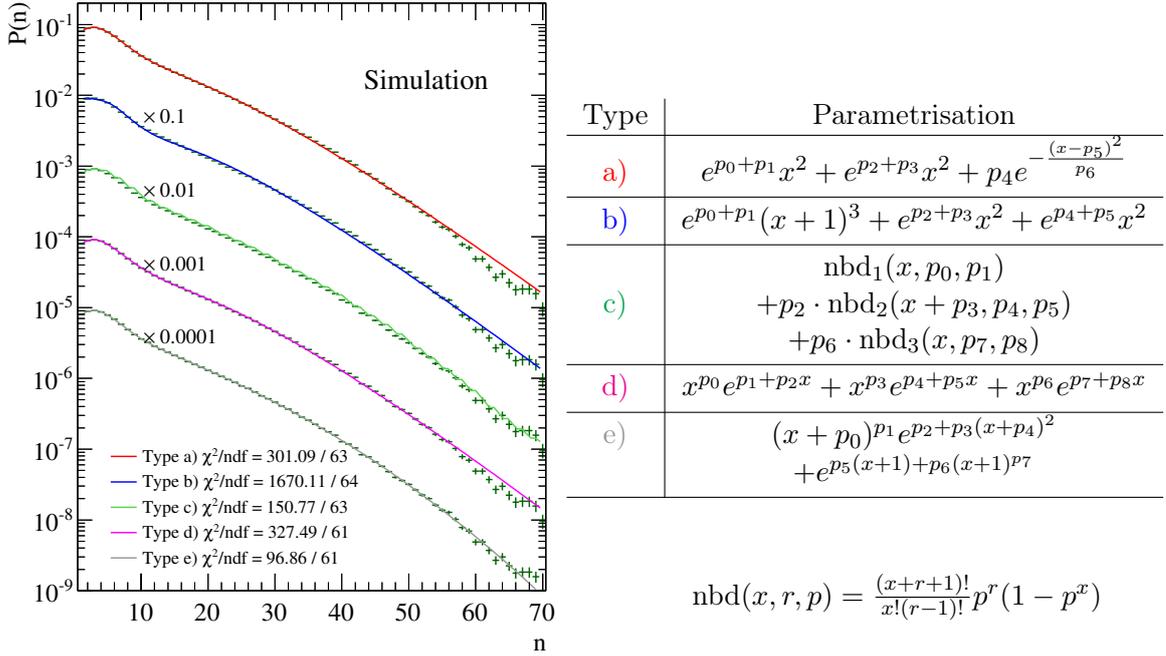


Figure 9.15.: Test of the phenomenological parametrisation models used for the full kinematic range. All models sufficiently describe simulated the charged particle multiplicity distributions. The used models are defined in the table on the right.

folding procedure based on a  $\chi^2$ -minimisation [113, 115] of the measured and the folded distribution  $R\bar{T}(n)$  for different hypotheses of the true distribution,  $\bar{T}(n)$ . The range of variations of  $\bar{T}(n)$  is constrained by parametrising the multiplicity distributions. In order to minimise introduced model dependencies to the unfolded result, five different phenomenological models are used. Four are based on sums of exponential functions combined with polynomial functions of different order in the exponent and as a multiplier. Another one is based on a sum of three negative binomial distributions (nbd) [116]. All models are tested to satisfactorily describe the multiplicity distribution in simulation, as shown in Fig. 9.15.

The floating parameters,  $p_i$ , of the particular hypothesis  $\bar{T}(n)$  are varied in order to minimise the  $\chi^2$ -function:

$$\chi^2(\tilde{T}) = \sum_m \frac{1}{E(m)^2} \left( M(m) - \sum_n R_{mn} \tilde{T}(n) \right)^2, \quad (9.33)$$

where  $E(m)$  represents the uncertainty of the measured distribution  $M(m)$ .

The unfolding procedure is best described by using Fig. 9.16, in which parametrisation model a) is used as an example. Starting point for the unfolding is the measured multiplicity distribution  $M(m)$ , obtained from the preceding correction steps. The displayed multiplicity range of up to  $\approx 50$  tracks has been selected for the background correction. A hypothesis  $\bar{T}(n)$  of the true distribution is folded with the detector response matrix and compared to  $M(m)$ . Both distributions are visualised in the *minimisation*

histogram on the left-hand side in Fig. 9.16, which compares the folded hypothesis  $R\bar{T}(n)$  with the measurement  $M(m)$ . The agreement between both distributions is calculated using the  $\chi^2$ -definition in Eq. 9.33 and is visualised by the pull distribution. The  $\chi^2$ -calculation is carried out in the range of  $m = [0, 45]$  tracks. The upper limit corresponds to the approximate statistical limit of the simulation which is still covered by the original detector response matrix, c.f. Fig. 9.13 a). By using the parametrised response matrix (Fig. 9.14 g), in principle, larger folded multiplicities  $m$  can be accessed. However, the contributing true multiplicities  $n$  reach very large values which are not fully mapped onto the folded distribution. As a result, for values of  $m > 45$  an underestimation of the folded distribution w.r.t. the measured distribution is expected. Furthermore, towards larger multiplicities the measured distribution is significantly affected by remaining background as observed in Sect. 9.2.5. Thus, the unfolding and the  $\chi^2$ -calculation is limited to the range specified above.

For the parameter set that minimises  $\chi^2$  for a given hypothesis, an unfolded multiplicity distribution is obtained. The unfolded distribution is displayed on the right-hand side in Fig. 9.16 and is compared to the true multiplicity distribution directly obtained from the event generator. The associated pull distribution shows overall a good agreement between both distributions, the deviation being less than  $3\sigma$ . However, the patterns in the pull distributions are not only related to random fluctuations but demonstrate the limitation of the used parametrised models. Towards large unfolded multiplicities  $n$ , the unfolded distribution overestimates the true distribution. This is a result of the bias which is already introduced during the background correction where a remaining fraction of fake tracks contaminates the corrected distribution, c.f. Fig. 9.7 in Sect. 9.2.5. It is confirmed that by using a perfect (background free) distribution for  $M(m)$  instead, an optimal unfolded distribution is obtained. As a consequence, only a limited range for the unfolded multiplicity distribution is quoted for the final result.

In addition to the presented parametrisation model a), displayed in Fig. 9.16, the results for the models b)-e) are presented in Fig. 9.17. The different models show different performances describing the multiplicity over the full range, which points to advantages

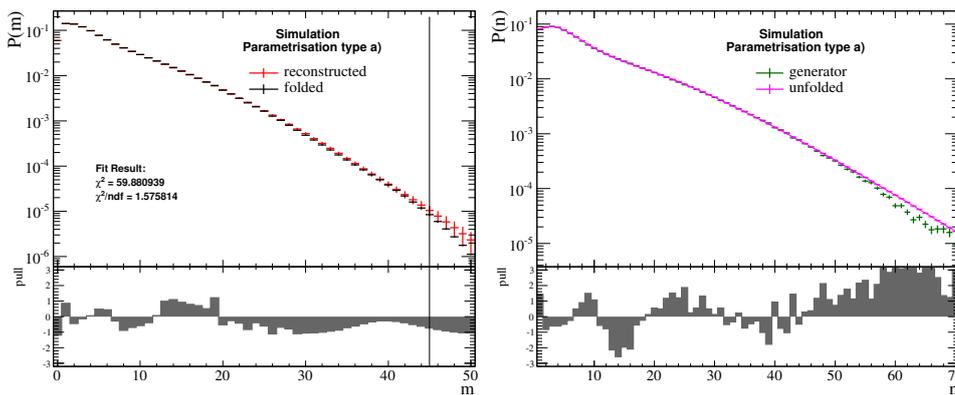


Figure 9.16.: Unfolding the true particle multiplicity distribution in simulation, using parametrisation model a). The minimisation histogram is shown on the left, the unfolded result on the right. The procedure is explained in the text.

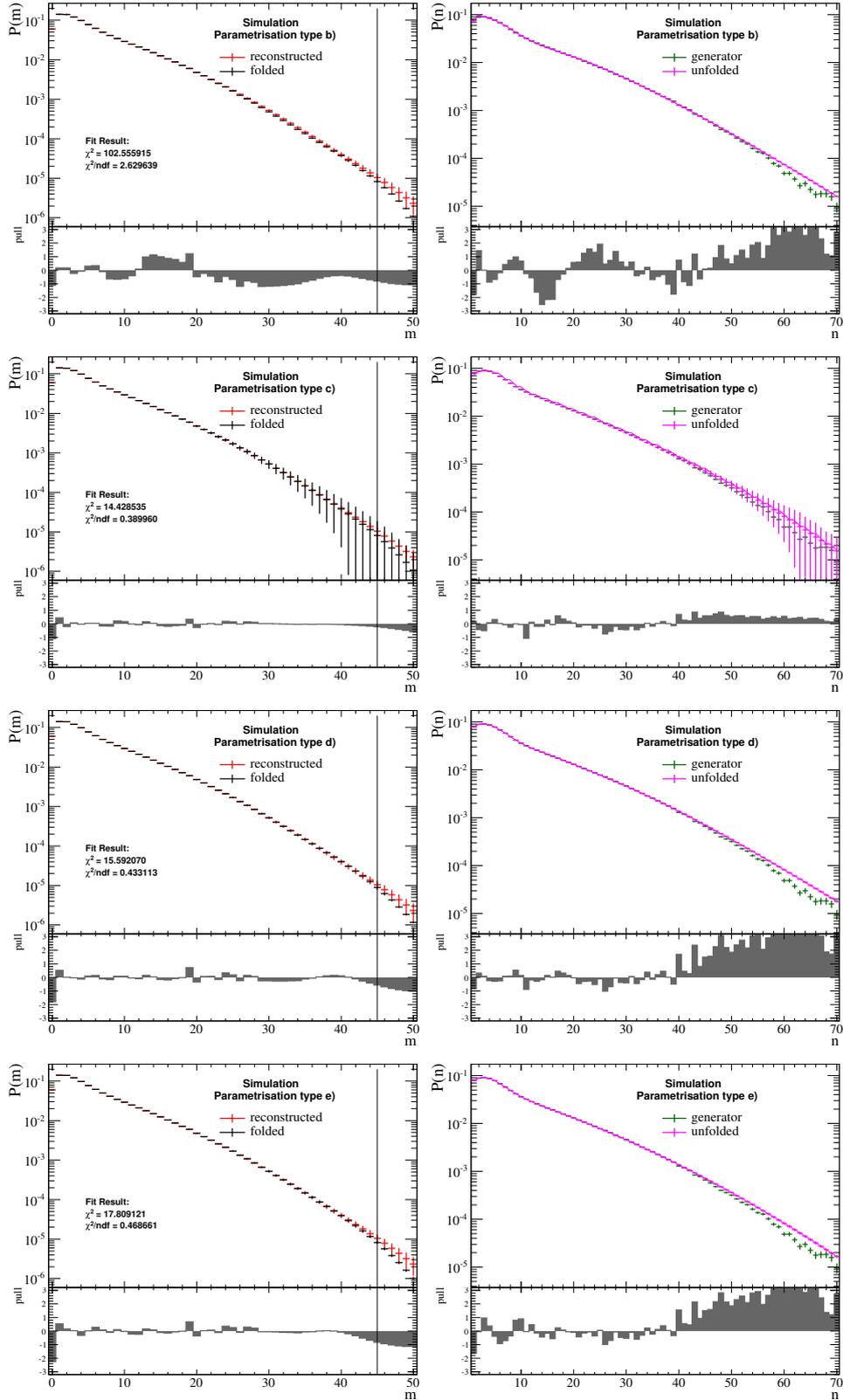


Figure 9.17.: Unfolding the true particle multiplicity distribution in simulation, using models b-e).

## 9. Data analysis and correction

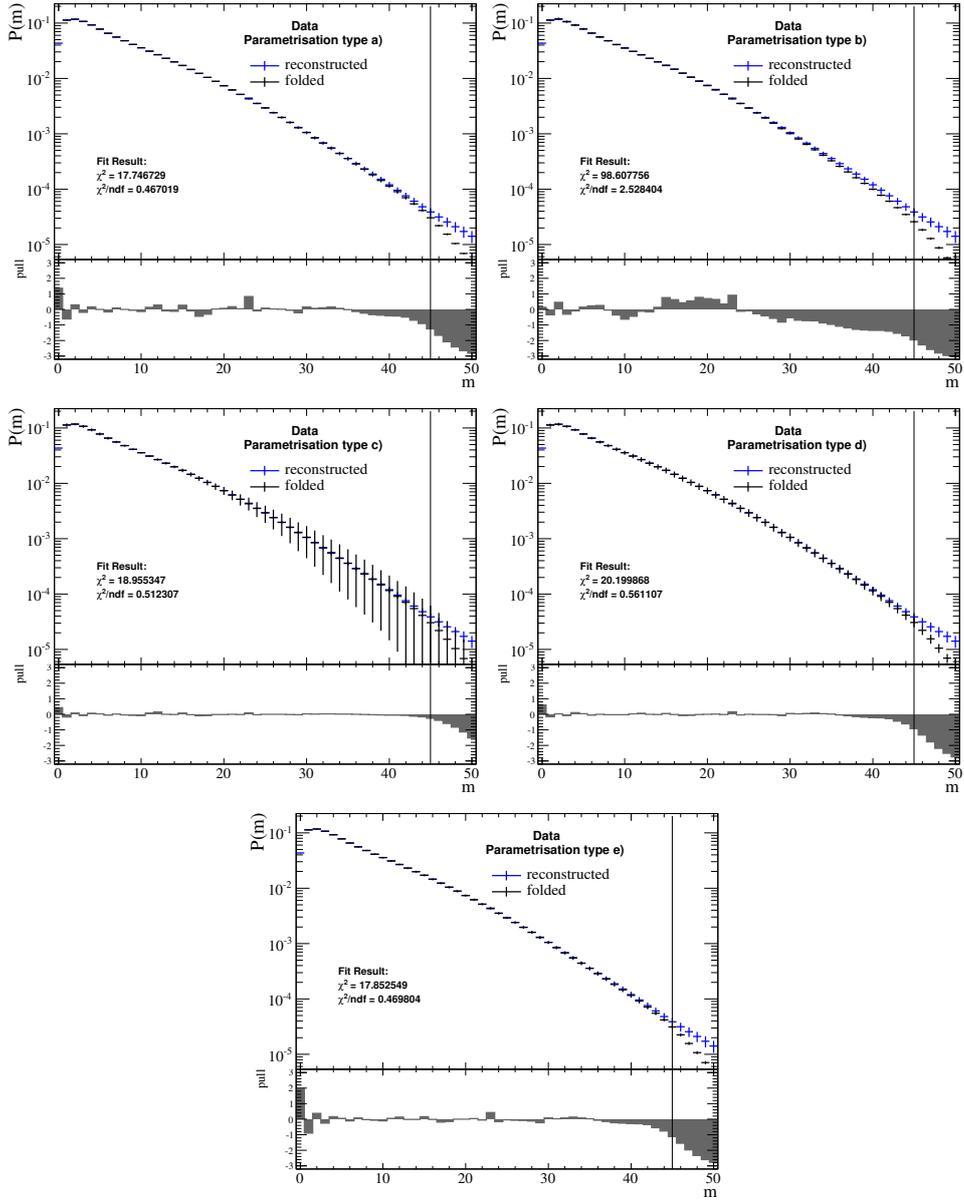


Figure 9.18.: Unfolding the true particle multiplicity distribution using data. The minimisation histograms show the background corrected measured distributions and the folded hypothesis for the parametrisation models a-e).

and disadvantages of the individual parametrisations. It should be emphasised again that the visible overestimate of the unfolded distribution towards large multiplicities is introduced by the previous background correction and not due to the unfolding procedure.

In simulation, the numerically best result is obtained by using model c), which is based on three negative binomial distributions. However, the uncertainties of the fit are large compared to the other models. A strong correlation of the parameters of the three

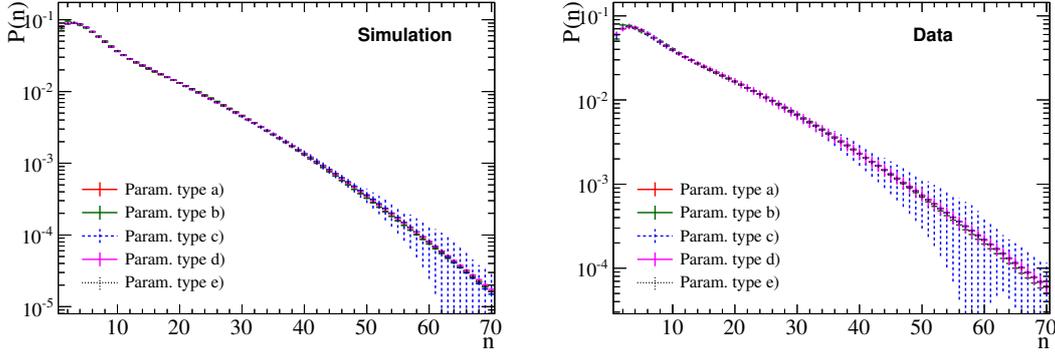


Figure 9.19.: Comparison of the unfolded multiplicity distributions using different parametrisation models in simulation (a) and in data (b).

negative binomial distributions causes an unstable fit and results in a large uncertainty estimate. Due to this, model c) only serves as a cross check. Instead, the unfolding is done by using model d) which represents the central result.

The data unfolding is presented in Fig. 9.18 by showing the minimisation histograms. In analogy to simulation, the same parametrisation models (a-e) are adopted and the same track-range ( $m = [0, 45]$ ) is used for the  $\chi^2$  minimisation. The resulting folded hypotheses  $R\bar{T}(n)$  are in good agreement with the measured distributions. Only towards large values of  $m$  the same discrepancy as in simulation is observed. This is the result of an incomplete mapping of the true to folded distribution and the remainders of the precedent background correction.

All unfolded multiplicity distributions are compared in Fig. 9.19, separately for simulation and data. The different parametrisations lead to consistent results. The displayed uncertainties of each model are obtained by varying all fitted parameters within their uncertainty ranges, the maximum variation within each bin is assigned as uncertainty. The large uncertainties of parametrisation type c) are present in simulation and in data. They reflect the fit instabilities and the large correlations in this model. The region that is most sensitive to the different models is found at very low multiplicities.

To account for systematic limitations by using a specific parametrisation model, the differences between the different unfolding models are later considered as a systematic uncertainty. As stated above, model d) performs best in the unfolding of both, simulated and real data. The respective unfolded distributions serve as the central result.

For the unfolded central multiplicity distribution, the uncertainties which are propagated through the unfolding procedure are recalculated. A more precise estimate is achieved by taking into account correlations of the fitted parameters. The obtained parameter set and the corresponding covariance matrix are rotated into its eigenbasis. The solution is smeared using the diagonalised covariance matrix and assuming Gaussian errors. After rotating back into the original basis, the modified multiplicity distribution is evaluated. This procedure is repeated 1000 times, the resulting variations of the multiplicity distributions are assigned as statistical uncertainty. The systematic uncertainties are discussed in the consecutive chapter.



The precision of the measurements of charged particle multiplicities and mean particle densities are limited by systematic effects. Several potential sources of systematic uncertainties that can bias the measurements are investigated in this chapter. The measured particle densities and particle multiplicities are affected by common underlying sources of uncertainties. These are mainly related to differences between data and simulation and can cause a systematic bias to the applied correction factors. In addition, there are correction specific uncertainties, such as model-dependent limitations in the data unfolding, which are evaluated separately.

All uncertainties that are found to be relevant for this analysis are discussed in the following. The combined total uncertainties are presented in the second to last section of this chapter. In the last section, a consistency check is provided, in order to test the applied unfolding procedure against a different Bayesian unfolding technique.

### **10.1. Fake track estimation on data and simulation**

Fake tracks represent the largest fraction of contaminations in the sample of reconstructed and selected tracks. Like for most of the corrections, the amount of fake tracks is determined by using fully simulated  $pp$  collisions. As a consequence, the precision of the determined correction factors relies on the quality of the simulation. The analysis is based on data collected during an early data taking period. At this time, only few comparisons between data and simulation were available. In order to compare the probability to reconstruct fake tracks in data and in simulation, a procedure, called *Veloflip* method, is developed that estimates the amount of fake tracks independent of using generator information.

As explained in Chap. 5, long tracks are reconstructed by combining VELO tracks with hits in the T-stations. In order to limit the number of different sources for creating fake tracks, only the forward tracking algorithm, as one of the two long track reconstruction algorithms, is used in this analysis. The probability of forming a fake track by assigning

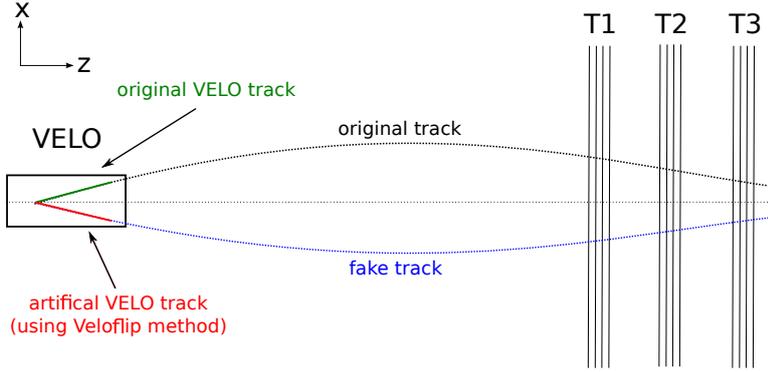


Figure 10.1.: Sketch visualising the idea of the Veloflip method.

suitable but unrelated hits is estimated from the following idea.

Figure 10.1 shows a simple sketch of the long track reconstruction. Ideally, a good VELO track is extrapolated through the magnetic field into the T-stations where the correct hits that correspond to the same particle are found and used to form a long track. Fake tracks are predominantly created by mismatching of good VELO tracks to unrelated T-hits. By extrapolating an artificial VELO track, which does not correspond to a real particle, into the T-stations the mismatching probability is probed. Using the artificial VELO track ensures that no truly related hits can be found. Every subsequently created long track must be a fake track that consists of mismatched T-hits.

To create a VELO track that accounts for the particular event topology an existing VELO track is modified. First, a random VELO track of the event is selected and duplicated. Afterwards, the sign of the track's  $x$ - and  $y$ -coordinate are changed (*flipped*) and a new artificial track is created. By repeating the forward reconstruction with all VELO tracks, including the artificial one, the event reconstruction is only changed minimally. The flipped VELO track is remembered by the algorithm and the resulting fake long track, if created from it, is identified.

Before studying the performance of the fake-track estimator, a few parameters are checked. In Fig. 10.2, three kinematic track parameters ( $\phi$ ,  $\eta$  and  $p$ ) and a quality parameter of the reconstruction are compared in four track categories, these comprise: all reconstructed long tracks, long tracks formed with the selected VELO track (serves as a cross-check), long tracks formed with the flipped VELO track, and all fake tracks identified by using generator information. Three conclusion are drawn from these histograms: (1) The selected VELO track is proven to be selected randomly. (2) A long track that is formed by using the selected VELO track (green) still behaves similar to all reconstructed long tracks (black). No significant bias is introduced to the original track by duplicating and modifying it. (3) The fake tracks created from a flipped VELO track (red) behave in a comparable way to regular fake tracks identified in simulation (blue). The momentum and the azimuthal angle distribution agree well, only in pseudorapidity a higher probability is found that artificial fake tracks are produced at larger  $\eta$ -values. However, the entire  $\eta$ -range can be probed with the Veloflip method. The distribution of the  $Q$ -value, used to identify *good quality* tracks in the reconstruction (c.f. Chap. 5.3), is also in satisfying agreement for regular and artificial fake tracks.

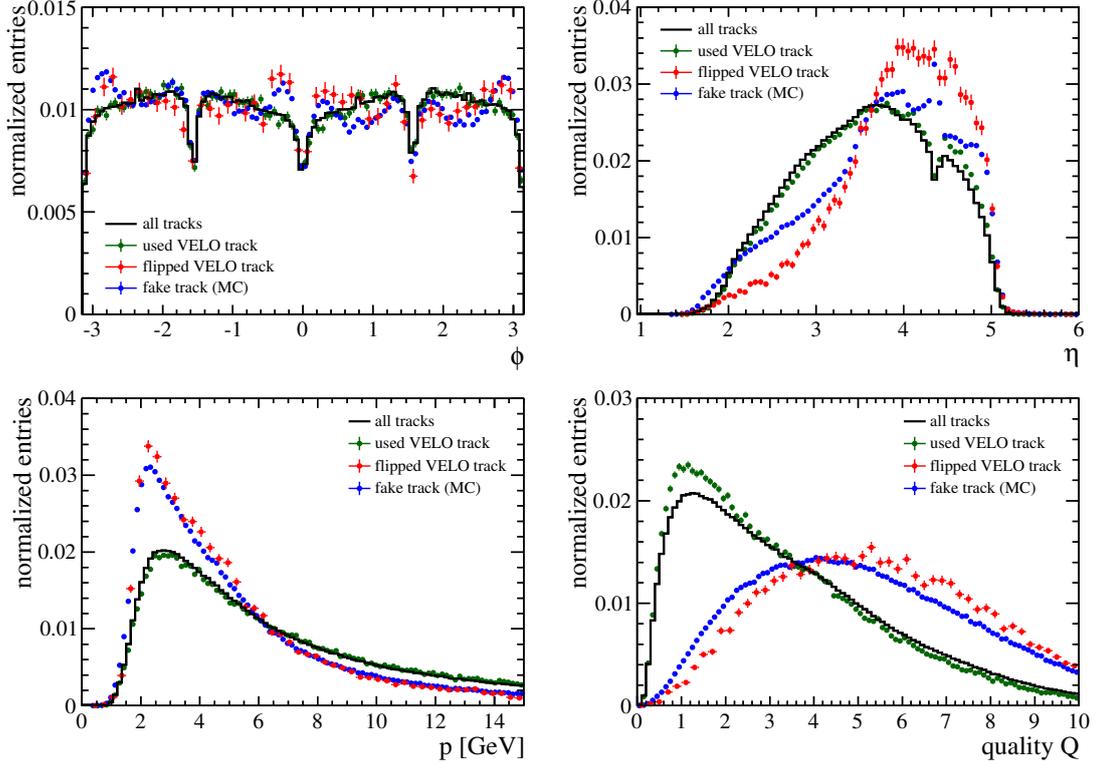


Figure 10.2.: Comparison of the track parameters  $\phi$ ,  $\eta$ ,  $p$  and the quality variable  $Q$  of the forward tracking algorithm, in different long track categories. Details are given in the text.

During the reconstruction sequence, typically more than one long track candidate is formed for each VELO track. The best candidate is identified by using a quality variable, called  $Q$  (c.f. Chap. 5.3). Usually the track candidate that corresponds to a real particle has a better  $Q$ -value than the fake tracks. However, there is a certain probability that a fake candidate prevails against the candidate of the true particle.

The situation is different for the artificially created tracks, since the long track candidate which is based on the flipped VELO track never represents a true particle. The created candidates are *all* fake tracks and there can be no competition with a good track candidate. As a result the estimated probability of creating fake tracks is too large. In order to reproduce the competition, the  $Q$ -value of the flipped VELO track candidate is compared to that of the original VELO track candidate as a reference. If the  $Q$ -value of the artificial candidate is better or equal to the reference candidate, the fake track is retained.

The output of the veloflip method represents the probability that an artificial fake track survives against the reference track. In Fig. 10.3 a), the output, as a function of the hit multiplicity in the T-stations, is shown as the yellow distribution. It behaves similar to the number of fake (long) tracks in simulation divided by the number of total VELO tracks per event, displayed by the blue distribution. By exploiting this fact, a fake-track estimator is obtained by multiplying the veloflip output with the number of long tracks

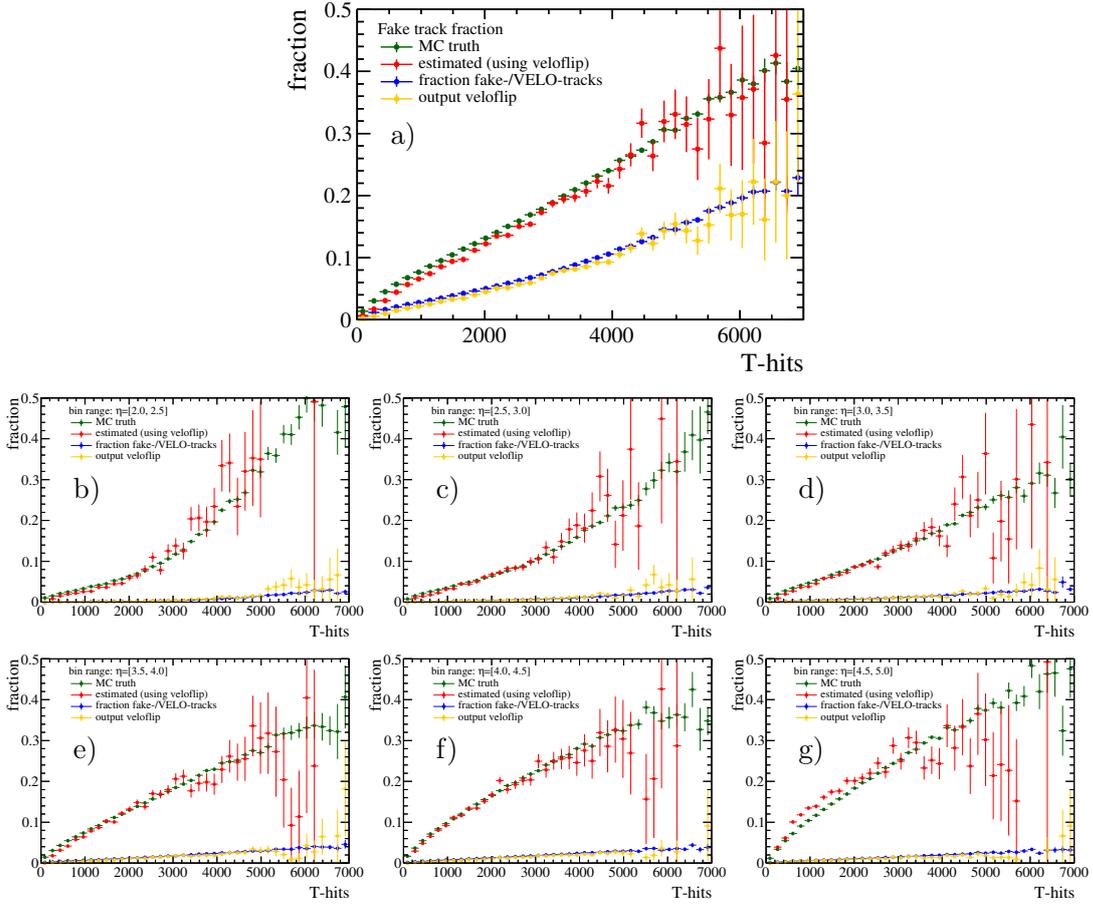


Figure 10.3.: Estimated and true fake track fraction in simulation, in the full kinematic range (a) and in subdivided  $\eta$ -ranges (b-g). The output of the veloflip method (yellow) follows the ratio of fake (long) tracks to VELO tracks (blue). The scaled veloflip output (red) is a good estimator for the true fake track fraction (green) obtained with generator information.

and dividing it by the number of VELO tracks. The resulting estimator (red) shows a remarkably good correlation to the actual fraction of fake long tracks, determined by using generator information (green). The same correlation is also achieved when dividing the full kinematic range in bins of pseudorapidity with a size of  $\Delta\eta = 0.5$ , see Fig. 10.3 b-g). Here, the estimator (red) only accounts for the number of long tracks and VELO tracks within the respective  $\eta$ -bin.

Since the fake-track estimator considered in this way is in good agreement with the actual fake-track fraction, the estimated fraction can be directly used to compare fake tracks in data and in simulation. In Fig. 10.4, the comparison of the fake-track fraction in both samples is depicted. As a function of hit multiplicity in the T-stations, an almost identical behaviour for data and simulation is found. The difference between both samples is fitted by a constant, within uncertainties the result is compatible with zero. The subdivided  $\eta$ -bins are tested separately. From bin to bin, the agreement between data and simulation varies between  $\pm 10\%$ . Each difference is also compatible with zero

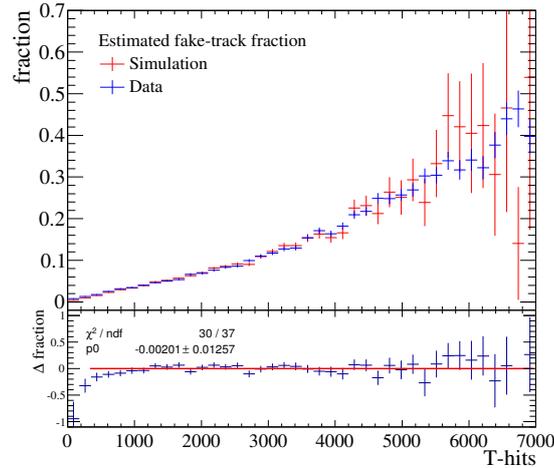


Figure 10.4.: Comparison of the estimated fake track fraction in data and simulation using the veloflip method. No significant difference is found when using T-hit multiplicity as parametrisation.

within less than three standard deviations and thus agrees with the result obtained in the full kinematic range.

It is concluded that no significant difference in the amount of fake tracks in data and simulation is expected, when parametrising the fake-track fraction as a function of hits in the T-stations. The sensitivity of the fake-track estimator in the total kinematic range is of the order of 1 – 2%. As a conservative estimate, a systematic uncertainty of 2% is assigned to the fake-track correction-factors, which have been calculated in simulation.

## 10.2. Cross-check for duplicate tracks

The probability that almost identical tracks are results of multiple reconstruction is determined in simulation. The systematic uncertainty on the correction factors, which are used to statistically correct the amount of duplicate tracks in data, is estimated in the following.

A pair of duplicate tracks typically has very similar track properties, since a large fraction of hits is shared among both tracks. In particular, the opening angle  $\varphi$  between a pair of duplicate tracks is small. The distribution of  $\varphi$ , calculated for all track pairs per event, shows an increase towards very small angles ( $\varphi \lesssim 0.003$ ). This peaking structure is identified to be related to duplicate tracks. A simple check for the number of tracks within the peaking region allows estimating the number of duplicate tracks in data and comparing it to simulation.

Candidates for duplicate tracks are defined, if the opening between two tracks is smaller than a certain cut-off value  $\varphi_{\text{dup}}$ . The resulting number of candidates is denoted as  $N_{\text{cand}}$ . In simulation, the purity  $p$  of this sample is determined as the ratio of actual duplicate tracks with respect to  $N_{\text{cand}}$ . Further, the efficiency,  $\epsilon$ , of selecting duplicate tracks is calculated as the fraction of actual duplicate tracks fulfilling the candidate definition and the total number of actual duplicate tracks. For a given number of probed

| $\varphi < \varphi_{\text{dup}}$ | efficiency $\epsilon$ | purity $p$         | $f_{\text{dup}}$ - fraction of duplicate tracks |                   |
|----------------------------------|-----------------------|--------------------|---|-------------------|
|                                  |                       |                    | simulation                                      | data (estimate)   |
| $\varphi < 0.003$                | $95.30 \pm 0.16\%$    | $22.06 \pm 0.19\%$ | $0.81 \pm 0.01\%$                               | $0.91 \pm 0.01\%$ |
| $\varphi < 0.002$                | $86.63 \pm 0.27\%$    | $26.28 \pm 0.25\%$ | $0.81 \pm 0.01\%$                               | $0.91 \pm 0.01\%$ |
| $\varphi < 0.001$                | $64.55 \pm 0.37\%$    | $31.87 \pm 0.36\%$ | $0.81 \pm 0.01\%$                               | $0.93 \pm 0.01\%$ |

Table 10.1.: The fraction of duplicate tracks in simulation is compared to the estimated fraction in data. Three values of the cut-off parameter  $\varphi_{\text{dup}}$  are evaluated which results in different sets of efficiency and purity. In total  $2 \times 0.5\text{M}$  events in data and simulation are analysed.

long tracks,  $N_{\text{long}}$ , the fraction of duplicate tracks,  $f_{\text{dup}}$ , follows the relation

$$f_{\text{dup}} = \frac{1}{N_{\text{long}}} \frac{p}{\epsilon} \cdot N_{\text{cand}}. \quad (10.1)$$

By implying that the values for efficiency and purity are the same in data and in simulation, the fraction of duplicate tracks can be also estimated for data. In Tab. 10.1, three different values of  $\varphi_{\text{dup}}$  are evaluated. Although  $\epsilon$  and  $p$  show a strong variation for different values of  $\varphi_{\text{dup}}$ , the estimated fraction of duplicate tracks in data is nearly constant.

A slightly larger fraction of duplicate tracks in data compared to simulation is found. This amounts to a relative difference of around 11% (absolute 0.1%). Thus, the simulation determined correction factors for duplicate tracks are assigned a relative systematic uncertainty of 11%. The total fraction of duplicate tracks in data is still below 1% and thus the absolute difference to simulation is almost negligible.

### 10.3. Non-prompt particles

The amount of wrongly selected non-prompt particles is also determined from fully simulated events. To evaluate systematic uncertainties on the correction factors, the composition of selected non-prompt particles is studied.

As mentioned in Chap. 9.2.3, the largest fraction ( $\approx 40\%$ ) of secondary particles passing the selection criteria arise from photon conversion. This effect is directly related to the amount of material in the detector. The amount, which is modelled in the simulation, is found to agree within 10%, on average. Another third of the selected non-prompt particles are charged decay products from  $K_{\text{S}}^0$  mesons. The  $K_{\text{S}}^0$  production cross-section has been measured by LHCb [117] to be in good agreement with simulation. Around 20% of secondary particles originate from decays of  $\Lambda$  baryons and hyperons. These have been measured [118] to disagree by approximately 40% to the production cross-section implemented in simulation.

Linearly combining these contributions results in a systematic uncertainty of  $\approx 12\%$  on the fraction of non-prompt particles .

## 10.4. Uncertainties on track finding and detector acceptance

In this section, systematic uncertainties on correction factors that account for detector acceptance and track finding efficiency are evaluated. For the measurement of charged particle densities individual uncertainties are assigned for both effects. In the multiplicity measurement both effects are accounted for in an unfolding procedure. This requires a different approach to propagate the uncertainties. First, the sizes of the systematic uncertainties are evaluated, before discussing the two separate approaches used propagate the uncertainties to the measurements.

### 10.4.1. Detector acceptance

The detector acceptance efficiency,  $\epsilon_{\text{acc}}$ , describes the fraction of particles that fulfil the reconstructibility criterion (c.f. Chap. 5.4.1). Two separate effects prevent particles from being reconstructible: (1) Particles can interact with detector material before reaching the end of the tracking system. (2) The trajectories of particles can point out of the geometric acceptance of the main tracking stations after the deflection by the magnetic field. Both components are described by a particles loss,  $\lambda$ , for which the following relation to the total acceptance efficiency applies:

$$1 - \epsilon_{\text{acc}} = a \cdot \lambda_{\text{geo}} + b \cdot \lambda_{\text{material}}. \quad (10.2)$$

The numbers  $a$  and  $b$  denote the fractions of particles that are lost due to the geometric acceptance,  $\lambda_{\text{geo}}$ , and material interaction,  $\lambda_{\text{material}}$ , respectively. Both components are displayed in Fig. 10.5 a), showing the particle loss as a function of the particles' momenta.

Of all non-reconstructible particles, a fraction of  $b \approx 20\%$  are lost due to material interaction within the detector. Besides a distinct increase of  $\lambda_{\text{material}}$  towards very low momenta, only a weak momentum dependence is found, as shown in Fig. 10.5 a). The dominant fraction of  $a \approx 80\%$  is not reconstructible, because the particles' trajectories do not remain within the geometric acceptance of the T-stations. The original flight

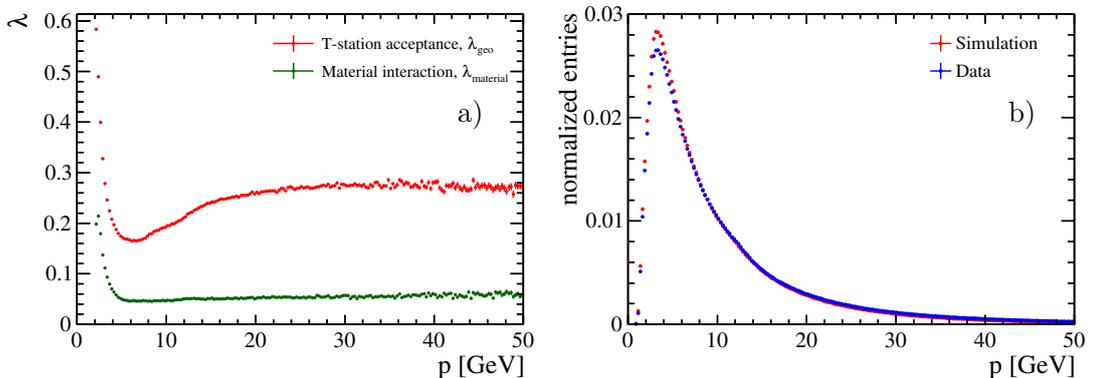


Figure 10.5.: a) The particle loss  $\lambda = (1 - \epsilon)$  presented as function of momentum. b) Comparison of the momentum spectra of selected tracks in data and in simulation.

direction of the particles is altered by the influence of the magnetic field and causes the momentum dependence of  $\lambda_{\text{geo}}$ .

The geometric acceptance is essentially defined by the detector, including the magnetic field, and the momentum of the respective particles. The material interaction predominantly depends only on the amount of material. Thus, no additional large uncertainties between data and simulation are expected for both effects. By assuming momentum to be the leading effect, a systematic uncertainty is estimated on the basis of the different momentum spectra for measured and simulated particles. The discrepancy between both spectra, see Fig. 10.5 b), appears to be small and only a minor systematic effect is expected. The average particle loss, related to the geometric acceptance, is calculated by using the normalised momentum distribution  $\rho(p)$  in data and simulation:

$$\overline{\lambda_{\text{geo}}} = \sum_p \rho(p) \cdot \lambda_{\text{geo}}(p). \quad (10.3)$$

This leads to an average loss of  $\overline{\lambda_{\text{geo}}^{\text{MC}}} = 23.70\%$  in simulation and  $\overline{\lambda_{\text{geo}}^{\text{data}}} = 23.59\%$  in data. By further accounting for the slight momentum dependence of  $\lambda_{\text{material}}$ , the total acceptance efficiency amounts to  $\overline{\epsilon_{\text{acc}}^{\text{MC}}} = 70.49\%$  and  $\overline{\epsilon_{\text{acc}}^{\text{data}}} = 69.85\%$  in simulation and data, respectively, and results in a relative difference of approximately 1%.

In the last step, the particle loss related to material interaction ( $b \approx 20\%$ ) is assigned a 10% uncertainty due to the knowledge of the material budget in the detector. This sums up to an average relative systematic uncertainty of 3% on the overall acceptance efficiency,  $\epsilon_{\text{acc}}$ .

#### 10.4.2. Track finding

To account for systematic differences between the actual track finding efficiency and that estimated from simulation, a global uncertainty of 4% per track is assigned. This comparatively large uncertainty applies to data, that have been collected in 2010, during the early data taking period [28, 119].

#### 10.4.3. Multiplicity unfolding with modified response matrix

For the measured particle multiplicity distributions, the limited efficiencies due to detector acceptance,  $\epsilon_{\text{acc}}$ , and track finding,  $\epsilon_{\text{tr}}$ , are corrected by using an unfolding procedure, based on a detector response matrix. The matrix accounts for the combined (in)efficiency,  $\epsilon_{\text{acc}} \cdot \epsilon_{\text{tr}}$ , which requires a collective treatment of the systematic uncertainty propagation.

The systematic uncertainties of the detector acceptance (3%) and the track finding (4%) can be considered as uncorrelated. This results in a quadratically combined total uncertainty of 5%. A modified response matrix with an efficiency decreased by this value is generated, as shown in Fig. 10.6. There is no separate matrix generated with an increased efficiency (+5%), since the size of the resulting systematic effect is approximately the same.

The whole unfolding procedure, as described in Chap. 9.4.4, is repeated with the modified response matrix. The full difference between the unfolded distributions, using

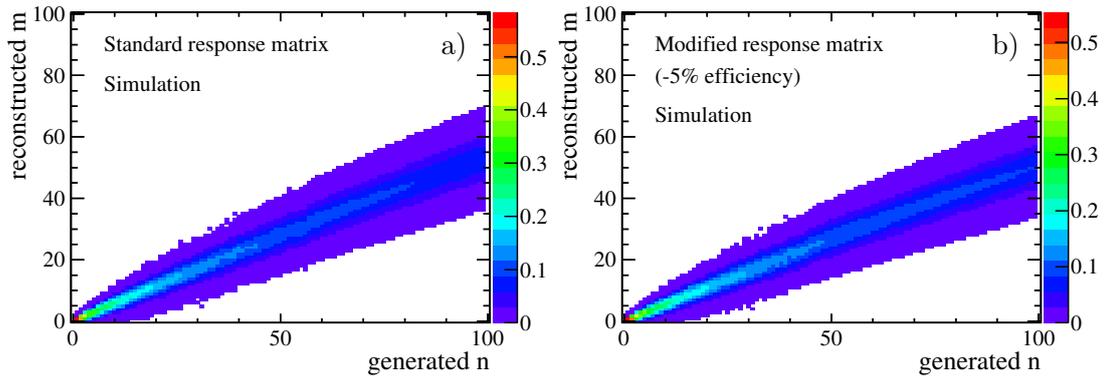


Figure 10.6.: Comparison of the standard detector response matrix (a) with the modified response matrix whose efficiency is artificially decreased by 5%.

the modified and the nominal response matrix, is assigned as a systematic uncertainty. The impact of this uncertainty to the multiplicity distribution is visualised in Sect. 10.6 when summarising all systematic effects.

#### 10.4.4. Efficiency uncertainties of particle densities

For the particle density measurements, the uncertainty on the acceptance efficiency is further evaluated by testing the impact of the binning scheme used to determine the correction factors, since an increasing variation is found towards the borders of the detector. The default fine two-dimensional binning scheme is changed by doubling the bin sizes and shifting the centre of the bins by half of its size. On average, the resulting impact of the alternative binning scheme to the measurement is the order of the previously stated 3% uncertainty. In the  $\eta$ -region around the beam pipe and close to the borders of the detector larger differences are found. As a result, the individual uncertainties due to the binning scheme are assigned a systematic uncertainty. Apart from that, also the aforementioned 4% uncertainty related to the track finding is considered.

### 10.5. Uncertainties to unobserved events

In this analysis, the data-driven approach described in Chap. 9.1 is used to correct for visible but unobserved events. The assumptions made in the calculation model result in a systematic overestimation of the fraction of unobserved events of  $\Delta_{\text{method}} \approx 13\%$ , c.f. Chap. 9.1. Although this bias is taken into account when calculating the correction factors for data, the limitation of the model is considered as a full systematic uncertainty.

The calculation of the number of unobserved events is based on the single-particle survival probability  $\mathcal{P}_{\text{sur}}$ . Similar systematic uncertainties, which are already discussed for the non-prompt correction and for the efficiency corrections, also affect the survival probability:

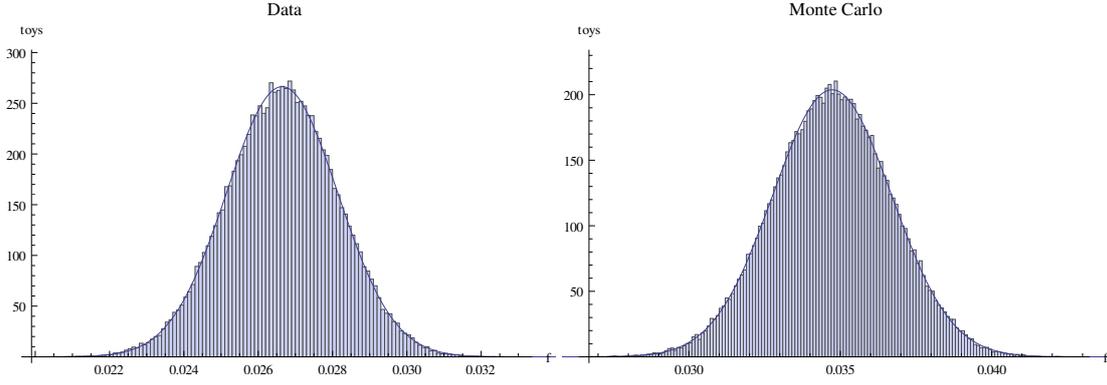


Figure 10.7.: Systematic uncertainty evaluation of  $f_{\text{unobs}}$  using  $100k$  Gaussian-smearred single-particle survival probabilities.

**Material uncertainty** According to Tab. 9.1 in Chap. 9.1, the dominant reason ( $\approx 57.9\%$ ) that prevents visible particles from being reconstructed is the interaction with detector material. The survival probability changes by  $\Delta_{\text{mat}} = 1.53\%$ , when varying the material by the previously introduced  $10\%$ .

**Detector acceptance** The second largest contribution ( $\approx 19.7\%$ ) affecting the survival probability are particles which are deflected out of the acceptance of the tracking stations while traversing the magnetic field. Systematic uncertainties are driven by the slightly different momentum distributions between measurement and simulation. As discussed in Sect. 10.4.1, this effect is covered by an average uncertainty of  $1\%$ . For the survival probability, this translates into an uncertainty of  $\Delta_{\text{acc}} = 0.06\%$ .

**Track finding** Another small contribution ( $\approx 6.7\%$ ) to the survival probability is the track finding efficiency. The average  $4\%$  uncertainty results in a systematic uncertainty of  $\Delta_{\text{tr}} = 0.07\%$  on  $\mathcal{P}_{\text{sur}}$ .

The remaining effect that contributes to the survival probability are particles that decay in flight. Uncertainties to this fraction are negligible.

All listed effects are considered uncorrelated. The **combined uncertainty** of  $\mathcal{P}_{\text{sur}}$  is dominated by the term related to the material and amounts to

$$\Delta_{\text{sur}} = \sqrt{\Delta_{\text{mat}}^2 + \Delta_{\text{acc}}^2 + \Delta_{\text{tr}}^2} = 1.54\%. \quad (10.4)$$

To evaluate the impact of this systematic uncertainty on the simulation-determined fraction of unobserved events,  $f_{\text{unobs}}$ , the survival probability is smeared by a Gaussian distribution with standard deviation  $\sigma = \Delta_{\text{sur}}$ . A sample of  $100k$  values of  $\mathcal{P}_{\text{sur}}$  is generated and then used to recalculate  $f_{\text{unobs}}$ . The resulting values of  $f_{\text{unobs}}$  follow again a Gaussian distribution, as displayed in Fig. 10.7 for data and simulation. The standard deviation represents the systematic uncertainty of  $f_{\text{unobs}}$  related to the survival probability and is further denoted as  $\Delta_{\text{sur}}$ .

| Data sample |            | $f_{\text{unobs}}$ | $\Delta_{\text{method}}$ | $\Delta_{\text{sur}}$ | $\Delta_{\text{tot}}$ | rel. syst. error |
|-------------|------------|--------------------|--------------------------|-----------------------|-----------------------|------------------|
| up          | Simulation | 3.07%              | 0.42%                    | 0.20%                 | 0.46%                 | 15.0%            |
|             | Data       | 2.33%              | 0.32%                    | 0.15%                 | 0.35%                 | 15.0%            |
| down        | Simulation | 3.07%              | 0.41%                    | 0.20%                 | 0.45%                 | 14.8%            |
|             | Data       | 2.35%              | 0.31%                    | 0.15%                 | 0.35%                 | 14.9%            |

Table 10.2.: Systematic uncertainties on the fraction of unobserved event, listed separately for different data sets.

The final systematic uncertainty,  $\Delta_{\text{tot}}$ , on the correction of unobserved events is composed of two uncertainties: the survival probability,  $\Delta_{\text{sur}}$ , and the calculation model,  $\Delta_{\text{method}}$ , mentioned in the beginning of this section. The calculated numbers are summarised in Tab. 10.2. A total systematic uncertainty of  $\approx 15\%$  is obtained. The result can be considered as a conservative estimate, since the value is dominated by the large value related to the method ( $\Delta_{\text{method}}$ ).

The uncertainty on  $f_{\text{unobs}}$  has a different impact on the particle density and particle multiplicity measurements, see Fig. 10.8. The particle densities are simply normalised by a different number of events. Since the correction itself is already small, the effect of the systematic uncertainty is barely visible. To propagate the uncertainty through the unfolded multiplicity distribution, the entire analysis is repeated with  $f_{\text{unobs}} \pm \Delta_{\text{tot}}$ . The multiplicity distribution is only changed in the range of a few particles, the tail of the distribution nearly remains unchanged.

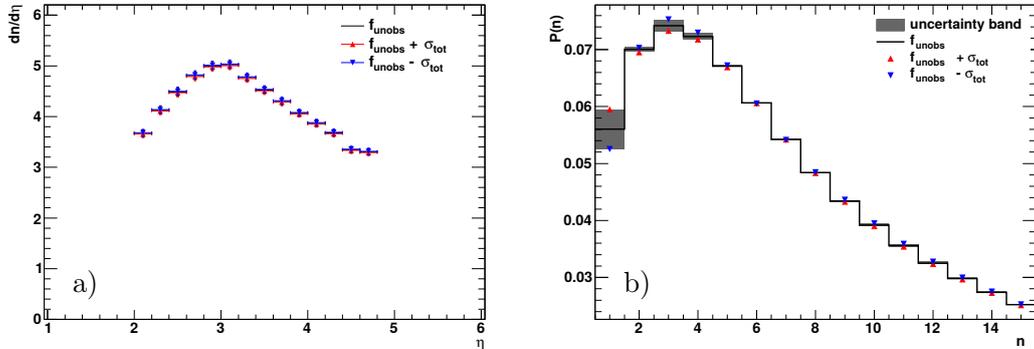


Figure 10.8.: Effect of varying  $f_{\text{unobs}}$  within its systematic uncertainty range. a) Particle density  $dn/d\eta$ , b) multiplicity distribution in the total kinematic range, showing only multiplicities from 1-15 particles.

## 10.6. Total systematics

The discussed uncertainties represent all relevant effects that have an impact on the measurements of this analysis. In several of the applied corrections a common effect, *e.g.* the uncertainty on the material budget, is responsible for the quoted systematic uncertainties. As a result, these uncertainties are treated as correlated.

The first correlation concerns fake tracks, duplicate tracks and the track finding efficiency. Fake and duplicate tracks are determined as function of VELO track multiplicity, which is correlated to the general track finding performance. Furthermore, parameters in the reconstruction software affect both, the efficiency of reconstructing particles and the probability of mis-reconstructing non-physical tracks. Thus, all three systematic uncertainties are treated as fully correlated.

Another correlation addresses the corrections for non-prompt particles and for the acceptance efficiency. The largest uncertainty on the amount of secondary particles is related to the material budget of the detector. Also, 20% of the acceptance efficiency are related to material interactions. The uncertainty on this fraction is treated as fully correlated to that of the non-prompt particles.

In addition to the explicitly mentioned systematic effects in this chapter, further effects are investigated. The precision of the iterative pile-up correction is systematically limited by the number of applied iterations. However, the effect of additional iterations is estimated to be smaller than 1‰, which results in a negligible systematic uncertainty compared to other contributions.

