UNIVERSITEIT ANTWERPEN

Faculteit Wetenschappen

Departement Fysica

Commissioning of the CMS pixel tracker and measurement of the charged particle multiplicity at the LHC

Ingebruikname van de CMS Silicium pixel detector en meting van de geladen deeltjes multipliciteit bij de LHC

Proefschift voorgelegd tot het behalen van de graad van doctor in de wetenschappen aan de Universiteit Antwerpen

Ph.D. Student:
Romain Rougny

Promotor:
Prof. Nick van Remortel

CERN-THESIS-2015-063
09/06/2015

Samenvatting

De structuur van materie op microscopische schaal is sinds meer dan een eeuw het onderwerp van wetenschappelijk onderzoek. Ironish genoeg vergt het moderne onderzoek naar de allerkleinste structuren de grootste apparatuur ter wereld. De Large Hadron Collider (LHC) is één van deze machines die op dit moment botsingen tussen protonen teweeg brengt bij de grootste massamiddelpunts energie. Deze versneller, die gebouwd is in dezelfde tunnel als zijn voorganger de LEP, is ontworpen om protonen frontaal te laten botsen bij een massamiddelpunts energie van 14 TeV. Rond de ring zijn vier experimenten opgesteld die het resultaat van die botsingen registreren. De Compact Muon Solenoid (CMS) detector is één van die experimenten en heeft een uitgebreid onderzoeksprogramma met als doel het fundamenteel onderzoek naar elementaire deeltjes. De zoektocht naar het laatste ontbrekende onderdeel van het Standaard Model van de elementaire deeltjes, het Higgs boson, heeft de ontwikkeling van de LHC en haar grote experimenten grotendeels gedreven, en met succes. In Juli 2012 leidde het tot de experimentele ontdekking van het Higgs deeltje met een Belgische Nobelprijs voor de theoretische ontwikkeling van het formalisme tot gevolg. Howel dit de voornaamste doelstelling was van RUN I, de eerste fase van het LHC programma, hebben de experimenten tal van niewe inzichten gegeven in de structuur van het allerkleinste.

De LHC werd in gebruik genomen in September 2008, doch werd vrij snel weer stilgelegd wegens een beschadiging van belangrijke componenten. Het jaar dat gespendeerd werd aan de herstellingen van de versneller liet de experimenten toe hun detectoren grondig te testen en te calibreren. Dit proefschrift beschrijft uitvoerig de gedetailleerde werking van de silicium pixel detector, de subdetector die het dichtst bij het interactiepunt gesitueerd is en deel uitmaakt van de sporenkamers van de CMS detector. De 66 miljoen sensoren die zij bevat laten toe om zeer nauwkeurig het traject van geladen deeltjes op te meten, alsook hun punt van oorsprong. Verschillende calibratie procedures worden behandeld, met de nadruk op de *gain* calibratie. Deze calibratie is nodig om de gecollecteerde lading uit elk van de pixel cellen te vertalen in elektron ladingen. Samen met de procedure om de calibratie gegevens per pixel te verwerken, worden de resultaten van verschillende calibratie metingen gepresenteerd, hun evolutie in de tijd gerelateerd aan de effecten van voortdurende bestraling van het instrument en de manier waarop dit door het bijsturen van de operationele parameters kan worden gecorrigeerd. De ingebruikname diende ook als grondige test van de volledige keten voor gegevensverwerking van het experiment aan de hand van het regsitreren van cosmische straling. In het bijzonder wordt hier de bepaling van de efficiëntie van de pixel sensors aan de hand van cosmische straling besproken. Met deze techniek vond men een gemiddelde efficiëntie van $(97.55 \pm 0.05)\%$ voor het *barrel* gedeelte van de pixel detector, gebaseerd op 95% van alle goed werkende modules. De techniek is echter niet bruikbaar om de efficiëntie van het meest voorwartse deel van de detector, de *endcaps*, te bepalen.

De LHC leverde in zijn beginjaren *pp* botsingen bij toenemende massimiddelpuntsenergie van 0.9 TeV, 2.26 TeV en 7 TeV. Deze laatste werd tijdens de eerste fase van het

programma als een veilig werkpunt beschouwd en gedurende het laatste jaar nog met 1 TeV verhoogd. De eerste gegevens bij de drie voorgenoemde energieën zijn in dit proefschrift verwerkt voor de bepaling van de geladen deeltjesmultipliciteit. Omwille van de vergelijking met historische data werd deze meting initieel uitgevoerd voor niet-diffractieve *pp* botsingen. Gebruik makende van een sporenreconstructie voor deeltjes met impulsen boven de 100 MeV werd de multipliciteitsverdeling gecorrigeerd voor de efficiëntie van de trigger en de aanwezigheid van diffractieve achtergronden. Om dan de werkelijke multipliciteitsverdeling te bekomen werd een Bayesiaanse onvouwingsprocedure gebruikt die de effecten van sporenreconstructie en detector efficiëntie en de aanwezigheid van secundaire vervallen in rekening brengt. De statistische onzekerheden werden bestudeerd met behulp van een Monte-Carlo resampling techniek. Een groot aantal systematische invloeden op de bepaling van de ware multipliciteit werden bestudeerd en uiteindelijk werden er drie beduidende invloeden gevonden die de nauwkeurigheid op het resultaat beïnvloeden: de model afhankelijkheid van het verwijderen van de diffractieve achtergronden, de trigger en selectie efficiëntie, en de behandeling van de efficiëntie van de sporen reconstructie. Tenslotte passen we een kleine correctie toe die de niet gereconstrueerde deeltjes met een impuls kleiner dan 100 MeV in rekening brengt.

De volledig gecorrigeerde geladen multipliciteitsverdelingen worden vervolgens grepresenteerd in verschillende pseudo-rapiditeitsintervallen gaande van $|\eta| < 0.5$ tot $|\eta| < 2.4$. De resultaten zijn in zeer goede overeenstemming met zowel resultaten bekomen bij andere LHC experimenten als met resultaten uit oudere experimenten. Een opmerkelijke eigenschap van de multipliciteitsverdeling is het ontstaan van een schouder structuur bij hoge multipliciteiten naarmate de bostsingsenergie toeneemt. De gemeten verdelingen worden ook vergeleken met fysische modellen geïmplementeerd in Monte-Carlo codes, waarbij de grootste discrepanties optreden bij grote deeltjesmultipliciteiten. Dit demonstreert de toenemende invloed van het ontstaan van meerdere partonische botsingen (*multiple partonic interactions, MPI*) bij 7 TeV die door de meeste modellen worden onderschat. De meting van de gemiddelde geladen deeltjesimpuls als functie van de multipliciteit, eveneens uitgevoerd bij de drie voorgenoemde energieën, onderzoekt dieper de dynamica van hoge multipliciteits botsingen. Deze hebben doorgaans een zachter impuls spectrum dan de meeste modellen die secundaire botsingen als jet productie processen beschrijven. Zoals verwacht is de impulsafhankelijkheid van de multipliciteit dezelfde bij de drie bestudeerde energieën. De toename van de gemiddelde multipliciteit met toenemende massamiddelpuntsenergie is ook steiler dan de voorspellingen van de meeste modellen. De multipliciteitsverdeling blijkt in strijd te zijn met het KNO schalingsprincipe voor de grootste pseudo-rapiditeitsintervallen, hetgeen bevestigd wordt aan de hand van de statistische momenten van de multipliciteitsverdeling. De schending van KNO schaling kan worden toegedragen aan de veranderingen in de verhoudingen van werkzame doorsnedes van alle processen die bijdragen tot de minimum bias gegevens, terwijk elk proces afzonderlijk KNO schaling waarschijnlijk respecteert. Daarnaast zorgt de toename van meerdere partonische interacties met toenemende massamiddelpuntsenergie voor een schending van het KNO principe. Desondanks blijkt KNO schaling onverwacht algemeen geldig te zijn voor kleinere pseudo-rapiditeitsintervallen hetgeen waarschijnlijk te wijten is aan de afwezigheid van lange afstandscorrelaties. De studie van statistische momenten van de multiplicitietsverdeling wijst immers op een drastische toename van triviale correlaties met de bostingsenergie, terwijl de intrinsieke correlaties te wijten aan de onderliggende kwantumtheorie meer onderdrukt zijn.

Na het vergaren van meer bostsingsgegevens met de LHC, alsook met het bestuderen van een tussenliggende energieschaal van 2.76 TeV die noodzakelijk was voor het zware ionen programma van de versneller, is de analyse van de multipliciteit herhaald met een

andere selectieprocedure. Het doel hiervan is tweevoudig: de modelafhanhelijkheid van de definitie van diffractieve gebeurtenissen werd verminderd en een meer coherente aanpak tussen de LHC experimenten werd ingevoerd, gebaseerd op inzichten verworven bij de eerste metingen. De resultaten tonen een ontzettend goede overeenstemming tussen alle experimenten en de fysische conclusies van de eerste resultaten worden bevestigd. De Monte-Carlo modellen werden ondertussen ook aangepast en vertonen een merkbaar betere overeenstemming met de gemeten waarden. Toch ontbreekt het nog steeds aan een volledig theoretisch inzicht in de Quantum Chromo Dynamica beschrijving van bostingen tussen protonen met een relatief kleine impulsoverdracht.

Contents

Page

Introduction **1**

1 Multiplicity **5**
1.1 The multiplicity distribution . . . 5
1.2 Moments . . . 6
1.3 Feynman scaling . . . 8
1.4 Koba-Nielsen-Olesen (KNO) scaling . . . 10
1.5 Multiplicity models . . . 10
1.5.1 Single Negative Binomial Distribution (NBD) fit . . . 10
1.5.2 Double NBD fit . . . 12
1.6 Processes in *pp* collisions . . . 13

2 The LHC collider and the CMS experiment **15**
2.1 LHC . . . 16
2.1.1 Description of the LHC Machine . . . 16
2.1.2 Performance Goals of the LHC . . . 17
2.1.3 The LHC Operation so far . . . 17
2.2 The Compact Muon Solenoid Experiment . . . 18
2.2.1 The Tracker . . . 19
2.2.2 The electromagnetic calorimeter . . . 21
2.2.3 The hadron calorimeter . . . 22
2.2.4 The Magnet . . . 23
2.2.5 Muon Detector . . . 23
2.2.6 The CMS Trigger . . . 24

3 The PIXEL detector: calibration and commissioning **27**
3.1 Detailed description of the pixel detector . . . 28
3.1.1 The detector geometry . . . 28
3.1.2 The pixel modules . . . 30
3.1.2.1 The sensor . . . 31
3.1.2.2 The Read-Out Chip . . . 32
3.1.2.3 The High Density Interconnect . . . 34
3.1.2.4 The Token Bit Manager . . . 35
3.1.3 The Analog Read-out . . . 36
3.1.4 Upgrade of the pixel detector . . . 38
3.2 The calibration of the PIXEL detector . . . 41
3.2.1 Pixel alive test . . . 41
3.2.2 Baseline calibration . . . 41
3.2.3 Clock adjustments and Delay scans . . . 42
3.2.4 Threshold calibration . . . 43

3.2.5 Address level calibration . . . 44
3.3 The gain calibration of the PIXEL detector . . . 46
3.3.1 Introduction . . . 46
3.3.2 Calibration Procedure . . . 47
3.3.3 Results . . . 55
3.3.4 Evolution of the gain/pedestal values over time . . . 56
3.3.5 Simulation . . . 59
3.4 The commissioning of the PIXEL detector . . . 60
3.4.1 The CRAFT run . . . 60
3.4.2 Several Results from cosmic run . . . 60
3.4.2.1 Hit distributions and charge collection . . . 62
3.4.2.2 Lorentz angle measurement . . . 63
3.4.2.3 Hit residuals . . . 64
3.4.2.4 Hit resolution . . . 65
3.4.2.5 Track impact parameter resolution . . . 66
3.4.3 Efficiency measurement . . . 67
3.4.3.1 Preliminaries . . . 68
3.4.3.2 Hit reconstruction efficiencies . . . 69
3.4.3.3 Results . . . 71
3.4.3.4 Conclusions . . . 75

4 Tracking in CMS **79**
4.1 Clustering . . . 79
4.2 Tracklets . . . 79
4.3 General description of CMS tracking . . . 80
4.3.1 Seeding . . . 80
4.3.2 Trajectory Building . . . 81
4.3.3 Trajectory Refitting . . . 81
4.3.4 Trajectory Cleaning . . . 81
4.3.5 Vertexing . . . 81
4.3.6 Beamspot . . . 82
4.4 Iterative tracking and its parameter selection . . . 83
4.4.1 Standard Tracking . . . 83
4.4.2 Tracking adapted to low-pt . . . 83

Introduction to the analysis **85**

5 Multiplicity analysis with Non-Single Diffractive events **87**
5.1 Introduction . . . 87
5.2 Models . . . 88
5.3 Data Sample . . . 89
5.4 Track reconstruction and vertex requirements . . . 91
5.5 Acceptance definition . . . 96
5.6 Corrections . . . 97
5.7 Systematic uncertainties . . . 101
5.8 Results . . . 104
5.8.1 Charged hadron multiplicity distributions . . . 104
5.8.2 Violation of KNO scaling . . . 108
5.8.3 Comparison with Monte-Carlo models . . . 109
5.8.4 Energy dependence of the mean multiplicity . . . 112

5.8.5 Energy evolution of the moments and cumulants of the multiplicity distribution . . . 113
5.9 Conclusion . . . 116

6 Multiplicity analysis with Inelastic events 117
6.1 Introduction . . . 117
6.2 Data Sample . . . 117
6.3 Track reconstruction and acceptance definition . . . 118
6.4 Corrections . . . 123
6.5 Cross-Checks and Systematic uncertainties . . . 126
6.5.1 Cross-Checks . . . 126
6.5.2 Systematic uncertainties . . . 128
6.6 Results . . . 131
6.6.1 Multiplicity distributions, $< p_{\mathrm{T}} >$ VS multiplicity . . . 131
6.6.2 KNO distributions, Moments . . . 132
6.6.3 Conclusion . . . 136

Summary 137

Acknowledgements 141
Terms and Abbreviations . . . 143

List of Figures 145

Appendices 153

A Gain Calibration Fitting Problems 155

B Numerical Values for Moments of the Multiplicity Mistributions 157
B.1 Non-Single-Diffractive Events . . . 157
B.2 Inelastic Events . . . 161

C Distributions of the different systematic studies 163
C.1 For NSD events . . . 163
C.2 For Inelastic events . . . 166

Bibliography 175

Introduction

14^{th} May 2012. A day that will be remembered by particle physicists. Not just because the CERN, the Conseil Européen pour la Recherche Nucléaire, the world's biggest and largest laboratory for high-energy experiments, is for the day the center of attention of our little planet, with journalists flooding in from all around the world, and even high ranking officials from several countries attending. Not just because it is, for many, the day they waited for years, decades, or even their entire career for the wisest among us. But maybe because, in a joint conference, two of the experiments from CERN announced the discovery of a new particle at 125 GeV, with properties consistent with the now famous Higgs boson.

But why is this so important to us particle physicists, to fundamental science in general, and even potentially - in the future - for mankind ? Because this particle might be the last missing piece of our particle physics puzzle, that was theorized years ago, but was yet eluding us. This particle, if it is indeed proven to be the particle proposed by Robert Brout, François Englert and Peter Higgs as part of the Standard Model (SM), will be the pinnacle of the particle physics research that spanned for more than a century, the SM theory taking shape years ago.

The Standard Model is the theory that describes the elementary particles and their interactions. It comprises three out of the four particle interactions: the strong interaction, the weak interaction and the electromagnetic interaction. Each of them is mediated through particles, the gauge bosons: the photons are responsible of the electromagnetic force, and are determined through the Quantum ElectroDynamics (QCD); the $W^{\pm}$ and Z^0 mediate the weak interaction and are governed by the ElectroWeak (EW) theory; finally the Quantum ChromoDynamics (QCD) describes the strong interaction and its corresponding mediators, the gluons. The elementary particles are the indivisible smallest particles as far as we know, from which all matter around us is built. They are divided into two families, the leptons and the quarks, each family having three generations. Only the lightest particles from the first generations are stable enough and hence compose the matter outside high-energy colliders: the electron and its neutrino (both leptons), and the two quarks up and down (fermions). The last piece of the Standard Model is the Higgs boson and its associated field, the Higgs field, who is responsible for the mass of the particles.

One of the remarkable properties of the Standard Model is its undefeated power of prediction: many of its predicted constituent particles where found through the years with the ever evolving colliders. The last one to complete the picture was the Higgs boson, who had yet to show himself despite the two decades of collision data analyzed. It was the main underlying goal that shaped the R&D of the latest collider and detectors of the CERN complex. Therefore, its discovery would again empower the Standard Model as a sturdy pedestal for our understanding of fundamental physics. Despite all its merits, the Standard Model is not perfect: for one, it does not include the gravitation interaction,

which is described by Einstein's theory of relativity. Other experimental facts are not accounted by the SM, which is still lacking a candidate for dark matter, or can't explain the neutrinos' mass. Nevertheless, the discovery of the Higgs boson confirms the Standard Model as a strong theory, in spite of the wonders its lack of discovery would have also brought to scientific community.

The work presented here is not about the Higgs discovery, but on the contrary represents one of the many small bricks that are needed when looking into the data to prove or disprove the existence of such new particles. Minimum-bias data, as well as Underlying Events (UEs), are always in the background of many important discovery channels of SM particles and beyond. Their study is therefore a necessary prerequisite to the proper understanding and measurement of such particles. Although rather basics in term of measurements, observables like the pseudorapidity density or the multiplicity distribution of minimum-bias events give therefore very important information on the dynamics of proton-proton interactions at unprecedented energies realized at the LHC. They also have a more fundamental role: as the EM and weak coupling are both small, one can apply perturbation theory to make accurate predictions and calculations. But this is not always possible for QCD as the large coupling of the strong interaction at long distances or low energy scales does not permit using perturbation techniques. This means that experimental results are mandatory to guide the theory in the right direction. Monte-Carlo generators, which are used to compare the data to what is expected, and are hence essential in any analysis aimed at discoveries, need to model the particle production correctly too. As the multiplicity distribution is heavily influenced by the mechanisms of particle production, its height can constrain the number of particles that models predict, while its shape is related to the detailed dynamics of the partonic interactions, and can thus give an insight on the mechanisms for particle production during proton-proton collisions. Furthermore, the high-energy collisions that the LHC collider provides are predominantly soft events where only a small amount of transverse momentum is exchanged between scatterers, where perturbative QCD is ineffective: predictions therefore rely on QCD phenomenology, which requires to be confronted to minimum-bias measurements. At the new expected collision energy, the interaction of multiple partons for each colliding proton - called Multiple Parton Interaction (MPI) - is expected to play an increasing role, and observables such as the multiplicity distribution or the evolution of the mean momentum with the multiplicity are useful to constrain the several available models. The study of the multiplicity distribution, which is the subject of this thesis, is consequently part of the stepping stone for discovery analysis on the experimental side, while it serves as one of the constraining observable of many phenomenological models on the theoretical side.

The LHC, CERN's latest collider, is the most advanced and powerful accelerator worldwide. It is the result of 60 years of experience as one of the few main particle physics laboratories in the world, from the first built in 1957 to now. The scientific community migrated from LEP, where the $W^{\pm}$ and Z^0 bosons were extensively studied, with a maximum center of mass energy of 900 GeV, to Tevatron (in the United-States) with c.o.m. of 1.8 TeV, which discovered the top quark, and aggregated data to either find or exclude the Higgs - unfortunately with no success either ways -, to finally the LHC which was expected to go to 14 TeV. This never achieved collision energy was finally diminished to 7-8 TeV for now, before the major machine updates of 2015. With this collider, several experiments were also built from the ground up, from which CMS - standing for Compact Muon Solenoid - is part of. Alongside the well-advertised goal of settling the presence of the SM Higgs boson, they were designed to achieve a vast physics program, with many new regions probed. Their implementation around the collider was the fruit of many years of work, from the design, to dedicated RD programs, before trying out complete sections

of the detectors. Finally, their constant pampering by the scientific crews allowed to start extracting some of the physics results from the data provided by the proton beams colliding at a never-achieved-before energy. The Ph.D. work related here explains part of the work done to understand the CMS detector, and with the knowledge gained provide one of the first results obtained with the LHC data. The obtained measurements are compared with previous work at lower and equivalent energies and with model predictions for multi-hadron production.

My involvement with actual particle physics dates from my bachelor years: during a short internship within the JANUS program of the in2p3, I studied - in a very simple and primitive form - the disintegration of the Z^0 boson into muon pairs, using preprocessed data from the DØexperiment at the Tevatron. Then at the end of my first year of Master, I was lucky to be able to participate to the CERN summer school. It was my first time getting introduced to the CERN complex, its inhabitants, its way of life, and the - huge - machines installed there for all the accelerators and experiments. My stay was prior to CMS (and the other LHC experiments) being installed in its cavern: only a portion of the tracker was mounted in the TIFF facility, for testing purposes. After that followed my first encounter with the CMS software, as well as a grasp of what calibrating a detector meant, initially with cosmic rays and later on in real running conditions. Finally, the mandatory internship of the second Master year was the opportunity to be involved in physics analysis, with the study of $t\bar{t}$ channels with CMS, albeit still with simulated data. All this work, although just introductory when compared to a Ph.D. and chosen rather at random or by "chance" along my study years, illustrate quite well the path that has been mine for the analysis presented here: first working with the detector, to understand how it functions, and to help to optimize its performances, leading to the actual physics analysis, once the data can be really well understood. This has been a very interesting learning curve, that I find am very fortunate to have been subjected to.

The first chapter describes the theoretical aspects of multi particle production and the properties of the multiplicity distribution. Relations between the multiplicity and its moments, as well as eventual scaling properties that have been discovered throughout the years with the increase in energy of the experimental apparatus are important to understand the results presented here. A brief mathematical overview of different distributions pertinent to the observed multiplicity is also provided, as it helps to grasp the physics involved in shaping the multiplicity observable.

The second chapter tries to give an overview of the LHC, the collider that produced the data on which the multiplicity analysis was produced. Alongside the accelerator machinery, several experiments have been installed, which CMS is part of. All the subcomponents of the CMS detector are briefly exposed, with the explanation of their purpose, and the technology they use to perform their task.

The third chapter goes in many details about the working principles of one of the CMS subdetectors, the one closest to the collisions: the pixel detector. Understanding precisely how it works is paramount to the comprehension of all the calibration work that is required for the pixel to perform in an optimal way. Each calibration is shortly presented, with an emphasis on the gain calibration that has been one of my main experimental tasks I have gained expertise in during my Ph.D. The analysis performed on cosmic ray data, while waiting for the LHC to be fully operational, and measuring the efficiency of the pixel detector, is also part of this chapter.

The fourth chapter contains a summary of the algorithms used by the CMS software to reconstruct the trajectories of particles crossing the layers of the detector. The differ-

ent objects produced, tracks and their vertices, are the base objects manipulated in the multiplicity analysis.

Finally, the fifth and sixth chapters contain the multiplicity analysis. The first analysis chapter contains results for non-single diffractive events, while the second chapter is the continuation of the previous analysis, with a different event selection not favoring any particular definition of diffraction and is therefore less model dependent. The corrections introduced to correct for experimental biases are fully explained, along with all the associated systematical errors. The multiplicity distributions with different selections is presented, along with their moments, and any eventual scaling is studied. Comparisons to results from previous and current experiments is done, while several models are confronted with the obtained measurements.

The thesis presented here is based on the following list of conferences attended, talked given and publications written:

Conferences:

- International CMS workshop on CRAFT, (Torino, Italy), 11-13/03/09 (attended)
- Lepton-Photon conference, (Hamburg, Germany), 17-22/08/09 (attended)
- 11th ICATPP conference, (Como, Italy), 5-9/10/09 (talk)
- SLHC-PP -2010 conference, (Madrid, Spain), 4-5/02/2010 (talk)
- QCD10 conference (Montpellier, France), 28/06-3/07/2010 (talk)
- ICHEP2010 conference (Paris, Montpellier), 21-28/07/2010 (poster)
- ISMD2010 conference (Antwerpen, Belgium), 21-25/09/2010 (attended)
- MPI@LHC meeting, (Glasgow, Scotland), 28/11-3/12/2010 (talk)
- XLI ISMD conference, (Miyajima, Japan), 26-30/09/2011 (talk)
- HCP2011 conference, (Paris, France), 14-18/11/2011 (talk)

Publications I especially worked on:

- "Commissioning of the CMS Pixel detector with Cosmic Rays", Proceeding of the 11th ICATPP conference, "astroparticle, particle and space physics, detectors and medical physics applications" proceeding, pp 904-909
- "Commissioning and performance of the CMS pixel tracker with cosmic ray muons", CMS Collaboration 2010 JINST 5 T03007
- "Pixel hit efficiency", CMS internal note, first author
- "Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s}$ = 0.9 and 2.36 TeV", The CMS Collaboration , Journal of High Energy Physics (2010) , ISSN: 10298479
- "Charged Hadron Spectra Measurement with the CMS detector in pp collisions at $\sqrt{s}$ = 0.9, 2.36 and 7 TeV", Proceeding of the QCD10 conference, Nuclear Physics B (Proc. Suppl.) 207-208(2010)
- "Charged particle multiplicities at $\sqrt{s}$ = 0.9, 2.36 and 7.0 TeV with the CMS detector at LHC", Proceeding of the ICHEP2010 conference, PoS (ICHEP 2010) 358
- "Charged particle multiplicities in pp interactions at sqrt(s) = 0.9, 2.36, and 7 TeV", J. High Energy Phys. 01 (2011) 079
- "Hadron production at LHC with the CMS detector", Proceeding of the XLI ISMD conference
- "Soft QCD from ATLAS and CMS experiments", Proceeding of the HCP2011 conference

Chapter 1

Multiplicity

The counting of the number of particles produced in each collision event, and its aggregation into a single observable, results in the multiplicity distribution. It holds in an integrated form information about the particle production and eventual correlations. Together with its rather simple extraction from data, its physical impact leads to a rather extensive study of its experimental properties and theoretical consequences over the years. This chapter will briefly introduce the mathematical aspects of the multiplicity distribution, and its representation through moments. Properties like scaling are discussed, along with some tentative explanations of its shape that constrain the particle production models.

1.1 The multiplicity distribution

The multiplicity distribution P_n is given for a multiplicity n by

$$P_n = \frac{\sigma_n}{\sigma_{inel}} , \tag{1.1}$$

where σ_n is the topological cross-section for the production of n charged particles, and the inelastic cross-section being hence the sum of all σ_n:

$$\sigma_{inel} = \sum_{n=0}^{\infty} \sigma_n \tag{1.2}$$

This is done over all possible values of the multiplicity, such that:

$$\sum_{n=0}^{\infty} P_n = 1 \tag{1.3}$$

Another way to obtain the multiplicity distribution is by using its generating function, which is defined by:

$$G(\eta, z) = \sum_{n=0}^{\infty} P_n(\eta)(1+z)^n, \tag{1.4}$$

with z being a dummy variable that is set to zero afterwards, while the pseudorapidity η is defined as:

$$\eta = \frac{1}{2}\ln\left(\frac{p+p_L}{p-p_L}\right) \tag{1.5}$$

$$= -\ln\tan\frac{\theta}{2} \tag{1.6}$$

where p (p_L) is the total (longitudinal) momentum of the particle and θ the angle between the particle and the beam axis.

1.2 Moments

The multiplicity distribution can also be represented using its moments, as all together they contain the whole information it carries. In practice, only the first few orders are computed since uncertainties on higher orders render them useless due to lack of statistics. The simpler low-order moments giving general characteristics of the multiplicity distribution are its mean $\langle n\rangle$ and the dispersion D. The mean is simply calculated by

$$\langle n\rangle = \sum_{n=0}^{\infty} n\cdot P_n \tag{1.7}$$

The D-moments are computed as follow:

$$D_q = \langle\ (n-\langle n\rangle)^q\ \rangle^{1/q}\quad , \tag{1.8}$$

where the D-moment of order 2, $D\equiv D_2=\sqrt{\langle n^2\rangle-\langle n\rangle^2}$, is usually referred to as dispersion, and estimates the width of the multiplicity distribution.

To ease the representation of the moments, it is common to use a normalization term dependent on their order, usually using $\langle n\rangle$. Hence every normalized moment of rank q is just the value of the normal - or unnormalized - moment divided by $\langle n\rangle^q$. To be consistent with conventions, in what follows the unnormalized moments will be written with a $\sim$ on top, to be easily distinguished from their normalized counterparts.

The ordinary statistical normalized C-moments are computed by the formulas

$$C_q = \frac{\sum_{n=0}^{\infty} P_n n^q}{\langle n\rangle^q} \tag{1.9}$$

$$= \frac{\langle n^q\rangle}{\langle n\rangle^q} \tag{1.10}$$

while the normalized factorial moments can be defined using

$$F_q = \frac{\sum_{n=0}^{\infty} P_n n(n-1)...(n-q+1)}{\langle n\rangle^q} \tag{1.11}$$

$$= \frac{\langle n(n-1)\cdots(n-q+1)\rangle}{\langle n\rangle^q}\quad . \tag{1.12}$$

Although the ordinary moments C_q are well known statistical quantities, their physical interpretation is not intuitive. The factorial moments F_q are better suited and much easier

to interpret, since for identical particles they correspond to the normalized phase-space integral of the q-particle inclusive density function ρ_q over a pseudorapidity domain in phase space denoted dη:

$$\begin{aligned}
\int \frac{1}{\sigma}\frac{d\sigma}{d\eta}d\eta &= \int \rho_1(\eta)d\eta = \langle n\rangle = \tilde{F}_1, \\
\int \frac{1}{\sigma}\frac{d^2\sigma}{d\eta_1 d\eta_2}d\eta_1 d\eta_2 &= \int \rho_2(\eta_1,\eta_2)d\eta_1 d\eta_2 = \langle n(n-1)\rangle = \tilde{F}_2, \\
&\vdots \\
\int \frac{1}{\sigma}\frac{d^q\sigma}{d\eta_1\cdots d\eta_q}d\eta_1\cdots d\eta_q &= \int \rho_q(\eta_1,\cdots,\eta_q)d\eta_1\cdots d\eta_q = \langle n(n-1)\cdots(n-q+1)\rangle = \tilde{F}_q.
\end{aligned} \tag{1.13}$$

Therefore, the factorial moments include in integrated form all correlations contained in the system of particles considered. Hence if all particles are produced independently, the former expressions lead to the multiplicity distribution being a Poissonian and $F_q = 1$ for all $q > 1$. If the particles are positively correlated, the P_n distribution becomes broader than a Poissonian and the normalized factorial moments F_q are larger than one.

The normalized cumulant moments K_q can be recursively derived from the factorial moments using the equation 1.14. They express the normalized phase-space integral over the q-particle correlation function, and as such reflect the genuine higher order correlations between q-particles.

$$K_q = F_q - \sum_{i=1}^{q-1}\binom{q-1}{i}K_{q-i}F_i \tag{1.14}$$

As both factorial moments and cumulants tend to increase rapidly for high q orders, it is convenient to also compute the H_q moments as ratio of K_q over F_q:

$$H_q = \frac{K_q}{F_q}; \tag{1.15}$$

They describe the genuine q-particle correlation relative to the q-particle density. They have been extensively studied before [1], and appear quite naturally as the solution of the QCD equations for the generating functions [2]. Higher-order QCD calculations also predict their quasi-oscillation around 0 with increasing order.

The factorial moments and cumulants can be derived from the generating function introduced in equation 1.4 using the relations

$$F_q = \frac{1}{\langle n\rangle^q}.\frac{d^q G(z)}{dz^q}|_{z=0}; \tag{1.16}$$

$$K_q = \frac{1}{\langle n\rangle^q}.\frac{d^q \ln G(z)}{dz^q}|_{z=0}; \tag{1.17}$$

Those can be inverted to infer the generating function $G(z)$ using every orders, yet another proof of them containing as a whole all the information held in the multiplicity distribution:

$$G(z) = \sum_{q=0}^{\infty} \frac{z^q}{q!} \langle n \rangle^q F_q; \tag{1.18}$$

$$\ln G(z) = \sum_{q=1}^{\infty} \frac{z^q}{q!} \langle n \rangle^q K_q. \tag{1.19}$$

The multiplicity distribution P_n and its ordinary moments C_q can likewise be derived from the generating function $G(z)$ using the formulas

$$P_n = \frac{1}{n!} . \frac{d^n G(z)}{dz^n}|_{z=-1}; \tag{1.20}$$

$$C_q = \frac{1}{\langle n \rangle^q} . \frac{d^q G(e^z - 1)}{dz^q}|_{z=0} \quad . \tag{1.21}$$

All moments are connected by definite relations that can be inferred from the equations above. For instance, cumulants can be computed using factorial moments through equation 1.14, and thus factorial moments can also be computed solely with cumulants, as follows for the first few orders

$$F_1 = K_1 = 1; \tag{1.22}$$

$$F_2 = K_2 + 1; \tag{1.23}$$

$$F_3 = K_3 + 3K_2 + 1; \tag{1.24}$$

$$F_4 = K_4 + 4K_3 + 3K_2^2 + 6K_2 + 1; \tag{1.25}$$

$$F_5 = K_5 + 5K_4 + 10K_3K_2 + 10K_3 + 15K_2^2 + 10K_2 + 1 \quad . \tag{1.26}$$

In the same way, factorial moments F_q and ordinary moments C_q can be derived from each other. They have been found to differ by terms depending only on the mean multiplicity $\langle n \rangle$, as illustrated for the order 2

$$F_2 = \frac{\langle n(n-1) \rangle}{\langle n \rangle^2} = C_2 - \frac{1}{\langle n \rangle} \quad , \tag{1.27}$$

and as such they both have a different energy dependence, since $1/\langle n \rangle$ is energy dependent.

1.3 Feynman scaling

In an article published in 1969 [3], Feynman predicted that for hadron collisions at very high energies, the mean total number of particles rises logarithmically with the center-of-mass energy $\sqrt{s}$ of the collisions (s being one of the Mandelstam variables):

$$\langle N \rangle \propto \ln \sqrt{s} \propto \ln W \quad \text{with} \quad W = \sqrt{s}/2 \quad . \tag{1.28}$$

His conclusions are based on phenomenological arguments concerning the exchange of quantum numbers between the colliding particles: for two body interactions, currents carrying quantum numbers - like the isospin for instance - are exchanged when reversing direction from one particle to the other. As the currents have fields as sources, those

fields are expected to radiate and produce particles, much like bremsstrahlung radiations. Using Lorentz transformation, at very high energies (W $\rightarrow \infty$) the fields get narrower in z, leading to a δ-function in the z direction. Therefore the field energy is uniformly distributed in p_z, and the number of particles with a given mass and transverse momentum per longitudinal momentum interval p_z depends on the energy E as

$$\frac{\mathrm{d}N}{\mathrm{d}p_z} \sim \frac{1}{E} \quad . \tag{1.29}$$

Extending the argument above, the probability P of finding a particle of kind i with mass m, transverse and longitudinal momentum p_T and p_z, and energy E, can be expressed as

$$P(p_T, x = \frac{p_z}{W}) \quad = \quad f_i(p_T, x)\frac{\mathrm{d}p_z}{E}\mathrm{d}^2p_T \tag{1.30}$$

where

$$E \quad = \quad \sqrt{m^2 + p_T^2 + p_z^2} \tag{1.31}$$

$$= \quad W\sqrt{x^2 + \left(\frac{m_T}{W}\right)^2} \quad , \tag{1.32}$$

$$m_T \quad = \quad \sqrt{m^2 + p_T^2} \tag{1.33}$$

The function $f_i(p_T, x)$ is a structure function describing the distribution of particles. Feynman's hypothesis states that f_i becomes independent of W for high energies. This assumption is known as *Feynman scaling* and f_i is called scaling function or Feynman function. The $x = p_z/W$ variable, now called *Feynman-x*, represents the fraction of longitudinal momentum p_z of the particle and the total energy W of the incident particle. Integration of equation 1.30, detailed here [4] in appendix, results in $\langle N \rangle \propto \ln W$.

Considering that the maximum reachable rapidity in a collision increases also with $\ln s$, and assuming the particles are evenly distributed in rapidity, it follows that $\mathrm{d}N/\mathrm{d}y$ is a constant, thus the height of the rapidity distribution at mid-rapidity - usually called plateau of the rapidity distribution - is independent of $\sqrt{s}$. Likewise, the pseudorapidity at mid-rapidity $\mathrm{d}N/\mathrm{d}\eta|_{\eta=0}$ is approximately constant under the condition that the Feynman scaling holds. However the transformation from rapidity to pseudorapidity needs to be taken into account, as it is dependent on the transverse mass which in turn depends weakly on the energy. This causes a dip around $\eta \approx 0$ in the pseudorapidity distribution which is not present in its rapidity counterpart.

In 1971, after the first colliding experiments, it was observed that the pseudorapidity density at mid-rapidity ($\eta \approx 0$) was in fact increasing with the energy, thus breaking the Feynman scaling. It suggested that hadron interactions can not be explained solely on weak parton interactions, which was confirmed later on with the introduction of contributions from gluon interactions.