The effect of non-zero beam crossing angles is determined to be insignificant, as well as the background induced by beam gas interactions.

In total, the limiting systematic uncertainty of this analysis is the 10% uncertainty of the knowledge of the material. This large number is probably a very conservative estimate which simply covers all potential uncertainties. However, so far no dedicated

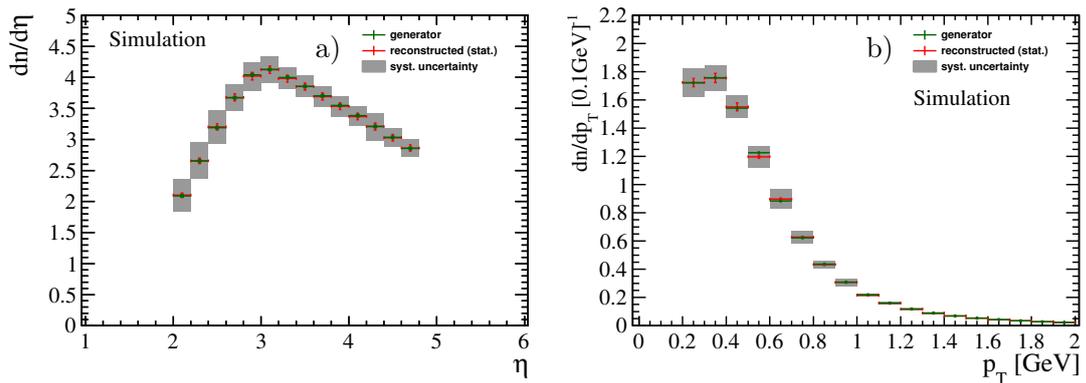


Figure 10.9.: Charged particle densities in simulation as function of pseudorapidity (a) and transverse momentum (b). Reconstructed corrected distributions are compared to the original generator input. The grey systematic uncertainty band exceeds the statistical error bars.

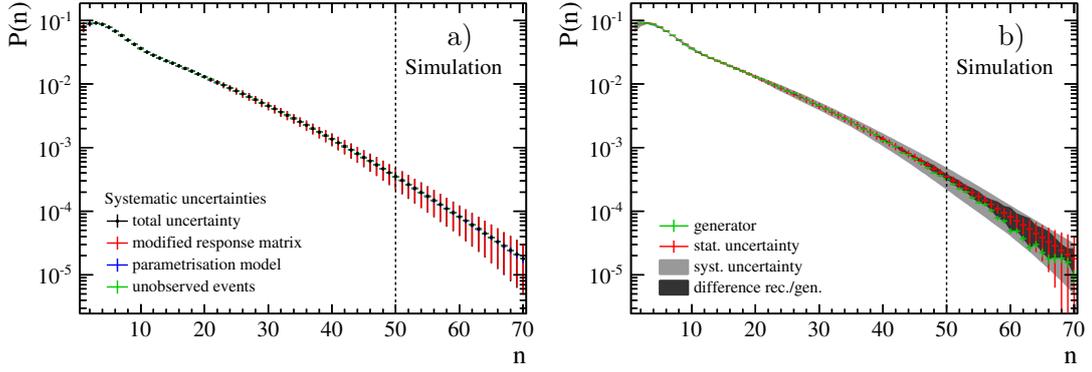


Figure 10.10.: Uncertainties on the simulated charged particle multiplicity distribution. Relevant systematic uncertainties are displayed in (a), details see text. The fully corrected reconstructed multiplicity distribution including all uncertainties is compared to the original generator distribution in (b). The size of the remaining differences after all corrections are indicated by the dark grey error band. The border indicates the maximum multiplicity considered for the measurement.

studies have been performed in order to obtain a better knowledge of the amount of material and its actual position within the detector.

The size of the combined systematic uncertainties, together with statistical uncertainties, are shown in Fig. 10.9 for reconstructed particle densities in simulation. The figures clearly show that the assigned systematic uncertainties limit the precision of the measurement. In regions where large correction factors are applied, *e.g.* small  $\eta$ , the uncertainties are naturally larger.

Due to the unfolding procedure, systematic uncertainties have a different effect on the particle multiplicity distributions. The contributing effects and their implementations are explained in the following. In Fig. 10.10 a) the unfolded multiplicity distribution for simulated events is displayed, showing the single contributing systematic uncertainties. Over almost the entire multiplicity range, the systematic uncertainties related to the efficiencies (described by the response matrix) are dominating. The uncertainty is obtained by comparing the central (best) unfolded distribution to the corresponding result obtained by using the modified response matrix, decreased by 5% efficiency. The difference between both unfolded distributions in each bin is considered as an uncertainty.

At small multiplicities, also the uncertainties related to the parametrisation models (blue) and the unobserved-event correction (green) are of relevance. The systematic error introduced by choosing a parametrisation model during the unfolding is estimated by comparing the unfolded distributions of each considered model. Within each bin, the value of the central unfolded result is compared to that of all other unfolded distributions. The maximum difference in each bin is treated as a full uncertainty.

The systematic uncertainty related to the correction of unobserved events to the final result is obtained by repeating the entire analysis with different values for the fractions of unobserved events:  $f_{\text{unobs}} \pm \Delta_{\text{tot}}$ . Only a small effect at very small multiplicities is observed, other systematic effects are dominant.

The final systematic uncertainty band is obtained by quadratically combining the discussed single uncertainties. Figure 10.10 b) shows the combined uncertainty (light grey band) together with the statistical uncertainty (red error bars). The final result is compared to the original multiplicity distribution obtained from the event generator. The dark grey uncertainty band indicates the small remaining bias between the generator distribution and the fully corrected reconstructed distribution. Towards larger multiplicities, the discrepancy between both distribution increases.

A multiplicity range of 1 – 50 particles is defined in which the sizes of the statistical and systematic uncertainties are moderate. Within the specified range, the truncated mean value,  $\mu$ , and the root-mean-square deviation,  $\sigma$ , are calculated for the reconstructed and the generated multiplicity distribution. The obtained values for  $\mu$  and  $\sigma$  are listed in Tab. 10.3. The results for the generated and the reconstructed simulation are in agreement within the systematic uncertainties. The multiplicity range of 1 – 50 particles is also used for the publication of the data. Results are presented in the next chapter.

| Simulation    | Mean value $\mu$            | RMS $\sigma$                |
|---------------|-----------------------------|-----------------------------|
| Generated     | $9.397 \pm 0.006$           | $8.367 \pm 0.004$           |
| Reconstructed | $9.414 \pm 0.006 \pm 0.012$ | $8.406 \pm 0.004 \pm 0.034$ |

Table 10.3.: Comparing generated and reconstructed charged particle multiplicities in simulation. The truncated mean value and root-mean-square deviation are calculated in a range of 1 to 50 particles. Given uncertainties are statistical and systematic.

## 10.7. Check of unfolding procedure

In this analysis, an unfolding procedure is applied, which is based on using a detector response matrix and a  $\chi^2$ -minimisation that quantifies the agreement between a folded solution and the measured multiplicity distribution. The implemented technique requires a parametrisation of the original not yet folded distribution and thus leads to a systematic limitation of the obtained result. However, systematic uncertainties are evaluated by propagating fit-parameter uncertainties to the assumed multiplicity distribution and by using different parametrisation models. Accounting for all potential effects leads to a significant uncertainty range.

In order to provide a supplementary check of the chosen unfolding procedure, an alternative approach is adopted, which does not require a parametrisation of the multiplicity distribution. In this approach a Bayesian unfolding technique is used. The method, described by D’Agostini [120], provides an iterative solution to invert the detector response matrix,  $R$ , using the notation of Chap. 9.4.4. The central calculation rule, used to determine the elements of the inverted matrix  $V$ , is

$$V_{n,m} = \frac{R_{m,n}T'(n)}{\sum_{n'} R_{m,n'}T'(n')}. \quad (10.5)$$

Here,  $T'(n)$  refers to an initial a priori distribution of the true multiplicity spectrum. The calculation of  $V_{n,m}$  obviously depends on the selected initial distribution  $T'(n)$ . If

using the actual true multiplicity distribution  $T(n)$ , the resulting inverted matrix fulfils the relation:

$$T(n) = \sum_m V_{n,m} M(m), \quad (10.6)$$

where  $M(m)$  denotes the measured multiplicity distribution. Since the true distribution,  $T(n)$ , is unknown, any other choice for  $T'(n)$  in Eq. 10.5 leads to a different matrix  $\tilde{V}$  and thus to a different solution  $\tilde{T}(n)$ . Following Ref. [120], this spectrum  $\tilde{T}(n)$  is a better estimate for the true spectrum compared to the initial choice. Typically a best-knowledge distribution is used as the initial distribution, however, it can be even set to a uniform distribution. The resulting solution  $\tilde{T}(n)$ , calculated by using Eq. 10.6, can be further used as an improved initial distribution in Eq. 10.5, which then results in a better solution for the unfolded multiplicity distribution.

It should be noted that this iterative procedure can start to interpret statistical fluctuations as physical structures of the multiplicity distribution. To prevent this behaviour, a regularisation parameter is required. A recommended and simple regularisation is to limit the procedure to a few iterations.

For the following check of the unfolding procedure, the existing software package `RooUnfold` [121] is adopted. It provides an implementation of the iterative Bayesian unfolding method which is used in the following. The nominal results of the multiplicity measurement are compared to the results obtained from the alternative method. The same measured multiplicity distributions are used, as well as the same detector response matrices. In the `RooUnfold` package, the implemented Bayesian approach uses a limited number of iterations as a regularisation. For this check, five iterations are chosen, although no significant differences are found when using up to ten iterations.

In Fig. 10.11, the Bayesian unfolded multiplicity distributions for simulation and data are compared to the respective final results of the analysis. For the Bayesian result, no proper statistical error propagation is performed and the statistical uncertainties are

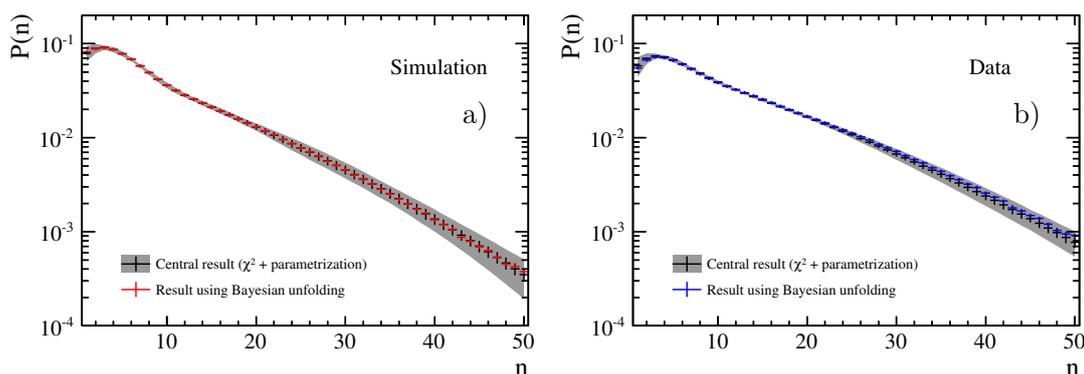


Figure 10.11.: Check of the unfolding procedure, showing particle multiplicity distributions in the full kinematic range. Nominal results (with statistical error bars and combined statistical and systematic uncertainty band) are compared to the unfolded distributions obtained with the Bayesian unfolding technique (no errors given). Consistent results are found for simulation (a) and for data (b).

underestimated. Therefore only the central values are considered.

In the simulation sample, Fig. 10.11 a), the results of the nominal and the alternative approach are almost identical. The parametrisations used in the nominal unfolding are designed and successfully tested to describe the true multiplicity distribution in simulation. The good agreement of the Bayesian unfolded result with the nominal result and thus the original generator distribution confirms the correct application of the implemented method. The results for data in Fig. 10.11 b) are also found to be in agreement with each other. The distribution obtained from using the Bayesian unfolding procedure shows a slight shift towards larger multiplicities. Within the quoted uncertainty range for the nominal result both distributions are fully consistent.

---

## Results of charged particle densities and multiplicities

---

In this chapter, the fully corrected measurements of charged particle densities and charged particle multiplicities are presented and compared to predictions from several Monte Carlo event generators.

The selected generators are grouped into two classes, those that have not been optimised using LHC data and those that have. The former includes the Perugia 0 and Perugia NOCR tunes of PYTHIA 6, the PHOJET generator, and an early version of PYTHIA 8, namely version 8.145. In more recent versions of PYTHIA 8, here represented by version 8.180, the default configuration of the generator is changed to Tune 4C, which is based on LHC measurements at central rapidity. The second class of recent event generators further comprises HERWIG++, which implements a cluster fragmentation model instead of the Lund string fragmentation used in PYTHIA. Predictions of two versions of HERWIG++ are selected, version 2.6.3 and 2.7.0, each operated in the minimum-bias configuration which uses the respective default underlying-event tune. More details about the above-named event generators and their tunes are given in Chap. 2.3.

In this chapter, the results of the particle density measurements are presented first, followed by the measured particle multiplicity distribution in the full kinematic range, and the particle multiplicity distributions for bins in pseudorapidity and transverse momentum. The presented figures and results are identical to those of the corresponding publication, Ref. [9].

### 11.1. Charged particle densities

The fully corrected measurement of charged particle densities in the full kinematic range of  $p > 2 \text{ GeV}$ ,  $p_T > 0.2 \text{ GeV}$  and  $2.0 < \eta < 4.8$  is presented as a function of pseudorapidity in Fig. 11.1 and as a function of transverse momentum in Fig. 11.2.

The measured particle densities show a characteristic drop towards large  $\eta$ . Towards central pseudorapidities, where typically a larger amount of charged particles are pro-

## 11. Results of charged particle densities and multiplicities

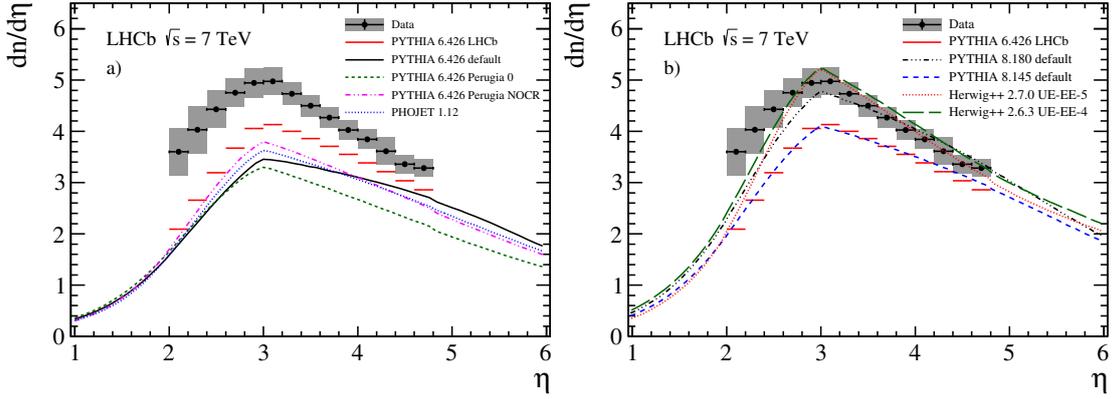


Figure 11.1.: Charged particle densities as a function of pseudorapidity. The LHCb data are shown in points with statistical error bars (smaller than the marker size) and combined systematic and statistical uncertainties as the grey band. The measurement is compared to predictions of several Monte Carlo event generators.

duced, the data points also show a falling edge for  $\eta < 3$ . This decrease is caused by the minimum momentum requirement in the analysis. This general shape of the  $dn/d\eta$  distribution is qualitatively reproduced by all considered event generators and their tunes.

The first group of generators that are compared to the measurement are different tunes of PYTHIA 6 and PHOJET, which are not optimised to LHC data. They are shown in Figs. 11.1 a) and 11.2 a). PYTHIA 6.426 in the default configuration clearly underestimates the amount of charged particles over the whole range of the measurement. While at large  $\eta$  the discrepancy is around 20%, it increases up to roughly 50% at small  $\eta$ . Also the descending slopes towards small and large pseudorapidities are insufficiently reproduced. The Perugia NOCR tune indicates a slight improvement in shape and in total particle density. For Perugia 0 the modelled shapes are similar but the predicted amount of charged particles is even smaller over the entire kinematic range. The second event generator, PHOJET, which models soft-particle production by using the dual-parton model, predicts similar results to that of the discussed PYTHIA 6 tunes. In the first group of generators, the LHCb tune of PYTHIA 6 provides the best agreement with the data. The charged-particle production rate is still underestimated by 10 – 40%. The same behaviour is also observed in the  $p_T$ -dependence. All aforementioned generator predictions underestimate the number of charged particles.

The picture changes when comparing the measured particle densities to the more recent Monte Carlo event generators PYTHIA 8 and HERWIG++, as shown in Figs. 11.1 b) and 11.2 b). The old PYTHIA version 8.145 was released before the default parameters were tuned to the first LHC measurements. As a result, these predictions are not better than those of the LHCb tune of PYTHIA 6. In contrast, PYTHIA 8.180 represents a version whose parameters are optimised using data from the LHC. The implemented default parameters set, Tune 4C, allows describing the LHCb data significantly better than the previous PYTHIA versions. The shapes of the measured particle densities in  $\eta$  and in  $p_T$  are not yet perfectly modelled, but the total amount of charged particles is almost reproduced. The largest discrepancies are found in describing the falling edge towards

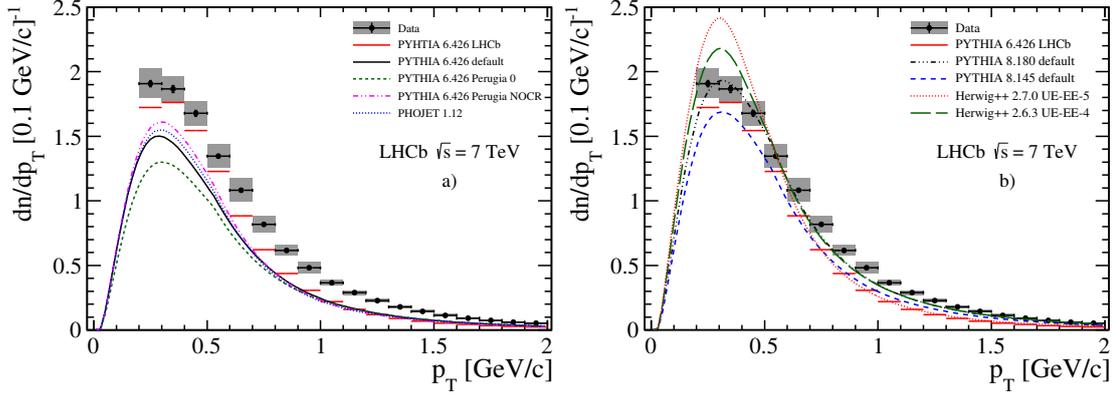


Figure 11.2.: Charged particle densities as a function of transverse momentum. The LHCb data are shown in points with statistical error bars (smaller than the marker size) and combined systematic and statistical uncertainties as the grey band. The measurement is compared to predictions of several Monte Carlo event generators.

small  $\eta$ , which is sensitive to the low-momentum cut off in this analysis. The predictions from HERWIG++ are also in reasonable good agreement with the measurement, only the charged-particle production at low  $\eta$  is also underestimated. Two version of the HERWIG++ generator are shown in this comparison. Version 2.7.0, which implements the underlying-event tune UE-EE-5-MRST (UE-5), overestimates the particle density in the low- $p_T$  range but underestimates it towards larger transverse momenta. The predictions of HERWIG++ 2.6.3, based on the UE-EE-4-MRST tune (UE-4), give a more complete picture of the data. Both tunes slightly differ in the effective cross section, which is used for double-parton scattering. However, the visible differences are small.

In summary, recent PYTHIA 8 and HERWIG++ event generators describe the LHCb data over a wide range, but some of the probed kinematic regions are still underestimated. Predictions of older event generators are significantly worse.

## 11.2. Charged particle multiplicities

The charged particle multiplicity distribution that is measured in the full kinematic range of the analysis is presented in Fig. 11.3. The data are compared to predictions of the same Monte Carlo event generators already used for the particle density measurements. The mean value,  $\mu$ , and the root-mean-square deviation,  $\sigma$ , of the measured distribution, truncated in the range from 1 to 50 particles, is determined to be  $\mu = 11.30 \pm 0.01 \pm 0.09$  and  $\sigma = 9.50 \pm 0.01 \pm 0.02$ , where the uncertainties are statistical and systematic, respectively. Using the full range gives consistent results with the values obtained from the particle densities.

Comparing the measurement with the predictions from event generators that do not rely on LHC data, Fig. 11.3 a), shows that the multiplicity distribution is significantly underestimated. The smallest probability to observe a high multiplicity event is predicted by the PHOJET generator, being in clear disagreement with the data. This can be understood since the PHOJET simulation mostly contains soft-scattering events. Also the

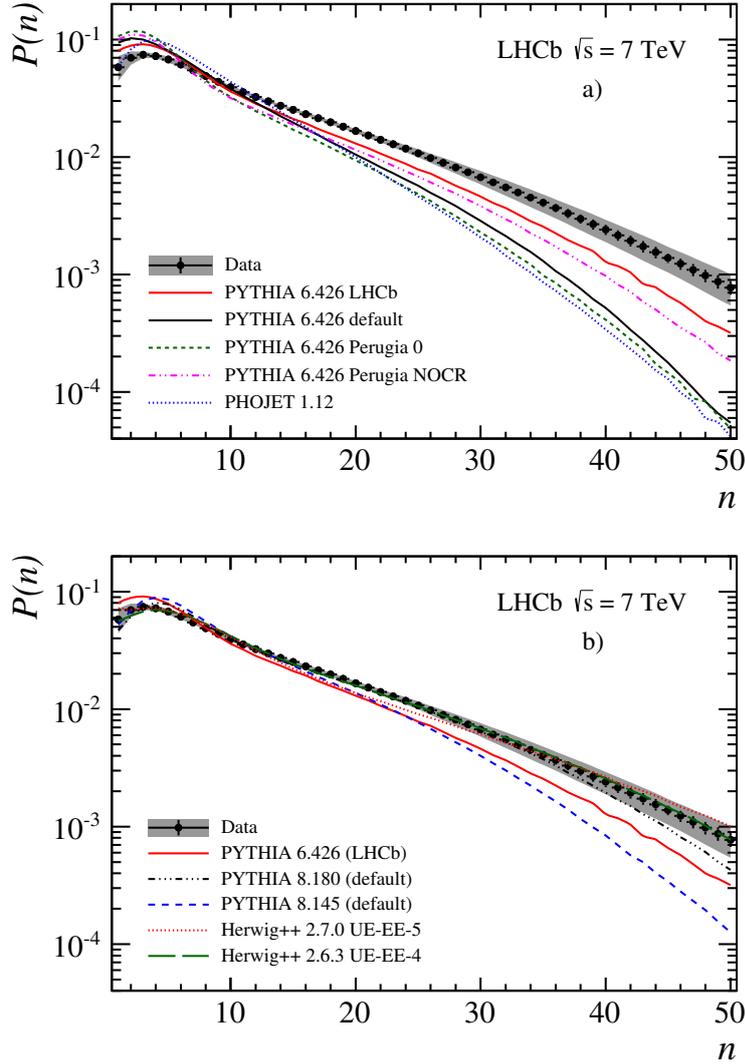


Figure 11.3.: Observed charged particle multiplicity distribution in the full kinematic range of the analysis. The error bars represent the statistical uncertainty, the error band shows the combined statistical and systematic uncertainty. The data are compared to predictions from several Monte Carlo event generators.

old PYTHIA 6 generator underestimates the production cross-section of charged particles in all tunes that are considered in this comparison. Closest to the data is the prediction of the LHCb tune, although the mean value is still predicted to be 15% lower than the measured data.

Recent event generators, which are depicted in Fig. 11.3 b), show the expected improvement as found for the particle density measurements. While PYTHIA 8.145 shows the same insufficient description of data as its predecessor, version 8.180 using Tune 4C predicts multiplicities being in reasonable agreement with the measurement. Also HERWIG++ using the UE-4 tune in version 2.6.3 reproduces the data with good

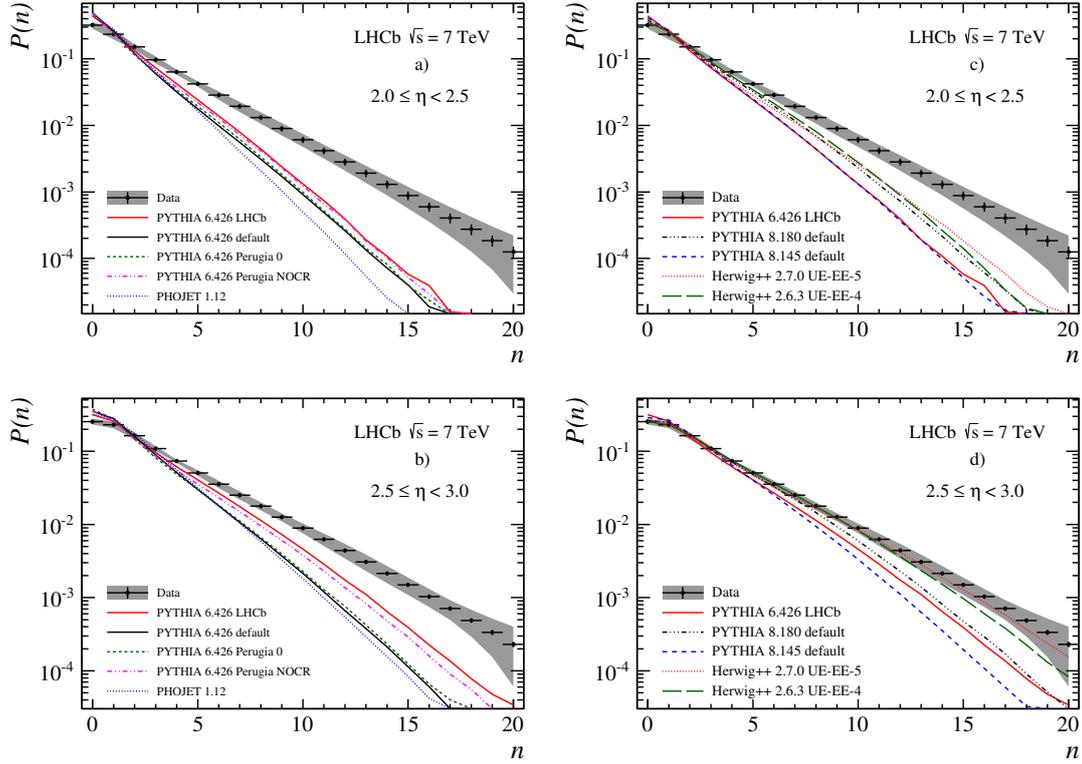


Figure 11.4.: Observed charged particle multiplicities in different  $\eta$  bins. Error bars represent the statistical uncertainty, error bands show the combined statistical and systematic uncertainties. The data are compared to several event generator predictions.

agreement. The UE-4 tune again describes the data slightly better than the more recent UE-5 tune used in HERWIG++ version 2.7.0.

**Charged particle multiplicities for bins of pseudorapidity** are depicted in Figs. 11.4 and 11.5. The comparison of the data results with predictions from Monte Carlo event generators shows the same general features as discussed for the integrated multiplicity distribution. Predictions of the old PYTHIA 6 and PHOJET generators all underestimate the measured multiplicities. The LHCb tune gives the best agreement with data, but the difference remains large. Towards small  $\eta$ , where the minimum momentum requirement reduces the amount of particles, the difference between data and predictions is most prominent. Recent generators predict multiplicity distributions that match the data much better. Both, PYTHIA 8 and HERWIG++, show good agreement with the data at larger pseudorapidities, only the range from  $2 < \eta < 3$  is still underestimated.

**Charged particle multiplicities for bins of transverse momentum** are displayed in Figs. 11.6 and 11.7. Comparing the observed multiplicities with the predictions from the event generators does not give a simple answer. For the lower  $p_T$  bins, the LHCb tune describes the measurements better than the other tunes. However, in the highest  $p_T$  bin, where the discrepancies between the LHCb tune and the data are most striking,

## 11. Results of charged particle densities and multiplicities

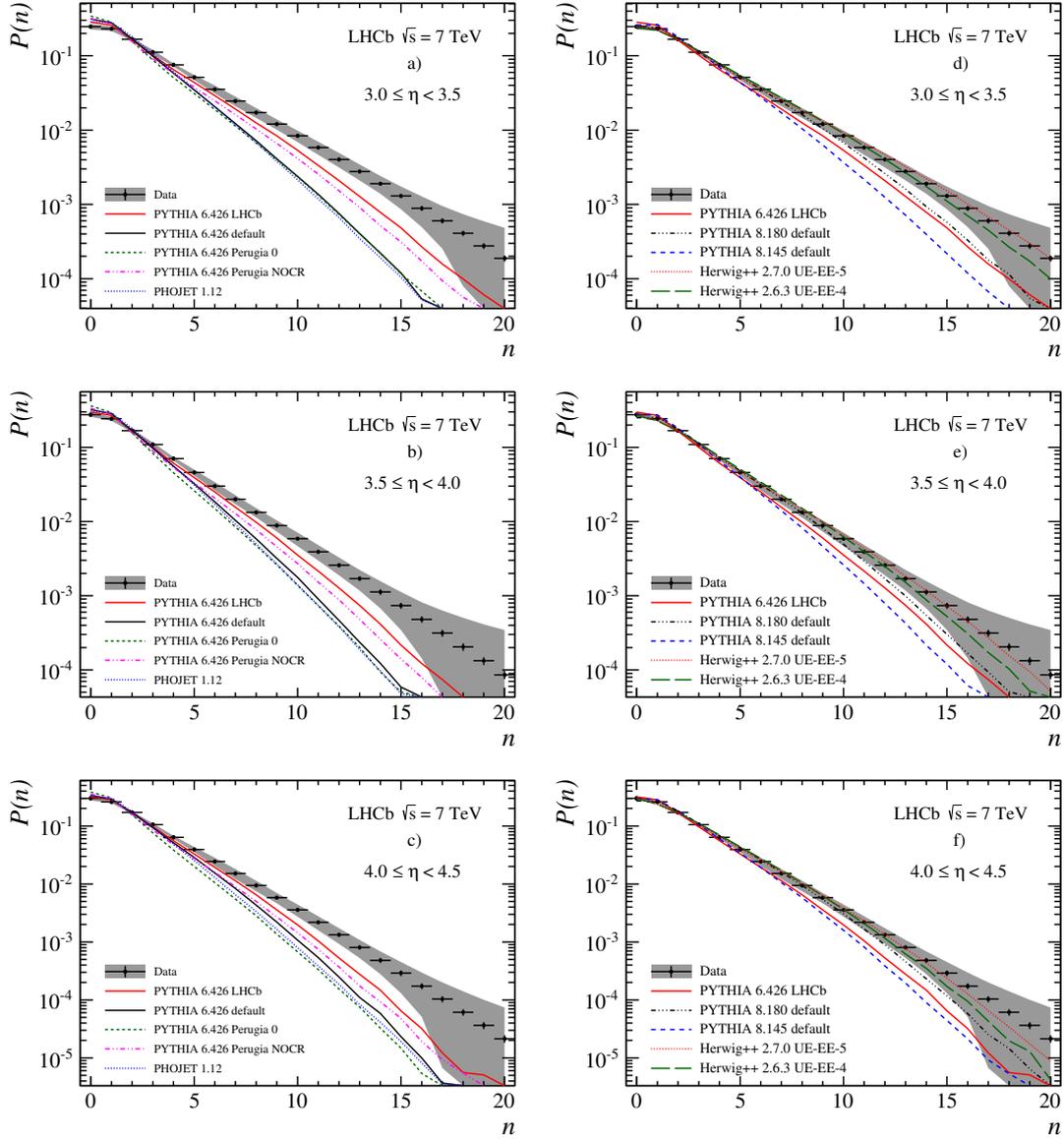


Figure 11.5.: Observed charged particle multiplicities in different  $\eta$  bins. Error bars represent the statistical uncertainty, error bands show the combined statistical and systematic uncertainties. The data are compared to several event generator predictions.

the default PYTHIA 6.426 configuration reproduces at least the shape of the measured multiplicity distribution. In general, all old generators and tunes, except the LHCb tune, again underestimate the charged particle production. The PYTHIA 8 generator in recent configuration shows a reasonable good agreement to data in the mid- and high- $p_T$  range, in which also the HERWIG++ generator describes the data. Predictions of HERWIG++ using the UE-4 tune are always closer to the measurement than predictions using the UE-5 tune. Towards large  $p_T$ , both HERWIG++ predictions underestimate the amount of particles while the PYTHIA 8 prediction is slightly better. To the contrary, PYTHIA 8

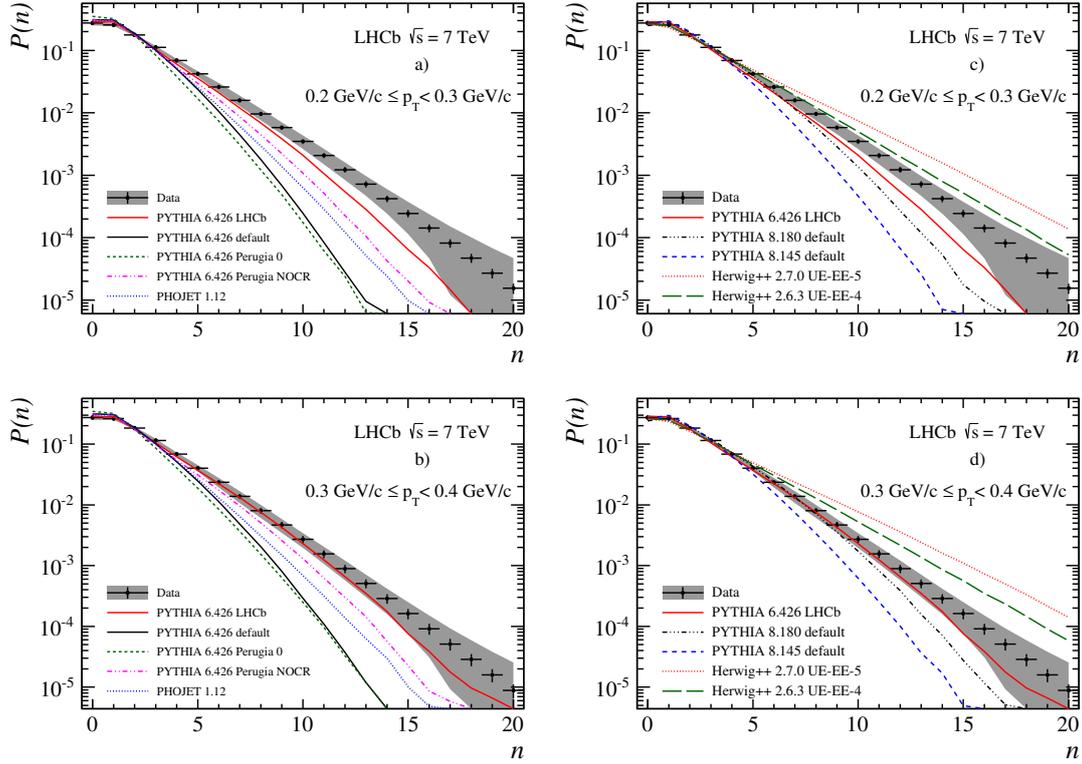


Figure 11.6.: Observed charged particle multiplicities in different  $p_T$  bins. Error bars represent the statistical uncertainty, error bands show the combined statistical and systematic uncertainties. The data are compared to several event generator predictions.

underestimates the data towards lower  $p_T$  and HERWIG++ overestimates it.

Summarising the results of the measurement in combination with the event generator predictions gives ambiguous conclusions. None of the presented generators and tunes is able to describe the entire range of measurements. As expected, the predictions based on generators that are not optimised to LHC measurements show the largest discrepancies. For the old event generators, the LHCb tune of PYTHIA 6 shows the best performance in describing the data from the forward region. However, large discrepancies are observed.

Recent event generators, such as PYTHIA 8 and HERWIG++, perform much better in describing the data. The improvements are strongly depending on the set of parameters that is used in the respective generator tune. PYTHIA 8 predictions using pre-LHC tunes are not better than previous PYTHIA 6 predictions. However, by tuning the generator to LHC measurements in the central rapidity region, reasonable good agreement with the LHCb data is obtained.

Both, HERWIG++ and PYTHIA 8 simulations, perform well in describing the particle multiplicities in some of the kinematic regions of the analysis. Towards higher  $p_T$ - and smaller  $\eta$ -ranges of the probed kinematic region the particle production is still underestimated and further optimisations of the models and its parameters are required.

## 11. Results of charged particle densities and multiplicities

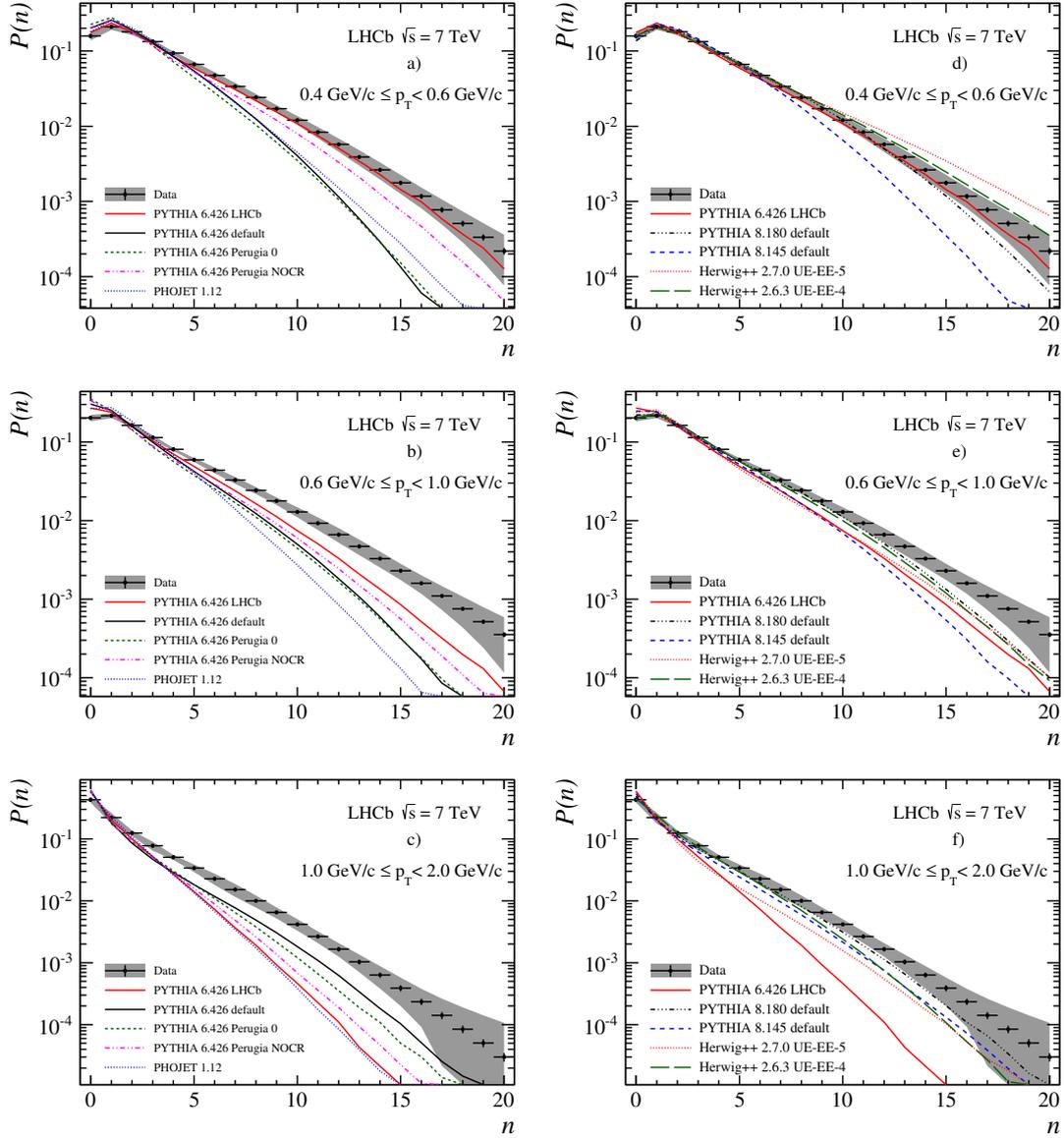


Figure 11.7.: Observed charged particle multiplicities in different  $p_T$  bins. Error bars represent the statistical uncertainty, error bands show the combined statistical and systematic uncertainties. The data are compared to several event generator predictions.

The presented measurement is also available in form of full data tables, which are published at [HepData](#) [122]. Further, a public Rivet plug-in [123] has been developed for this analysis, which provides a common tool-kit for future validations of Monte Carlo event generators.

## Part II.

Two-particle correlations in  
 $p+\text{Pb}$  and  $\text{Pb}+p$  collisions at  
 $\sqrt{s_{NN}} = 5.02 \text{ TeV}$



---

## An introduction to two-particle correlations

---

Angular correlation studies of particles produced in high-energy collisions are an important experimental tool to test particle production mechanisms and to investigate collective effects arising in the dense environment of a collision. In heavy-ion physics, correlation measurements are found to be very valuable in order to characterise the properties of the strongly interacting medium produced in ultra-relativistic nuclear collisions. Recent observations of unexpected long-range correlations in high-multiplicity proton-proton collisions, previously only found in heavy-ion collisions, triggered new interest in the understanding of non-perturbative particle production.

In this second part of the thesis, an analysis of two-particle correlations is presented. The differences in pseudorapidity,  $\Delta\eta$ , and azimuthal angle,  $\Delta\phi$ , between particle pairs are used to define two-dimensional correlation functions. The analysis is performed with proton-ion data recorded by the LHCb experiment in 2013, at a nucleon-nucleon centre-of-mass energy of  $\sqrt{s_{NN}} = 5.02$  TeV.

As an introduction, this chapter first gives a brief overview of the history of two-particle correlation studies performed at various experiments. This is followed by a more detailed explanation of the properties that are visible in the correlation functions, and a section dedicated to the theoretical interpretation of *the ridge* phenomenon.

The next chapters are focusing on the experimental part and the results of this measurement: The strategy of this analysis is presented in Chap. 13 in order to give an overview of the following chapters, which are related to the data preparation (Chap. 14) and the analysis procedure (Chap. 15). The results are presented in Chap. 16, followed by a discussion about the impact of systematic uncertainties in Chap. 17. This analysis is concluded, together with the charged particle multiplicity measurement (part I), in the final chapter of this thesis, Chap. 18.

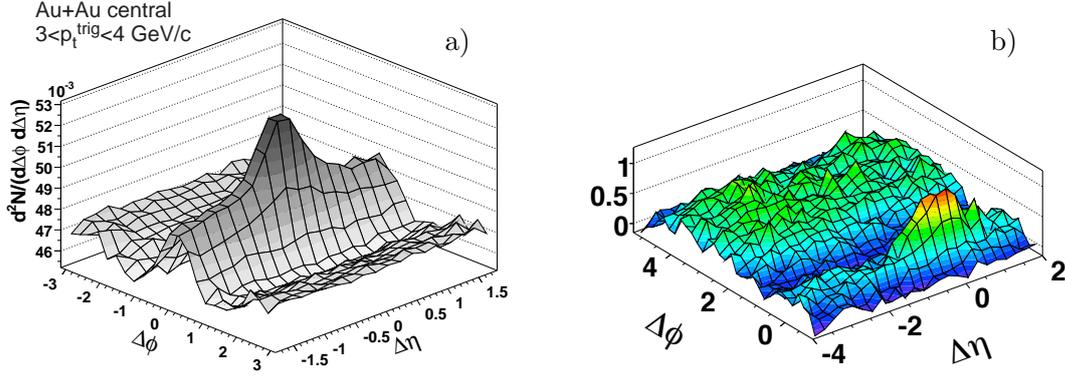


Figure 12.1.: Particle correlations measured in central AuAu collisions at  $\sqrt{s_{NN}} = 200$  GeV by the STAR [128] (a) and PHOBOS [129] (b) experiments. The STAR measurement shows the first observation of the near-side ridge in nucleus-nucleus collisions, indicating an collective long-range correlation that has not been expected from  $pp$  collisions at the same energy.

## 12.1. Brief review of recent two-particle correlation measurements

Already in the 1970s and 80s, experiments at CERN (see *e.g.* Refs [124, 125]) used measurements of final-state particle correlations in rapidity or azimuthal angle to investigate the mechanisms of particle production.

Experiments at the *Relativistic Heavy Ion Collider* (RHIC) established correlation studies as a very useful tool in analysing particle production, especially in the context of a high-density medium created in nuclear collisions. In such correlation studies, final-state particles that are directly produced in the collision are grouped into all possible combination of pairs. For each of these particle pairs, the relative azimuthal angle,  $\Delta\phi$ , and the separation in pseudorapidity,  $\Delta\eta$ , is calculated. When considering a large number of pairs, conclusions can be drawn about the angular correlations between these particles.

Early analyses performed with ultra-relativistic nuclear collisions (*e.g.* Refs [126, 127]) were investigating particle correlations one-dimensional, either as a function of  $\Delta\phi$  or  $\Delta\eta$ . These important measurements showed that the yield of correlated recoil particles from a high- $p_T$  particle produced in nuclear collisions is suppressed for high-transverse momenta, but enhanced in the low- $p_T$  range. This softening of fragmented jets suggested the presence of a medium that is opaque for particles produced in the direction pointing to the centre of the collision.

Later analyses were then studying correlations in terms of an associated yield of particles using two-dimensional  $(\Delta\eta, \Delta\phi)$ -correlation functions. The STAR collaboration published an analysis [128] of charged di-hadron correlations in gold-gold (AuAu) collisions at  $\sqrt{s_{NN}}=200$  GeV. The displayed correlation function in Fig. 12.1 a) shows the presence of a narrow peak at small angular separation, which is related to decay products

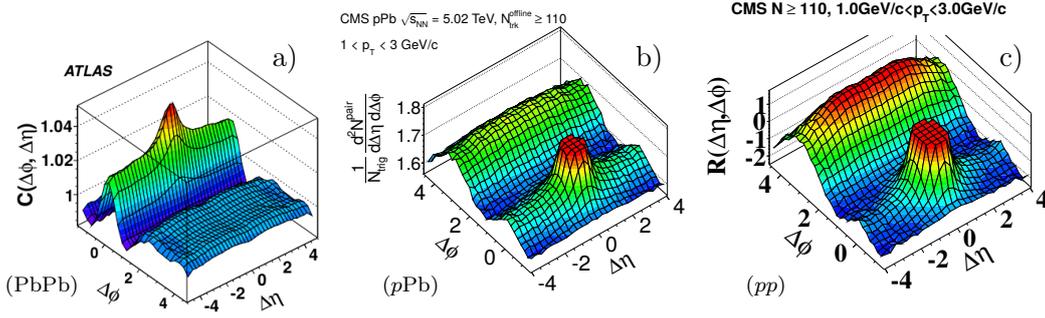


Figure 12.2.: Two-dimensional particle correlations measured in different collision systems at the LHC. (a) Central (0-1%) PbPb (ATLAS) [130], (b) high-multiplicity  $p$ +Pb (CMS) [8], and (c) high-multiplicity  $pp$  (CMS) collisions [7]. All collision systems show basic analogue structures even though they can be of different origin.

from jet-like objects. This peak has also been confirmed in deuterium-gold (dAu) and  $pp$  collisions. In addition, a new and not-understood structure at  $\Delta\phi \approx 0$  (*near-side*) which is elongated in  $\Delta\eta$  over almost two units has been observed, but only in central AuAu collisions. This phenomenon was henceforth referred to as **the ridge**. A more detailed introduction to the visible correlation properties is given in the subsequent section.

The PHOBOS experiment [129] extended this measurement of long-range correlations up to four units in pseudorapidity, see Fig. 12.1 b). The histogram also shows a broad long-range structure, centred around  $\Delta\phi \approx \pi$  (*away-side*), which is understood as representing the momentum balance of the jet-related peak.

Analyses on angular correlations have been continued at the LHC for various colliding systems ( $pp$ , PbPb and  $p$ Pb) and at different centre-of-mass energies. An exemplary collection of correlation measurements in central or high-multiplicity collisions is given in Fig. 12.2. Focusing only on the near-side ridge, the observation of this phenomenon in AuAu collisions has also been found in PbPb collisions (a) at the LHC [130–133]. In addition to the expected rediscovery in heavy-ion collisions, the CMS experiment [7] has also found a less pronounced but similar shaped ridge structure in  $pp$  collisions with very high particle multiplicities (c). This intriguing result is in conflict to theoretical ideas describing the near-side ridge in heavy-ion collisions in the context of hydrodynamic flow [134–138] of the produced strongly interacting medium. After finding this unexpected discovery in  $pp$  collisions, a similar long-range correlation has also been observed in  $p$ Pb collisions [8] consisting of a mixed beam configuration of protons and lead-ions (b).

While all existing measurements at the LHC are probing the central rapidity region up to  $|\eta < 2.5|$ , the LHCb detector allows particle correlation studies in  $pp$  and  $p$ Pb collisions<sup>1</sup> to be performed in the forward region ( $2 < \eta < 5$ ). This kinematic range has not been probed before at LHC energies.

<sup>1</sup>LHCb did not record PbPb data due to the very high particle multiplicity in PbPb collisions. The resulting occupancies in the tracking detectors are too large to perform an event reconstruction.

## 12.2. Properties of the two-dimensional correlation function

The visualisation of  $(\Delta\eta, \Delta\phi)$ -correlation functions are rich in structure. These depend on the global activity of the event, which is related to peripheral or central collisions, and on the kinematic range of the probed particles. The general shapes that can be found in correlation plots are basically similar for colliding protons, heavy-ions or mixed systems of protons and ions. To the contrary, the physical explanation of the features may not be the same and are also not yet fully understood. In common notation, one distinguishes between *short-range* ( $|\Delta\eta| \lesssim 2$ ) correlations and *long-range* ( $|\Delta\eta| \gtrsim 2$ ) correlations.

The short-range correlations can be phenomenologically described by the concept of *cluster emission* or an *independent cluster model* (ICM), see *e.g.* Refs. [140–143]. Within this approach, hadrons are not produced individually but in clusters which decay isotropically in their rest frame into a certain number of final-state particles. This approach has been successfully applied to many two-particle correlation measurements in the past, *e.g.* Refs. [7, 124, 144, 145], where parameters of this phenomenological description have been determined. However, no insights on the underlying mechanisms that form these clusters could be derived from this approach.

Long-range correlations have different sources. The correlations of jet-like objects and momentum conservation of particles are responsible for structures that are visible back-to-back in azimuth. These evident kinematic properties are included in Monte Carlo event generators and are easily reproduced in simulations.

An additional, more complex, long-range structure is observed when analysing an intermediate  $p_T$ -range of particles in events with a very high event activity. The high particle density that is reached in these events is the prerequisite for the mechanism that form the near-side ridge. The origin of this mechanism is still under discussion and not yet fully understood.

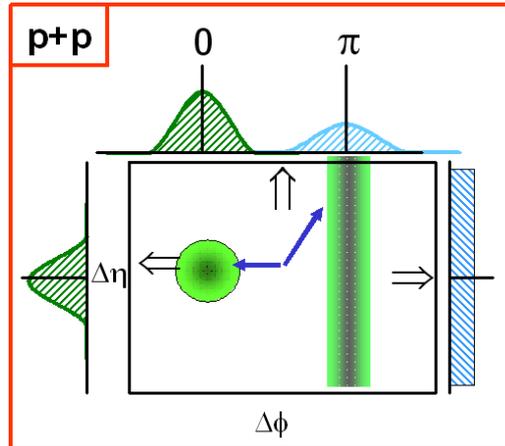


Figure 12.3.: Schematic illustration of the jet-induced correlation signal for  $pp$  collisions in the  $(\Delta\eta, \Delta\phi)$ -plane. A similar signal is also found in  $pPb$  collisions. The green circle at  $\Delta\eta = \Delta\phi = 0$  indicates the near-side peak, the green-grey structure at  $\Delta\phi = \pi$  which is elongated in  $\Delta\eta$  marks the away-side ridge. More details are explained within the text, the figure is taken from Ref. [139].

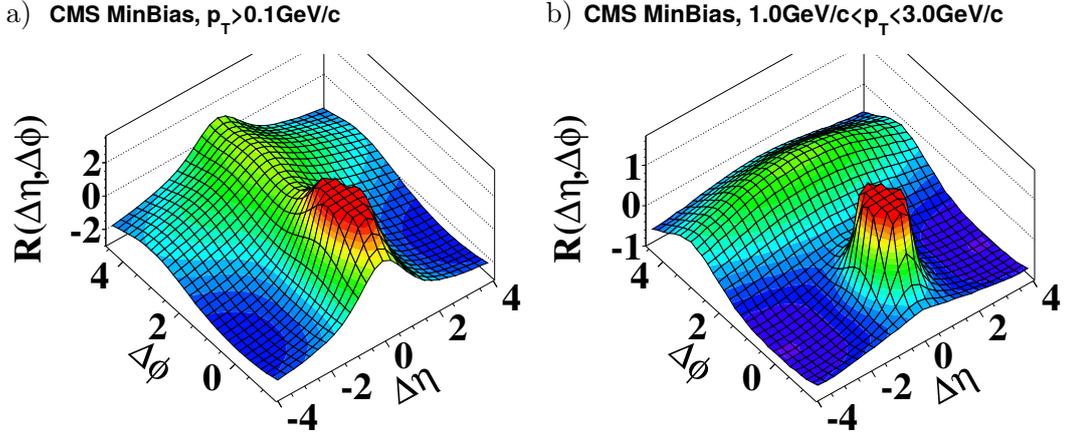


Figure 12.4.: Two-particle correlations in minimum-bias  $pp$  collisions at 7 TeV centre-of-mass energy, measured by CMS. In (a) all particles with an inclusive  $p_T$ -range of  $p_T > 0.1$  GeV are considered, in (b) only particles in the range of  $1 < p_T < 3$  GeV are selected. Figures are taken from Ref. [7].

In the following, the most prominent features that are visible in the two-dimensional correlation function are explained by using results from CMS [7, 8, 131] on  $pp$ , PbPb and  $p$ Pb collisions. A schematic illustration of jet-related contributions manifesting in the  $(\Delta\eta, \Delta\phi)$ -correlations is depicted in Fig. 12.3 by using  $pp$  collisions as an example. This section is intended to give a qualitative introduction, a more quantitative discussion follows when presenting the results of this analysis using the LHCb proton-ion data.

**Near-side peak at  $(\Delta\eta, \Delta\phi) \approx (0, 0)$**  A narrow peak typically dominates the complex structure in each correlation function around the origin. The peak arises from particles originating from the same jet-like hard process, which has a relatively high transverse momentum compared to the remaining underlying soft event. These particles are aligned along the leading- $p_T$  particle and only have a very small angular separation to each other. Comparing the result of the  $p_T$ -inclusive selection (Fig. 12.4 a) with the restricted intermediate- $p_T$  range of  $1 - 3$  GeV (Fig 12.4 b), a narrowing of the peak is observed. The appearance of the jet-like structure in correlation histograms is also indicated in the schematic illustration in Fig. 12.3. Accordingly, this structure is also observed in proton-ion and ion-ion collisions. The magnitude of the near-side peak varies depending on the configuration of the colliding system and the respective event activity. Also, the particle's momentum affects the size of this peak. In most of the cases, the characteristic of the peak dominates all other structures and thus, the peak is usually truncated in histograms.