1.4 Koba-Nielsen-Olesen (KNO) scaling

Although Feynman's scaling was observed to be violated in data coming from the first colliding experiments, Koba, Nielsen and Olesen proposed in 1972 a new scaling of the multiplicity distribution [5], that they derived from the Feynman scaling and the observation of data. It was later on proven that although Feynman scaling is sufficient to infer this new scaling, it was in no way necessary. It is usually referred to as KNO scaling, in deference to its authors, and is derived by calculating the expression

$$\langle n(n-1)\cdots(n-q+1)\rangle = \int f^{(q)}(x_1, p_{T,1}; ...; x_q, p_{T,q}) \frac{\mathrm{d}p_{z,1}}{E_1} \mathrm{d}p_{T,1}^2 ... \frac{\mathrm{d}p_{z,q}}{E_q} \mathrm{d}p_{T,q}^2 \quad (1.34)$$

$f^{(q)}$ describing q-particle correlations (q particles with energy E_q , longitudinal momentum $p_{z,q}$, transverse momentum $p_{T,q}$, and Feynman-x x_q). Integration by parts is performed for all x_i and it is proven that the resulting function is uniquely defined by moments, as it is the expression for the $\tilde{F}_q$ used in equation 1.11 without the normalization. This yields a polynomial in $\ln s$, and substituting it with $\langle n\rangle$ - since Feynman scaling is assumed to hold - results in the multiplicity distribution P_n scaling as

$$P_n = \frac{1}{\langle n\rangle} \Psi\left(\frac{n}{\langle n\rangle}\right) + \mathcal{O}\left(\frac{1}{\langle n\rangle^2}\right) \quad (1.35)$$

with $z = n/\langle n\rangle$ and $\Psi(z) = \langle n\rangle P_n$ being called KNO variables. $\Psi(z)$ is thus independent of the energy, and plotting the multiplicity distribution using the KNO variables at different center of mass energies lead to all curves being on top of one another. However, $\Psi(z)$ is dependent on the type of reactions as well as the kind of particle measured.

As specified, the moments define the function $\Psi(z)$ uniquely: the ordinary moments C_q can be expressed as

$$C_q = \int_0^\infty z_q \Psi(z) \mathrm{d}z \quad (1.36)$$

where replacing z by $n/\langle n\rangle$ leads to equation 1.9. If the KNO scaling holds, the moments are independent of the energy, which will be used later here to test whether or not the scaling holds at the LHC energies.

1.5 Multiplicity models

1.5.1 Single Negative Binomial Distribution (NBD) fit

Very quickly after the first colliding experiments, it was observed that the multiplicity distribution deviated from a Poissonian, as correlations were involved between the emitted particles. Therefore, new distributions were explored to explain the shape of the multiplicity, and it was strikingly the negative binomial distribution that described well the multiplicity, this for a wide range of processes and energies.

The negative binomial distribution gives the probability of n fails and $k-1$ successes in a sequence of independent Bernoulli experiments before the k'th success, with a probability p of success :

$$P(n;p;k) \quad = \quad \binom{n+k-1}{n} (1-p)^n \; p^k \quad . \tag{1.37}$$

For multiplicity distributions, the two independent parameters used are generally the mean multiplicity $\langle n \rangle$, and the parameter k which describes the shape of the distribution. Using $p = k/(k + \langle n \rangle)$, the NBD formula can hence be expressed as

$$P(n;\langle n \rangle;k) \quad = \quad \binom{n+k-1}{n} \left(\frac{\langle n \rangle}{\langle n \rangle + k}\right)^n \left(\frac{k}{k+\langle n \rangle}\right)^k \tag{1.38}$$

The Figure 1.1 shows several examples of NBD distributions for a fixed $k = 2$ on the left, and for a fixed $\langle n \rangle = 10$ on the right. From equation 1.38, various types of distributions can be obtained depending on the parameter k, and are represented in Figure 1.1: a Poissonian distribution is obtained in the limit of $k \to \infty$, $k = 1$ gives a Bose-Einstein distribution, while a negative integer value generates a binomial.

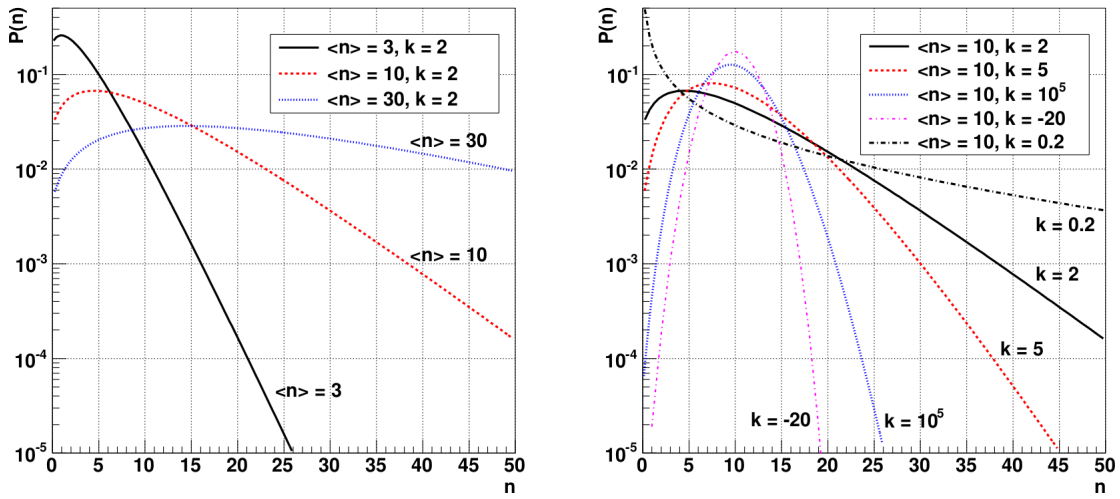

Figure 1.1: Several examples of negative binomial distributions taken from [4]

In order to use negative integer values for the k parameter, the binomial factor in equation 1.38 is described using Gamma functions:

$$\binom{n+k-1}{n} \quad = \quad \frac{(n+k-1)!}{n!\;(k-1)!} = \frac{\Gamma(k+n)}{\Gamma(n+1)\Gamma(k)} \tag{1.39}$$

$$\quad = \quad \frac{(n+k-1)\cdot(n+k-2)\cdot\ldots\cdot k}{\Gamma(n+1)} \quad , \tag{1.40}$$

which leads to another often seen formula for the negative binomial distribution,

$$P(n;\langle n \rangle;k) \quad = \quad \frac{\Gamma(k+n)}{\Gamma(n+1)\Gamma(k)} \left(\frac{\langle n \rangle}{k}\right)^n \left(1+\frac{\langle n \rangle}{k}\right)^{-n-k} \quad . \tag{1.41}$$

Its generating function, expressed in function of the same two parameters, leads to a rather simple relation:

$$G^{(NBD)}(z) = \left(1 - \frac{z\ \langle n \rangle}{k}\right)^{-k} \quad . \tag{1.42}$$

From this equation, the factorial moments, the cumulants and their ratio, all described in Section 1.2, can be obtained:

$$F_q = \frac{\Gamma(k+q)}{k^q \Gamma(k)} \tag{1.43}$$

$$K_q = \frac{\Gamma(q)}{k^{q-1}} \tag{1.44}$$

$$H_q = \frac{\Gamma(q)\ \Gamma(k+1)}{\Gamma(k+q)} \quad . \tag{1.45}$$

Their shape and behavior for different sets of parameters can be seen in [2]. It is interesting to notice from these equations that those moments are expressed solely using the k parameter, the mean $\langle n \rangle$ never intervening: for instance the first factorial moment, of order $q = 2$, can be simply expressed as $F_2 = 1 + \frac{1}{k}$. Therefore the KNO scaling for a NBD only depends on the behavior of k: if it is independent of the energy, then the KNO scaling holds. This is why many experiments observed its evolution with $\sqrt{s}$, in place or in addition to computing the moments. The relation between KNO scaling and k can also be inferred from the NBD formulation in terms of the KNO variables previously defined:

$$\Psi^{(NBD)}(z) = \frac{k^k}{\Gamma(k)} z^{k-1} e^{-kz} \quad . \tag{1.46}$$

It is not yet understood why the negative binomial fits so well the multiplicity distribution at low energy. Several models, such as the clan [6–8] model, tried to initiate an explanation.

1.5.2 Double NBD fit

As aforementioned, older experiments have their multiplicity data very well described by the negative binomial distribution. However the multiplicity distribution started to deviate from a NBD when collisions occurred at higher energies: the CERN experiment UA5 observed a shoulder developing for $\sqrt{s} = 546\,\text{GeV}$ which became even more pronounced for $\sqrt{s} = 900\,\text{GeV}$ [9], while the same feature was discovered in e^+e^- collisions later on and thoroughly discussed by Giovannini [10]. The multiplicity distributions for higher energies were found to be well fitted by a combination of two NBDs, and even three NBDs for the latest data. The lower energy data seem to agree slightly better with the two NBD hypothesis too. There are several attempts to explain it, from which two are detailed here: a division between soft and semi-hard components, or the multiple parton interactions (MPI) framework.

After double NBDs were proven to describe the high energy multiplicity distribution in data, Giovannini and Ugoccioni argued that each NBD represented a different component

of the collisions [11]. The first NBD describes the soft part of the collision, while the second one - resulting in the shoulder structure - describes the semi-hard part. It has to be understood as two different classes of events, rather than a difference in their production mechanism: the semi-hard events contain mini-jets (described by the UA1 collaboration as a group of particles having a total transverse energy larger than 5 GeV) whereas the soft events are devoid of them. Consequently, they are both independent from each other, no interfering terms needing to be considered, and the final distribution is just the sum of the two NBD distributions:

$$\begin{aligned} P_n &= \alpha_{soft} \cdot P_n^{(NBD)}(\langle n \rangle_{soft}; k_{soft}) \\ &\quad + (1 - \alpha_{soft}) \cdot P_n^{(NBD)}(\langle n \rangle_{semi-hard}; k_{semi-hard}) \end{aligned} \tag{1.47}$$

The soft component follows Feynman scaling as well as KNO scaling, whereas the semi-hard part is strongly energy dependent and violates KNO scaling. When applying this to observed data, the mean multiplicity for the first NBD, $\langle n \rangle_{soft}$, was measured to be about half that of the second one, $\langle n \rangle_{semi-hard}$.

Another way to explain two NBDs shaping the multiplicity distribution is to introduce multiple parton interactions during a single *pp* collision. At LHC energies, parton-parton interactions carrying a considerable momentum fraction of the colliding protons, called hard scattering, contribute significantly to the total charged particle multiplicity. They result in QCD-jets - much like mini-jets introduced by UA1 - and can be described by perturbative QCD. On the other hand, soft interactions, usually defined by a transverse momentum smaller than a given threshold $p_{T,0}$, require either calculation techniques to counter the divergences at $p_{\mathrm{T}} \to \infty$, or specific models for soft-particle production [12–14]. In the latter case, two or more parton-parton scatterings take place within a single proton collision. These models explain many observed features of *pp* collisions, and MPIs have lately been widely accepted to interpret underlying events (UE, for which a comprehensive depiction can be found in [15]). The several NBDs result from those multiple parton interactions, sometimes directly like in the IPPI model [16] where each independent parton interaction leads to a NBD contributing to the total multiplicity, or in the dual parton model [17] where the number of soft gluon exchanges - represented by virtual pomerons - are again directly correlated to the number of NBDs.

1.6 Processes in *pp* collisions

The majority of proton-proton collisions end-up as soft interactions, divided into elastic and inelastic scatterings. The elastic scatterings are of no use in this thesis, as they are usually not recorded, the protons following a very forward trajectory without much particle production but the eventual radiations. The soft inelastic interactions are then divided into diffractive processes and the non-diffractive ones. During the scattering, one of the partons can dissociate into a shower of particles, which are overall bound by the conservation of quantum numbers. A diagram explaining such occurrences is shown in Figure 1.2 (left), and is called Single-Diffractive (SD). When both partons dissociate due to their interactions, the system is called Double-Diffractive (DD), and is represented in Figure 1.2 (right). The diffraction process always produces a large gap between the two resulting systems, which is devoid of any particles. The remaining Non-Diffractive (ND) events are cases where large transfers of momentum as well as quantum numbers happen between the two partons, and usually results in more particles created, with no pseudorapidity gap apparent.

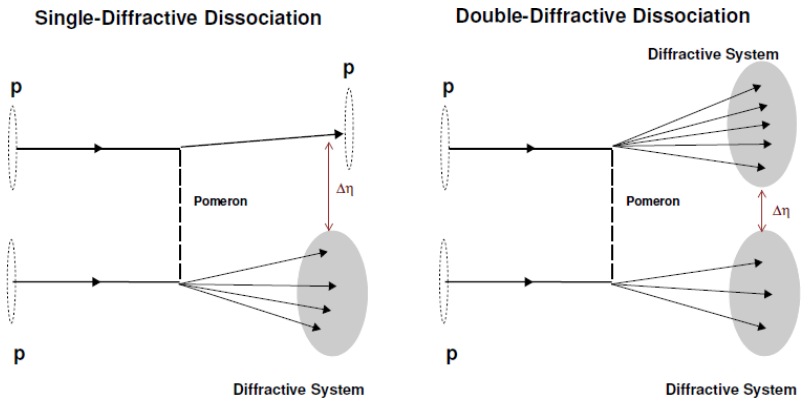

Figure 1.2: Schematic representation of two protons colliding and dissociating into a single-diffractive system (left) and a double-diffractive system (right). Taken from [18].

The total cross-section can therefore be calculated as:

$$\sigma_{tot} = \sigma_{el} + \sigma_{inel} = \sigma_{el} + \sigma_{SD} + \sigma_{DD} + \sigma_{ND} \tag{1.48}$$

Chapter 2

The LHC collider and the CMS experiment

In the aftermath of the second world war, world-wide cooperation grew bigger in all domains, and international organizations such as the UN, or UNESCO, started blooming everywhere, to strengthen relations between people and begin rebuilding countries that were devastated be the war. In Europe, several renowned scientists took this opportunity to propose a plan for a european laboratory for nuclear physics, which would not only unite scientists, but also help them share the increasing costs of this field of research. The CERN, Conseil Européen pour la Recherche Nucléaire, was officially borne on 29th September 1954, after its convention was ratified by the 12 founding Member States. The first particle accelerator, the SynchroCyclotron (SC), built in 1957, was rapidly followed by others: the Proton Synchrotron (PS) in 59, the Intersecting Storage Rings (ISR) in 71, the Super Proton Synchrotron (SPS) in 76 which was CERN's first giant ring accelerator, and the Large Electron-Positron (LEP) in 89, for which a 27km long tunnel was excavated under the Franco-Swiss border. Many of those ground-breaking machines are still in action, and through technological upgrades, constitute CERN's chain of particle acceleration. Along with major improvements in many fields related to particle accelerators and particle detection, discoveries such as the W and Z bosons -which was rewarded with a Nobel Prize- were achieved throughout the years. Nowadays, CERN is run by 20 member states, while many non-member states and organizations participate, leading in more than 10000 scientists from 608 universities and 113 nationalities using the CERN facilities for their research.

In December 1994, the CERN council approved the LHC project, which would require to solve many technological issues in order to build it and achieve its physics program. In 2000, after fevered discussions on whether more time should be given to either find or exclude the Higgs boson, the LEP was stopped to make place to the new LHC accelerator which would use its already built tunnel. Along with the accelerator development, R&D for 4 detectors started: the ALICE detector dedicated to heavy-ion physics, the LHCb for b-quark physics and CP violation studies, and 2 multi-purpose detectors, ATLAS and CMS. While ALICE and LHCb were installed in previously dug caverns, ATLAS and CMS needed the construction of 2 new caverns to accommodate their need of space. Two much smaller experiments, TOTEM and LHCf, were later approved: they are mainly dedicated to forward physics, and were respectively integrated in the caverns of CMS and ATLAS.

In the following section, I will give a short description of the LHC machine, then I will enter in detail in the CMS experiment apparatus.

2.1 LHC

2.1.1 Description of the LHC Machine

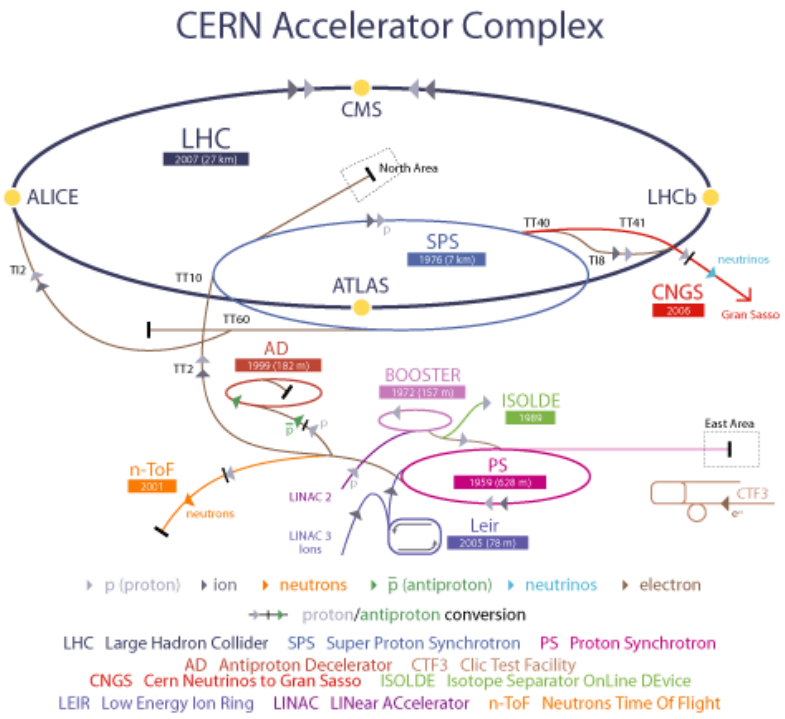

Figure 2.1: Schematic view of the CERN accelerator chain and all experiments present on site.

As shown in Figure 2.1, the CERN complex accommodates many physics accelerators and experiments. The Large Hadron Collider, though being the mightiest of all accelerators, is only the last element in the chain of acceleration that leads to collisions. As it is a *pp* collider, the first step is to strip bare hydrogen atoms from their electrons. Then starts the acceleration process: first in the LINAC linear accelerator, to the Booster, to the PS where the protons are separated in bunches, to the SPS where they reach an energy of 451 GeV per beam, before finally being injected into the LHC rings.

The LEP tunnel, which now hosts the LHC accelerator, is 27km long, and runs underground at around 100 meters deep, depending on the location. It is composed of eight straight sections and eight arc sections. As the LHC is a hadron collider, it emits much less synchrotron radiations than the LEP accelerator did as it was an *ee* collider, and as such could have had longer arc sections. But using the current tunnel with its inadequacies was a good compromise when compared to the cost and time of making a whole new tunnel suited for a hadron collider. Four main caverns holding the experiments are present in the straight sections, were the particle beams collide.

One of the main feature of the LHC accelerator is its magnets. There is a total of 9600 magnets around the collider ring, each with its specific purpose. The main magnets are the 1232 cryodipoles located in the arcs: they are the ones bending the proton beams around the tunnel, and are the results of many years and breakthroughs in R&D. The choice of using same sign hadrons to accelerate, contrary to Tevatron for instance, was dictated by the very high beam intensity required, which is not achievable with anti-particle beams.

As such, the 2 beams of hadrons going in the opposite direction cannot use the same magnetic field, and would demand having 2 separated magnets and their cooling system, which the space in the tunnel would not allow. Therefore NbTi super-conducting twin bore magnets were preferred, permitting the 2 beams to be hosted in a single magnet. Those dipoles measure 14.2 meters long, and produce a magnetic field of 8.3T. They require to be cooled down to 1.9K using super-fluid helium. This very low temperature is needed to achieve the required magnetic field, but at the expense of provoking easily more quenches, as the threshold temperature is simpler to attain. Built in 3 different countries, they were harshly tested at CERN before being put in production: each dipole must produce similar magnetic fields down to the 10^{-4} level. The other magnets have different types and functions: many quadrupoles are used to focus and defocus the beams, while sextupoles and octupoles are scattered all over the tunnel.

The acceleration process is executed by 400 MHz klystrons, with parts like the power converters being reused from the LEP era. Many beam instrumentations are also present all along the accelerator, for purposes such as beam position measurement, transverse and longitudinal profile measurement, luminosity monitors.

2.1.2 Performance Goals of the LHC

The LHC was meant to achieve collisions with a total center-of-mass energy of 14 TeV, but this goal was downsized to 7 TeV in 2009-2011 and 8 TeV in 2012, as doubts about the magnets were raised. The goal for *pp* collisions was to hold 2808 bunches at a time per beam in the LHC ring, with a 24.95ns bunch spacing, and 1.15x10^{11} protons per bunches. Those three parameters are the main parameters used to calculate the instantaneous luminosity (see Equation 2.1, where f is the revolution frequency of a proton around the LHC ring, n_b is the number of bunches per beam, N_b is the number of protons per bunch, and A is the effective overlap area of the beams), and as such maximizing any of them would increase the luminosity, which is extremely important in rare particle searches. This luminosity increased throughout the LHC operation, as the experts controlled much better their machines, and were slowly able to increase all parameters closer to their nominal values, up to the design goal of $L = 10^{34} cm^{-2} s^{-1}$. For the experiments, high luminosity leads to more proton interactions for each bunch collisions, which also increased throughout the years to reach more than 25 Pile-Up (PU) interactions per collisions.

$$L = f n_b \frac{N_b^2}{A} \tag{2.1}$$

For the heavy ion design goals, fully stripped lead ions accelerated to 2.76TeV per nucleon for a total center-of-mass energy of 0.15PeV were meant to achieve a nominal luminosity of 1.0x$10^{27} cm^{-2} s^{-1}$.

2.1.3 The LHC Operation so far

The official start of the LHC accelerator was on the 10 September 2008, were the first beams circulated at the SPS injection energy of 450 GeV, and the CMS recorded their first collisions. But on the 19 September 2008, a magnet quench due to a faulty electrical connector between two dipoles created a huge helium leak wave in the tunnel, resulting in 53 magnets needing repairs. The inquiry of this incident lead to the discovery of poorly manufactured connectors, which resulted in the decision, to decrease the

collision energy from initially 10 TeV to maximum 7 TeV (and later in 2012 8 TeV), as it was deemed too dangerous to operate higher. The experiments used the one-year it took to repair everything to commission their detectors and to analyze the Cosmic Rays: it is during this period that my work on the pixel commissioning started, along with the pixel efficiency measurements I participated in. The LHC operations restarted on the 20 November 2009,with the first 7 TeV collisions happening the 30 March 2010. At the end of the year, during November, the first run with lead ions started, up to the winter shutdown beginning of December. The 2011 run started mid-march, during which the LHC became the world's highest-luminosity hadron accelerator achieving a peak luminosity of $4.67\text{x}10^{32} cm^{\hat{a}\acute{L}\check{S}2} s^{\hat{a}\acute{L}\check{S}1}$, beating the Tevatron's previous record. The first inverse femtobarn of data collected by ATLAS and CMS was reached on the 17 Jun 2011, while the 5 fb^{-1} was reached on the 23 Oct 2011. The first collisions with stable beams in 2012 started after the winter shutdown in April, with an energy increased to 4 TeV per beam.

2.2 The Compact Muon Solenoid Experiment

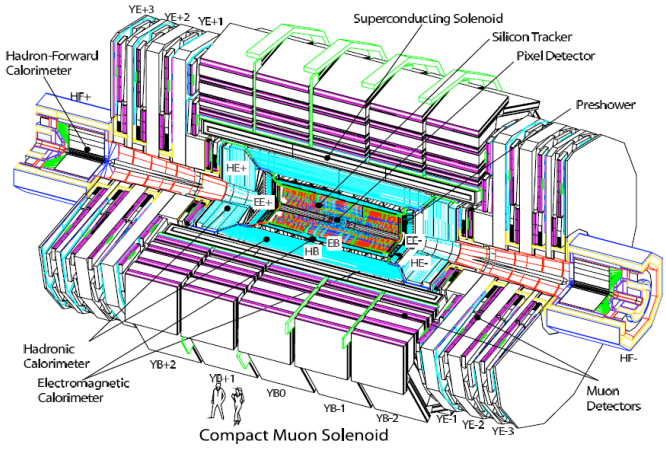

Figure 2.2: Schematic view of the CMS experiment with all its subdetectors.

The Compact Muon Solenoid (CMS) experiment is with ATLAS one of the two multi-purpose experiments at the LHC. It is located at point 5 along the LHC ring, at the opposite side of ATLAS, in a cavern built exclusively for it. It weighs 12500 tons, is 21.6 m long and 14.6 m high. It uses the common onion-layer layout typical for such experiments, with layers of subdetectors forming barrels, and disks assembled in end-caps on each side to close the structure and provide maximum coverage. Figure 2.2 presents a global view of the CMS detector, while Figure 2.3 shows a slice in the plane perpendicular to the beam propagation of all subdetectors and the particles they each detect. As it name suggests, the central component of CMS is a solenoid, providing the magnetic field needed to bend the particles and be able to measure their momentum. It encases a tracker, which is the detector closest to the collision point and which gives a very precise trajectory of the charged particles. Then comes the calorimeters, first the electromagnetic one which detects photons and electrons, and the hadronic one which also permits to detect neutral hadrons. Both calorimeters are also encompassed in the solenoid. The muon chambers alternated with return yokes to contain the magnetic field complete the layout of the CMS

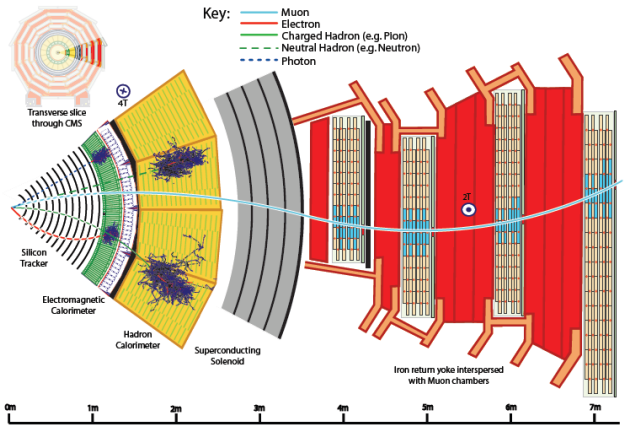

Figure 2.3: Slice view of the CMS detector presenting each subdetectors and the particles they detect.

detector. As they are exposed to high luminosity, especially for the tracker, all electronic had to be radiation-hard.

Many of those subdetectors were first assembled on different sites, in order to test their components. For instance, part of the tracker was built in the TIFF facility at CERN, and my first encounter with the CMS experiment and its research was during the CERN summer school of 2006, where I first had to understand and play with part of the tracker commissioning. Before the first collisions, each subdetectors had to be lowered to the cavern, which proved to be quite challenging, as each of them were very sensitive and had to be lowered for 100 meters down a not-so-big shaft.

CMS uses a global coordinate system: the origin of the axis is the Interaction Point (IP); the x-axis is directed towards the inside of the LHC ring; the y-axis points upwards; finally the z-axis is parallel to the beam direction to the jura. The θ and ϕ are the standard polar angles defined with respect to the xyz-axis. The pseudorapidity η is often used in place of *theta*, and has been previously defined in Equation 1.5.

2.2.1 The Tracker

The CMS tracker is the subdetector the closest to the impact point. As such, it is the one submitted to the harshest environment, which was a great concern during its conception. Its main purpose is to detect charged particle, measure with great precision their trajectories and their momentum, reconstruct the location of the primary vertices the interaction came from, identify several primary vertices as the luminosity, thus the PU, delivered by the LHC machine increases over time, and finally possibly find secondary vertices from long-life decaying particles. To achieve this, the high granularity of the tracker was of primary importance, and use of on-board electronic chips as well as local cooling was chosen, though it had to be balanced with the need of using as less non-sensitive material as possible. Indeed, increasing the material budget of the tracker has drastic effects like multiple scattering, bremsstrahlung, photon conversions and nuclear interactions, which would create unnecessary noise while also impoverishing the energy resolution of the next subdetectors.

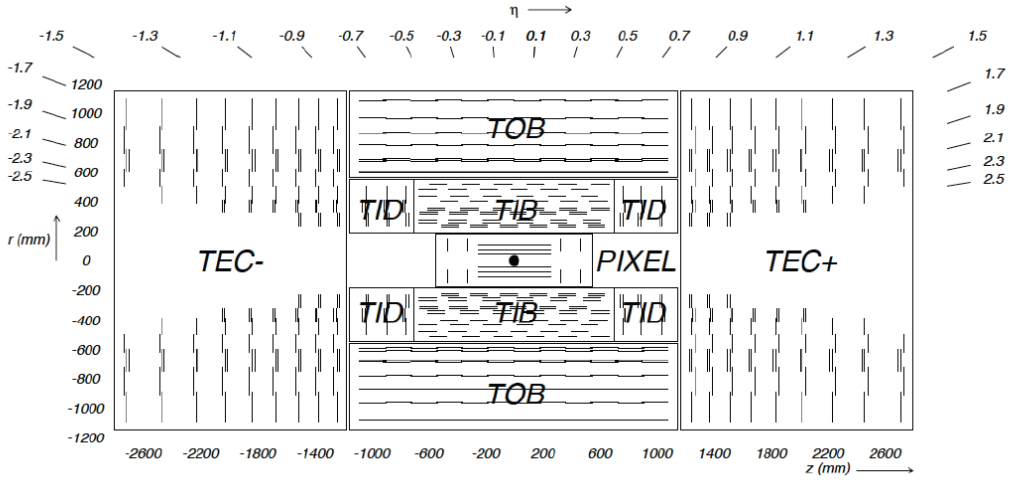

Figure 2.4: Longitudinal sketch of the tracker subdetector. Each line represents a module, while double lines indicate back-to-back modules delivering stereo hits.

The CMS people decided to use a full silicon-based tracker, which was quite a feat at the time as silicon material was much more expensive than it is now. Reuse of military-grade technology for radiation-hard components was employed to ensure a correct lifetime for the detector. The tracking system is divided in two subdetectors: the pixel detector and the Silicon Strip Tracker (SST) detector. The sketch in Figure 2.4 presents the tracker detectors with all its subdivisions, as well as the layout of the modules. The tracker covers the complete ϕ range, and up to $|\eta| < 2.4$.

The pixel detector comprises a barrel with three 53 cm long layers at a radii of 4.4, 7.3 and 10.2 cm respectively, and end-caps with two disks on each side at $|z| = 34.5$ cm and 46.5 cm. The most basic unite of sensitive material is the pixel, which is a 100×150 μm^2 np semi-conductor, where a high voltage is applied to amplify and migrate to the detector part the charge deposited by the particle crossing the silicon material. There are 66 millions of such pixels in the pixel detector, which is far too much to read at once, especially at the high rates of 40MHz the LHC is designed for. Therefore the pixel detector uses a double-column zero-suppression read-out, were the information contained in double-columns of pixels are retrieved only if at least one pixel collected a charge. It can measure very precisely the vertex position, and was arranged such that at least two layers/disks are crossed by any particle with $|\eta| < 2.2$. A more detailed description of how the pixel detector works can be found in Section 3.

The SST detector is also divided into barrel and end-caps, each of them being also splitted in two subcomponents. The Tracker Inner Barrel (TIB) has 4 layers, while the Tracker Outer Barrel (TOB) is separated into 6 layers. Concerning the end-caps, the Tracker Inner Disks (TID) comprises of 3 disks per side, while the Tracker Enc-Caps (TEC) is made of 9 disks on each side. It uses also an np semiconductor with strips, with module thicknesses of 300 and 500 μm for the inner and outer barrels respectively, with again a high voltage applied to collect the drifting charges. Presence of stereo modules, with strips on each side having an angle in order to determinate much more precisely the position of the particle passing through, are present in several layers, and are represented as double lines in Figure 2.4. The SST consists of more than 10 million channels, which makes an total detector surface of close to 198 m^2.

As the tracker is subjected to such harsh environment, its life-time and efficiency decreases significantly with increasing total luminosity. That is why during the long shut-down of the LHC in 2014-2015, the tracker has been modified before its complete upgrade planned in 2017: parts of the pixel and SST were replaced with fresh modules, while a fourth pixel layer is planned to be added (see Pixel Upgrade in Section 3.1.4).

2.2.2 The electromagnetic calorimeter

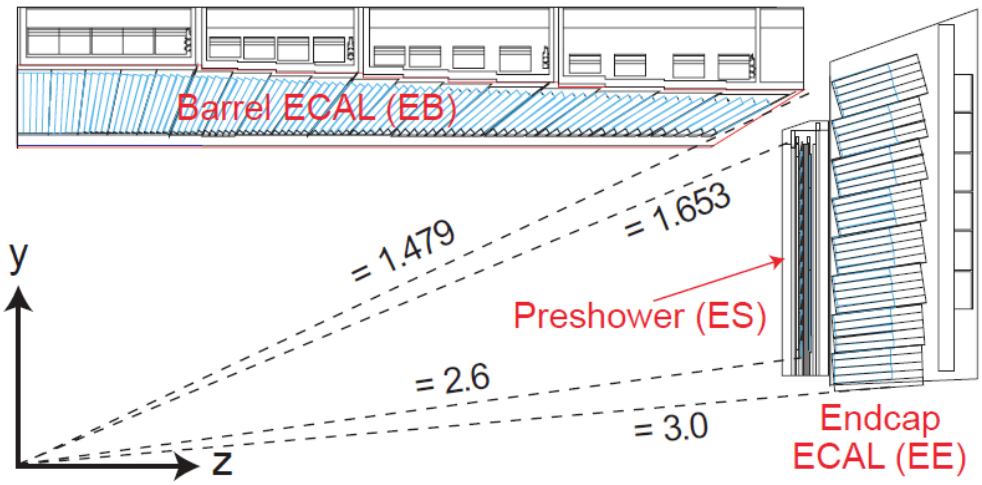

Figure 2.5: Longitudinal sketch of one quarter of the ECAL subdetector.

The purpose of the electromagnetic calorimeter is to collect the energy of the particles interacting electromagnetically: the photons, which are not detected at all in the tracker, and the electrons. The design of the ECAL was driven by the goal of making the best calorimeter possible to look efficiently for the Higgs boson decaying in two photons. Thus a fine granularity and a good resolution were required, leading to the choice of lead tungstate ($PbWO_4$) crystals as the scintillator base material: its small Moliere radius of 2.2 cm and short radiation length of 0.89 cm allowed the showers to be contained in a very compact yet with high granularity calorimeter. The light emitted by the shower in then collected with Avalanche Photo-diodes (APDs) in the barrel, while Vacuum Photo-triodes (VPTs) were used in the end-caps.

The barrel (EB) contains 61200 crystals, separated into 36 supermodules of 1700 crystals each. Its inner radius is 129 cm, and it covers a pseudorapidity range of $|\eta| < 1.479$. The end-caps (EE) hold 7234 crystals, divided in two Dees made of supercrystals containing 5x5 crystals. They are located at $|z|$ =314 cm, and cover a pseudorapidity range of $1.479 < |\eta| < 3.0$. A preshower (ES) device lays in front of the end-caps. It is made of silicon-strip sensors behind a lead absorber, and helps identify neutral pions, discriminates electrons from Minimum Ionizing Particles (MIP) and improves the photon and electron resolution. Figure 2.5 shows a sketch of the ECAL subdetector and its subcomponents.

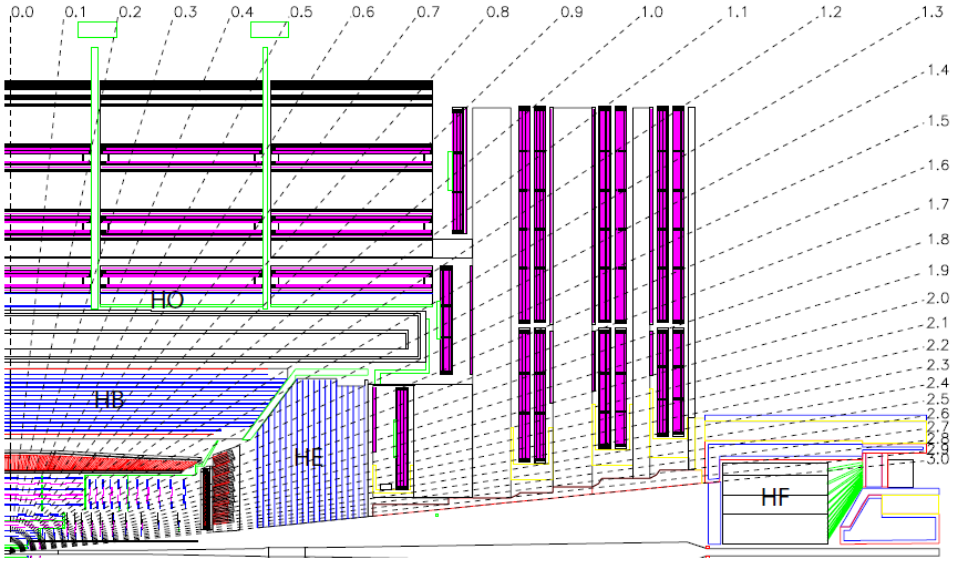

Figure 2.6: Longitudinal sketch of one quarter of the HCAL subdetector and its subcomponents, which are all explained in the text.

2.2.3 The hadron calorimeter

The CMS hadron calorimeter (HCAL) is a brass/scintillator sampling hadron calorimeter. Its aim is to collect the energy from the shower produced by hadrons interacting with material, which is usually much bigger than electromagnetic showers, thus the much more imposing size of HCAL with respect to ECAL. The HCAL is the last subdetector present inside the homogeneous field of the solenoid, and as such takes up all the rather limited space left between ECAL and the magnet. It is divided in four sections: the barrel (HB), the end-caps (HE), the forward hadron calorimeter (HF) and the outer calorimeter (HO), as detailed in Figure 2.6.

The barrel covers a pseudorapidity range of $|\eta| < 1.4$. It is composed of layers of brass absorber (70% copper, 30% zinc) alternating with plastic scintillators. The light is then collected with wave-length shifting (WLS) fibers. As the space for absorber material inside the solenoid is not enough to catch every time the whole shower, the HO scintillator was based outside the magnet, which was then used as an additional absorber layer. It collects the tails of the showers, and has the same pseudorapidity coverage as the barrel. The end-caps cover the range $1.3 < |\eta| < 3.0$, and uses cartridge brass (C26000) absorber, as they are subject to much more radiation, and are situated at the end of the solenoid were no magnetic material can be used. Finally, the HF calorimeter is placed in a much more forward position at 11.2m of the IP, and covers the range $2.9 < |\eta| < 5.2$. It uses steel/quartz fibers to channel the Cherenkov light emitted by the hadron shower in the fiber. As it is a very forward detector, it was used as part of the trigger requirements to either select or reject diffractive events, this in several analysis such as the ones presented here in Chapter 5.

2.2.4 The Magnet

The CMS solenoid is a 13 meters long, 6 meters inner diameter NbTi superconducting magnet, and provides an homogeneous magnetic field of 3.8T to the subdetectors it supports. To reach such field, the superconducting state is achieved by cooling it down to 4.2K with superfluid helium. Outside of the solenoid, the return yokes made of iron and holding the muon subdetector gets saturated with the remaining magnetic field. It allows bending of particles, even those with TeV scale energy, and as such provides the possibility to measure very precisely their momentum. It also impacts the charges collected in the different subdetectors, as they drift along the magnetic field in their respective mediums, which is of course taken into account and corrected for.

2.2.5 Muon Detector

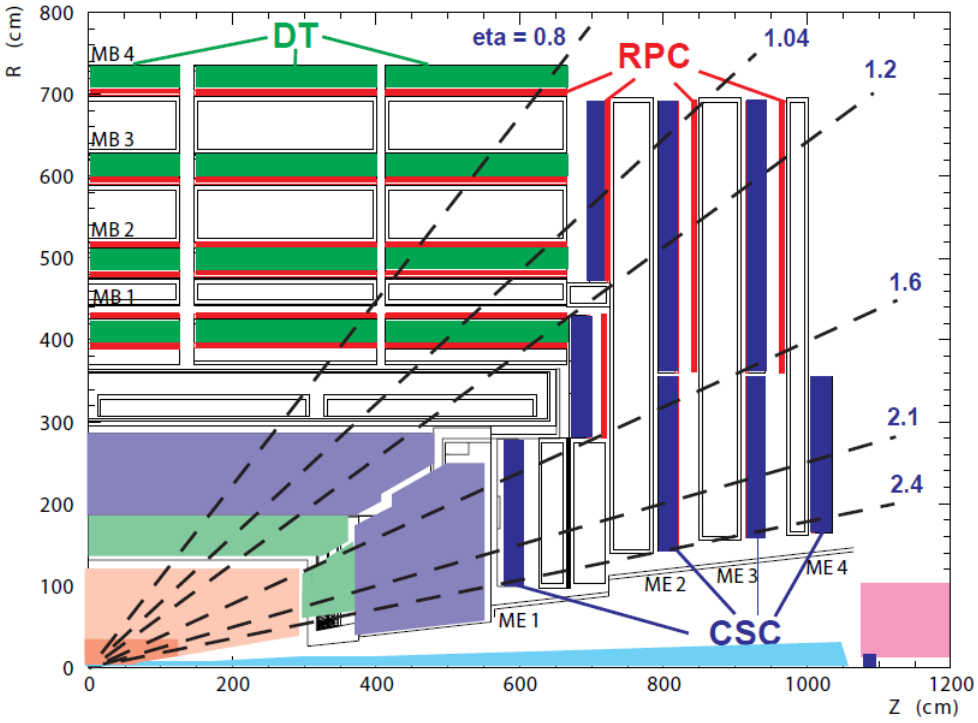

Figure 2.7: Longitudinal sketch of one quarter of the muon chambers and its subcomponents, which are all explained in the text.

As emphasized in its name, the muon detector is an important part of the CMS detector. It measures muon trajectories and their momentum, as in high energy collisions they only deposit minimal energy in the material they cross through and have encountered up to reaching the muon chambers. The muon detectors are gaseous detectors, and their cost was required to be rather small as the detector surface to cover was quite huge. Therefore, three different types of chambers were used to acclimate to their location.

The barrel region used Drift Tube (DT) chambers filled with a mixture of Argon and CO_2 gas. They cover the pseudorapidity range of $|\eta| < 1.2$. As the muon rate is pretty low in this region, and the magnetic field is quite homogeneous, this technology was a

particularly good fit. In the end-caps region, the muon rate is much higher, and thus Cathod Strip Chambers (CSC) were preferred, covering $0.9 < |\eta| < 2.4$. Those two type of gaseous chambers are interesting to use for triggering as they measure pretty effectively the p_{T} of the muons, but the rather long time needed for the charges from the ionized gas to reach the captors make it hard to distinguish from which exact collisions the detected muons come from. Therefore Resistive Plate Chambers (RPC) were placed in both the barrel and end-caps regions to allow identification of correct bunch crossing, of course at the price of a worsened position resolution.

2.2.6 The CMS Trigger

The BPTX and the BSC detectors for triggering

As mentioned in Section 2.1.1, many machines are scattered over the LHC ring to monitor the beam conditions. Two bptx devices are located at 175 m from the interaction point on both sides of CMS. They provide timing of the incoming beams at the level of 0.2 ns, as well as detailed bunch structure information. They are used by the CMS triggering system to differentiate bunch crossings were both beams were present, thus where a collision could happen, and bunch crossings with only one beam present.

The Beam Scintillator Counter (BSC) is the main detector used to trigger with during the first collisions. It was installed rather late, with some sectors being installed the night before the actual collisions. This situation is due to the fact the LHC and CMS have design working luminosities quite high, so no real effort and expenses went into designing a device for triggering during very low luminosity runs, which anyway did not last long. The two BSCs are located at 10.86 m of the IP, and are mounted against the inner surface of the HF calorimeters. The flight time for particles coming from the interaction point is around 36.5 ns, and they cover a pseudorapidity range of $3.23 < |\eta| < 4.65$. They are made of scintillators read out through photo-multipliers via wavelength shifting fibers. As they are pretty forward detectors, they allow to detect and trigger on particles coming from an actual collision, even for very soft collisions with very few activity in the central region. Using its very small time resolution of 3 ns, thus its very precise timing, it also helps rejecting beam halo events leaving traces in the tracker though no collisions happened.

The BPTX and the BSC are the two main detectors used to collect the zerobias and minimum-bias data samples used in the analysis presented later on.

The Trigger

At its nominal parameters, the LHC accelerator would collide protons at CMS' interaction point every 25 ns. Although in an ideal world we would like to record every single events, it is not possible for the detectors to send for each and every events all the data for all their channels, and no computing equipment would allow to write and store the quantity of data received at that speed. Therefore some events need to be discarded, which is the role of the trigger: carefully choose which events are worth keeping, and rejecting all the others. The CMS trigger is divided in two levels: the first one called the Level 1 (L1) trigger, which reduces the data rate from 40MHz to about 100kHz, and the second one named High Level Trigger (HLT), which later reduces the event rate to a mere 100Hz.

The L1 is the first trigger to make a decision. This decision process is designed to take only 3.2μs, from the data being sent by the subdetectors and processed by the L1 trigger, to the reception of the decision by all subdetectors. This very short decision time has several implications. First, as sending data and decision takes time just for them to travel and be processed, the L1 are custom-designed and highly programmable electronics

located in situ. Then, only data from the muon and calorimeters are used, as tracker information takes too much time to be transferred and to be processed. Lastly, as each subdetectors do not know if the information for a particular event is interesting, it has to be stored along with all the next events coming during the decision process: this is done with onboard local buffers tailored to store the data of each channels for enough events while the decision is taken.

If a decision to keep the event has been reached by the L1 trigger, the data for the event is sent to be processed by the HLT. Contrary to the L1, the HLT is a software trigger, using more than thirteen thousand CPUs to reach a decision in an acceptable time. It makes use of all available information from all subdetectors, starting with the simplest information to try and discard the event, while a partial and quicker reconstruction is operated only if needed.

Chapter 3

The PIXEL detector: calibration and commissioning

Pixel detectors are a specific type of detector, often present in particle physics experiments, but also other branches of physics looking to detect precisely the passage of a charged particle or eventually photons. Although their concept is similar to the sensor present in every digital camera one has at home, their required shape and the environment they need to function in usually means they are the fruit of long years of research, and all the current knowledge and know-how is poured into their development. As usual, one of the main constraints when making such detectors is its cost, especially for particle physics which tends to use rather large designs, with the ideal to cover as much surface as possible with sensitive material. It is of course not the only thing that restricts the making of such detectors: albeit being great at measuring very precisely particles' trajectories, one has to deal with their accordingly high number of channels, along with their response time. Although this is true for all types of detectors, their characteristics make them perfect to install them close to the interactions, which means they need to endure and keep working as long as possible under the excessive flow of particles crossing them. It is why much research is being done for pixel detectors, with several different designs, from full silicium ones to gas and grid layouts. Therefore many technologies exist, with different costs per surface of sensitivity, readout speed or even durability to choose from.

At CERN, more specifically for what concerns the LHC collider, all four experiments have a portion of pixel subdetector to measure the charged particles, although different technologies or designs were selected to accommodate at best each of the physics goals of each experiment. As detailed in Chapter 2, in CMS the choice was made to have a full silicium tracker, which the pixel is part of. The detector was principally designed, tested and assembled in two region of the world: Switzerland for the central part, and the US for the more forward section. The first section of this chapter will describe in much details the geometry of the pixel detector, along with the inner workings of all components and electronics. This will be followed by a brief summary of some of the main calibration procedures required to keep the detector in good health, with a longer explanation of what the gain calibration entails. Finally, the first measurements produced by the complete pixel detector after its installation in the CMS cavern will be presented, using Cosmic Ray data, with more emphasis on the pixel efficiency results.

3.1 Detailed description of the pixel detector

The pixel detector is the part of CMS that is the closest to the interaction point, hence it required certain specific approaches when it was designed. Its main purpose is to provide very accurate charged particle trajectory measurements, in order to reconstruct precisely the primary vertices, as well as finding secondary vertices, a task especially hard in the harsh environment close to the interaction point. It also needs to allow for track seeding, a mandatory step in the reconstruction of the full trajectories of the particles, which is explained in detail in section 4. Albeit being part of the full tracker of CMS, it was conceived to provide to the High Level Trigger a very fast and as accurate as possible tracking and vertexing by itself, as using the information from the whole tracker would take too much time at this level of decision. Being so close to the interaction point also means dealing with a very high number of particles crossing the detector layers, which leads to high irradiation. With that in mind, radiation-hard components were developed, so that the pixel detector could withstand the radiation damage during several years and keep operating in an optimal way. All those requirements had to be balanced with building a detector with as small material budget as possible, to minimize unnecessary multiple scattering of the particles that would complicate the task for further detector layers. Meanwhile, low noise electronics were used to reduce the number of fake hits, provide minimal hit losses, while ensuring costs were still affordable.

3.1.1 The detector geometry

As described before in section 2.2.1, the pixel part of the tracker detector comprises of a barrel with three layers (often shortened to BPIX) surrounded by two end-caps disks on each side (often shortened to FPIX, for Forward PIXel, as it measures essentially particles with high pseudorapidity). A schematic view of the detector is offered in Figure 3.1, along with the pseudorapidity acceptances of each layers/disks.

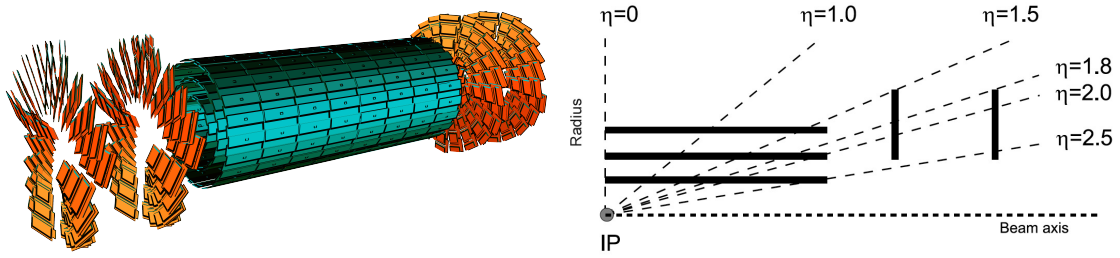

Figure 3.1: Schematic view of the pixel detector (left), with the three barrel layers and two end-cap disks on each side, along with the pseudorapidity acceptance of each layers/disks (Right).

The Pixel Barrel

The innermost barrel layer has a radius of 4.4 cm, while for the second and third layers the radii are 7.3 cm and 10.2 cm, respectively. All three of them have a length of 53 cm. To allow for the pixel detector to be taken out and facilitate the work around the beam-pipe, each layer is divided in two half-shells. Indeed, since the full LHC beam-pipe will be under vacuum, it is not possible to just take out the part of the beam-pipe which resides inside

CMS. Moreover, each time the pixel detector is moved, either to be placed in or taken out, a lot of precision and patience are required even with the division in two half-shells, as breaking the very thin layer of the beam-pipe would mean undoing the vacuum, which would in turn delay considerably the LHC operations. The layers are further divided into ladders, with half-ladders at the edge of the half-shells, ensuring an overlap at the juncture to provide a full hermetical $r\phi$ coverage. Those ladders are made of thin carbon fibers of 0.22 mm thickness, and are glued on aluminum cooling pipes. Those have a trapezoidal shape, with a wall thickness of only 0.3 mm. This cooling structure represents the main skeleton frame of the barrel, and it is filled with liquid fluorocarbon (C_6F_{14}), the coolant chosen by the whole CMS detector, and keeps the detector at a temperature of $-10\,^{\circ}\mathrm{C}$. There are 8, 14 and 20 ladders on the respective half-layers, while the number of cooling pipes taking part in the mechanical structure are 9, 15 and 21, respectively.

The three half-shells are mounted on the endflange, and form half of the pixel barrel. The endflange is the main support structure of the barrel. Finally, four 2.2m supply tubes extend along the beam-pipe out of the detector, connecting it to the outside world. They carry everything the barrel detector needs to function: the cooling fluid, all the electric power lines, as well as the optical fibers and electronics for readout and control. The connection between the detector and the supply tubes is done via a six layer PCB, which is mounted on the endflange.

The electronic board containing the active sensor area is called a module, and is discussed at length in section 3.1.2. There are 8 modules fixed on every ladder, while half-modules were also made to accommodate the half-ladders. Therefore the barrel region is composed in total of 672 full modules and 96 half modules, which makes for a sensitive area covering 0.78 m^2.

The Pixel End-caps

The forward pixel detector completes the barrel part of the pixel, by extending its pseudorapidity. The end-cap disks extend from 6 to 15 cm in radius, and are placed at $z = \pm 35.5\mathrm{cm}$ and $z = \pm 48.5\mathrm{cm}$. As for the bpix, the disks are split into half-disks to permit taking in and out the detector. Each half-disk consists of 12 trapezoidal blades arranged in a turbine-like geometry. The blades are slightly tilted, to permit charge sharing between pixel neighbors: as the electric field in the sensor is not parallel to the magnetic field, the ionization charge deposited by the particles in the sensor is subject to the Lorentz force, which distributes the charge among several pixel. Knowing the direction of this Lorentz force, which is one of the tasks performed during calibration and detailed in section 3.2, allows for a better hit resolution measurement. Each blade is composed of a sandwich of two back-to-back panels around a U-shaped cooling channel. Rectangular sensors, identical to the modules in the barrel but of five different sizes to adapt to the geometrical constraints, form the so-called plaquettes. The front panel contains three plaquettes, while the back one contains four of them, all arranged with overlap to provide full coverage for charged particles originating from the interaction point. The end-cap disks include a total of 672 plaquettes, forming a sensitive area of 0.28 m^2. Pairs of half-disks are finally mounted on a carbon-fiber half service cylinder, which contains all the mechanical and electrical structure to support, position, cool, power, control and read out the fpix detector. Figure 3.2 gives a representation of all the subdivisions of the end-cap disks and their respective arrangements and numbers, from plaquettes, to panels, blades and half-disks.

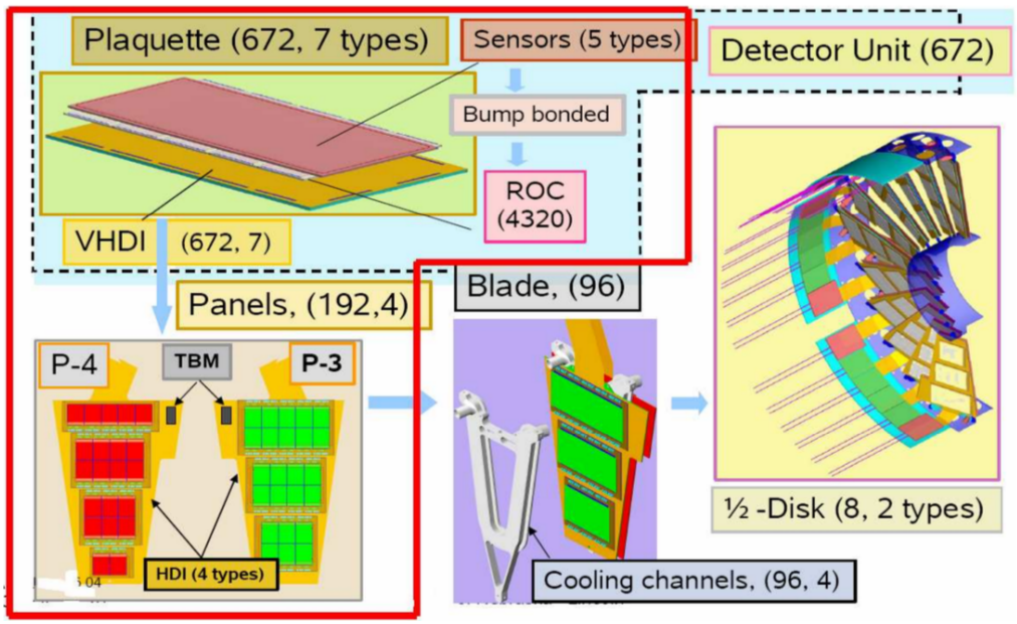

Figure 3.2: The flow diagram of the steps in the construction of the FPix detector.

3.1.2 The pixel modules

A photograph of a full module and a half module can be found in Fig.3.3, accompanied by a schematic view of all its components. A complete module has the dimensions 66.6x26.0 mm^2, and weighs 2.2 g plus up to 1.3 g for cables, depending on the position of the module and consequently the length of the cables serving it.

The modules are made of the following components:

- **2 basestrips :** made out of silicon nitride, they provide the necessary mechanical rigidity to the module, and are used to fix it on to the ladder support and the cooling structure.

- **16 Read-Out Chips (ROCs):** each chip is composed of 80x52 Pixel Unit Cells (PUCs), connected to the sensor pixels with an indium bumped bond. Its purpose is to read out the charge collected in the pixel sensor, to store it for each bunch crossing while the trigger decision is performed, and to eventually send out all the information it buffered. A detailed description is in section 3.1.2.2.

- **The sensor :** a silicon-based sensor divided into pixels of sizes 100 μm and 150 μm along the $r\phi$ and z coordinates, respectively, and a thickness of 285 μm resulting in a deposit of 23 ke^- for a perpendicularly incident Minimum Ionizing Particle (MIP). A detailed description is in section 3.1.2.1.

- **The High Density Interconnect (HDI) :** a low mass flexible printed circuit board, that distributes power and control signals to the different chips. It also transmits the readout data of the ROCs to the Token Bit Manager. A detailed description is in section 3.1.2.3.

- **The Token Bit Manager (TBM) :** located on the HDI, it controls the readout of the ROCs and sends the data to the Front-End Drivers (FED). Each TBM can handle up to 24 ROCs. A detailed description is in section 3.1.2.4.

- **A power cable :** it consists of 6 copper coated aluminum wires. Soldered to the HDI, it brings the analog, digital and high voltages needed by the module.

- **A signal cable:** it is a two layer Kapton/copper cable with 21 traces, wire-bonded to the HDI. It is meant for sending in the control signals and reading out the analog output.

The half-modules are identical to the full modules, but contain only one row of 8 ROCs. This makes for a total of 11520 ROCs present in the bpix, with around 47.9 million pixel sensors.

The plaquettes in the fpix are similar to the bpix modules described here, but due to their different number of shapes to adapt to the geometrical constraints of their position on the panels, they have different number of ROCs on each of them, from 2 to 10 ROCs as can be seen from Fig.3.2. Their sizes and weight are therefore also different from the values given for the standard barrel module, but their inner workings are identical as they are just different accumulations of ROCs. The fpix is composed of 4320 ROCs, for a total of almost 18 million pixels.

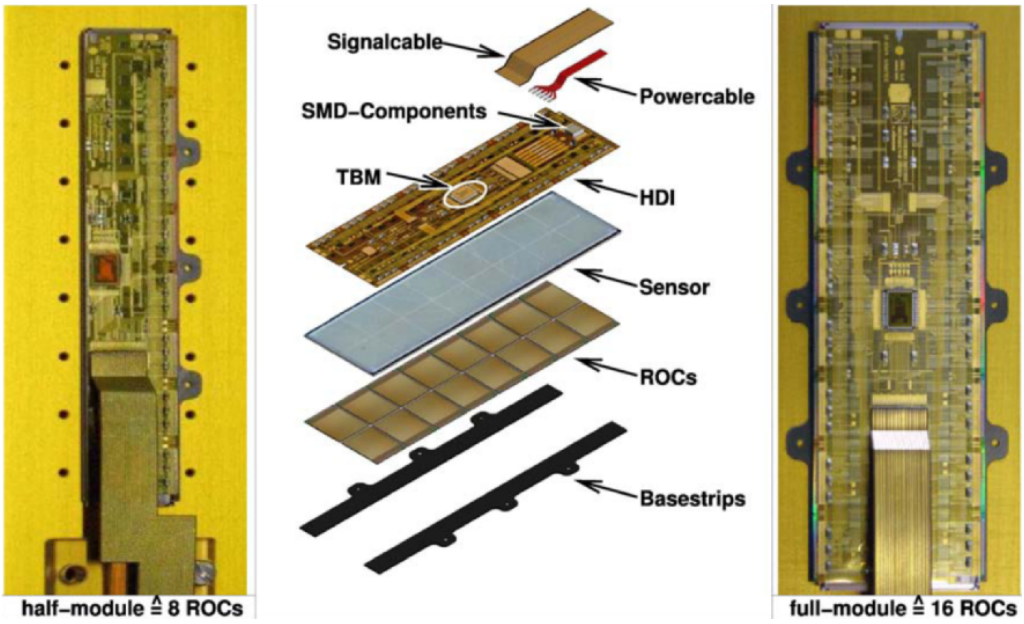

Figure 3.3: Picture of a bpix half module (left) and full module (right). Center: The components of a pixel barrel detector module (from top to bottom): the Kapton signal cable, the power cable, the HDI, the silicon sensor, the 16 ROCs and the basestrips

3.1.2.1 The sensor

The principal that follows any semi-conductor sensor such as the tracker and the pixel detector is as follows: when traveling through certain type of materials, charged particles lose some of their energy mainly due to elastic scattering with electrons. This leads to formation of electron-hole pairs: for instance in silicon the energy required to create such pair is 3.6 eV. The charge carriers can then be moved by depleting the sensor using an electrical field, which will lead to an induced current detectable in the electrodes by the readout chips.

The sensor of the CMS pixel modules is made of crystalline silicon, using a double-sided design with high dose n+-pixels on an n-doped substrate. The back side of the sensor consists of a large p-implant forming the pn-junction that depletes the sensor. A moderated p-spray technique was also developed to help with interpixel isolation. A schematic view of a pixel sensor attached to the ROC can be viewed in Figure 3.4.

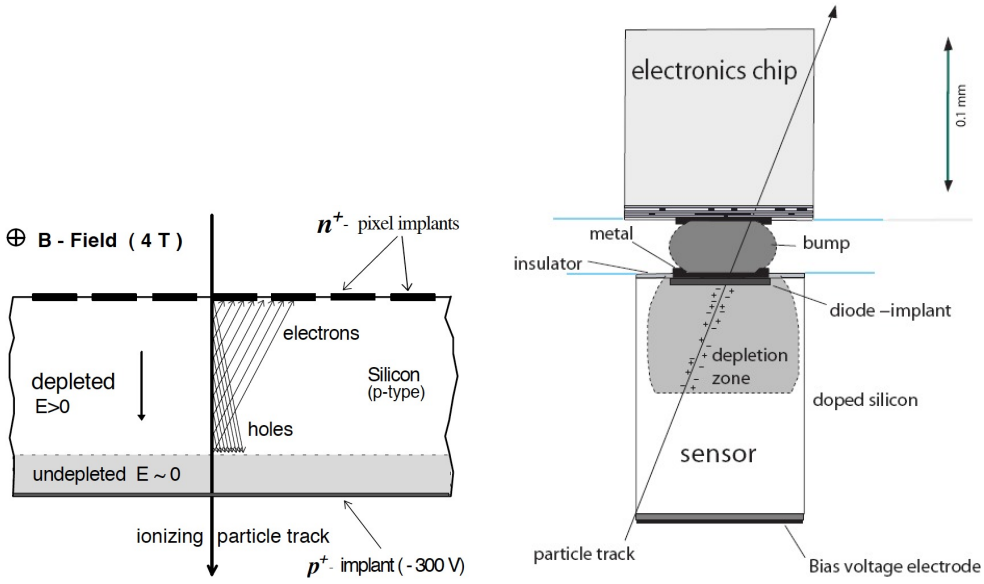

Figure 3.4: Charge deposition in the sensor (left), and attached to a Pixel Unit Cell (right).

Although the cost of this double sided "n-in-n" concept is higher than the standard "p-in-n" one, it had several advantages that weighted the balance in its favor: it assures a high signal charge even under moderated bias voltages, as well as when being heavily irradiated; the Lorentz angle drift is larger due to the high mobility of electrons, which increases the spatial resolution of the detector; and due to the need of a structured back side, guard rings could be implemented to keep all sensor edges at ground potential, allowing a stable operation even at very high bias voltages. At the beginning of its life, when the pixel detector will not have yet suffered from irradiation, the sensors will be operated at bias voltages around 150 V. Those voltages will increase over time to up to 600 V, to keep the sensors damaged by irradiation still fully depleted and working with the same characteristics. The silicon pixel sensors description and the result of their testing are given in [19] for the barrel and in [20] for the forward pixels.

3.1.2.2 The Read-Out Chip

The goal of a readout chip is to read the amount of charge produced by ionization in the sensor for every bunch crossing, so ultimately at a rate of 40 MHz, and store it while the L1 trigger takes its decision. The chip is 180 μm thin, measures 7.9x9.8 mm^2, and draws around 120 mW of power. Both the ROC chip and the TBM chip are made of 0.25 μm CMOS radiation-tolerant technology [21], which means the lifetime of the module is mainly limited by the lifetime of its sensor component. The ROCs are composed of three main functional units, shown in Figure 3.5. The active portion is made of Pixel Unit Cells (PUC), one per pixel sensor, with 52 columns and 80 rows of PUCs per ROC for a grand total of 4120. The readout chain is organized by pair of columns, therefore

the standard unit is often labeled as the double column, with 26 of them in each ROC. As can be seen in Figures 3.4 and 3.6, the PUC is bump bonded to the sensor: made of indium, the 25 μm bump-bonds where specifically developed by a CMS team at PSI in Switzerland, as this size could not be achieved through standard industrial processes. The two other parts of the ROC are the double column periphery, containing the charge buffer, the timestamps buffer, and the chip periphery with the control interfaces. Each ROC can be programmed using a fast I^2C interface with a 10 bit format. In order to optimize the signal processing and the readout of the ROCs, as well as to compensate the unavoidable chip-to-chip variations, there are 26 Digital to Analog Converters (DACs), and 3 registers controlling the voltages and currents on the PUC and double column periphery.

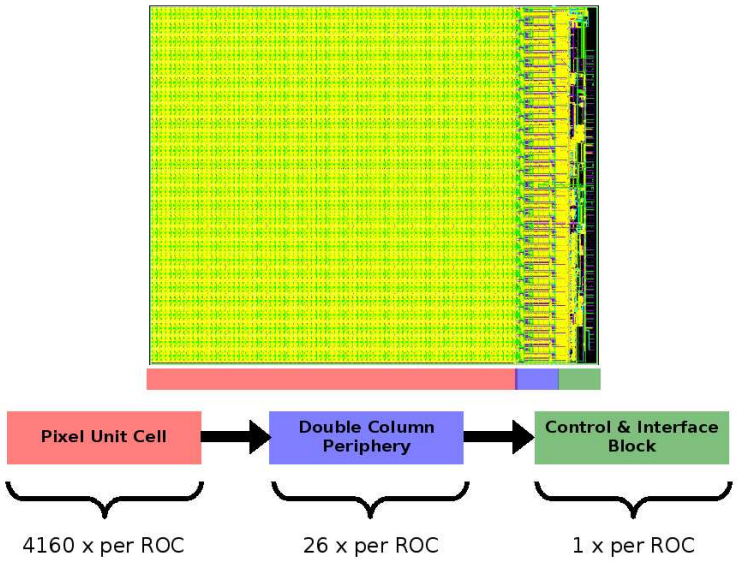

Figure 3.5: Diagram of a ROC with the 3 parts highlighted: the active part containing the double columns and all the PUCs, the double column periphery containing the data and timestamp buffers, and the chip periphery. A more detailed color-coded schematic view of the electronics in each section can be viewed in Figure 3.7.

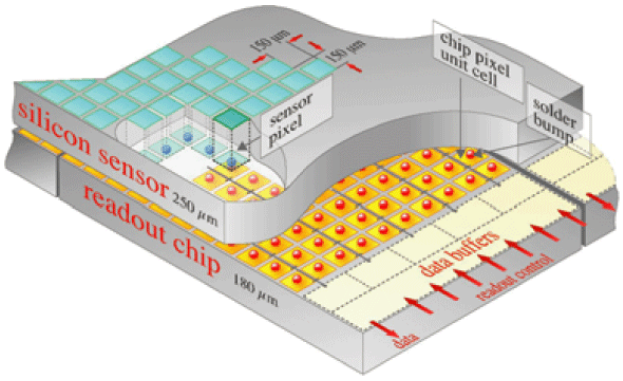

Figure 3.6: Schematic view of a portion of the sensor on top of a ROC with its PUC, double column periphery and chip periphery.

The ROC uses a double column zero-suppression readout: only charges above a wanted

threshold are collected, and a whole double column is read only when there is at least one hit in one pixel of the double column. This mechanism is necessary to minimize the amount of data sent for each bunch crossing, as it is technically not feasible (and a waste of time and resources) to read out and send data for every one of the 66 million pixels in the detector. The whole diagram of the electronical readout layout of the PUC, the double column periphery and the chip periphery as well as DAC and register settings are drawn in Figure 3.7.

The readout starts in the sensor: each time a charged particle crosses the sensor, a signal voltage is induced in the PUC. There, the signal passes through a preamplifier and a shaper, before arriving at the comparator, where the zero-suppression is applied by comparing it to a reference signal. If the charge is higher than the set threshold, the double column periphery is notified, and it will start reading itself. This threshold is tunable by a DAC setting giving it the same value for an entire ROC, and a per pixel correction can be applied using four trim bits available on each PUC, which decreases the threshold value with an effect depending on the global DAC setting of the ROC. This threshold is consequently quite important and needs to be calibrated as explained in Section 3.2 in order to make the pixel response uniform throughout the detector. Once notified, the double column is read using the column drain mechanism: it starts with its bottom outer left pixel (closest to the periphery), goes up the column, and goes back down the right column. The pixels with a hit go in inactive mode, then send an analog pulse height and their pixel address to the periphery, which stores them in the data buffer along with the corresponding bunch crossing number in the timestamp buffer, before setting themselves active again. Finally, the data is sent to the TBM when a token pass is received, at which time the stored timestamp of the data is compared to the L1 value received from the TBM to check if they match. Along with the four trim bits controlling the per pixel threshold, each PUC also has a mask bit that can disable its readout by toggling on or off the comparator: this allows for pixels not behaving properly - for instance noisy pixels with constantly saturating signals, see Section 3.4 - to be left out of the readout data stream.

In order to test the quality of a module without any external sources, which would require a complicated setup and could hardly be achieved at will, each PUC has a specific circuitry allowing the injection of an internal calibration signal, in place of the signal received from the sensor. This permits to test each pixel's charge readout chain, and to calibrate each of them to correct the pixel-to-pixel variations within a ROC, as well as between ROCs and modules. This calibration work is further detailed in Section 3.2, as it is one of the pixel tasks I dedicated time to during my thesis. The amplitude injected can be modified through the $Vcal$ DAC setting, and its timing is set with the $CalDel$ DAC. Both parameters are used to sample the charge sent to the FEDs, and compute offline the correction to apply to each pixel. There are two ranges of amplitude that can be applied: they are often labeled high and low ranges, with a ratio of Vcal high over low equal to 7. One Vcal low unit amounts to around 65 e^-, while one Vcal high unit is around 455 e^-. The switching of the two ranges is controlled by the $CtrlReg$ control register.

3.1.2.3 The High Density Interconnect

The HDI is the component of the module that connects it to the outside world: it receives the voltages needed through the power cable, receives the control signals and sends the data using the Kapton signal cable. The power cable provides three types of voltages, to accommodate both analog and digital parts of the module: analog and digital

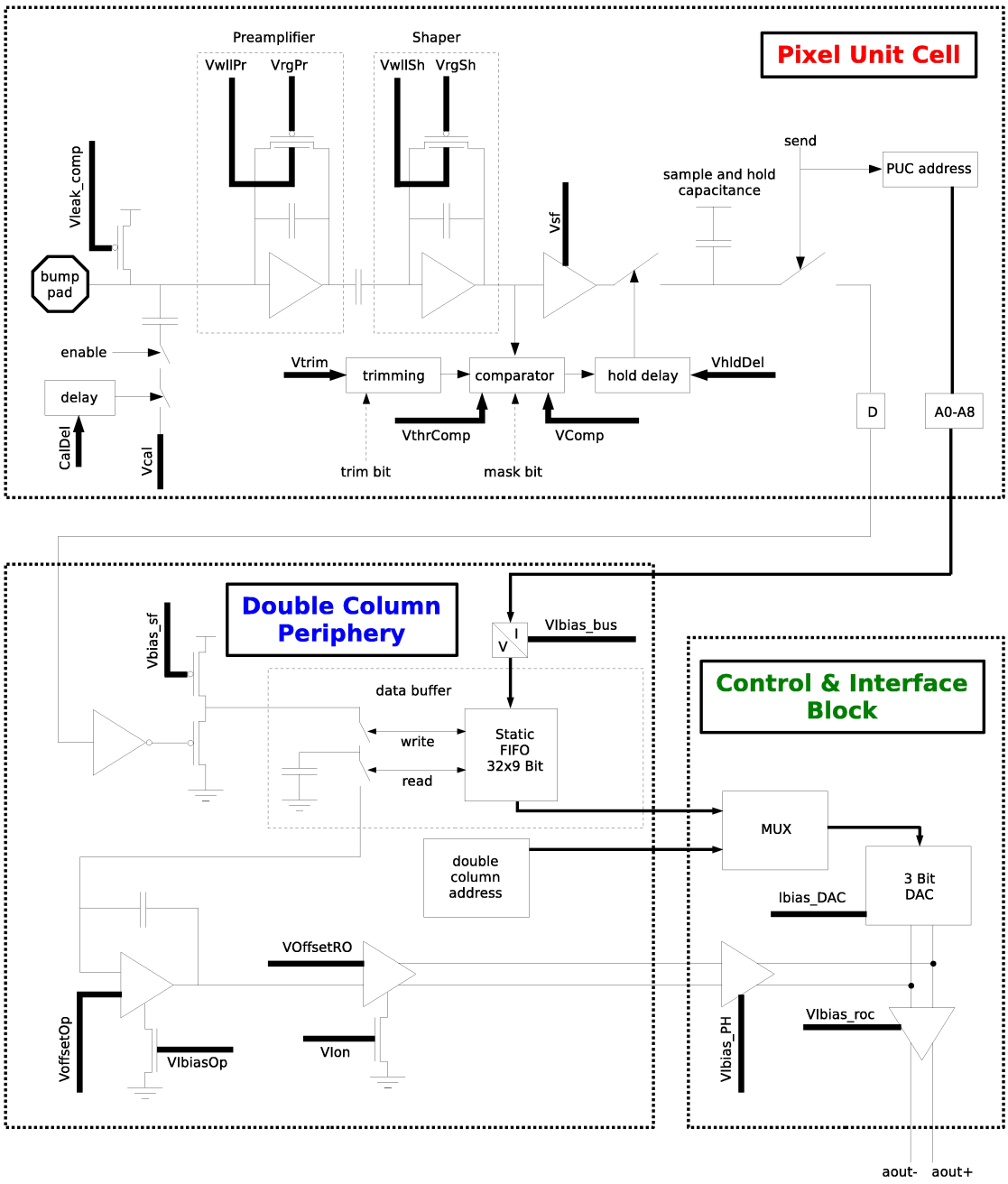

Figure 3.7: Schematic overview of all the electronical components and their layout within a ROC. The different DAC settings available are also presented.

voltages, plus the high voltage applied to the sensor. The ROCs and the TBMs are connected to the HDI circuit board using wire-bonds.

3.1.2.4 The Token Bit Manager

The first two layers of the bpix have two optical links each to send the data out of the detector, as they are the ones closest to the IP, with therefore the highest number of tracks crossing them, so the higher number of hits thus data to retrieve from. Each chip contains

therefore two TBMs, which are sometimes named TBM A and B. Each TBM is capable of controlling up to 24 ROCs. In the barrel region, they control either 8 (half-modules and first two layers) or 16 ROCs (third layer), while one chip per blade in the fpix means a TBM for 21 (front panel, see Figure 3.2) or 24 (back panel) ROCs. The main purpose of the TBM is to organize the readout of the data. It receives the L1 triggers, which can be stored in case of high burst rates: up to 32 triggers are stacked in buffers, awaiting the start of the data readout from the ROCs. Each time the TBM receives a trigger from the L1 telling it to retrieve the event data, a token pass is issued to the first controlled ROC initiating its readout sequence. The token pass is passed from ROC to ROC, the last one in the readout chain sending it back to the TBM. The TBM adds a header, an 8-bit encoded event number, and a trailer containing an 8-bit TBM error status, after multiplexing the data sent by each ROCs, to finally drive the whole data stream in analog form through the data link. The TBM is also responsible of passing the L1 trigger and the clock it receives to the ROCs. A low power design approach was used, leading to the choice of a TBM chip needing a power supply of only 2.5 V to function. The role of the TBM is schematized in Figure 3.8.