**Gaussian ridge at  $\Delta\eta \approx 0$**  A broad ridge structure centred at  $\Delta\eta \approx 0$  but extending over the entire  $\Delta\phi$ -range emerges when including particles with lower  $p_T$  in the correlation function, such as done in Fig. 12.4 a). The structure covers the entire  $\Delta\phi$ -range and

becomes broader in  $\Delta\eta$  with increasing  $\Delta\phi$ . Particles from decays of clusters with low- $p_T$  (e.g. from soft-QCD string fragmentation) or decays of low- $p_T$  resonances are produced isotropically in azimuth and lead to such short-ranged correlations. Simulation studies demonstrate that these correlations are qualitatively reproduced by low-mass resonances within the independent cluster model, c.f. Ref. [124], or by isotropic cluster decays (see Refs [144, 145]). Within the ICM, correlations as a function of  $\Delta\eta$  can be parametrised by a Gaussian distribution where the parameters refer to the average number of particles within the cluster. The Gaussian behaviour of the short-range correlations is eponymous for this ridge. This phenomenological parametrisation has recently been applied in analyses, for example in Refs [7, 145].

**Away-side ridge at  $\Delta\phi \approx \pi$**  Another pronounced ridge is present on the away-side ( $\Delta\phi \approx \pi$ ) and elongates over a broad range in  $\Delta\eta$ . This ridge is present over the entire  $p_T$ -range, but is more pronounced towards larger  $p_T$ . The away-side balances the transverse momentum of the near-side and can be interpreted as back-to-back jet correlations in azimuth. This is displayed in the schematic illustration of Fig. 12.3. In general, the recoil of a particle is balanced by another particle with an azimuthal angle shifted by  $180^\circ$ . This structure is more pronounced when balancing two jet-like objects (back-to-back jets). Depending on the  $p_T$  of the jets and the boost due to the total momentum of the colliding partons, the pseudorapidity separation of the jets and their particles can be of any value. Thus, only the acceptance of the detector is limiting the  $\Delta\eta$ -range and the away-side ridge is visible over the entire measured range, as shown in Figs 12.4 and 12.5 a).

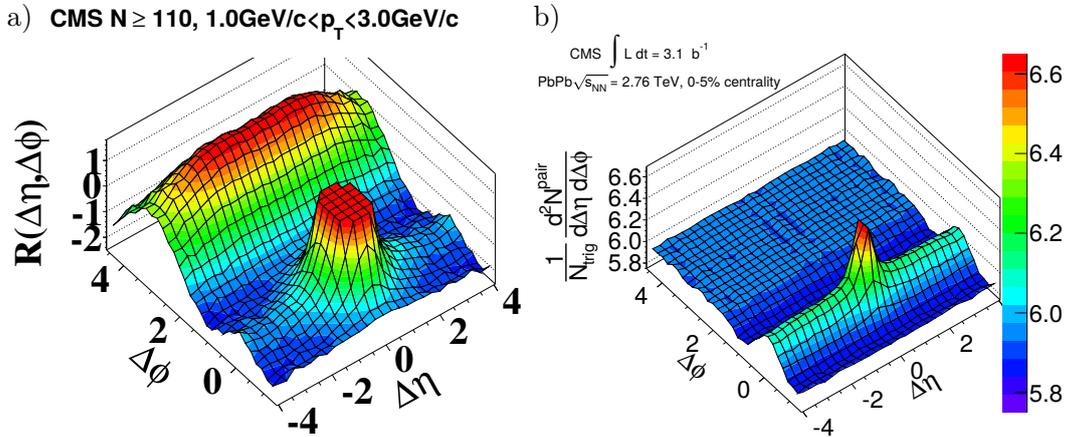


Figure 12.5.: The near-side ridge ( $\Delta\phi \approx 0$ ) measured by CMS for particles in an intermediate  $p_T$ -range in high-multiplicity  $pp$  collisions (a) and in central PbPb collisions (b), Figures taken from Refs [7] and [131], respectively.

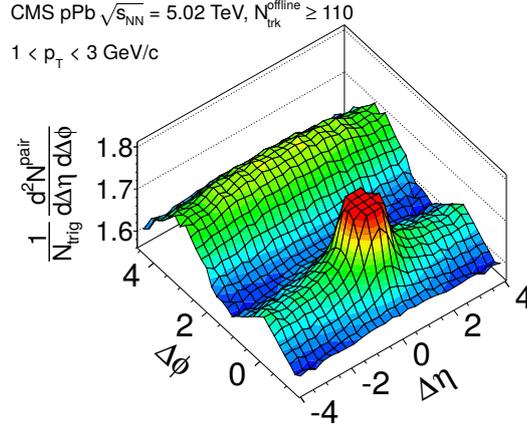


Figure 12.6.: The near-side ridge measured in proton-ion collisions with a very large track-multiplicity. Only particle pairs in an intermediate  $p_T$ -range are considered. Figure taken from Ref. [8].

**Near-side ridge at  $\Delta\phi \approx 0$**  An additional long-range correlation on the near-side can be seen in Fig. 12.5 a) for  $pp$  collisions with a very high particle multiplicity and for particles in an  $p_T$ -range of  $1 < p_T < 3 \text{ GeV}$ . The observation of the near-side ridge in  $pp$  collisions was unexpected even though the structure is reminiscent of *the ridge* seen in nucleus-nucleus collisions, as exemplified in Fig. 12.5 b) showing PbPb collisions. There is yet, no consistent explanation for this correlation comprising high energetic  $pp$  collisions and more complex systems of ultra-relativistic nuclei. In the heavy-ion community, the collective behaviour seen in nucleus-nucleus collisions is assumed to originate from the hydrodynamical flow of the created medium or from interactions between hard scattering processes and the medium. However, in  $pp$  collisions, the particle density is expected to be not high enough to create such a medium that could account for these long-range effects. Recent observations of this ridge structure also in proton-ion collisions at the LHC, as presented in Fig. 12.6, can help in understanding the common origin of this correlation. Apparently, also in collisions where no hot nuclear medium is created but still a very large number of particles are produced, some common collective behaviour is observed. The subsequent section is focussing on the theoretical interpretation of the near-side ridge in  $pp$  and  $pPb$  collisions and briefly explains the most popular models.

Up to now, the near-side ridge in proton-proton and proton-ion collisions has only been found in the central rapidity region. The target of this analysis is to also search for the ridge in the kinematic region accessible by the LHCb detector, the forward region. Proving or disproving its existence gives experimental input to theoretical explanations.

### 12.3. Theoretical interpretations of the ridge

The microscopic dynamics in collisions of  $pp$ ,  $p\text{Pb}$  and  $\text{PbPb}$  collisions, in which a huge amount of particles is produced, are not yet fully understood. Long-range correlations, and in particular the (near-side) ridge have been observed in all three colliding systems, but with a different characteristic strength. However, a similar origin or mechanism for this correlation structure is likely, but no common theoretical concept has been confirmed so far.

Correlations of final-state particles require a causal influence between the particles at an earlier time. Since the final-state correlation in the ridge structure are elongated over a long range in  $\Delta\eta$ , the causality constraints require that the correlation must be established shortly after the interaction [146]. Therefore, understanding the mechanism of the long-range correlation gives insights of the initial-state structure of the collisions and probes quantum fluctuations at very short time scales.

In heavy-ion collisions, long-range correlations, including the near-side ridge, are interpreted in the picture of hydrodynamical flow of the strongly interacting medium that is arising in the collision [148]. The collisions of nuclei are often described in two planes, the reaction plane defined by the  $x$ - and  $z$ -axis<sup>2</sup>, and the corresponding transverse plane, see Fig.12.7. By looking at the transverse plane it is found that the overlap region has an almond-like shape. As a result, the pressure in that region is anisotropic w.r.t the azimuthal angle,  $\phi$ , and depends on the eccentricity of the collision. This anisotropy produces a radial flow and results in an azimuthal momentum distribution, which can be expanded into a Fourier series with harmonic components,  $\cos(n\Delta\phi)$  [149]. These components are sensitive to the geometry and the initial state of the colliding system [150, 151] and probe the transport properties of the created medium using hydrodynamic models [152–154]. Each of the Fourier terms can lead to a maximum at  $\Delta\phi = 0$  but being independent of  $\Delta\eta$ . The sum of the components then build a ridge-like structure as it is observed in experimental data [133]. The dominant contributions are related to the  $n = 2$  and  $n = 3$  terms, which are referred to as elliptic ( $n = 2$ ) and triangular ( $n = 3$ ) flow, respectively.

<sup>2</sup>A right-handed coordinate system is assumed. For simplification, the colliding ions are further assumed to travel along and against the  $z$ -axis. In general, the reaction plane considers the impact parameter and the momentum vector of the projectiles.

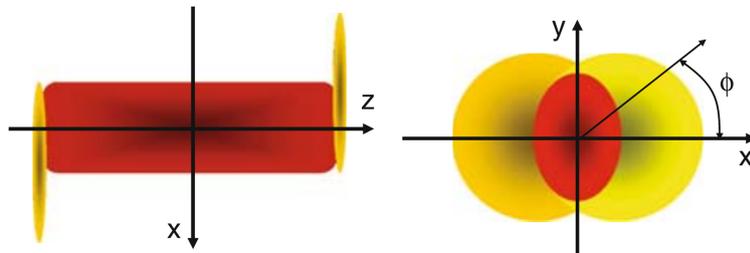


Figure 12.7.: Illustration of a heavy-ion collisions in the reference plane (left) and in the transversal plane (right). Figure is taken from Ref. [147].

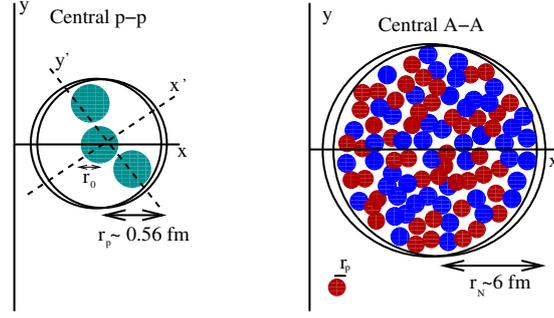


Figure 12.8.: Schematic view of the overlap region in  $pp$  (left) and nucleus-nucleus (right) collisions, displayed in the transverse plane. The coloured circles indicate hadronically active regions. Depending on the number and size of these regions, large eccentricities can result, even in central collisions. Figure taken from Ref. [156].

In proton-proton and also in proton-ion collisions, the mechanism that is responsible for long-range correlations on the near-side is less clear and still debated. Three different popular approaches are briefly presented in the following, based on the review in Ref. [155].

The idea of hydrodynamical flow can be also applied to the much smaller system of two colliding protons, see Fig. 12.8, when making the following assumptions: Collisions which produce a very large number of final-state particles are assumed to also have a large overlap, either in form of a geometrical overlap of the projectiles, or in terms of overlapping proton wave-functions. The hadronic activity in such a collision is clustering in a few "hot spots", even if emerging from a large number of multi-parton interactions. These hot spots can be identified with the constituent quarks within the protons, or can be also understood as local high parton densities. These are necessary for producing a large number of MPIs. The resulting density profile in the overlap region is thus generally not smooth nor isotropic. With the spatial limited interaction regions, c.f. Fig. 12.8, an initial-state eccentricity is generated, differently for each event. By assuming interactions between the final-state partons or hadrons, the initial-state fluctuations can lead to ridge-like two-particle correlations, similar to those of heavy-ion collisions. A theoretical calculation [157] of two-particle correlations in high-multiplicity  $pp$  collisions, based on the described hydrodynamical approach, results in a similar near-side ridge structure with a comparable magnitude. Further discussions about different order flow-effects arising in  $pp$  and  $pPb$  collisions are found in Refs [158–160] and [161–163], respectively.

The experimentally observed ridge correlation in high-multiplicity  $pp$  and  $pPb$  collisions shows a similar dependence on the  $p_T$  of the probed particles as found in heavy-ion collisions, for which the collective flow description was originally proposed. A rising strength of the correlation is observed in the intermediate  $p_T$ -range with a maximum of approximately 3 GeV, followed by a rapid decrease towards larger  $p_T$ . This similar behaviour supports the idea of a hydrodynamical effect that also arises in high-multiplicity  $pp$  and  $pPb$  collisions.

A second approach for explaining the observed ridge phenomenon is given in the framework of a colour-glass condensate (CGC) and the effect of gluon saturation at

small Bjorken- $x$ . [164, 165] In the picture of a heavy-ion collisions, the initial-state of the interaction is described by colour connections forming between the remnants of the two passing nuclei. These colour tubes are stretched in longitudinal direction and can produce particles which are naturally correlated over a long range in rapidity. According to CGC model calculations the final-state particles are intrinsically collimated in  $\Delta\phi$  over a long range in  $\Delta\eta$ . In contrast to the hydrodynamic model, the ridge is not driven by a radial flow.

Long-range correlation observed in  $dAu$  collisions are successfully described within the CGC approach [166]. Recent publications (*e.g.* Refs [167–169]) also claim an excellent agreement of calculation in the CGC framework with the ridge observations in high-multiplicity  $pp$  and  $pPb$  collisions at the LHC. This agreement includes the observed correlation strength as a function of the  $p_T$  of the probed particles as well as its dependence on the track-multiplicity of the events. Model calculations for heavy-ion collisions can be found in Refs [146, 170, 171].

The third category of theoretical interpretations considers a jet-induced formation of the ridge, also known as jet-medium interactions. Since hydrodynamics is not used in this model, the description can be applied to  $pp$ ,  $pPb$  and heavy-ion collisions. The idea of the underlying mechanism is the following [172]: It is implied that also in  $pp$  collision soft partons are created with a high density. Semi-hard partons (jets) can induce local fluctuation in these soft-parton fields by energy loss, when traversing this dense medium. Azimuthal anisotropy is produced by the initial geometric configuration, also valid in the small system of  $pp$  collisions for non-zero impact parameters. As a final consequence,  $\phi$  asymmetries are obtain which manifest in a ridge structure.

Another related mechanism is the Momentum Kick Model [173]. In this model, the ridge arises from soft partons that are produced in the dense medium created in the collision. These so-called *medium partons* collide with jets, which thereby lose some of their energy. On the other hand, the medium partons acquire a momentum kick along the flight direction of the jet.

In both jet-models it is predict that the ridge structure should persist towards very high transverse momenta. However, current measurements are not supporting this predictions, although no statement for  $p_T > 6$  GeV is possible due to statistical limitations.

The experimental prove or disprove of the discussed theoretical models is difficult, since they are arranged to describe existing observations and do not predict many model-specific observables. By decomposing the ridge structure (as a function of  $\Delta\phi$ ) in  $pPb$  collisions by a Fourier series, the measured elliptic and triangular flow coefficients are qualitatively in agreement with hydrodynamic calculations. Furthermore, the observed  $p_T$ -dependence of the near-side ridge in  $pPb$  collisions resembles the flow behaviour in heavy-ion collisions. However, also predictions from CGC models about the  $p_T$  and multiplicity dependence of the ridge are in good agreement with measurements. Only the jet-related models postulate a ridge-like correlation at high- $p_T$ . This provides a good signature, which is accessible with future correlation measurements.

---

## Strategy of the analysis

---

In this chapter the strategy used for the analysis of two-particle correlations in proton-ion collisions is presented.

The aim of this study is to measure angular correlations of charged particles that are directly produced in proton-ion collisions. For this purpose, two-dimensional  $(\Delta\eta, \Delta\phi)$ -correlation functions are investigated, where  $\Delta\eta$  is the difference in pseudorapidity and  $\Delta\phi$  is the difference in azimuthal angle between two particles. These particle pairs are formed for all possible unique combinations of particles in a pre-defined kinematic range. To obtain a statistically significant statement several hundred million events are analysed.

The analysis in a nutshell comprises four parts which are briefly explained in the following:

In the first part, Chap. 14, the preparation and selection of the data and the simulation are discussed. The used  $p$ +Pb and Pb+ $p$  data samples correspond to unselected samples containing all events recorded by the detector. The only requirement during data taking is the presence of at least a single reconstructed track segment per event. Without a pre-selection these data samples contain a significant number of background interactions. Besides proton-ion collisions, various parasitic interactions between particles of the LHC beams and material or residual gas molecules are recorded. The event selection aims for a suppression of events containing background interactions in order to provide a high-purity sample of proton-ion collisions. Afterwards, the selection of prompt particles is presented. Besides the kinematic restrictions in the analysis, quality requirements are defined with the objective of a small fraction of reconstruction artefacts in the track sample while retaining a high reconstruction efficiency.

The second part, Chap. 15, is dedicated to corrections applied to the track sample and the actual measurement of the two-particle correlations. The amount of contaminating mis-reconstructed and mis-selected tracks, as well as the quantification of inefficiencies are evaluated using a special  $p$ Pb simulation. Each track is assigned an individual correction factor, when calculating the two-particle correlation functions. Since correlation properties depend on the particle density of the respective events, a classification of

the event-activity is given. Two different approaches are implemented in this analysis that are either based on the hit-multiplicity or the track-multiplicity of an event. The results obtained in activity-classes of hit-multiplicity are considered as the central results, while results using the track-based classification serve as a cross-check. Afterwards, the definition of the correlation function is discussed. Two-particle correlations are measured for particles within the same event, referred to as *signal* correlations. *Background* correlations are modelled by mixing particles from different events. Artificial correlations induced by a limited detector acceptance or by reconstruction artefacts are removed by dividing the signal by the background distributions.

In the third part, Chap. 16, the final results of the analysis are presented, separately for  $p+\text{Pb}$  and  $\text{Pb}+p$  collisions. The data in each beam configuration are analysed independently for magnet-up and magnet-down polarity. This also includes individual correction factors which are determined from simulations with the respective beam and magnetic field configuration. Since the physical correlations are expected to be the same for data of both magnet polarities, their corrected results are averaged according to their statistical power. The rise of the near-side ridge in  $p+\text{Pb}$  and  $\text{Pb}+p$  collisions is presented in more detail. Furthermore, also a one-dimensional quantification of the correlation yield is given.

The fourth part, Chap. 17, is related to the estimation of systematic uncertainties. The utilised correlation function is already very robust against detector effects. No qualitative differences are expected when comparing results using raw-data or the fully-corrected data.

---

Preparation and selection of data and simulation

---

As discussed in Chap. 4.5, LHCb has recorded a large data set of proton-ion collisions at a nucleon-nucleon centre-of-mass energy of  $\sqrt{s_{NN}} = 5.02$  TeV, corresponding to a total integrated luminosity of  $\mathcal{L} = 1.7 \text{ nb}^{-1}$ . The collected data are divided into four sub-samples which differ by the direction of the two colliding beams, either  $p+\text{Pb}$  or  $\text{Pb}+p$ , and by the polarity of the magnetic field. In the course of this chapter, only data recorded in one magnetic field configuration, magnet-down, are presented, since there are no qualitative differences between both polarity samples.

The LHCb detector is designed to record  $pp$  collisions with a moderate instantaneous luminosity,  $\mathcal{L}_{inst}$ , at a low average interaction rate,  $\mu$ . During the  $pp$  data taking, the detector was already operated above design parameters with, on average,  $\mu = 1.7$   $pp$  collisions per bunch-crossing, corresponding to a luminosity of around  $\mathcal{L}_{inst} = 4 \cdot 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ . The specifications during the proton-ion ( $p\text{Pb}$ ) data taking are very different compared to that of proton-proton, because single  $p\text{Pb}$  collisions cause much larger detector occupancies and track multiplicities than single  $pp$  collisions. The average instantaneous luminosity of  $\mathcal{L}_{inst} \approx 3.5 \cdot 10^{27} \text{ cm}^{-2}\text{s}^{-1}$  and the average interaction rate in the range of  $\mu = [0.015, 0.040]$  are much smaller compared to that of the  $pp$  data taking. Furthermore, the event topology of one high-multiplicity  $p\text{Pb}$  collision is different from a high-multiplicity  $pp$  event with several pile-up interactions, in which the produced particles do not all originate from a single vertex. Depending on the beam configuration, either  $p+\text{Pb}$  or  $\text{Pb}+p$ , the maximum occupancy of the detector reaches much larger values. The data taking, in particular the trigger system which is optimised for  $pp$  collisions, was modified in order to allow an efficient recording of  $p\text{Pb}$  collisions.

Figure 14.1 shows the hit-multiplicity distributions in the SPD and in the VELO sub-detectors for the recorded  $p\text{Pb}$  collisions in both beam configurations. No event selection has been applied at this point. In contrast to an expected smooth distribution, the data samples in Fig. 14.1 show a step-like structure at 10000 VELO-hits. This is the result of a faulty implementation of the trigger, which affects events with very large

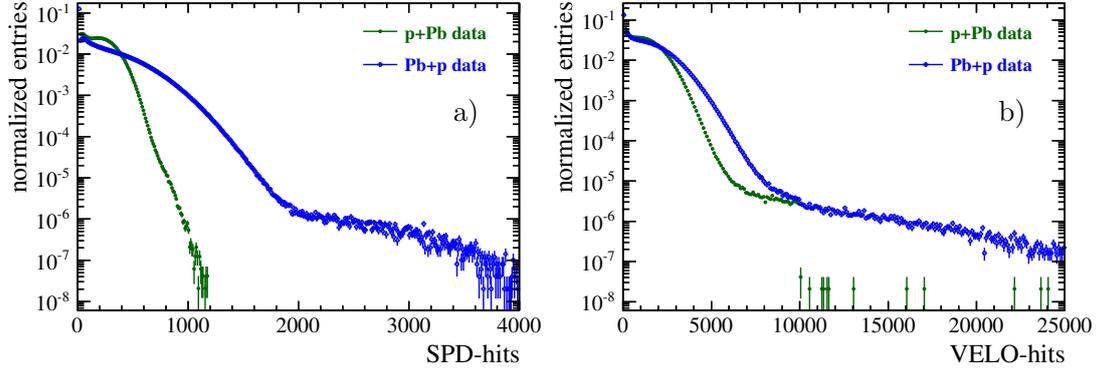


Figure 14.1.: Hit-multiplicity distributions in the SPD (a) and the VELO (b) for  $p+Pb$  and  $Pb+p$  collisions before any selection. Besides the different occupancies in the  $p+Pb$  and  $Pb+p$  configuration a clear step at 10000 VELO hits is visible, which is introduced by a trigger mis-configuration. The tail of the distributions is related to background interactions, which are discussed in the course of this chapter.

detector occupancies. The ridge structure intended to be found by this analysis, is expected to evolve in events with a high particle density. Thus, the analysis could be affected by this trigger mis-configuration which is investigated among other effects in the following sections.

This chapter addresses the preparation and selection of the recorded proton-ion data. At first the data sample used throughout this analysis are defined. Afterwards, the configuration of the trigger and the resulting problems in the high multiplicity range are discussed. This is followed by the event selection. The aim of this selection is to suppress beam-related backgrounds that pollute the event samples. In this context, the parameters of the custom-made simulation sample are presented. The chapter closes with the track selection, which defines the fiducial range of the analysis and also reduces the contamination by reconstruction artefacts.

| Data sample    | Beam configuration | Magnet polarity | Number of events    | Integrated luminosity $\mathcal{L}$ |
|----------------|--------------------|-----------------|---------------------|-------------------------------------|
| minimum-bias   | $p+Pb$             | down            | $50.5 \times 10^6$  | $22 \mu\text{b}^{-1}$               |
|                | $p+Pb$             | up              | $56.7 \times 10^6$  | $27 \mu\text{b}^{-1}$               |
|                | $Pb+p$             | down            | $52.4 \times 10^6$  | $23 \mu\text{b}^{-1}$               |
|                | $Pb+p$             | up              | $51.4 \times 10^6$  | $23 \mu\text{b}^{-1}$               |
| high-occupancy | $p+Pb$             | down            | $561.2 \times 10^6$ | $240 \mu\text{b}^{-1}$              |
|                | $p+Pb$             | up              | $477.1 \times 10^6$ | $226 \mu\text{b}^{-1}$              |
|                | $Pb+p$             | down            | $382.3 \times 10^6$ | $161 \mu\text{b}^{-1}$              |
|                | $Pb+p$             | up              | $306.5 \times 10^6$ | $137 \mu\text{b}^{-1}$              |

Table 14.1.: Overview of the  $pPb$  data samples used throughout this analysis. The nucleon-nucleon centre-of-mass energy is  $\sqrt{s_{NN}} = 5.02 \text{ TeV}$  in all samples.

## 14.1. The data samples

The entire recorded data sample of  $p\text{Pb}$  collisions recorded with the LHCb detector amounts to an integrated luminosity of approximately  $\mathcal{L} = 1.7 \text{ nb}^{-1}$ , corresponding to  $\mathcal{L} = 1.1 \text{ nb}^{-1}$  in the  $p+\text{Pb}$  configuration and  $\mathcal{L} = 0.6 \text{ nb}^{-1}$  in the  $\text{Pb}+p$  configuration. The proton-ion data are all collected by using a minimum-bias trigger configuration, which is explained in more detail in the next section. For this analysis only smaller subsets of the full data sample are used. In order to exploit the large available statistics, two different data samples are employed, as explained in the following:

**Minimum-bias sample:** The main data sample used in this analysis comprises four sub-samples of at least 50 million events in each beam configuration,  $p+\text{Pb}$  and  $\text{Pb}+p$ , and in magnet-up and magnet-down polarity. The events are randomly selected out of the entire data sample and are henceforth referred to as the *minimum-bias* data sample. In total, the data correspond to an integrated luminosity of  $\mathcal{L} \approx 95 \mu\text{b}^{-1}$ .

**High-occupancy sample:** As stated in the previous chapter, the near-side ridge correlation arises only in events with a large particle density. In the minimum-bias data sample, the fraction of high-multiplicity events is naturally only small. In order to increase the amount of high-multiplicity events in the analysis, an additional *high-occupancy* data sample is used, again consisting of four samples for each beam and magnet configuration. A simple pre-selection based on the number of hits in the VELO is applied to select events with a large number of particles. It is required that events have at least 2200 hits in the VELO. The choice for this selection is motivated by the classification of high-multiplicity events, which is discussed in Chap. 15.2. No further pre-selection is applied. The high-occupancy sample corresponds to an integrated luminosity of  $\mathcal{L} \approx 764 \mu\text{b}^{-1}$ .

Throughout this document, the minimum-bias samples are used unless stated differently. The high-occupancy samples are used for the presentation of the final results, in order to increase the statistical power for analysing events with large particle multiplicities. The sizes of all individual samples are summarised in Tab. 14.1.

## 14.2. Trigger configuration

The principle of data acquisition and processing of proton-ion collisions is performed in the same way as for proton-proton collisions. However, the aforementioned different interaction rates and detector occupancies during the  $p\text{Pb}$  and  $pp$  data taking required to operate the trigger in a different configuration. Due to the low interaction rate of proton-ion collisions a simplified trigger configuration is used, which is comparable to that of the low-luminosity  $pp$  data taking. The hardware implemented L0-trigger was operated in a pass-through mode forwarding all events to the high-level software trigger. The HLT was configured to select and store all events with a minimum activity in the detector. However, an unintended relict of the  $pp$  data acquisition rejected  $p\text{Pb}$  events

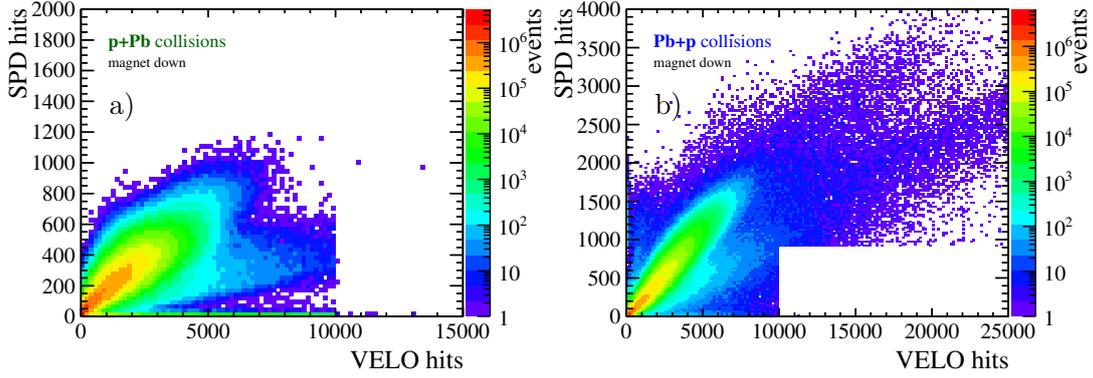


Figure 14.2.: Correlation between hit-multiplicities in the VELO and the SPD, in  $p+Pb$  (left) and  $Pb+p$  data (right). Cuts in the histograms indicate a trigger mis-configuration, as explained in the text. Note the logarithmic scale and the different  $x$ -axis ranges.

with very high detector occupancies. The effect of this mis-configuration can be seen in Fig. 14.2, which displays the hit-correlation between the SPD and VELO sub-detectors for data recorded in  $p+Pb$  and  $Pb+p$  beam configuration. In both data samples a sharp cut-off at a VELO-hit multiplicity of 10000 is present. This is identified as a remaining *Global Event Cut* (GEC) which was implemented in the trigger during the standard LHCb data taking.

Several GECs have originally been implemented in order to protect the high-level trigger and the offline track-reconstruction from processing of events that are too busy. In particular during the high-luminosity  $pp$  data taking, the trigger is required to work fast and to take a decision within microseconds. In contrast, during the  $pPb$  data taking, no GECs are needed. Due to the low collision rate essentially all events can be processed and no event-specific trigger is required. However, technically a trigger system is necessary to operate the detector, but also to preselect specific events for exclusive analyses. As a consequence, the trigger has been modified in order to select as many events as possible and to not introduce any intentional selection. The still present GEC in the VELO, however, was missed to be removed.

In the  $pPb$  data taking, the software trigger basically consists of a few inclusive triggers, which only use global event quantities instead of particle or decay-specific properties to take a decision. In addition, a few exclusive physics triggers have been retained but are not included in this discussion. All events selected by these exclusive trigger lines only build sub-samples of the inclusive triggered events sample.

The entire recorded data samples, suitable for physics analyses, have been selected by only two trigger lines. There is also a small number of random triggered events used for luminosity determinations or trigger performance studies, but which are not of interest for this analysis. Due to the low interaction rate, in principle a single micro-bias trigger, which selects events with at least one reconstructed track segment in the VELO<sup>1</sup>, is sufficient to identify all interesting events. However, Fig. 14.2 indicates that only events with at maximum 10000 VELO-hits are selected by this specific trigger. Beyond this limit, one of the GECs aborts the decoding of the VELO. The subsequently carried out

<sup>1</sup>the specific trigger line is the `Hlt1MBMicroBiasVelo`.

| Trigger line        | Trigger condition      | Global event cut |
|---------------------|------------------------|------------------|
| Hlt1MBMicroBiasVelo | $\geq 1$ track in VELO | 10000 VELO-hits  |
| Hlt1ActivityVelo    | $\geq 6000$ VELO hits  | 10000 VELO-hits  |
| Hlt1ActivitySPD     | $\geq 900$ SPD hits    | no limit         |

Table 14.2.: Relevant trigger lines in the proton-ion data taking. Any other (not listed) trigger line only selects a subset of events.

online track reconstruction in the HLT does not see any detector hits and thus cannot reconstruct tracks. As a result, the track-based micro-bias trigger discards these events, since no tracks are reconstructed.

To avoid that potentially interesting events with more than 10000 VELO-hits are rejected, two *activity* triggers are included in the HLT. Instead of performing a time-consuming track reconstruction, these trigger lines only require a certain hit-multiplicity in a given sub-detector, namely in the VELO or in the SPD. The VELO-based activity trigger selects events with  $> 6000$  VELO-hits and complements the track-based micro-bias trigger. As being affected by the same GEC in the decoding algorithm, also the VELO-based activity trigger failed in selecting the high-occupancy events.

The remaining SPD-based activity trigger works independently of the aforementioned triggers, it selects events with  $> 900$  SPD-hits. The lower boundary of the SPD-activity trigger is clearly visible in the hit-correlation of the Pb+*p* data sample, depicted in Fig. 14.2 b). Events with a hit-multiplicity that is high in the VELO but moderate in the SPD remain undetected, only events with more than 900 SPD-hits are recorded.

For the *p*+Pb data sample, Fig. 14.2 a), the situation is different, because in general smaller occupancies are accessed compared to Pb+*p* collisions. While the limit of 10000 VELO-hits is also reached, there are almost no events found with a number of SPD-hits beyond the threshold of the SPD-activity trigger. Hence, the GEC of 10000 VELO-hits effectively marks the occupancy limit of the *p*+Pb data sample. The requirements of the three trigger lines are summarised in Tab. 14.2.

The presence of the unintended GEC during the data acquisition affects the high-occupancy events, which are intended to be studied in this analysis. However, the displayed hit-correlations point to the presence of background contributions which manifest in independent correlation structures towards higher occupancies. Already at this point, it should be noted that essentially no events relevant for this analysis are lost due to the trigger misconfiguration. The event selection, which suppresses these beam-related backgrounds, is presented in the following section.

### 14.3. Event selection

The recorded and reconstructed data used for physics analyses are typically pre-selected in order to suppress apparent backgrounds and to reduce the amount of data to a reasonable size. This presented proton-ion analysis cannot benefit from such an existing pre-selection. There is no experience from previous analyses or from simulation studies in order to prepare the data for this measurement. Also, the aim of this correlation measurement is very different from typical LHCb analyses in which specific particle

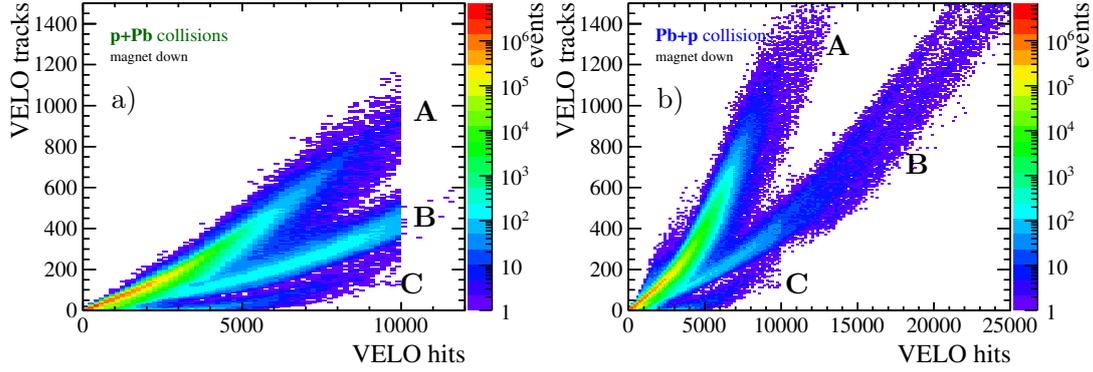


Figure 14.3.: Correlation of VELO-track to VELO-hit multiplicity for  $p+Pb$  (left) and  $Pb+p$  collisions (right). Capital letters indicate three different correlations which are discussed within the text. Note the logarithmic scale and different  $x$ -axis ranges.

decays are studied. For the presented analysis, basically all events that contain a proton-ion interaction are useful.

The data samples correspond to unselected inclusive sample of all recorded events. Along with the actual proton-ion interactions the event samples contain a significant amount of various backgrounds. In Figure 14.3, track-to-hit correlations in the Vertex Locator for the  $p+Pb$  and  $Pb+p$  data samples are depicted, which visualise the potential background events. Naively, only one correlation pattern is expected which links the increasing number of reconstructed tracks to the also increasing number of detector hits. The recorded data in both beam configurations, however, reveal three separate correlations, which are indicated by the capital letters A, B and C. Two of these will be identified as background. Please note, that the number of events in the presented histograms is given in a logarithmic scale.

The purpose of the event selection is to suppress all background interactions which contaminate the sample of proton-ion collisions, especially in the high multiplicity region, in which the ridge structure is expected to be present. The single selection criteria are discussed in the following.

**Bunch-crossing type** Both,  $p+Pb$  and  $Pb+p$  data show unexpected structures when looking at the track-to-hit correlation in Fig. 14.3. Provided that proton-ion collisions show a similar fundamental detector response like proton-proton collisions, the multi-correlation pattern has to originate from different kinds of background interactions. To investigate this hypothesis, sub-data samples with different bunch-crossing types are studied, which contain events with irregular beam interactions. Both beams which are brought to collisions at the LHCb interaction point are filled by bunches of protons or ions. These so-called buckets are organised in a certain filling scheme, but not all possible slots are assigned by a filled bucket. As a consequence, four types of bunch-crossings are possible: nominal interactions with a filled bunch in both beams, called *beam-beam* crossings; beam-gas interactions, where either of the two beams has no filled proton/ion bunch (*beam-empty*, *empty-beam* crossings); and *empty-empty* bunch crossings, where

neither of the two beams has a filled bunch at the time of the recorded interaction.

Irregular bunch-crossing, which are typically not used for physics analyses, should contain no  $p$ Pb collisions but instead various types of background interactions: Beam-gas interactions, where a colliding particle of one beam hits remaining gas molecules within the VELO; beam-splash interactions where a particle from one beam interacts with material of the detector's or accelerator's support structure; interactions between a nominal and a previous bunch; or interactions of a regular-filled bunch with an originally empty bunch that has been migrated by single particles from neighbouring bunches during the acceleration. These kinds of background interactions can be characterised by two facts. The energy which is available to produce particles is smaller compared to that of nominal collisions, and the average interaction point is displaced from the nominal point. This results in different track-to-hit correlations, as visible in Fig. 14.3, but also in different hit-to-hit correlation between different sub-detectors, as previously shown in Fig. 14.2.

The three structures found in the track-to-hit correlations are studied separately for events marked by the different bunch-crossing types. Looking first at beam-beam events in Fig. 14.4 a),b), the dominant correlation in both beam configurations is structure A. This specific correlation is denominated as signal. The structures B and C are less pronounced by several orders of magnitude, as indicated by the logarithmic scale.

By considering events recorded in empty-empty configuration (c, d), these structures are likely to be identified as background. The probability to have a regular proton-ion collision in this configuration is extremely small and thus the signal structure is only barely populated. Structures B and C can be related to any of the previously named background interactions. Detailed studies, preferably in simulation, would be required to assign the correct sources of these kinds of collisions, but for the presented analysis this is not necessary.

Studying the empty-beam and beam-empty events in Fig. 14.4 e-h), something interesting can be noticed. In beam-empty events (e-f), only the beam pointing upstream the detector is filled with particles. It is likely that the recorded events are caused by background interactions, most probably beam-gas interactions, which occur either in front of or within the VELO. Structures A and B are both strongly pronounced and, besides the larger occupancy tail in the Pb+ $p$  data, there is no fundamental difference between both beam configurations.

Looking at empty-beam events (g-h), in which only the downstream beam is present, the picture is different. If the proton-beam is present (h), only few events with low multiplicities are recorded. In the presence of the ion-beam (g) lots of activity is recorded. The lead-ions have a larger cross-section than the protons and thus also have a larger probability to interact with gas molecules in the VELO. Furthermore the ion-beam seems to be less collimated compared to the proton-beam, which results in a larger probability to interact with material. Obviously, also the available energy to produce particles is greater in ion- compared to proton-interactions.

The track-to-hit correlations are also checked in the proton-ion simulation. Only the signal structure A is present in both beam configurations. A more detailed study is not possible, since only actual proton-ion interactions and no beam backgrounds are included in the simulation.

14. Preparation and selection of data and simulation

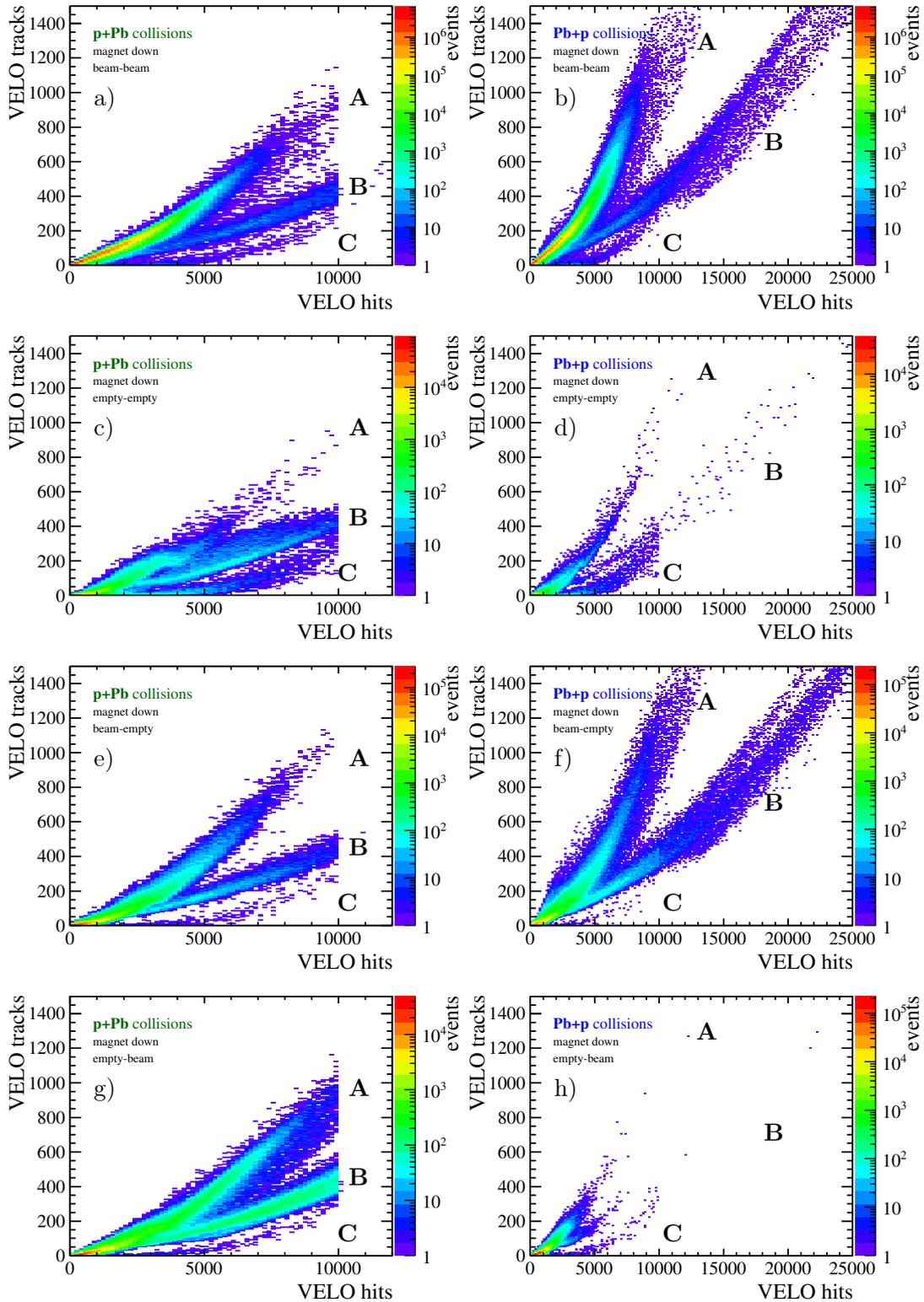


Figure 14.4.: Correlation of VELO-track to VELO-hit multiplicity in  $p+Pb$  (left) and  $Pb+p$  collisions (right). The structures labelled by capital letters are referred to in the text. Note the logarithmic scale and the different  $x$ -axis ranges.

In the further course of the analysis, the selected events are restricted to be recorded in the beam-beam bunch-crossing configuration. Concluding from the previous overview, the structures B and C are identified as background. Their contamination to the beam-beam sample (Fig. 14.4 a,b)) is less pronounced compared to the inclusive sample (14.3). The remaining contaminations are suppressed by other selection criteria.

**Primary vertex requirement** Another important criterion to select actual proton-ion collisions is to require the presence of exactly one reconstructed primary vertex per event. As aforementioned, the very low average interaction rate during the entire data taking leads essentially to either one or no  $p$ Pb collision per bunch-crossing. The maximum recorded peak interaction rate of  $\mu = 0.04$  corresponds to a pile-up probability of less than 2%. The number of pile-up events is negligible and also suppressed by permitting only events with one reconstruct primary vertex.

The algorithm used to reconstruct primary vertices requires a minimum number of three tracks originating from a common vertex. Due to the large number of particles produced in proton-ion collisions a high reconstruction efficiency is expected. The efficiency is tested by using simulated proton-ion samples, which contain only events with exactly one  $p$ Pb collision. The two beam configurations are studied separately. In the  $p$ +Pb simulation only 16 out of  $2.02 \times 10^6$  events, and in the Pb+ $p$  simulation 24 out of  $1.99 \times 10^6$  events have no reconstructed primary vertex. This corresponds to a negligible fraction of  $\approx 10^{-5}$  in which the algorithm is unable to find a primary vertex, and a reconstruction efficiency of effectively 100% can be assumed.

In the following, only events with exactly one reconstructed PV are considered in the analysis. Background interactions taking place outside the VELO produce particles whose trajectories do not intersect within the VELO acceptance. The reconstruction of a primary vertex is therefore prevented. The presence of more than one primary vertex points to additional pile-up and beam-gas interactions. Since a dedicated study of these contaminations is not possible with the existing simulation, events with multiple PVs are rejected. The impact of including multiple PVs in the measurement is considered as a systematic uncertainty.

**Luminous region** In addition to the PV requirement, the position of the reconstructed vertex is important. By constraining the position of the primary interaction vertex to a small region two advantages are obtained. First, particles that are produced in collisions whose interaction points are close together in space traverse the same detector regions and are affected by similar inefficiencies. Second, actual proton-ion collisions cluster around the nominal interaction point defined by the beam parameters, whereas background interactions, such as beam-gas collisions, are distributed along the entire beam line. Their amount is reduced by selecting events from a small defined region.

For this purpose, a luminous region is defined in a similar way as it is done in the multiplicity analysis, c.f. Chap. 8.2. The distributions of reconstructed PVs in all three space coordinates are depicted in Fig. 14.5, showing data and simulation in both beam configurations. Each distribution is typically composed of a Gaussian distributed

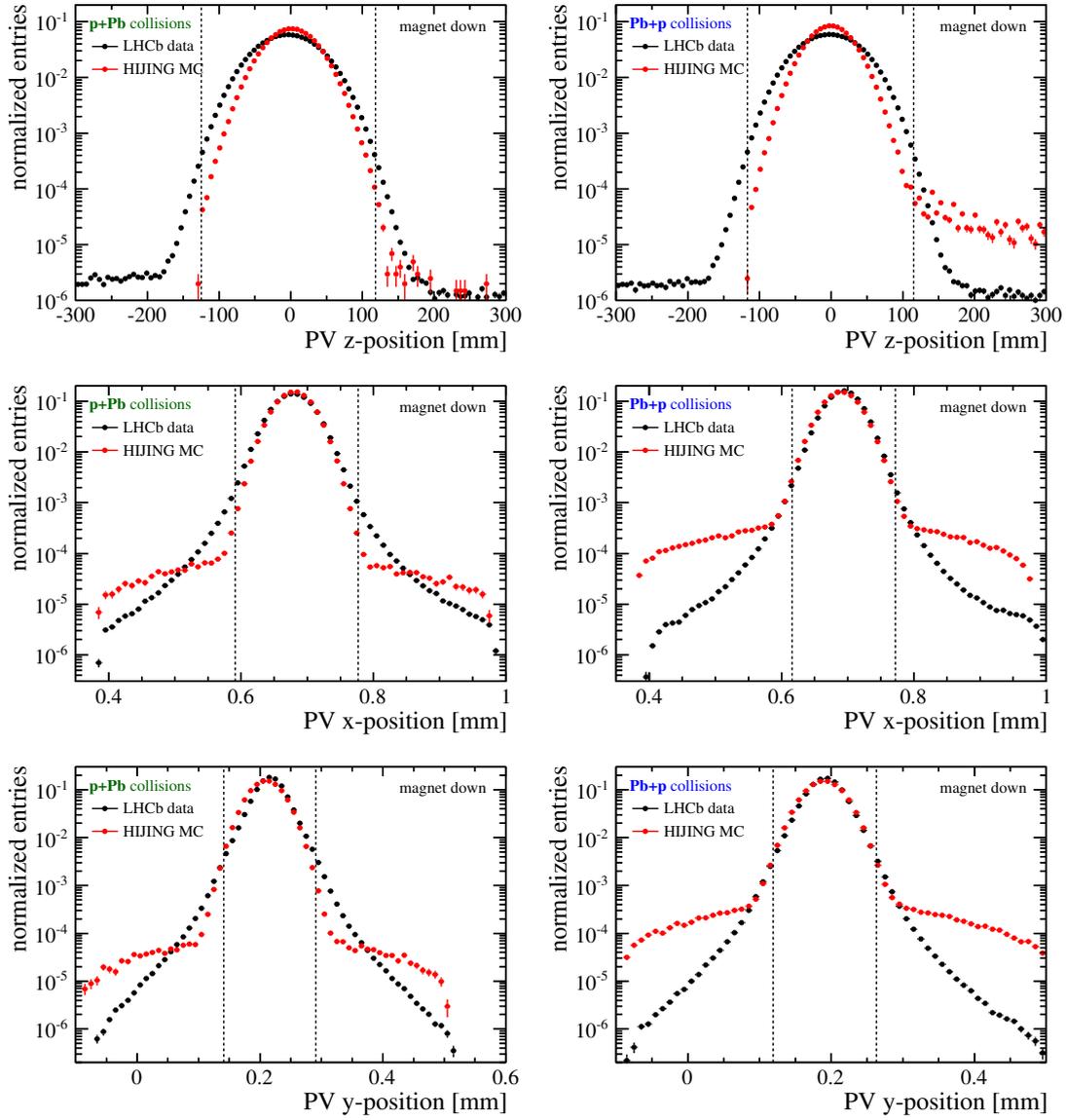


Figure 14.5.: Distribution of reconstructed primary vertices in  $p$ +Pb (left) and Pb+ $p$  (right) collisions for all three space-coordinates. Only magnet-down data samples are displayed. The data in black are compared to the simulation in red. The dotted lines indicate the borders of the luminous region.

bulk region and differently shaped background components. This is best visible for the longitudinal component, where the bulk region of actual  $p$ Pb collisions is nicely separated from a flat background. For the  $x$ - and  $y$ -component of the primary vertex position, the transition between the Gaussian distribution and the overlaid background is smeared out more strongly.

To define the luminous region, the mean value,  $\mu$ , and the standard deviation,  $\sigma$ , are determined from the distribution of the primary vertices within the Gaussian bulk

| Beam and magnet configuration |      | PV x-position [mm] |       | PV y-position [mm] |       | PV z-position [mm] |        |
|-------------------------------|------|--------------------|-------|--------------------|-------|--------------------|--------|
|                               |      | mean               | sigma | mean               | sigma | mean               | sigma  |
| $p$ +Pb                       | Down | 0.684              | 0.031 | 0.216              | 0.025 | -3.276             | 40.540 |
| $p$ +Pb                       | Up   | 0.697              | 0.025 | 0.209              | 0.023 | +3.859             | 60.920 |
| Pb+ $p$                       | Down | 0.694              | 0.026 | 0.191              | 0.024 | -1.064             | 38.660 |
| Pb+ $p$                       | Up   | 0.688              | 0.025 | 0.199              | 0.025 | +3.209             | 60.515 |

Table 14.3.: Overview of the parameters used to define the interaction region of proton-ion collisions in the custom-made HIJING Monte Carlo simulation.

region. Each dimension is considered separately. Originally, a range of three standard deviations around the respective mean value in each component is used to define the luminous region:

$$|x_{\text{PV}} - \mu_x| < 3\sigma_x, \quad |y_{\text{PV}} - \mu_y| < 3\sigma_y, \quad |z_{\text{PV}} - \mu_z| < 3\sigma_z. \quad (14.1)$$

During the development of the analysis more and more data have been included which results in an increasing spread of the PV positions. Given that the parameters  $\mu$  and  $\sigma$  obtained from a smaller data sample are also used to configure the interaction region in the custom-made HIJING simulation for this analysis (see below, Sec. 14.4), these old parameters are still used to define the luminous region. The set of parameters implemented in the simulation are listed in Tab. 14.3.

Figure 14.5 also compares the position of the reconstructed PVs in data and in simulation. The increased data sample size is reflected by a broader data distribution in the  $z$ -component compared to the simulated distribution. In the  $x$ - and  $y$ -component the agreement within the bulk region is good. Apparently, the tails of the distributions in simulation are modelled differently compared to those in data. These events are not used in the analysis.

The dashed lines in Fig. 14.5 indicate the borders of the luminous region. All event outside these limits are not further considered. The agreement between data and simulation within this region is important for the determination of the simulation-based correction factors, which are discussed in the subsequent chapter.

**Energy-to-hit correlation in ECAL and VELO.** The final selection criterion used to separate proton-ion collisions and beam-related background interactions is the energy-to-hit correlation between the electromagnetic calorimeter and the Vertex Locator. Assuming an interaction of two colliding particles with a fixed energy results in a well-defined correlation between the amount of energy deposited in the electromagnetic calorimeter and the number of hits in the tracking detectors.

The observed energy-to-hit correlations for preselected data events and for simulated  $p$ Pb collisions are depicted in Fig. 14.6. The correlation patterns in data and simulation have a common shape. However, additional outliers are found in data, which correspond to events with a large number of tracking hits but with a comparatively low energy deposit. This signature is more likely to originate from lower energetic beam-gas interactions or

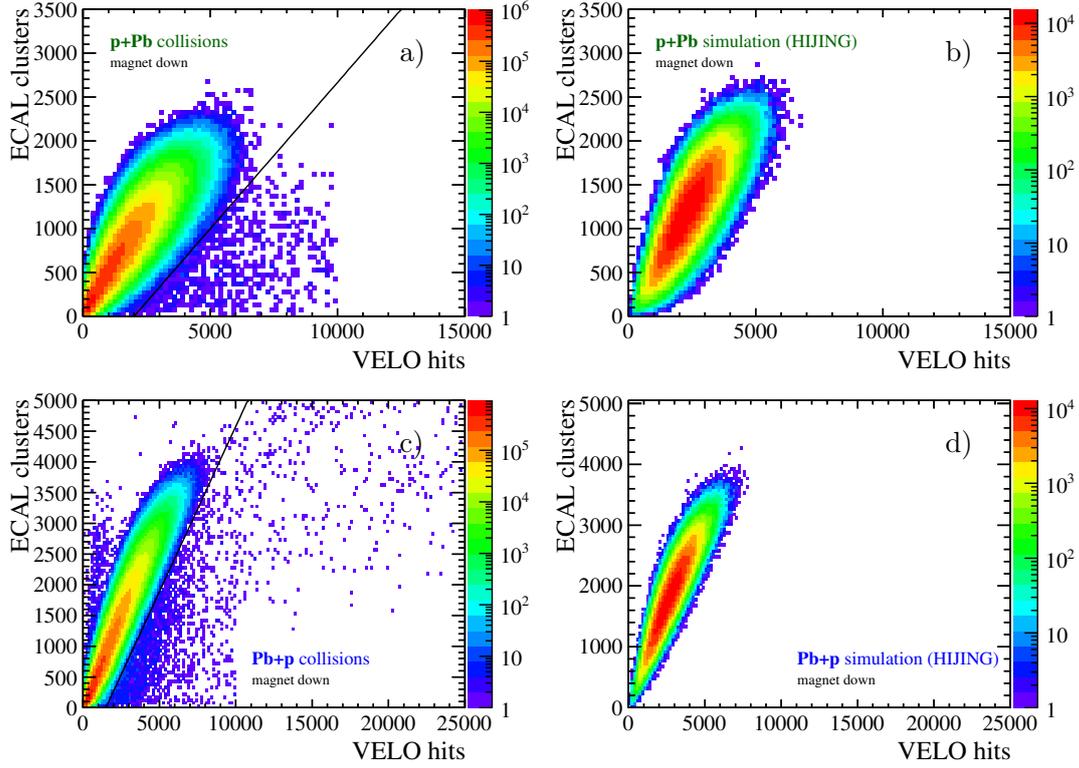


Figure 14.6.: Energy-to-hit correlation between the ECAL and the VELO for data (left) and simulation (right).  $p$ +Pb collisions are shown in the top row, Pb+ $p$  collisions in the bottom row. The data are preselected in terms of bunch-crossing type, PV requirement and luminous region. The simulation only contains single proton-ion collisions, no selection is applied. The lines indicates the borders used to separate background from actual proton-ion collisions. Note logarithmic scale.

interactions of beam particles with support structure (beam-splash) instead of proton-ion collisions. All discussed background interactions can potentially produce a similar amount of particles and detector hits compared to  $p$ Pb collisions, but the total energy of all created particles must be lower. Motivated by the compact correlation pattern that is found in simulation, a linear function is defined in order to reject events with an energy-to-hit ratio that is too small:

$$\mathcal{N}_{ECAL} < a \times \mathcal{N}_{VELO} - b, \quad (14.2)$$

where  $\mathcal{N}_{ECAL}$  and  $\mathcal{N}_{VELO}$  represent the number of clusters in the ECAL and the number of hits in the VELO, respectively. The parameters  $a$  and  $b$  are chosen such that the bulk of the correlation, which represents the region of actual  $p$ +Pb and Pb+ $p$  collisions in the nominal interaction region, remains unaffected, see Fig. 14.6.