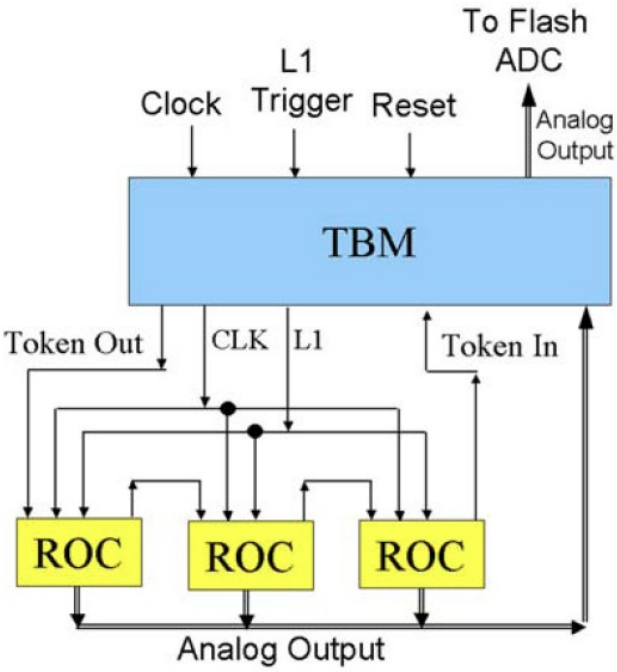

Figure 3.8: Schematics of the TBM roll in the readout chain.

3.1.3 The Analog Read-out

In the module

Once the pulse height for each pixel is sent by the ROCs to their master TBM, the data of all the ROCs attached to the TBM is multiplexed, a header and trailer are added, and it is sent in analog form through the data link, transmitting a pulse for each clock cycle where the ADC charge value encodes the data. An example of a TBM data sample where the only hit is in a pixel of ROC 0 is shown in Figure 3.9: the TBM header,

the data for the 16 ROCs and the TBM trailer are clearly visible and span over 70 clock cycles. The TBM header consists of eight clock cycles: it starts with three very low analog levels called Ultra-Black (UB), and a black level defining the zero level of the differential analog signal. The event number information is encoded using different ADC values in an additional four clock cycles. The readout sequence of a ROC consists of at minimum a UB level, followed by a black level and a level called last DAC comprising the value of the DAC addressed by the last programming command. The hit information, if any, follows those three clock cycles: it is transmitted in 6 clock cycles per hit, encoding the pixel address and the pulse height information. The pixel address is coded using six different discrete analog levels (so in base-6), where the first two values encode the double column index (leaving 36 possibilities for the 26 double columns), while the pixel address within the double column requires three clock cycles (2*52 pixels encoded with 216 possibilities). This address encoding is often calibrated and briefly explained in Section 3.2, as an error in the ADC pulse read could lead to mis-identifying the pixel location. Following its address, the hit charge collected in the sensor is sent in one clock cycle. The last cycles of the readout are for the TBM trailer: two UB levels and two black levels are followed by four clock cycles transmitting the TBM error status. In total, for a bunch crossing where nothing happened, a module of the third barrel layer - where each TBM controls 16 ROCs - will send its data over a minimum of 64 clock cycles.

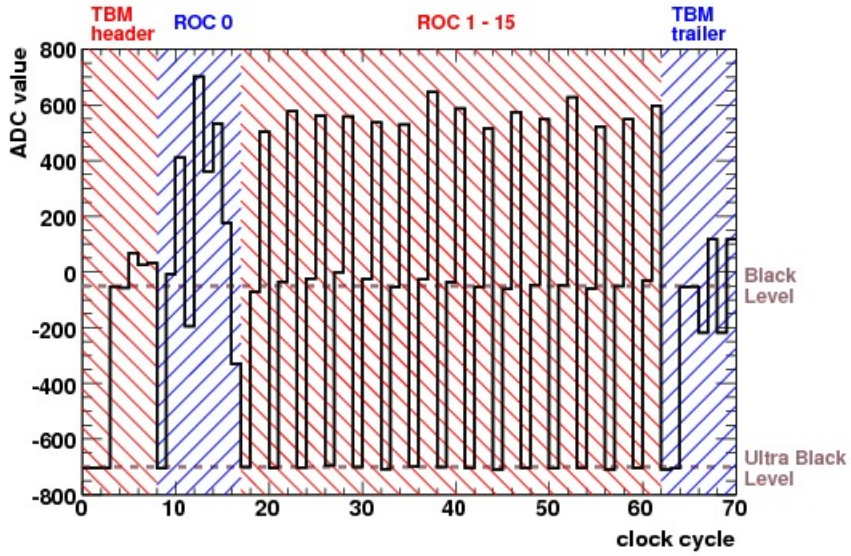

Figure 3.9: The analogue readout of a module with only one hit, in the first ROC.

In the supply tube

As explained before, the analog signals are transmitted outside the module by the TBM through the Kapton cables. There are 1120 such readout links for the barrel and 192 for the end-caps, corresponding to the number of active TBM chips. The data is brought to printed circuit boards called analog opto-boards, located outside the pixel detector, in the supply tubes, on which Analog Optical Hybrids (AOHs) are placed. There, the electric analog signals are amplified in an Analog Level Translator chip, then converted into 40 MHz analog optical signals inside the AOH. The 6 lasers equipped on every AOH

then transmit the data through 60 m long optical fibers, which connect the AOH to the Front-End Driver (FED) located in the CMS underground service room near the CMS cavern.

The supply tubes also host several other components essential for the readout to happen. For instance, the digital opto-boards contain Digital Optical Hybrids (DOH) chips that are connected to the Front-End Controller (FEC) with 4 optical fibers: two to receive signals such as the L1 trigger and the clock - both first split up by a Phase Locked Loop (PLL) chip -, as well as control signals - with a DELAY25 chip to adjust their relative phases -, that are all passed down to the modules, and two fibers for sending data. There are 2 types of Front-End Controllers: the standard one also employed in the Strip Tracker, which exploits a slow control link to configure the readout electronics hosted on the supply tube (AOH, DOH, PLL, DELAY25 chips) and to allow communication with the Communications and Control Units (CCUs), and the specific pxFEC with a digital control link and a modified I^2C protocol to send the L1 trigger, the clock and the DAC settings and registers through the DOH to the modules. The pixel detector front-end control system consists of four CCU boards equipped with 9 CCU chips - where one chip is used for redundancy, as it is a critical component: if one AOH fails, it is a whole portion of the detector that cannot be read - which, located at the outer end of the supply tubes, is responsible for dispatching the respective control signals (which are again the clock, the L1 Trigger, and the control data).

In the service room

The optical fibers transmitting the data arrive to the FEDs located in the service room, near the CMS cavern containing the detector and still underground. It is the room that holds the more computer-like equipments for every CMS subdetector. The pixel subdetector takes three whole 9U crates to contain all the 40 pixel FEDs - 32 for the barrel, 8 for the end-caps -, and every crate is controlled by one PixelFEDSupervisor application. Each FED is composed of 36 optical inputs with each an optical receiver and an ADC. The role of the FED is to receive the analog signals, to digitize all the data at the LHC frequency of 40 MHz, and to decode the base-6 pixel address. The digitized data is then sent down S-Link cables to the Data Acquisition System (DAQ), which constructs the event. The FED data can also be read out via VME by the PixelFEDSupervisor, as it also accepts data in RAW format with then all the FED work being bypassed. This way of reading the data is needed during calibration, as the readout of the detector is usually rather specific and quite different from an event where the two beams colliding send a whole bunch of particles through the detector.

3.1.4 Upgrade of the pixel detector

The LHC is planned to have two big upgrade phases, where the beam characteristics will be improved to provide to the experiments more useful collisions. The first upgrade, called Phase 1, should happen in 2016, and will see the collider go from LHC to Super LHC (SLHC). The total energy of the beams will be brought up to 14 TeV, which was the designed goal of the actual LHC but could not be achieved after the accident of 2009, which severely underlined the machines weak points, and the peak luminosity reached will be doubled to $L = 2\text{x}10^{34}\, cm^{-2}s^{-1}$. The Phase 2, planned for after 2021, will seek to reach a luminosity of $L = 1\text{x}10^{35}\, cm^{-2}s^{-1}$, so 5 times higher than Phase 1. Those upgrade will result in more challenging environments for the detectors, and particularly

for the pixel detectors - both the CMS and ATLAS ones - which are the closest to the interaction point. There will be a higher occupancy, and the detectors will be submitted to even more radiation than they already are. Moreover, the current detectors have already been running for several years now, and while they perform greatly their task with the current beam conditions, they accumulate radiation damages. This was of course planned when they were designed - hence their still remarkable working conditions - and their longevity is reaching its intended end. It will therefore be time to upgrade the detectors, with un-irradiated components and with newer designs that should be able to cope with the harsher environment - as a reminder, the first designs for the current pixels are dated 15 years ago -, and even if possible out-best the performances of the current generation and improve the physical observables used in the physics analysis.

The design of the new pixel detector of CMS was achieved with several key points in mind. First, the high hit detection efficiency that the current detector had attained would have to be kept, while still preventing data losses to occur. The same goes for the good track seeding and the pattern recognition performances: as it is one of the main jobs of a pixel detector, the goal was to provide the tracking system with at least 4 pixel hits per tracks in its whole acceptance volume, compared to the current 3 pixel hits (2 at $\eta > 2.1$). The track parameter resolutions were also a key observable to improve with the new design: for this, the material budget in the tracker volume was greatly reduced, the radius of innermost layer was decreased, while the radial acceptance was increased. Finally, careful thoughts were brought to the late Phase 1 run luminosity challenges that would occur concerning radiation damages: the sensor and electronics had to, again, perform as well as in the beginning even after several years of collisions with increased luminosity. To achieve all this goal, many changes were introduced, the main ones being: the upgrade would bring along new geometry, new readout chips were developed, new cooling system based on two-phase CO_2 would be used, and new powering system based on DC-DC converters would be put in place.

Concerning the detector geometry, the Figure 3.10 shows the new layout of the pixel barrel and end-caps. The barrel will change to contain one more cylindrical layer, to hence have 4 layers. The first layer will have a much smaller radius of R= 29 mm (previously 44 mm), which will only be possible thanks to the upgrade of the beam-pipe, which will see its radius decreased as well. The third layer will be at the same place (R=102 mm), and the added fourth layer will have a radius of 160 mm. This will lead to an increased number of modules, from 768 to 1184, for a total of 79 M pixels, compared to the 48M channels it currently has. There will also be no need anymore for half modules: all of them will integrate two rows of 8 ROCs.

There will also be changes made to the end-cap geometry: from two disks per side, the upgraded fpix will hold three disks on each side of the barrel. Each disk is planned to be made of two concentric rings, an inner and an outer one, which would allow for easy replacements. They will hold in total the same number of modules as the current subdetector (672), but the ring layout will no longer require specific modules adapting to the geometry of the panels: every modules will be identical to the bpix ones, with 2x8 ROCs, increasing the number of pixel channels from currently 18M to 45M. This means a total increase of 13856 ROCs and 57M pixels in the whole pixel detector, which of course also requires improvements in the readout of the detector.

New readout chips, quite similar to the ones in use right now but with slight improvements, have been developed and are currently being tested. Their main changed characteristic is that the readout will no longer be analog at 40 MHz, but rather digital with a frequency of 320 MHz (160Mb/sec since digital), which would be able to cope with

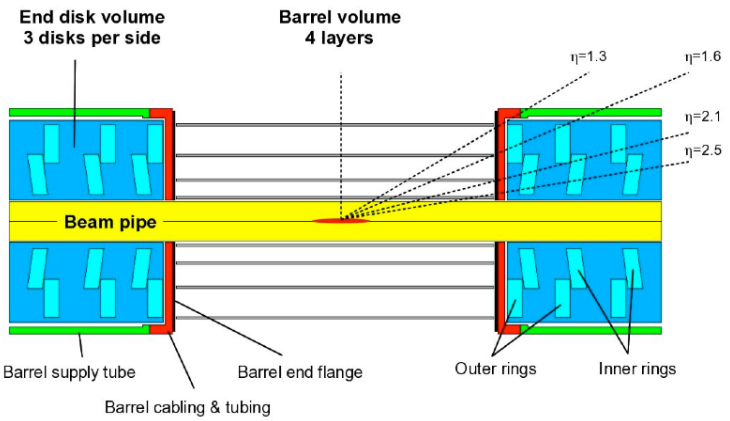

Figure 3.10: Layout of the updated pixel detector, with the four layers and three disks on each side, and their global pseudorapidity acceptance.

the high rates and expected PU of the coming LHC data delivery. The ROC will also hold more data and timestamp buffers, and the zero-suppression threshold should decrease to an effective 2000 e^-. The TBM, the optical hybrids (POH) and the FEDs, all essential components in the readout sequence, will also be adapted to the new digital data stream.

As mentioned previously, the material budget of the total pixel detector will be reduced, even with the addition of the fourth layer and third end-cap disks. This will be achieved by a newer and lighter material in the support structure, while the cooling will still be part of the mechanical structure but will decrease in size, as the usage of CO_2 cooling fluid will require smaller pipe thanks to its smaller mass flow needs. For instance, the total weight of the layer 1 mechanics (this includes support, cooling pipes and cooling fluid) will decrease from a current 400 g to around 100 g.

All the physics observables also look improved compared to the already good values obtained in the past data taking. The vertex resolution is planned to gain around 20%. The track seeding efficiency will increase, while fake rate will drop. The b-tagging - thus the secondary vertices measurements -, which will be even more essential to many analysis planned for the future, will be greatly improved.

The upgrade of the LHC machinery, as well as the different detectors, will take time, and are planned during several Long Shutdowns (LSs) scheduled over the coming years. During the recently completed shutdown (LS1, 2013/2015) the beampipe has been modified, and several new modules have been inserted to test the whole chain under realistic conditions and to get a first glimpse at their performances. Then, during one of the regular winter shutdowns, the new pixel will be inserted to be ready for Phase 1 starting after LS2. On the long term, the LS3 planned in 2021 will see the SLHC become the HC-LHC, and the trackers will need to be fully replaced. Already new ideas are emerging for this next pixel upgrade, with for instance new modules called p_T modules, or ways to exploit the full tracker info for triggering.

As with the first round of designs of the CMS detector, a Technical Design Report (TDR) has been published [22], containing details of the new detector upgrade, as well as simulation and first results of their expected improvements.

3.2 The calibration of the PIXEL detector

The different components were fully tested before their implementation at CERN in the CMS cavern. Indeed, the barrel section and the end-cap section were respectively built at PSI in Switzerland and FNAL in the US. There, several testing procedures were implemented to calibrate and get a first look at the readout chain characteristics and behavior while the detector constituents were being put together. This was done in larger scale at the Tracker Integration Facility (TIF) in CERN - where I spent the summer of 2007 learning about the SST tracker under the CERN summer school program -, with major portions of the tracker parts being fully commissioned and tested in real conditions, before the final shipping of all detectors down the CMS shaft to the underground cavern during summer 2008, in time for the start of the LHC collisions. Although it started well, the collisions stopped abruptly after the accident, which was the end for a time being of the physics goal for the experiments. But this was not lost time for the detectors: CMS took the more than 12 months without beam to fully calibrate all the subdetectors, and even take cosmic ray data (results are detailed in Section 3.4). This lead to a much better understanding of all the detectors, and allowed CMS to be well prepared for the actual collisions starting November 2009. This year was the opportunity for me to work with the pixel detector, in its calibration with the gain calibration, and with the Cosmic Ray data by measuring the pixel efficiency.

3.2.1 Pixel alive test

This is one of the most basic calibrations possible. A charge, way above the comparator threshold, is injected in the pixels, and the pixel is said *alive* (compared to *dead*) when a charge is readout. This is the most effective way to test problems in the readout chain for each pixels. Of course the goal was, even after all the travels and work done on the detector, to have as close to zero dead channels as possible. Since this calibration makes use of the internal charge injection, which is inserted just before the PUC pre-amplifier (see Figure 3.7), it cannot test neither defects in the sensor, the bump-bonds nor the power supply voltage to the sensor. This is why data is used to test the whole pixel channel, by comparing the number of hits they recorded to the total number of pixel hits. Those numbers change over time, as ROCs could start behaving badly; In the fpix, power supplies were also the source of detector parts not being able to collect data. Cases where complete sets of pixels - like complete double columns, ROCs or even modules - were dead have been the subject of repair trials whenever possible, to ensure the number of dead pixels stays as low as possible. Figures have been given in Section 3.4 for the year 2009.

3.2.2 Baseline calibration

This calibration is used to correct the signal of the AOH. Indeed, the analog optical hybrids used to convert the electrical analog signal of the TBM to an optical one is very temperature sensitive: a change of just one degree in temperature shifts the signal received by the FEDs by more than 50 ADC counts, out of 1024. This means the temperature has to be carefully monitored, and an automatic procedure has been developed to correct the baseline shift. This correction has been tested - when taking data - to work correctly for small changes of one degree maximum, which therefore requires to first have a very uniform baseline distribution, hence the calibration. The baseline calibration sets the optical input to be ± 5 ADC from a defined value, usually 450. As it cannot be performed during data

taking, this 5 minute long calibration is executed very often: at every fill of the LHC before colliding beams, and whenever necessary. A too harsh change in the temperature signifies the correction cannot be automatically applied, and requires the pixel detector to be taken out of data taking, the calibration to be carried out before the pixel can be put back in the run.

3.2.3 Clock adjustments and Delay scans

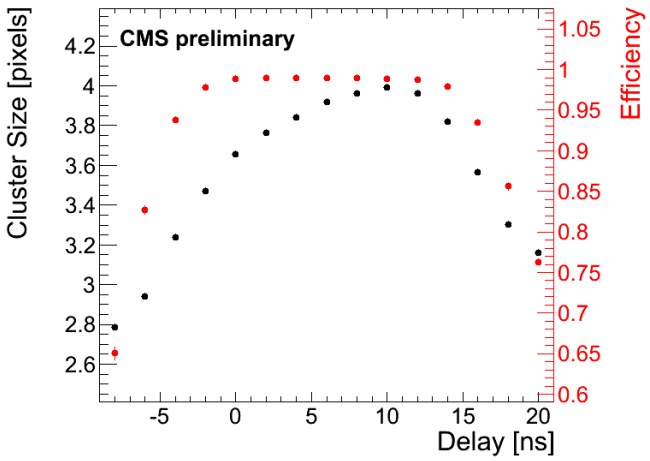

Figure 3.11: Result of the delay scan, with 2 ns steps. The best value is with 10 ns, where the efficiency is high and the cluster size the highest.

The internal pixel clock and the LHC clock received by the onboard chips need to be adjusted to each other. Indeed, the LHC creates collisions with a very precise timing, and the clock sent by the LHC to all the detectors is received with a certain delay after the actual beam crossing, due to the electronics as well as the cable lengths, which both impact the path the clock signal has to take to arrive for instance in the pixel detector. The delay has been firstly estimated to be 153, in unit of LHC clock (so in 25 ns windows), in order to attribute the trigger to the correct bunch crossing, but had to be shifted to 152 once installed to get the pixel detector in sync with the trigger. This data to trigger delay value is set in the WBC register of each ROC.

After this coarse determination, the clock delay needs to be finely tuned within the 25 ns representing the window between two LHC clock ticks. This is done using data from collisions: the pixel clock is shifted by 2 ns steps with respect to the LHC clock, and observables pertaining to the pixel clusters are used to determine the optimal delay. Those measurements are the pixel cluster size and the cluster efficiency, which are represented in function of the clocks' delays in Figure 3.11. The cluster size is relevant and sensitive to this delay as small charges, due to time-walk, would take more time to get collected and to build up a sufficient charge to overcome the comparator's threshold, thus being wrongly assigned to the next window, decreasing the number of pixels included in the current cluster and diminishing its charge. The efficiency is also dependent on the small charges, as tracks could see some of their hits assigned to the next bunch crossing, which would register as inefficiencies. The goal of this calibration is therefore to choose the setting with a maximized cluster size, while staying in the most cluster efficient plateau region, which was achieved during the 2009 data runs for a delay of 10 ns between the

global LHC clock and the pixel one, as seen in Figure 3.11. During the year 2010, there was time to calibrate the detector with more data, and this calibration was performed for different sectors of the pixel detector, leading to an adjustment of maximum 3 ns in the barrel and 8 ns in the end-caps.

3.2.4 Threshold calibration

As other calibrations, the thresholds are measured with a special calibration run. While knowing their value has its use in itself, being able to measure them frequently is paramount to the need of having them as low as possible, as they impact directly the charge that can be recorded. The thresholds can be modified with two global DAC settings that influence each ROC value, while the threshold for each pixel is changed with four trim bits, in order to uniformize inside a ROC the values of all pixel thresholds. The goal of the threshold calibration is to measure the value of the thresholds, which is performed with a calibration run: pixels are injected with increasing charges, and the efficiency of the readout is calculated for each of those value. It resembles a pixel alive test, with its actual efficiency when variating the amount of the charge that is injected, inside a rather low VCAL window to sample both regions where the efficiency is 0% and 100%. The result of such a calibration is presented in Figure 3.12 (left), where the distribution is fitted with an error function: the threshold is then defined as the charge with a 50% efficiency, while the noise can also be measured, since it is represented as the width of the rising portion. Another equal way of measuring those observables is by differientiating the error function, the threshold being given as the mean of the resulting Gaussian, with the noise being its width.

The threshold extracted from the S-Curve represents the absolute threshold, as the readout mechanism induces two types of thresholds, relevant in different physics aspects:

- The **absolute threshold**, which is in effect the real value of the comparator threshold, with no sense of time involved.
- The **in-time threshold**, which is the charge needed to overcome the comparator inside a single bunch crossing.

This differentiation is due to the time-walk effect: charges need a certain amount of time to pass the amplifier and thus register in the comparator, with small charges taking more time than high ones. This means that for collision use, as the pixel detector collects data within a single bunch crossing window of 25 ns, some charges will be affected to the next bunch crossing. Therefore the in-time threshold is particularly relevant for data taking, as it will have a direct effect on the smallest charge that can be recorded, while the absolute threshold is still the real detector threshold, relating to the actual pixel's electronics. The in-time threshold is always higher - by around 1000 electrons for the calibrations of 2008 -, meaning charges in-between the two types of threshold need more than the alloted 25 ns to exceed what the comparator needs to accept the signal. The in-time threshold has therefore an effect on the reconstruction, mainly on the position resolution - more pixels in a cluster increases the resolution - as well as the hit efficiency, while on the contrary the absolute threshold dictates the hit occupancy, with a small threshold decrease possibly leading to sharp hit rate increases.

The absolute threshold is dependent on the electronics of the pixel, especially the noise and the cross-talk. Therefore they have been set just above the ROC cross-talk value, as if set lower it would completely fill the column buffers with spurious hits, preventing real hits to be recorded. To do this, an iterative method is used: the thresholds are first put

to a quite high value, and decreased by 300 electrons steps; for each step, a pixel alive test is run, where failures are assumed to be filled buffers, leaving no place to store the actual result of the injected charges, meaning the thresholds are too low. The values of a previous step are then used, hence setting the minimum threshold to avoid cross-talk in the most sensitive channel of the ROC. This trimming of the thresholds is very time consuming, even if some methods to increase its efficiency, like sampling only 81 pixels per ROCs, have been implemented.

For the year 2008, for the first set of calibrations, there was enough time to trim the thresholds of the end-caps, while the bpix settings were taken from module testing performed prior to the arrival of the detector at P5. Figure 3.12 (right) shows the distribution of the thresholds values for both bpix and fpix. The absolute threshold is determined to be 3829 e^- for the pixel barrel, with as expected a smaller value for the end-caps of 2941 e^-. The noise was measured at 141 and 85 e^- for the barrel and the forward pixel respectively, with the noise distributions having their RMS at 35 and 26 e^- respectively, as is detailed in [23]: this accounts only for the readout electronics of the pixel, the other ones adding around 300 e^- to this signal. It is much smaller than the cross-talk value, which is why the latter one is set as limit for the threshold, meaning the noise is also negligible when compared to a signal charge.

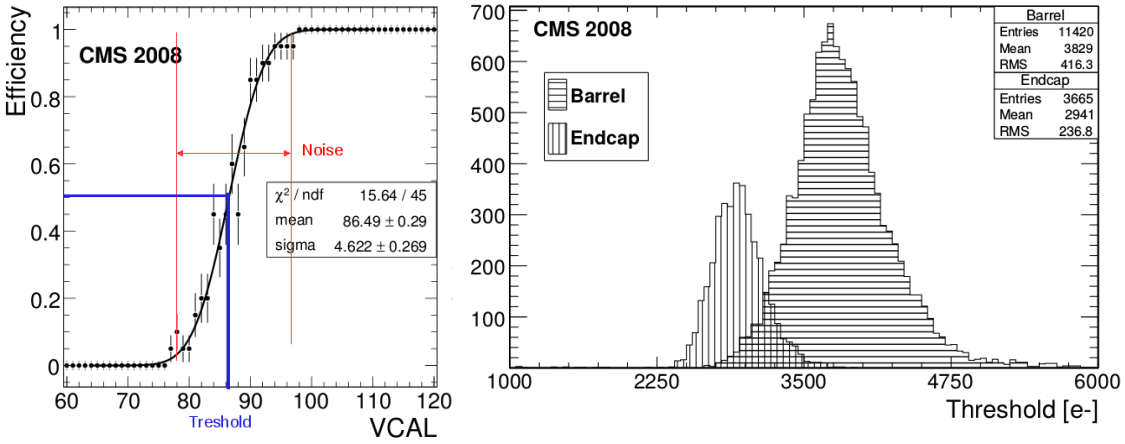

Figure 3.12: (Left) S-Curve for one pixel, with the fit giving the threshold and noise values. (Right) Distribution for both bpix and fpix of the thresholds.

3.2.5 Address level calibration

As explained previously, during the readout sequence, the address of every pixel is encoded using a base-6 scheme, with a 10-bit analog ADC signal sent over five clock cycles. The distribution of the chosen six discrete values to code each clock cycle, for each pixel of one ROC, can be seen in Figure 3.13 (left plot) for an address calibration run done before the Cosmic Ray Run in 2008. The goal of this calibration is to make sure the peaks are clearly identifiable, and well separated: an overlap would mean that sometimes the discrete value sent can be misidentified with another one, thus rendering the address decoding error-prone. To perform the calibration, the charge circuitry on each pixel is used, and all addresses are decoded by the FEDs. In order to qualify the health of the address coding, the RMS of each peak for all active ROCs is computed (see Figure 3.13 central plot): the mean of all the RMS is at only 3 ADC count, whereas the tail only

has values around 7 ADC maximum, to compare to the 1024 ADCs possible in the 10-bit signal. The right plot depicts the separation between two adjacent peaks for every ROCs by computing sigma, the quadrature sum of the adjacent peaks RMS: with a mean around 20, it demonstrates the good splitting of the coding base, and ensures the pixel addresses will be decoded correctly.

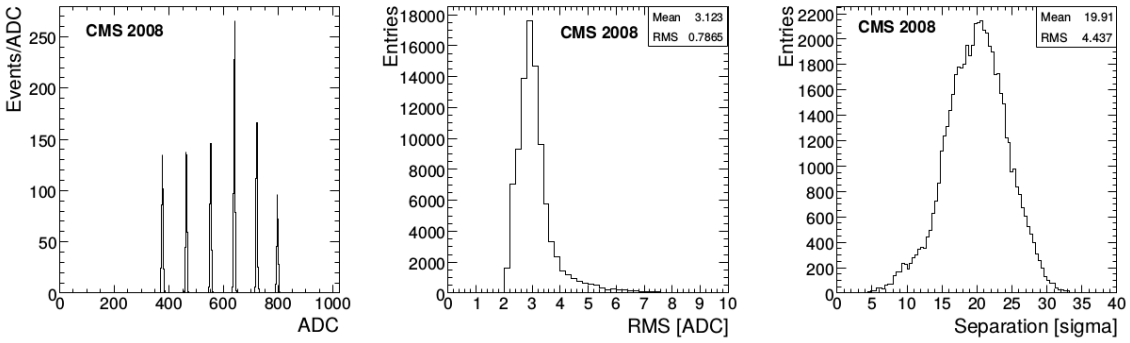

Figure 3.13: The left plot presents the distributions for the six different ADC levels used to encode the pixels address within one ROC. The central plot presents the RMS in ADC of each peak for all the active ROCs. The peaks are all separated, as shown in the right plot, where the separation between the mean of two adjacent peaks is plotted against sigma, the quadrature sum of their RMS.

3.3 The gain calibration of the PIXEL detector

This section will detail the first and one of the main parts of my research devoted to the pixel detector: the so-called gain calibration. It is part of the calibration procedures that are run regularly, and therefore requires expert people to run the calibration and understand the data. Contrary to the previous calibrations that are merely summarized here, a complete description of what the calibration entails will be presented, followed by a comprehensive explanation of its results, along with their evolution with time. Finally, its aspects within the CMS simulation will be detailed. Although quite long and maybe technical, this section is for now the only written record describing precisely this calibration, as unfortunately no specific literature pertaining to the CMS pixel gain calibration has of yet been published.

Although the purpose of the gain calibration is to convert charges stored in ADC counts back to electrons, the main reason it is performed is to keep the charge response of the pixel detector as uniform as possible between all pixels, as this has a considerable impact on several other measurements. This is due to the charge interpolation techniques that are applied during the reconstruction of clusters of pixels resulting from a single particle crossing the sensor. The reconstruction of the shape of the clusters makes use of templates [24], which are pre-simulated shapes of clusters done with a detailed simulation and depending on the incoming angle of the particle's trajectories. They considerably improve the resolution measurements, which where found to be much smaller than the width of one pixel (see Section 3.4.2.4), leading to much better accuracy in finding secondary vertices for instance. Charge sharing between pixels is therefore of primordial importance and has been enhanced whenever possible, which explains for instance why the blades in the end-caps were slightly tilted as is mentionned earlier, to allow charges to drift because of the Lorentz force. The gain calibration is hence necessary to ensure the charges in each pixel of a cluster is measured in an uniform way.

3.3.1 Introduction

As explained previously during the description of the pixel detector, the charge deposited by the particle crossing the sensitive region of each pixel is a number of electron, but the read-out chain of electronics stores the information in ADC counts. In order to determine the conversion factor between charges in electrons and in ADC counts, each pixel has to be sampled over a wide range of possible charge deposit. It is why a specific injection circuit was built in the pixels' circuitry, so that the conversion factor and its calibration could be obtained at will. Figure 3.14 shows such a sampling, where the pulse height measured in ADC is plotted in function of the injected charge. The charge that can be injected through the electronics is not known in number of electrons, but is quantified in number of VCal units. To obtain the value of one VCal unit in electrons, x-ray tests were performed on some of the modules, and measured the mean value per ROCs:

$$Q[e^-] = 65.5 \times Q[VCAL] - 414 \tag{3.1}$$

$$\Updownarrow$$

$$Q[VCAL] = \frac{Q[e^-] + 414}{65.5} \tag{3.2}$$

On the other hand, the curve giving the ADC value in function of the VCal charge is nicely fitted by a tanh parameterization:

$$Q[ADC] \quad = \quad p_3 + p_2 \cdot \tanh\left(p_0 \cdot Q[VCAL] - p_1\right) \tag{3.3}$$

where each four parameters has its own influence on the distribution. For instance the p_1 parameter influences the linearity at low VCAL, which can be of great importance. As can be seen, the rising portion of the curve is mostly linear: this means the response of the detector to charges deposited in the sensor can be approximated to a simple linear function, reducing the number of parameters to two: the slope and the y-intercept. The whole goal of the gain calibration, for the pixel detector, is then to determinate for each pixel those two parameters, in order to transform the detected charges stored in ADC units back to charges in electrons. The parameters of the fits have specific names related to this calibration: the gain is defined as the inverse of the slope, while the y-intercept is called the pedestal. The relation between ADC and VCAL charges is hence obtained with:

$$Q[ADC] \quad = \quad \frac{Q[VCAL]}{gain} + pedestal \tag{3.4}$$

The example of conversion curve in Figure 3.14 demonstrates such linear fit, resulting in a gain of 3.09 and a pedestal of 24.7 ADC counts.

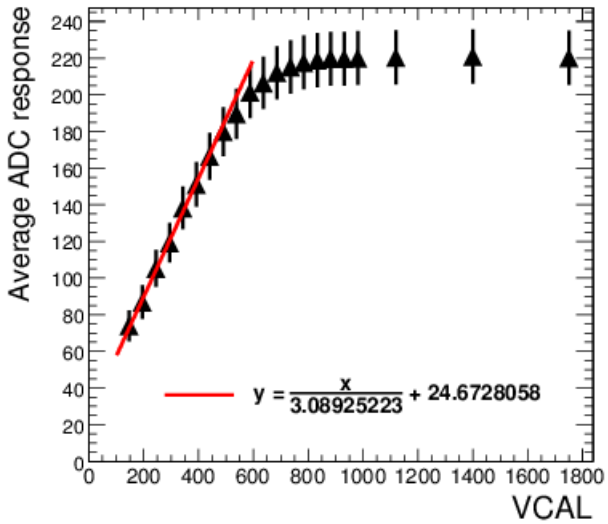

Figure 3.14: Example of a fit of the gain calibration curve for one pixel that was successful.

3.3.2 Calibration Procedure

The Data Taking

The first step of the gain calibration is to perform the run calibration. It is a rather long process: each pixel needs to have the whole VCAL range sampled to test its readout. This is done by injecting pixels with a charge of known VCAL value, using the ROC readout

sequence to send the data to the FED, and repeating the operation while increasing the injected charge. Each VCAL point is sampled five times per pixel: this is done to diminish the impact of misbehavings and ensure a correct ADC readout value by the FEDs, the error on the computed mean for each VCAL input giving then information on the reliability and stability of the charge injection mechanism. A whole run takes more than five hours to perform, which means they cannot be performed on a whim, as it takes a lot of detector time that cannot be spent on something else.

As each pixels gets tested many times - between 10 to 20 different charge values, 5 times per value -, it is not possible to inject every pixel channel in the detector every time: a scheme has been developed in order to efficiently sample pixels within a ROC without creating bottlenecks with the readout speed. Each pixels of a ROC are thus fully sampled over many clock cycles, which explains the duration of the calibration run. The data read by the FEDs are then directly sent to VME PCs, bypassing the standard CMS reconstruction, as it is quite specific data that can't be analyzed with the main software.

The data is stored in 40 files, one per FED, and is analyzed offline using software designed specifically by the pixel team. A special calib.dat file is also produced for every run, and details the injection scheme that has been used to perform the calibration.

The Offline Analysis

In the following sections, the packages mentioned in italic *are part of the global CMSSW software.*

The offline code makes use, for historical reasons, of the Data Quality Monitoring (DQM) packages. It uses a specific digitizer plugin, that takes in the RAW information from the calibration run, and transforms it into digis, packaging the information by module detid. The gain calibration analysis is performed on top of it, iterating over every module. This offline code is localized in the *CalibTracker/SiPixelGainCalibration* package, while the python configuration files needed to start the analysis are in the *DQM/SiPixelCommon* package. Through the years, I wrote many scripts to automatize as much as possible the analysis of the gain calibration, in order to first shorten the time needed to go from running the calibration to having the results and decide whether or not to put to use the data, and secondly to minimize the amount of manual work needed for each analysis, as quite a few of them had to be done, which also had the positive side effect of standardizing the workchain and how all steps were done, meaning it was simpler for others to perform the offline analysis in case I could not.

The first step of the workchain is to transfer the 40 root files containing the RAW data stored by the 40 pixel FEDs, along with the dat file containing the injection scheme: they need to be moved from the VME crates in P5 to the central CASTOR storage system, in order to be used by the CAF queue system. This is done to allow for parallelizing of jobs: the VME crates are not powerful enough to run simultaneously 40 times the offline analysis code on each root file, which is the reason for the CASTOR storage and the CAF queue. This queue works such that as many jobs can be sent in one go, and the system balances the number of simultaneous jobs per user, managing at best its humongous number of available cpus to keep the waiting time at a minimum while giving everyone a chance. Therefore, from more than a day if the analysis had to be run on one file after another, the time required drops to around one hour when all jobs are started by the system at the same time, providing a very short period between end of calibration and first look at the output. While the CASTOR storage system is present for every CERN member to use, the

CAF queue is meant specifically for calibration workflow, with dedicated servers to ensure those mandatory analysis required to look after the CMS detector have priority above the standard physics analysis, and usage permission is hence delivered with parsimony.

The main goal of the gain calibration code - after the DQM transformation is operated - is, for each pixel, to create the calibration curve (example in Figure 3.14), to perform the fit to get the gain and the pedestal, store all the relevant data, and finally produce results to first permit an understanding of the quality of the calibration when compared to previous one, understand the changes if any, and eventually allow for insertion of the constants in the database to be used further on during reconstruction.

The production of the calibration curve is done as follows: first, as each charge in VCAL has been injected several times, the ADC response for each VCAL data point is obtained by computing the mean of all data sets, along with its related error. Then comes the most complex task of this calibration: the selection of the range of the fit, and which VCAL points should be included and excluded. The main purpose of this step is to fit as best as possible the rising part of the curve - which represents the main region reflecting the charges that are deposited by charged particles in the sensor - with as less influence as possible coming from the saturation portion at high VCAL or the small curving at low VCAL. It requires first to detect the plateau ADC value, and exclude all data points with an ADC ordinate higher than a certain percentage of this plateau value. It is therefore important to get a precise estimation of the saturation value: as some more code was implemented to improve this measurement, the exclusion value went from 80% to 90% of the ADC plateau. This was made possible with a major rewrite of the plateau detection code, making it less error prone and more adaptable to the various calibration curves possible (see Appendix A for examples of odd behaviors). The effect of the small curving at low VCAL was solved in two ways: first, after its effect - though small - was brought to light, several DAQ settings influencing this portion of the curve where small charges are injected were modified; secondly the analysis code was also adapted to exclude if still present those few points, making sure enough points remain to perform the fit adequately. The work carried out on those two aspects greatly improved the fit, mainly by increasing the number of valid data points that could be used, but also by increasing the amount of pixels with less standard gain calibration curves that were consequently correctly fitted.

Once the data points that should be used to perform the fit are selected, a new calibration curve is drawn containing only those values. The linear function is set with defaults of 50 for the intercept and 0.25 for the slope. The fit is then executed with the `ROOT` framework, and up to ten iterations are completed until both free parameters have been found. Once this is done, several parameters on the fit are extracted, and stored within 2D histograms representing a whole module, which are themselves stored using the DQM booking system: each module has a its own directory, containing one 2D histogram per variable, and the output root file directory tree follows the physical structuring of the pixel detector, with shell/ladders for the barrel, and half-disks/blades/panels for the end-caps. The observables stored for each pixels are detailed, with their definition, in what follows:

- **Gain:** is the inverse of the slope of the linear fit.
- **Pedestal:** intercept value (y value of the line for $x = 0$) of the linear fit.
- **Error on the Gain** value computed during the fit.
- **Error on the Pedestal** computed during the fit
- χ^2**/NDOF:** value of the normalized χ^2 of the fit, where the Number of Degree Of Freedom (NDOF) is the number of points included in the fit minus the number of fitted parameters - here 2 since it is a linear fit.

- χ^2 **Probability** computed after the fit, using NDOF.
- **Number of points** used in the fit.
- **Lowest Point** in VCAL where the fit starts.
- **Highest Point** in VCAL where the fit ends.
- **Dynamic Range** of the fit, which is the difference in ADC between the ordinates of the highest and lowest fitted points.
- **The Plateau** value, defined by the flat portion of the gain calibration curve at high VCAL values, and due to the saturation of the ADC response of the pixel readout.

An example of all those variables can be seen in Figure 3.15, for a module in one of the end-cap disks. The two rows of 4 ROCs are clearly identifiable, as behaviors between ROCs can differ, and gain values for pixels at the edge of the chips are higher than the average, delimiting the border of the ROCs. Although this is only one module, it confirms the relative stability of both gain and pedestal values between a column, which is crucial for the database storage explained in the next section.

The gain calibration curve cannot be stored for each pixel, as their sheer number would by far exceed the space capabilities required to hold on to all curves. This is particularly penalizing when the output of the offline analysis shows differences between the new calibration and previous ones, or when unexpected features appear, be it for just one ROC, or even full modules, or a general trend - which can all come from either misbehavings or even failures of some detector components, changes in the settings of the pixel detector, often done to improve other domains, or modifications not specific to the pixel that still impact the calibration -, and that require some specific scrutinizing. To help with the detection of specific features, and to increase the efficiency of their comprehension, I implemented two approaches in the gain code:

- The first one added was a **logging capability**. Text files are produced identifying the module, then the row and column of pixels for which the fit was not successful. This allowed for an easy parsing of specific parts of the detector that had problems, and a first grasp on what it was. A final summary log is produced detailing the number of modules, pixels that the code was run on - just as a very basic but necessary input on the quality of calibration - and, once the status of the calibration was implemented (see below), the number of pixels with specific status. The mere number of valid fits in the barrel and end-caps is already a good information on how well the calibration went, and can already dissuade from any further study of the results, in which case another gain calibration run would often be scheduled whenever possible.
- Over time, a **fit status** was also implemented, as recurring behaviors and pixel curve with failed fits were identified. Indeed, all gain calibration curves are alas not as well formed as the one in Figure 3.14: the fit failing can ensue from several types of bad curves, like some noisy ones which always saturate the readout, or some with only a few data points. Those features are then categorized, and each pixel fit is given a value stored as are the other relevant variables, in a 2D histogram per module, with positive status value if the fit was successful, and negative one if it failed. In order to decrease the time needed to identify and understand the failed fits, a few pixel curves are stored per module for each value of the status, as running the whole analysis again to produce any new plots to comprehend the features would be time consuming. Although the gain curve could be more or less induced from all the

correlated variables that are stored, this usually makes it possible with a glance to deduce what happened, even if the way to correct it is less obvious, either with code modifications or changing conditions of the calibration run. Some problems encountered during the fitting procedure are detailed in Section A, along with examples of such odd behaviors, and hence a detail on the statuses used to identify them.

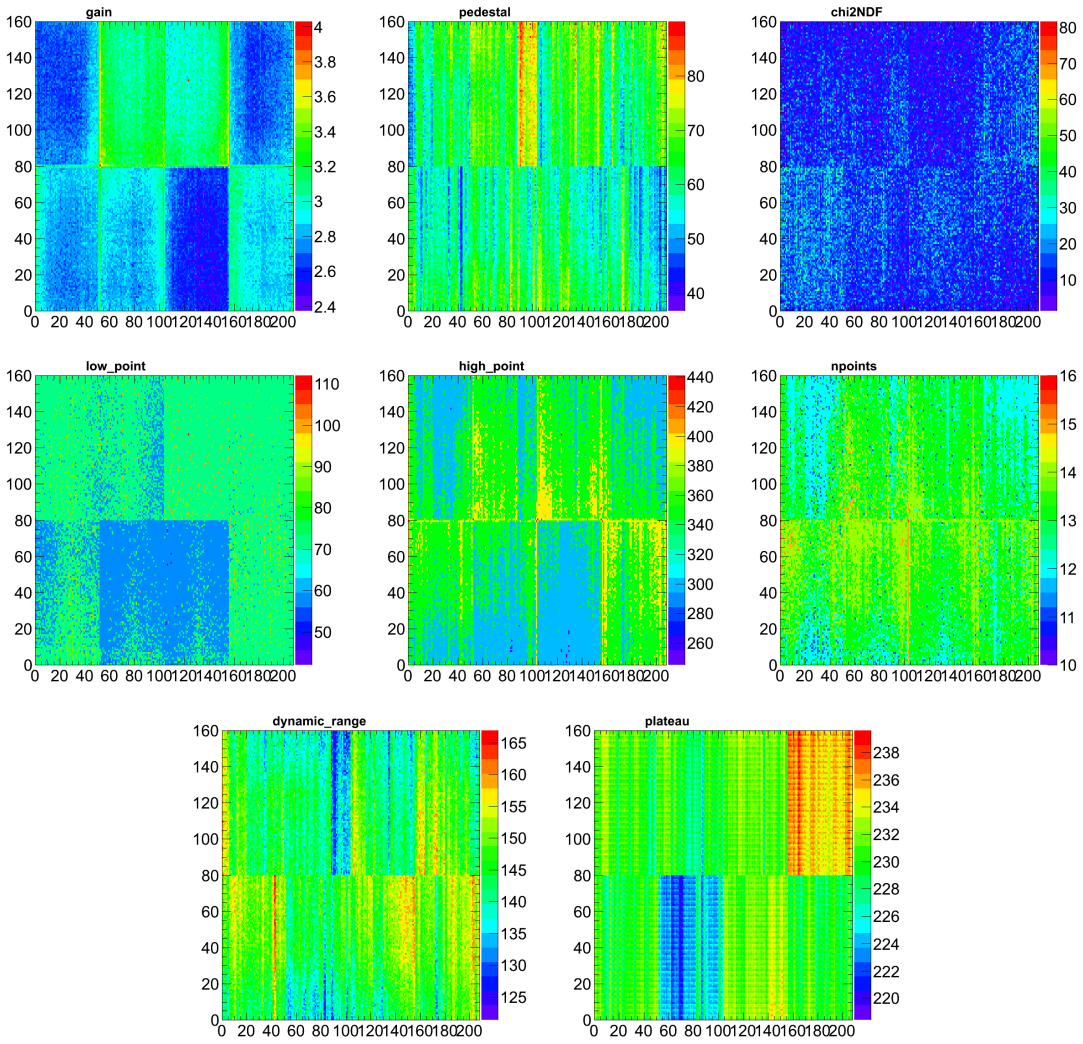

Figure 3.15: 3D plot of several variables - obtained during offline analysis of a gain calibration run - for a module in one of the end-cap disks, with two rows of 4 ROCs. Each point is the value for one pixel of the module, where its x and y coordinates are the pixel's column and row number. The variables presented here are respectively: the gain, the pedestal, the χ^2/NDOF of the fit, the abcisse (in VCAL) of the lowest point used in the fit, the abcisse of the highest point used in the fit, the number of points used in the fit, the ADC range between the lowest and highest point in the fit, and the computed ADC value of the plateau of the calibration curve.

Once the 40 jobs are finished, the 40 output root files containing the results are concatenated into a single one, with an average size of 1.5 GB. The final task is then to shape those results such that they are easily understandable, and to carefully look at all variables and identify possible troubles. Examples of the result of the analysis is shown

in Section 3.3.3.

Storing in the Database

As explained before, the ultimate goal of the gain calibration is to provide for each pixel the conversion constants to go from ADC counts back to charges in number of electrons. This is used during the reconstruction, specifically in the clusterization step: the charges of each pixel is recorded in ADC counts, and its value is then calculated in electrons, along with the corrections applied from the Lorentz angle and the pixel templates. In order for this conversion to happen, the constants obtained from the calibration run must be used. This is done in CMS using a specific database (DB), where calibration constants are stored for each subdetectors. This database has therefore specific constraints: as it will be used for every job running the reconstruction, it needs to be easily accessible every time, to not cause bottlenecks and delays or increase the processing time alloted to the reconstruction step. This means the database cannot be too big space-wise, this for two reasons: first its transfer requirement, and secondly its memory imprint. Indeed, every computer needs to have access to it to get the constants it wants, hence it has to be transfered every time, as all different versions acquired over time of all the different sets of calibration constants cannot be stored on every computer just in case one of the constant sets *might* be used. It is therefore mandatory that the transfer of the database information from the squid servers - the ones that actually do hold all versions, and that are used solely to serve the correct variant of the constants to the machine hosting the jobs asking for them - through the network up to the grid servers does not hog up all the bandwidth, as it it absolutely not the only data transfer that is operated down to the grid machines. This puts a first cap on the database size, although it is not its main constraint: once on the server that required the data, it is stored in memory for efficient access. As machines have a very finite amount of memory, that cannot be extended painlessly, and since servers host many jobs in parallel to maximize the efficiency of cpu usage, it was decided that reconstruction jobs could not use more that 1 GB of memory. This is a very strong cap for any set of calibration constants, especially for detectors with millions of channels. As it is not the only object that uses memory during the reconstruction work, it means that all DB constants for all the subdetectors must be quite small, well under the GB limit.

Ideally, the gain calibration would be fitted by a tanh function with four parameters, as explained previously. Although this would fit perfectly the calibration curve, it was deemed enough to perform a linear fit of the portion of the curve that rises, as the pixel detector was designed to have the collected charges within those particular values. From four parameters, the linear fit allows to decrease to two parameters to be stored, which is already a big improvement, especially in view of the database size constraints. To possibly reduce even more the space used by the gain calibration constants, the effect of storing only the mean values per column - as the pixel detector has a column drain mechanism, and thus pixel share part of their readout chain within a same column - was studied: it was found that the gain value was very stable inside a ROC column, with pixels within 3% of the average value, which is smaller than the uncertainty of the injected calibration charge. The Figure 3.15, which granted is only for one module, is still a prime example: it is clear that many of the observables plotted have a rather stable value within a ROC column, but can differ in a visible way between columns. The pedestal variation in a column is higher than for the gain, but still rather small. Those considerations, explored during the test beam of the pixel detector before it was introduced in the CMS cavern, were extremely beneficial when deciding what granularity to input in the database.

The code used to read the root file produced by the offline calibration analysis, extract the gain and pedestal constants, perform the average calculation for cases where the column granularity is required, and finally store everything in a db object ready to be uploaded to the central CMS condition database, is located in the *CondTools/SiPixel* package. In order to anticipate any possible future changes to the current decisions, the code was written such that granularity on the pixel or column level could both be chosen for the gain and/or the pedestal constants. For the offline reconstruction, a granularity on the column level was chosen for the gain, while the pedestal information was kept on the pixel level. This results in a db file weighting around 70 MB, the majority of which is taken by the pedestal data as there are 80 pedestal constants stored for one gain constant - remember, each ROC column contains 80 pixels. The same consideration were applied concerning what granularity to store for the reconstruction used by the High Level Trigger. Since it differs from the standard one mainly by requiring a much faster reconstruction as the trigger decision has to be made within hundreds of milliseconds to not loose any detector data, we ended with deciding to use a smaller database object, therefore storing both gain and pedestals on the column level. As the HLT is only there to select relevant events, the effect of using column averages rather than the per pixel value when reconstructing the electron cluster charge was deemed negligible, especially in light of the gain a lightweight database object brings. The database file for the HLT holding only column averages is thus only 1.8 MB, considerably smaller than the offline one.

The storing of this information inside the database file is also carefully thought, with the key concept being to use as less space as possible. For this, each constant, be it the gain or the pedestal, is encoded using an 8 bit scheme, leading to 256 possible values for them, from 0 to 255. The values 254 and 255 were employed to encode whether a pixel (or column if the value represents a column average) is noisy or dead respectively. While this feature was not actively used, the implementation was still required just in case. The values of the gain and pedestal were therefore encoded making use of their maximum and minimum, dividing the possible range into 254 values to ensure all could be encoded. This means that the maximum and minimum of both gain and pedestal are rather important, as it influences directly on the stored values: if the range is too large, like from 1 to 25 for the gain which is not uncommon as it takes just one bad fit, this would lead to the bulk of the gain values, which are usually around 3, to be encoded with just a few of the 254 possible values, with no differences within a 0.1 spread. This is one of the reasons to detect wrong fits and implement a status, as pixels flagged with bad statuses can then be excluded from the maximum search and have their value set to the mean of the column/module or excluded if not needed when averaging on a column. The code hence contains instructions to prevent too wide ranges, with hardcoded boundaries whenever a value goes beyond them.

This specific 8-bit choice was operated to once again minimize the storage space: it represents one byte, which is the storage size needed for a character in `C++`. This ensures no space is wasted when storing the information: each encoded gain and pedestal value is transformed into a character, thus without any lost bit. The database object is then composed of blobbed[1] vectors for each modules containing the gain and pedestal constants, storing along the module detid to identify it. The ranges of both gain and pedestals are also kept inside the database, in order to make it possible to decode the values when reading the database.

[1] a blob, in computing, is an object for which the container - here the database - has no knowledge about, unlike integers, floats or characters. Complex objects like images and video can be stored this way inside databases.

Finally, when the constants have been validated, the DataBase experts need to be informed that new constants exist for the pixel gain calibration, and after a presentation during a weekly meeting to show what it entails, the specific database file is pushed to the central CMS condition database, which is then replicated on every squid host to be used by everyone. When putting new constants inside the global database, an Interval Of Validity (IOV) has to be specified: the constants are indeed pertinent only for a set period of time, as they usually replace older ones because changes in the detector conditions mean the previous constants do not reflect the current state of the detector. This IOV is defined in internal CMS run number, as they always are defined in time, are always increasing and are widely used to set the boundaries of the datasets involved in physics analysis.

Validation of the Calibration Constants

The validation of the constants extracted from the calibration is a compulsory step before they can be used widely by the CMS community. It ensures the good quality of the calibration data, and detects the possible hardware misbehaviors or even component failures. The best way to perform this task is by looking at the output of the offline analysis: it is why an important amount of time has been devoted to the presentation of the gain and pedestal constants, with many different plots, so that behaviors deviating from the normal could be easily spotted, their location identified, and if possible a first understanding provided through the correlation of all the different information available. Another way to check the quality of the data is to compare it with a previous calibration, which should have already been worked on and therefore have all possible features understood. In order to automatize this process, I built the code constructing the result histograms such that with a simple switch, the macro would use the two root files to create exactly the same plots as for a single output, with for each variables looked at, the difference between the values for each calibration instead. If the calibration runs were similar, all variables would be close to zero, making it very simple to spot new features, or old ones disappearing. A comparison of the status pixel by pixel also allows to see rapidly whether or not the calibration has improved overall.

The study of the output of the calibration was the main validation required to guarantee its quality, and was often the only one performed, especially when changes in the detector conditions meant the current database conditions were anyway obsolete: as long as the new calibration was deemed good, without any problems that could be fixed with a new calibration run, and reflecting the state of the detector whatever it is, it was considered mandatory to push it into the condition database. But it is also important to check the effect of the new constants during reconstruction, to ensure a continuity within the reconstructed data: creating artifacts solely due to new calibration constants being implemented is not an option, when doing physics analysis. This is why a second validation process was often asked for and appreciated: data would be reconstructed with the current constants from the database and with the new extracted ones, and basic pixel observables like cluster charge and size would be compared. A small framework to automatize this task, along with the creation of the resulting observables was developed. Contrary to the first validation step, this one increases considerably the period needed between running the calibration and inserting new constants to the database, as first the reconstruction is quite time-consuming, and secondly it requires data adapted to the new detector conditions to be relevant, and therefore means one had to wait for either cosmic or collision data to be acquired after the calibration was finished. This validation step was also often the subject of fierce discussions, as it was not always deemed relevant, since both sets of constant could not be necessarily applied to a same set of data, the conditions with which

they were taken often leading to one calibration being inadequate. Moreover, what to do when the calibration was understood and reflecting the detector's state, but introduced slight differences within the reconstructed data, was not established as a new calibration run would produce the same results.

3.3.3 Results

The output of the gain calibration's offline analysis is a ROOT file with more than 10 histograms stored per module. Knowing there are around 1400 modules in total, it is impossible to look at every plot, and uncover potential issues with the calibration. In order to help visualize the result of the analysis, the observables are plotted in many configurations, that either summarize the information, or lay it out to be able to discern abnormal behaviors. This is performed with a macro that produces plots, a general summary, shapes everything inside a pdf using LATEX, and stores everything on a twiki for an easy access for everybody. Some examples, albeit only a fraction of what is available, are presented in Figure 3.16, where the gain constant is shown in different ways:

- First with the mean of each barrel layers and end-cap disks, showing it is not that uniform between sections of the pixel detector.
- Second with the distribution of the gain for each pixel in the barrel, showing the majority of the pixels have a gain between two and five, with the bulk being around three. This kind of pixel distribution - for barrel, end-cap and encompassing both - is also present with the ordinate scale in logarithmic form, in order to observe the presence of tails going to higher values, or groupments signifying a whole column, ROC or module shares a similar trait, for which cases distributions with column and even ROC averages have been implemented as well.
- Third with 2D correlation histograms, where the number of pixels - plots with number of columns or ROCs are also available to detect whether features belong to a column/ROC or are on the contrary spread everywhere - in bins of gain and pedestal values is shown. It clearly indicates that fractions of the detector have behaviors diverging from the normal. Correlation plots with the gain or pedestal and their respective error are also present, along with the correlation of the errors of the gain and pedestal extracted from the fit.
- Fourth with the mean of the gain for each module, in function of the module's detid, unique number referencing it. Although the detids are not readable on a fixed plot, the regions of the detector they belong to can be intuited by expert eyes, as the detids are nevertheless ordered. Moreover, the ROOT framework allows to zoom in in any directions, making it possible to know the detids of specific modules if needed. This type of distribution also simplifies the comparison with other calibrations, as plotting several of them on top of one another makes it easy to spot evolutions per module. It is clear from this distribution that there are variations between modules, and even regions, which is another way to draw the first histogram, although this one is more detailed.

All the variables are also shown in 2D maps, where each module is represented by its average value, with one plot per layer and one per disk: Figure 3.17 presents those 3+4 maps for the gain followed with the pedestal. The layout of the modules follows the online convention, in order to be the usual for the pixel experts: modules are divided between each ladder, and numbered with respect to the CMS frame. For the end-cap disks, the x axis uses a combination of blade and panel number to differentiate each one of them,

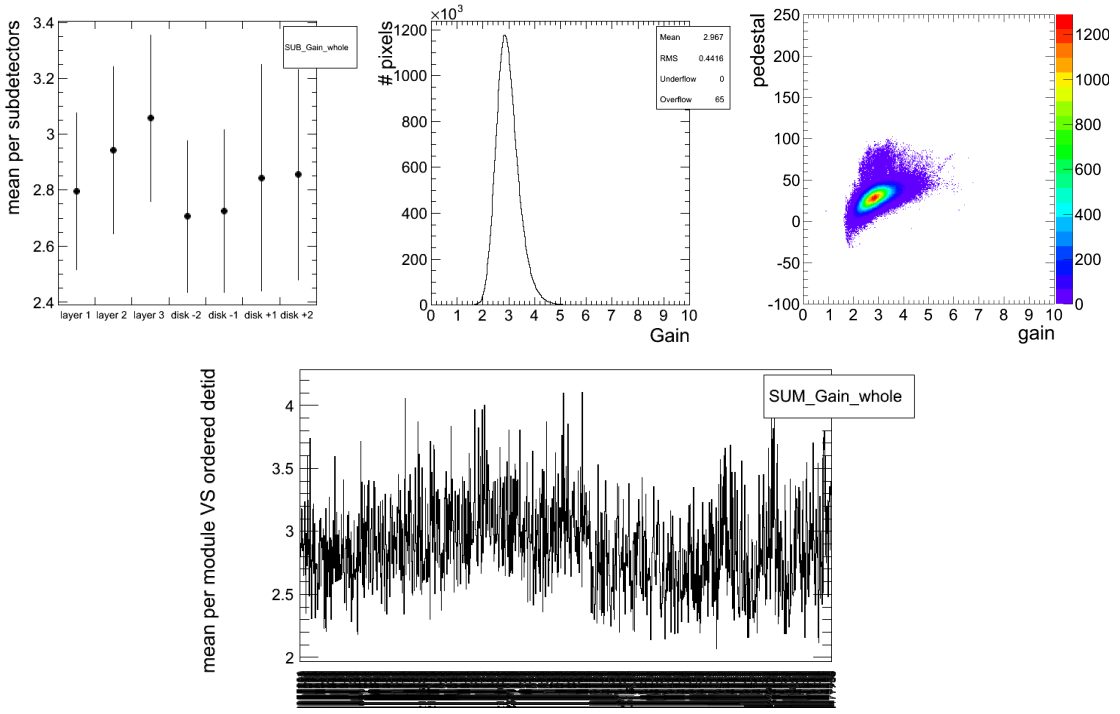

Figure 3.16: Different representations of the extracted gain value: the first plot shows the mean per subcomponent of the pixel detector (3 barrel layers and 4 end-cap disks), the second one is the distribution of the gain values for all pixels in the barrel, the third plot shows the correlation of the gain with the pedestal with one entry per column (one entry for the computed mean for each column of the gain (x-axis) and pedestal (y-axis)), the fourth plot shows the mean gain of the module in function of its detid for the whole pixel detector.

while the y axis is the plaquette number, with either three or four plaquette per panel (see Section 3.1.1 for the particular geometry of the forward pixel). It is effortless to spot from those colorful 2D maps the modules that have different constants, and find out the value of their other observables, to try and understand the feature. For instance, there is a section on disk +2 (first disk plot), between 2*blade+panel-1 equal 15 and 20, with an average pedestal twice as high as the usual: such divergence is also seen in the maps representing the ADC range of the fit, which is smaller than average for those blades. This aspect was thoroughly inspected, and it was uncovered that those blades did not consistently present this odd behavior: therefore it was decided to use data from a calibration where they were not failing, until they were fixed. This simple example shows how easy it is to spot such features, which appear in several plots.

All those results were presented to the weekly pixel Offline Software meeting, and the quality of the gain calibration was discussed each time, before a decision to use them as new database conditions was eventually taken.

3.3.4 Evolution of the gain/pedestal values over time

Over the years leading up to this manuscript, many gain calibrations were performed and analyzed. Only a few of them though were inserted inside the condition database: they first had to be deemed of very good quality, and with significant differences enough

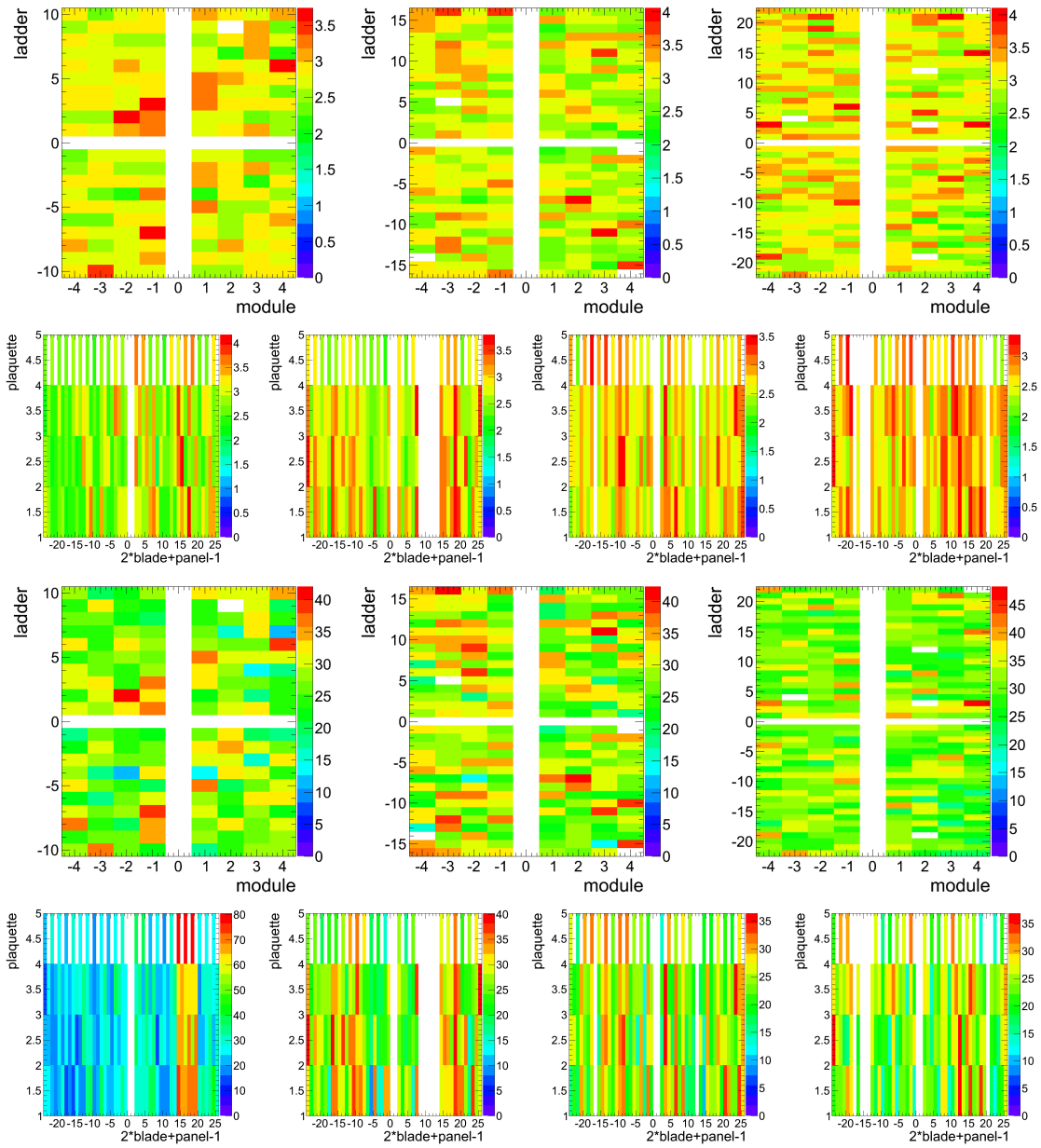

Figure 3.17: 2D representation of the mean gain (first 2 rows) and mean pedestal (last 2 rows) per module, with the whole pixel detector module layout: respectively the layer 1,2 and 3 of the barrel, and disks 1 and 2 of $+z$ side of the end-caps followed by disks 1 and 2 on the -z side. The modules on the layers are represented in function of their ladder and their position within one, while the disks use the plaquette number and a combination of the blade number and the panel number to represent the more complex geometry of the end-caps.

compared to the ones currently in use to justify changing the constants. Obviously, runs were done each time the detector conditions changed, as this would mean a change in the detector's response, even if for just a small portion of the pixel detector, leading to new constants describing better the detector. Plots showing the evolution of all variables were eventually added: Figure 3.18 displays such progression for the gain, followed by the pedestal, with one histogram for the barrel and one for the end-caps. Within the

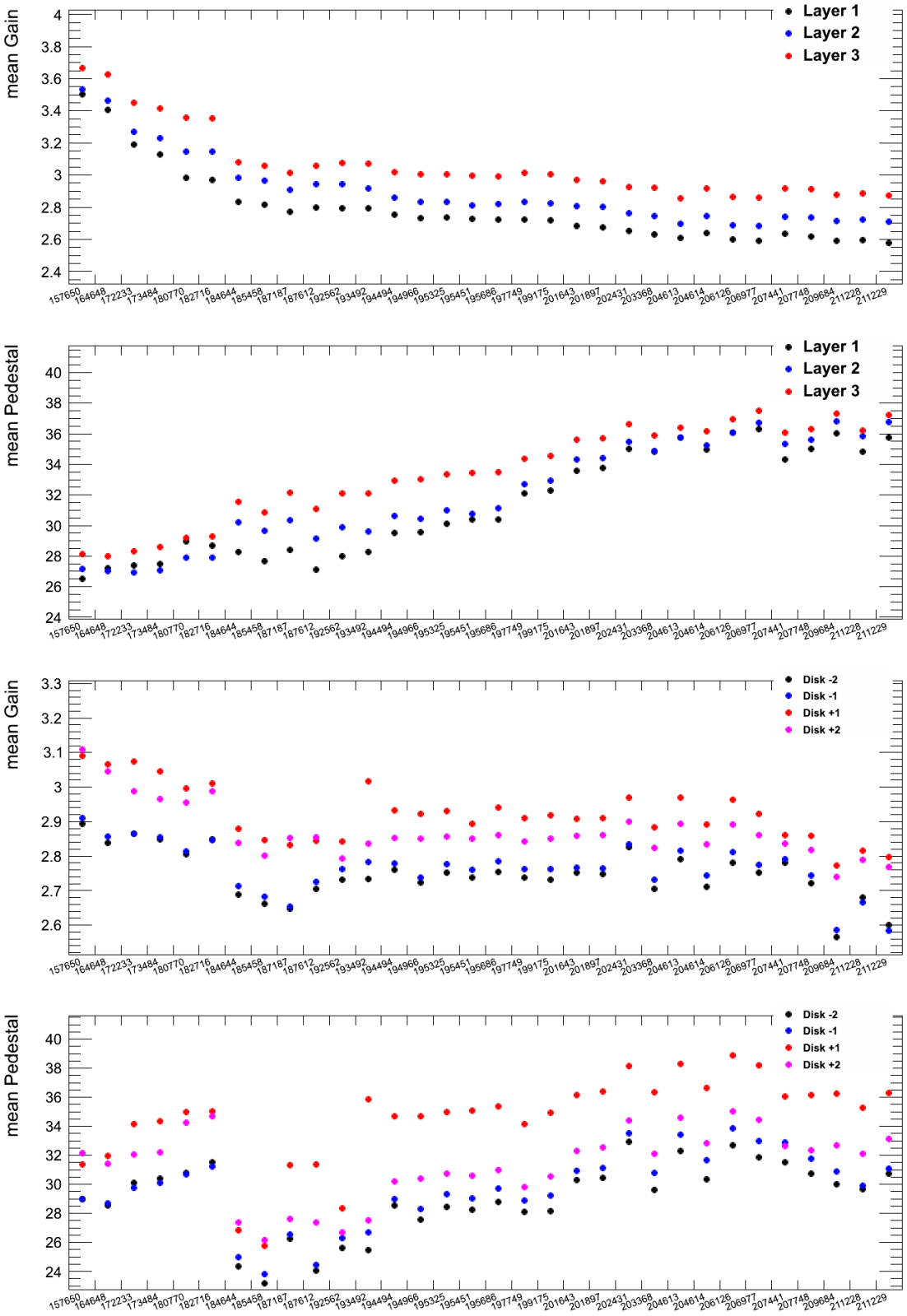

Figure 3.18: Evolution over several years of the mean of the gain and pedestal for all 3 layers of bpix and 4 disks of fpix. Each abcisse number represents an internal CMS gain calibration run number.

subdetector plots, each layers or disks is respectively presented. The evolution is rather smooth, with however clear changes resulting from big modifications of the detector's conditions, as its behavior was better understood and well under control, meaning it was heavily calibrated to outbest its performances. The gain values have a decreasing tendency, while on the contrary the pedestal values were getting higher. It is also clear form this evolution plots that layers and disks all had different values for either gain or pedestal constants, albeit the effect of time was consistent between each of the subdetectors, all values keeping their same gap within one calibration. Better calibration could decrease all the differences between layers and disk, with potentially the same value overall, but this would have required an insane amount of careful and patient work, making sure the changes would not impact negatively other calibrations, and was never started being seen as not very profitable and quite useless, the current status-quo being perfectly acceptable.

This general evolution, making abstraction of the brusk changes, is the result of the pixel detector aging: although the detector is better understood, the sensors and electronics degrade with time, as for the record the pixel detector is the closest to the collision point, and is therefore subjected to the harshest environment. The radiation has a lasting effect overall, that even accumulates, with the gain calibration being one of the analysis where it materializes, giving the evolution plot its general smooth trend. Those results where put together with other studies to get a complete picture of the irradiation damages on the sensors, as it is very important information with respect to the current pixel detector's health, as well as any future detector development as they will have to endure even more radiations, for longer periods of time.

3.3.5 Simulation

The CMS simulation is quite a complex machinery, that is not explained here. Only three small portions of it are anyway relevant to the gain calibration:

- The simulation of the charge deposition and its drift inside the sensor, which is performed by the `GEANT` package included within CMSSW. The charge is simulated in number of electrons.
- The simulation of the electronics of the pixel detector, and the transformation of the charge in electron to a signal in VCAL then ADC, performed in CMS by the readout chain of the pixel detailed earlier. This conversion of the charge from electrons to ADC counts is done during the digitization step, the code being in the *SimTracker/SiPixelDigitizer* package, where the full tanh parameterization using four parameters is used to go from VCAL to ADC: two sets of parameters are implemented, one for the barrel and one for the end-caps, as the constants from the gain calibrations are always different. This means the conversion from electrons to ADC is done with exactly the same formula within the whole bpix or fpix, with no differentiations whatsoever between pixels.
- The inverse operation converting ADC back to electrons - with an intermediate VCAL step, where the last conversion is always done with the same electron value for a VCAL extracted from test beams -, is as usual done during the reconstruction, more precisely in the clusterization step. In order to do the reconstruction, a database with constants specific to the simulation is made, using the *CondTools/SiPixel* package, both for the offline and HLT jobs. There are two main types of conditions that one would want to simulate: one with ideal conditions - named *IDEAL* Global Tag -, meaning it should reflect the health of the pixel detector when in perfect condition, and another

simulating the real detector's condition, with its imperfections - named *STARTX* Global Tags -. As the reconstruction code is the exact same one as the one used for data, the offline databases file must contain column granularity for the gain and pixel level for the pedestal, and weight the same as the data one, while the HLT holds column granularity for both constants. Those gain and pedestal constants are simulated given a gaussian distribution, with its mean and spread extracted from data, with again an overall set of parameters for the barrel and another one for the end-caps. In order to simulate data as close as collisions as possible, the START global tags have gains and pedestal smeared, and make use of the dead pixel flag (see Section 3.3.2) to insert a certain amount of dead pixels. On the other hand, IDEAL constants have no smearing, hence use only the mean of the gain and pedestal extracted from data for bpix and fpix, and of course do not simulate dead pixels.

Overall, this way of simulating the response of the pixel detector, shows a very good agreement with the data obtained with cosmics or collisions. While some features are not completely understood, no modifications were required over time, and the simulation was kept very stable through all the collection of data.

3.4 The commissioning of the PIXEL detector

At the end of 2008, after the LHC incident and the start of the year long no-collision period, the CMS experiment decided to go through a data taking of cosmic rays. After the calibrations performed subsequently to the pixel installation in the cavern, it was the opportunity to commission the CMS detector with the data it could get: while being 100 meters under earth, there were still cosmic rays traversing the detector layers. With the CMS tracking solenoid operating at a magnetic field strength of 3.8 Tesla, the run was named Cosmic Run At Four Tesla 2008 (CRAFT08). During this CRAFT08, the pixel detector was completely inserted and acquired data for the first time in conjunction with the whole CMS detector.

3.4.1 The CRAFT run

Among a total of 270 million cosmic triggers, only a few percent of the collected events had tracks crossing the tracker volume, of which around 85 thousand were reconstructed crossing the pixel detector. Those triggers correspond to the 19 days of data taking with the magnet on, while the situation also allowed to have a few runs without any magnetic field in CMS, providing a chance for a rough but straight-forward alignment of the pieces of the detector - although the modules move a few millimeters under the influence of the high magnetic field - as tracks are then just straight lines, and giving a few analysis the opportunity to do some cross-checks.