### Summary of the event selection

The presented selection criteria used to define the event sample are summarised in bullets:

- Events triggered either by Hlt1MBMicroBiasVelo or Hlt1ActivitySPD
- Events marked with the bunch-crossing flag *beam-beam*
- Requiring the presence of exactly one reconstructed PV per event
- The position of the PV must be within a 3-dimensional luminous region
- Requiring a minimum energy-to-hit ratio

Starting with the multi-pattern structure seen in the track-to-hit correlation in Fig. 14.3, the presented event selection significantly reduces the background in the event sample. The resulting track-to-hit correlation in data, after applying the full selection, is given in Fig. 14.7 a),c) together with the respective correlation found in the simulation samples (b,d), for which no selection is applied. In data, essentially only the correlation structure labelled as 'A' is remaining, this is qualitatively in agreement with the structure obtained from simulation. The single outliers that are still visible in the Pb+p data sample are remaining background events, but their amount is negligible.

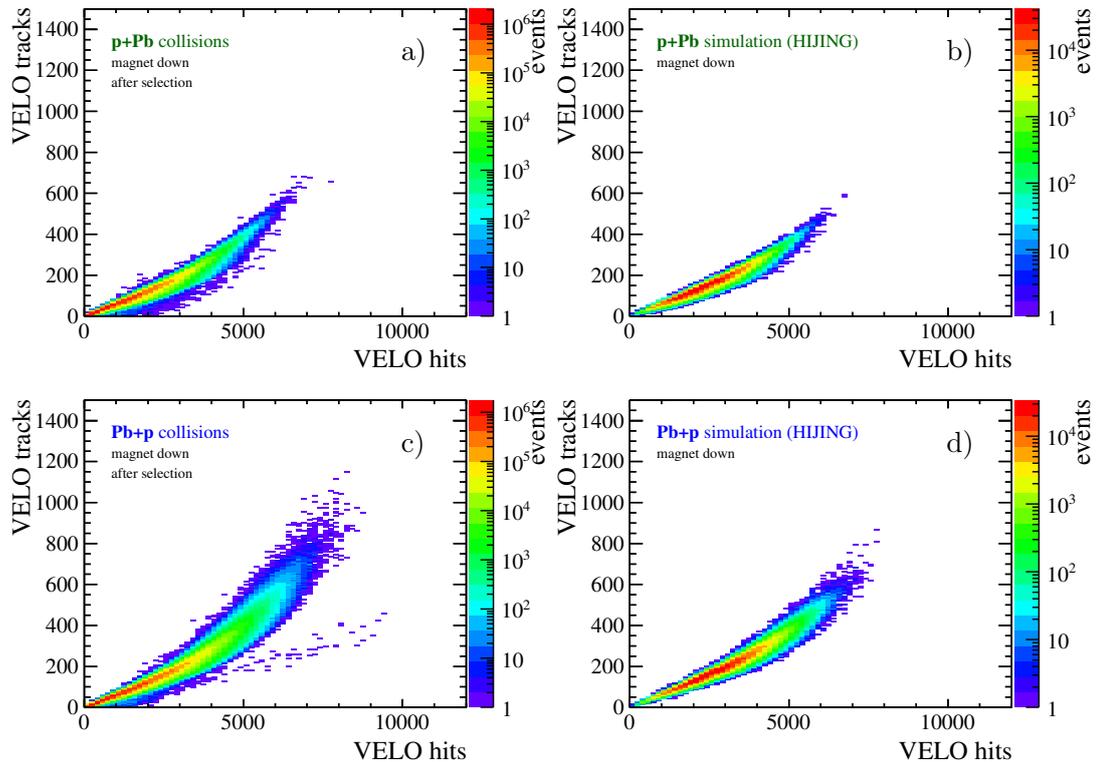


Figure 14.7.: Correlation of VELO-track to VELO-hit multiplicity in data (left) and in simulation (right) after the full event selection. Note the logarithmic scale.

It is important to note that for the selected events, depicted in Fig. 14.3, the VELO-hit multiplicities are all well below 10000. Since the correlation structures do not approach this limit, neither in the  $p$ +Pb nor in the Pb+ $p$  data, the GECs in the trigger do not affect the measurement.

## 14.4. HIJING simulation

This section is dedicated to the proton-ion Monte Carlo simulation which is used in this analysis. The simulation is required in order to determine correction factors, which are used to account for detector and reconstruction related biases to the measurement.

Since the particle density and the event topology of  $p$ Pb collisions is different compared to that of  $pp$  collisions, the standard LHCb  $pp$  simulation cannot be used for this analysis. As an example, the performance of the reconstruction strongly depends on the occupancy of the detector, which is known to be much larger in  $p$ Pb collisions compared to  $pp$  collisions. By increasing the number of pile-up  $pp$  collisions in simulation, an occupancy similar to that of a  $p$ Pb collisions can be achieved. However, the created particles do not originate from the same interaction vertex and also have a different momentum spectrum. In order to obtain reliable correction factors, the use of a dedicated  $p$ Pb simulation is essential.

At the time of this analysis, the LHCb collaboration did not provide or support an official proton-ion simulation. Therefore, a private production is used which is briefly discussed in the following. For the simulation of the physics collisions, the HIJING [174] event generator in version 1.383bs is utilised. It does not only provide the possibility of simulating  $pp$  or nucleus-nucleus collisions but also mixed systems, such as protons and lead-ions ( $p$ Pb). Two beam configurations, namely  $p$ +Pb and Pb+ $p$ , are realised in the simulation, both with a nucleon-nucleon centre-of-mass energy set to  $\sqrt{s_{NN}} = 5.02$  TeV. The nominal interaction point and the one-sigma deviation in the  $x$ -,  $y$ -, and  $z$ -component are defined according to Tab. 14.3. Depending on the polarity of the LHCb dipole, the interaction region is different. The values are determined from reconstructed primary vertex positions in data, but using only smaller sub-samples. As a result, the simulated vertex positions do not fully match the broader distributions for the large data samples.

The orientation of the magnetic field in the detector also has a small impact on the beam line and thus affects the crossing angles of the colliding beams. In the simulation, the internal LHCb crossing angles has been set to  $-0.236$  mrad at magnet-down polarity and  $+0.236$  mrad at magnet-up polarity, according to Ref. [175].

The HIJING simulation is adjusted such that exactly one proton-ion collision is taking place in each event and no pile-up interactions are included. Further, there are no background interactions, such as beam-gas, included. All simulated events correspond to clean proton-ion collisions with beam parameters taken from real data.

This custom HIJING simulation phase has been implemented into the LHCb software framework. All subsequent reconstruction steps are identical to official Monte Carlo simulation productions<sup>2</sup>. The statistics of the simulation samples is listed in Tab. 14.4.

---

<sup>2</sup>The used software versions for the private HIJING MC production are: GAUSS v45r7 (simulation), BOOLE v26r3 (digitalization), MOORE v20r4 (L0-trigger), MOORE v14r7 (HLT-trigger), BRUNEL v44r8p1 (reconstruction) and DAVINCI v33r9 (analysis).

| Beam configuration | Magnetic field | Number of events   |
|--------------------|----------------|--------------------|
| $p+\text{Pb}$      | down           | $2.02 \times 10^6$ |
| $p+\text{Pb}$      | up             | $1.76 \times 10^6$ |
| $\text{Pb}+p$      | down           | $1.99 \times 10^6$ |
| $\text{Pb}+p$      | up             | $2.02 \times 10^6$ |

Table 14.4.: Overview of the generated simulation sample using HIJING. The nucleon-nucleon centre-of-mass energy is  $\sqrt{s_{NN}} = 5.02$  TeV in all samples.

It should be emphasised that no physics results are concluded from these simulation samples. The purpose of the simulation is to determine the detector acceptance and to quantify the amount of fake and duplicate tracks in the high-multiplicity environment of proton-ion collisions. For this purpose, the relevant detector simulation and the track reconstruction are adopted from the official software framework and remain unchanged.

The properties of the simulation samples are not separately discussed at this point. A comparison of basic track properties in data and simulation is given at the end of this chapter. The determination of correction factors extracted from the simulation is explained in detail in Chap. 15.1.

## 14.5. Track selection

This section describes the track selection and further discusses the resulting track samples in data and simulation. The selection is applied to reconstructed tracks from events passing the previous event selection. There is no difference in treatment for simulated and recorded data, both are processed and reconstructed with identical software versions and settings.

In this analysis, inclusive track samples are used, which contain charged (quasi-) stable particles that are directly produced in the primary interaction. No separation for different particle species is done<sup>3</sup>. In order to prove or disprove the existence of the near-side ridge for particles in an intermediate  $p_T$ -range (around  $1 < p_T < 3$  GeV), momentum information is needed in this analysis. Similar to the analysis on charged particle production in Part I of this thesis, again only *long tracks* are considered. In contrast to the previous analysis, no modification to track reconstruction is done, both long track algorithms are used, as described in Chap. 5.3.

### 14.5.1. Definition of the kinematic range

For this measurement it is important to cover a large range in pseudorapidity. The relevant two-particle correlation features only become visible when probing particles with a large separation in pseudorapidity,  $\Delta\eta$ . Hence, the kinematic range that is accessible with long tracks is exploited to a maximum. In pseudorapidity, this is equivalent to a range of  $2.0 < \eta < 4.9$ . Also the lower boundaries for the total and transverse momentum are selected to be as low as possible w.r.t. the acceptance of the detector, which are

<sup>3</sup>All charged particles are reconstructed under the hypothesis of a pion. The used `StandardParticle` type in the reconstruction is called `StdAllNoPIDsPions`.

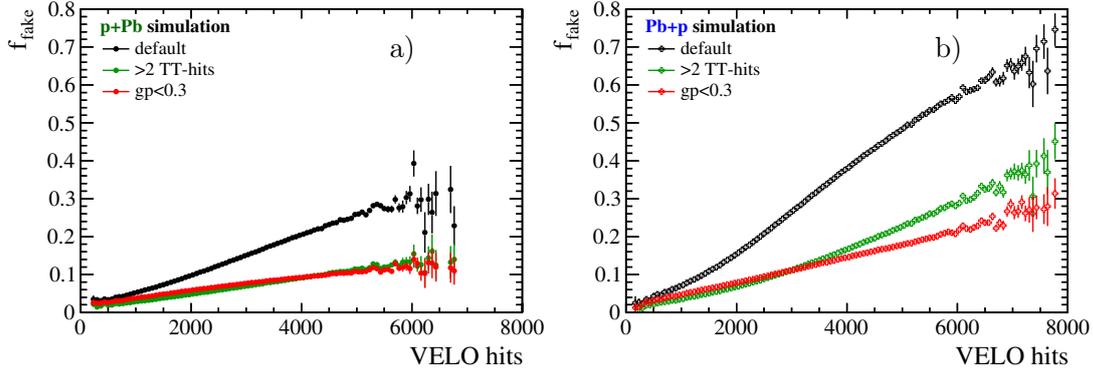


Figure 14.8.: Fake (ghost) track fraction determined in simulation for the  $p+Pb$  and  $Pb+p$  beam configuration.

$p > 2.0$  GeV and  $p_T > 150$  MeV, respectively. The transverse momentum is further limited to a maximum value of up to 5.0 GeV. Tracks with a higher  $p_T$  are excluded, since only the low and medium  $p_T$ -range is of interest for this measurement, but also the number of high- $p_T$  tracks is too small to obtain significant results.

Compared to the kinematic range used in the multiplicity analysis the  $\eta$ -range and low- $p_T$  requirements are less stringent. The selected pseudorapidity range represents the maximum detector acceptance which allows charged particles to be reconstructed with momentum information. With the improvements of the reconstruction software this larger kinematic range can be used without suffering from too large uncertainties towards the borders of the detector acceptance. A quantitative discussion of the reconstruction artefacts is given in the subsequent chapter.

### 14.5.2. Suppression of reconstruction artefacts

Following the conclusion of a strongly rising contamination of fake tracks with detector occupancy (c.f. Chap. 9.2.1 in Part I), a huge amount of these mis-reconstructed tracks can be expected. The high occupancies of the tracking detectors that are reached in  $p+Pb$  and even more in  $Pb+p$  collisions, requires to consider an active suppression of fake tracks in the reconstruction.

To simplify the discussion, the fraction of fake-tracks in Fig. 14.8 is only presented as a function of hits in VELO. Both simulated samples,  $p+Pb$  (a) and  $Pb+p$  (b), show the expected strong increase of fake tracks in the track sample. With the default track reconstruction (black markers) the fraction of fake tracks in  $p+Pb$  rises up to a large but still moderate level of approximately 30% at 6000 VELO hits. In contrast to this, the  $Pb+p$  simulation shows a percentage of fake tracks that is twice as large as the one found in the  $p+Pb$  simulation at the same occupancy, and finds an even larger maximum fraction of more than 70% at 8000 VELO hits. The different occupancy behaviour in each sample can be explained by the higher local track densities that are reached in  $Pb+p$  collisions. Furthermore, differences in the momentum spectra of the created particles contribute to observed different performance. This observation also demonstrates the need of a dedicated  $pPb$  simulation for this analysis, since this particular fake-track

behaviour cannot be reproduced with the  $pp$  simulation. The origins of fake tracks are discussed in more detail in Chap. 15.1 when explaining the correction method that is applied to the track sample.

With more than half of the tracks in the sample being mis-reconstructed, a significant measurement is almost impossible. Additional quality requirements are necessary in order to increase the purity of the track sample. The default track reconstruction already requires the quality of the track fit to be better than  $\chi_{\text{track}}^2/\text{nDoF} < 3$ . This is a standard requirement for all LHCb analyses and already reflects a high track quality. A further tightening of this criterion was tested to be inefficient, because a large fraction of good tracks is discarded as well.

By design, the amount of material in the tracking system of LHCb is small, which allows low-momentum particles to be reconstructed with a high efficiency. As described in Chap. 5.4, the resulting small number of measurements related to a particle favours the presence of mis-reconstructed tracks. Long tracks are mainly affected by combining wrong track segments of the VELO and the T-stations. Additional information of the Tracker Turicensis (TT) located in front of the dipole magnet, can reduce this combinatoric effect, but is not considered in the default reconstruction. Requiring a long track to have at least one assigned TT-hit in each of the detection layers drastically reduces the amount of fake tracks, as shown by the green distributions in Fig. 14.8. However, the slightly smaller acceptance of the TT restricts the accessible pseudorapidity range to  $\eta \lesssim 4.7$ . Thus, the explicit TT-requirement is not applied for this measurement.

A multivariate classifier combines information of the TT and all other tracking detectors to the so-called *ghost probability* ( $GP$ ). The separation power to distinguish between reconstruction artefacts and good tracks is additionally obtained by adding information of kinematic variables, event properties and parameters that are accessed during the reconstruction sequence [106]. The ghost probability is available for all reconstructed long tracks. Requiring the  $GP$  to be smaller than 0.3, see red distributions in Fig. 14.8, suppresses fake tracks to a maximum fraction of 10% in  $p+\text{Pb}$  and 30% in  $\text{Pb}+p$  collisions.

Another kind of reconstruction artefact is due to duplicate tracks. While in the analysis of charged particle multiplicities the contamination of duplicate tracks is significant, the default track reconstruction used in this analysis already includes an active suppression for these kinds of mis-reconstructed tracks. This is achieved by using the so-called *Kullback-Leibler* (KL) *clone distance* [176]. It is a measure of the shared information between two tracks. The larger the KL distance of two tracks is the less likely they are duplicate tracks. The default cut value<sup>4</sup> rejects duplicate tracks to a fraction of less than 1‰ in the track sample. Their contribution is negligible for the further analysis.

### 14.5.3. Selection of prompt particles

In order to reject reconstructed tracks from non-prompt particles, a limit for the impact parameter between the tracks and the primary vertex ( $\text{IP}_{\text{PV}}$ ) is defined. Figure 14.9 shows the impact parameter distributions of reconstructed prompt and secondary particles in the  $p+\text{Pb}$  and  $\text{Pb}+p$  simulation, where each distribution is normalised to the respective total number of reconstructed tracks. Around an impact parameter

<sup>4</sup>In the used reconstruction software, BRUNEL v44r8p1, the  $KL$  distance must be larger than 5000.

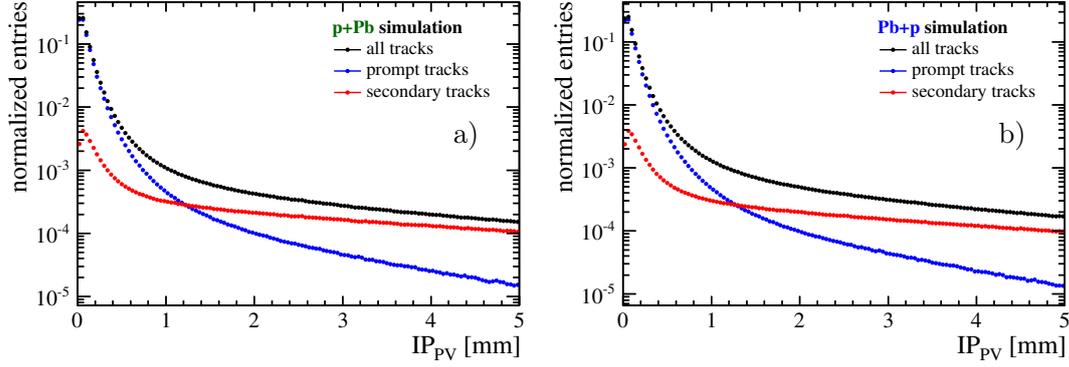


Figure 14.9.: Impact parameter distribution of reconstructed tracks determined in simulated  $p+\text{Pb}$  (a) and  $\text{Pb}+p$  (b) collisions. Each distribution is normalised to the total number of reconstructed tracks.

of 1.20 – 1.25 mm the prompt and secondary particle distributions intercept and the fraction of non-prompt particles dominates for larger distances. In order to use a common selection for the  $p+\text{Pb}$  and  $\text{Pb}+p$  data samples, the maximum impact parameter is chosen to be  $\text{IP}_{\text{PV}} = 1.2$  mm for both samples. This corresponds to a purity of 96.48% and 96.51% in the  $p+\text{Pb}$  and  $\text{Pb}+p$  configuration, respectively.

## 14.6. Selection summary and overview of data samples

All discussed selection criteria are again listed in Tab. 14.5. The selection is identically applied to the data and simulation samples in both beam configurations ( $p+\text{Pb}$  and  $\text{Pb}+p$ ) and both magnetic field orientations (up and down). In order to check for potential difference in the data samples an overview of the most important kinematic quantities and angular distributions of the selected tracks is given in Fig. 14.10. The data are compared to simulation, separately for the  $p+\text{Pb}$  and  $\text{Pb}+p$  data samples. The important conclusions are:

- Comparison of different magnetic field polarities:  
The  $\eta$  and  $\phi$  distributions indicate only very small differences since the detector's left- and right-hand side are basically symmetric. No relevant detector parts were broken or offline during the data taking, which would be reflected in both angular distributions. The measured momentum distributions,  $p$  and  $p_{\text{T}}$ , are identical for both field configurations as the absolute power of the magnet remains unchanged.
- Comparison of data and simulation:  
The general shape of the angular distributions are well reproduced in the simulation. The position of distinctive spots of inefficiencies are in satisfactory agreement. This is crucial for the later correction of detector effects in the track sample. The  $p$ - and  $p_{\text{T}}$ -distributions show that the HIJING simulation slightly underestimates the amount of higher (transverse) momentum particles, while the low momentum

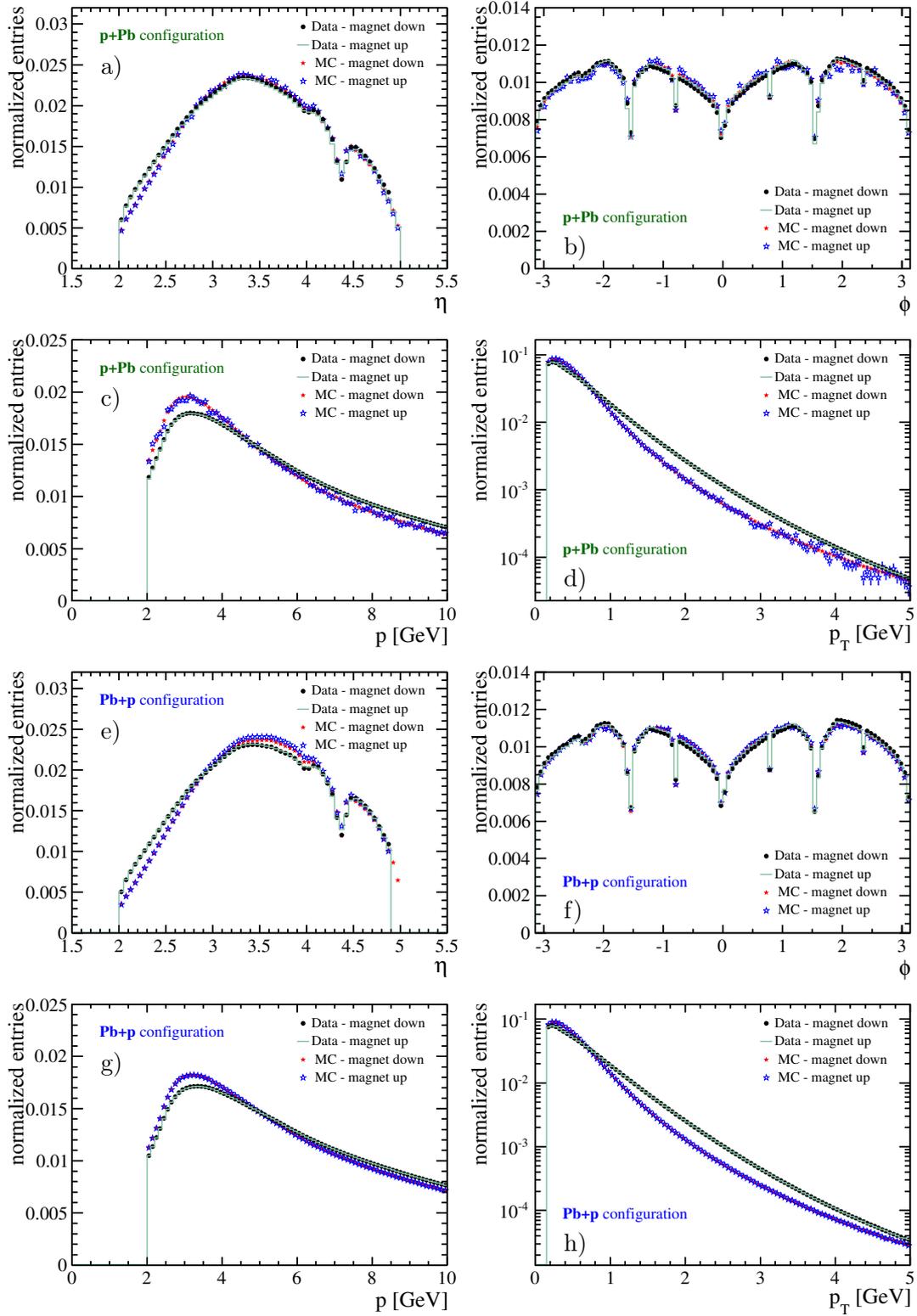


Figure 14.10.: Comparison of track properties in data and simulation, for both magnetic field configurations. The distributions are presented separately for the  $p+Pb$  (a-d) and  $Pb+p$  (e-h) data samples. Results are discussed within the text.

region is overestimated. This tendency is present in both beam configurations ( $p+\text{Pb}$  and  $\text{Pb}+p$ ) but not of importance for this analysis, since the simulation is only used to determine detector related correction factors.

- Comparison of  $p+\text{Pb}$  and  $\text{Pb}+p$  data:

During the  $p+\text{Pb}$  and  $\text{Pb}+p$  data taking the detector was in an almost identical condition. Hence, the positions of distinctive structures in the  $\eta$ - and  $\phi$ -distributions of both data sets are nearly the same. The origin of the dips visible in the  $\phi$ -distributions is related to the geometry of the VELO sensors, the decline which is observed at  $\eta \approx 4.35$  is due to the shape of the beam pipe. These effects are explained in more detail in the next chapter. A closer look at the pseudorapidity distributions shows that the peak is located at slightly different positions in the  $p+\text{Pb}$  and  $\text{Pb}+p$  samples and that the entire distribution is shifted. This is the result of the asymmetric beam configuration which causes a boost of the collision system into the direction of the proton. As the basic shape of the measured distribution is defined by the detector acceptance the differences are only barely visible. The  $p$ - and  $p_{\text{T}}$ -distributions are also slightly different which is related to the probed different rapidity ranges in the centre-of-mass system of the collision.

The resulting track samples in data and simulation represent the basis for the following measurement of two-particle correlations. Remaining fake tracks and mis-selected secondary tracks as well as inefficiencies related to the reconstruction are quantified in the subsequent chapter, in which also the analysis method itself is described.

|                         | Selection Parameter               | Cut value                     |
|-------------------------|-----------------------------------|-------------------------------|
| quality cuts (default)  | track fit $\chi^2/nDoF$           | $< 3$                         |
|                         | KL clone distance                 | $> 5000$                      |
| quality cuts (explicit) | ghost probability $GP$            | $< 0.3$                       |
| kinematic selection     | pseudorapidity $\eta$             | $\in [2.0, 4.9]$              |
|                         | momentum $p$                      | $> 2.0 \text{ GeV}$           |
|                         | transv. momentum $p_{\text{T}}$   | $\in [0.15, 5.0] \text{ GeV}$ |
|                         | impact parameter $IP_{\text{PV}}$ | $< 1.2 \text{ mm}$            |

Table 14.5.: Common track selection for reconstructed tracks in the  $p+\text{Pb}$  and  $\text{Pb}+p$  data samples. The first two selection criteria are already applied for standard LHCb tracks, the subsequent selection is explicitly applied for this analysis.

---

## Measurement of angular correlations in proton-ion collisions

---

After the selection of  $p$ Pb events and the definition of the track samples, this chapter is dedicated to the analysis method which is applied for the measurement of two-particle angular correlations. This chapter is arranged in four sections, starting with the quantification and correction of contaminations in the selected track sample and different inefficiencies induced by the reconstruction. Afterwards, a discussion about the activity within a proton-ion collision is given and different approaches of how to classify this activity are presented. In the next section, the two-dimensional correlation function is defined, which represents the central ingredient of this analysis. This is followed by a section about the one-dimensional representation of correlation properties. Afterwards a brief look at two-particle correlation properties that are implemented in the HIJING simulation is given. In the last part, a cross-check is presented which shows that no artificial angular correlation remain after applying the correction procedure chosen for this analysis.

### 15.1. Data correction for reconstruction artefacts and mis-selection

The main challenges to obtain an unbiased track sample, free of detector effects, are discussed in Chap. 5. The contamination of fake tracks and the inefficiencies related to the reconstruction and the detector acceptance have already been determined for the analysis on charged particle multiplicities (see part I). However, proton-ion collisions have a different event topology compared to  $pp$  collisions. The general detector occupancy can be similar to  $pp$  events with moderate or large pile-up, but the local track density is always different, since all prompt particles originate from a single common vertex. Furthermore, the official event reconstruction of the proton-ion raw data uses more advanced tracking algorithms, including a set of new default quality requirements, as discussed in the previous section. As a consequence, an independent re-evaluation of the correction factors is required.

The strategy of the following corrections is to quantify the single contaminations and inefficiencies and then assign per-track weighting factors that account for these effects.

The reconstructed and selected tracks only include two significant categories of impurities, which are fake tracks and non-prompt particles. The fraction of duplicate tracks in the track sample is less than 0.1‰ and thus negligible. In addition, inefficiencies induced by the detector acceptance and the track reconstruction must be accounted for. In the following, each of these effects is discussed and quantified, before the correction procedure is explained.

It should be pointed out that the main goal of the correction procedure is to remove any detector induced structures in the angular distributions,  $\eta$  and  $\phi$ . These can introduce artificial patterns to the correlation function and thus bias the result of the actual physical correlation. Since the correlation measurement is intended to be performed in different classes of event activity and for particles in different  $p_T$ -ranges, the correction factors have to further consider global event properties and particle kinematics.

All presented results rely on the customised HIJING Monte Carlo simulation, c.f. Chap. 14.4. The respective simulation samples for different beam configurations,  $p+\text{Pb}$  and  $\text{Pb}+p$ , and magnetic field polarities are studied separately. The overall event topology in simulation is not of particular importance, since all corrections are determined differentially in bins of relevant track or event quantities. As a result, no large event-generator dependence is expected.

### 15.1.1. Fake-track contamination

Since the amount of duplicate tracks is negligible the only significant contamination in the track sample due to mis-reconstruction are fake tracks. Even after applying the strict requirement on the ghost probability in the selection, the fraction of fake tracks w.r.t selected tracks, the *fake-track fraction*  $f_{\text{fake}}$ , reaches up to 30% in high multiplicity events and requires a dedicated correction.

Besides the rising combinatorial probability to create a fake track in events with more detector hits, also local properties of the tracking detectors are important. Figure 15.1 shows the fake-track fraction determined in simulation as functions of the azimuthal angle, the pseudorapidity, the transverse momentum and the number of detector hits in the VELO,  $\mathcal{N}_{\text{VELO}}^{\text{hit}}$ . The distributions for  $\text{Pb}+p$  compared to  $p+\text{Pb}$  collisions are shifted towards larger fake-track fractions, since the events have a different average detector occupancy.

The distributions in  $\eta$  show the expected structure related to the shape of the beam pipe, c.f. Chap. 4.2. Around a value of  $\eta = 4.35$  a peak of the fake-track fraction is visible, which corresponds to a polar angle with which particles must traverse a greater amount of material. An interconnection between two parts of the beam pipe is responsible for this additional material. The rise of  $f_{\text{fake}}$  towards large values of pseudorapidity is related to the missing acceptance of the TT sub-detector. As there is no TT-information available for tracks with  $\eta \gtrsim 4.7$  the ghost probability is less powerful in separating good and fake tracks.

Studying fake tracks as a function of the azimuthal angle  $\phi$  shows a moderate rise towards  $\phi = \pm 90^\circ$ . This is related to the increasing amount of material due to the

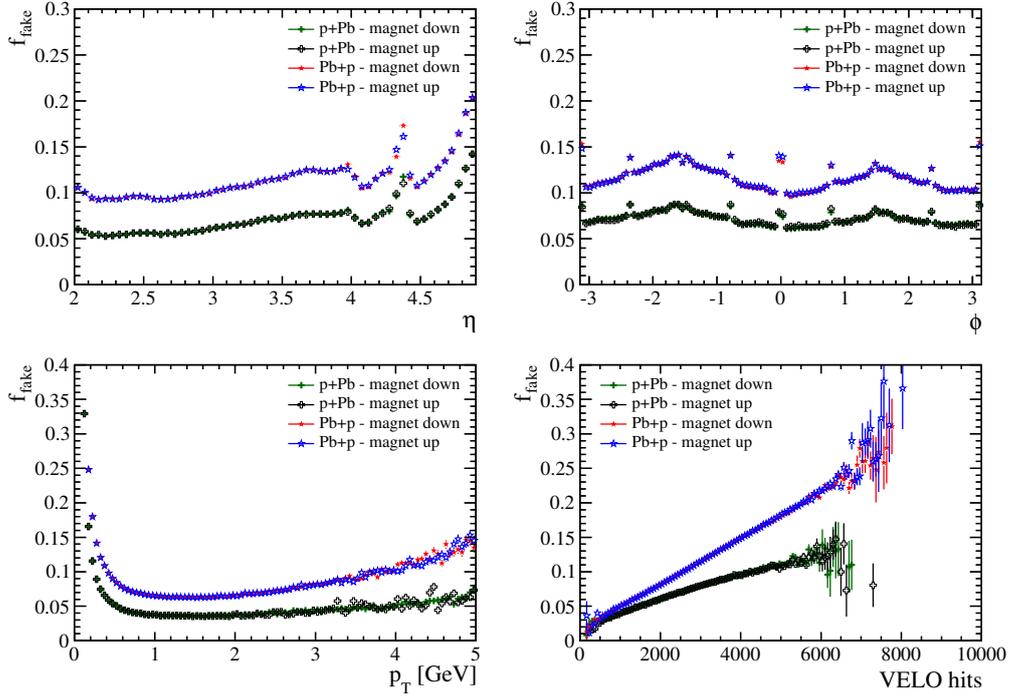


Figure 15.1.: The fraction of fake tracks (*ghost tracks*) studied in one dimension as functions of  $\eta$ ,  $\phi$ ,  $p_T$  and VELO-hit multiplicity in simulation. Results are presented for  $p$ +Pb and Pb+ $p$  collisions in magnet-up and -down configurations. Pb+ $p$  events have a higher average detector occupancy compared to  $p$ +Pb events.

overlap of the VELO discs and the RF-foil that separates the different vacua between the beam pipe and the VELO sensors, c.f. VELO design in Chap. 4.1.2. Since the VELO R-sensors are arranged in  $45^\circ$  segments also a repeating small peak structure is found in the fake-track fraction.

The  $p_T$ -dependence shows a slight increase of  $f_{\text{fake}}$  towards larger transverse momenta. The probability to reconstruct a fake track due to combinatorics remains the same, also in the higher  $p_T$ -range. However, the number of particles decreases, which results in an effectively larger fake-track fraction. Towards low  $p$  and  $p_T$ , the effect of multiple scattering drastically increases and leads to a steep rise of mis-reconstructed tracks.

An almost linear increase of the fake-track fraction with the hit multiplicity in the VELO is observed. This rise accounts for the higher combinatorial probability to form random tracks from unrelated detector hits.

As discussed in the previous analysis, the probability to create fake tracks depends on various detector parameters. In this analysis the fake track correction has the focus of providing an optimal correction of the angular distributions. In order to account for fine structures in the  $\eta$ - and  $\phi$ -distribution a fine parametrisation scheme in these quantities is necessary. In addition, also the  $p_T$ -dependence needs to be considered. Since only the lower  $p_T$ -range shows a strong variation, a non-equidistant parametrisation with a small number of bins is sufficient.

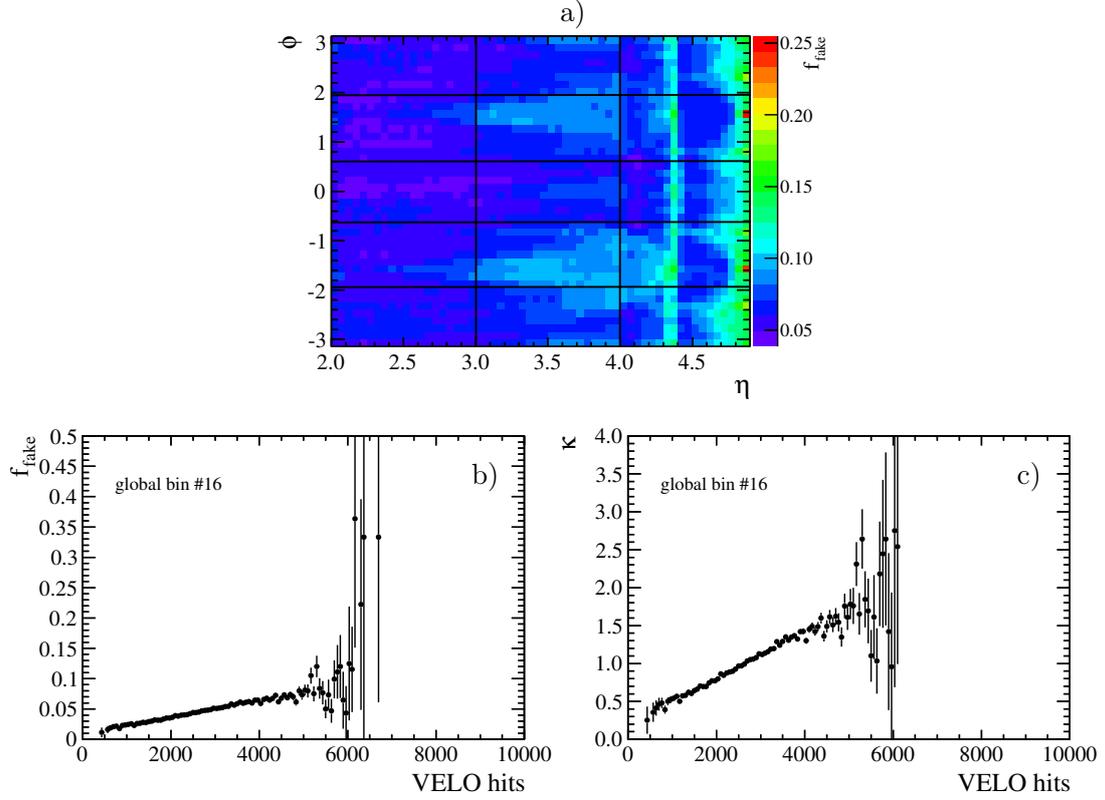


Figure 15.2.: Visualisation of (3+1)-dimensional parametrisation of the fake-track fraction. The top histogram shows the  $(\eta, \phi)$ -projection of  $f_{\text{fake}}$  in the fine binning scheme, overlaid with the borders of the simple binning scheme. The bottom left histogram (b) shows  $f_{\text{fake}}$  as a function of the VELO-hit multiplicity in an exemplary  $(\eta, \phi, p_{\text{T}})$ -bin. After accounting for the average hit-multiplicity of the simulation sample the binned scaling factor  $\kappa$  is obtained (c).

|   | Variable                                 | Number of bins | Bin range                               |
|---|--|----------------|---|
| $\mathcal{P}_{\text{fake}}^{3\text{D}}$ | $\eta$                                   | 60             | $[2.0, 5.0], \Delta\eta = 0.05$         |
|   | $\phi$                                   | 42             | $[-\pi, +\pi], \Delta\phi = 2\pi/42$    |
|   | $p_{\text{T}}$                           | 5              | $[0.15, 0.4, 1.0, 2.0, 3.0, 5.0]$ GeV   |
| $\kappa_{\text{fake}}^{3+1\text{D}}$    | $\eta$                                   | 3              | $[2.0, 3.0, 4.0, 4.9]$                  |
|   | $\phi$                                   | 5              | $[-\pi, -1.89, -0.63, 0.63, 1.89, \pi]$ |
|   | $p_{\text{T}}$                           | 5              | $[0.15, 0.4, 1.0, 2.0, 3.0, 4.9]$ GeV   |
|   | $\mathcal{N}_{\text{VELO}}^{\text{hit}}$ | 150            | $[0, 10000]$                            |

Table 15.1.: Binning scheme used for the determination of fake tracks. The first block corresponds to the fine binning scheme optimised for the angular variables. The second block shows the simple binning scheme with an additional fine parametrisation of the hit-multiplicity in the VELO.

Ideally, all three mentioned variables and the global hit-multiplicity dependence should be used in a 4-dimensional parametrisation of the fake-track probability. Starting with a proper parametrisation of the kinematic variables, c.f. Tab. 15.1, already results in a total number of 3150 bins. The limited sizes of the simulation samples do not permit the use of fourth parametrisation variable. Consequently, the  $\eta$ ,  $\phi$  and  $p_T$  dependences are first combined to a fine 3-dimensional parametrisation of the probability to create fake-tracks,  $\mathcal{P}_{\text{fake}}^{3\text{D}}(\eta, \phi, p_T)$ , while the hit-multiplicity dependence is treated separately.

An almost linear dependence of the fake-track fraction with the hit-multiplicity in the VELO is observed. However, this slope also depends slightly on  $\eta$ ,  $\phi$  and  $p_T$  and thus cannot be factorised. A mixed (3+1)-dimensional parametrisation is found to be feasible and is implemented as follows: Instead of the aforementioned fine binning scheme, the fake-track probability is again determined using a simple binning scheme of only 75 bins for the three variables  $\eta$ ,  $\phi$  and  $p_T$ . The small number of bins allows using an additional fine parametrisation as a function of the hit-multiplicity in the VELO, leading to  $\mathcal{P}_{\text{fake}}^{3+1\text{D}}(\eta, \phi, p_T; \mathcal{N}_{\text{VELO}}^{\text{hit}})$ .

To benefit from both parametrisation approaches,  $\mathcal{P}_{\text{fake}}^{3+1\text{D}}(\eta, \phi, p_T; \mathcal{N}_{\text{VELO}}^{\text{hit}})$  is also determined for the average number of VELO-hits (per event) in the simulation sample,  $\overline{\mathcal{N}_{\text{VELO}}^{\text{hit}}}$ . From this, a scaling factor  $\kappa_{\text{fake}}^{3+1\text{D}}(\eta, \phi, p_T; \mathcal{N}_{\text{VELO}}^{\text{hit}})$  is extracted, with which the hit-multiplicity dependence of the fake-track probability can be modelled:

$$\kappa_{\text{fake}}^{3+1\text{D}}(\eta, \phi, p_T; \mathcal{N}_{\text{VELO}}^{\text{hit}}) = \mathcal{P}_{\text{fake}}^{3+1\text{D}}(\eta, \phi, p_T; \mathcal{N}_{\text{VELO}}^{\text{hit}}) \cdot 1 / \mathcal{P}_{\text{fake}}^{3+1\text{D}}(\eta, \phi, p_T; \overline{\mathcal{N}_{\text{VELO}}^{\text{hit}}}). \quad (15.1)$$

It is checked that this scaling factor, determined in the simple  $(\eta, \phi, p_T)$  binning, factorises sufficiently well with the fine parametrisation  $\mathcal{P}_{\text{fake}}^{3\text{D}}(\eta, \phi, p_T)$  and is therefore used as binned scaling factor:

$$\mathcal{P}_{\text{fake}}(\eta, \phi, p_T, \mathcal{N}_{\text{VELO}}^{\text{hit}}) = \mathcal{P}_{\text{fake}}^{3\text{D}}(\eta, \phi, p_T) \cdot \kappa_{\text{fake}}^{3+1\text{D}}(\eta, \phi, p_T; \mathcal{N}_{\text{VELO}}^{\text{hit}}). \quad (15.2)$$

With this approach, the hit-multiplicity dependence is combined with the fine parametrisation in  $(\eta, \phi, p_T)$ , where the latter is important for a precise correction of the angular observable used in the correlation measurement.

For visualisation, Fig. 15.2 a) shows the fine  $(\eta, \phi)$ -binning of  $\mathcal{P}_{\text{fake}}$ , overlaid with the corresponding simple binning scheme. For the sake of convenience, the  $p_T$ -dependence is ignored in this example. The hit-multiplicity dependence within an exemplary large  $(\eta, \phi)$ -bin is depicted in Fig. 15.2 b). After taking into account the mean hit-multiplicity of the simulation sample, the corresponding scaling factor  $\kappa$  is obtained, as given in Fig. 15.2 c). Any of the fine  $(\eta, \phi)$ -bins that are comprised within one of the large  $(\eta, \phi)$ -bin can be then scaled by the corresponding factor  $\kappa$ , which describes the hit-multiplicity dependence.

### 15.1.2. Non-prompt contribution

Another kind of contamination in the track sample are mis-selected non-prompt tracks. The average fraction of secondary particles,  $f_{\text{sec}}$ , in both simulation samples,  $p+\text{Pb}$  and  $\text{Pb}+p$ , amounts to approximately 3.5%.

In analogy to fake tracks, the probability to find a non-prompt track is studied as functions of the same quantities  $(\eta, \phi, p_T)$  and  $\mathcal{N}_{\text{VELO}}^{\text{hit}}$ , as illustrated in Fig.15.3. The

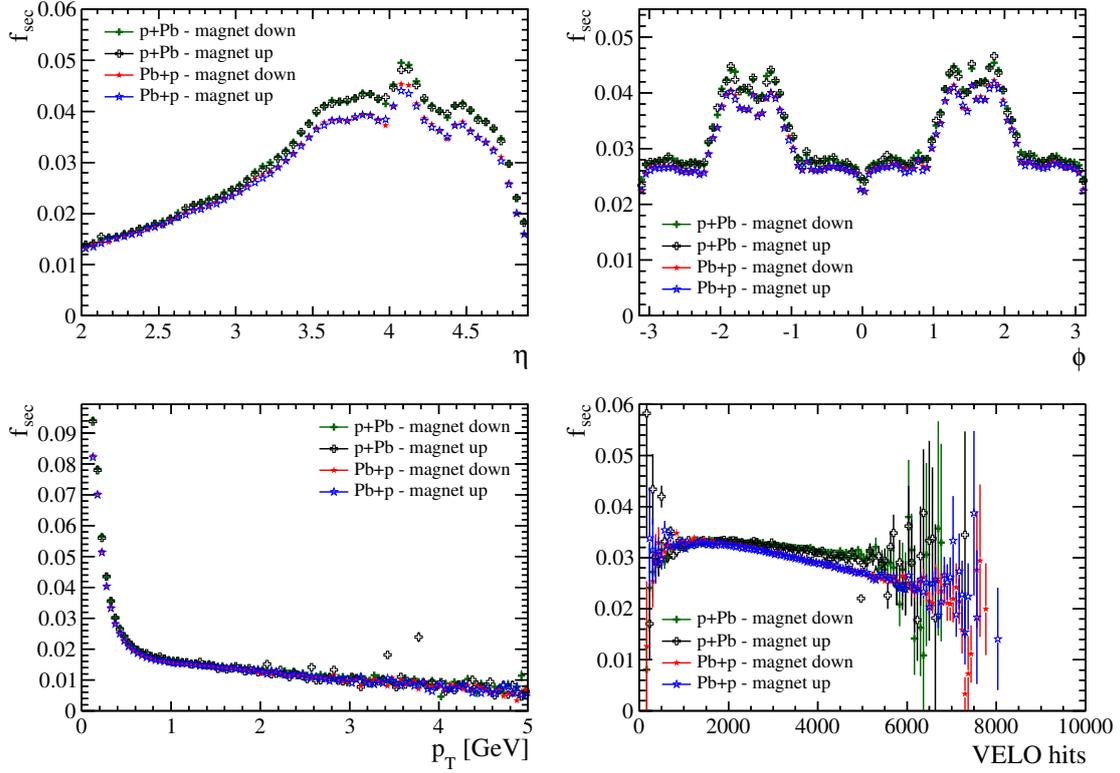


Figure 15.3.: The fraction of reconstructed non-prompt particles (*secondaries*) studied in one-dimension as functions of  $\eta$ ,  $\phi$ ,  $p_T$  and VELO-hit multiplicity in simulation. Results are presented for  $p+Pb$  and  $Pb+p$  collisions in magnet-up -down configurations.

different patterns that are visible in the  $\eta$  and  $\phi$  parametrised distributions are primarily related to the material that is seen by the particles when traversing the detector. The more material is traversed, the larger is the probability to interact with it and to produce secondary particles. The increase with pseudorapidity reflects the rising number of VELO sensors that are passed through. Towards larger  $\eta$ , again the material of the beam pipe is responsible for the visible complex structure. For particles with  $\eta \gtrsim 4.5$  again only few VELO sensors and less material are traversed.

The  $\phi$ -dependence shows the additional material related to overlapping two half-disc sensors and the RF-foil within the Vertex Locator, which are present in the range

|                                 | Variable                                 | Number of bins | Bin range                              |
|---------------------------------|--|----------------|--|
| $\mathcal{P}_{\text{sec}}^{4D}$ | $\eta$                                   | 60             | $[2.0, 5.0]$ , $\Delta\eta = 0.05$     |
|                                 | $\phi$                                   | 42             | $[-\pi, \pi]$ , $\Delta\phi = 2\pi/42$ |
|                                 | $p_T$                                    | 3              | $[0.15, 0.3, 0.6, 5.0]$ GeV            |
|                                 | $\mathcal{N}_{\text{VELO}}^{\text{hit}}$ | 3              | $[0, 1800, 2600, 10000]$               |

Table 15.2.: 4D binning scheme used to determine the probability to select non-prompt tracks.

of around  $\phi = |60 - 120|^\circ$ . As for fake tracks, there is a larger probability to find non-prompt tracks in the low- $p_T$  region, since low-momentum particles are more affected by material interaction. The decreasing fraction of non-prompt tracks towards high detector occupancies accounts for the in total larger number of particles in these events.

The comparatively small fraction of selected non-prompt tracks and the only slight variation with the number of hits in the detector permits the use of a simplified binning scheme, compared to the fake-track parametrisation. A moderate number of bins in  $p_T$  and  $\mathcal{N}_{\text{VELO}}^{\text{hit}}$ , combined with a fine parametrisation in  $\eta$  and  $\phi$ , results in a regular 4-dimensional probability,  $\mathcal{P}_{\text{sec}}(\eta, \phi, p_T, \mathcal{N}_{\text{VELO}}^{\text{hit}})$ , to find mis-selected non-prompt tracks. The detailed binning scheme is listed in Tab. 15.2.

### 15.1.3. Detector acceptance

Another important aspect which must be accounted for is limited efficiency related to the acceptance of the detector. The detector acceptance efficiency,  $\epsilon_{\text{acc}}$ , is defined in analogy to the multiplicity analysis, c.f. Chap. 9.4.1. For particles fulfilling the kinematic requirements of the track selection,  $\epsilon_{\text{acc}}$  describes the fraction that reach the end of the main tracking stations. Particles that do not reach the last T-station either interact with material or are deflected out of the detector by the magnetic field. This condition

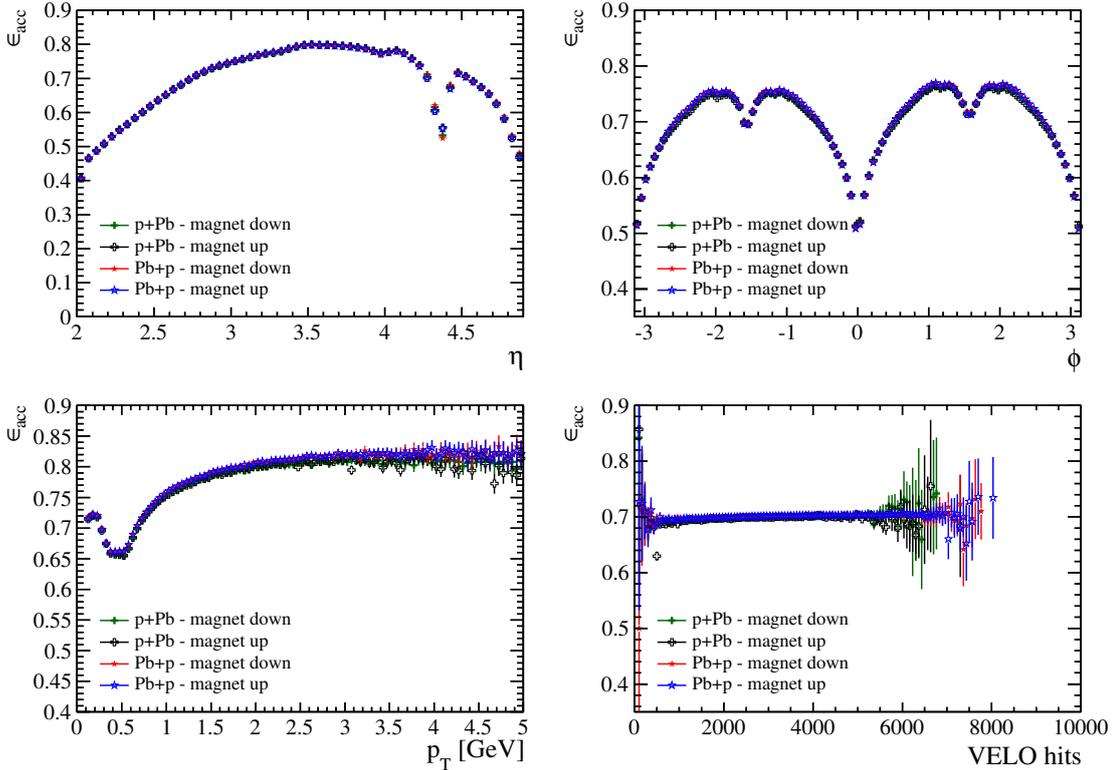


Figure 15.4.: Acceptance efficiency studied in one dimension as functions of  $\eta$ ,  $\phi$ ,  $p_T$  and VELO-hit multiplicity in simulation. Results are presented for  $p$ +Pb and Pb+ $p$  collisions in magnet-up and -down configurations.

|                                     | Variable                                 | Number of bins | Bin range                           |
|-------------------------------------|--|----------------|-------------------------------------|
| $\epsilon_{\text{acc}}^{\text{4D}}$ | $\eta$                                   | 60             | $[2.0, 5.0], \Delta\eta = 0.05$     |
|                                     | $\phi$                                   | 42             | $[-\pi, \pi], \Delta\phi = 2\pi/42$ |
|                                     | $p_{\text{T}}$                           | 4              | $[0.15, 0.35, 0.6, 1.2, 5.0]$ GeV   |
|                                     | $\mathcal{N}_{\text{VELO}}^{\text{hit}}$ | 3              | $[0, 1800, 2600, 10000]$            |

Table 15.3.: 4D binning scheme used for the determination the acceptance efficiency.

is the prerequisite for particles to be reconstructed by the tracking algorithms.

In Fig. 15.4 the acceptance efficiency is presented as functions of  $\eta$ ,  $\phi$ ,  $p_{\text{T}}$  and the hit-multiplicity in the Vertex Locator. The shape of the acceptance in pseudorapidity can be explained by two effects: the reduced efficiency due to multiple scattering in the region around the beam-pipe flange ( $\eta \approx 4.35$ ), and the increasing fraction of particles that are deflected out of the detector when approaching the geometrical borders of the detector.

The same effect explains the decreasing acceptance efficiency in the azimuthal angle distribution. Particles in the horizontal  $xz$ -plane ( $\phi$  around  $0^\circ$  and  $180^\circ$ ) can be deflected into the beam pipe or interact with its material. At  $\phi = \pm 90^\circ$ , in the vertical plane, the overlap of the sensor discs and the resulting larger amount of material induces a slight decline of the efficiency.

The acceptance efficiency also shows a dependence on the particle's transverse momentum. For an increasing  $p_{\text{T}}$ , and thus increasing  $p$ , the deflection by the magnetic field is less strong and an asymptotic behaviour of the efficiency is observed. With the same argument, the efficiency decreases towards low- $p_{\text{T}}$ . However, at even lower  $p_{\text{T}}$ , again a larger fraction of particles stays within the acceptance. This can be explained as follows: For two particles with the same momentum but different  $p_{\text{T}}$  the deflection by the magnetic field is the same. Due to the smaller opening angle (w.r.t. the beam axis) the particle with the lower  $p_{\text{T}}$  stays within the detector, while the higher  $p_{\text{T}}$  particle does not.

Since only geometric and kinematic properties define the acceptance, no dependence on the hit- or track-multiplicity of the events is observed. Hence, no parametrisation in these quantities is necessary. Nevertheless, for an easier implementation the same simple binning in VELO-hit multiplicity in analogy to  $\mathcal{P}_{\text{sec}}$  is adopted.

The parametrisation of the acceptance efficiency also has the focus on the angular variables,  $\eta$  and  $\phi$ , which are considered in a fine binning. Together with a simple binning in  $p_{\text{T}}$  and VELO-hit multiplicity, a 4-dimensional parametrisation of the acceptance efficiency,  $\epsilon_{\text{acc}}(\eta, \phi, p_{\text{T}}, \mathcal{N}_{\text{VELO}}^{\text{hit}})$ , is realised. The total binning scheme listed in Tab. 15.3.

#### 15.1.4. Track finding efficiency

The tracking efficiency,  $\epsilon_{\text{tr}}$ , describes the fraction of reconstructible particles (c.f. acceptance efficiency) that are in fact reconstructed by the track finding algorithms. The performance of the reconstruction is again evaluated as functions of  $\eta$ ,  $\phi$ ,  $p_{\text{T}}$  and the VELO-hit multiplicity, and is depicted in Fig. 15.5.

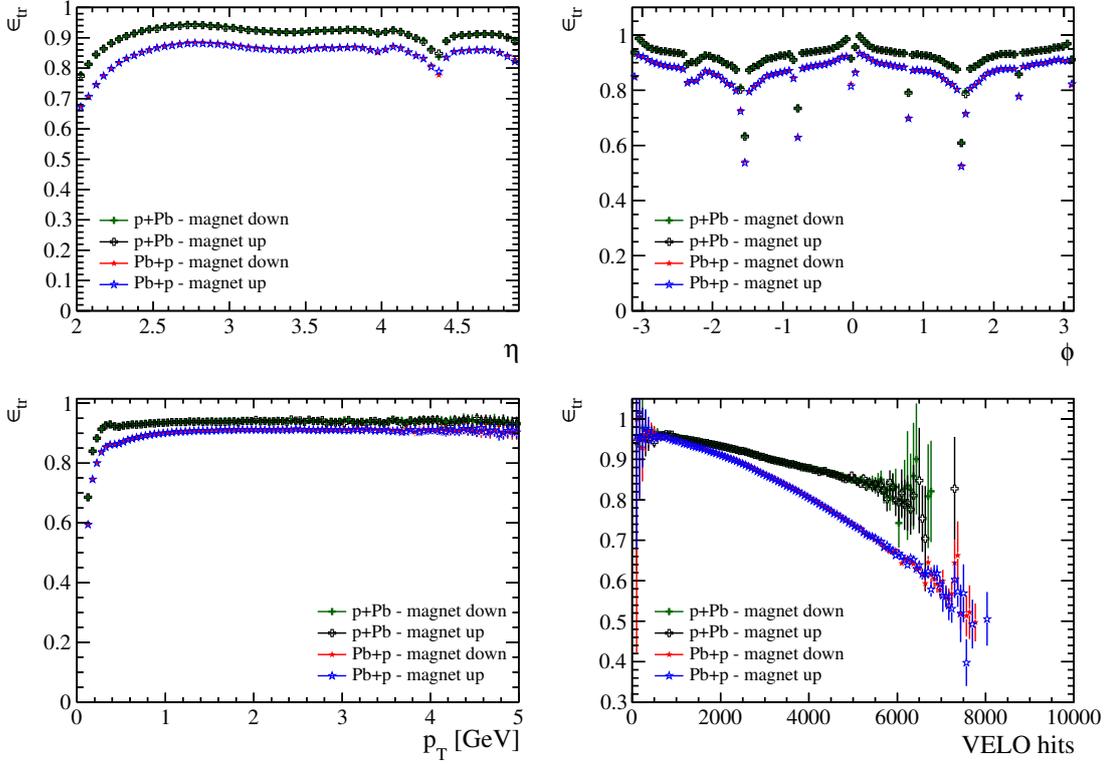


Figure 15.5.: Tracking efficiency studied in one dimension as functions of  $\eta$ ,  $\phi$ ,  $p_T$  and VELO-hit multiplicity in simulation. Results are presented for  $p+Pb$  and  $Pb+p$  collisions in magnet-up and -down configurations.

|                        | Variable                   | Number of bins | Bin range                              |
|------------------------|----------------------------|----------------|--|
| $\epsilon_{tr}^{3D}$   | $\eta$                     | 60             | $[2.0, 5.0]$ , $\Delta\eta = 0.05$     |
|                        | $\phi$                     | 42             | $[-\pi, \pi]$ , $\Delta\phi = 2\pi/42$ |
|                        | $p_T$                      | 3              | $[0.15, 0.3, 1.0, 5.0]$ GeV            |
| $\epsilon_{tr}^{3+1D}$ | $\eta$                     | 2              | $[2.0, 3.5, 5.0]$                      |
|                        | $\phi$                     | 2              | $[-\pi, 0, \pi]$                       |
|                        | $p_T$                      | 2              | $[0.15, 0.3, 1.0, 5.0]$ GeV            |
|                        | $\mathcal{N}_{VELO}^{hit}$ | 150            | $[0, 10000]$                           |

Table 15.4.: Binning scheme used for the determination of the tracking efficiency. The first block corresponds to the fine binning scheme optimised for the angular variables. The second block shows the simple binning scheme with an additional fine parametrisation of the hit-multiplicity in the VELO.

The angular distributions show the already discussed patterns induced by the segmentation of the VELO sensors (every  $45^\circ$ ) or induced by additional material, either from the beam pipe ( $\eta \approx 4.35$ ), or the RF-foil ( $\phi \approx \pm 90^\circ$ ).

From the  $p_T$ -distribution follows that for  $p_T \gtrsim 1$  GeV the tracking efficiency is almost independent of the transverse momentum of the particles. Only the low- $p_T$  region, where the particles' trajectories are disturbed by multiple scattering, shows a lower reconstruction efficiency.

Of particular importance for the performance of the tracking algorithms is the hit-multiplicity in the tracking detectors. With the increasing VELO-hit multiplicity and the resulting larger combinatorics, the efficiency of the pattern recognition in tracking drops down to  $\approx 50\%$  in extremely busy Pb+p collisions. The difference in performance between p+Pb and Pb+p collisions originates from the higher particle density which is present in the direction of the lead-ion remnant.

In order to properly account for the occupancy dependence of the tracking efficiency, a (3+1)-dimensional binning scheme, similar to that of the fake-track fraction, is applied. First,  $\epsilon_{\text{tr}}^{3\text{D}}(\eta, \phi, p_T)$  is defined by a fine parametrisation in  $\eta$  and  $\phi$ , in combination with two  $p_T$ -bins. Second, the occupancy dependence is accounted for by a binned scaling factor,  $\kappa_{\text{tr}}^{3+1\text{D}}(\eta, \phi, p_T; \mathcal{N}_{\text{VELO}}^{\text{hit}})$ , which uses a simpler binning in  $\eta$ ,  $\phi$  and  $p_T$ , but an additional parametrisation as a function of VELO-hit multiplicity. The exact binning scheme is summarised in Tab. 15.4. Combining both approaches results in a (3+1)-dimensional parametrisation of the track reconstruction efficiency,

$$\epsilon_{\text{tr}} = \epsilon_{\text{tr}}^{3\text{D}}(\eta, \phi, p_T) \cdot \kappa_{\text{fake}}^{3+1\text{D}}(\eta, \phi, p_T; \mathcal{N}_{\text{VELO}}^{\text{hit}}). \quad (15.3)$$

### 15.1.5. Correction procedure

The statistical correction of the track sample is performed by a two-step weighting procedure. Each track is assigned with two individual weighting factors,  $\omega_p$  accounts for the average purity of the track,  $\omega_\epsilon$  accounts for the average efficiency of reconstructing it. Both weighting factors depend on the kinematic properties of the respective track and on the global activity of the event, which is represented by the number of VELO-hits.

The purity weighting factor consists of the probability to reconstruct a fake track and to mis-select a non-prompt track:

$$\omega_p = 1 - \mathcal{P}_{\text{fake}}(\eta, \phi, p_T, \mathcal{N}_{\text{VELO}}^{\text{hit}}) - \mathcal{P}_{\text{sec}}(\eta, \phi, p_T, \mathcal{N}_{\text{VELO}}^{\text{hit}}). \quad (15.4)$$

In analogy, the efficiency weighting factor is the inverse of the product of the acceptance and the tracking efficiency:

$$\omega_\epsilon = 1/(\epsilon_{\text{acc}}(\eta, \phi, p_T, \mathcal{N}_{\text{VELO}}^{\text{hit}}) \cdot \epsilon_{\text{tr}}(\eta, \phi, p_T, \mathcal{N}_{\text{VELO}}^{\text{hit}})). \quad (15.5)$$

The final correction factor comprises the product of both weighting factors,  $\omega_p \cdot \omega_\epsilon$ .

The correction procedure is tested in simulation by comparing the reconstructed track samples after applying the correction factors with the corresponding distributions that are obtained by using generator information. Figure 15.6 shows the performance

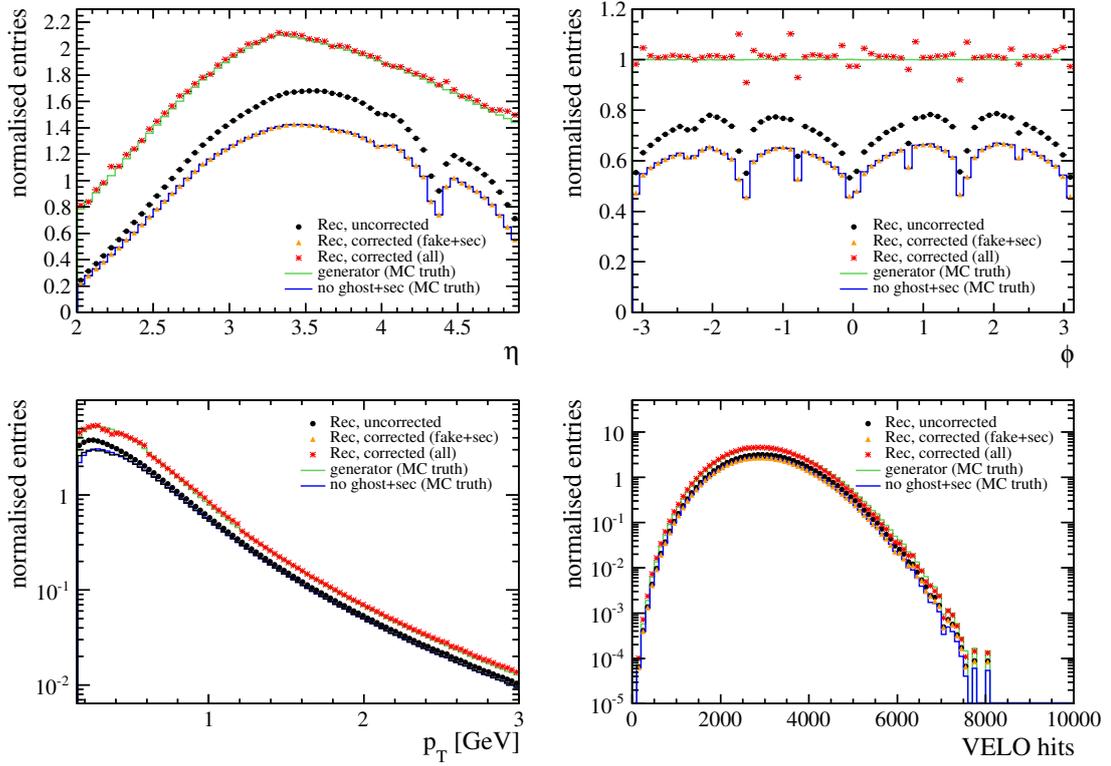


Figure 15.6.: Performance of the correction method, evaluated as functions of  $\eta$ ,  $\phi$ ,  $p_T$  and VELO hit-multiplicity by using simulation. Only results for Pb+p collisions are presented, similar results are obtained for p+Pb collisions. The histograms compare reconstructed data after different correction steps with the corresponding distributions using generator information.

after the two consecutive weighting steps. The distribution of the reconstructed but uncorrected simulation (black points) shows the characteristic patterns induced by the detector, as explained in the previous section. After weighting by purity factors (orange triangles) the obtained distributions are in excellent agreement with the corresponding distributions for which fake tracks and non-prompt particles are removed with generator information (blue lines).

After the additional weighting that accounts for the detector acceptance and the track efficiency the fully corrected distributions (red stars) are in satisfactory agreement with the original generator distributions (green lines). Within the precision required for this analysis, the corrected angular distributions are free of detector induced structures. Remaining fluctuations are related to the finite bin sizes of the weighting factors. The analysis method itself, presented in the following sections, furthermore includes an intrinsic correction of the measured two-particle correlations and can compensate remaining effects.

The corrected  $p_T$ -distribution shows a step-like structure, which is introduced by the  $p_T$ -binning of the correction factors. Since this variable is only used for the classification of kinematic ranges, these structures are not of importance for the further analysis.

The hit-multiplicity distribution is well corrected. It should be pointed out that the lower and higher occupancy ranges of the simulation sample are statistically under-represented in comparison to the data. The chosen ansatz of a binned scaling factor that accounts for the occupancy dependence minimises the statistical limitation of the simulation sample.

As stated before, the correction procedure is based on a statistical weighting of the individual tracks. So far, potential correlations between contaminating tracks, or between particles that are not reconstructed are not explicitly considered. It is assumed that no artificial angular correlation structures are introduced by any of the discussed effects. In order to test this assumption, a simulation study is performed which compares correlation structures obtained by using particles directly from the event generator with correlations obtained by using reconstructed tracks. The study is presented in Sect. 15.6, after introducing the procedure used for measuring the particle correlations. It is concluded that no significant correlation structures remain after applying the statistical corrections.

## 15.2. Event Classification

As presented in Chap. 12, the properties and structures that can be observed in two-particle angular correlations are different depending on the total event activity. In (heavy-) ion collisions, event classes are often categorised by the *centrality* of the colliding particle system. There is no fundamental definition of centrality for a mixed beam configuration, such as proton-ion. In ion-ion collisions, centrality can be understood as the geometric overlap of the two colliding nuclei. It can be defined by the impact parameter, which is the distance between the centres of the colliding nuclei, see Fig. 15.7. Theoretically, this distance allows to relate the centrality with the number of participating nucleons in the collision. In practise, this is achieved by using simulations, *e.g.* based on the Glauber model [178, 179]. The impact parameter is experimentally not a directly

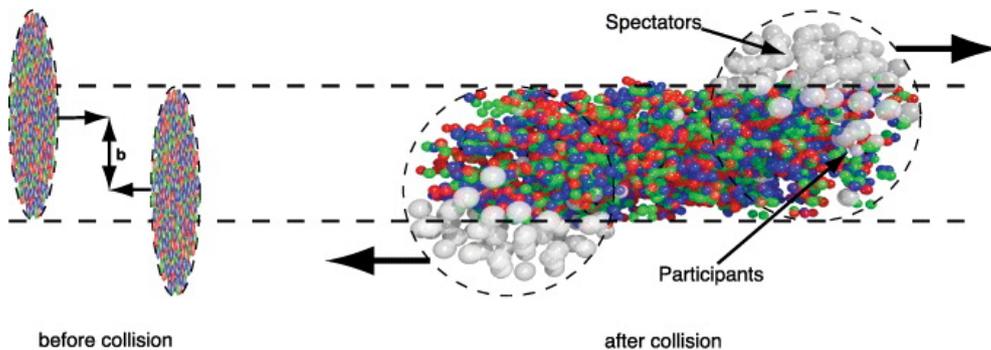


Figure 15.7.: Visualisation of a heavy-ion collision. The centrality is typically defined by the impact parameter  $b$  of the colliding nuclei and can be related to the number of participating nucleons in the collision. The figure is taken from Ref. [177]

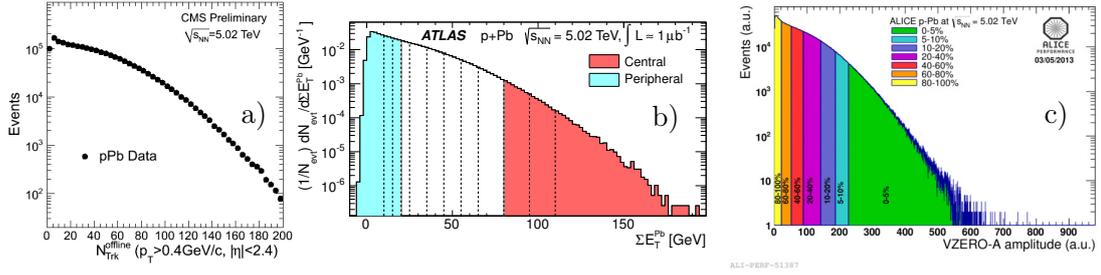


Figure 15.8.: Different approaches of defining the event activity of proton-ion collisions among the LHC experiments. Within CMS [8, 181] the activity is categorised by the number of reconstructed tracks (a), ATLAS [182] uses the transverse energy deposit in a forward calorimeter in direction of the lead beam (b), and ALICE [183] uses charge deposit in a forward oriented scintillator (c).

accessible quantity. Therefore, centrality is inferred by the number of charged particles being produced in the collisions, see *e.g.* Ref. [180]. Percentiles of a charged particle multiplicity distribution can be then used to define the centrality of the collision.

In a simple picture, the colliding proton in a proton-ion collision sees a larger amount of nucleons when facing the centre of the nucleus instead of only grazing the edge area. On average, *central* collisions will also produce a larger number of prompt particles compared to *peripheral* collisions, since not only the number of interacting nucleons is larger, but also the probability to undergo hard-scattering increases. Even though a distinction between more central and more peripheral proton-ion collisions is adequate, the aforementioned concept of centrality is only defined for colliding nuclei. As a consequence, proton-ion collisions are classified in terms of an abstract *event activity*. Using an experimental approach to measure the activity of the collision allows a multitude of observables to be used. A comparison among LHC experiments which have performed particle correlation studies in proton-ion collisions demonstrates that there is no unique or optimal choice.

The first analysis performed by the CMS collaboration [8] is based on an event classification by track multiplicity. In this analysis, all primary tracks of an event are counted which are reconstructed within the main fiducial region of the CMS detector ( $|\eta| < 2.4$  and  $p_T < 0.4$  GeV). From the track multiplicity distribution, as depicted in Fig. 15.8 a), events are grouped in different multiplicity classes. These are obtained by defining quantiles of the total distribution.

A different analysis presented by the ATLAS collaboration [182] relies on a complementary approach. Instead of using reconstructed tracks, events are classified by the total transverse energy,  $\sum E_T$ , which is measured in the forward calorimeter (coverage:  $3.1 < \eta < 4.9$ ), see Fig. 15.8 b). It is important to point out that also different pseudorapidity ranges are considered. While the event classification applied in the CMS analysis is based on an observable probing the same central pseudorapidity range as the actual measurement, the ATLAS analysis utilises a separate forward  $\eta$ -range, which has no geometric overlap with the range of the measurement.

In the third LHC analysis, performed by the ALICE collaboration [183], event classes are defined by the total energy deposition in the forward mounted VZERO

detectors ( $2.8 < \eta < 5.1$  and  $-3.7 < \eta < -1.7$ ). As for the ATLAS analysis, calorimeter information in the forward region is used, while the particle correlations are measured in the central pseudorapidity region.

The LHCb detector is only instrumented in the forward region. Hence, defining event classes by a measure that probes a complementary pseudorapidity region w.r.t. the actual correlation measurement is not possible. However, several event measures probing a similar kinematic region can be considered for defining the event activity. The advantages and disadvantages of three suitable measures are discussed in the following:

**Track multiplicity:** In analogy to the CMS analysis and motivated by the fact that track multiplicity is directly related to the number of prompt particles in a collision, track multiplicity distributions of all LHCb track-types are an adequate choice. Aspects in its disfavour are the discussed impurities due to fake tracks and non-prompt particles, as well as inefficiencies in the reconstruction. These effects contribute non-linearly to the number of tracks per event. Furthermore, the performance of the track reconstruction depends on the particular algorithms and are not necessarily comparable between different time periods. By accounting for inefficiencies and contaminations in the track sample a more appropriate particle-multiplicity distribution can be obtained. From the physics point-of-view, each track type covers a different kinematic range of particles and thus probes different properties of the collision. From the list of track types (c.f. Chap. 5.2) VELO tracks and long tracks are best suited. VELO tracks access the largest range in pseudorapidity and provide the best reconstruction performance. Long tracks only probe a sub-range of VELO tracks, due to the kinematic separation by the dipole magnet. As a result, using VELO tracks for the event classification includes low-momentum particles ( $p < 2 \text{ GeV}$ ), while using long tracks does not. On the other hand, long tracks probe the same kinematic range that is also accessed in the correlation measurement.

**Hit multiplicity / detector occupancy:** In contrast to reconstructed tracks, raw hit-multiplicities are a robust measure of the event activity, because no reconstruction software is involved. The hit-multiplicity in a sub-detector is proportional to the number of charged particles created in a collision. Each sub-detector again probes different properties of the collision. As the VELO is located directly around the primary interaction the number of secondary particles and thus the amount of hits induced by them is small. Furthermore, the large  $\eta$ -coverage in the forward and also in the backward direction gives the most inclusive measure of the event activity. The other tracking detectors, such as the TT, can only probe sub-ranges in pseudorapidity. The main tracking stations located downstream the dipole magnet can only access particles with a higher momentum. In conclusion, the VELO hit-multiplicity is the most comprehensive hit-based measure that probes prompt particles.

**Calorimeter cluster:** The advantage of a calorimeter based quantity is naturally the inclusion of neutral particles. All sub-detectors of the LHCb calorimeter system are located downstream of the dipole magnet. As a result, the calorimeter can only measure neutral particles, or charged particles with at least 2 GeV momentum. In addition to the more complex electromagnetic and hadronic calorimeters, the scintillator pad detector, which is particularly designed to distinguish pions and photons, can also provide a measure of the event activity. However, it is only sensitive to charged particles, a classification by using SPD-hits is thus comparable to using T-station information.

From the discussion follows that there are several good observables to classify the activity of an event. In order to examine different sorts of event activity and by keeping the analysis in a reasonable scope, two opposing measures are selected:

- The hit-multiplicity of the Vertex Locator ( $\mathcal{N}_{\text{VELO}}^{\text{hit}}$ )
- The corrected track-multiplicity of long tracks ( $\mathcal{N}_{\text{long}}^{\text{trk}}$ ) that are passing the track selection (c.f. Tab. 14.5 in Sec. 14.5), including correction factors.

While the VELO-hit multiplicity probes the event activity most comprehensively, the corrected long-track multiplicity directly quantifies the amount of charged particles within the kinematic range of the correlation measurement. Figure 15.9 shows a broad correlation between both activity measures, displayed separately for  $p+\text{Pb}$  and  $\text{Pb}+p$  collisions. The corresponding 1-dimensional distributions of both variables are given in Fig. 15.10. From these distributions quantiles are defined, where each corresponds to a discrete event-activity class. The  $p+\text{Pb}$  and  $\text{Pb}+p$  data samples are categorised separately, since in  $\text{Pb}+p$  collisions the lead beam is pointing in the forward direction and much higher track- and hit-multiplicities in the detector are reached. The event classes used in the course of the analysis are defined as follows:

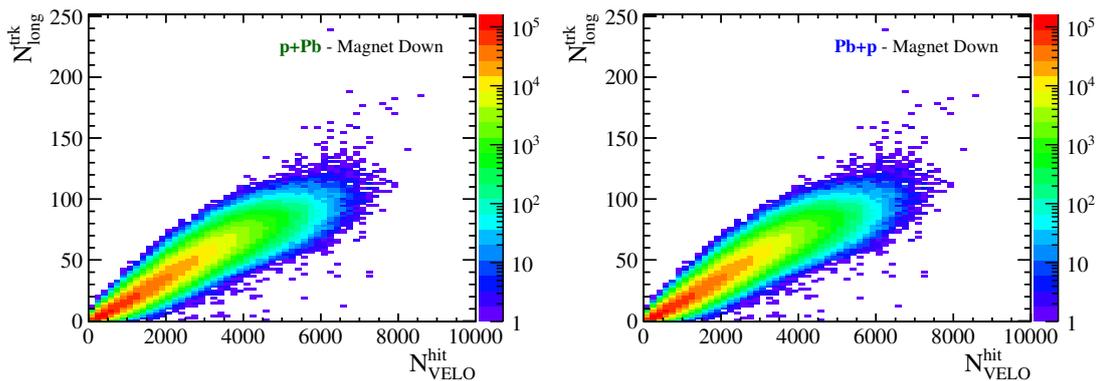


Figure 15.9.: Correlation between long-track and VELO-hit multiplicity in selected  $p+\text{Pb}$  and  $\text{Pb}+p$  collisions. Only long tracks in the kinematic range of  $2.0 < \eta < 4.9$ ,  $p > 2 \text{ GeV}$  and  $p_{\text{T}} > 150 \text{ MeV}$  are considered. The track multiplicity is corrected for fake tracks, non-prompt particles, detector acceptance and tracking efficiency.

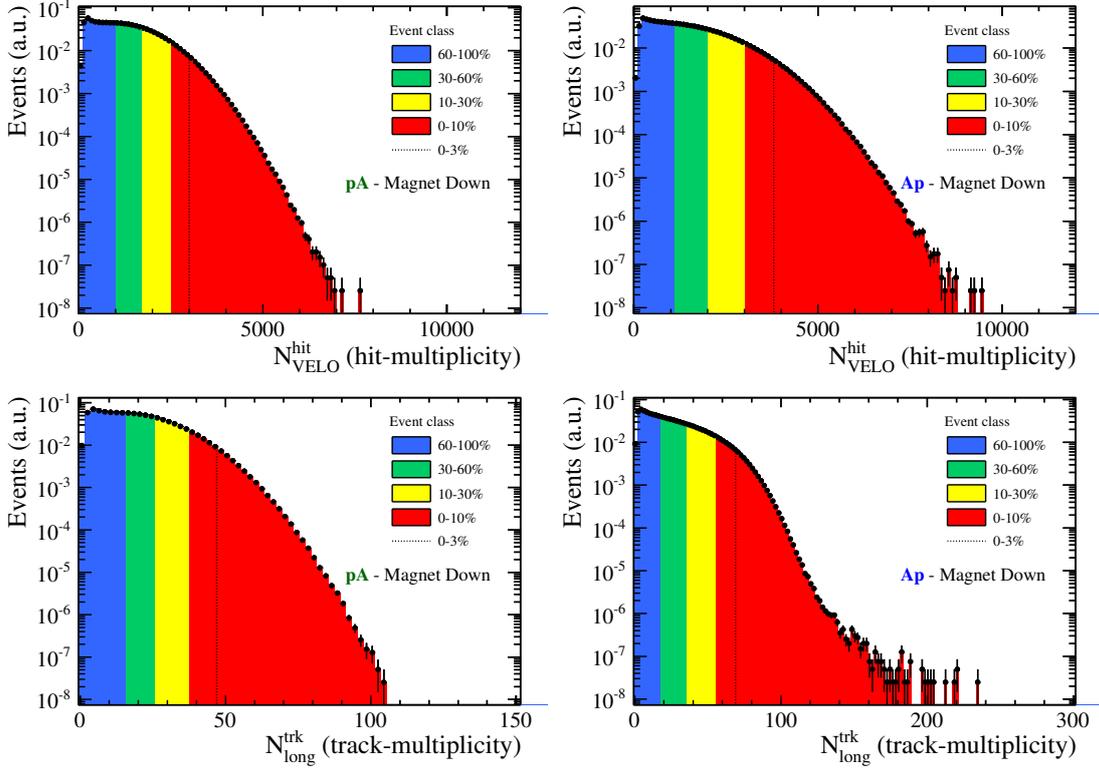


Figure 15.10.: Multiplicity distributions of VELO-hits ( $\mathcal{N}_{\text{VELO}}^{\text{hit}}$ ) and long tracks ( $\mathcal{N}_{\text{long}}^{\text{trk}}$ ) for selected  $p+\text{Pb}$  and  $\text{Pb}+p$  events, as described in the text. Activity-classes are defined as quantiles of the full distributions.

| Data          | Event Class  | Range ( $\mathcal{N}_{\text{VELO}}^{\text{hit}}$ ) | $\langle \mathcal{N}_{\text{long}}^{\text{trk}} \rangle$ | Range ( $\mathcal{N}_{\text{long}}^{\text{trk}}$ ) | $\langle \mathcal{N}_{\text{long}}^{\text{trk}} \rangle$ |
|---------------|--------------|--|--|--|--|
| $p+\text{Pb}$ | minimum-bias | all  | 27.4   |  |  |
|               | 60 – 100%    | [ 0, 925]  | 11.9   | [ 0, 20]   | 10.8   |
|               | 30 – 60%     | [ 925, 1675]                                       | 28.3   | [20, 36]   | 27.7   |
|               | 10 – 30%     | [1675, 2450]                                       | 42.0   | [36, 52]   | 43.0   |
|               | 0 – 10%      | [2450, max]  | 57.0   | [52, max]  | 62.1   |
|               | 0 – 3%       | [3000, max]  | 64.0   | [64, max]  | 72.2   |
| $\text{Pb}+p$ | minimum-bias | all  | 38.9   |  |  |
|               | 60 – 100%    | [ 0, 1050]   | 14.2   | [ 0, 17]   | 13.6   |
|               | 30 – 60%     | [1050, 1950]                                       | 37.9   | [17, 35]   | 38.4   |
|               | 10 – 30%     | [1950, 3000]                                       | 62.2   | [35, 55]   | 63.4   |
|               | 0 – 10%      | [3000, max]  | 87.4   | [55, max]  | 90.7   |
|               | 0 – 3%       | [3750, max]  | 96.8   | [69, max]  | 103.5  |

Table 15.5.: Definition of the event-activity classes for  $p+\text{Pb}$  and  $\text{Pb}+p$  collisions.

- The *low*-activity class comprises all events with a multiplicity within the lowest 40% fraction of the respective distribution. In the nomenclature of centrality, this range is referred to as 60 – 100% and corresponds to peripheral collisions.
- The next two activity classes with an increasing multiplicity are the *medium-low* (30 – 60%) and the *medium-high* (10 – 30%) activity class. These intermediate classes comprise events in the transition from low to high event-activities. Hence, exclusive high-activity correlation features should slowly emerge for events within this classes.
- The *high*-activity class (0 – 10%) includes all events within the 10% tail of the respective distributions and predominantly contain central proton-ion collisions.
- A sub-sample of the high-activity class is defined, which selects events from the highest 3% tail of the multiplicity distributions. It is referred to as the *highest*-activity class (0 – 3%).

All exclusive event classes are indicated as coloured fractions in Fig. 15.10, the threshold of the 3%-sub-sample is indicated by a dashed line. The numerical definition of each event class, either relying on hit- or on track-multiplicity, is given in Tab. 15.5, together with the corresponding mean track-multiplicity,  $\langle \mathcal{N}_{\text{long}}^{\text{trk}} \rangle$ , of each class.

The high-occupancy data sample defined in Chap. 14.1, pre-selects events with at least 2200 VELO-hits. This threshold is chosen such that the high- and highest-activity classes based on  $\mathcal{N}_{\text{VELO}}^{\text{hit}}$  are fully covered by this data sample, the track-based classes are not considered.

### 15.3. The correlation function

To access the collective behaviour of many particles in an event, the correlation between two particles of many particle-pairs in the event are analysed. The track sample, which represents prompt particles, can be understood as groups of *trigger* and *associated* particles. Selecting a random trigger particle results in a multitude of particle-pairs between the trigger particle and the remaining group of associated particles.

Historically, the method applied for studying particle correlations has been introduced for analysing heavy-ion collisions. In these analyses, the trigger particles were selected by the highest transverse momenta. The high- $p_T$  particles were used to define the direction of jets, in which all other associated particles have lower momenta. In general, the groups of trigger and associated particles can be either defined for independent  $p_T$ -ranges or as a subset of the same  $p_T$ -range. Among the LHC experiments, again no common approach is considered in performing proton-ion correlation studies. On the one hand, the ALICE collaboration uses separate  $p_T$ -ranges, where a few high- $p_T$  particles form the group of trigger particles and a larger set of lower- $p_T$  particles form the group of associated particles. On the other hand, the ATLAS and CMS collaborations utilise a common  $p_T$ -range for both groups. As shown *e.g.* in Refs [155, 182], particle correlation studies using separate definitions of  $p_T$ -ranges for trigger and associated particles give qualitatively similar results compared to analyses using common  $p_T$ -ranges. The more general approach of defining separate  $p_T$ -intervals, which may overlap, provide more

possibilities to analyse properties of the near-side ridge. Since this analysis is intended to first confirm its existence also in the forward region, it is sufficient to use a common  $p_T$ -interval for trigger and associated particles, following the approach used for the first observation of the near-side ridge in  $pPb$  collisions [8].

In the introduction to two-particle correlation measurements (Chap. 12.2) it is shown that the choice of the analysed kinematic range probes different types of correlations. The proton-ion measurements at the LHC indicate that the most interesting  $p_T$ -range for correlation studies is up to approximately 4 – 5 GeV. The selected tracks in this analysis are distributed over four intervals in transverse momentum:  $0.15 < p_T < 1.0$  GeV,  $1.0 < p_T < 2.0$  GeV,  $2.0 < p_T < 3.0$  GeV and  $3.0 < p_T < 5.0$  GeV. Due to the strongly decreasing number of particles with higher  $p_T$ , the range of the highest bin is chosen to be twice as large as the others.

Referring to the discussion about event activity and their classification, two-particle correlations are measured separately for each event-activity class. In addition, the tracks in each of the above given  $p_T$ -ranges are also considered separately. By using a common  $p_T$ -range for trigger and associated particles, each track of an event within a given interval is considered as a trigger particle. The total number of these tracks is further denoted as  $N_{trig}$ . When selecting one of these trigger particles the remaining particles in the same  $p_T$ -interval form the group of associated particles. Each trigger particle,  $T$ , then forms a particle pair with one particle  $A$  out of the group of associated particles. By avoiding double-counting of identical pairs this leads to a total number of  $N_{pair} = \binom{N_{trig}}{2}$  different particle pairs.

For each pair, the relative angles in azimuth,  $\Delta\phi = |\phi_T - \phi_A|$ , and in pseudorapidity,  $\Delta\eta = |\eta_T - \eta_A|$ , are calculated. Since the leading sign of the relative angles only depends on the chosen order of the particles,  $\Delta\eta$  and  $\Delta\phi$  are always taken to be positive. In the case of pseudorapidity, where a range from  $\eta = [2.0, 4.9]$  is probed, the range of values for  $\Delta\eta$  is  $[0, 2.9]$ , accordingly. The probability to find large values of  $\Delta\eta$  decreases linearly, because only a limited section of the  $\eta$ -space is covered.

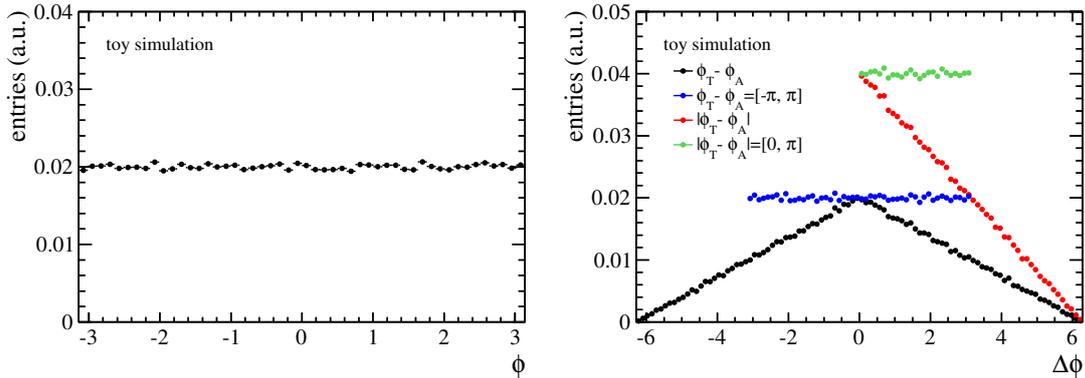


Figure 15.11.: Visualisation of different possibilities to calculate  $\Delta\phi$ . The left panel shows a flat distribution of particles in azimuth. The right panel shows the corresponding  $\Delta\phi$ -distribution for different calculation rules as explained in the text.

For the azimuthal angle the situation is slightly more difficult and is explained with the help of Fig. 15.11. The angle  $\phi$  is defined in the range of  $[-\pi, +\pi]$ , as displayed in a). The corresponding  $\Delta\phi$ -distributions for different calculation rules are shown in b). The simple difference  $\Delta\phi$  of two random particles, displayed in black, can have values in the range of  $[-2\pi, +2\pi]$ , with a larger probability for small compared to large differences. Consequently, also the absolute value of  $\Delta\phi$  (red), defined in the range of  $[0, +2\pi]$ , includes the same decreasing probability towards large angular differences. However, in contrast to  $\eta$ , the measured range in  $\phi$  covers the entire azimuthal space. As a result, the probability to find any value of  $\Delta\phi$  should be the same, when assuming a uniform distributed number of particles in  $\phi$ . Defining the range of values to be  $\Delta\phi = [-\pi, \pi]$  (blue), the calculation rule becomes:

$$\Delta\phi = \left\{ \begin{array}{ll} \phi_T - \phi_A & \text{for } \phi_T - \phi_A \in [-\pi, +\pi] \\ +2\pi + (\phi_T - \phi_A) & \text{for } \phi_T - \phi_A < -\pi \\ -2\pi + (\phi_T - \phi_A) & \text{for } \phi_T - \phi_A > +\pi \end{array} \right\}. \quad (15.6)$$

In the last step, only the absolute value  $|\Delta\phi|$  is considered (green), since the correlations are symmetric to the origin.

With the calculated differences in both angles,  $\Delta\eta$  and  $\Delta\phi$ , the correlation function is then defined as the *associated differential yield per trigger particle*:

$$\frac{1}{N_{trig}} \frac{d^2 N_{pair}}{d\Delta\eta d\Delta\phi} = B(0, 0) \times \frac{S(\Delta\eta, \Delta\phi)}{B(\Delta\eta, \Delta\phi)}. \quad (15.7)$$

Accordingly, also the 1-dimensional yield is introduced:

$$\frac{1}{N_{trig}} \frac{dN_{pair}}{d\Delta\phi} = B(0) \times \frac{S(\Delta\phi)}{B(\Delta\phi)}. \quad (15.8)$$

The functions  $S(\Delta\eta, \Delta\phi)$  and  $B(\Delta\eta, \Delta\phi)$ , as well as their 1-dimensional equivalents, denote *signal* and *background distributions*, respectively. The factors  $B(0, 0)$  and  $B(0)$  represent the yield of the background distribution at the origin and are used for normalisation of the background distribution, as explained later.

The signal distribution is the per-trigger associated yield for particles obtained from the same event,

$$S(\Delta\eta, \Delta\phi) = \frac{1}{N_{trig}} \frac{d^2 N_{same}}{d\Delta\eta d\Delta\phi}, \quad (15.9)$$

where  $N_{same}$  is the number of pairs given in a certain  $(\Delta\eta, \Delta\phi)$ -range. The index 'same' points out that all particle pairs are only obtained from particles of the same event. When summing over the events within a given event class (and the particles of a given  $p_T$ -range),  $N_{trig}$  and  $\frac{d^2 N_{same}}{d\Delta\eta d\Delta\phi}$  are determined separately before their ratio is calculated, following the analysis technique of Ref. [183]. An alternative approach, *e.g.* applied in Refs [8, 182], first calculates the entire signal function for each event and then averages it over the number of events. This latter approach introduces an additional multiplicity dependence in the pair yields, since events with different amounts of trigger particles are not weighted equally. Thus, the former implementation of an equal event weighting is adopted in this analysis.

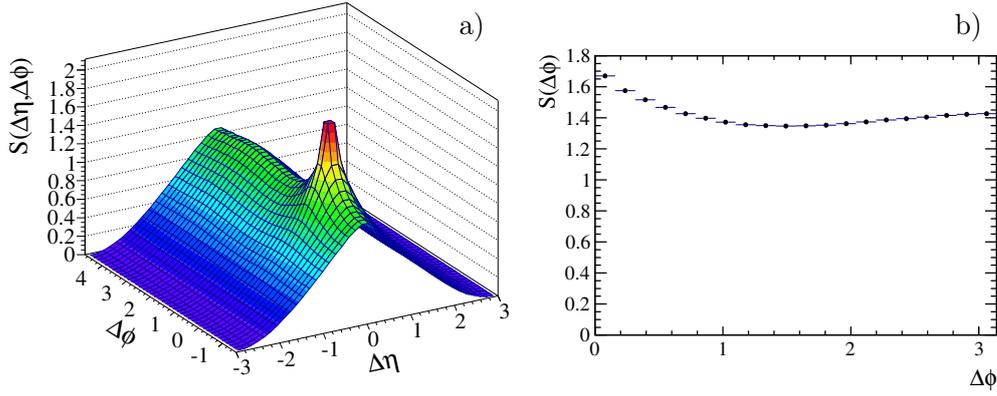


Figure 15.12.: Examples for two- and 1-dimensional signal distributions  $S(\Delta\eta, \Delta\phi)$  and  $S(\Delta\phi)$ , respectively.

Examples for 2- and 1-dimensional signal distributions are given in Fig. 15.12. By choosing all relative angles to be positive, only one quadrant  $(+\Delta\eta, +\Delta\phi)$  of the 2-dimensional histogram is actually computed. To obtain a better visualisation of the correlation properties the remaining three quadrants are filled by reflection. This results in a symmetric distribution around the origin. Following the discussion about two-particle correlation properties in Chap. 12.2, the interesting correlation features focus around relative azimuthal angles of  $\Delta\phi \approx 0$  and  $\Delta\phi \approx \pi$ . Hence, the displayed  $\Delta\phi$ -range in all 2-dimensional correlation histograms is shifted by a quarter of a period to a range of  $\Delta\phi = [-\frac{1}{2}\pi, +\frac{3}{2}\pi]$ .

The signal distribution can be understood as a per-trigger yield of particle pairs within the same event. The examples in Fig.15.12 show that there is a distinctive peak for particle pairs around the origin. While in the 1-dimensional distribution only a moderate increase towards  $\Delta\phi = 0$  is found, in the  $(\Delta\eta, \Delta\phi)$ -distribution the peak is much more pronounced. In addition, the entire 2-dimensional histogram is dominated by a triangle-shaped structure in  $\Delta\eta$ . As discussed previously, the shape is the result of the decreasing combinatorial probability to find large values of  $\Delta\eta$  when probing only a limited range in pseudorapidity. Thus, this dominant structure can be identified as an underlying correlation due to the acceptance of the detector.

The background further includes variations found in the correlation function that are introduced by a non-uniform detector acceptance in the angular variables  $\eta$  and  $\phi$ . Any artificial variation in these variables is naturally reflected in the angular differences. However, due to considering a large statistical sample, the effect must be strongly attenuated. In order to understand and correct for these effects, the background distribution is defined in a similar way as the signal distribution:

$$B(\Delta\eta, \Delta\phi) = \frac{1}{N_{\text{trig}}} \frac{d^2 N_{\text{mix}}}{d\Delta\eta d\Delta\phi}. \quad (15.10)$$

In this formula  $N_{\text{mix}}$  denotes the number of particle pairs obtained by combining all trigger particles of one event with particles from a different event (*mixed event*). Only the particles of the same  $p_T$ -range are combined in both events. The mixed event must

be of the same event-activity class and is required to have a similar primary vertex position: The longitudinal primary vertex position,  $z_{\text{pv}}$ , of the original and the mixed events have to agree within a 2 cm wide range,  $|z_{\text{pv}}^{\text{original}} - z_{\text{pv}}^{\text{mixed}}| < 2 \text{ cm}$ . This ensures, that particles in both events probe the same regions of the detector and have similar inefficiencies.

The mixed-particle pairs are physically unrelated to each other and only include the following background correlations which are not of interest in the analysis: correlations induced by the detector acceptance and correlations due to the general event topology. The latter includes *e.g.* correlations due to a non-flat particle density which decreased towards larger  $\eta$ . Since the measured events are all collimated in the forward direction, any resulting general correlation has to be accounted for.

For the background function it is convenient to increase the number of mixed events in order to avoid statistical limitations. In particular, this can affect low-activity events, since their particle multiplicities are typically also very low. A fixed number of 5 mixing events is required for all event classes. Considering more events does not show a sizeable impact on the results.

To find a proper normalisation for the background function, the pair yield at the origin,  $B(0,0)$ , is used. Particle pairs travelling into the same direction with relative angles of  $\Delta\eta \approx 0$  and  $\Delta\phi \approx 0$  have, per definition, the maximum possible geometric pair acceptance. By using the yield of the zero-bin as a normalisation factor, the resulting distribution,  $B(\Delta\eta, \Delta\phi)/B(0,0)$ , can be understood as a pair-acceptance efficiency, which is set to unity at the origin.

Figure 15.13 shows examples for the normalised 2- and 1-dimensional background distributions. With the applied corrections from the previous chapter, an essentially uniform coverage of the entire azimuthal angle is obtained. Thus, in both distributions in Fig. 15.13 the pair-acceptance efficiency in  $\Delta\phi$  is found to be almost flat. Depending on the considered event-activity and  $p_{\text{T}}$ -interval, small remaining structures in  $\Delta\phi$  can be still observed. In pseudorapidity difference, the aforementioned triangle shape is reproduced until reaching the maximum range of  $|\Delta\eta| < 2.9$ . Since the pseudorapidity distribution of particles is far from being uniform, the triangle shape is smeared.

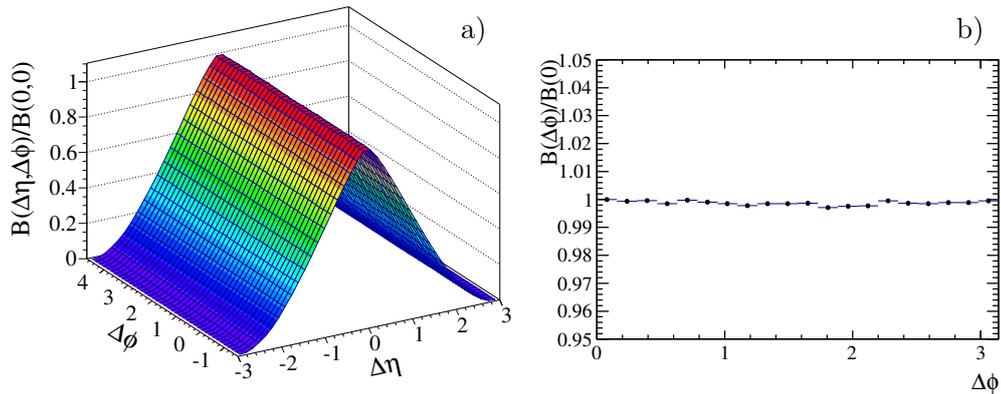


Figure 15.13.: Examples for normalised background distributions  $B(\Delta\eta, \Delta\phi)/B(0,0)$  and  $B(\Delta\phi)/B(0)$  in two (a) and one (b) dimensions, respectively.

The final correlation function is obtained by dividing the signal distribution,  $S(\Delta\eta, \Delta\phi)$ , by the normalised background distribution,  $B(\Delta\eta, \Delta\phi)/B(0,0)$ , as introduced in Eq. 15.7. This is visualised in Fig. 15.14. The method corrects the signal distribution for acceptance and background correlations and the resulting correlation histogram does no longer show the afore present global structures. Instead, previously hidden correlation patterns are highlighted. The still pronounced near-side peak at the origin is truncated in order to make other structures to become more visible. The final correlation function,  $\frac{1}{N_{\text{trig}}} \frac{d^2 N_{\text{pair}}}{d\Delta\eta d\Delta\phi}$ , now quantifies the per-trigger-particle associated yield.

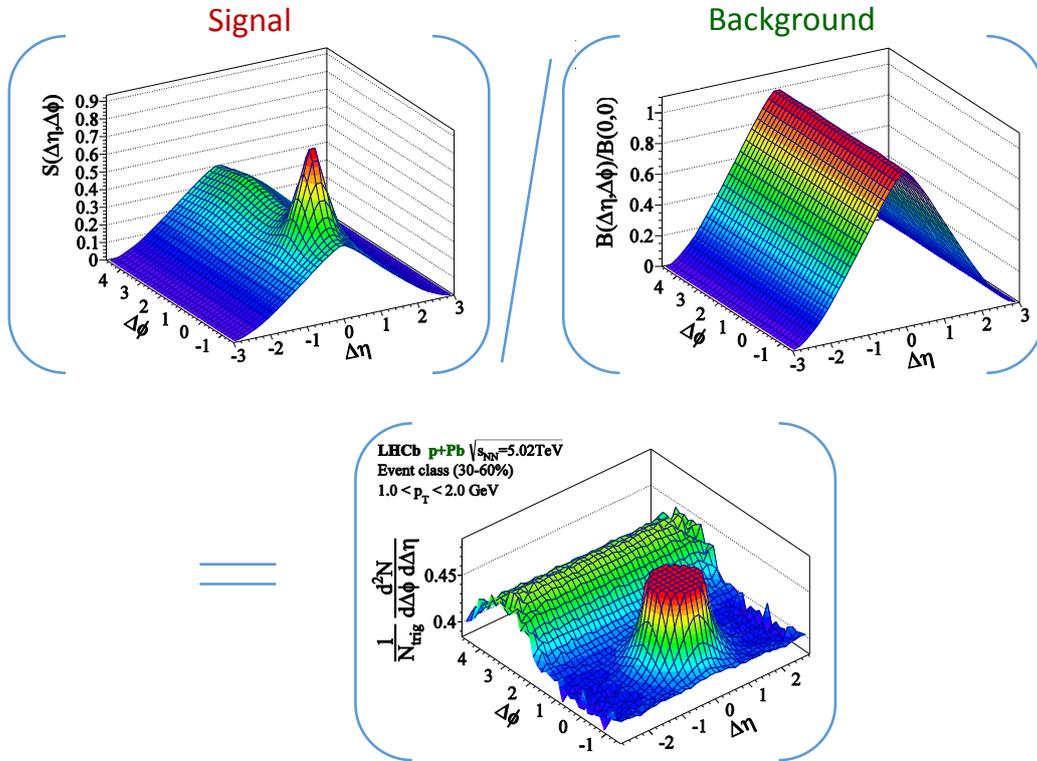


Figure 15.14.: Visualising Eq. 15.7, which is used to obtain the final correlation function. As an example data from  $p$ +Pb collisions are used. The signal function,  $S(\Delta\eta, \Delta\phi)$ , is divided by the normalised background function,  $B(\Delta\eta, \Delta\phi)/B(0,0)$ , in order to remove background correlations, such as induced by the detector acceptance.

## 15.4. One-dimensional long-range correlations in $\Delta\phi$

The two-dimensional representation of the correlation function allows a good qualitative comparison of the correlation properties. In order to make a quantitative statement, especially for the long-range correlations, it is useful to average the 2D function over a range in  $\Delta\eta = [\Delta\eta_a - \Delta\eta_b]$ , and obtain the 1D function  $Y(\Delta\phi)$ :

$$Y(\Delta\phi) := \frac{1}{N_{\text{trig}}} \frac{dN_{\text{pair}}}{d\Delta\phi} = \frac{1}{\Delta\eta_b - \Delta\eta_a} \int_{\Delta\eta_a}^{\Delta\eta_b} \frac{1}{N_{\text{trig}}} \frac{d^2 N_{\text{pair}}}{d\Delta\eta d\Delta\phi} d\Delta\eta. \quad (15.11)$$

The short-range correlations (*e.g.* the near-side peak) are present in a range of  $|\Delta\eta| \lesssim 2$  and dominate the less pronounced long-range correlations on the near-side, which are only visible for  $|\Delta\eta| \gtrsim 2$ . By averaging the two-dimensional function over a range of  $2 < \Delta\eta < 2.8$ , as indicated by the coloured region in Fig. 15.15 a), the one-dimensional yield,  $Y(\Delta\phi)$ , free of short-range correlations, is obtained.<sup>1</sup>

Figure 15.15 b) shows the resulting 1D yield as a function of  $\Delta\phi$ . In both representations of the yield, 1D and 2D, it is visible that the interesting correlation structures appear relative to an offset. In analogy to previous analyses, *e.g.* Refs [8, 182, 184], the ridge is investigated relative to the minimum of the total yield. Multiple sources of correlations produce a flat pedestal whose magnitude can be determined and afterwards subtracted by using the *zero-yield-at-minimum* (ZYAM) method [185]. In order to locate the position of the minimum,  $\Delta\phi_{\text{min}}$ , the averaged 1D yield,  $Y(\Delta\phi)$ , is fitted by a second-order polynomial in the range of  $0.1 < \Delta\phi < 2$ , where  $\Delta\phi_{\text{min}}$  is expected in the presence of a near-side ridge. Otherwise the yield simply decreases from its maximum at  $\Delta\phi = \pi$  towards  $\Delta\phi = 0$ . The yield at the fitted position of the minimum is henceforth referred to as the constant  $C_{\text{ZYAM}} = Y(\Delta\phi_{\text{min}})$ . An exemplary fit is depicted in Fig. 15.15 b). The obtained value of  $C_{\text{ZYAM}}$  is afterwards subtracted from  $Y(\Delta\phi)$  as

<sup>1</sup>Only the positive  $\Delta\eta$ -range is used for the averaging, since the negative part of the correlation function was obtained by mirroring the positive side.