3.4.2 Several Results from cosmic run

The CRAFT run was the opportunity to test the detector as well as some of the software that would be later used for LHC collisions. Therefore, on top of the usual calibrations, several pixel analysis were performed using the cosmic tracks instead of collisions, and managed to give quite a good overview of the stance of the pixel detector - and more globally of many other subdetectors like the strip tracker and the muon chambers detailed in other publications - along with a reassuring feel on its excellent performance

and readiness for collisions. Beside the basic description of the inactive portions of the pixel and the listing of the noisy pixels, the first hit and charge distribution of a particle - here muons - crossing the assembled detector was looked at, along with measurements of the Lorentz angle due to the drift of the charges inside the 3.8T magnetic field. The hit residuals and their position resolution were assessed. Making use of the more than 85 thousand tracks, the track residuals were also measured. Finally, it was the occasion for **Luca Mucibello** [15] and myself to measure the efficiency of the pixel modules, analysis depicted at length in the last section and hence containing relevant details also pertaining to the other analysis summarized here. All those results obtained during the first CRAFT data taking are all fully described in a publication [25] of the pixel community, that went out along with many others from all the subdetectors[2].

Inactive Sections

During the CRAFT data taking of 2008, about 99% of the barrel and 96% of the end-caps were operational. The biggest part of the dead modules, which were in the forward part of the pixel, was due to one defective high voltage and a low voltage line, impacting several modules, and which could not be repaired with CMS being closed. During the following winter 2008 shutdown of the LHC, CMS used the down time to re-open the whole detector, and the pixel's end-caps were carefully extracted. The faulty power supplies managed to get repaired, and the pixel end-caps were quickly reinserted. As mentioned before, the ease of the extraction of the pixel system was implemented in its design, and proved as such to be very effective. After the recommissioning of the pixel started, 99% of the total detector showed to be operational, although unfortunately some ROCs started to show problems, which led for the following CRAFT run of 2009 to an operational efficiency of 99.9% of the barrel modules, and about 97.5% for the end-caps. A table detailing the different problems and the number of ROCs affected for both bpix and fpix can be found in [25].

Noisy pixels in the ReadOut

The noisy pixel channels are identified using the DQM package. The analysis can either be done on the fly using the online DQM tools, or offline using the complete set of data or even a specific reconstructed dataset. To identify noisy pixels, their event rate is calculated, dividing the number of events in which they produce a registered signal above the threshold by the total number of events considered. Whenever this digi event rate is higher than one in a thousand, the corresponding pixel is considered noisy, as it is highly unlikely so many tracks would cross the detector at this precise position, and more probable the pixel is having some troubles and registering a charge even when nothing happened inside the detector. Using this cutoff value of 10^{-3}, there were 253 noisy pixels identified in the barrel, and a merely 17 in the end-caps. Those pixels were masked during the data taking using the *mask* bit from the DAQ settings, as they would pointlessly increase the data sent from the ROCs to the FEDs and might ultimately clog the readout stream of the pixel detector, as well as potentially confuse the tracking algorithms by creating hits unrelated to passing tracks. Even when applying a looser cutoff value of one in ten thousand for the event rate of noisy pixels, only 13 more pixels were tagged as noisy - 8 in the barrel, 5 in the end-caps -, which compared to the total amount of pixels in

[2]Presented at the eleventh ICATPP conference in como, Italy.

CMS means the number of noisy channels was negligible during the 2008 CRAFT data taking: less than $3.8 \cdot 10^{-4}\%$.

Along the years of operation, the masked noisy pixels were further modified to adapt to the changes in the detector: as many parameters can affect the pixels, such as changes in their calibration parameters, or hardware changes like power supplies or even full module repairs, the list of noisy pixels has fluctuated. They were also identifiable in other calibrations, like the gain calibration described previously in A with only saturating values for the gain curve, and thus even if not masked, they would likely not intervene in the reconstruction process.

3.4.2.1 Hit distributions and charge collection

From the 85000 tracks recorded crossing through the pixel detector under the 3.8 Tesla magnetic field, around 270000 clusters were formed in both bpix and fpix and examined. Clusters are defined as groups of nearby pixel signals, and each one of them is the actual result of a single particle passing through the sensor at this position. As described in Section 4, clusters are the basis used in the tracking algorithm to predict the particle trajectories, and although algorithms specific to cosmic rays were developed - the collision ones were designed for particles coming from the beampipe and radiating from the center of CMS to the detector outskirts, while cosmics come from the outside of the CMS detector, thus the need to seriously adapt the tracking to correctly map their trajectory to the data collected -, it was important at the pixel level to get a look at them. Figure 3.19 shows for the three layers of the barrel subdetector the 2D hit occupancy per ROC: each data point represents the total number of on-track clusters collected for the corresponding readout chip. Sections of the bpix - be it entire modules or single ROCs - that were not participating in the data taking, due to either DAQ set-backs during data acquisition or hardware problems described previously (see Section 3.4.2), are depicted by the white bins. On average, each ROC acquired around 60 hits during the CRAFT08 data taking. Although it is not a lot, due to the small rates of cosmics and the even smaller probability of one crossing the modest pixel volume, and especially compared to the expected values from collisions and its design operation capabilities, it was enough to derive the first standard distributions to gauge the readiness of the pixel detector.

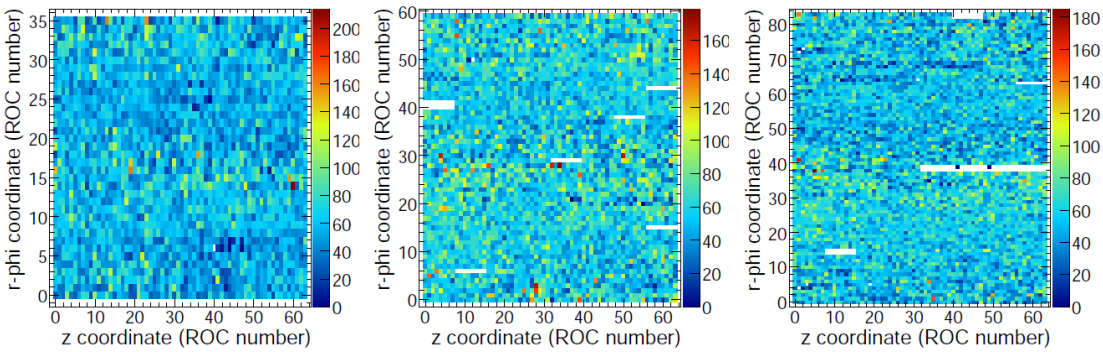

Figure 3.19: 2D occupancy map of the first (left), second (center) and third (right) barrel layers, with each point representing the number of hits associated to a track within a ROC. White bins represent the ROCs that where taken out of data taking. The plot origin corresponds to $\phi = 0$ and $z = -26.7\,\text{cm}$.

After the translation of the cluster charge from ADC back to electrons using the gain calibration data as reported previously, only clusters with a charge higher than 5000 electrons were used. Their charge was further corrected to accommodate for the variating path length of the particle in the sensor material, which strongly depends on the incidence angle of the particle trajectory with respect of the module surface. To keep the incident angle somewhat similar to collisions coming from the IP, only tracks with angles in the transverse plane smaller than 12 ° from normal incidence were kept. Some further selection to ensure the good quality of the clusters were performed: only clusters composed of more than one pixel were selected, and clusters with edge pixels were rejected as those have subpar behaviors compared to others, as is pretty clear from the 2D gain module map taken during calibration (see Figures 3.15 were the ROC edges are clearly distinct). Finally, clusters are omitted if another one is found in the same module/plaquette. The result of this selection gives the charge distribution drawn in Figure 3.20, where the data is plotted with points over the simulation in solid line. The data and the simulation agree pretty well, the lingering differences being explained by features specific to cosmic data and not collision data - thus not needed and not implemented in the simulation -, such as the time-walk effect and random arrival time with respect to the LHC clock, both reducing the charge collected inside a single bunch crossing, or the high number of broken clusters.

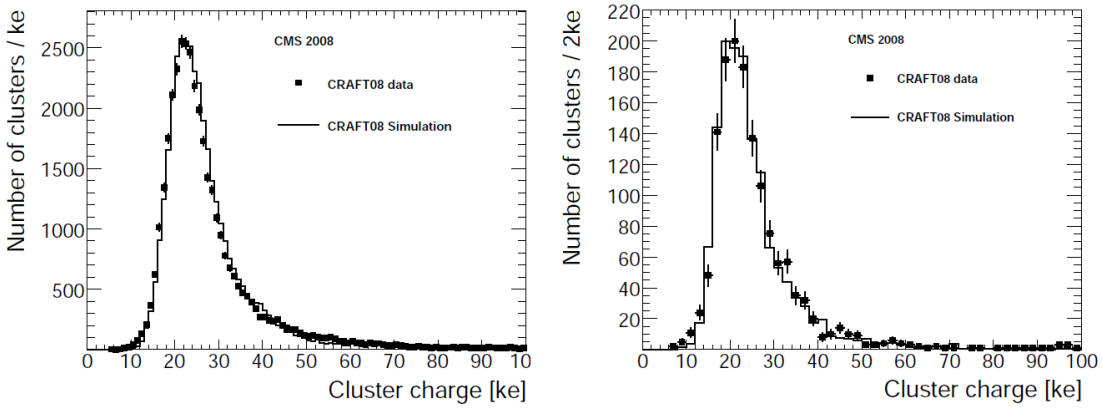

Figure 3.20: Distribution of the cluster charge for the barrel (left) and end-caps (right). The dots are the data points and the line represents the simulation. The clusters must have more than one pixel, no edge pixel, and their charge is corrected to the particle path length in the sensor, dependent on its incidence angle.

3.4.2.2 Lorentz angle measurement

As mentioned before, there are two main forces that influence the charges left by passing particles inside the sensor: the electric field, applied to make the charges drift to the n^+ side of the sensor to collect the signal, and the magnetic field, which is mainly there to curve the trajectories of particles to allow a measurement of their momentum, but has the side effect of introducing another force on the moving charges inside the sensors. This Lorentz force deflects the charges on their way to the cathode with an angle called Lorentz angle, θ_{LA}, which of course depends on the strength of the magnetic field bathing the sensors, but also on the bias voltage applied, as well as the temperature and irradiation endured by the detector. The magnetic field produced inside the CMS solenoid is uniform, which results in the barrel subdetector to the magnetic field vector being anti-parallel to the local y of every modules, leading to charges drifting strongly in the local x direction,

meaning charges will spread over several pixels due to this Lorentz effect. As explained in the description of the pixel end-caps, the panels were deliberately tilted at an angle of 20 °, to allow there too for charges to drift under the influence of the magnetic field. Indeed, the sharing of the charge among close-by pixels was intended in the design of the pixel detector, as it enhances the spatial resolution of clusters: the interpolation of the different pixel's charges constituting the cluster results in its more precise position measurement. The charge is then also corrected using this Lorentz angle value in a second step, as knowledge of the track's incidence angle is required to apply this correction, therefore requiring a first cluster measurement to compute the tracks. During CRAFT08, the Lorentz angle was obtained by means of the minimum cluster size method. Later on, another method relying on tracks with grazing angles [26] was performed alongside the first one, and results were found to agree perfectly [27].

The minimum cluster size method was developed on the principle that the charges would drift not influenced by the magnetic field when produced by particles arriving with an incidence angle that corresponds to the Lorentz angle. Therefore, when looking at a variable representing the drift, like the cluster transverse size - the number of pixels the charge was finally spread on - in function of the incidence angle in the plane of the drift, there would be a minimum cluster size for the Lorentz angle. To allow the distribution to be easily fitted, the cluster transverse size has been plotted in function of the cotangent of the α angle, and the Lorentz angle is obtained deriving the minimum of the fit, as can be seen in Figure 3.21 for the barrel (left) and the end-caps (right). The minimum obtained for the barrel and end-caps are respectively -0.462 ± 0.003 and -0.074 ± 0.005, which correspond to a Lorentz angle of $(24.8 \pm 0.2)\,^\circ$ in the barrel detector and $(4.2 \pm 0.3)\,^\circ$ in the forward detector. The drift length in the transverse direction due to the Lorentz force is therefore $65.9 \pm 0.5\,\mu$m for the barrel modules with sensor thicknesses of $285\,\mu$m, and $9.9 \pm 0.7\,\mu$m in the end-caps where the sensors are $270\,\mu$m thick. The Lorentz angle measurement was one of the analysis that also benefited on the short period of data taking with the CMS solenoid being offline, leaving the charges without any influence of the magnetic field. Using the same cluster size method, the minimum were derived at 0 T and correspond to the second set of data visible on Figure 3.21: as one should expect, the results are compatible within uncertainties to a Lorentz angle of 0 ° for both subdetectors.

The Lorentz angle values were also simulated using `PIXELAV` [28–31], a simulation package developed only for the pixel detector to eventually replace expensive test beams, and which contains more details in the characteristics and specifics of the pixel module as well as the physics of charge deposition and drift inside a silicon sensor, when compared to the global CMS simulation which is less detailed but much faster. This simulation allows to accurately simulate effects such as the Lorentz force applied to charges within the pixel module, and gives Lorentz angle summarized in Table 3.1 which are compatible within uncertainties to the evaluation from data.

3.4.2.3 Hit residuals

The hit residuals are measured in both local x and y by comparing the predicted position of where the track crosses the sensor layer, with the actual position of the measured charge deposition. To obtain the track position's prediction, the trajectory is refitted without the hit studied. The X and Y residuals in function of the local track angles α and β and its corresponding cluster charge are presented in Figure 3.22, where each bin is the result of a Gaussian fit, the width of the fit being the residual between the track projection and the hit. Along with the other analysis, only tracks with a momentum higher

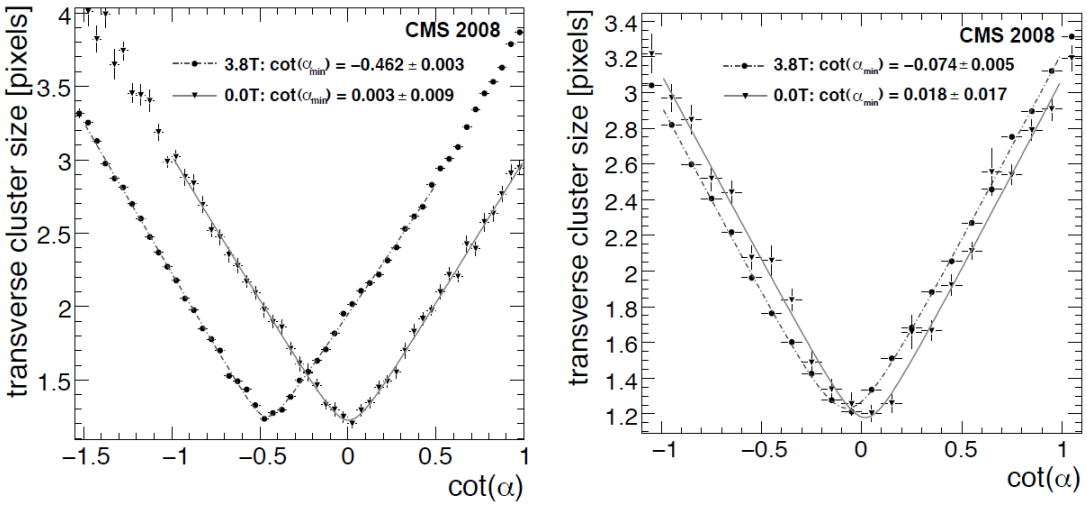

Figure 3.21: Distribution of the size of the clusters in the transverse plane - the module's local x direction - in function of the cotangent of the local α angle - defined as the angle between the track and the surface of the module in the xz plane- for the barrel (left) and the end-caps (right), with two sets of data: with and without the 3.8 T magnetic field. The data points are fitted in order to derive the minimum of the curve defining the Lorentz Angle.

Table 3.1: Values of the minimum of the cluster transverse size curve and the corresponding Lorentz angle, for both bpix and fpix. Data with the 3.8 T magnetic field and without it are shown. The simulation from `PIXELAV` is also given and agrees well with the data.

		barrel		end-caps	
		Minimum	Lorentz angle	Minimum	Lorentz angle
3.8 T	data	-0.462 ± 0.003	$(24.8 \pm 0.2)\,^\circ$	-0.074 ± 0.005	$(4.2 \pm 0.3)\,^\circ$
	simu	-0.452 ± 0.002	$(24.3 \pm 0.1)\,^\circ$	-0.074 ± 0.004	$(4.2 \pm 0.2)\,^\circ$
0 T	data	-0.003 ± 0.009	$\sim 0\,^\circ$	-0.018 ± 0.017	$\sim 0\,^\circ$

than 10 GeV are kept, while hits with edge pixels or composed of only one pixel with a registered charge smaller than 10000 e^- were left out. From the results, one can conclude that the tracks with the smaller residuals - thus best precision - are the ones arriving with a normal incident angle with respect to the modules' surface, and when depositing a charge of 30 ke^-. The measurements presented here are the total residuals: they include the track extrapolation error, the detectors - both pixel and strip tracker - intrinsic resolutions, contributions from multiple scattering, as well as the eventual uncertainties due to residual detector misalignments.

3.4.2.4 Hit resolution

During CRAFT08, the presence of overlapping modules in bpix was used to perform a measurement of the hit resolution in the pixel detector [32]. Indeed, although few in number, some modules in the barrel are installed such that they overlap within the same

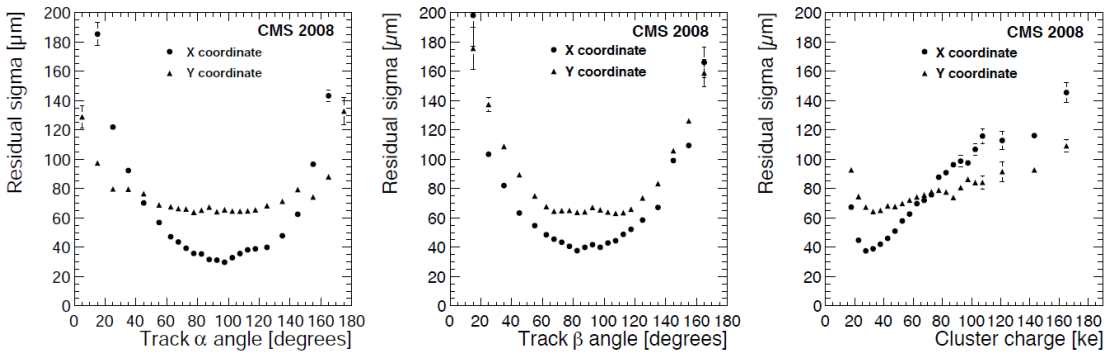

Figure 3.22: Hit residuals along the x and y coordinates obtained through Gaussian fits of each bins in α (left), β (center) and charge (right).

layer, providing two sensor crossings at very close proximity. One can then calculate the double difference between the predicted positions on the overlapping modules and the actual measured hit position - so for the x coordinate, $\Delta X_{pred} - \Delta X_{hit}$ - , and fit by a gaussian the distribution obtained accumulating tracks on each overlap site, the width of the fit giving the resolution on the measurement which is then independent of the translational misalignment of the modules. As there are two crossed layers at very close range, the predicted position of the track crossing the modules uses a combination of backward and forward prediction to increase the accuracy: the layer under study is excluded from the track information in order not to bias the result, and the position of the track crossing the module's surface is inferred using both forward prediction and backward extrapolation of the trajectory starting from the overlapping module not under scrutiny. Albeit using a combination of the predicted state, the predicted position difference is still by far less precisely known compared to the hit position difference.

Such measurement is shown in Figure 3.23, where the width of the gaussian for the double difference in both local x and y directions are plotted for each overlap occurrence. In order to ensure a correct result, several restrictions on studied clusters and tracks were applied: to lessen the impact of multiple scattering, only tracks with a momentum of at least 5 GeV were selected, while requiring their incidence angle to be greater than 30 °; to minimize the presence of charges collected within consecutive bunch crossing due to the time-walk effect, and resulting in broken clusters biasing the hit position, clusters with charges under 10000 e^- were cut out, as well as those containing edge pixels. To obtain the hit resolution, the track prediction uncertainty was quadratically subtracted to the width of the gaussians fit. After taking into account only overlapping modules that collected at least 30 tracks, the hit resolution was found to be $18.6 \pm 1.9\,\mu$m along x and $30.8 \pm 3.2\,\mu$m along y. A simulation using `PIXELAV` was performed for similar tracks, concluding in a very good agreement with the results extracted from the cosmic data.

3.4.2.5 Track impact parameter resolution

The impact parameter of the tracks, defined as the difference between the closest point of the trajectory to the - here virtual - beamline and the beamline, is one of the first measurements performed to qualify the tracking. Obviously, it has much less significance for cosmics as they are not coming from collisions of particles rushing inside the beampipe along the expected beamline, and therefore can have very large transverse (d_{xy}) or longitudinal (d_z) impact parameters that have no physical meaning or experimental value,

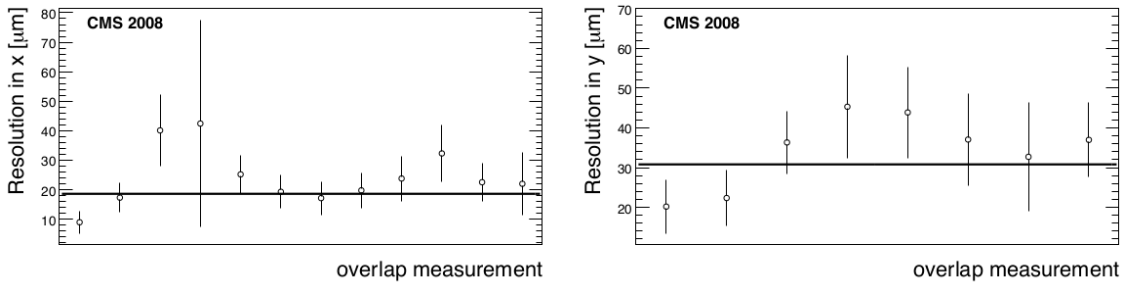

Figure 3.23: Hit resolution along the local coordinates x (left) and y (right) for the overlapping modules with enough data. The average over all overlapping sites gives $18.6 \pm 1.9\,\mu$m along x and $30.8 \pm 3.2\,\mu$m along y.

contrary to collisions which come from a single point in space - called the beamspot, see Section 4.3.5 - thus allowing a characterization of both the collision point features and the tracking algorithm performances. For the CRAFT data, the leg splitting technique [33] was used, which permits a measurement of the residuals of the track impact parameters, giving an input at least on the performances of the tracking. This method consist of splitting tracks in two at their point of closest approach to the beamline, producing two independently measured trajectories of the same particle, and then calculate the difference in d_{xy} and d_z for the two predicted track paths. To ensure a still optimal quality on the legs' trajectories, several traits were required from the tracks: they should have at least 10 hits in the tracker (pixel + SST), 3 hits in the barrel of the pixel detector, have a χ^2 smaller than 100, and their momentum must exceed 4 GeV. Figure 3.24 encompasses the residuals of such approach for the transverse (left) and longitudinal (right) impact parameters. Both distributions are fitted by a Gaussian, with a width of $23.3 \pm 0.3\,\mu$m for d_{xy} and $39.4 \pm 0.5\,\mu$m for d_z.

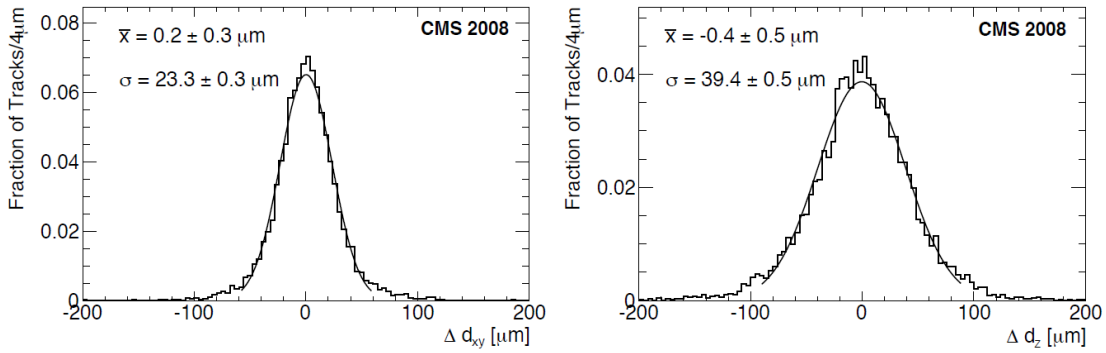

Figure 3.24: Transversal (left) and longitudinal (right) track impact parameter residuals obtained through the splitting method [33]. Distributions are fitted with a Gaussian (solid line).

3.4.3 Efficiency measurement

The cosmic data acquired during CRAFT08 was the occasion to measure the efficiency of the pixel detector, which is detailed in what follows.The aim of this note is to analyze the efficiency with which the pixel sensors have been able to properly collect track hits

during CRAFT08. It embraces a description of the datasets used for the analysis and software aspects, the definitions, cuts and methods used in the analysis, as well as the results extracted from the measurements.

3.4.3.1 Preliminaries

This first section here will describe the peculiarity of cosmics when detected by an apparatus designed for collisions, and their relevance and impact on the measurement of the pixel efficiency, followed by the basic specifics of the data and events used in the analysis.

Cosmic Ray triggering and timing

The pixel acquisition of cosmics presents some features of intrinsic inefficiency not related to sensor behavior. As described previously, the pixel detector operates at the LHC frequency, collecting charges in a narrow window of 25 ns at a fixed rate of 40 MHz. Each of those windows is correlated to one and only one expected bunch crossing. Likewise, the read-out sequence of the pixel detector works such that each time an external trigger is received, the data is retrieved from the detector with the precise information about the bunch crossing it refers to, encoded into the data stream. But the peculiarity of cosmics is that they arrive randomly in time, and consequently it is possible that a signal which is seen in a certain bunch crossing window from the muon subdetectors, is stored by the pixels as belonging to the subsequent bunch crossing.

Another relevant effect is related to the time-walk, defined as the time it takes for a signal of a given amplitude to cross the comparator threshold: it strongly depends on the signal amplitude deposited in the sensor and drifting to the cathode, with large signals crossing the threshold faster than the low ones. This implies that low amplitude signals could come too late and be shifted to the next bunch crossing. In such a case, it would be lost from detection and contribute to pixel inefficiencies.

The pixels at the borders of the readout chips do not have the same dimension as the other ones, in order to accomodate the edges between readout chips, and as such can be source of strange behaviors. This is the reason they are usually excluded from the analysis, especially for the first ones published, during CRAFT08, as it was our first time playing with a completed pixel detector.

One of the important observables for cosmic tracks is their muon arrival time, which is measured by the muon drift tubes detectors and defined as the time difference between the arrival of the muon in the CMS muon subsystems and the 40 MHz LHC clock. In essence, it characterizes how much off from the designed collision time the cosmic track under study is. Figure 3.25 (left) shows the efficiency of finding tracks passing through the pixel detector in function of the muon arrival time. The efficiency is defined applying an ad hoc selection: the denominator consists of tracks with $p_{\mathrm{T}} > 10$ GeV, a closest approach to the nominal interaction point in the xy-plane $d_0 < 9\,\mathrm{cm}$ ($d_0 = d_{xy}$ and are just two ways to write the same variable), a closest approach to the nominal interaction point in z coordinate $d_z < 30\,\mathrm{cm}$, a maximum uncertainty on the muon arrival time of 5 ns and a χ^2 probability larger than 0.5, while the numerator is the subset of these tracks showing at least one pixel hit. The solid line represents the simulation, while the data are represented divided in two datasets, as a change in the pixel timing with respect to the LHC clock was operated. It shows that the track efficiency is quite high when the tracks arrive at the expected LHC clock time, giving the tracks time to arrive to the pixel detector and the charges the maximum 25 ns window they need to reach the comparator threshold, while

it can drop consequently when arriving out-of-time. This clearly indicates the need for a selection on this muon arrival time to ensure an unbiased efficiency measurement, which is described later on.

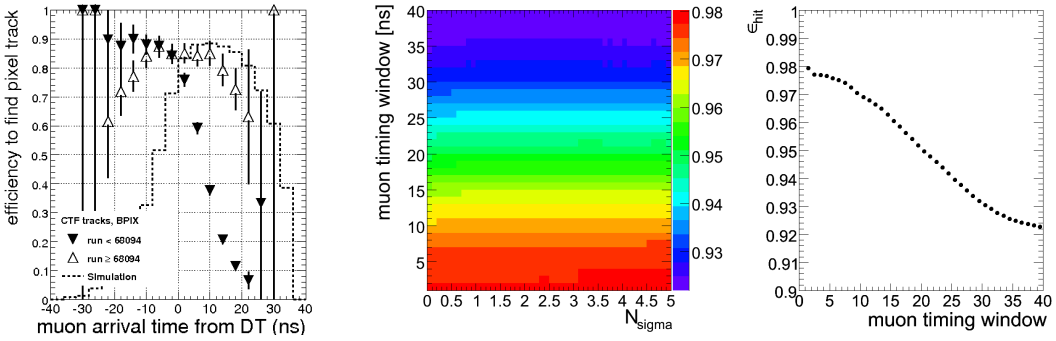

Figure 3.25: (left) Efficiency of finding tracks passing through the pixel subdetector. The points refer to CRAFT data, the change in the timing set-up of the pixel subdetector during data taking period is reported with two different set of points. The dashed line refers to Monte-Carlo simulation.
(center) Efficiency ϵ_{hit} for different values of the edge cut (number of standard variations of the propagated state on the layer for the x axis) and of the muon time cut (track muon arrival time window for the y axis). All the other cuts are already applied.
(right) Slice of central plot showing efficiency dependence on the muon arrival time window for a fixed edge cut choice of 1 standard variation.

All the considered features - referred to subsequently as the muon timing problem - are inherent to cosmic data, and clearly affect the results of the efficiency measurement, leading to constraints applied specifically to account properly for those features.

Event reconstruction, data sets and software

The data used in the analysis was produced in the second global reprocessing of the CRAFT08 data, where alignment and calibration constants were updated using the knowledge gained from the first processing of the events. The second reprocessing, as well as the present analysis, have been centrally produced using the `CMSSW` software (version `CMSSW_2_2_5`).

The CRAFT08 data have been reconstructed assuming a constant magnetic field of 3.8 T. In practice cosmic data have been also collected during the ramping on and off of the magnet, as well as during periods without magnetic field, but this analysis used only the events where the magnetic field was on and stable. Moreover, only runs where all pixel Front-End Drivers were read out were accepted, which was ensured by the `fedInRunFilter` module. Events with more than one track were discarded since it was observed that in a large fraction of the CRAFT08 two-track events, the tracks were sharing some hits which would most likely lead to an increase of the fake track rate.

A list of well known bad modules was also used for a further cleanup, taking into account modules with faulty power supplies and/or broken wire-bonds. Dead and half dead modules were also completely excluded from the measurement.

3.4.3.2 Hit reconstruction efficiencies

Definitions

The first step of the pixel efficiency measurement was the aggregation of pixel charges into a cluster. Here, all pixels with a charge above the hardware readout threshold were considered during the clusterization. Clusters were built around a seed pixel with a charge above 3000 electrons, adding adjacent pixels which overcame a charge threshold of 2500 e^-. Finally, the cluster was accepted only if its total charge was more than 5500 e^-.

The pattern recognition, or trajectory building of the cosmic particles, was based on a combinatorial Kalman filter method [34], which differs from the standard CMS tracking developed for collisions (see Chapter 4). The filter proceeds iteratively from the seed layer, starting from a coarse estimate of the track parameters provided by the seed and including the information of the successive detection layers one by one. With each iteration the track parameters are updated, decreasing the uncertainty on the successive predicted states. Since several hits on the new layer may be compatible with the predicted trajectory, several new trajectory candidates are created, one per hit. One further candidate is created in which no measured hit is used to account for the possibility that the track did not leave any hit on that particular layer, accounting for detector inefficiencies, mandatory for this measurement. The window for the search of a hit on a layer is defined to be approximately five times larger than the uncertainty of the predicted state on the layer under observation. The better the precision on the predicted state, the smaller the search window becomes.

With this scheme in mind, two kinds of hits are defined:

- **Valid hits:** when during the propagation on the successive layer, a hit is found inside the search window.
- **Missing hits:** when during the propagation on the successive layer, no hit is found inside the search window.

This results in a hit finding efficiency ϵ_{hit} given by:

$$\epsilon_{hit} = \frac{n_{valid}}{n_{valid}+n_{missing}}$$

where n_{valid} is the total number of pixel valid hits in the analyzed sample and $n_{missing}$ the total number of pixel missing hits. This definition is hereafter referred to as Method I, as a second one follows.

The definition of this efficiency makes it dependent on the track reconstruction algorithm, as although the 5σ window to look for a hit around the predicted state isn't arbitrary and gives the best performances between finding tracks and rejecting fakes, changing this value would modify the results of the efficiency. To palliate this effect, we defined another independent measure of ϵ_{hit} by propagating ourselves each track towards a successive pixel layer starting from the strip layers. Around the intersection between the propagated track and the layer, a search was performed for a pixel cluster inside a window ΔR of radius R. The hit finding efficiency $\epsilon_{hit}(\Delta R)$ is hence now dependent on the size of the search window:

$$\epsilon_{hit}(\Delta R) = \frac{n_{cluster}(\Delta R)}{n_{intersection}},$$

where $n_{intersection}$ is the total number of analyzed intersection between layers and propagated states and $n_{cluster}(\Delta R)$ is the number of occurrences when at least one cluster is found inside the window ΔR around the studied intersection. This definition is less biased by the tracking algorithm efficiency, and is from here on referred to as Method II. This efficiency is expected to converge to zero for small window sizes, and to develop a stable plateau when the window size approximates the standard Kalman filter search window size.

Hit selection

Some cuts are imposed during the selection of analyzed tracks and hits in order to ensure that the observed inefficiencies are actual hardware inefficiencies and as less as possible coming from the particular configuration of the cosmic acquisition. A first set of selections is applied at the track level. First, in order to be consistent with other analysis, tracks with transverse momentum less than 10 GeV/c are rejected. Then a cut is performed on the muon arrival time of the track, as described previously: only tracks within a time window of ±5 ns around the maxima of the timing distributions of Figure 3.25 are kept. This cut is referred to as the *muon time cut* in what follows, to allow for an easy identification.

A second set of selections is applied at the hit level. In order to avoid border effects at the edges of pixel detector modules, a hit is skipped when it is near the border of the sensor for less than one standard deviation of the propagated state of the trajectory on the sensor. This is referred to as the *edge cut*. Because of the muon timing issue, it is also required that for each considered hit there is at least one other pixel valid hit on the track in the region above the beam line (upper region) and one other pixel valid hit in the region below the beam line (lower region). In case the requirement is not fulfilled the hit is skipped. This is referred to as the *telescope cut*. This cut is unbiased since the hit under study does not contribute to the selection of the hit itself. A sketch of this selection is represented within the pixel detector in Figure 3.26, with the first track containing one hit satisfying the telescope cut, while the other track would be dismissed.

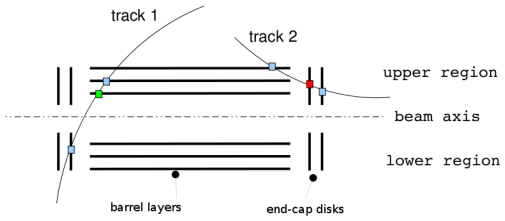

Figure 3.26: Schematic view of the telescope cut. In track number 1 the green hit under analysis satisfies the telescopic requirement due to the presence of at least one hit in the upper region and at least one hit in the lower region. In track number 2 the red hit under analysis does not satisfy the telescope since no hit is found on the same track in the lower region of the detector.

3.4.3.3 Results

This section contains the results of the analysis on the CRAFT08 data, with first distributions pertaining to the optimization of the selection cuts applied, followed by the main results obtained with the first definition along with the efficiency with respect to all four local variables x, y, α and β and their detailed study, and finishing with the produce of the second method.

Preliminary observations: selection optimization

Figure 3.25 (center) shows the variation of the efficiency ϵ_{hit} as function of the muon time cut window and the number of standard variations N_σ chosen for the edge cut, while all other cuts have already been applied. For each given value of the timing window, the efficiency demonstrates its independence on the edge cut within 0.2%. A reference

value of one standard variation is therefore chosen for the edge cut. On the other hand, a strong variation in the efficiency can be observed for different values of the timing window at fixed edge cut. As for the right plot of Figure 3.25, it displays in more detail the behavior of the efficiency with respect to the muon time cut for an edge cut of one sigma. A window of 10 ns has been chosen which optimizes the statistics and stability of the efficiency calculation, by having as many hits as possible while staying in the plateau of the efficiency with respect to the muon timing, ensuring the measure would depend as less as possible on the cut applied.

The selections employed decrease strongly the sample size: Table 3.2 references the amount of hits surviving the cut sequence separately for the barrel and the end-caps. It is clear that the amount of surviving hits in the pixel end-caps is not enough to make quantitative statements, hence the analysis provides here only results for the barrel of the pixel detector.

Table 3.2: Number of observed hits after applying the sequence of cuts shown in the first column.

cuts	barrel	end-caps
no cuts	356719	59489
+ $p_T > 10$ GeV	228154	39073
+ edge cut	215263	31158
+ muon time cut	70299	8341
+ telescope cut	48673	21

The selection cuts affect also some of the cluster properties. One of the most important of those properties is the cluster charge, measured in number of electrons, and defined as the total amount of charge stored in the pixels composing the cluster. Another important relevant variable is the cluster dimension, which is the number of pixels that are merged together to construct the cluster. This dimension is by design sensitive to the Lorentz drift explained previously, as well as to the track incident angle with respect to the surface of the module. Figure 3.27 (left) shows a two-dimensional histogram of those cluster observables for data without any cut applied, with the number of clusters for each bin of cluster charge and cluster dimension. The charge distribution presents a large peak around 25 ke^- and a smaller one around 6 ke^-, with the smaller low charge peak corresponding to single pixel clusters. Studies were done specifically to understand better the origin of these single pixel clusters, which were found to come from broken clusters and muon timing issues. Applying the cut on transverse momentum and the edge cut do not change substantially the charge distribution, as can be seen from Figure 3.27 (center), contrary to the muon time cut which removes the smaller charge peak and decreases the contribution of single pixel clusters (Figure 3.27 right), while the telescope cut decreases the high charge and large cluster size tails, making the final charge distribution - after all selection cuts are applied - narrower around the mean of 25 ke^-.