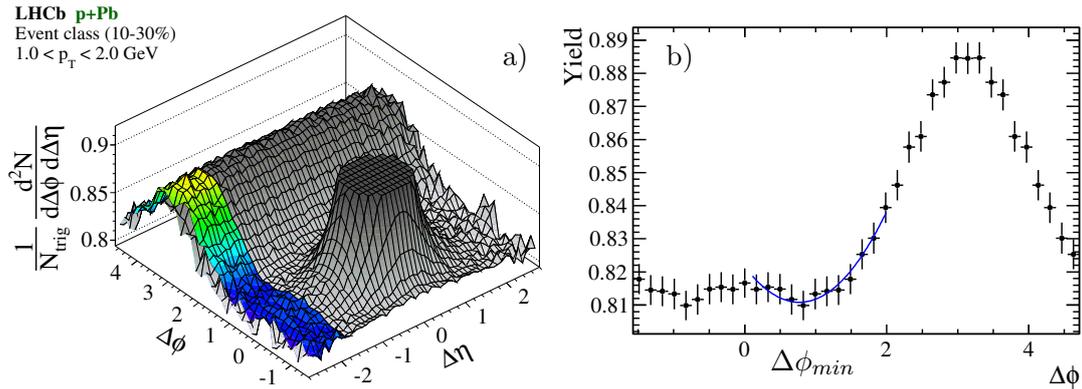


Figure 15.15.: Determination of  $Y(\Delta\phi)$  using the ZYAM-method. The two-dimensional associated yield (a) is averaged over a certain range in  $\Delta\eta$ , as indicated by the colourful region. The  $\Delta\eta$ -range is chosen such that no short-range correlations contribute to  $Y(\Delta\phi)$ . The obtained 1D yield (b) is fitted with a second-order polynomial to determine  $C_{\text{ZYAM}}$  at the position of the minimum.

a constant background. As a result, the height of the minimum is shifted to be at zero yield:  $Y(\Delta\phi_{min}) - C_{ZYAM} = 0$ .

The method allows a quantitative comparison of the long-range yields among different event classes and for different kinematic ranges.

## 15.5. Correlations in HIJING simulation

In this section, a brief look at the correlation function in the HIJING simulation is taken. The analysis of simulated events is carried out in the same way as explained for data.

The size of the generated proton-ion samples amounts to around  $2 \times 2\text{M}$  events (magnet-up & magnet-down) in each beam configuration,  $p+\text{Pb}$  and  $\text{Pb}+p$ , c.f. Tab. 14.4 in Chap. 14.4. The limited event statistics does not allow a separation in terms of event-activity classes. Hence, the correlation function is computed for all events in the samples, irrespective of the event activity.

A detailed study of the correlation properties is not possible by using the existing samples and would require a large simulation production. For this analysis, no quantitative comparison to the data results is intended. Instead, a qualitative overview of the visible correlation pattern is given in the following.

In Fig. 15.18, the results for the intermediate  $p_T$ -range of  $1 - 2\text{ GeV}$  is presented as an example. The  $p+\text{Pb}$  (a) and the  $\text{Pb}+p$  (b) samples both show the same correlation properties. Following the nomenclature of Chap. 12.2, the most prominent structure is the near-side peak, which is truncated in both histograms. This peak, related to jet-like objects, is located around the origin. It is accompanied by the away-side ridge ( $\Delta\phi \approx \pi$ ) which balances the momentum of the near-side. The third structure is the Gaussian ridge centred at  $\Delta\eta = 0$  and elongated over the entire  $\Delta\phi$  range. It is induced by decays of low- $p_T$  clusters or resonances. The inclusive event samples for both beam

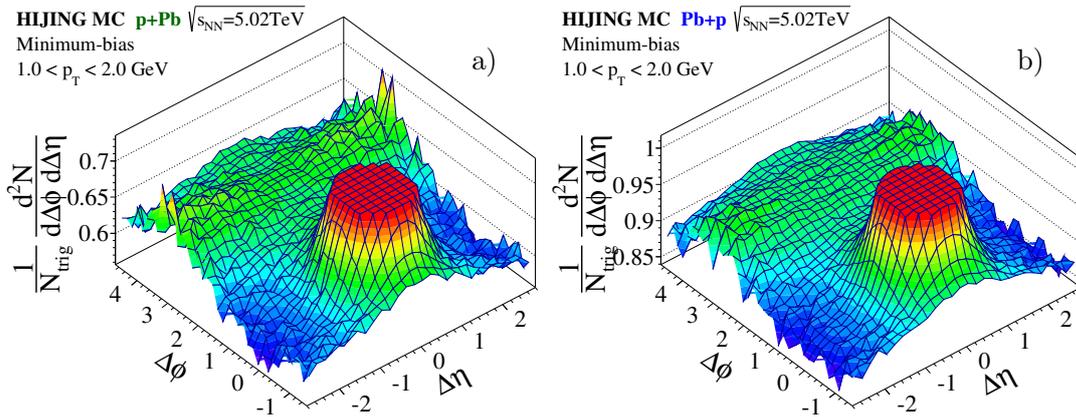


Figure 15.16.: Two-particle correlations in simulated proton-ion collisions using the HIJING event generator. Results are presented for the  $p+\text{Pb}$  (a) and  $\text{Pb}+p$  (b) beam configuration, probing the  $p_T$ -range of  $1 - 2\text{ GeV}$ . There is no classification in terms of event-activity applied. The simulation shows the well-known structures of the near-side peak, away-side ridge and Gaussian ridge. There is no sign of the ridge in the near-side.

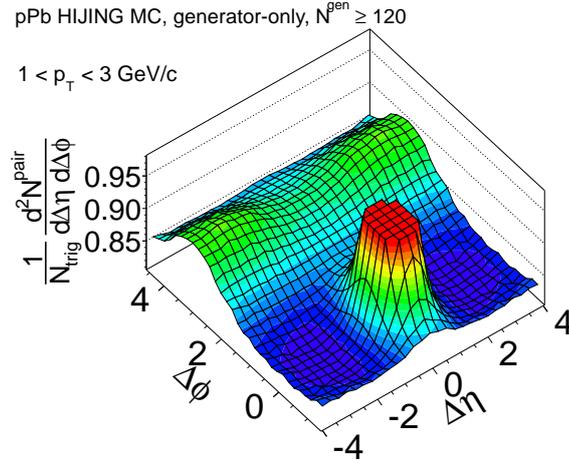


Figure 15.17.: Two-particle correlations in simulated proton-ion collisions at  $\sqrt{s_{NN}} = 5.02 \text{ TeV}$  using the HIJING event generator. The sample is generated for high-multiplicity events and probes the central rapidity region that is accessible by the CMS detector. The high-multiplicity sample shows no qualitative difference compared to the inclusive samples in Fig. 15.18, there is also no sign of a near-side ridge. Figure is taken from Ref. [181].

configurations contain many events with lower activity. A large fraction of particles are produced by soft processes, isotropically over the entire azimuthal angle. Thus, the broad Gaussian ridge is visible in the presented inclusive samples.

The near-side ridge is not visible in the HIJING simulation sample. No mechanisms are implemented in this event generator which are able to reproduce this long-range correlation. The low event statistics prevents for testing a high-activity sub-sample, in which the near-side ridge is observed in data. However, a similar HIJING sample containing only high-multiplicity events has been simulated by the CMS collaboration for the central rapidity region, the result is depicted in Fig. 15.17. Even though a slightly different kinematic range is tested, there are no qualitative differences visible compared to the samples produced for this analysis. There is also no indication for the presence of a long-range correlation on the near-side in high-multiplicity events.

## 15.6. Cross-check for artificial correlation structures

The statistical background correction, described in Sect. 15.1, accounts for fake and mis-selected tracks, as well as for inefficiencies due to the track reconstruction by applying per-track weighting factors. By comparing track-parameter ( $\eta$ ,  $\phi$  and  $p_T$ ) and hit-multiplicities directly obtained from the event generator with the corresponding reconstructed values after the weighting, a good agreement of the distributions is found. However, it is not yet checked, if artificial correlations remain in the angular observables,  $\Delta\eta$  and  $\Delta\phi$ , after this statistical correction. As an example, fake tracks could be correlated either with each other or with certain other particles, and thus introduce an additional non-physical correlation pattern.

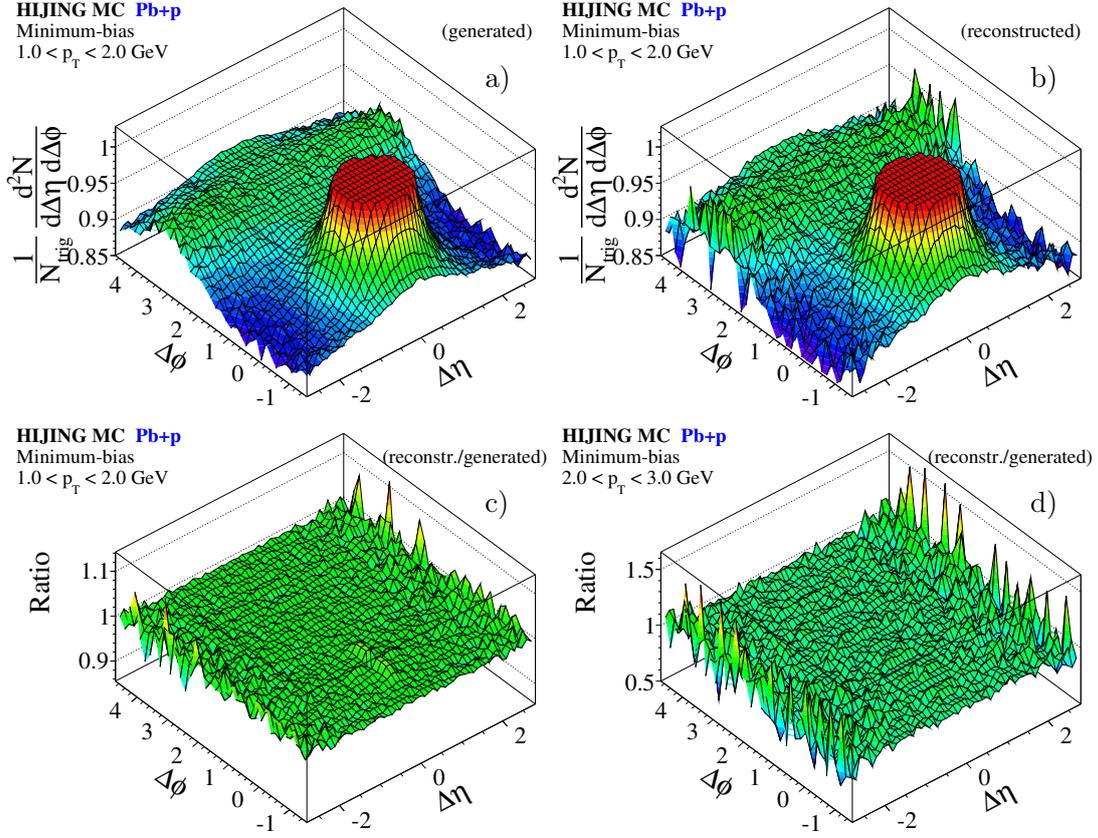


Figure 15.18.: Test for artificially induced angular correlation structures. Two-particle correlations are presented for minimum-bias selected Pb+*p* collisions in simulation. Results based on generator particles (a) are compared to results based on reconstructed tracks including correction factors (b), both for particles in the range of  $1 < p_T < 2$  GeV. After dividing both results (c), a flat distribution in  $(\Delta\eta, \Delta\phi)$  is obtained. The corresponding ratio for particles in the  $p_T$ -range of  $2 < p_T < 3$  GeV also shows no indication of artificial correlations (d).

In order to explicitly check that no artificial correlation structures are present in the fully corrected correlation histogram the following test is performed in simulation. The two-particle correlations obtained by using reconstructed tracks including all correction factors are compared to the corresponding original correlations found by using particles from the event generator. Since the size of the simulation sample does not allow a study for different event-activity classes, the simulated events are used altogether for this test. For further quantitative tests the production of larger simulation samples is required. The result of this test is depicted in Fig. 15.18 for Pb+*p* collisions, similar results are found for *p*+Pb collisions. The ratio of the generated and the reconstructed angular correlations is approximately flat in both probed  $p_T$ -ranges. No sign of a significant artificial correlation structure is observed.

This test shows that potential correlations of contaminating background tracks or inefficiencies of the reconstruction have no qualitative impact on the measurement of two-particle angular correlations.

In this chapter the results of the two-particle correlation functions are presented for the forward ( $p+\text{Pb}$ ) and backward ( $\text{Pb}+p$ ) beam configuration. All results comprise the combination of the magnet-up and magnet-down samples. The given event-activity classes are based on hit-multiplicity. The track-based approach is discussed in the next chapter as a systematic check.

The discussion of the results for both beam configurations is split into three parts. First, a general qualitative discussion of the properties visible in  $(\Delta\eta, \Delta\phi)$ -correlations is given for all event-activity classes and  $p_{\text{T}}$ -ranges. Second, the focus is set on the long-range correlations, in particular the search for the near-side ridge. Third, a discussion of the corresponding integrated yield as a function of  $\Delta\phi$  is provided.

## 16.1. Results from $p+\text{Pb}$ data

The final results of the angular correlations are first discussed for the  $p+\text{Pb}$  data, where the proton beam is in the forward direction. The results for the backward configuration ( $\text{Pb}+p$ ) are presented afterwards in Sect. 16.2.

### 16.1.1. Overview of 2D correlations in $p+\text{Pb}$

The correlation function is measured for different event-activity classes and transverse momentum ranges. An overview of the 2D correlations for the  $p+\text{Pb}$  data sample is given in Fig. 16.1, showing all four exclusive event-activity classes. The highest-activity class, which is obtained as a sub-sample of the high-activity class, is not included in this overview but will be discussed in the subsequent section (Sect. 16.1.2).

The results of the lower activity classes, represented in Fig.16.1 a)-f), are obtained by using the *minimum-bias* sample. The results of the higher activity classes (g-l), rely on the *high-occupancy* sample in order to provide a sufficient statistical power. Details of the data samples are discussed in Chap. 14.1.

All 2D correlations are rich in structure, which changes with the event classes and the  $p_T$ -ranges. As a general observation, the per-trigger-particle associated yield ( $\frac{1}{N_{\text{trig}}} \frac{d^2 N_{\text{pair}}}{d\Delta\eta d\Delta\phi}$ ), which is represented by the correlation function, increases when comparing the results starting from the top row going towards the bottom row in Fig. 16.1. This is a result of the increasing number of particles towards higher event-activity classes.

By briefly comparing the correlations in all different bins one can find that structures within the same  $p_T$ -range (columns) are present in all activity classes (rows). In contrast, within a certain activity class the visible structures vary significantly depending on the  $p_T$ -range.

The general features can be explained by using the example of Fig. 16.1 a) in the low event-activity class (60 – 100%). The depicted low- $p_T$  range is dominated by the Gaussian ridge which is centred at  $\Delta\eta = 0$  and elongated over the whole  $\Delta\phi$ -range. The low- $p_T$  particles in this bin are mostly created by soft-QCD processes. The shape can be interpreted as due to several independent soft processes creating low- $p_T$  clusters. Afterwards, these clusters decay and emit particles isotropically, which leads to an uniform correlation over the  $\Delta\phi$ -range. Also particles produced in decays of less boosted resonances can lead to the same uniform correlation. In addition to the Gaussian ridge, the near-side peak is present around  $\Delta\eta = \Delta\phi = 0$ . Any boosted object that decays will produce particles with a small separation in both angles,  $\eta$  and  $\phi$ . Finally, another structure is visible around  $\Delta\phi = \pi$  in all  $p_T$ -bins. This away-side ridge accounts for the recoil of particles on the near-side and balances the momentum. On average, the balancing particles are back-to-back in azimuth and thus are shifted by  $180^\circ$ . There is no preferred direction in pseudorapidity, since particles produced in the primary interaction can be randomly boosted, depending on the momentum of the interacting partons. As a result, the away-side ridge is elongated over the entire measured  $\Delta\eta$ -range.

By comparing the different  $p_T$ -ranges in the low event-activity class (60 – 100%) a strong change of the visible structures is observed. The Gaussian ridge disappears, when probing higher  $p_T$ -ranges, due to a less significant contribution of soft-QCD processes. The near-side peak in the low  $p_T$ -bin (a) is broader compared to those of the higher  $p_T$ -bins (b,c), since the total momenta are on average smaller and thus, also decay products of jet-like objects are less collimated. The structure in higher  $p_T$ -bins (b,c) are rather similar, only the statistical power in the high- $p_T$  range is strongly limited.

The medium-low (30 – 60%) and medium-high (10 – 30%) activity-classes of the  $p$ +Pb data sample in Fig. 16.1 d)-i) basically show the same qualitative correlations as found in the low activity-class. Comparing the  $1 < p_T < 2$  GeV bins (e,h), only the near-side peak and the away-side ridge are present. The high- $p_T$  bins (f,i) are qualitatively similar to the corresponding medium- $p_T$  bins. Only the yields are smaller, being the result of the lower average number of particles in the high- $p_T$  bins. Towards higher event activities the away-side ridge becomes more pronounced, which accounts for the higher probability to form jets. In both intermediate event classes, yet, no evidence of a ridge structure on the near-side is found, which would be visible around  $\Delta\phi = 0$  and  $|\Delta\eta| > 2$ . In this range, only a flat minimum is present.

Focusing on the high event-activity class ( $0 - 10\%$ ) reveals that the near-side ridge is evolving in the higher  $p_T$ -bins (k,l). Comparing all three  $p_T$ -ranges, there is no sign of the ridge at low transverse momenta (j) and only a slight indication in the range from  $2 < p_T < 3$  GeV (l). In the intermediate  $p_T$ -range of  $1 < p_T < 2$  GeV (k), the ridge

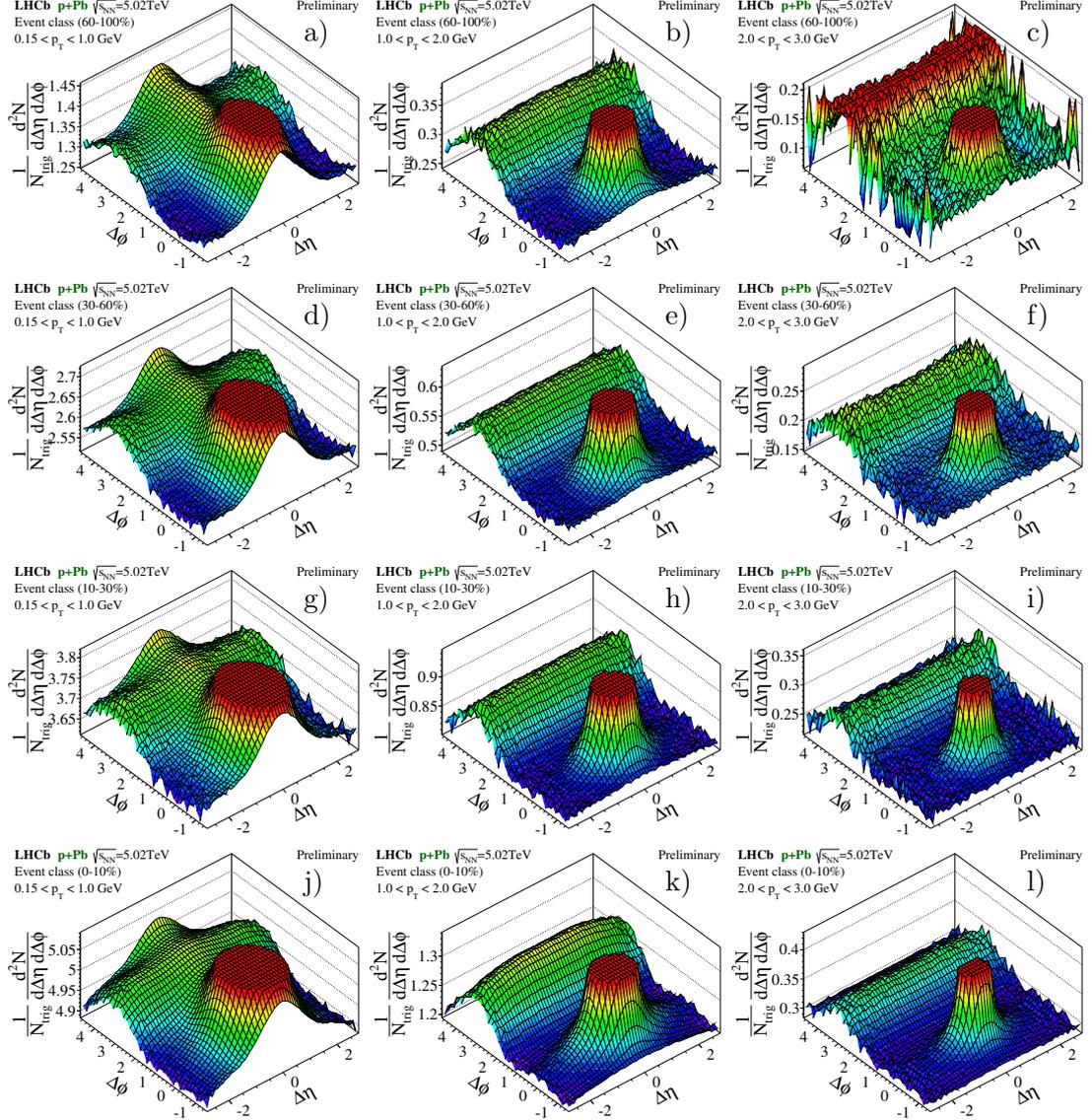


Figure 16.1.: Two-particle correlation functions for  $p$ +Pb collisions at  $\sqrt{s_{NN}}=5.02$  TeV. Each column corresponds to a separate  $p_T$ -range in increasing order from left to right. Each row represents a certain event-activity class starting with low activity on the top towards high activity on the bottom. In all histograms, the near-side peak is truncated to allow a better visualisation of other properties. Details are discussed within the text.

is not very pronounced but visible when comparing to the same  $p_T$ -range of the other activity classes (b,e,h). All other correlation structures are qualitatively the same as in the corresponding lower activity-classes.

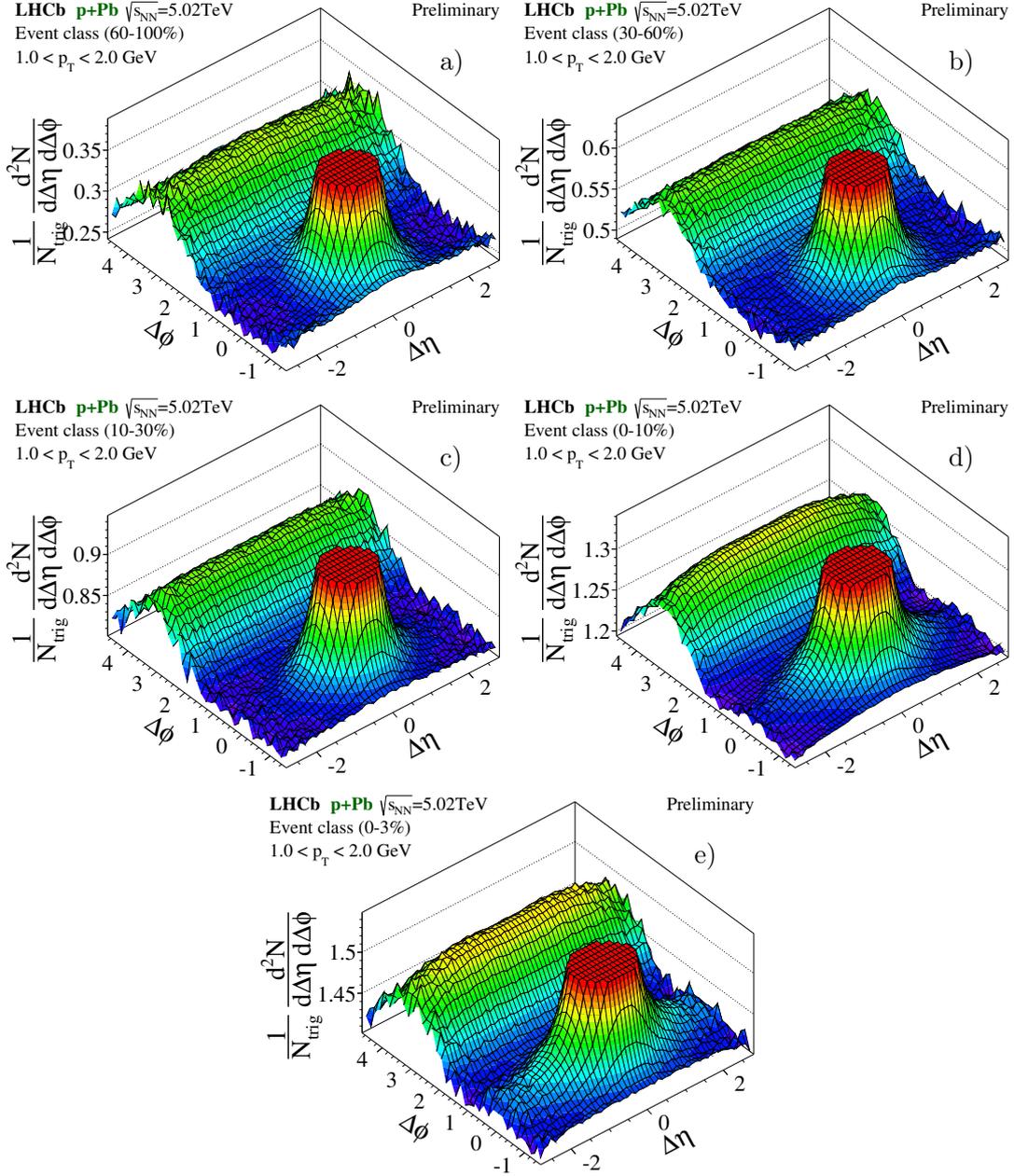


Figure 16.2.: Evolution of the near-side ridge in  $p+Pb$  collisions in the most interesting  $p_T$ -range of  $1 < p_T < 2 \text{ GeV}$ . Activity classes are defined by VELO-hit multiplicity. The low (a) and intermediate (b,c) activity classes show no indication of the ridge on the near-side, in the high activity class (d) a slight ridge structure is found. The highest-activity sub-sample (e), clearly proves the presence of the near-side ridge in the  $p+Pb$  data sample.

### 16.1.2. Long-range correlation in $p+\text{Pb}$

Long-range azimuthal correlations comprise the well-known ridge structure on the away-side but also the ridge on the near-side. While both correlations emerge over the entire measured  $\Delta\eta$ -range, the near-side ridge is always shadowed over a wide range by the presence of the near-side peak. As discussed previously, the near-side ridge is most pronounced in an intermediate  $p_{\text{T}}$ -range. Therefore the most interesting region to investigate long-range correlations, is the  $p_{\text{T}}$ -range of  $1 < p_{\text{T}} < 2 \text{ GeV}$ , which is presented in more detail in Fig. 16.2 for different event-activity classes.

In the low and intermediate activity classes, Fig. 16.2 a)-c) one can see that there is no indication of a ridge on the near-side. The away-side ridge is uniformly distributed in all bins and more distinct towards higher event-activity. The short-range correlations, such as the near-side peak, are restricted to a range of around  $|\Delta\eta| < 1.5$ . At larger  $\Delta\eta$ , only a flat valley is present. In the high-activity class, Fig.16.2 d), the same region accommodates the near-side ridge, although it is still not very distinct. This ridge is separated from the near-side peak, when comparing to the lower activity bins. By taking advantage of the special high-occupancy data sample, the large event statistics allows further analysing a sub-sample of events with even higher average hit-multiplicities. The *highest*-activity sample (0 – 3%) comprises the 3% fraction of events in the tail of the hit-multiplicity distributions, c.f. Chap. 15.2. This event class is presented in Fig. 16.2 e) and clearly shows the emergence of the near-side ridge. Compared to the (0 – 10%)-class, its magnitude increases, which confirms the observation of a smooth turn-on of this correlation structure. An approximately linear increase of the yield of the near-side ridge with track multiplicity is reported in Ref. [8].

### 16.1.3. 1D associated yield versus $\Delta\phi$ in $p+\text{Pb}$

In order to quantitatively examine the long-range correlations, the one-dimensional  $\Delta\phi$ -yield is calculated, as described in Chap. 15.4. The 2D correlation function is averaged over a limited region in  $\Delta\eta$ , which is not biased by the strong short-range correlations. The histograms in Figs 16.1 and 16.2 show that the peak-region is limited to a range of around  $|\Delta\eta| < 1.5$ , depending on the  $p_{\text{T}}$ -range. An integration region at even larger  $\Delta\eta$  values of  $2 < \Delta\eta < 2.8$  is chosen, in order to minimise potential remaining contaminations of short-range effects. The obtained 1D correlation functions for the entire kinematic range measured for all event-activity classes is summarised in Fig. 16.3. The pedestal yield of multiple other correlations is determined by the ZYAM-method (c.f. Chap. 15.4), the subtracted constant background,  $C_{\text{ZYAM}}$ , is quoted in each bin. The one-dimensional yields,  $Y(\Delta\phi)$ , show two interesting regions:

- The away-side ridge is centred around  $\Delta\phi = \pi$ , and is visible in each of the bins.
- The near-side ridge, if present, is centred around  $\Delta\phi = 0$ .

Focussing on the intermediate  $p_{\text{T}}$ -range of  $1 < p_{\text{T}} < 2 \text{ GeV}$  as an example, not only the underlying pedestal yield increases with the event-activity, but also the relative development of the away-side ridge. On the near-side, there is no indication of the ridge in the low and intermediate activity-classes.

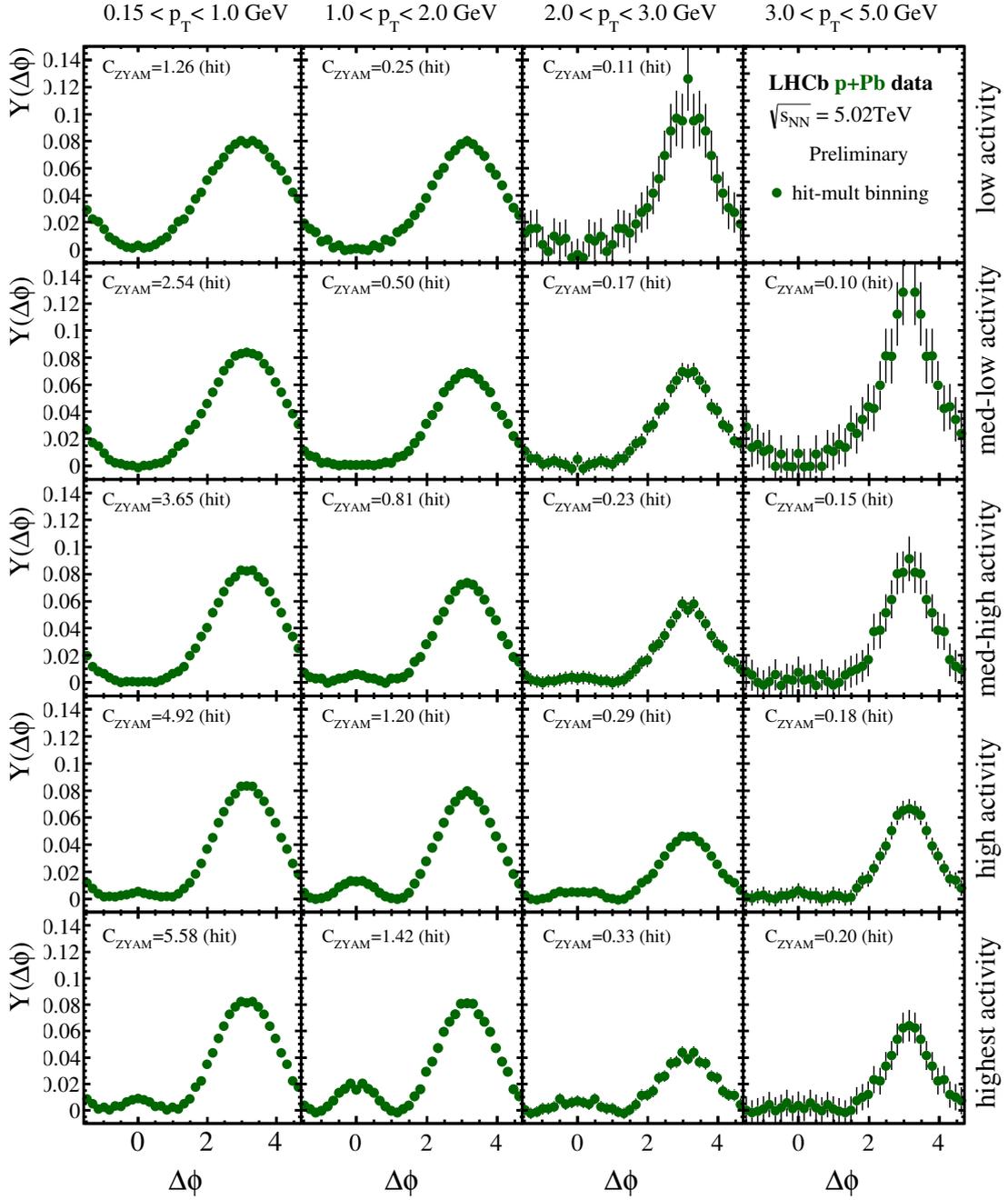


Figure 16.3.: Integrated one-dimensional correlation yield as a function of  $\Delta\phi$  in  $p$ +Pb collision at  $\sqrt{s_{NN}} = 5.02$  TeV. Results of all measured  $p_T$ -ranges and event-activity classes as displayed. The applied integration region is  $2 < \Delta\eta < 2.8$ . The final 1D yield is obtained by using the ZYAM-method, the subtracted constant is listed in each panel. All errors are statistical only.

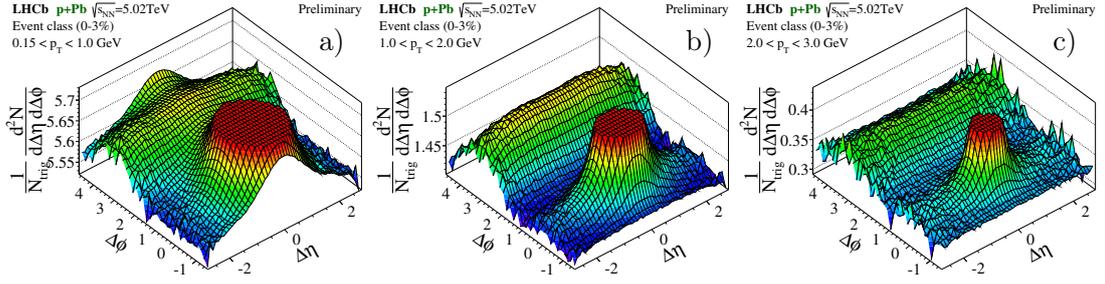


Figure 16.4.: Two-particle correlation functions for  $p$ +Pb data in the highest event-activity class using the high-occupancy data sample. Event-activity classes are defined by using VELO-hit multiplicity. The low- $p_T$  range (a) does not show an indication of the near-side ridge. In the intermediate- $p_T$  range (b), the ridge is clearly visible. At high- $p_T$  (c), also a slight ridge-like structure is found on the near-side, being in agreement with the observation of the integrated 1D yield of Fig. 16.3.

When probing high event-activities, the near-side ridge is present in a  $p_T$ -range of  $1 < p_T < 2$  GeV, as observed in the 2D correlations. For the highest activity-class, a maximum yield of around  $Y(0) \approx 0.02$  is reached on the near-side. The corresponding away-side is more pronounced by a factor of four, with a yield of  $Y(\pi) \approx 0.085$ .

Looking at the neighbouring  $p_T$ -bins, no definite statement can be made. The low- $p_T$  bins do not accommodate a near-side ridge. The tiny increase around  $\Delta\phi = 0$  in the low- $p_T$  bin cannot be related to a ridge structure, when comparing to the corresponding 2D correlation function that is given in Fig. 16.4 a). The situation for the high- $p_T$  bin ( $2 < p_T < 3$  GeV) is different. The slight excess in the  $\Delta\phi$ -correlation on the near side can be matched to a ridge-like structure in the respective 2D histogram in Fig. 16.4 c). However, the statistical power is limited.

The  $\Delta\phi$ -correlations in the highest  $p_T$ -bins ( $3 < p_T < 5$  GeV) of Fig. 16.3 show no indication for a ridge structure. All event classes are statistically limited due to the very small number of particles in this high- $p_T$  range.

## 16.2. Results from Pb+p data

After the discussion of the  $p$ +Pb data sample, the results in the backward configuration (Pb+ $p$ ) are presented. In this data sample, the lead remnant is travelling upstream through the LHCb detector, while the proton beam is pointing to the backward direction. As a result, the observed track multiplicity and thus also the particle densities in the accessible region of the detector are larger than in the  $p$ +Pb data sample. Since the mechanism which is responsible for the near-side ridge is sensitive to a very high particle density, a more pronounced ridge can be expected in the Pb+ $p$  data compared to what is found in the  $p$ +Pb data.

### 16.2.1. Overview of 2D correlations in Pb+p

In analogy to the  $p$ +Pb collisions, the correlation function for Pb+ $p$  collisions is measured in the same kinematic ranges and the same kind of event-activity classes. As a reminder, the activity classes for the Pb+ $p$  data are determined in the same quantiles as for the  $p$ +Pb data, but the hit- and track-multiplicity distributions reach higher values and thus lead to different thresholds, c.f. Chap. 15.2.

An overview of all measured 2D correlation functions in the Pb+ $p$  data sample is given in Fig. 16.5. The results of the low (a-c) and medium-low (d-f) activity classes are obtained from the minimum-bias sample, the medium-high (g-i) and high (j-l) activity classes rely on the high-occupancy sample. The results of the highest-activity sub-sample are discussed separately when focusing on the long-range correlations in Sect. 16.2.2.

The general structures that are visible in the 2D correlations of the Pb+ $p$  data in Fig. 16.5 are similar to what is found in the  $p$ +Pb data (Fig. 16.1).

The low- $p_T$  range, in particular the low-activity class (a), is dominated by the Gaussian ridge around  $\Delta\eta = 0$ . As in  $p$ +Pb data, this ridge is related to soft particle production and remains present in the corresponding higher activity classes (d,g,j), but becomes superimposed by other effects.

Also in Pb+ $p$  collisions, the near-side peak is the most distinct correlation structure in all categories. Its shape and magnitude is comparable to the results obtained in the  $p$ +Pb data.

The away-side ridge, which balances the momenta of particles on the near-side is also present in all measured correlation functions. Compared to the  $p$ +Pb data, the same behaviour is found. The away-side ridge can be best described by using the  $p_T$ -range of  $1 < p_T < 2$  GeV. At low (b) and medium-low (e) event-activities, the away-side ridge is already present and smoothly distributed along the entire  $\Delta\eta$ -range. Starting from medium-high (h) activities, the ridge on the away-side becomes more distinct. At high activities (k) the previously flat ridge decreases towards larger  $\eta$ -separations. It should be pointed out that this behaviour is not related to inefficiencies of the detector. The same effect has been also observed by other experiments and is again discussed in Sect. 16.3. The away-side ridge in the higher  $p_T$ -range but with same event-activity (l) is less affected and remains almost flat in  $\Delta\eta$ .

Of particular interest is the evolution of the ridge on the near-side. Figure 16.5 k) shows a very pronounced ridge structure, which significantly exceeds the ridge found in the  $p+Pb$  data. Already in the medium-high activity class (h), its magnitude is larger than the ridge seen in the highest activity class in  $p+Pb$  data. Also in the high- $p_T$  range (i,l), the near-side ridge is clearly visible as a steady structure elongated over the entire  $\Delta\eta$ -range. In both  $p_T$ -ranges a common turn-on point in event-activity is concluded.

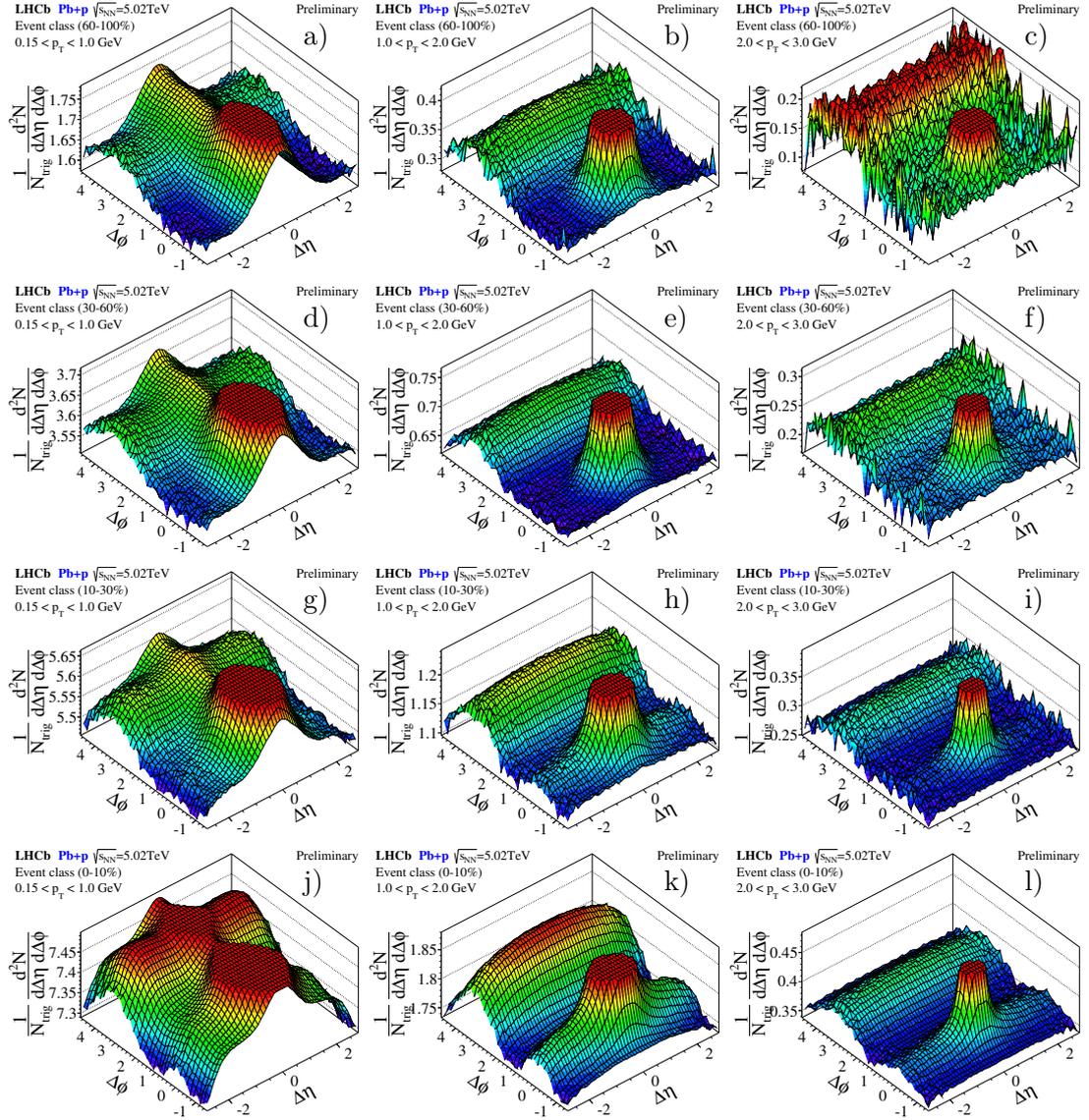


Figure 16.5.: Two-particle correlation functions for Pb+p collisions at  $\sqrt{s_{NN}}=5.02$  TeV. Each column corresponds to a separate  $p_T$ -range in increasing order from left to right. Each row represents a certain event-activity class, starting order with low activity on the top towards high activity on the bottom. In all histograms, the near-side peak is truncated to allow a better visualisation of other properties. Details are discussed within the text.

### 16.2.2. Long-range correlation in Pb+p

The long-range correlations are discussed again in more detail on the basis of Fig. 16.6. The displayed correlations in the  $p_T$ -range of  $1 < p_T < 2$  GeV are not influenced by the Gaussian ridge and nicely show the development of the long-range effects as a function of the event-activity.

In sub-figures a) and b), only the long-ranged correlations on the away-side are present. The near-side still does not show any ridge-like structure and only accommodates the short-range peak which is restricted to small  $\Delta\eta$  values. With the medium-high activity class (c), this region becomes additionally populated by the near-side ridge. Hence, in the Pb+p configuration this long-range correlation starts to evolve at lower relative event-activities than in the p+Pb setup. Since the activity-classes are defined separately for the p+Pb and Pb+p data, the chosen event-bins probe different absolute activity-ranges. The activity-classes in which the near-side ridge can be first seen are the 10–30% class in Pb+p and 0–10% class in p+Pb, respectively. The corresponding hit-multiplicity boundaries probe a range of 1950–3000 VELO-hits in Pb+p data and  $> 2450$  VELO-hits in p+Pb data. Due to the overlap in both ranges, this suggests a common turn-on point of this long-range correlation in both data samples.

Looking at the high-activity class, two interesting features are observed. The ridge at the away-side becomes much more distinct but also decreases slightly towards  $|\Delta\eta| \gtrsim 2.0$ . It is interesting to note, at even higher event-activities (e), this decrease seems to be slightly less pronounced again. As mentioned before, this behaviour can be also seen in other publications (*e.g.* CMS [8]), but no explanation is given. A direct comparison to other results from other experiments is given in Sect. 16.3.

The near-side ridge in the high activity-class is already very striking but still increases when probing the special highest-activity sub-sample shown in Fig. 16.7. Comparing both long-range effects shows that the near-side ridge reaches relative magnitudes of up to around 50% of the away side. Further quantification of the results are given in the subsequent section.

### 16.2.3. 1D associated yield versus $\Delta\phi$ in Pb+p

Long-range correlations in Pb+p data are studied in the same way as for p+Pb data. The one-dimensional function  $Y(\Delta\phi)$  is obtained by integrating the 2D correlation function over a  $\Delta\eta$ -range of  $2.0 < \Delta\eta < 2.8$ . This range only probes long-range correlations, the short-ranged peak is excluded. The constant background due to multiple other correlations is again subtracted by the ZYAM-method. The resulting one-dimensional  $\Delta\phi$ -correlation functions for all measured bins are presented in Fig. 16.7, together with the subtracted constant  $C_{ZYAM}$ .

The single 1D yields reveal the same basic properties as previously seen in p+Pb data. However, the observed magnitudes of both ridges are very different. The yield at the away-side ( $\Delta\phi = \pi$ ) shows the general tendency to increase with the event-activity and reaches a yield of up to  $Y(\pi) \approx 0.1$  in the lower and intermediate  $p_T$ -ranges. In addition, a decrease of the away-side yield towards larger  $p_T$  is observed, even though results in the high- $p_T$  range are often statistically limited.

The near-side, around  $\Delta\phi = 0$ , shows no sign of a ridge induced yield in the low and medium-low activity-bins. In the medium-high activity-bin a ridge induced yield of

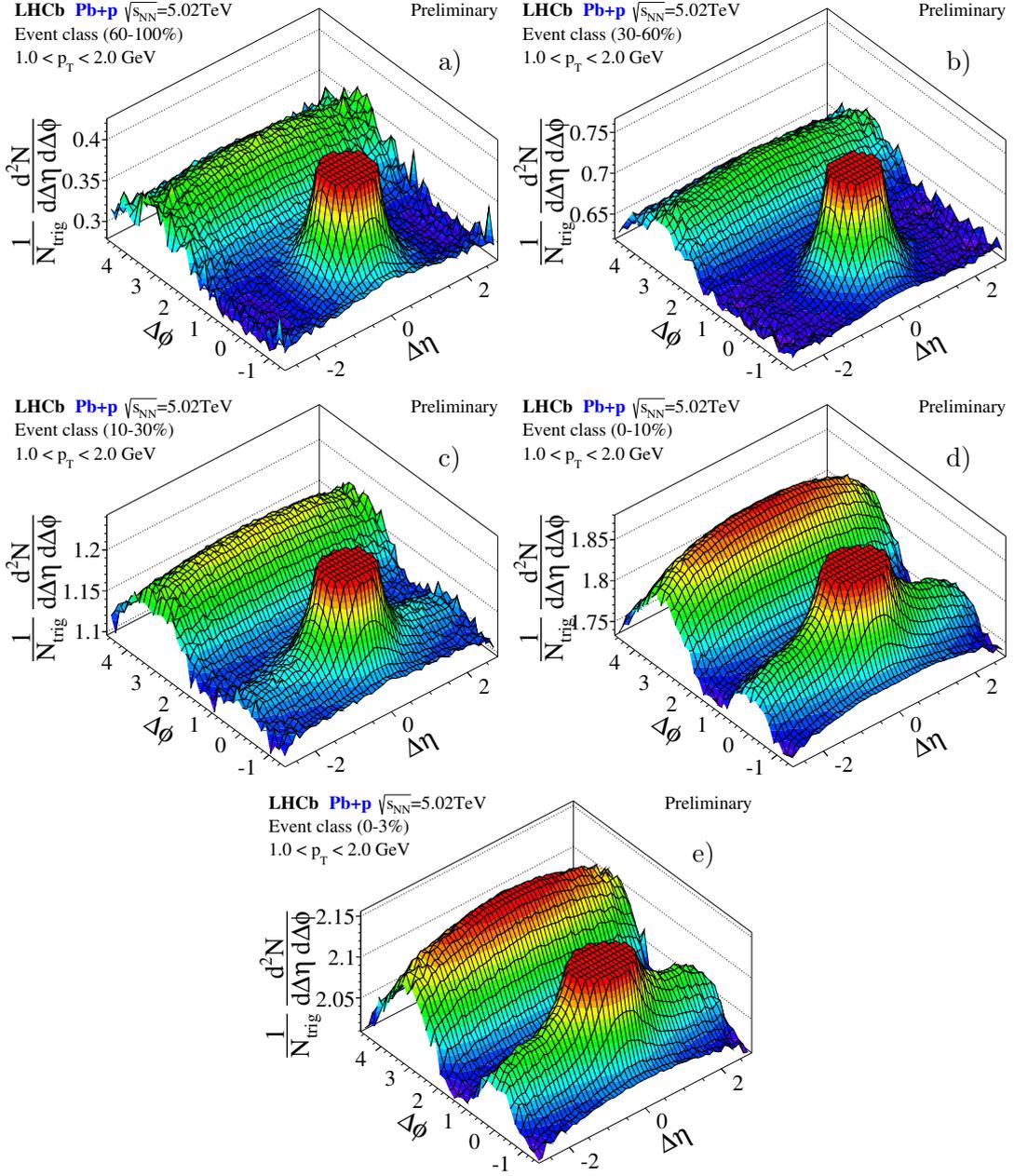


Figure 16.6.: Evolution of the near-side ridge in Pb+p collisions in the most interesting  $p_T$ -range of  $1 < p_T < 2 \text{ GeV}$ . Activity classes are defined by VELO-hit multiplicity. As in p+Pb data, the low (a) and medium-low (b) activity classes show no sign of the near-side ridge. In contrast, already in the medium-high activity class (c) the ridge is clearly visible and confirms its existence in the Pb+p data. In the high-activity class (d), the ridge structure is much more distinct compared to the p+Pb data, the highest-activity sub-sample (e) shows an even more pronounced near-side ridge.

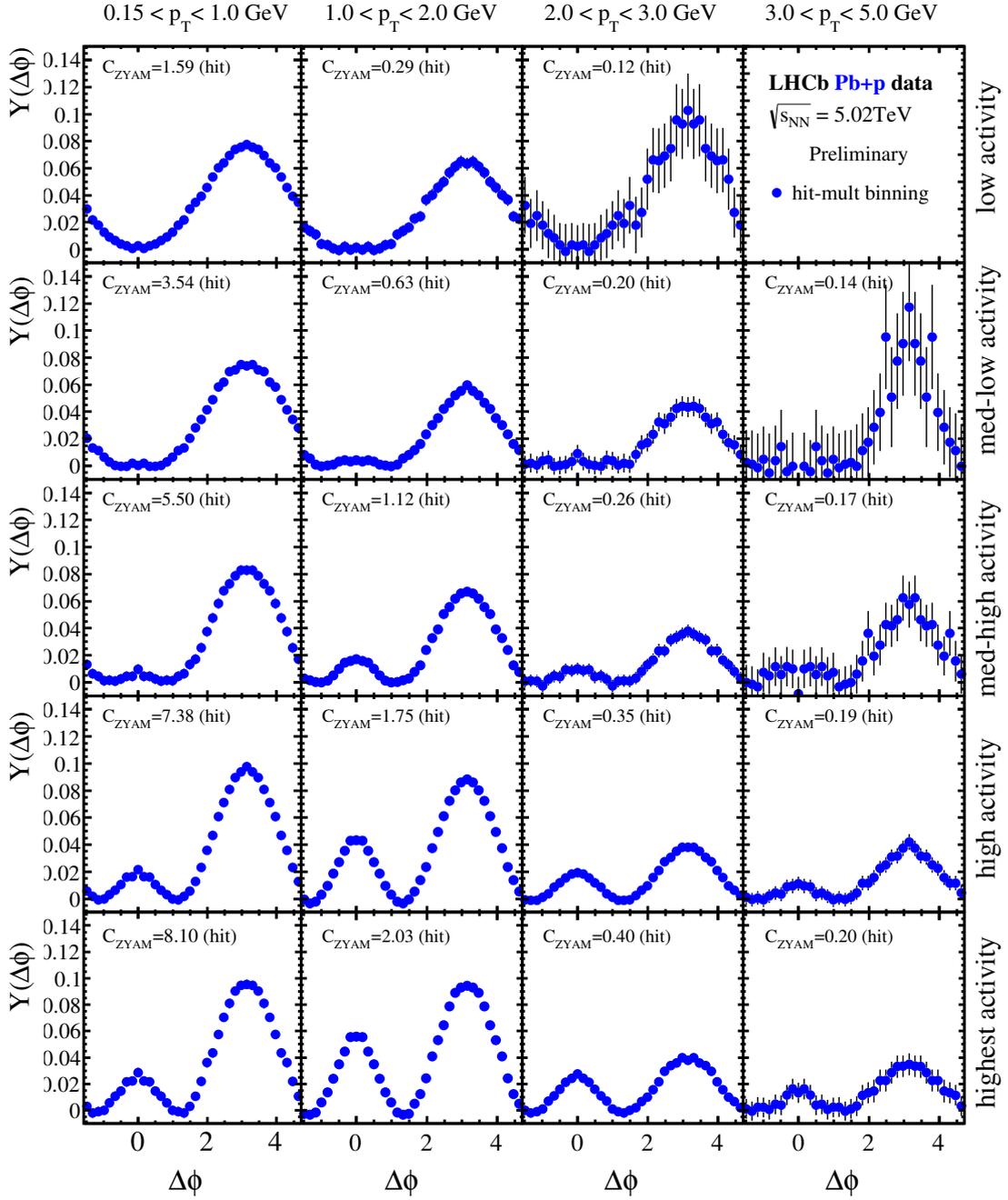


Figure 16.7.: Integrated one-dimensional correlation yield as a function of  $\Delta\phi$  in Pb+p collision at  $\sqrt{s_{NN}} = 5.02$  TeV. Results of all measured  $p_T$ -ranges and event-activity classes as displayed. The applied integration region is  $2 < \Delta\eta < 2.8$ . The final 1D yield is obtained by using the ZYAM-method, the subtracted constant is listed in each panel. All errors are statistical only.

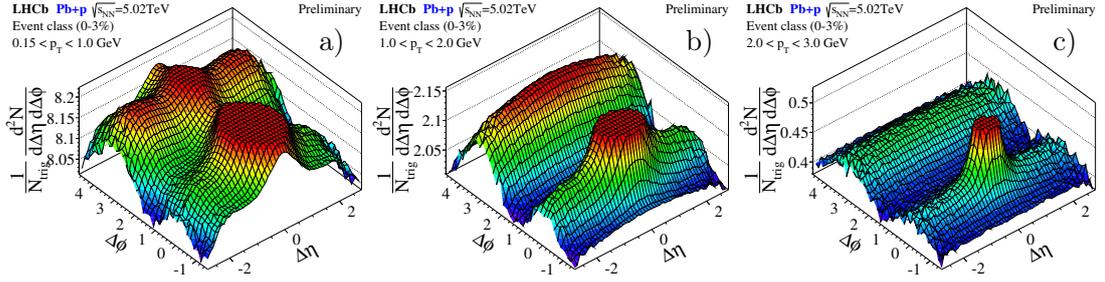


Figure 16.8.: Two-particle correlation functions for Pb+ $p$  data in the highest event-activity class using the high-occupancy data sample. Activity-classes are defined by using VELO-hit multiplicity. The near-side ridge appears to be present in all  $p_T$ -ranges at this multiplicity regime. In the low- $p_T$  range (a) various structures are present in the correlations function, also a near-side ridge appears to be present. In the intermediate- (b) and high- $p_T$  ranges (c) the near-side ridge is clearly visible.

$Y(\Delta\phi = 0) \approx 0.015$  is found in the  $1 < p_T < 2$  GeV range, which is already larger than the corresponding yield in high-activity  $p$ +Pb collisions. Looking at the Pb+ $p$  data in the high-activity bin in the  $p_T$ -range of  $1 < p_T < 2$  GeV, the near-side yield reaches a maximum of  $\approx 0.50$  compared to a yield of  $\approx 0.90$  at the away-side. It follows that in the most busy Pb+ $p$  collisions, the correlation strength of the near-side ridge becomes more than half the size of the away-side ridge.

In the higher  $p_T$ -region of  $2 < p_T < 3$  GeV the absolute yields at the near- and away-side are significantly lower, but the ratio between both ridges increases up to a value of around 75%.

Looking at the near-side yield in the two higher activity-classes, a ridge like behaviour is seen in all  $p_T$ -ranges. Focussing only on the highest activity-class, the 1D yield can always be related to a ridge structure in the corresponding 2D correlation functions, which are depicted in Fig. 16.8. In the intermediate- $p_T$  range the ridge is clearly visible. By taking advantage of the high statistics of the data sample, the near-side ridge can also be confirmed in the high- $p_T$  range. At low- $p_T$ , various structures are present in the correlation function. However, in contrast to the  $p$ +Pb data, a ridge-like structure centred around  $\Delta\phi = 0$  appears to be present, see Fig. 16.8 a).

### 16.3. Qualitative comparison of the ridge among LHC experiments

All four LHC experiments participated in the proton-ion data taking and have performed two-particle correlations measurements. This section is intended to give a qualitative comparison of the main results obtained from this analysis (LHCb) to the results from the CMS, ALICE and ATLAS collaborations.

The analysis strategies adopted by the different experiments are all comparable but not the same. As mention in Chap. 15.2, the experiments use different approaches for defining the activity of the recorded events. CMS utilises the track-multiplicity of reconstructed tracks in the fiducial region of their measurement. ALICE and ATLAS choose the hit-multiplicity of a forward oriented sub-detector, which has no overlap

## 16. Results for angular correlations

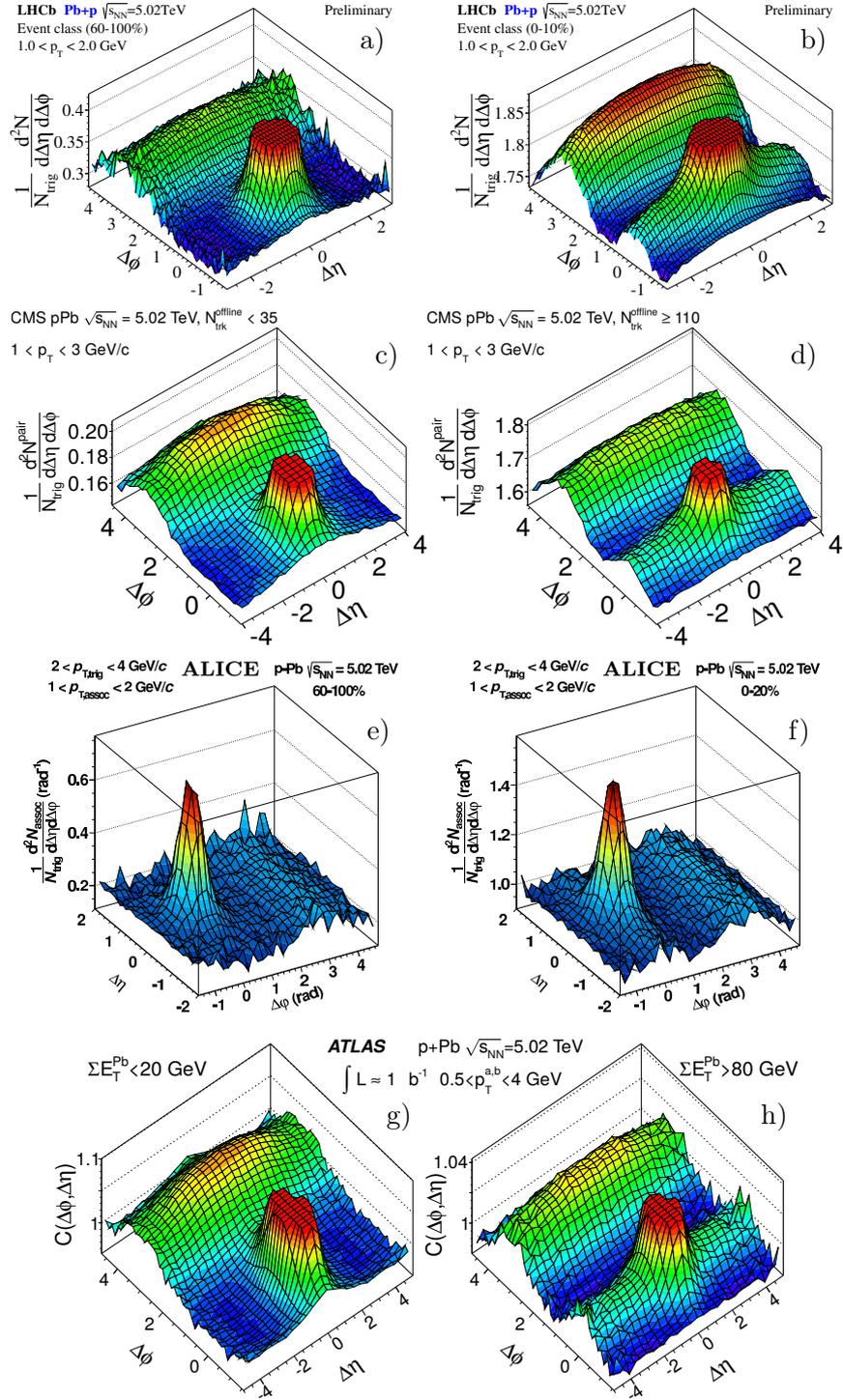


Figure 16.9.: Comparison of two-particle correlations in proton-ion collisions at  $\sqrt{s_{NN}} = 5.02\text{ TeV}$  among the four big LHC experiments. Results are presented for events with low (left) and high (right) event-activity, more details see text. LHCb results (a-b) in Pb+p configuration probing the forward region,  $\eta = [2.0, 4.9]$ ; CMS results [131] (c-d) probing the central region,  $|\eta| < [2.0]$ ; ALICE results [183] (e-f) probing the central region,  $|\eta| < [1.2]$ ; ATLAS results [182] (g-h) also probing the central region,  $|\eta| < [2.5]$ . Figures taken from respective reference.

with the pseudorapidity range accessed in their actual correlation measurement. As a reminder, the definition employed in LHCb is also based on hit-multiplicity. The utilised detector, the VELO, has by construction a large overlap with the pseudorapidity range of the correlation measurement. Since all approaches are very similar, the resulting event classes have a large overlap and lead to comparable results.

Furthermore, the  $p_T$ -ranges that are probed by the experiments are also different. CMS and ATLAS use a common  $p_T$ -range for trigger and associated particles, as it is adopted in the analysis presented here. The respective ranges are different, but all probe an intermediate  $p_T$ -range of a few GeV. ALICE uses exclusive  $p_T$ -ranges for trigger and associated particles but being in overlap with the probed ranges of the other three experiments.

The differences of the analyses, in particular the definitions of the correlation yield and the respective detector acceptances, do not allow a quantitative comparison of the results among the experiments. However, the qualitative statements of the results are comparable. It should be pointed out again, that ALICE, ATLAS and CMS all probe the central region, at a pseudorapidity of maximum  $|\eta| < 2.5$ . LHCb provides the only measurement in the forward region, in the pseudorapidity range of  $\eta = [2.0, 4.9]$ .

Figure 16.9 compares the  $(\Delta\eta, \Delta\phi)$ -correlations from all four experiments. The presented histograms are divided into results for events with low (left) and high event-activity (right).

The correlation function for low-activity events shows the same fundamental properties among all experiments. Dominant structures are the near-side peak and the away-side ridge. The ridges found in all measurements have a similar shape when comparing the overlapping  $\Delta\eta$ -ranges. Towards larger  $\Delta\eta$ -values, a decline of the yield is visible. Depending on the probed  $p_T$ -range also a relict of the Gaussian ridge is visible. The ALICE measurement shows no indication of this, since the correlations are calculated with respect to a trigger particle with higher  $p_T$ .

Focussing on the results for high-activity events, a prominent ridge on the near-side accompanies the away-side ridge and the jet-peak. The near-side ridge found in the forward region looks similar to the ridge seen in the central region. Due to the different definition of "high activity" the yields cannot be compared directly. However, the ridge in the forward region seen by LHCb seems to be at least similarly distinct compared to the ridge found in the central rapidity region seen by the other LHC experiments.



In the following chapter, systematic effects are discussed that can have an impact on the results presented in the previous chapter. For a better visualisation, the integrated  $\Delta\phi$ -correlations (c.f. Chap. 15.4) are presented to compare the size of systematic effects. The analysis method used to obtain the final correlation function is by construction robust against detector effects, such as the acceptance or mis-reconstruction. This fact is demonstrated in the following for several systematic effects.

First, the results of the data with different magnetic field polarities are compared to each other. Second, the impacts of reconstruction artefacts and inefficiencies to the correlation measurement are estimated. Third, the effect of using the alternative event-activity definition is evaluated by comparing results of the track-based classification to the nominal results based on hit-multiplicities. Afterwards, several aspects related to the track and event selection are discussed and the influence to the correlation results are evaluated.

### **17.1. Consistency check - magnetic field polarity**

The LHCb detector provides the possibility of reversing the direction of the magnetic field of the dipole magnet. There are no large systematic differences expected, since there is no different treatment for positively or negatively charged particles in this analysis. Furthermore, the difference of the interaction points as a result of the different field polarities is small compared to the normal variation from event to event. Thus, the data samples with both field polarities can be used as a simple cross-check of two independent samples.

The analysis is carried out identically for the magnet-up and magnet-down samples, only the applied correction factors were determined separately by using the simulation samples with the respective field configuration. Since the correction factors from the magnet-up and magnet-down simulation samples are in good agreement, no effect on the correlation function is expected.

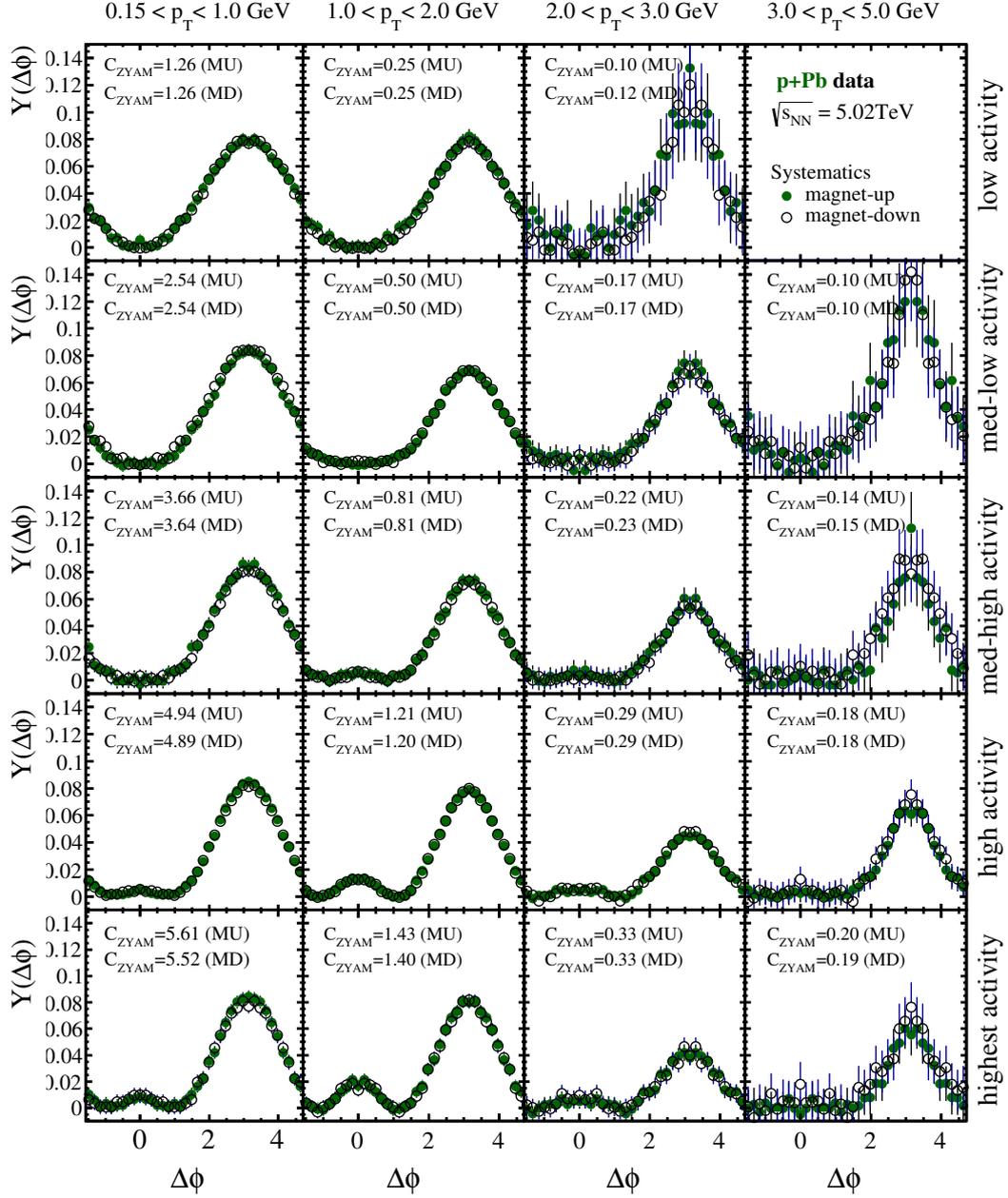


Figure 17.1.: One-dimensional correlation yield as a function of  $\Delta\phi$  for  $p$ +Pb collision at  $\sqrt{s_{NN}} = 5.02$  TeV, presented separately for magnet-up and magnet-down. The integration region is  $2 < \Delta\eta < 2.8$ , the final yield is subtracted by the constant listed in each panel using the ZYAM-method. All errors are statistical only.

The results obtained for the  $p$ +Pb data are presented in Fig. 17.1. Both data samples are in excellent agreement. The background correlation yield obtained by the ZYAM-method (c.f. Chap. 15.4), as well as the qualitative structures show no difference in behaviour over the entire measurement. The slight excess of the near-side ridge towards the highest-activity bin is similar in both samples.

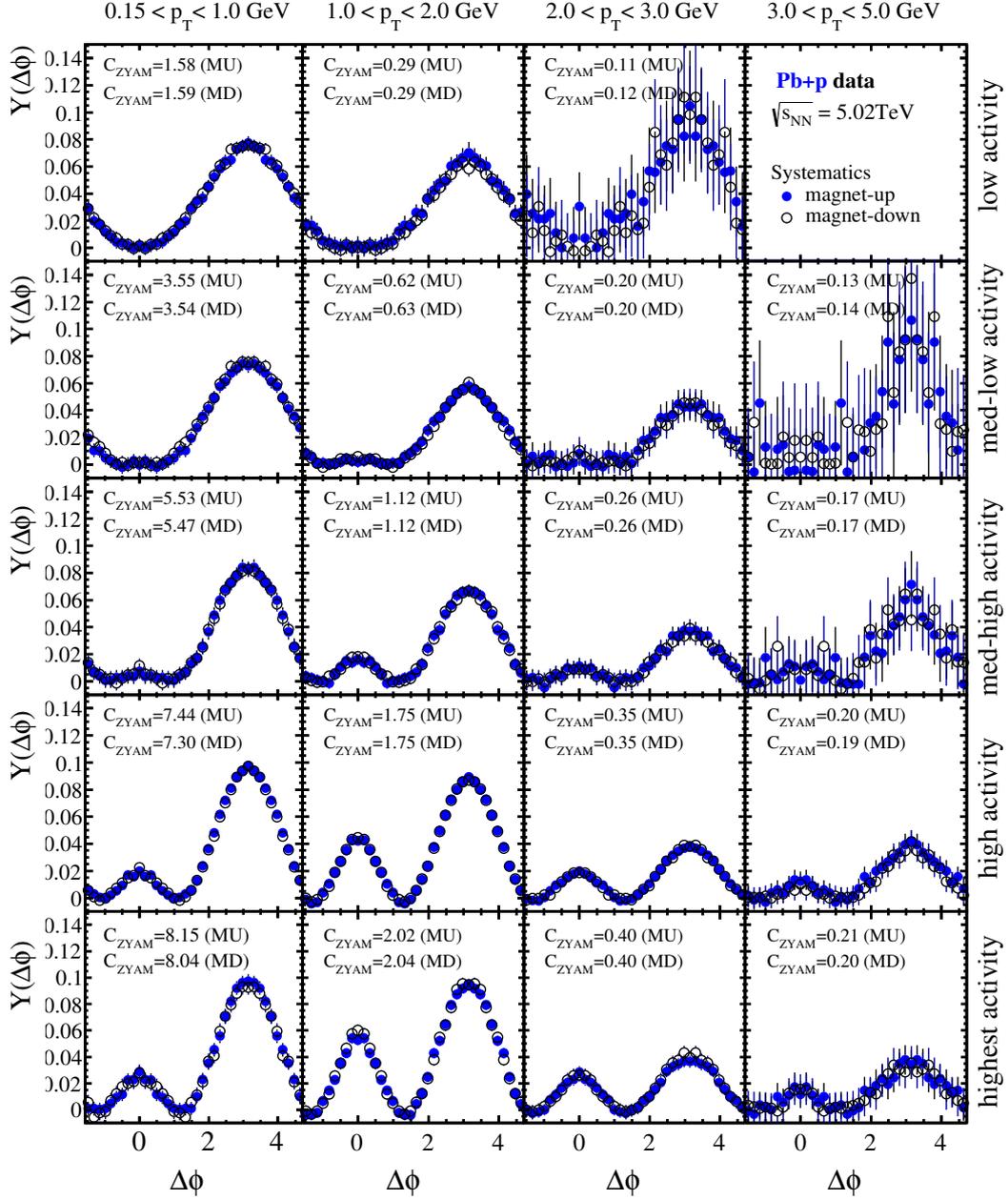


Figure 17.2.: One-dimensional correlation yield as a function of  $\Delta\phi$  for Pb+ $p$  collision at  $\sqrt{s_{NN}} = 5.02$  TeV, presented separately for magnet-up and magnet-down. The integration region is  $2 < \Delta\eta < 2.8$ , the final yield is subtracted by the constant listed in each panel using the ZYAM-method. All errors are statistical only.

The same picture appears in Pb+ $p$  data, the results are depicted in Fig. 17.2. The data samples with different magnetic field polarities give compatible results confirming the strong characteristics of the near-side ridge in events with high activity.

Due the similar behaviour of the magnet-up and magnet-down samples, both have been averaged for the final results presented in the previous chapter.

## 17.2. Alternative event-activity classification

In Chapter 15.2, it is discussed that there are several different approaches to classify the activity of an event. Two choices are realised: the nominal classification is based on the hit-multiplicity in the VELO, the alternative approach is based on track-multiplicity of reconstructed long tracks and is used as a cross-check. The results using both classifications are compared to each other in the following. For this, only results using the minimum-bias sample are presented. The high-activity sample is preselected according to hit-multiplicity and thus would bias the classification by track multiplicity. Furthermore, only magnet-down data are considered in this comparison.

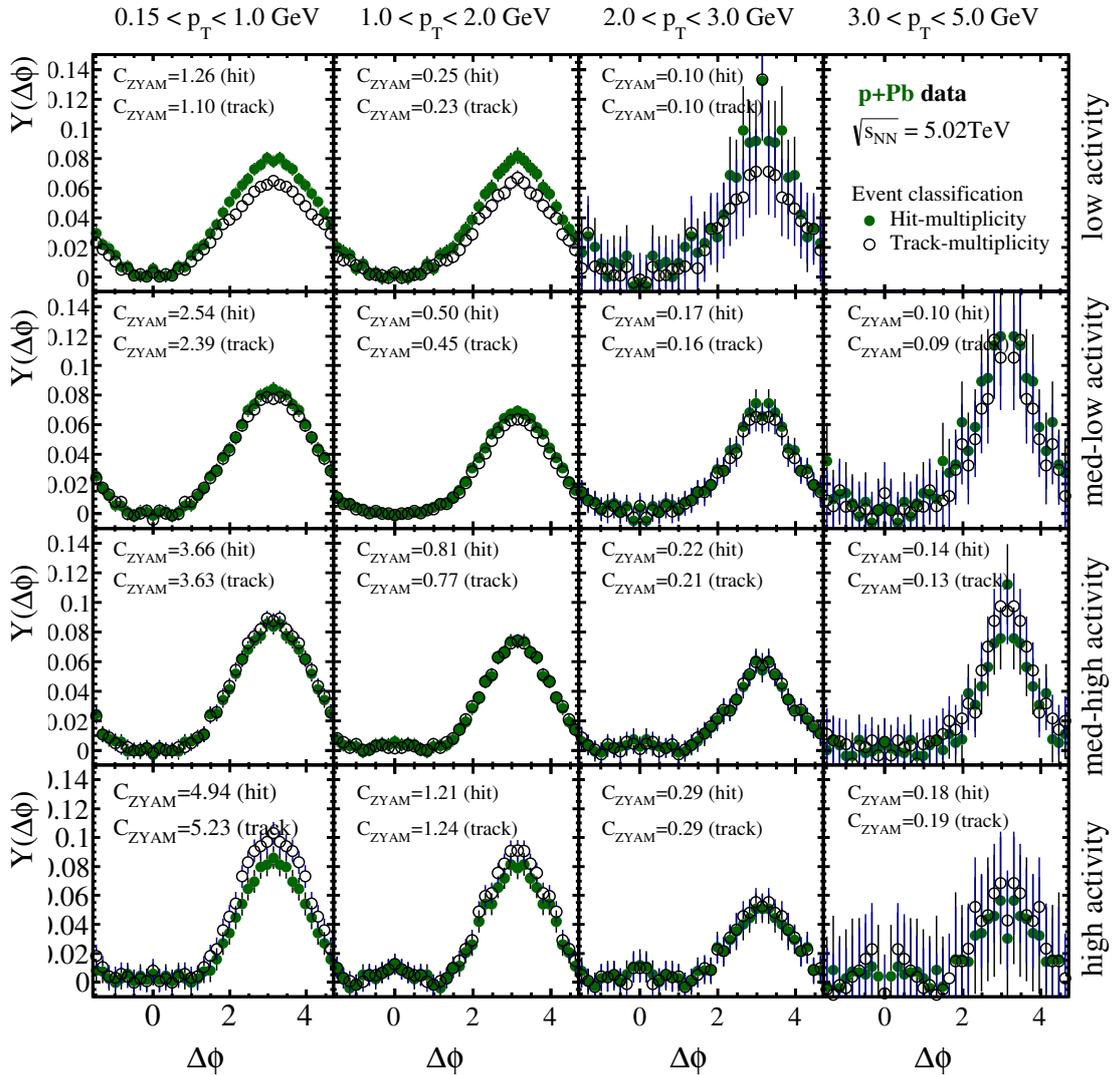


Figure 17.3.: One-dimensional correlation yield as a function of  $\Delta\phi$  for  $p+Pb$  collision at  $\sqrt{s_{NN}} = 5.02$  TeV, comparing results based on different event-activity classifications. The integration region is  $2 < \Delta\eta < 2.8$ , the final yield is subtracted by the constant listed in each panel using the ZYAM-method. All errors are statistical.

Due to the strong correlation between hit- and track-multiplicity, it can be expected that results of both event classifications are not fundamentally different. The direct comparison of both results is given in Fig. 17.3 for the  $p$ +Pb data. Both classifications are leading to similar results. Only in the low- and high-activity classes small quantitative differences of the yields are visible. Low-activity events using the hit-based approach show larger yields compared to that of the track-based approach. In the high-activity class, one can see a result to the contrary, where a larger yield in the track-multiplicity classification is found.

Looking at the results of the Pb+ $p$  data in Fig. 17.4 a similar situation is found. The low-activity class defined by using hit-multiplicity shows slightly larger yields compared

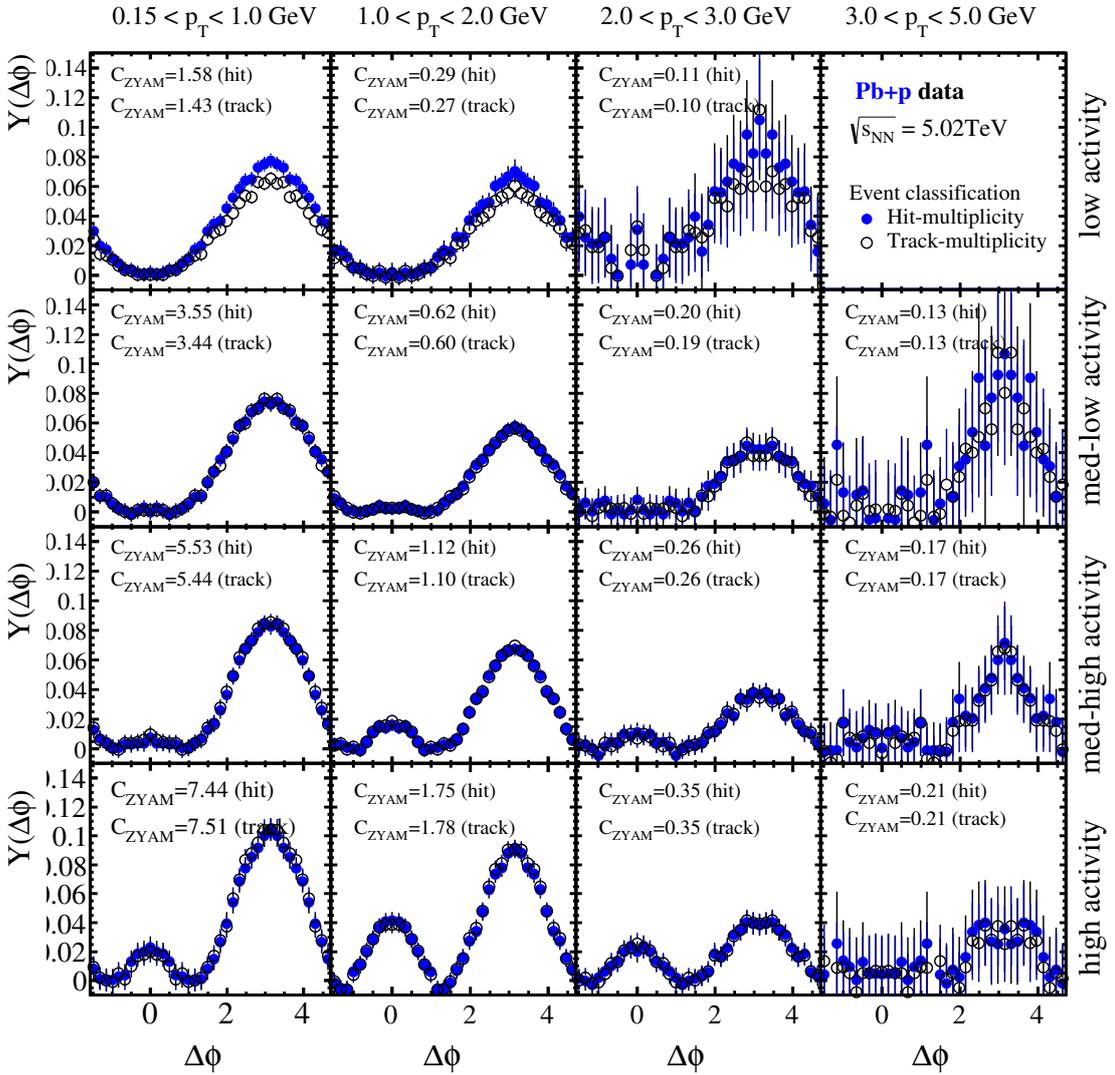


Figure 17.4.: One-dimensional correlation yield as a function of  $\Delta\phi$  for Pb+ $p$  collision at  $\sqrt{s_{NN}} = 5.02$  TeV, comparing results based on different event-activity classifications. The integration region is  $2 < \Delta\eta < 2.8$ , the final yield is subtracted by the constant listed in each panel using the ZYAM-method. All errors are statistical.

to that of the track-multiplicity approach. In the high-activity class, the differences are only marginal.

In summary, there is no difference in the qualitative properties of two-particle correlations found if using the track- or the hit-multiplicity approach to classify the event-activity. In particular, the yield of the near-side ridge, which is found in high-activity events, is compatible in both approaches.

It should be emphasised again that both ideas for an event-classification are equally right and valid, but probe slightly different properties of the events. Hit-multiplicity is the more robust quantity and is further not used for the actual correlation measurement. These are the main motivations, for the use of hit-multiplicity as the nominal classifier. However, the alternative track-based approach results in similar qualitative conclusions.

### 17.3. Influence of correction factors

Another systematic test estimates the impact of the correction factors to the final results, which are applied as weighting factors in the course of the analysis. It was already stated that the correlation function, based on a definition of dividing a signal by a background distribution, is robust against detector artefacts and inefficiencies. To visualise the impact of the correction factors, the one-dimensional correlation yield,  $Y(\Delta\phi)$ , is compared for results with either applying no corrections, applying only factors for fake and secondary tracks, or applying all correction factors including for inefficiencies. Figures 17.5 and 17.6 show the result of this comparison for  $p+Pb$  and  $Pb+p$  data, respectively. All event-activity classes based on the hit-multiplicity approach are presented, but only the minimum-bias data samples are considered.

First of all it is notable that there are no qualitative differences arising when applying any correction factors or not. In particular, the impact of fake and secondary corrections is only reflected by the constant yield, which is described by  $C_{ZYAM}$ . As expected, the track weighting according to the purity of the respective tracks decreases the yield. The zero-yield-at-minimum varies by 5 – 18% depending on the event-bin and the  $p_T$ -range. The remaining subtracted yield is almost insensitive to the weighting. There is no change in shape, only a tiny decrease of the away-side yield is found.

Studying the effect of the correction factors related to detector and reconstruction efficiencies shows a slightly different behaviour. By assigning the additional weighting factors to compensate for efficiencies, the resulting yield naturally increases. The constant yield is changed by approximately 40% at most, with respect to the yield without corrections. The structures in the displayed subtracted yield,  $Y(\Delta\phi)$ , are strengthened as being visible by the more pronounced away- and near-side ridges. It is important to point out that no additional structures are induced nor suppressed.

There is no change of the qualitative results obtained by applying the correction factors or not. The presence of the near-side ridge, which is arising for tracks in an intermediate  $p_T$ -range in high-activity  $pPb$  collisions, is unaffected. To the contrary, the quantitative yield is of course sensitive to these corrections. The difference of the ZYAM-subtracted yield is only moderate and is suggesting a very small final systematic

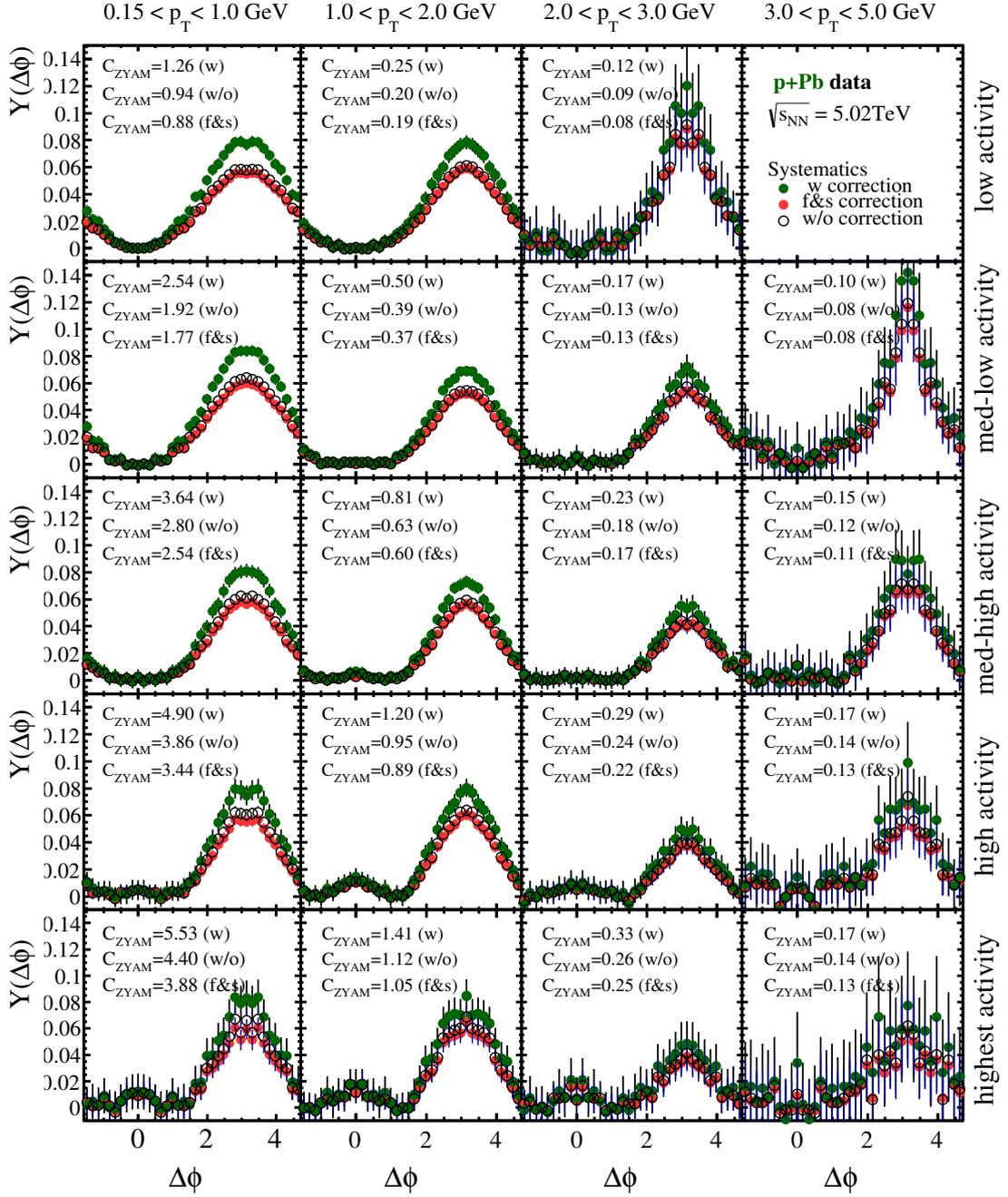


Figure 17.5.: One-dimensional correlation yield as a function of  $\Delta\phi$  for  $p$ +Pb collision at  $\sqrt{s_{NN}} = 5.02$  TeV. Results are compared for applying different correction factors: with all corrects, without any corrections and only corrections related to fake and secondary tracks. The integration region is  $2 < \Delta\eta < 2.8$ , the final yield is subtracted by the constant listed in each panel using the ZYAM-method. All given errors are pure statistical.

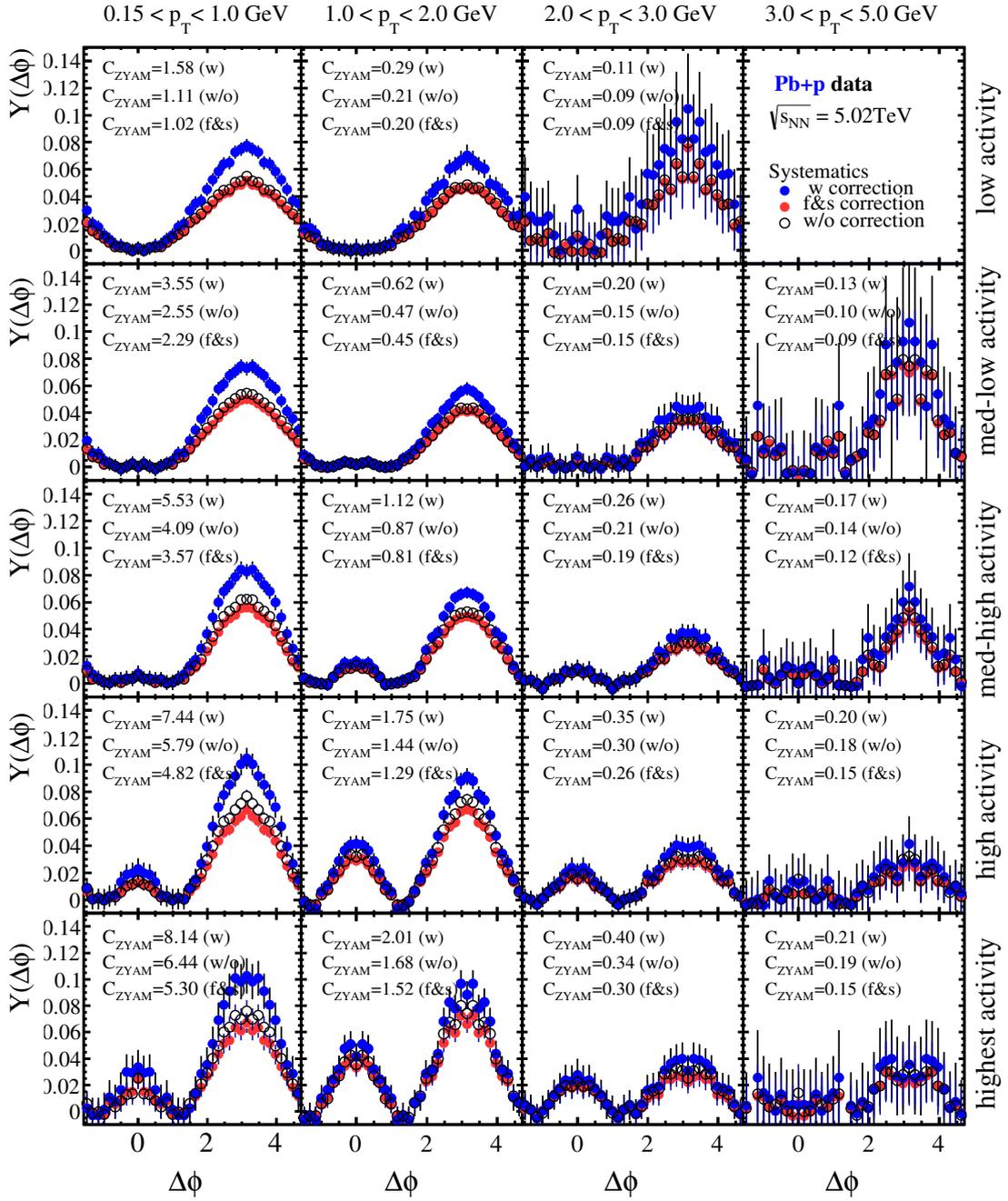


Figure 17.6.: One-dimensional correlation yield as a function of  $\Delta\phi$  for Pb+ $p$  collision at  $\sqrt{s_{NN}} = 5.02$  TeV. Results are compared for applying different correction factors: with all corrects, without any corrections and only corrections related to fake and secondary tracks. The integration region is  $2 < \Delta\eta < 2.8$ , the final yield is subtracted by the constant listed in each panel using the ZYAM-method. All given errors are pure statistical.

uncertainty. To evaluate this uncertainty, further studies are required which should comprise a different Monte Carlo simulation in order to cross-check the correction factors obtained from the HIJING simulation. Since the differences between results with and without any corrections are moderate, no large final systematic uncertainty is expected. The qualitative statement of this current analysis is unaffected.

## 17.4. Uncertainties related to multiple PVs

An important requirement of the event selection in Chap. 14.3 is that only events containing exactly one reconstructed primary vertex are selected. The PV requirement suppresses triggered background events in which no proton-ion collision were taking place. Since the efficiency of the vertex reconstruction is determined to be effectively 100% for  $p$ Pb collisions, systematic effect due to undetected  $p$ Pb collisions is negligible. However, the strict requirement of selecting only events with exactly one reconstructed PV can also bias the results, because events with multiple PVs are rejected.

Even though the maximum pile-up probability during the proton-ion data taking was less than 0.8‰, events with several reconstructed PVs have been recorded. Besides pile-up  $p$ Pb collisions also additional background interactions, *e.g.* beam-gas interactions, or an incorrect event reconstruction can cause multiple PVs in a single event. The PV algorithm is optimised to reconstruct  $pp$  collisions with a comparatively small number of prompt particles w.r.t. to  $p$ Pb collisions. In high-multiplicity events, it is possible that a single vertex is reconstructed as multiple different vertices which are close together in space. These kind of event are also rejected during the event selection, together with other background events that are unrelated to actual  $p$ Pb interactions.

In order to estimate potential systematic biases to the results of the correlation measurement, the analysis procedure is repeated with a slightly modified event selection. Instead of requiring exactly one primary vertex, all events containing at least one PV are selected. The resulting one-dimensional yields found in the  $p$ +Pb and Pb+ $p$  data samples for both event selections are compared in Fig. 17.7 and Fig. 17.8, respectively. The minimum-bias and high-occupancy samples are considered but only using the magnet-up configuration.

The obtained result in the  $p$ +Pb and Pb+ $p$  samples are almost identical for both event selections. The low- and medium-activity classes are completely unaffected by allowing additional events with more than one reconstructed PV. The high-activity events show a minimal variation of the quantified yields, but the effect is negligible.

It can be concluded that the systematic uncertainty due to additional reconstructed primary vertices, either due to background interactions or due to mis-reconstructed  $p$ Pb collisions, have no impact on the final results.

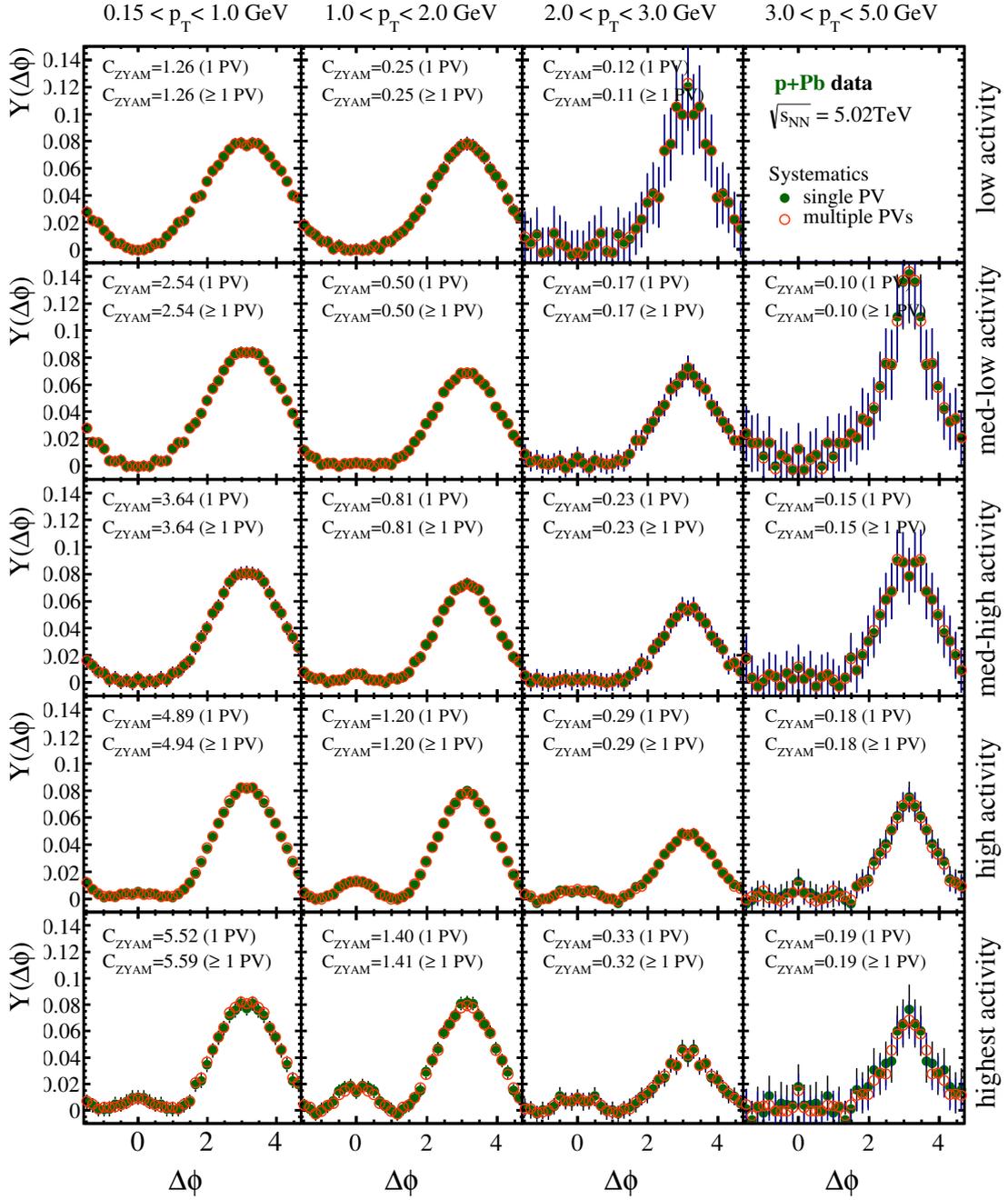


Figure 17.7.: One-dimensional correlation yield as a function of  $\Delta\phi$  for  $p$ +Pb collision at  $\sqrt{s_{NN}} = 5.02$  TeV. Results are compared for selected events containing exactly 1 PV (nominal) and events containing  $\geq 1$  PVs. The integration region is  $2 < \Delta\eta < 2.8$ , the final yield is subtracted by the constant listed in each panel using the ZYAM-method. All given errors are pure statistical.

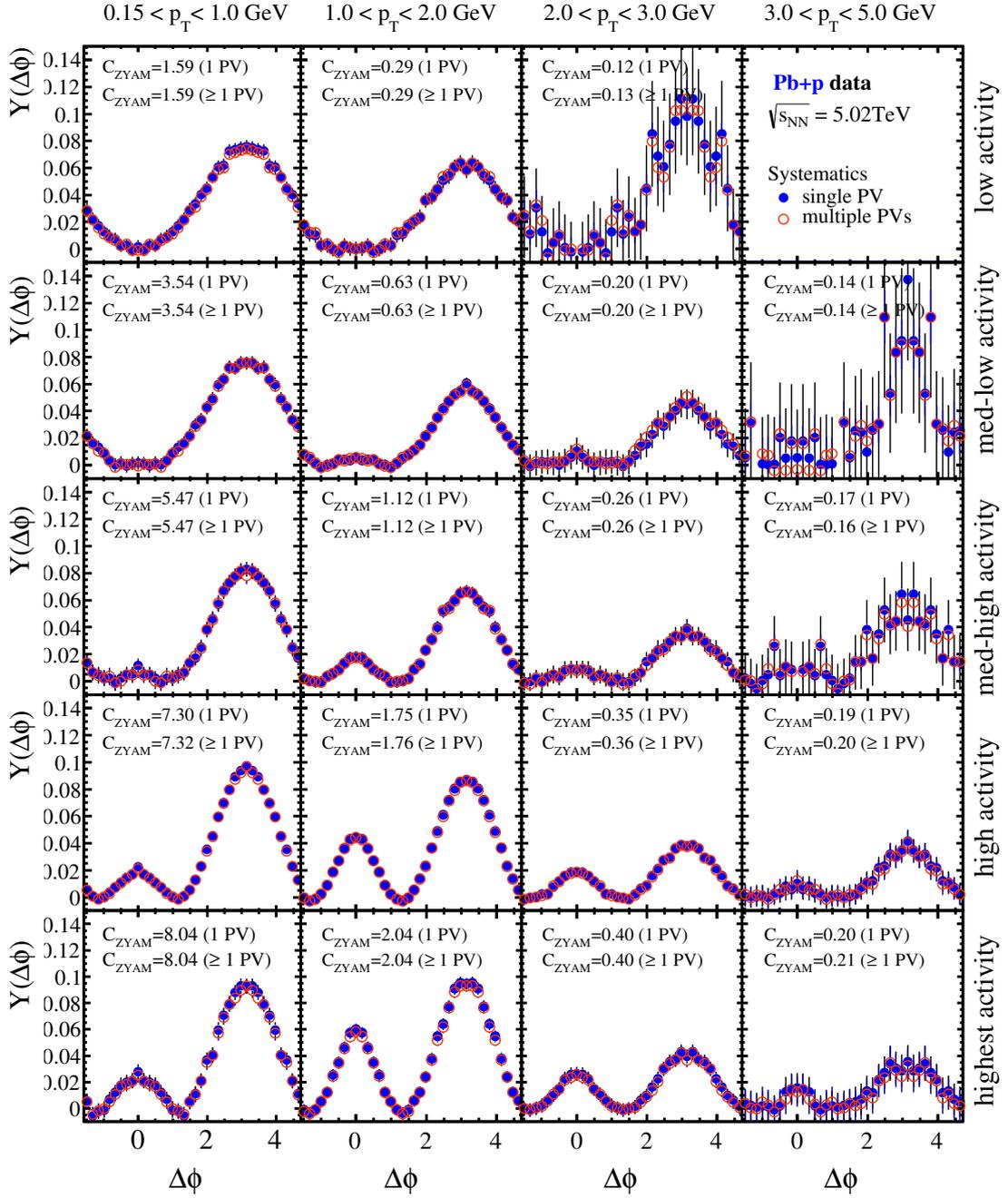


Figure 17.8.: One-dimensional correlation yield as a function of  $\Delta\phi$  for Pb+p collision at  $\sqrt{s_{NN}} = 5.02$  TeV. Results are compared for selected events containing exactly 1 PV (nominal) and events containing  $\geq 1$  PVs. The integration region is  $2 < \Delta\eta < 2.8$ , the final yield is subtracted by the constant listed in each panel using the ZYAM-method. All given errors are pure statistical.

## 17.5. Uncertainties related to track selection

The measured two-particle correlations in this analysis are performed on samples of prompt charged particles. Consequently, the obtained results depend on the purity of the track samples.

In the track selection, Chap. 14.5, prompt particles are selected by a condition set on the impact parameter of a track w.r.t the primary vertex. The default condition requires an  $IP_{PV} < 1.2$  mm to select a track. The remaining contamination of tracks from secondary particles is accounted for by applying weighting factors, which are determined in simulation where the identical condition is set. As a result, potential discrepancies between data and simulation will affect the correlation measurement. In order to estimate the size of this effect, the IP requirement in simulation is changed to  $IP_{PV} < 2.4$  mm and the resulting correction factors are applied to data for which the nominal IP requirement is retained.

Another crucial point in the track selection is the quality requirement used to suppress fake tracks. As for non-prompt tracks, also the remaining fake tracks in the track sample are accounted for by weighting factors which are determined in simulation. The nominal criterion requires a ghost probability of less than 0.3. In order to estimate uncertainties related to differences between data and simulation, the nominal value is changed to  $GP < 0.25$  and to  $GP < 0.35$  in the simulation. The resulting new correction factors are applied to data, for which again the default selection is retained.

The impact of the modified correction factors to the correlation measurement is presented by showing the integrated correlation yield  $Y(\Delta\phi)$  in Fig. 17.9 for the  $p+Pb$  data, and in Fig. 17.10 for  $Pb+p$  data. The results for the nominal selection compared to the results using the modified impact-parameter and ghost-probability criteria are fully compatible. Differences between the distributions are smaller than the used marker sizes and are qualitatively identical. The change of the impact parameter selection has no significant effect to the measurement.

A small difference is observed in the subtracted constant yield ( $C_{ZYAM}$ ), when comparing the results using different ghost probabilities. Requiring a stricter ghost probability ( $GP < 0.25$ ) in the simulation results in a larger yield in data, because the amount of fake tracks is under-corrected. For a less strict requirement the effect is vice versa.

The changes in the selection have no impact on the qualitative results of the measurement. In particular the observed shapes of the distribution are insensitive to a different amount of fake tracks in the track sample.

The systematic checks in this chapter show that the analysis procedure is very robust against various kinds of backgrounds. Varying the amount of contaminating background tracks has effectively no influence on the final results. It is even shown, that the analysis carried out without any correction factors still gives the same qualitative results. The relative structures become slightly damped being the result of a limited detector and reconstruction efficiency, but remain unchanged in shape. Further including events with multiple PVs, which mainly adds background interactions to the event sample, shows also no sizeable effect on the measurement. An alternative approach to identify events with low and high activity produces the same qualitative results, also with a very similar quantitative statement.