Method I

The values of the efficiency for each layer using the first definition are $(96.94 \pm 0.10)\%$, $(97.54 \pm 0.08)\%$ and $(96.18 \pm 0.10)\%$ respectively for the first, second and third layer of the barrel. The measure on the whole pixel barrel is $(96.91 \pm 0.05)\%$. Excluding all the well-known misconfigured modules, the values of the efficiency per layer increases to

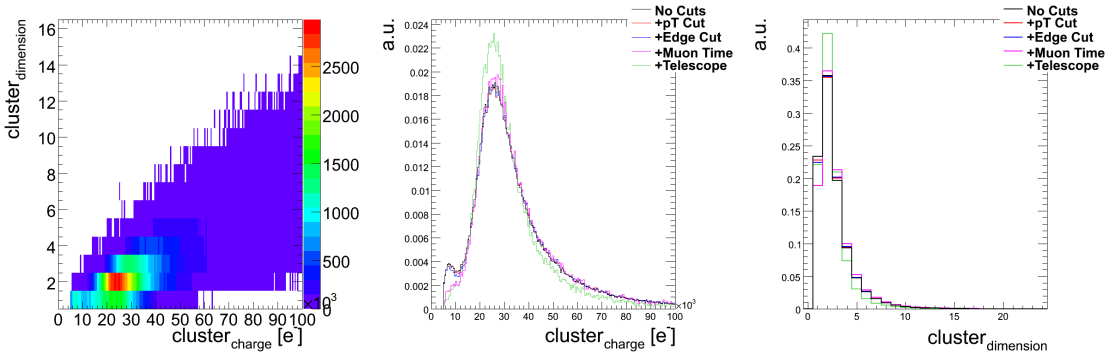

Figure 3.27: (Left) Pixel cluster distribution of charge versus dimension without applying any selection on hits nor on tracks. Cluster charge (center) and dimension (right) after applying the specified cut sequence, with each distribution normalized to unity.

$(97.50 \pm 0.10)\%$, $(97.79 \pm 0.08)\%$ and $(97.33 \pm 0.09)\%$ respectively for the first, second and third layer and to $(97.55 \pm 0.05)\%$ for the whole pixel barrel. The errors quoted and presented here are calculated using standard binomial statistics. Figure 3.28 shows the efficiency by module for those with a relative error smaller than 10% - to make sure there was enough statistics to perform the measurement -, while Figure 3.29 shows the relative error for each module. The misconfigured modules are clearly identifiable as their efficiency is much lower than the average, and were made recognizable by a solid border. Dead modules or half-dead ones - where just one of the TBMs was disfunctionning - were also left out, as they would decrease the real efficiency measurement without any reason. The third layer of the barrel is the one most affected by the lack of statistic: the two horizontal bands with high relative errors are particularly visible and correspond to vertical modules, less likely to encounter a particle as cosmics tend to "fall" from above.

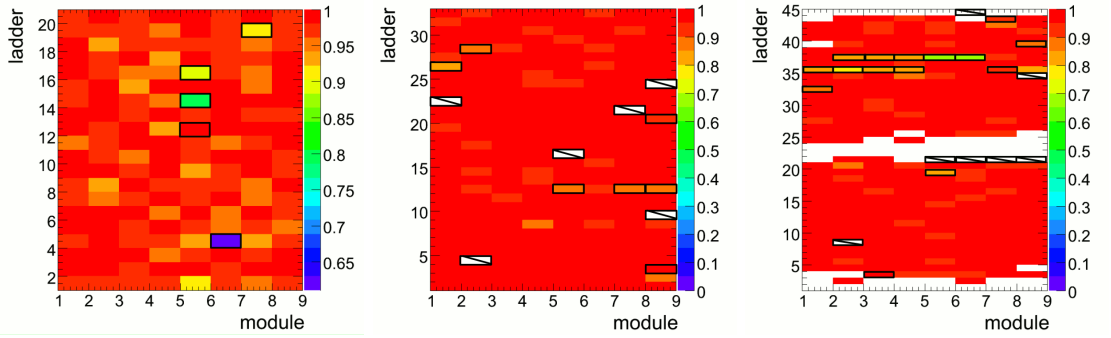

Figure 3.28: Hit detection efficiency measured in each sensor of the first (left), second (center) and third (right) barrel layer. The white cells with diagonal strip correspond to dead/half-dead modules, the colored cells with a border correspond to misconfigured modules. The white modules have relative errors higher than 10% on the calculated efficiency.

The efficiency was also computed as function of the local coordinates of the hits inside the sensor module. The local position x_{local} and y_{local} are the position in centimeter on the surface of a module, taking the center of the module as axis origin. As for the z-axis, it is defined as orthogonal to the module surface. In the barrel detector, the local xz-plane coincides with the $r\phi$-plane of the CMS global reference frame, while the local y-axis is

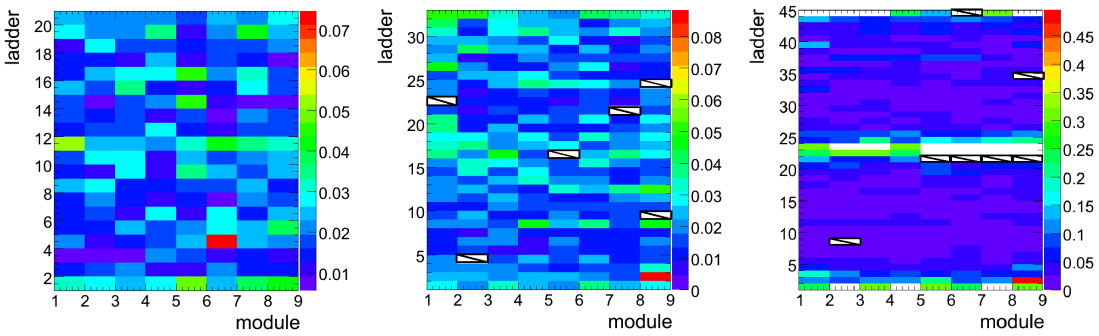

Figure 3.29: Relative errors for the hit detection efficiency measured in each sensor of the first (left), second (center) and third (right) barrel layer. As for Figure3.28, the dead/half-dead modules correspond to white cells with diagonal strip.

parallel to the beam and the magnetic field direction. The homogeneous behavior of the efficiency in function of y_{local} (Figure 3.30 top right) is not confirmed when plotted in function of x_{local} (Figure 3.30 top left), where the efficiency reaches its minimum in the center of the module. This behavior, still present when splitting the observation between big and small modules - the big modules having a x-length of 0.8 mm and being the standard ones, while half-modules having only one row of ROCs in the barrel mean their length along the x coordinate is only 0.4 mm - was not understood.

The local angle α_{local} (β_{local}) is defined in the module as the angle of the track projection on the local xz (yz) plane. The β_{local} efficiency distribution in Figure 3.30 (bottom right) is compatible with a flat distribution, with a small decrease of efficiency for very grazing tracks. On the other hand, the α_{local} distribution presented in Figure 3.30 (bottom left) indicates clearly a minimum for a specific value of α, which happens to be the value of the Lorentz angle, $\alpha_L = (\pi - 24.8)^\circ = 1.138$ rad - see Table 3.1 for the results obtained during CRAFT08 for the pixel barrel. The behavior of the α_{local} efficiency distribution was not expected, and it took some time to find its correlation with the Lorentz angle.

In order to understand the two features present in x_{local} and α_{local}, all the different possible correlations between local variables have been studied. The only one showing some particular features is the efficiency distribution with respect to α_{local} versus x_{local}, seen in Figure 3.31 (left). Even in the small amount of missing hits, one can recognize a region of inefficiency not uniformly distributed in the α_{local}/x_{local} plane. To better understand the shape of this region, the same plot has been produced on data without applying any selection. On the Figure 3.31 (right), one can identify a central region of higher inefficiency approximated by an ellipse: inside this region the inefficiency is mostly focused around values of the Lorentz angle. The border effects observable in this figure - for x around ± 0.8 cm- are not discussed here, as they deal mostly with track reconstruction algorithm features.

These results indicate inefficiency contributions connected to the Lorentz drift mechanism in the particular acquisition configuration used during CRAFT08. The idea is reinforced by the fact that the β_{local}, defined with respect to the y-axis parallel to the magnetic field, does not show such strange behaviors. Lots of different parameters, such as the pixel thresholds, are involved in such consideration. The effects of the Lorentz drift could be further studied using data from the end-caps, where the Lorentz effect is very small, provided that enough data would be collected in this area, but as this was not pertinent for collision data, no additional effort was put to use for this.

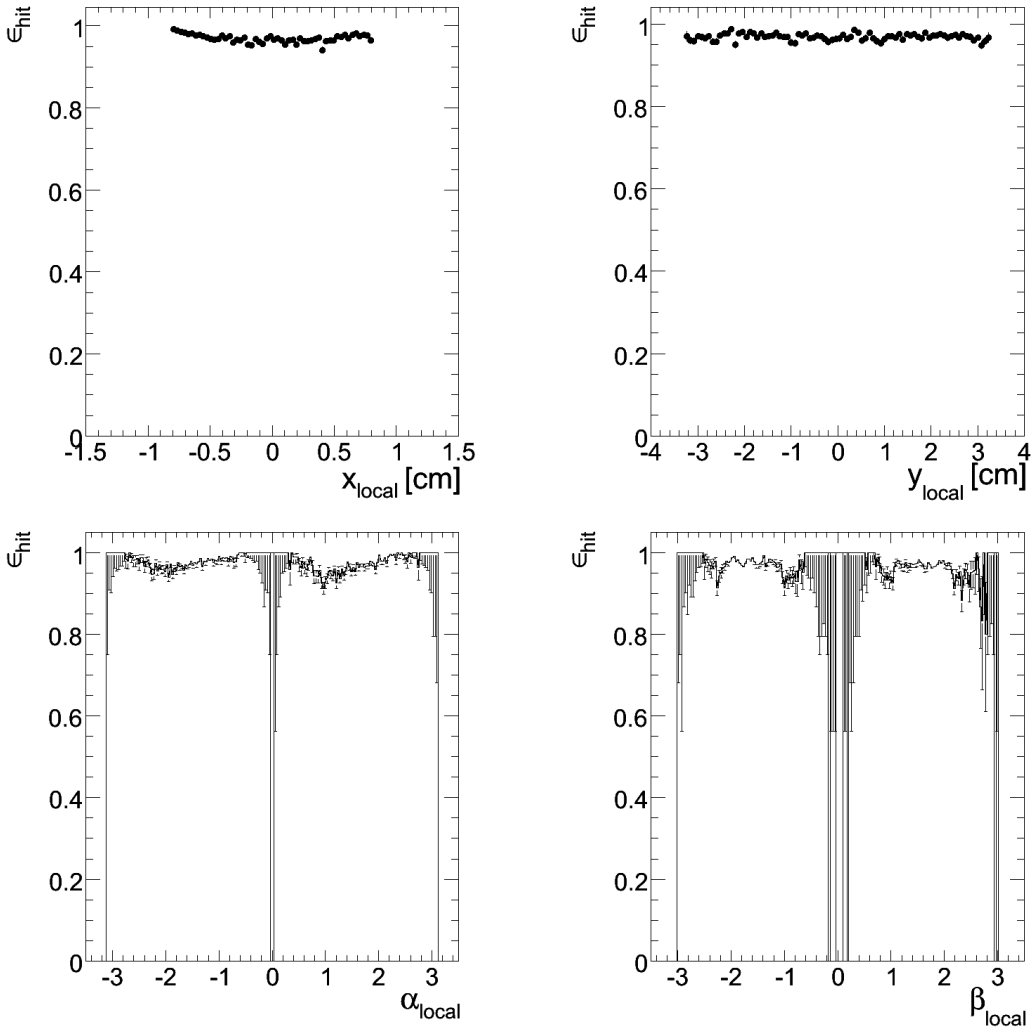

Figure 3.30: Efficiency in function of local variables (a) x_{local}, y_{local}, α_{local} and β_{local}.

Method II

Figure3.32 reports the results obtained using the window-depending efficiency definition: $\epsilon_{hit}(\Delta R)$ reaches a plateau around 300 μm, which is compatible with the observed typical residual of the pixel detector. The maximum size of the window shown in the plot is taken to be one order of magnitude larger than the typical residual to allow for the plateau to appear for certain, and to ease the measure of its value. The asymptotic values of the efficiency are 96.93% and 97.49%, respectively observing all the modules and excluding all the misconfigured ones. The efficiency values are completely compatible with the ones earlier obtained, showing no particular contribution of inefficiencies coming from the tracking algorithm, as one would hope.

3.4.3.4 Conclusions

The efficiency with which the pixel sensors have been able to properly collect track hits during CRAFT08 has been carefully studied. An average value of $(97.55 \pm 0.05)\%$ was obtained for the barrel region, using data from about 95% good working modules, while unfortunately no information was available for the end-cap region due to lack of statistics,

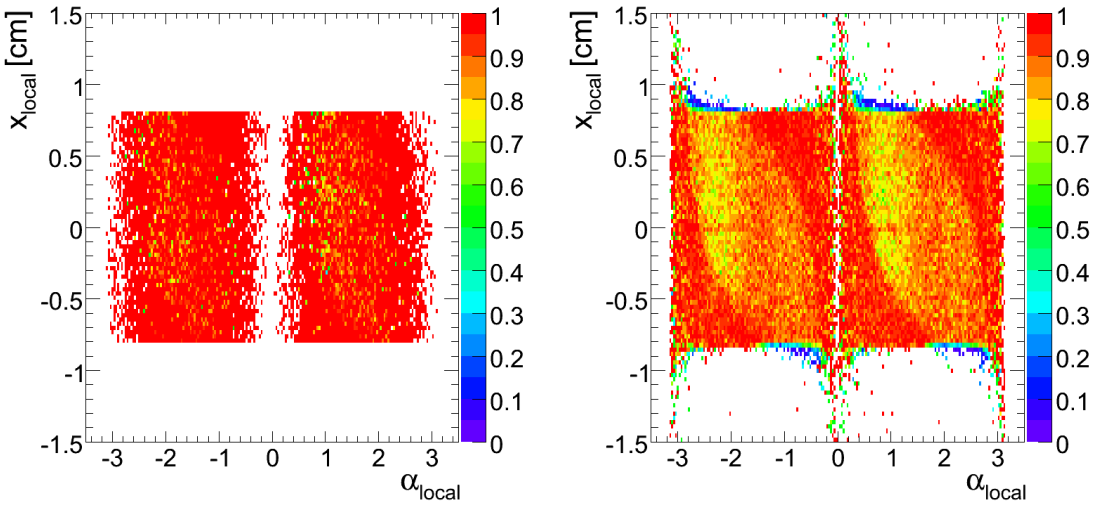

Figure 3.31: Efficiency distribution of α_{local} versus x_{local} for barrel modules with 2 rows of ROCs. The figure on the left is obtained applying all the selection cuts while the right one is obtained without applying any selection cut.

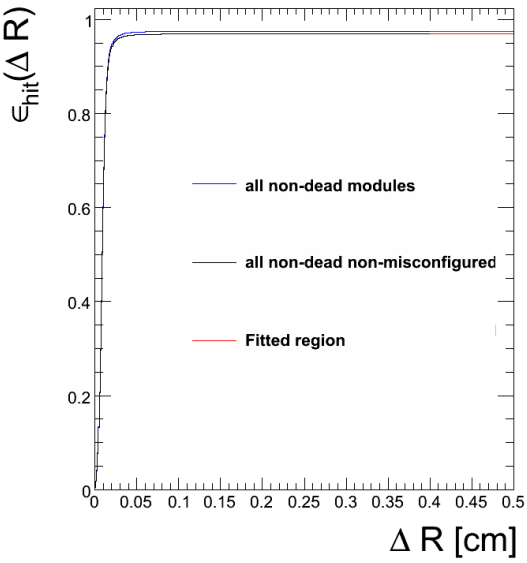

Figure 3.32: Efficiency $\epsilon_{hit}(\Delta R)$ in function of the window search as defined in method II. The line in black refers to the observation done on all the non-dead modules, the line in blue refers to the observation done excluding all the dead and misconfigured modules. The asymptotic values reported of respectively 96.93% and 97.49% of efficiency are obtained fitting the red part of the curve with an horizontal straight line.

after the mandatory selections were applied. The efficiency in function of several variables of interest were shown, and some features caused by the Lorentz drift of charge carriers were observed. The measurement and observations refer here to the particular acquisition of cosmic rays, and thus caution is advised when applying or comparing these numbers with LHC beam data. The measurement was also carried out on simulated data: the features discovered where not present in the results, as expected as those effects were

not introduced in the simulation code, the result of the efficiency measurement for the barrel was around 99.49%. This higher value compared to data is the consequence of the simulation not reproducing all the effects of the cosmic data acquisition.

I also performed the same measurement presented here when the first collisions happened in 2009. Looking at this efficiency had two positive aspects: it was first one of the quick analysis that could be performed on the data just hours after recording the collisions, and thus could provide along with DQM - where I later implemented it for this exact purpose - one of the basic input on the quality of the collected data and the performances of the pixel detector. Secondly, during the first week of collisions, a special task force was implemented, with people physically at CERN - with meetings I took part in in the CERN offline control room -, with the goal to provide nearly real-time information on the fast-incoming data, along with the first physic analysis CMS would publish. The basic and most immediate observable that was completed was the $\mathrm{d}n/\mathrm{d}\eta$ distribution, analysis prepared before hand but that still took many people working nearly around the clock to obtain, and was the first CMS publication [35]. One of the systematic errors to contribute to this $\mathrm{d}n/\mathrm{d}\eta$ measurement was the pixel efficiency, which therefore needed to be calculated: this was done using two separate measurement, the one described here which I provided and one with a simpler tracking - tracklets, see Section 4.2 for more details -, and was observed to be higher than in cosmic data, above 99%.

The pixel efficiency measurement was also useful later on to determinate the best delay for the pixel internal clock with respect to the LHC one. This calibration is explained previously in Section 3.2.3, and makes use of the pixel efficiency value as well as the cluster size to find the optimal clock delay.

Chapter 4

Tracking in CMS

As detailed in chapter 2, CMS has a very complex tracking detector, composed of the pixel subdetector and the silicon strip tracker subdetector. They are both used to measure very precisely the position and momentum of all charged particles passing through. The trajectories are then reconstructed from hits, to cluster, seeds, tracks and their vertices. Since tracks are the main objects used in my analysis (no jets are used, and particles such as electrons, muons, etc are not specifically identified), this chapter details the methods and algorithms used in their reconstruction.

4.1 Clustering

The first step to tracks is to aggregate the collected nearby charges on each pixel and SST modules to form clusters or hits. When a particle crosses any sensitive material of the tracker, the resulting charge is often shared between several pixels/strips, meaning each pixel/strip collecting a charge does not correspond each time to a particle crossing the detector. To build clusters, a fairly similar algorithm is used in both the pixel and strip detectors. First, a pixel (strip) with a signal to threshold (noise) ratio higher than a specific value is searched for. This minimum ratio has not been constant over time, in order to adjust and adapt the clustering to all the changing conditions and keep its performance at its best. Then adjacent pixels/strips are added to the cluster, as long as their charge also reaches a specific value, to keep noisy channels out. Finally, a cut on the total cluster charge is applied. Several other parameters, such as Lorentz angle and channel gain calibration are used to correct the total charge and get the cluster local position and its uncertainty on each modules.

4.2 Tracklets

Tracklets are pairs of pixel hits from two out of three pixel layers. As they require very few hits from the layers closest to the IP, they can be used to detect very low-pt tracks, down to 50 MeV. On the other hand, their position measurement is poorer than the full tracking algorithm, and their p_{T} cannot be accurately measured. Since disks on either side of the pixel barrel are not used, its pseudorapidity range goes only to $|\eta| < 2.0$. The tracklets are reconstructed in four steps: the primary vertex is first reconstructed with the method described below; Then for each reconstructed hit, its pseudorapidity is calculated using the primary vertex location; Starting with a hit in the first pixel layer and looping

over the reconstructed hits in the second chosen layer, all possible combinations are saved as proto-tracklets; Lastly the proto-tracklets are sorted by their difference $\Delta\eta$ between the two hits, and in cases where the second hit is used several times only the proto-tracklet with the smallest $\Delta\eta$ is kept.

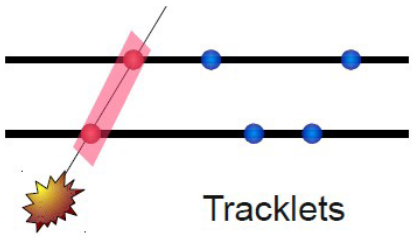

Figure 4.1: Schematic view of the making of tracklets from hits on pixel layers and a vertex.

It is one of the methods used in the first publication of CMS where the pseudorapidity distribution was measured. It also served as a very helpful verification tool for the low energy multiplicity results published by CMS and described in Chapter 5. A detailed description of the same results obtained using tracklets can be found in [18].

4.3 General description of CMS tracking

The tracking used by the CMS collaboration is composed of four steps: first seeds are created to serve as input to the pattern recognition step, which then propagates the seed trajectory to find more hits, then the track is fitted to improve its parameters, and is finally filtered to get rid of doubles and poorly reconstructed tracks.

4.3.1 Seeding

The seeding step is the one providing the actual first signs of what the tracks will be. Seeds are constructed as a collection of hits, that can originate from different sections of the tracker detector, depending on the type of tracks that one wants to find. As they are only the prototypes of tracks, they are composed of only pairs of hits, or eventually triplets, the later showing better constraints on the determined trajectory, but with the downside of a smaller reconstruction efficiency due to eventual detector inefficiencies (see Section 3.4.3 for results on the pixel efficiency measurements) making it easier to reconstruct portion of the tracks with only two hits rather than three - remember, the pixel has three layers, which means seeds built from triplets would require 100% detector efficiency. As the tracker is submerged inside a uniform magnetic field, the trajectory of charged particle is assumed to form a helix.

The majority of tracks are initiated with a seed made of pixel hits only, to minimize the material budget that can pollute the reconstruction and make use of the lower occupancy and faster readout of the pixel detector, with eventually a beamspot constraint to increase the accuracy of the seed parameters measurement for triplets, or allow a calculation of the momentum by adding a mandatory third point for hit pairs. Other tracks are seeded with hit pairs coming from the first two layers of the strip tracker: although it requires reading the information from the strip, thus is inadequate for fast reconstruction used for instance in the HLT, it is helpful to reconstruct tracks not crossing the pixel volume.

4.3.2 Trajectory Building

The building of the complete trajectories inside the CMS tracker is performed using a Kalman filter algorithm, the Combinatorial Track Finder (CTF) algorithm [36]. It works by using a seed as starting point, and propagating the path of the particle up the next potential detector layers. If any compatible hits are found around the projection location, the track parameters are updated with the new information, the hit is added to the track collection, and the propagation continues further out. For each possible hit, a new trajectory candidate is added to the collection: every candidates will be propagated, although only a few best candidates are kept to ensure the efficiency of the tracking. It is also possible, in order to account for possible detector inefficiencies, to allow the propagation of one candidate even without compatible hits, which is also added to the track collection, but as an invalid hit.

Each new propagation refines the trajectory measurement and its precision, until no more detector layers are available, or a stopping condition is met. Several parameters ensure the flexibility of the algorithm, which can therefore be tailored to one's need: for instance the minimum p_{T} of the track, the minimum number of tracker hits, the minimum number of hits, or the maximum number of invalid hits can be tuned separately.

4.3.3 Trajectory Refitting

When all hits are finally included to the track collection, a fit is performed using the Kalman filter combined with a smoother. While the algorithm starts from the hit closest to the interaction point and goes through the list of all hits assigned to the track, the position and its corresponding uncertainty is re-evaluated one last time, taking into account the energy loss and multiple scattering effects. The initial propagation of the seeds is carried out from the center of the detector to the outward layers, in an "in-out" manner, therefore a final fit needs to be applied starting from the most outward hit, to ensure no bias is introduced by building the trajectory with the innermost layers first.

4.3.4 Trajectory Cleaning

There are many particles crossing the detector layers that do not come from the collision point: these can be secondary particles coming from the decay of long-lived particles, or the result of primary particles interacting with inert material from the tracker like cables or electronics. They increase the number of fake tracks reconstructed, are are generally not compatible with the interaction vertex. Several quality cuts are therefore applied on the collection of trajectories, in order to decrease this fake rate, while maintaining the high efficiency of the tracking algorithm. Some of the variables involved in this cleaning are: the ^{2}/ndof of the track, d_{xy} and its associated uncertainty $\sigma_{d_{xy}}$, d_z and its associated uncertainty σ_{d_z}

4.3.5 Vertexing

The determination of the origin of the tracks, which is called a vertex, is very important as a whole during the reconstruction process. Its accurate measurement was one of the physics imperative during the design of the detector, and one of the reason of the pixel subdetector's presence and layout: the resolution of the vertex finding must be small enough to be able to separate several collisions happening during the same bunch crossing,

as well as differentiate secondary vertices coming from decays further down the shower initiated by the initial collision.

The vertex reconstruction is obtained through an adaptative vertex filter [37]: the tracks are grouped according to their z impact parameter within a z_{sep} window of 1 cm. Their coordinates in the transverse plane is not considered, in order to fasten the process, and z_{sep} was later on reduced to 2 mm, when the instantaneous luminosity delivered to the experiments was increased, to ensure the optimal behavior of the vertex finding. Then a selection of those tracks is performed, and finally their point of origin is iteratively computed with a Kalman fit updating the vertex parameters after each iteration. The fit minimizes the separation between the tracks and the vertex position to find the correct one. The associated tracks are weighted from 0 to 1, with fractions allowed, leading to a softer track-vertex inclusion, as tracks are not rejected but assigned a value close to nil if identified as outliers. This has the advantage of allowing outliers without breaking, but is hence more dependent on mis-measured tracks, as well as those coming from another vertex. As the weights are also re-computed during iteration, a first estimate at initialization of the weights or number of outliers is not needed. In order to avoid any local minima during the computation, a geometric annealing is put in place. Later on, a deterministic annealing scheme was developed when PU increased, to cope with the high number of vertices initiated by primary collisions, change which resulted in a noticeable increase in the reconstruction efficiency of the primary vertices.

4.3.6 Beamspot

The CMS experiment makes use of two different ways to determinate the beamspot position. The first one, the most obvious, is by making the distribution of the primary vertices, which would give the "average" 3D location of the interaction point where the collisions happen. This way, the luminous region delivered by the LHC machine for CMS can be mapped with just a few events, as 35 vertices are enough to get a first estimate of the beamspot location and its spread. The other method put in place by CMS exploits the correlation of the transverse impact parameter and the ϕ_0 angle of the tracks when their point of origin is different than expected. This correlation is inserted inside an iterative χ^2 fit [38] to retrieve the beamspot location and parameters, for which a minimum of 150 reconstructed tracks are required.

The beamspot measurement is an important observable, especially to inspect the quality of the collisions delivered by the machine, as the experiments prefer a collision point stable in time, with collisions as close as possible from each other between events, which is measured by the beamspot width. As the CMS experiment - as well as the others - has been designed with an onion layer shape, the interaction is meant to come from its center, in which cases the efficiency of each apparatus is maximized: the beamspot location and its displacement with respect to the CMS center is therefore carefully monitored, as a beamspot too far off would increase the uncertainties in many physics analysis or involve consequent corrections. Hence the beamspot is measured very frequently, and is given as feedback to the LHC operation team, in order for them to monitor their beam and if necessary adjust its parameters to deliver the best collisions possible to the experiments. As the full reconstruction cannot be run on-the-fly for all events, the two methods mentioned above use non-standard objects, delivered by a fast reconstruction: tracks are reconstructed only inside the pixel volume, thus consisting of only pixel hits, which sacrifices accuracy of the trajectory to improve greatly the reconstruction speed, while primary vertices are constructed with only those pixel tracks. An example of the evolution of the beamspot characteristics for a whole LHC fill can be found in [39].

4.4 Iterative tracking and its parameter selection

The tracks used by the CMS analysis are obtained by iterating several times the procedure described in the previous section. After each iteration, the hits that can be unambiguously assigned to a track are removed from the collection of hits available to the next iteration. The seed parameters change between iterations, as well as the cleaning cuts applied. This allows for reconstructing tracks with a broad set of parameters, as each iteration can be tailored for a specific range of trajectories, ensuring the algorithm keeps a high efficiency while reducing the number of fakes to a minimum. It also permits to adapt the complete tracking step to the collision conditions provided by the LHC: this was especially important during the first months of collisions, as both the teams responsible for the proton beams' behavior and the experiments were still learning to provide stable and identical collision parameters over a long period. For instance, the requirements on the impact parameters were loosened during the first days of collisions as the machine provided at first a rather large beamspot, which was later on reduced by a factor ten. The tracking environment was also severely altered with the increase of luminosity, and while the Pile-Up increased, so did the complexity of the tracking. Therefore the iterative steps were adapted and the seeds and quality parameters modified, in order to cope with the higher number of particles measured inside the tracker volume, while ensuring the uniformity over time of the resulting tracks.

4.4.1 Standard Tracking

The tracking used by the vast majority of the CMS analysis is called general tracking. A complete description of each iterations with the value of their parameters can be found in [15], while the study of resulting tracks with the first LHC data is fully detailed in [40].

The first two iterations use pixel triplets and pixel pairs as seeds, to reconstruct tracks with a $p_{\mathrm{T}} > 0.9\,\mathrm{GeV}/c$. The next iteration uses pixel triplets as seeds to find the low-momentum particles. The fourth iteration is meant to reconstruct displaced tracks, and is seeded by a mixture of pixel and strip hits. The last two steps intend on reconstructing tracks that have no pixel hits, either because of inefficiencies in the pixel detector, or because of the geometry of the detectors, and as such are seeded with a combination of strip hit pairs.

4.4.2 Tracking adapted to low-pt

A specific tracking was developed to reconstruct a maximum of low-momentum tracks, to provide the first measurements on minimum-bias data. While it is not as reliable as the general tracking at high luminosity, it was well suited to the first data provided by the LHC, and allowed to reconstruct tracks with a momentum as low as 100 MeV, which are otherwise lost during the standard tracking.

The seeds are built from triplets of pixel hits, constrained by the x and y positions of the beam spot and the z coordinate of the primary vertex. These clean pixel seeds are used in the first iteration. The second iteration uses pixel triplet seeds as well, but does not require a vertex constraint and has a looser transverse impact parameter requirement. Finally, the last iteration finds primary tracks seeded by two hits in the pixel detector, with a requirement of at least three hits for a track to be accepted. Tracks found during the three iterative steps are collected and a second iteration of the primary vertex reconstruction is

performed to refine its position determination. Finally, all the tracks are refitted with the corresponding vertex constraint, hence improving their resolution in η and p_{T}.

Those tracks, hereafter referred to as minbias tracks (for minimum-bias tracks), were used as such during the first few days of data collection at 0.9, 2.36 and 7 TeV. After longer periods of data taking were performed in 2010, it was decided to create a track collection that combines the two tracking types, in order to ensure all low-momentum particles were measured while guaranteeing the tracking quality provided by the standard iterative steps. The combination made particularly sure that tracks present in both collections were not kept twice.

Introduction to the analysis

The chapters that follow encompass the physics data analysis that was performed during my Ph.D. It focuses on the charged particle multiplicity of the *pp* collisions, ie the number of charged particles produced by the collision of the partons from the protons each time the two LHC bunches cross pass. There are several reasons that lead to this subject being the physics portion of this Ph.D. First, it is a rather "simple" analysis that require few events to be collected, and that can hence give a first insight on the data in a timely fashion. As such, it has been performed at every colliders experiment before, from ISR to Tevatron, which provides for a handful of previous results to compare with. This makes for a pertinent benchmark of expectations when discovering collisions at new heights of energy, with consequent predictions made for the LHC energies with Monte-Carlos tuned to previous results from the lower energy experiments. Its measurement provides an insight on the dynamical process of particle production in the collision of two energetic protons, eventually constraining the theoretical models that try to explain it. As will be shown, several can be derived from the multiplicity. Moreover, the inelastic proton-proton cross-section is several orders of magnitude larger than for instance Higgs production, therefore studying the bulk of the physics events and their characteristics is of primordial importance when also looking for more exotic processes that are extremely rare.

The analysis is divided into two sets of results, presented in the next two chapters. Although the second analysis chapter might seem similar to the results explained in the first one, there are significant differences that, first and foremost, warranted the time and efforts spent to produce such results, as well as its consequent place here. What differentiates them most is their anchor in time, as they were not done simultaneously but rather one after the other. This leads to underlying differences, whatever similar the studied observable is. For instance, the first set of results were intended to be published fast, and hence used the collisions acquired in a small period of time, while the second analysis makes use of much more data that was available after the year long of LHC running. The experience and knowledge gathered while performing the first physics analysis was of course very helpful when looking at the new data. Another difference, related to the usage of the distributions by others and comparisons between experiments, was introduced after collaboration-wide discussions between experiments took place, leading to a different event selection in the second analysis.

Chapter 5

Multiplicity analysis with Non-Single Diffractive events

5.1 Introduction

The charged particle multiplicity, or charge multiplicity, n, is an essential observable in hadron collisions. It is the result of the counting of all charged particles produced by the primary interaction. In proton-proton collisions at the LHC, the events collected by minimum bias triggers contain predominantly soft interactions and produce mostly particles with low transverse momenta. In order to understand the dynamics of hadron production, one studies among other quantities the multiplicity spectrum of minimum bias events. In addition, more exclusive measurements of the multiplicity in the presence of one or more hadron jets [41, 42] with a large transverse momentum can provide additional insight into the mechanism of multiple parton interactions [43]. The focus of this chapter is on the inclusive measurement of charged particle multiplicities for minimum bias events in various phase space domains.

A seen previously, the dependence of the multiplicity distribution on $\sqrt{s}$ contains information on the scaling properties of the collision mechanism. Koba, Nielsen and Olesen [5] have shown that the function $\Psi(z) =< n > P_n$, with $z = n/ < n >$ becomes asymptotically independent of $\sqrt{s}$. This scaling was observed to hold at ISR energies [44, 45], and was shown to be violated [46–50] at high-energy $p\bar{p}$ collisions up to at least 2 TeV [51, 52]. The LHC extended the energy domain attained by previous hadron colliders from 0.9 TeV to a center of mass energy of 2.36 TeV and 7.0 TeV. It is therefore of primary interest to verify the validity of KNO scaling at this new energy frontier, verify the energy dependence of the average charged particle multiplicity and test the predictions of Monte Carlo generators that were tuned to lower energy data.

In addition, the shape of the charged particle multiplicity spectrum, $P_n = \sigma_n/\sigma$ is sensitive to the particle production mechanism. Independent emission of single particles leads to a Poisson multiplicity distribution. Deviations from this shape, therefore, reveal correlations and dynamics. Comprehensive reviews on the subject can be found in [1, 4, 53]. Previous measurements were well described by Negative Binomial distributions, indicating short range positive correlations between particles produced close in rapidity space. Quantitatively, these correlations can be studied by means of normalized factorial moments, cumulants and the so-called H-moments.

It is also known that the average transverse momentum of charged particles grows roughly linearly for events with increasing multiplicities. This growth should be fairly

independent of the center of mass energy, but is very sensitive to the Monte Carlo modeling of particle production. The PYTHIA V6.2 [54] generator and subsequent fragmentation model tuned to CDF $p\bar{p}$ data [41, 55] at 1.8 TeV, hereafter called PYTHIA D6T is used as a baseline model to simulate minimum bias collisions that include a soft single diffractive and soft double diffractive component of the inelastic cross section. In order to correct the very high multiplicity data observed at $\sqrt{s}$=7.0 TeV, a technical tune called PYTHIA ATLAS was established that is capable of generating high multiplicity events. It should be noted that this tune is not the result of any comparison with experimental data.

All produced particles are subjected to the CMS detector simulation and reconstruction algorithms as described in Chapter 4, in order to obtain corrections for tracking acceptance, algorithmic efficiencies and tracks that do not originate from the primary interaction. We correct our measurement for these effects by using a statistical unfolding procedure [56]. Ideally this correction procedure should be independent of the underlying physics model used in the Monte Carlo generator, as is demonstrated in Section5.6. Finally the data are corrected for acceptance, trigger and event selection efficiencies.

Various alternate tunings that differ mainly in the modeling of multiple parton interactions [41, 55, 57–60] will be compared with our measurements where relevant. PHOJET [61,62] is used as an alternative event generator that differs mostly in the treatment of the diffractive events. These alternative models will also be used to test the model dependence of our correction procedure and to derive corresponding systematic uncertainties.

Due to the limited acceptance of the tracking devices of most collider experiments, the charge multiplicity spectrum is usually measured in a limited phase space domain of (pseudo)rapidity and/or transverse momentum of the measured particles. A high performance tracking algorithm capable of reconstructing tracks with transverse momenta down to 100 MeV/c will allow us to extrapolate to a full p_{T} acceptance with controllable uncertainties. Due to the limited tracking acceptance of the CMS detector ($|\eta| < 2.4$), an attempt to correct for full phase space will not be made.