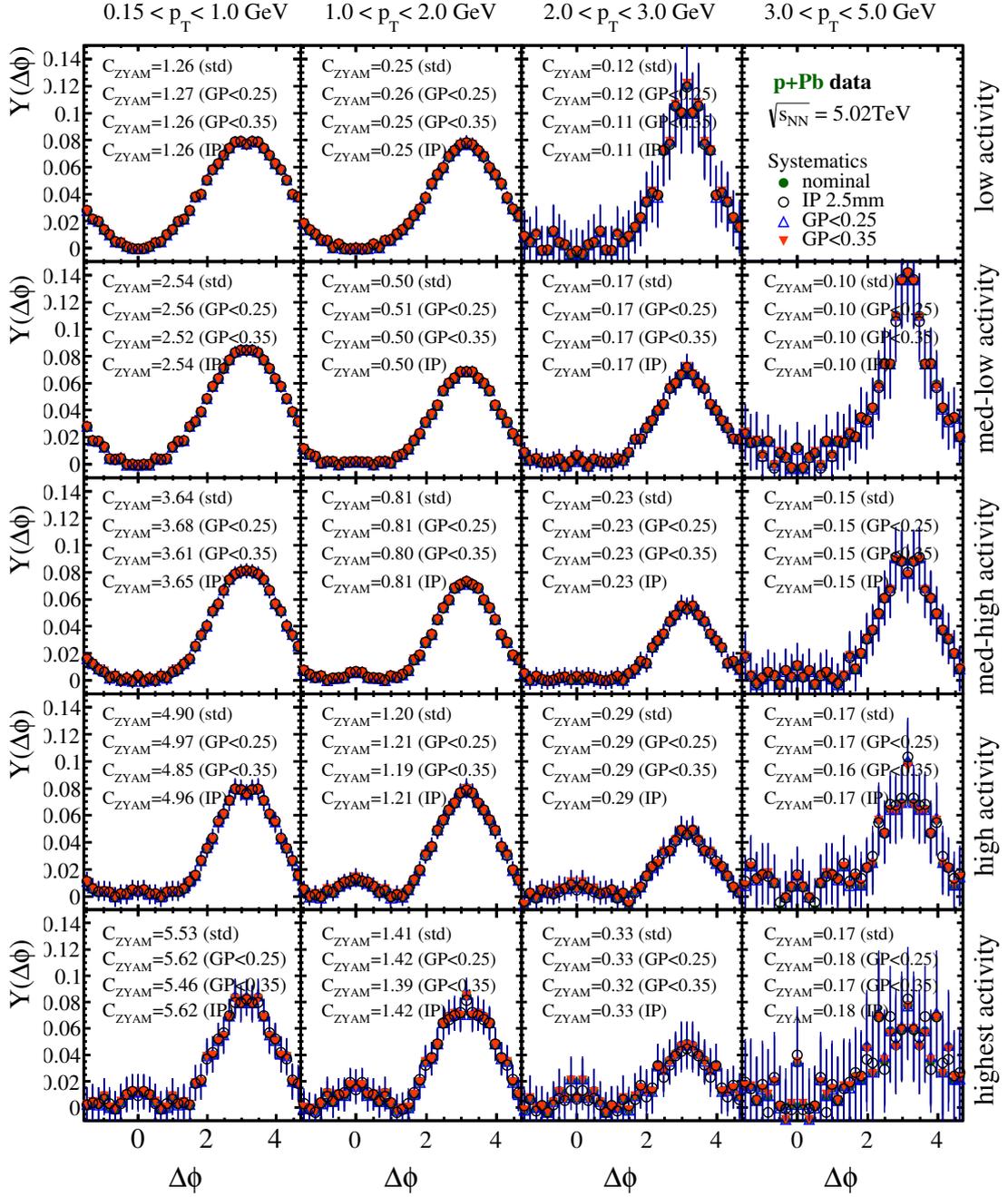


Figure 17.9.: One-dimensional correlation yield as a function of  $\Delta\phi$  for  $p$ +Pb collision at  $\sqrt{s_{NN}} = 5.02$  TeV. Results are compared for correction factors with different selection criteria: nominal selection, modified impact parameter of  $IP_{PV} < 2.4$  mm, modified ghost probabilities of  $GP < 0.25$  and  $GP < 0.35$ . The integration region is  $2 < \Delta\eta < 2.8$ , the final yield is subtracted by the constant listed in each panel using the ZYAM-method. All given errors are pure statistical.

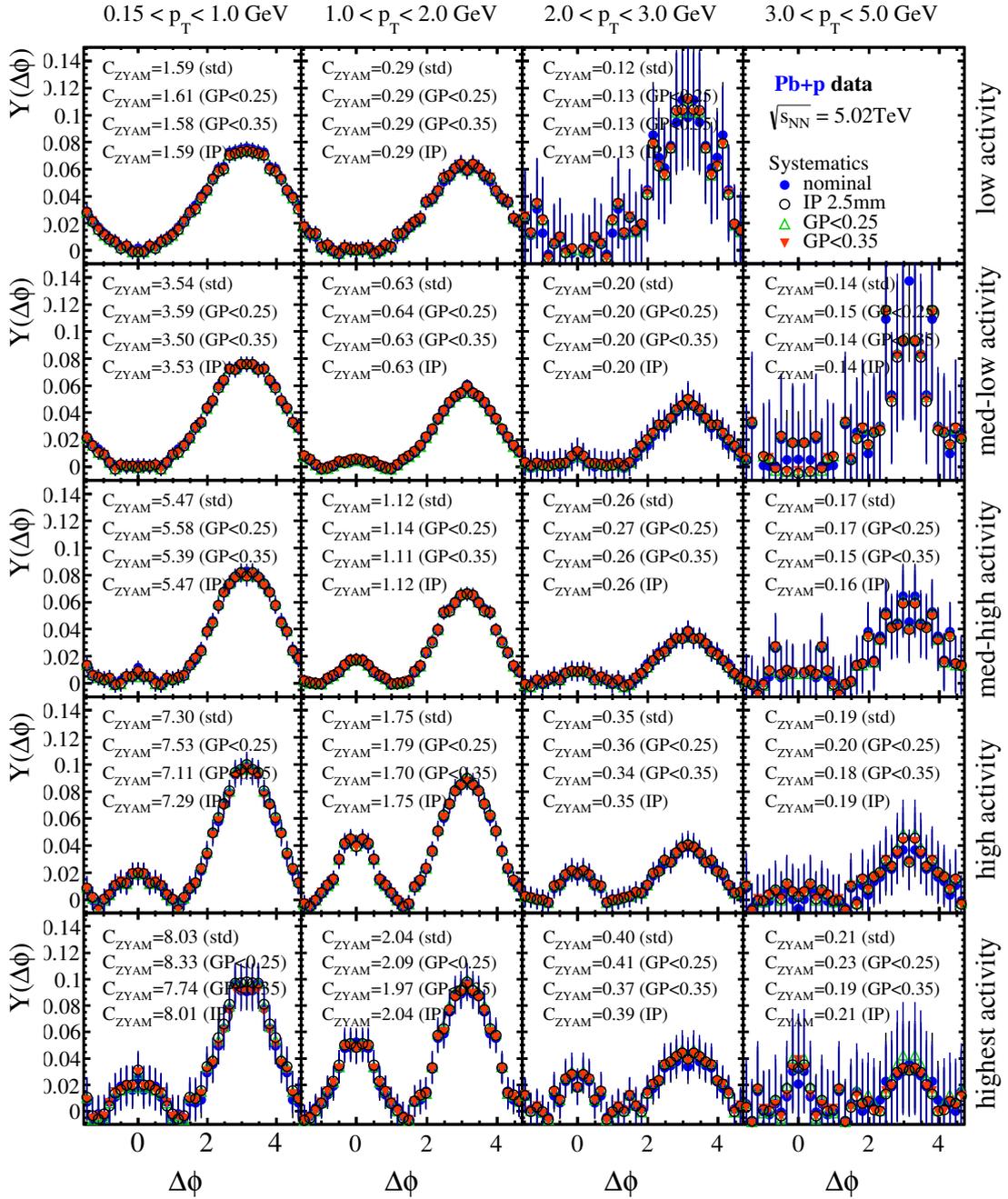


Figure 17.10.: One-dimensional correlation yield as a function of  $\Delta\phi$  for Pb+p collision at  $\sqrt{s_{NN}} = 5.02$  TeV. Results are compared for correction factors with different selection criteria: nominal selection, modified impact parameter of  $IP_{PV} < 2.4$  mm, modified ghost probabilities of  $GP < 0.25$  and  $GP < 0.35$ . The integration region is  $2 < \Delta\eta < 2.8$ , the final yield is subtracted by the constant listed in each panel using the ZYAM-method. All given errors are pure statistical.

Measurements of particle production in  $pp$  and two-particle correlations in  $p\text{Pb}$  collisions are presented in this thesis, by using data collected by the LHCb experiment.

In the first analysis, particle multiplicities and particle densities are measured for prompt charged particles produced in  $pp$  collisions at a centre-of-mass energy of  $\sqrt{s} = 7 \text{ TeV}$ . With the LHCb forward spectrometer a kinematic range of  $2.0 < \eta < 4.8$  in pseudorapidity, momenta greater than  $2 \text{ GeV}$  and transverse momenta greater than  $0.2 \text{ GeV}$  are accessed in this measurement. The analysis is performed for events with at least one charged particle in this kinematic acceptance.

The prediction of the soft component in hadronic interactions, the so-called underlying event, is important for many kinds of physics analyses. Since soft-QCD processes such as light particle production cannot be calculated perturbatively, they are described by models implemented in Monte Carlo event generators. The modelling of the fragmentation and hadronisation processes are treated differently in the various event generators. These models contain several parameters which need to be optimised according to the collision energy and the colliding particle species. The tuning process is usually performed using measurements of soft-QCD processes. Among others [186], the measurements of charged particle multiplicities and densities provide a fundamental input for event generator optimisations.

The main challenge of this analysis is to obtain a clean track sample that precisely reflects the number of prompt charged particles per event. The track-multiplicity of an event, which is a simple counting observable, is biased by two contrary effects. As a result of the LHCb detector design, which minimises the amount of material in the tracking devices, the measured tracks exhibit a large fraction of wrongly reconstructed tracks. Inefficiencies due to the detector acceptance and the track reconstruction prevent a large fraction of low-momentum particles from being detected. To account for these effects a precise quantification and control of these inefficiencies is required. Since single tracks in an event cannot be identified, a statistical correction of the track sample is performed. For the measurement of particle densities, it is sufficient to apply simple

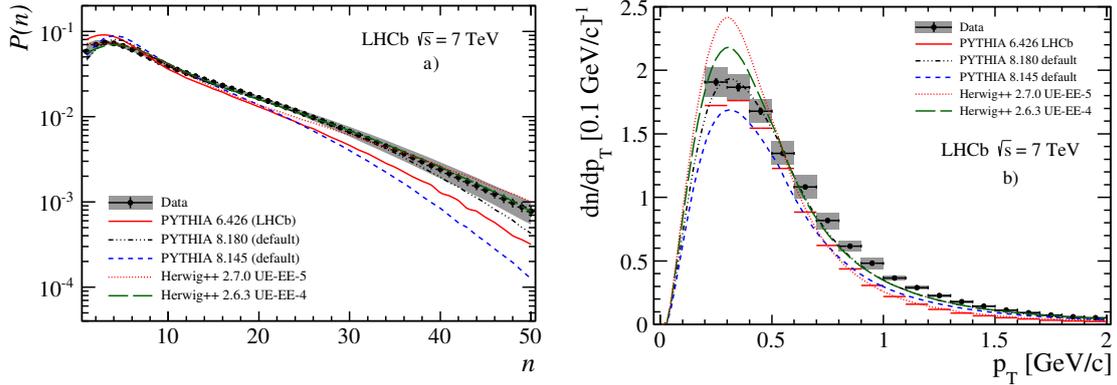


Figure 18.1.: Measured charged particle multiplicity in the full kinematic range (a) and charged particle density  $dn/dp_T$  as a function of transverse momentum (b) in  $pp$  collisions at  $\sqrt{s} = 7$  TeV. The LHCb data are shown as points with statistical error bars and combined systematic and statistical uncertainties as the grey band. Measurements are compared to predictions of several Monte Carlo event generators.

average correction factors. The particle multiplicity determination requires a specific two-step approach. After accounting for the number of contaminating tracks per event, the physical particle multiplicity distribution is obtained by using an unfolding technique to account for reconstruction and detector inefficiencies.

The analysis is performed by using a visible event definition that allows a direct comparison of the measurement with Monte Carlo generator predictions, without requiring a full detector simulation. The measured charged particle multiplicity distribution in the kinematic range of the analysis is depicted in Fig. 18.1 a). The corresponding mean value,  $\mu$ , and the root-mean-square deviation,  $\sigma$ , truncated in the range from 1 to 50 particles, are determined to be

$$\mu = 11.30 \pm 0.01 \pm 0.09 \quad \text{and} \quad \sigma = 9.50 \pm 0.01 \pm 0.02, \quad (18.1)$$

where the uncertainties are statistical and systematic, respectively. All measurements are compared to predictions from different event generators and tunes. Older event generators (PHOJET and PYTHIA 6.4) that are not optimised to LHC data, significantly underestimate the particle production. Predictions from recent event generators (PYTHIA 8.1 and HERWIG++) tuned to LHC data obtained in the central rapidity region are also in reasonable good agreement with the LHCb data in the forward region, as shown in Fig. 18.1 a).

In addition to the full kinematic range, particle multiplicities are measured in bins of  $\eta$  and  $p_T$ . Depending on the respective generator tune, the single  $\eta$ ,  $p_T$  bins can be described by the simulation, however, there is no model that can reproduce all bins together. Observed discrepancies are particularly large in the low- $p_T$  range.

A qualitatively similar statement can be drawn from the measured charged particle densities,  $dn/d\eta$  and  $dn/dp_T$ , where the latter measurement is depicted in Fig. 18.1 b). Recent versions of PYTHIA 8.1 and HERWIG++ show in general a reasonable good agreement with data, but the results are strongly depending on the respective tuning of the generator.

---

Among the LHC, this analysis represents the only measurement in the forward region that provides momentum dependent results of charged particle multiplicities and densities. The measurement complements other results from the ALICE, ATLAS and CMS experiments performed in the central rapidity region, as well as a momentum inclusive particle density measurement by the TOTEM experiment, in which even larger pseudorapidities are accessed ( $5.3 < \eta < 6.4$ ).

The presented measurement is a reference for future generator optimisations. All results are published as full data tables. Furthermore, the analysis is provided in form of a generator plug-in, which includes an implementation of the analysis selection on generator level, together with data results.

Within the LHCb collaboration, the measurement is used for future generator tuning, in order to replace the current LHCb tune of PYTHIA, which is found to significantly underestimate charged particle production in  $pp$  collisions at  $\sqrt{s} = 7$  and 8 TeV. In flavour physics analyses, the quantitative modelling of charged particles as a background in the event has a significant impact on the performance to identify quark flavours. For the upcoming  $pp$  data taking period in 2015 at  $\sqrt{s} = 13$  TeV, the tuning process of the simulation is currently being prepared.

Further multiplicity analyses with LHCb data are possible and planned for the future. There are existing data samples of  $pp$  collisions at centre-of-mass energies of 2.76 TeV and 8 TeV, which are suitable for similar particle multiplicity measurements. In combination with a measurement at  $\sqrt{s} = 13$  TeV, which is planned with the first data collected in 2015, it is possible to study the energy-scaling of charged particle production over almost one order of magnitude in  $\sqrt{s}$ . In addition to results for inclusive charged particles, identified particles (in particular  $p$ ,  $\pi$  and  $K$ ) are interesting to study, since they provide a more specific test for hadronisation models. Thanks to an excellent particle identification system, LHCb provides a good environment for these types of measurements.

With the successful data taking of proton-ion collisions, particle multiplicity and density measurements can be extended to this mixed system. LHCb has the unique possibility to probe particle densities in  $p$ Pb collisions over a large range in pseudorapidity of the nucleon-nucleon centre-of-mass system of  $1.5 < \eta_{cms} < 4.5$  and  $2.5 < \eta_{cms} < 5.5$  for the  $p$ +Pb and Pb+ $p$  configuration, respectively.

The second analysis presented in this thesis is based on proton-ion collisions, recorded at nucleon-nucleon centre-of-mass energy of  $\sqrt{s_{NN}} = 5.02$  TeV. Two-particle angular correlations are studied for inclusive charged particles that are directly produced in the collision. This analysis is motivated by the striking observation of a near-side ridge structure in high-activity  $pp$  and  $p$ Pb collisions, which was previously known only from heavy-ion collisions. From the theory point-of-view, this correlation structure in heavy-ion collisions is qualitatively well described by several models. However, its origin is not yet understood from first principles. In all models describing the ridge, the strongly interacting medium created in a heavy-ion collision plays an important role. In these models, the ridge is attributed to interactions between jets and this medium, or to the medium itself. Strong evidence for the creation of such a medium is the observation of an anisotropy in the momentum distribution of the particles. The hydrodynamic flow in the medium can be decomposed into discrete Fourier harmonics, whose components,

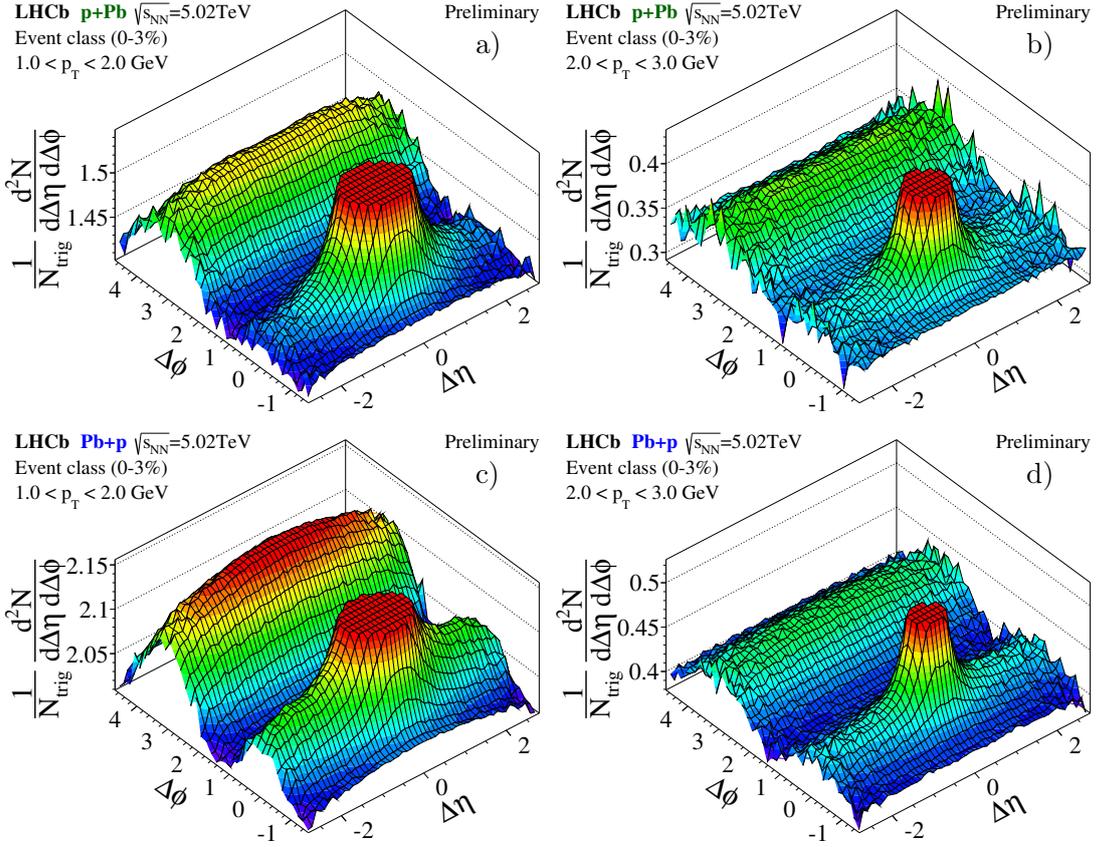


Figure 18.2.: The near-side ridge ( $\Delta\phi = 0$ ) measured in in the forward region of  $\eta = [2.0, 4.9]$  for proton-ion collisions at  $\sqrt{s_{NN}} = 5.02$  TeV. Results in  $p+\text{Pb}$  (a,b) and  $p+\text{Pb}$  (c,d) collisions are presented for events with high event-activity. Particles in  $p_T$ -ranges of  $1 < p_T < 2$  GeV (a,c) and  $2 < p_T < 3$  GeV (b,d) are considered.

*e.g.* related to elliptic and triangular flow, have been measured.

The observed ridge phenomenon in high-activity  $pp$  and  $p\text{Pb}$  collisions cannot be explained by the same theoretical considerations, since the formation of a strongly interacting medium as in  $\text{PbPb}$  collisions is not expected. A variety of models have been proposed to explain the near-side ridge in the  $pp$  and  $p\text{Pb}$  system. Among several hydrodynamical approaches, similar to those applied in heavy-ion collision, models describing the ridge without a flow driven mechanism are proposed.

Since the ridge has been only observed in the central pseudorapidity region, up to  $|\eta| < 2.5$ , the presented analysis is intended to verify this correlation structure additionally in the forward region, for  $2.0 < \eta < 4.9$ . Using LHCb data with two different beam configurations, the near-side ridge is clearly observed in  $p+\text{Pb}$  and  $\text{Pb}+p$  collisions for events with high activity and particles in an intermediate  $p_T$ -range, see Fig. 18.2. While the ridge structure is only slightly visible in the  $p+\text{Pb}$  data (a-b), a very distinct ridge is found in  $\text{Pb}+p$  data. The difference of the ridge strength is primarily related to the respective activity in the events, in this analysis described by the hit-

---

multiplicity in the Vertex Locator. In Pb+ $p$  collisions, larger absolute hit-multiplicities are reached compared to  $p$ +Pb collisions. By selecting events with the 3% highest hit-multiplicity in each respective data sample, larger absolute hit-multiplicities are probed in the Pb+ $p$  data sample resulting in a more pronounced ridge.

Comparing the result for  $p$ +Pb and Pb+ $p$  collisions among all activity classes, indicates a common turn-on point in hit-multiplicity for the near-side ridge correlation. Further studies are ongoing, which probe common activity ranges for  $p$ +Pb and Pb+ $p$  collisions.

As observed at central rapidities, the formation of the ridge in the forward region shows a strong dependence on the  $p_T$ -range of the considered particles. In a range of  $1 < p_T < 2$  GeV, the ridge is most pronounced in both data samples (a,c) and also remains present towards larger  $p_T$  ( $2 < p_T < 3$  GeV), as shown in (b,d). A significant decrease of the ridge is observed towards increasing  $p_T$ . Since only a few theoretical models predict that the ridge remains at high transverse momenta,  $\mathcal{O}(10$  GeV), further investigations in this direction are suggested.

This measurement proves the feasibility of two-particle correlation measurements at LHCb. Thus, a similar analysis with  $pp$  collisions is suggested, in order to confirm the ridge observation in high-activity  $pp$  collisions.

Future two-particle correlation analyses in proton-ion collisions should include the measurement of the Fourier coefficients, describing the long-range correlation structures, since different values are predicted by the various models. In addition, similar two-particle correlation analyses of identified particle spectra are a good probe to support the idea of the formation of a dense, highly interacting system which can be responsible for the hydrodynamical flow. Depending on the masses of the particles, a different  $p_T$ -dependence is expected.

Both measurements presented in this thesis are unique results, since an exclusive kinematic range at the LHC is probed with the LHCb detector. The results are complementary to that of other LHC experiments and give input to the theory community and can be used to probe existing theory models. Also for the future, the LHCb experiment offers excellent opportunities to contribute to the understanding of soft-QCD physics in  $pp$  collisions and to the study of proton-ion collisions at the LHC.



- [1] S. Glashow, *Partial Symmetries of Weak Interactions*, Nucl.Phys. **22** (1961) 579–588.
- [2] A. Salam and J. Ward, *Electromagnetic and weak interactions*, Physics Letters **13** (1964), no. 2 168 – 171.
- [3] S. Weinberg, *A Model of Leptons*, Phys.Rev.Lett. **19** (1967) 1264–1266.
- [4] ATLAS Collaboration, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys.Lett. **B716** (2012) 1–29, [arXiv:1207.7214](#).
- [5] CMS Collaboration, S. Chatrchyan *et al.*, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys.Lett. **B716** (2012) 30–61, [arXiv:1207.7235](#).
- [6] Planck Collaboration, P. Ade *et al.*, *Planck 2013 results. I. Overview of products and scientific results*, [arXiv:1303.5062](#).
- [7] CMS Collaboration, V. Khachatryan *et al.*, *Observation of Long-Range Near-Side Angular Correlations in Proton-Proton Collisions at the LHC*, JHEP **1009** (2010) 091, [arXiv:1009.4122](#).
- [8] CMS Collaboration, S. Chatrchyan *et al.*, *Observation of long-range near-side angular correlations in proton-lead collisions at the LHC*, Phys.Lett. **B718** (2013) 795–814, [arXiv:1210.5482](#).
- [9] LHCb collaboration, R. Aaij *et al.*, *Measurement of charged particle multiplicities and densities in pp collisions at  $\sqrt{s} = 7$  TeV in the forward region*, Eur.Phys.J. **C74** (2014) 2888, [arXiv:1402.4430](#).
- [10] M. Meissner, T. Ruf, and U. Uwer, *Measurement of charged particle multiplicities and densities in pp collisions using "forward tracks"*, LHCb-ANA-2011-084 (Oct, 2011).

- [11] M. Meissner, *Two-particle angular correlations in proton-ion collisions at  $\sqrt{s_{NN}} = 5.02$  TeV*, LHCb-ANA-2014-091 (Oct, 2014).
- [12] G. Altarelli, *The Standard model of particle physics*, arXiv:hep-ph/0510281.
- [13] G. Altarelli, *Collider Physics within the Standard Model: a Primer*, arXiv:1303.2842.
- [14] I. C. Brock and T. Schorner-Sadenius, *Physics at the Terascale*. 2011.
- [15] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.
- [16] Particle Data Group, K. A. Olive *et al.*, *Review of Particle Physics (RPP)*, Chin.Phys. **C38** (2014) 090001.
- [17] LHCb collaboration, R. Aaij *et al.*, *Observation of the resonant character of the  $Z(4430)^-$  state*, Phys.Rev.Lett. **112** (2014) 222002, arXiv:1404.1903.
- [18] R. Placakyte, *Parton Distribution Functions*, arXiv:1111.5452.
- [19] James Stirling, Imperial College London, *Parton luminosity and cross section plots*, <http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html>.
- [20] G. Antchev *et al.*, *First measurement of the total proton-proton cross section at the LHC energy of  $\sqrt{s} = 7$  TeV*, Europhys.Lett. **96** (2011) 21002, arXiv:1110.1395.
- [21] V. Barone, *High-energy particle diffraction*. 2002.
- [22] L. Harland-Lang, V. Khoze, M. Ryskin, and W. Stirling, *Standard candle central exclusive processes at the Tevatron and LHC*, Eur.Phys.J. **C69** (2010) 179–199, arXiv:1005.0695.
- [23] E. Nurse and S. Sen, *Methods to Select Soft Diffraction Dissociation at the LHC*, arXiv:1107.2688.
- [24] Chris Blanks, *Article: Strangeness at LHCb*, <http://www.imperialhep.blogspot.de/2011/08/strangeness-at-lhcb.html>.
- [25] E. Norrbin and T. Sjostrand, *Production and hadronization of heavy quarks*, Eur.Phys.J. **C17** (2000) 137–161, arXiv:hep-ph/0005110.
- [26] C. Elsasser (Universität Zürich), *Feynman Diagram Library (FDL)*, <http://www.physik.uzh.ch/~che/FeynDiag/>. (CC BY-SA license).
- [27] P. Nason *et al.*, *Bottom production*, arXiv:hep-ph/0003142.
- [28] LHCb Collaboration, R. Aaij *et al.*, *Measurement of  $\sigma(pp \rightarrow b\bar{b}X)$  at  $\sqrt{s} = 7$  TeV in the forward region*, Phys.Lett. **B694** (2010) 209–216, arXiv:1009.2731.
- [29] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput.Phys.Commun. **178** (2008) 852–867, arXiv:0710.3820.

- 
- [30] A. Donnachie and P. Landshoff, *Total cross-sections*, Phys.Lett. **B296** (1992) 227–232, arXiv:hep-ph/9209205.
- [31] T. Regge, *Introduction to complex orbital momenta*, Nuovo Cim. **14** (1959) 951; T. Regge, *Bound states, shadow states and Mandelstam representation*, Nuovo Cim. **18** (1960) 947–956.
- [32] P. Collins, *An Introduction to Regge Theory and High-Energy Physics*. 1977.
- [33] CTEQ collaboration, H. L. Lai *et al.*, *Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions*, Eur.Phys.J. **C12** (2000) 375–392, arXiv:hep-ph/9903282.
- [34] J. Pumplin *et al.*, *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012, arXiv:hep-ph/0201195.
- [35] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152–155.
- [36] I. Belyaev *et al.*, *Handling of the generation of primary events in GAUSS, the LHCb simulation framework*, Nuclear Science Symposium Conference Record (NSS/MIC) **IEEE** (2010) 1155.
- [37] UA5 Collaboration, G. Alner *et al.*, *Scaling of Pseudorapidity Distributions at c.m. Energies Up to 0.9-TeV*, Z.Phys. **C33** (1986) 1–6.
- [38] UA5 Collaboration, K. Alpgard *et al.*, *Comparison of  $p\bar{p}$  and  $pp$  Interactions at  $\sqrt{s} = 53\text{-GeV}$* , Phys.Lett. **B112** (1982) 183.
- [39] CDF Collaboration, F. Abe *et al.*, *Pseudorapidity distributions of charged particles produced in  $\bar{p}p$  interactions at  $\sqrt{s} = 630\text{ GeV}$  and  $1800\text{ GeV}$* , Phys.Rev. **D41** (1990) 2330.
- [40] S. Todorova-Nova, *Colour reconnection in string model*. DELPHI 96-158 PHYS 651, 1996.
- [41] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys.Rev. **D82** (2010) 074018, arXiv:1005.3457.
- [42] R. Corke and T. Sjostrand, *Improved Parton Showers at Large Transverse Momenta*, Eur.Phys.J. **C69** (2010) 1–18, arXiv:1003.2384.
- [43] R. Corke and T. Sjostrand, *Interleaved Parton Showers and Tuning Prospects*, JHEP **1103** (2011) 032, arXiv:1011.1759.
- [44] R. Engel, *Photoproduction within the two-component dual parton model: amplitudes and cross-sections*, Z. Phys. **C66** (1995) 203–214.
- [45] R. Engel and J. Ranft, *Hadronic photon-photon interactions at high-energies*, Phys.Rev. **D54** (1996) 4244–4262, arXiv:hep-ph/9509373.

- [46] A. Capella, U. Sukhatme, C.-I. Tan, and J. T. T. Van, *Dual parton model*, Physics Reports **236** (1994), no. 4–5 225–329.
- [47] S. Navin, *Diffraction in Pythia*, arXiv:1005.3894.
- [48] M. Bahr *et al.*, *Herwig++ Physics and Manual*, Eur.Phys.J. **C58** (2008) 639–707, arXiv:0803.0883.
- [49] M. Bahr, S. Gieseke, and M. H. Seymour, *Simulation of multiple partonic interactions in Herwig++*, JHEP **0807** (2008) 076, arXiv:0803.3633.
- [50] S. Gieseke, P. Stephens, and B. Webber, *New formalism for QCD parton showers*, JHEP **0312** (2003) 045, arXiv:hep-ph/0310083.
- [51] D. Amati and G. Veneziano, *Preconfinement as a Property of Perturbative QCD*, Phys.Lett. **B83** (1979) 87.
- [52] S. Gieseke, C. Rohr, and A. Siodmok, *Colour reconnections in Herwig++*, Eur.Phys.J. **C72** (2012) 2225, arXiv:1206.0041.
- [53] K. Arnold *et al.*, *Herwig++ 2.6 Release Note*, arXiv:1205.4902.
- [54] J. Bellm *et al.*, *Herwig++ 2.7 Release Note*, arXiv:1310.6877.
- [55] HERWIG++, *Minimum-bias and underlying-event tunes*, [https://herwig.hepforge.org/trac/wiki/MB\\_UE\\_tunes](https://herwig.hepforge.org/trac/wiki/MB_UE_tunes).
- [56] M. H. Seymour and A. Siodmok, *Constraining MPI models using  $\sigma_{eff}$  and recent Tevatron and LHC Underlying Event data*, JHEP **1310** (2013) 113, arXiv:1307.5015.
- [57] A. Sherstnev and R. Thorne, *Parton Distributions for LO Generators*, Eur.Phys.J. **C55** (2008) 553–575, arXiv:0711.2473.
- [58] LHCb Collaboration, *LHCb webpage*, [http://lhcb.web.cern.ch/lhcb/speakersbureau/html/Material\\_for\\_Presentations.html](http://lhcb.web.cern.ch/lhcb/speakersbureau/html/Material_for_Presentations.html), March, 2014.
- [59] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.
- [60] ATLAS Collaboration, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.
- [61] CMS Collaboration, S. Chatrchyan *et al.*, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.
- [62] ALICE Collaboration, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.
- [63] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.
- [64] TOTEM Collaboration, G. Anelli *et al.*, *The TOTEM experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08007.

- 
- [65] LHCf Collaboration, O. Adriani *et al.*, *The LHCf detector at the CERN Large Hadron Collider*, JINST **3** (2008) S08006.
- [66] MoEDAL Collaboration, J. Pinfold *et al.*, *Technical Design Report of the MoEDAL Experiment*, Tech. Rep. CERN-LHCC-2009-006. MoEDAL-TDR-001, CERN, Geneva, Jun, 2009.
- [67] LHCb Collaboration, *LHCb material for presentations*, [http://lhcb.web.cern.ch/lhcb/speakersbureau/html/Material\\_for\\_Presentations.html](http://lhcb.web.cern.ch/lhcb/speakersbureau/html/Material_for_Presentations.html).
- [68] LHCb collaboration, *LHCb reoptimized detector design and performance: Technical Design Report*, CERN-LHCC-2003-030. LHCb-TDR-009.
- [69] LHCb collaboration, *LHCb magnet: Technical Design Report*, CERN-LHCC-2000-007. LHCb-TDR-001.
- [70] J. Van Tilburg, *Track simulation and reconstruction in LHCb*. CERN-THESIS-2005-020, 2005.
- [71] LHCb collaboration, *LHCb VELO (Vertex Locator): Technical Design Report*, CERN-LHCC-2001-011. LHCb-TDR-005.
- [72] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, arXiv:1405.7808. to appear in JINST.
- [73] LHCb collaboration, *LHCb inner tracker: Technical Design Report*, CERN-LHCC-2002-029. LHCb-TDR-008.
- [74] LHCb Silicon Tracker, *Material for Publications*, <http://lhcb.physik.uzh.ch/ST/public/material/index.php>.
- [75] LHCb collaboration, *LHCb outer tracker: Technical Design Report*, CERN-LHCC-2001-024. LHCb-TDR-006.
- [76] D. van Eijk *et al.*, *Radiation hardness of the LHCb Outer Tracker*, Nucl. Instrum. Meth. **A685** (2012) 62–69.
- [77] R. Arink *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9** (2014) P01002, arXiv:1311.3893.
- [78] LHCb collaboration, *LHCb RICH: Technical Design Report*, CERN-LHCC-2000-037. LHCb-TDR-003.
- [79] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, arXiv:1211.6759.
- [80] P. Cerenkov, *Visible Emission of Clean Liquids by Action of Radiation*, Doklady Akad. Nauk SSSR **2** (1934) 451.
- [81] LHCb collaboration, *LHCb calorimeters: Technical Design Report*, CERN-LHCC-2000-036. LHCb-TDR-002.

- [82] R. Aaij *et al.*, *Performance of the LHCb calorimeters*, LHCb-DP-2013-004. in preparation.
- [83] O. Deschamps, *LHCb Calorimeters - Commissioning & performance*, LHCb-TALK-2010-053.
- [84] LHCb collaboration, *LHCb muon system: Technical Design Report*, CERN-LHCC-2001-010. LHCb-TDR-004.
- [85] F. Archilli *et al.*, *Performance of the muon identification at LHCb*, JINST **8** (2013) P10020, [arXiv:1306.0249](https://arxiv.org/abs/1306.0249).
- [86] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, JINST **8** (2013) P04022, [arXiv:1211.3055](https://arxiv.org/abs/1211.3055).
- [87] LHCb collaboration, *LHCb trigger system: Technical Design Report*, CERN-LHCC-2003-031. LHCb-TDR-010.
- [88] LHCb Collaboration, *The Gaudi project*, <http://proj-gaudi.web.cern.ch/proj-gaudi/>.
- [89] LHCb Collaboration, *The Gauss project*, <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/gauss/>.
- [90] P. Golonka and Z. Was, *PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays*, Eur.Phys.J. **C45** (2006) 97–107, [arXiv:hep-ph/0506026](https://arxiv.org/abs/hep-ph/0506026).
- [91] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: a simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.
- [92] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans.Nucl.Sci. **53** (2006) 270.
- [93] LHCb Collaboration, *The boole project*, <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/boole/>.
- [94] LHCb Collaboration, *The Moore project*, <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/moore/>.
- [95] LHCb Collaboration, *The Brunel project*, <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel/>.
- [96] LHCb Collaboration, *The DaVinci project*, <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/>.
- [97] M. De Cian, U. Straumann, O. Steinkamp, and N. Serra, *Track Reconstruction Efficiency and Analysis of  $B^0 \rightarrow K^{*0} \mu^+ \mu^-$  at the LHCb Experiment*. PhD thesis, Zurich U., Sep, 2013. Presented 14 Mar 2013.
- [98] LHCb collaboration, R. Aaij *et al.*, *Measurement of charged particle multiplicities in pp collisions at  $\sqrt{s} = 7$  TeV in the forward region*, Eur. Phys. J. **C72** (2012) 1947, [arXiv:1112.4592](https://arxiv.org/abs/1112.4592).

- 
- [99] D. Hutchcroft, *VELO Pattern Recognition*, Tech. Rep. LHCb-2007-013. CERN-LHCb-2007-013, CERN, Geneva, Mar, 2007.
- [100] O. Callot, *FastVelo, a fast and efficient pattern recognition package for the Velo*, Tech. Rep. LHCb-PUB-2011-001. CERN-LHCb-PUB-2011-001, CERN, Geneva, Jan, 2011. LHCb.
- [101] M. Benayoun and O. Callot, *The forward tracking, an optical model method*, Tech. Rep. LHCb-2002-008, CERN, Geneva, Feb, 2002. revised version number 1 submitted on 2002-02-22 17:19:02.
- [102] O. Callot and S. Hansmann-Menzemer, *The Forward Tracking: Algorithm and Performance Studies*, Tech. Rep. LHCb-2007-015. CERN-LHCb-2007-015, CERN, Geneva, May, 2007.
- [103] M. Needham, *Performance of the Track Matching*, Tech. Rep. LHCb-2007-129. CERN-LHCb-2007-129, CERN, Geneva, Oct, 2007.
- [104] E. Rodrigues, *Dealing with clones in the tracking*, Tech. Rep. LHCb-2006-057. CERN-LHCb-2006-057, CERN, Geneva, Nov, 2006.
- [105] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, arXiv:1408.1251. submitted to JINST.
- [106] J. Brehmer, J. Albrecht, and P. Seyfert, *Ghost probability: an efficient tool to remove background tracks*, Tech. Rep. LHCb-INT-2012-025. CERN-LHCb-INT-2012-025, CERN, Geneva, Sep, 2012.
- [107] ALICE Collaboration, K. Aamodt *et al.*, *Charged-particle multiplicity measurement in proton-proton collisions at  $\sqrt{s} = 0.9$  and 2.36 TeV with ALICE at LHC*, Eur.Phys.J. **C68** (2010) 89–108, arXiv:1004.3034.
- [108] ALICE collaboration, K. Aamodt *et al.*, *Charged-particle multiplicity measurement in proton-proton collisions at  $\sqrt{s} = 7$  TeV with ALICE at LHC*, Eur. Phys. J. **C68** (2010) 345–354, arXiv:1004.3514.
- [109] ATLAS Collaboration, G. Aad *et al.*, *Charged-particle multiplicities in pp interactions at  $\sqrt{s} = 900$  GeV measured with the ATLAS detector at the LHC*, Phys.Lett. **B688** (2010) 21–42, arXiv:1003.3124.
- [110] ATLAS collaboration, G. Aad *et al.*, *Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC*, New J. Phys. **13** (2011) 053033, arXiv:1012.5104.
- [111] CMS collaboration, V. Khachatryan *et al.*, *Charged particle multiplicities in pp interactions at  $\sqrt{s} = 0.9$ , 2.36, and 7 TeV*, JHEP **01** (2011) 079, arXiv:1011.5531.
- [112] N. Brook, M. Meissner, T. Ruf, and M. Schmelling, *Measurement of Charged Particle Multiplicities in pp-collisions at a centre-of-mass energy of  $\sqrt{s} = 7$  TeV in the forward region*, LHCb-ANA-2011-044 (Jan, 2013).

- [113] V. Blobel, *Unfolding methods in high-energy physics experiments*, DESY-84-118 (Dec, 1984) 40 p, [arXiv:0208022v1](#).
- [114] J. F. Grosse-Oetringhaus and K. Reygers, *Charged-Particle Multiplicity in Proton-Proton Collisions*, J.Phys. **G37** (2010) 083001, [arXiv:0912.0023](#).
- [115] G. Zech, *Comparing statistical data to Monte Carlo simulation: Parameter fitting and unfolding*, DESY-95-113 (1995).
- [116] Wolfram MathWorld, *Negative Binomial Distribution*, <http://mathworld.wolfram.com/NegativeBinomialDistribution.html>.
- [117] LHCb Collaboration, R. Aaij *et al.*, *Prompt  $K_s^0$  production in pp collisions at  $\sqrt{s} = 0.9$  TeV*, Phys.Lett. **B693** (2010) 69–80, [arXiv:1008.3105](#).
- [118] LHCb collaboration, R. Aaij *et al.*, *Measurement of  $V^0$  production ratios in pp collisions at  $\sqrt{s} = 0.9$  and 7 TeV*, JHEP **08** (2011) 34, [arXiv:1107.0882](#).
- [119] A. Jaeger, P. Seyfert, M. De Cian, J. van Tilburg, and S. Hansmann-Menzemer, *Measurement of the track finding efficiency*, Tech. Rep. LHCb-PUB-2011-025, CERN-LHCb-PUB-2011-025, CERN, Geneva, Apr, 2012.
- [120] G. D’Agostini, *A Multidimensional unfolding method based on Bayes’ theorem*, Nucl.Instrum.Meth. **A362** (1995) 487–498.
- [121] T. Adye, *Unfolding algorithms and tests using RooUnfold*, ArXiv e-prints [arXiv:1105.1160](#).
- [122] HepData - analysis ins1281685, *Measurement of charged particle multiplicities and densities in pp collisions at  $\sqrt{s} = 7$  TeV in the forward region*, <http://hepdata.cedar.ac.uk/view/ins1281685>.
- [123] Rivet project, *Homepage*, <http://rivet.hepforge.org/>.
- [124] K. Eggert *et al.*, *Angular Correlations Between the Charged Particles Produced in pp Collisions at ISR Energies*, Nucl.Phys. **B86** (1975) 201.
- [125] UA5 Collaboration, R. Ansorge *et al.*, *Charged Particle Correlations in  $\bar{P}P$  Collisions at c.m. Energies of 200-GeV, 546-GeV and 900-GeV*, Z.Phys. **C37** (1988) 191–213.
- [126] STAR Collaboration, C. Adler *et al.*, *Disappearance of back-to-back high  $p_T$  hadron correlations in central Au+Au collisions at  $\sqrt{s_{NN}} = 200$ -GeV*, Phys.Rev.Lett. **90** (2003) 082302, [arXiv:nucl-ex/0210033](#).
- [127] PHENIX Collaboration, A. Adare *et al.*, *System Size and Energy Dependence of Jet-Induced Hadron Pair Correlation Shapes in Cu+Cu and Au+Au Collisions at  $s(NN)^{1/2} = 200$  and 62.4-GeV*, Phys.Rev.Lett. **98** (2007) 232302, [arXiv:nucl-ex/0611019](#).

- 
- [128] STAR Collaboration, B. Abelev *et al.*, *Long range rapidity correlations and jet production in high energy nuclear collisions*, Phys.Rev. **C80** (2009) 064912, [arXiv:0909.0191](#).
- [129] PHOBOS Collaboration, B. Alver *et al.*, *High transverse momentum triggered correlations over a large pseudorapidity acceptance in Au+Au collisions at  $s(NN)^{1/2} = 200$  GeV*, Phys.Rev.Lett. **104** (2010) 062301, [arXiv:0903.2811](#).
- [130] ATLAS Collaboration, G. Aad *et al.*, *Measurement of the azimuthal anisotropy for charged particle production in  $\sqrt{s_{NN}} = 2.76$  TeV lead-lead collisions with the ATLAS detector*, Phys.Rev. **C86** (2012) 014907, [arXiv:1203.3087](#).
- [131] CMS Collaboration, S. Chatrchyan *et al.*, *Long-range and short-range dihadron angular correlations in central PbPb collisions at a nucleon-nucleon center of mass energy of 2.76 TeV*, JHEP **1107** (2011) 076, [arXiv:1105.2438](#).
- [132] CMS Collaboration, S. Chatrchyan *et al.*, *Centrality dependence of dihadron correlations and azimuthal anisotropy harmonics in PbPb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV*, Eur.Phys.J. **C72** (2012) 2012, [arXiv:1201.3158](#).
- [133] ALICE Collaboration, K. Aamodt *et al.*, *Harmonic decomposition of two-particle angular correlations in Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV*, Phys.Lett. **B708** (2012) 249–264, [arXiv:1109.2501](#).
- [134] STAR Collaboration, B. Abelev *et al.*, *Three-particle coincidence of the long range pseudorapidity correlation in high energy nucleus-nucleus collisions*, Phys.Rev.Lett. **105** (2010) 022301, [arXiv:0912.3977](#).
- [135] P. Kolb, P. Huovinen, U. W. Heinz, and H. Heiselberg, *Elliptic flow at SPS and RHIC: From kinetic transport to hydrodynamics*, Phys.Lett. **B500** (2001) 232–240, [arXiv:hep-ph/0012137](#).
- [136] B. Back *et al.*, *The PHOBOS perspective on discoveries at RHIC*, Nucl.Phys. **A757** (2005) 28–101, [arXiv:nucl-ex/0410022](#).
- [137] STAR Collaboration, J. Adams *et al.*, *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. **A757** (2005) 102–183, [arXiv:nucl-ex/0501009](#).
- [138] PHENIX Collaboration, K. Adcox *et al.*, *Formation of dense partonic matter in relativistic nucleus-nucleus collisions at RHIC: Experimental evaluation by the PHENIX collaboration*, Nucl.Phys. **A757** (2005) 184–283, [arXiv:nucl-ex/0410003](#).
- [139] J.-y. Jia, *How to Make Sense of the Jet Correlations Results at RHIC?*, Eur.Phys.J. **C62** (2009) 255–264, [arXiv:0810.0051](#).
- [140] F. Henyey, *Elastic scattering from the multiperipheral model*, Phys.Lett. **B45** (1973) 469–474.

- [141] E. L. Berger, *Rapidity Correlations at Fixed Multiplicity in Cluster Emission Models*, Nucl.Phys. **B85** (1975) 61.
- [142] J. Meunier and G. Plaut, *Azimuthal Correlations in  $p$  Or  $q$ -Factorized Cluster Production*, Nucl.Phys. **B87** (1975) 74.
- [143] C. Michael, *Link Correlated Models for Multiparticle Production*, Nucl.Phys. **B103** (1976) 296.
- [144] PHOBOS Collaboration, B. Alver *et al.*, *Cluster properties from two-particle angular correlations in  $p + p$  collisions at  $s^{1/2} = 200\text{-GeV}$  and  $410\text{-GeV}$* , Phys.Rev. **C75** (2007) 054913, [arXiv:0704.0966](#).
- [145] PHOBOS Collaboration, B. Alver *et al.*, *System size dependence of cluster properties from two-particle angular correlations in  $\text{Cu}+\text{Cu}$  and  $\text{Au}+\text{Au}$  collisions at  $s(NN)^{1/2} = 200\text{-GeV}$* , Phys.Rev. **C81** (2010) 024904, [arXiv:0812.1172](#).
- [146] A. Dumitru, F. Gelis, L. McLerran, and R. Venugopalan, *Glasma flux tubes and the near side ridge phenomenon at RHIC*, Nucl.Phys. **A810** (2008) 91–108, [arXiv:0804.3858](#).
- [147] S. Sarkar, H. Satz, and B. Sinha, *The Physics of the Quark-Gluon Plasma*, Lecture Notes in Physics **785** (2010).
- [148] J.-Y. Ollitrault, *Anisotropy as a signature of transverse collective flow*, Phys. Rev. D **46** (Jul, 1992) 229–245.
- [149] S. Voloshin and Y. Zhang, *Flow study in relativistic nuclear collisions by Fourier expansion of Azimuthal particle distributions*, Z.Phys. **C70** (1996) 665–672, [arXiv:hep-ph/9407282](#).
- [150] J.-Y. Ollitrault, *Anisotropy as a signature of transverse collective flow*, Phys.Rev. **D46** (1992) 229–245.
- [151] B. Alver and G. Roland, *Collision geometry fluctuations and triangular flow in heavy-ion collisions*, Phys.Rev. **C81** (2010) 054905, [arXiv:1003.0194](#).
- [152] B. H. Alver, C. Gombeaud, M. Luzum, and J.-Y. Ollitrault, *Triangular flow in hydrodynamics and transport theory*, Phys.Rev. **C82** (2010) 034913, [arXiv:1007.5469](#).
- [153] B. Schenke, S. Jeon, and C. Gale, *Elliptic and triangular flow in event-by-event  $D = 3 + 1$  viscous hydrodynamics*, Phys. Rev. Lett. **106** (Jan, 2011) 042301.
- [154] Z. Qiu, C. Shen, and U. Heinz, *Hydrodynamic elliptic and triangular flow in  $\text{Pb-Pb}$  collisions at  $\sqrt{s} = 2.76\text{ATeV}$* , Phys.Lett. **B707** (2012) 151–155, [arXiv:1110.3033](#).
- [155] W. Li, *Observation of a 'Ridge' correlation structure in high multiplicity proton-proton collisions: A brief review*, Mod.Phys.Lett. **A27** (2012) 1230018, [arXiv:1206.0148](#).

- 
- [156] J. Casalderrey-Solana and U. A. Wiedemann, *Eccentricity fluctuations make flow measurable in high multiplicity p-p collisions*, Phys.Rev.Lett. **104** (2010) 102301, arXiv:0911.4400.
- [157] K. Werner, I. Karpenko, and T. Pierog, *"ridge" in proton-proton scattering at 7 tev*, Phys. Rev. Lett. **106** (Mar, 2011) 122004.
- [158] E. Avsar, C. Flensburg, Y. Hatta, J.-Y. Ollitrault, and T. Ueda, *Eccentricity and elliptic flow in proton-proton collisions from parton evolution*, Phys.Lett. **B702** (2011) 394–397, arXiv:1009.5643.
- [159] E. Avsar, Y. Hatta, C. Flensburg, J. Ollitrault, and T. Ueda, *Eccentricity and elliptic flow in pp collisions at the LHC*, J.Phys. **G38** (2011) 124053, arXiv:1106.4356.
- [160] W.-T. Deng, Z. Xu, and C. Greiner, *Elliptic and Triangular Flow and their Correlation in Ultrarelativistic High Multiplicity Proton Proton Collisions at 14 TeV*, Phys.Lett. **B711** (2012) 301–306, arXiv:1112.0470.
- [161] P. Bozek and W. Broniowski, *Collective dynamics in high-energy proton-nucleus collisions*, Phys.Rev. **C88** (2013), no. 1 014903, arXiv:1304.3044.
- [162] P. Bozek, *Collective flow in p-Pb and d-Pd collisions at TeV energies*, Phys.Rev. **C85** (2012) 014911, arXiv:1112.0915.
- [163] P. Bozek and W. Broniowski, *Correlations from hydrodynamic flow in p-Pb collisions*, Phys.Lett. **B718** (2013) 1557–1561, arXiv:1211.0845.
- [164] E. Iancu and R. Venugopalan, *The Color glass condensate and high-energy scattering in QCD*, arXiv:hep-ph/0303204.
- [165] F. Gelis, E. Iancu, J. Jalilian-Marian, and R. Venugopalan, *The Color Glass Condensate*, Ann.Rev.Nucl.Part.Sci. **60** (2010) 463–489, arXiv:1002.0333.
- [166] J. Albacete and C. Marquet, *Azimuthal correlations of forward dihadrons in d + Au collisions at rhic in the color glass condensate*, Phys. Rev. Lett. **105** (Oct, 2010) 162301.
- [167] K. Dusling and R. Venugopalan, *Evidence for BFKL and saturation dynamics from dihadron spectra at the LHC*, Phys.Rev. **D87** (2013), no. 5 051502, arXiv:1210.3890.
- [168] K. Dusling and R. Venugopalan, *Explanation of systematics of CMS p+Pb high multiplicity di-hadron data at  $\sqrt{s_{NN}} = 5.02$  TeV*, Phys.Rev. **D87** (2013), no. 5 054014, arXiv:1211.3701.
- [169] K. Dusling and R. Venugopalan, *Comparison of the color glass condensate to dihadron correlations in proton-proton and proton-nucleus collisions*, Phys.Rev. **D87** (2013), no. 9 094034, arXiv:1302.7018.
- [170] Y. V. Kovchegov and D. E. Wertepny, *Long-Range Rapidity Correlations in Heavy-Light Ion Collisions*, Nucl.Phys. **A906** (2013) 50–83, arXiv:1212.1195.

- [171] B. Arbuzov, E. Boos, and V. Savrin, *Cms ridge effect at lhc as a manifestation of bremsstrahlung of gluons off quarks accelerated in a strong color field*, The European Physical Journal C **71** (2011), no. 9 .
- [172] R. C. Hwa and C. Yang, *Ridge Formation Induced by Jets in pp Collisions at 7 TeV*, Phys.Rev. **C83** (2011) 024911, [arXiv:1011.0965](#).
- [173] C.-Y. Wong, *Momentum Kick Model Description of the Ridge in (Delta-phi)-(Delta eta) Correlation in pp Collisions at 7 TeV*, Phys.Rev. **C84** (2011) 024901, [arXiv:1105.5871](#).
- [174] X.-N. Wang and M. Gyulassy, *HIJING: A Monte Carlo model for multiple jet production in p p, p A and A A collisions*, Phys.Rev. **D44** (1991) 3501–3516.
- [175] R. Versteegen, *Implications of LHCb polarity change*, LPC meeting (Jan 28th, 2013).
- [176] S. Kullback, *Letter to Editor: the Kullback-Leibler Distance*, The American Statistician **41** (1987) 340. The use of the Kullback-Leibler distance at LHCb is described in M. Needham, *Clone Track Identification Using the Kullback-Leibler Distance*, LHCb-2008-002.
- [177] R. Snellings, *Elliptic Flow: A Brief Review*, New J.Phys. **13** (2011) 055008, [arXiv:1102.3010](#).
- [178] R. J. Glauber, In Lectures in Theoretical Physics, Ed. by W. E. Brittin and L. G. Dunham (Interscience, New York, 1959) **1** 315.
- [179] M. L. Miller, K. Reygers, S. J. Sanders, and P. Steinberg, *Glauber modeling in high energy nuclear collisions*, Ann.Rev.Nucl.Part.Sci. **57** (2007) 205–243, [arXiv:nucl-ex/0701025](#).
- [180] ALICE Collaboration, K. Aamodt *et al.*, *Elliptic flow of charged particles in Pb-Pb collisions at 2.76 TeV*, Phys.Rev.Lett. **105** (2010) 252302, [arXiv:1011.3914](#).
- [181] CMS Collaboration, *PhysicsResultsHIN12015*, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsHIN12015>.
- [182] ATLAS Collaboration, G. Aad *et al.*, *Observation of Associated Near-side and Away-side Long-range Correlations in  $\sqrt{s_{NN}}=5.02$  TeV Proton-lead Collisions with the ATLAS Detector*, Phys.Rev.Lett. **110** (2013) 182302, [arXiv:1212.5198](#).
- [183] ALICE Collaboration, B. Abelev *et al.*, *Long-range angular correlations on the near and away side in p-Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV*, Phys.Lett. **B719** (2013) 29–41, [arXiv:1212.2001](#).
- [184] PHENIX Collaboration, A. Adare *et al.*, *Dihadron azimuthal correlations in Au+Au collisions at  $s(NN)^{1/2} = 200$ -GeV*, Phys.Rev. **C78** (2008) 014901, [arXiv:0801.4545](#).

- [185] N. N. Ajitanand *et al.*, *Decomposition of harmonic and jet contributions to particle-pair correlations at ultrarelativistic energies*, Phys. Rev. C **72** (Jul, 2005) 011902.
- [186] LHCb, M. Meissner, *Studies of Soft QCD at LHCb*, PoS **EPS-HEP2013** (2013) 444.