5.2 Models

The PYTHIA 6 [54] generator and its fragmentation model tuned to CDF data [41,57], hereafter called PYTHIA D6T, is used as a baseline model to simulate inelastic pp collisions. However, at 7 TeV a dedicated PYTHIA tune [58] describing better the high multiplicities is used for correcting the data. Alternative tunings that differ mainly in the modeling of multiple parton interactions have also been considered [57, 59, 60]. PHOJET [61, 62] is used as an alternative event generator that differs mainly in the underlying dynamical model for particle production. While PYTHIA contains at least one hard scatter per event, particle production in PHOJET, which is based on the dual-parton model [17], is predominantly soft and contains in general multiple-string configurations derived from the dual-parton model with multi-Pomeron exchanges [17]. Each Pomeron exchange gives rise to two strings stretched between either valence or sea partons. At low energies the dominant process is single-Pomeron exchange, which leads to two strings stretched between valence quarks and diquarks. With increasing energy, additional Pomeron exchanges occur, forming strings stretched between sea partons. Because the sea partons carry on average only a small fraction of the momentum of the incident hadrons, these strings are concentrated in the central rapidity region. These extra strings [63,64] are needed to explain the KNO scaling violations observed at high energies [46,65], the increase of the central particle density with increasing energy [46,50,65], the $\langle p_T \rangle$ versus n dependence [66,67], and long-range rapidity

correlations [68]. The distribution of the number of exchanged Pomerons can be obtained from Reggeon calculus and unitarity, by means of fits to the measured total, elastic, and diffractive cross sections as described in [69–71]. Other alternatives based on the Colour Glass Condensate (CGC) picture with saturated gluons [72] also make predictions for the multiplicity dependence of $\langle p_{\mathrm{T}} \rangle$ and for the long-range rapidity correlations.

We also compare our measurements with a new fragmentation model implemented in PYTHIA 8 [73]. In addition to having p_{T}-ordered parton showers, rather than an ordering by virtuality, it differs from its predecessor in the modeling of multiple-parton interactions and the treatment of beam remnants and diffraction. The detailed MC simulation of the CMS detector response is based on GEANT4 [74]. Simulated events were processed and reconstructed in the same manner as collision data.

5.3 Data Sample

Near the end of 2009, the CMS experiment collected two datasets of proton-proton collisions at center-of-mass energies of 0.9 and 2.36 TeV. In March 2010 a new running period at $\sqrt{s} = 7$ TeV started, of which data collected in the first days have been analyzed here. The corresponding inelastic proton-proton interaction rates were about 11, 3, and 50 Hz, respectively, for these datasets. At these rates, the fraction of Pile-Up events, i.e. bunch crossings in which two or more minimum-bias collisions occurred is negligible [35,75].

The trigger and offline event selection is nearly identical to other minimum bias event analysis published by CMS [35, 75], however a factor of three times more events were used in this analysis which were still subject to offline reconstruction and reprocessing when [35] was published. Events were selected by a trigger signal in any of the forward or backward BSC scintillator counters, in coincidence with any of the two BPTX detectors indicating the presence of at least one proton bunch crossing from the IP. Beam halo backgrounds are reduced by using the timing information from the BCS counters at opposite ends of the CMS detector. Beam induced backgrounds are further removed by requiring the cluster sizes in the pixel detector to be compatible with a single primary vertex, as is described in [35].

Non-Single Diffractive (NSD) events are selected by requiring a coincidence of at least one HF calorimeter tower with more than 3 GeV of total energy on each of the positive and negative sides of the forward hadron calorimeter detectors.

Finally a well contained pixel vertex was required within 15 cm around the position of the nominal interaction point along the beamline. The number of these vertices found in each event is shown in Figure 5.1. If more than one vertex was found, the vertex with the highest number of associated tracks was chosen to be the primary vertex.

A total of 130k events were retained for final analysis at $\sqrt{s}$=0.9 TeV, 12k at $\sqrt{s}$=2.36 TeV and 440k at $\sqrt{s}$=7.0 TeV. Event yields at various selection stages for both center of mass energies are shown in Table 5.1.

The total efficiency of the triggers and event selection is shown in Figure 5.2 as function of the true charged particle multiplicity. All measured multiplicities will be finally corrected by their corresponding efficiency.

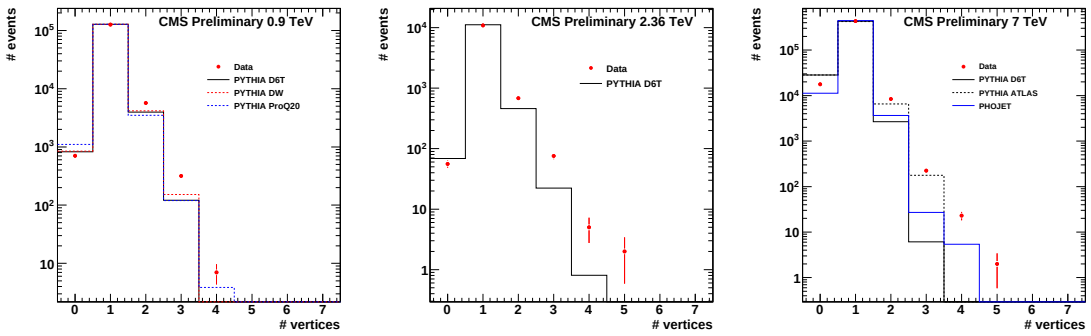

Figure 5.1: The number of reconstructed vertices in each event at each center of mass energy.

Table 5.1: Event yields in each data sample after sequential trigger and event selection.

Selection	$\sqrt{s}$ (TeV)		
	0.9	2.36	7
beam background rejection + L1 trigger	254666	18739	610549
all preceding + forward calorimeters	146658	12019	500077
all preceding + primary vertex	132294	11674	441924

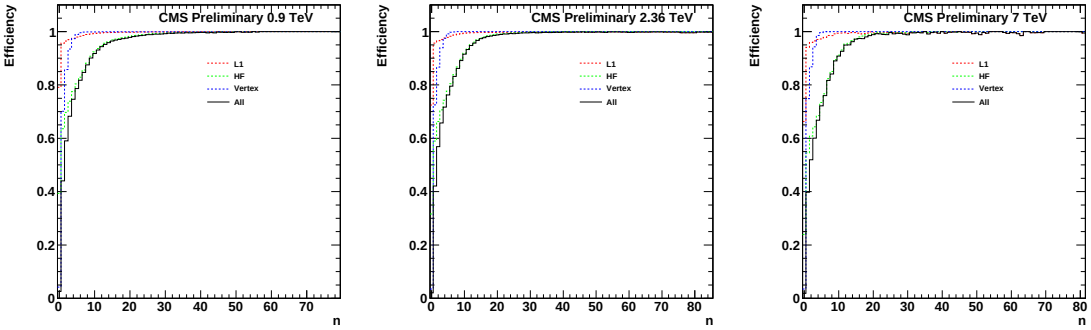

Figure 5.2: The efficiency of the trigger and event selection as function of the true charged particle multiplicity in simulated NSD events at each center of mass energy.

5.4 Track reconstruction and vertex requirements

Both the Pixel and SST detectors were used in the reconstruction of tracks within an acceptance of $|\eta| < 2.5$ and contain information both from the barrel and the end-cap sensors. The standard track reconstruction algorithm of CMS is based on a combinatorial track finder that performs multiple iterations. This iterative procedure was further optimized for primary track reconstruction with transverse momenta down to 100 MeV/c in minimum bias events [35, 76]. The whole track reconstruction is explained in detail in the Chapter 4. The performance of the minimum-bias tracking exceeds the standard tracking, in terms of reconstruction efficiency and fake rate, for track momenta lower than 0.4 GeV/c and down to 0.1 GeV/c. The reconstruction efficiency, estimated by means of the detector simulation, drops below 70% for tracks with a transverse momentum lower than 0.15 GeV/c when using the minimum-bias tracking. For the standard tracking the efficiency drops below 70% at transverse momenta lower than 0.25 GeV/c. Fake rates are comparable between the two algorithms and are kept below 5% for momenta lower than 0.4 GeV/c. At high transverse momenta the reconstruction efficiencies rise to 95% for both algorithms. A comparison of both tracking algorithms in terms of their reconstruction efficiency and fake rate is shown in Figure 5.3 as function of the transverse momentum of the tracks.

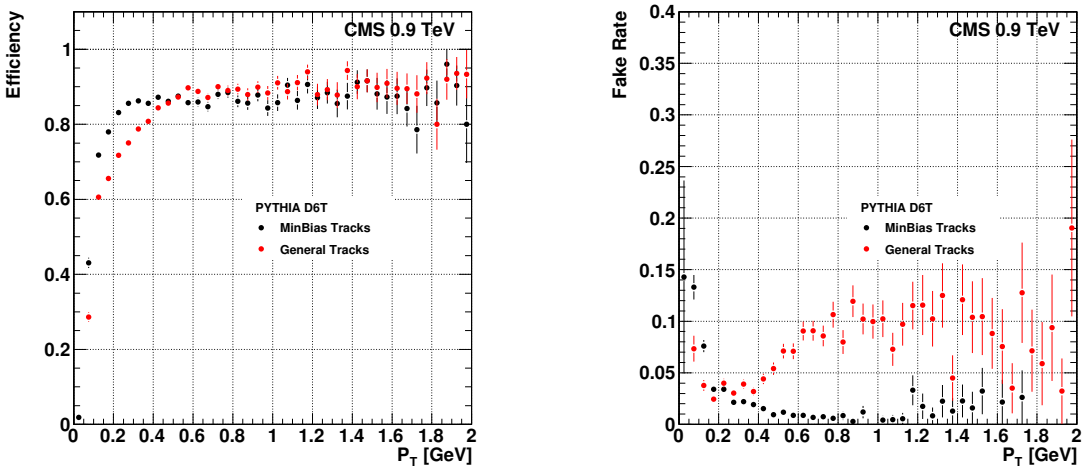

Figure 5.3: A comparison of the track reconstruction efficiency (left) and fake rate (right) between the standard CMS tracking and the minimum bias tracking algorithm as function of the transverse momentum the tracks.

After three iterations of the combinatorial track finder, the position of the primary vertex is recomputed which is then used as an additional constraint in a refit of all previously reconstructed tracks, thus improving the overall resolution in η and p_{T}. An agglomerative clustering algorithm followed by a Gaussian mixture model [77] has been applied in order to optimize the efficiency for finding multiple interaction vertices.

The contamination due to decays of long-lived particles (denoted as V^0 decays) is dominated by charged pions and protons originating from K^0_S and Λ decays. The K^0_S production rate is roughly 5% of that for all charged particles [78], and the resulting charged secondaries amount to 7% of all $\pi^{\pm}$. The Λ production is measured to be 43% of K^0_S [79,80] in this kinematic domain, yielding another 3% of charged secondaries consisting of protons and pions.

In order to reduce the contamination of these V^0 decay products and secondaries produced in interactions of charged particles with the material of the detector, we require all reconstructed tracks to be associated with the primary vertex. This is done by selecting tracks with a small impact parameter with respect to the primary vertex position both in the transverse plane and along the z-axis [35]. The latter implies that for the impact parameter with respect to the beamspot in the transverse plane, d_{xy}, we require

$$d_{xy} < \min(0.2\,\text{cm}\ ,\ 4 \cdot \sigma_{xy}), \tag{5.1}$$

where

$$\sigma_{xy} = \sqrt{\sigma^2_{d_{xy}} + w_x \cdot w_y}, \tag{5.2}$$

with $w_{x,y}$ the width of the beamspot in the x and y direction, respectively. In the direction of the beam axis, we require the point of closest approach of the track with respect to the primary vertex, d_z, to satisfy

$$d_z < 4 \cdot \sigma_z \tag{5.3}$$

where

$$\sigma_z = \sqrt{\sigma^2_{d_z} + \cosh\eta \cdot w_x \cdot w_y}, \tag{5.4}$$

Only 1.5% of K^0_S and Λ decay products pass these selections, resulting in a final contamination of 0.2%. The $K^\pm/\pi^\pm$ ratio is known to be fairly constant over a wide range of center-of-mass energies and is between 8 and 12% [78]. The $K^\pm$ have a lower reconstruction efficiency at low p_T than pions and mis-modeling of the $K^\pm/\pi^\pm$ ratio could result in a change in the multiplicity distribution. However a doubling of the $K^\pm/\pi^\pm$ ratio yields a negligible shift of 0.25% in the multiplicity average. This is expected because the $\langle p_T \rangle$ of $K^\pm$ is substantially higher than $\pi^\pm$, and $K^\pm$ therefore contribute very little in the p_T range where the reconstruction efficiencies differ. Finally, only tracks with a relative uncertainty on their measured transverse momentum smaller than 10% are selected; this requirement rejects mainly low-quality and badly reconstructed tracks.

A comparison between data and MC of the relative momentum error and of both impact parameter significances is given in Figures 5.4 to 5.6, before any selection cuts are applied to the tracks. After application of the aforementioned track quality and vertex association cuts, an overall good agreement between data and MC simulation is found in transverse momentum, pseudorapidity, number of tracker hits used in the track fit and the χ^2 of the track fit, as is shown in Figures 5.7 to 5.9. Note that none of the MC models used is describing the pseudorapidity and the lower end of the transverse momentum distribution very well. This is not due to an inadequate simulation of the detector response done by the CMS software, but rather a feature of the MC modeling of minimum bias data in the event generators. Reference [35] reported the same observations, implying the need for retuning of the MC generators, which is discussed in more details in Section 5.8.

The *raw* multiplicity spectra are obtained by counting all charged particles that are accepted by the track selection cuts and fall completely within the acceptance of the tracking detector, i.e. with $|\eta| < 2.4$. These spectra, compared with MC models, are shown in Figure 5.10 at the three center-of-mass energies. They serve as input for the unfolding and correction procedures described below.

Due to the limited trigger acceptance for SD events in multi-purpose particle physics experiments, the charge multiplicity has historically been mainly measured for non-single diffractive events. Therefore the remaining SD component will need to be subtracted from the data, based on the predictions of the Monte-Carlos, and is further explained below.

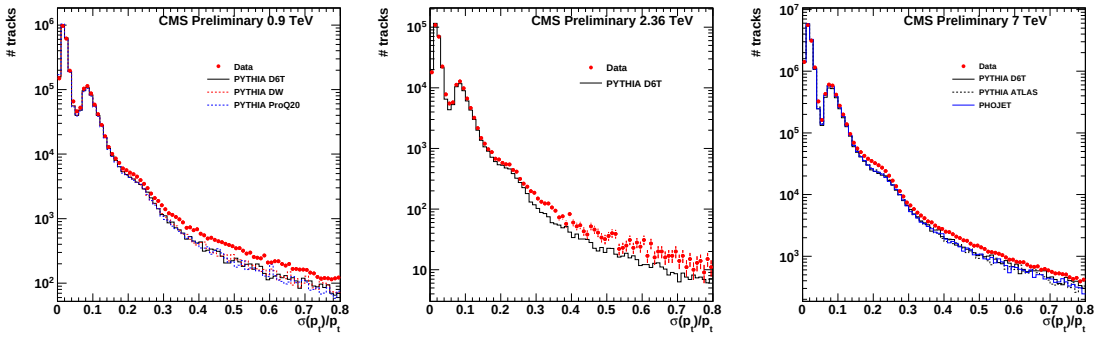

Figure 5.4: The relative error on the transverse momentum measurement at each center of mass energy.

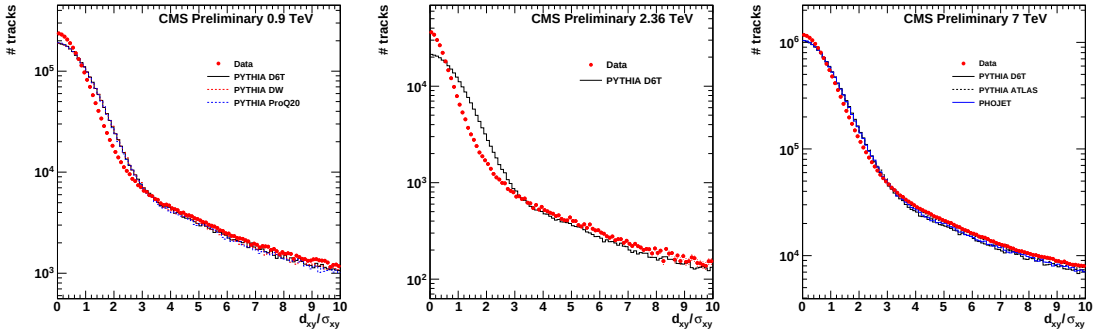

Figure 5.5: The impact parameter significance in the plane perpendicular to the beam pipe at each center of mass energy.

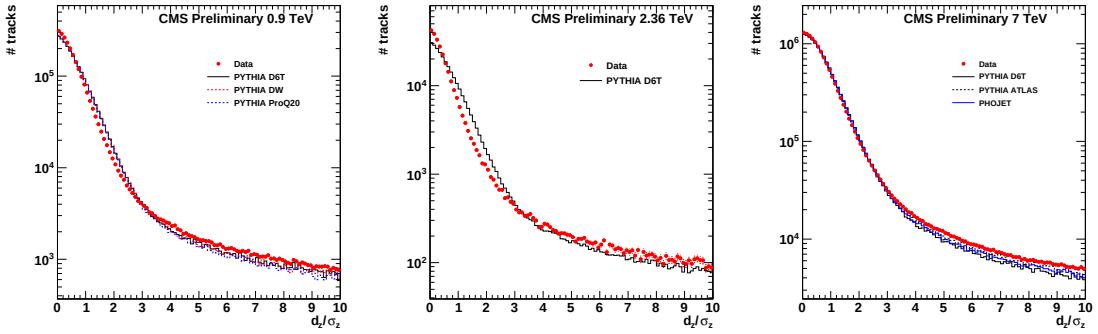

Figure 5.6: The impact parameter significance along the direction of the beam pipe at each center of mass energy.

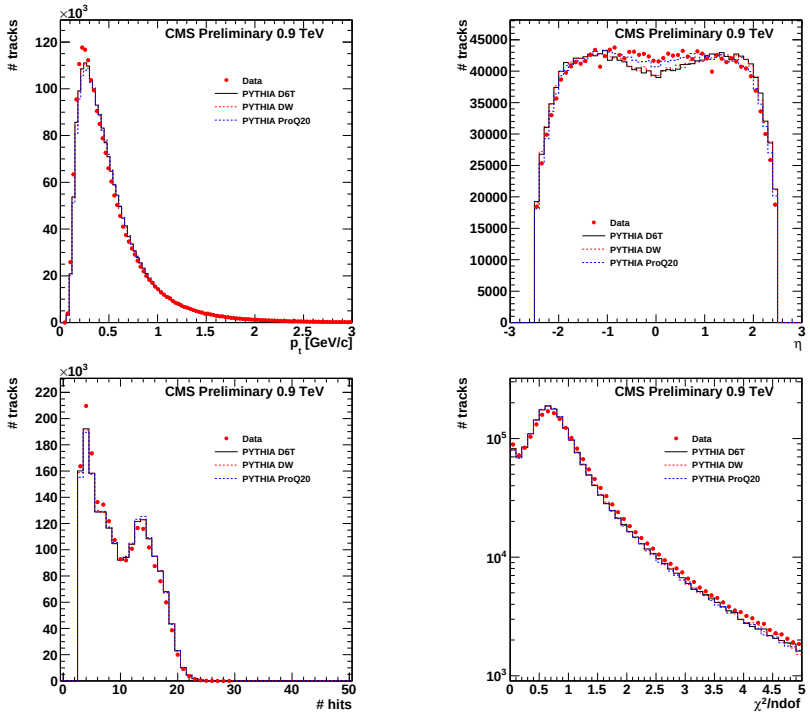

Figure 5.7: Comparison between data and MC at $\sqrt{s}$=0.9 TeV of the transverse momentum, pseudorapidity, number of track hits and the goodness of fit.

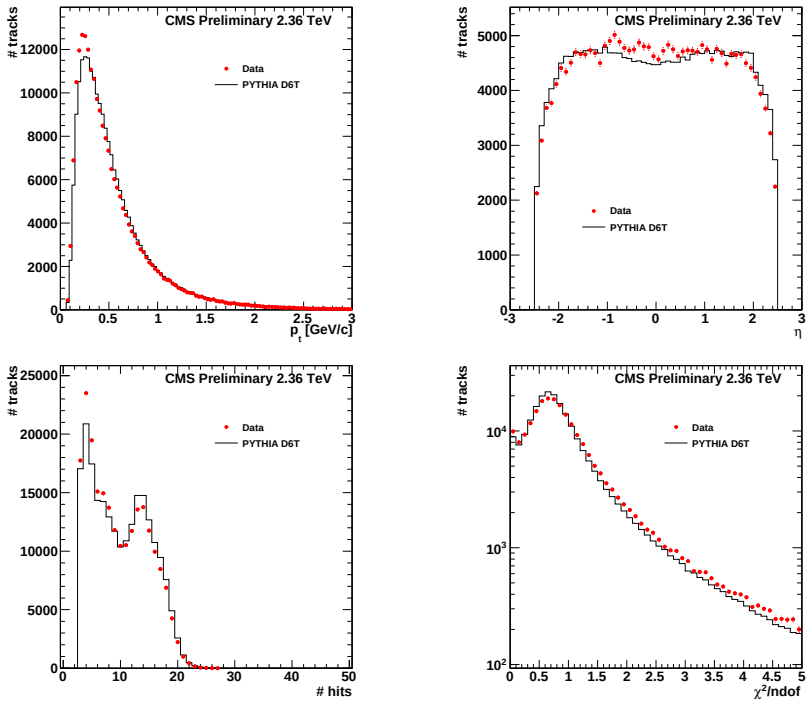

Figure 5.8: Comparison between data and MC at $\sqrt{s}$=2.36 TeV of the transverse momentum, pseudorapidity, number of track hits and the goodness of fit.

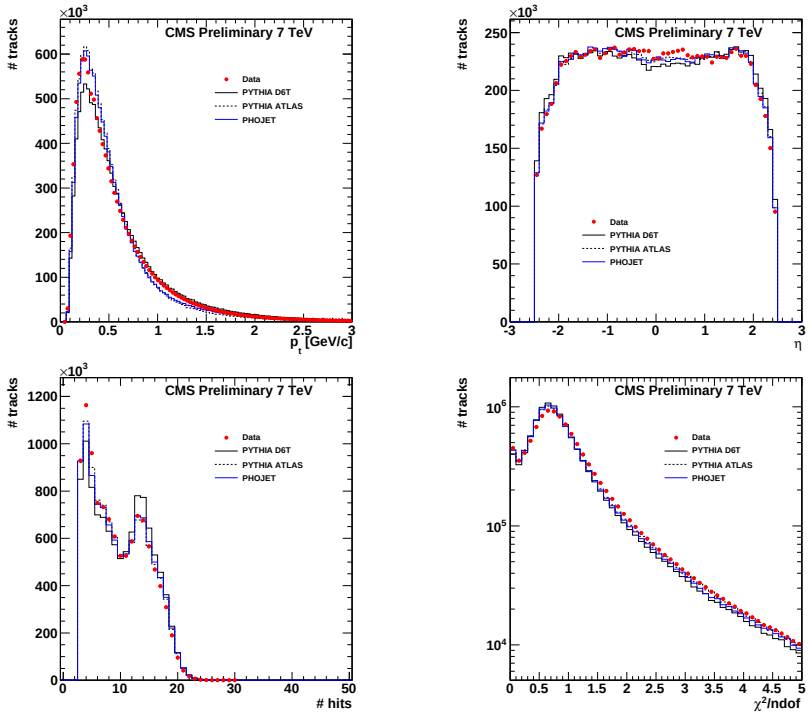

Figure 5.9: Comparison between data and MC at $\sqrt{s}$=7.0 TeV of the transverse momentum, pseudorapidity, number of track hits and the goodness of fit.

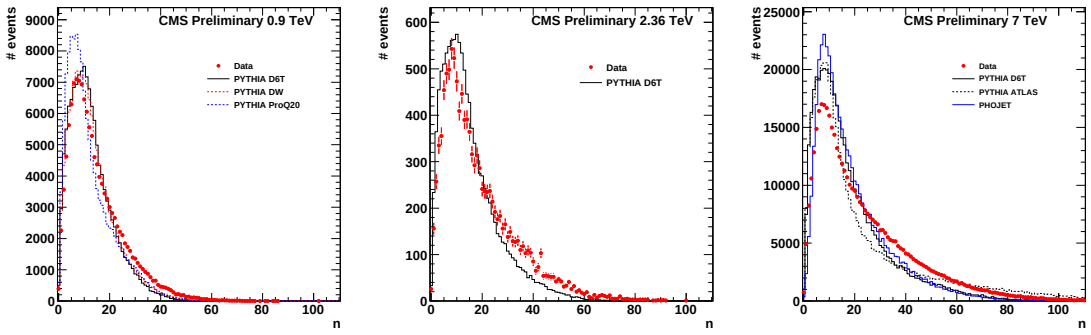

Figure 5.10: Comparison between data and MC models of the raw multiplicity distribution at $\sqrt{s}$= 0.9, 2.36 and 7.0 TeV.

5.5 Acceptance definition

The CMS tracking detector extends physically up to $|\eta| < 2.5$, but at the border of the detector the reconstruction efficiency drops drastically. We therefore limit our acceptance to $|\eta| < 2.4$. The multiplicity distribution will also be reported in smaller pseudorapidity domains down to $|\eta| < 0.5$. Unless stated otherwise, we correct for a full p_{T} acceptance down to 0.0 MeV/c.

Events that have zero reconstructed charged tracks are events that passed the trigger and event selection criteria but have no reconstructed tracks within the acceptance range. The same definition applies to fully corrected primary charged hadron multiplicities. Non-collision backgrounds are predominantly concentrated in this bin, however it was found in the first CMS analysis [35] that their total contribution is of the per mill level.

Due to the requirement of a significant hadronic activity on both sides of the forward calorimeters, the event selection acceptance for single diffractive events is small: 16% at 0.9 TeV increasing to 27% at 7.0 TeV. As the fraction of SD events without any acceptances issues is already small, at most 20% depending on the model used to describe the diffraction process, the SD contribution is therefore simply subtracted to the data using simulated events. The predictions of PYTHIA and PHOJET of the relative fraction of SD events at each center of mass energy is shown in Table 5.2. PHOJET typically predicts a lower amount of SD events than PYTHIA, but recent investigations have shown that the measured fractions in data tends more towards the PYTHIA predictions [75]. The generator level multiplicities for inelastic events simulated by PYTHIA is shown in Figure 5.11. The SD and DD components contribute mostly to the low multiplicities within the acceptance of the CMS tracking detector. Before unfolding, the expected contribution of SD events is subtracted from the observed raw multiplicity distribution using the PYTHIA prediction: this results in a multiplicity spectra representing only NSD events. The effect of this correction is illustrated in Figure 5.12.

Table 5.2: The fraction of single diffractive events at each center of mass energy as predicted by PYTHIA and PHOJET.

$\sqrt{s}$	0.9 TeV	2.36 TeV	7.0 TeV
Generator	SD/INEL (%)	SD/INEL (%)	SD/INEL (%)
PYTHIA	22.5	21.0	19.2
PHOJET	18.9	16.2	13.8

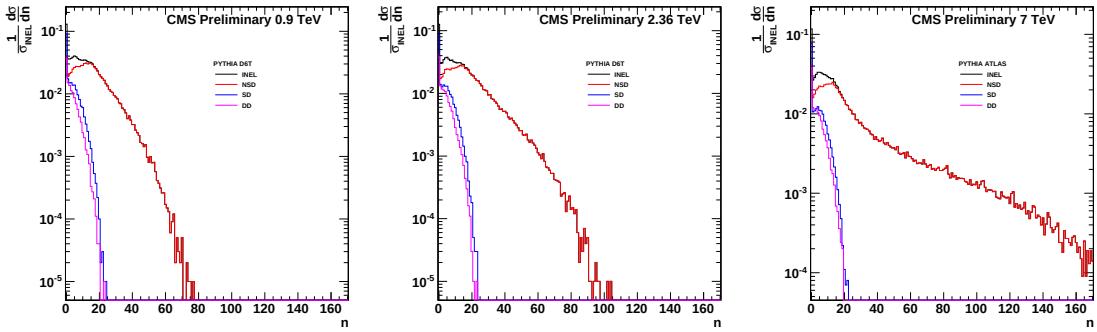

Figure 5.11: The diffractive and non-diffractive contributions to the multiplicity spectrum in $|\eta| < 2.4$ at $\sqrt{s}$=0.9, 2.36 and 7.0 TeV, as predicted by PYTHIA

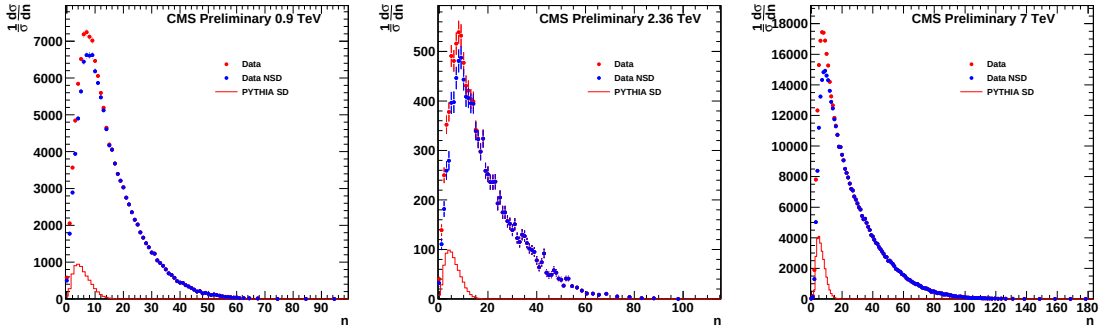

Figure 5.12: The contribution of SD events that are subtracted from the raw data multiplicity before unfolding.

5.6 Corrections

Due to the requirement of significant activity in both ends of the HF, the event selection acceptance for SD events is small: 5% at 0.9 TeV and 7% at 7 TeV. This contribution is therefore subtracted based on simulated PYTHIA events. The PYTHIA and PHOJET predictions of the initial fractions of SD events differ substantially, as seen from Table 5.2. The difference of the two predictions is taken as the systematic uncertainty related to the SD subtraction.

It is customary to normalize the charged hadron multiplicity distribution $P_n = \sigma_n/\sigma$, where σ_n denotes the cross section for a fixed multiplicity n, to either the total inelastic cross section or the NSD cross section. For the results presented in this paper the normalization factor σ corresponds to the latter.

The minimum-bias trigger and NSD selection unavoidably introduce a bias in the measured charged hadron multiplicity. Furthermore, a fraction of the events are removed by the requirement of a good quality primary vertex. These effects result in an accepted multiplicity given by

$$T_n = \epsilon_n \cdot P_n, \tag{5.5}$$

where ϵ_n is the trigger and event reconstruction efficiency for multiplicity n. This efficiency is close to 100% for multiplicities larger than $n = 20$ and drops gradually to 40% for $n = 1$, as shown in the examples of Figure 5.2.

Due to inefficiencies in track reconstruction and acceptance, the creation of secondary particles by the interaction of primaries with the beam pipe and the detector material and the presence of decay products of long lived hadrons, one will in general not measure the true accepted multiplicity, n, but a statistically related quantity, m. If O_m is the statistical distribution of this observed multiplicity, it can be related to the true accepted multiplicity distribution by a linear relationship

$$O_m = \sum_n P_{m,n} \cdot T_n, \tag{5.6}$$

The problem of inverting $P_{m,n}$ is well known and extensive literature on the topic exists. When dealing with limited statistics, a algebraic inversion of $P_{m,n}$ is often impossible or leads to unstable results. Therefore an iterative unfolding procedure is applied. We apply a Bayesian unfolding, as described in [56]. The unfolding matrix, $P_{m,n}$, is shown as an example in Figure 5.13 for NSD events with $|\eta| < 2.4$. The average transverse momentum, $< p_{\mathrm{T}} >$ in each multiplicity bin was measured as well. For each bin in raw multiplicity, a Monte Carlo based correction factor $< p_{\mathrm{T}}^{gen} > / < p_{\mathrm{T}}^{rec} >$ was applied to convert the measured $< p_{\mathrm{T}} >$ to the corresponding value for primary charged hadrons. Subsequently, the unfolding matrix $P_{m,n}$, was applied to correct the multiplicities and repopulate the $< p_{\mathrm{T}} >$ values.

The correctness of the unfolding procedure was verified using MC events where we compare the generated hadron multiplicity with the unfolded reconstructed tracks. At all energies, a perfect agreement with the generated hadrons is achieved after unfolding, as can be seen in Figure 5.14 and 5.15. After unfolding, we correct for the trigger and event selection efficiency using the curves shown in Figure 5.2. The effect of these corrections is also shown in Figure 5.14.

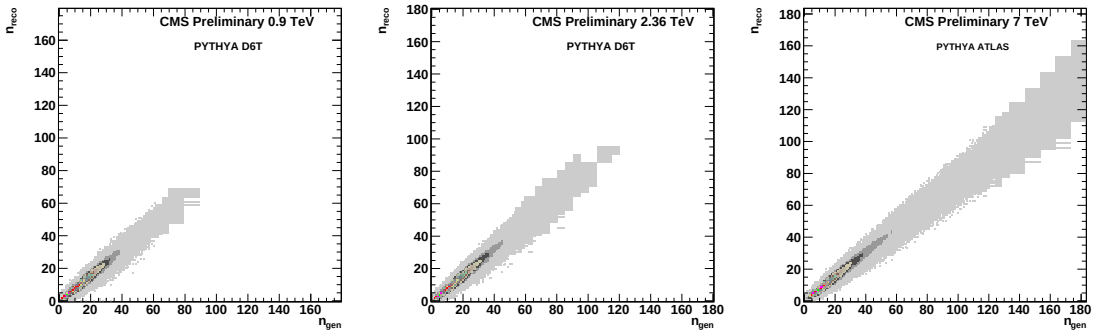

Figure 5.13: Matrices used for unfolding.

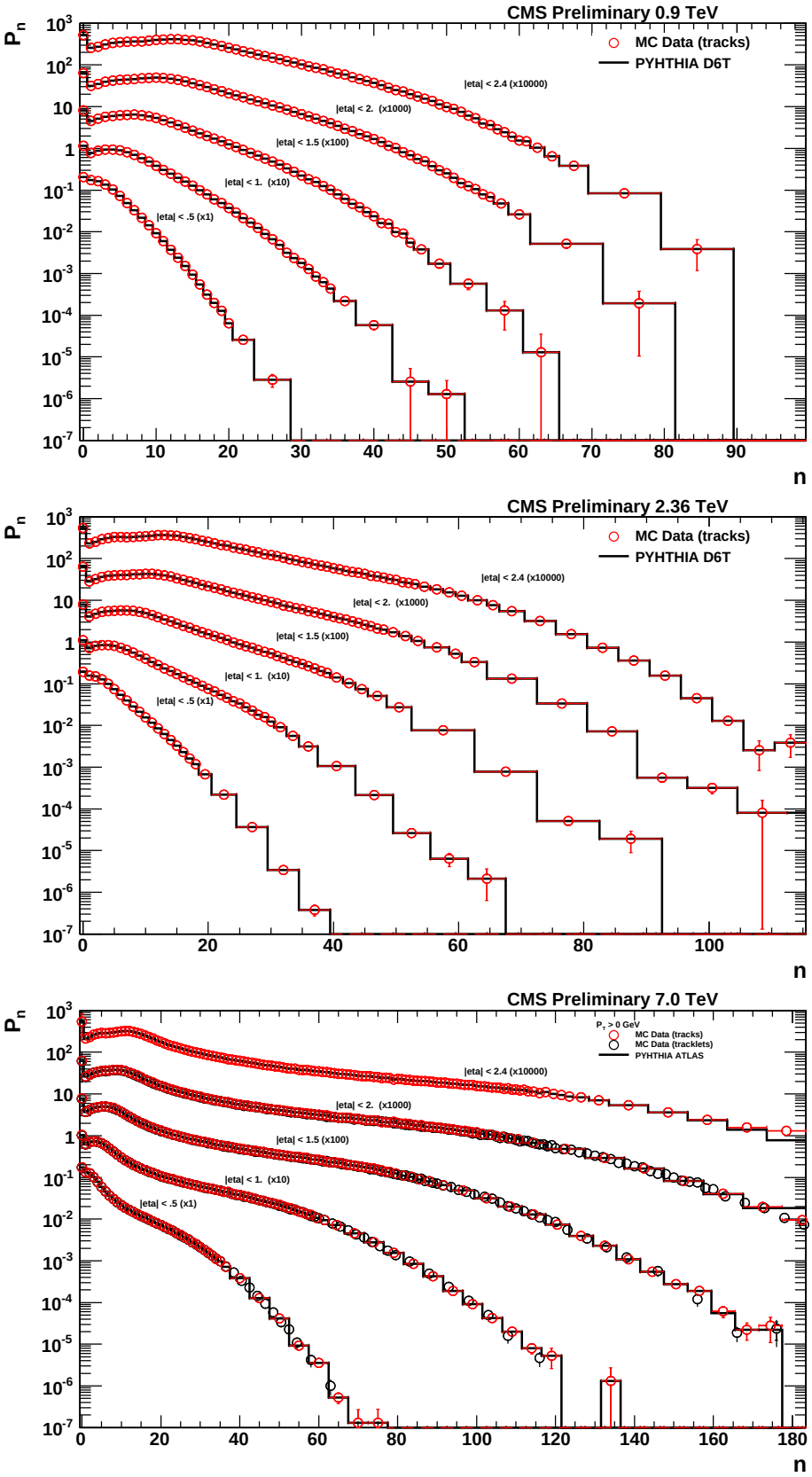

Figure 5.14: A sanity check of the unfolding procedure of the charged particle multiplicity distribution, P_n, correcting reconstructed tracks back to the hadron level. The efficiency corrections for trigger and event selection criteria is also applied.

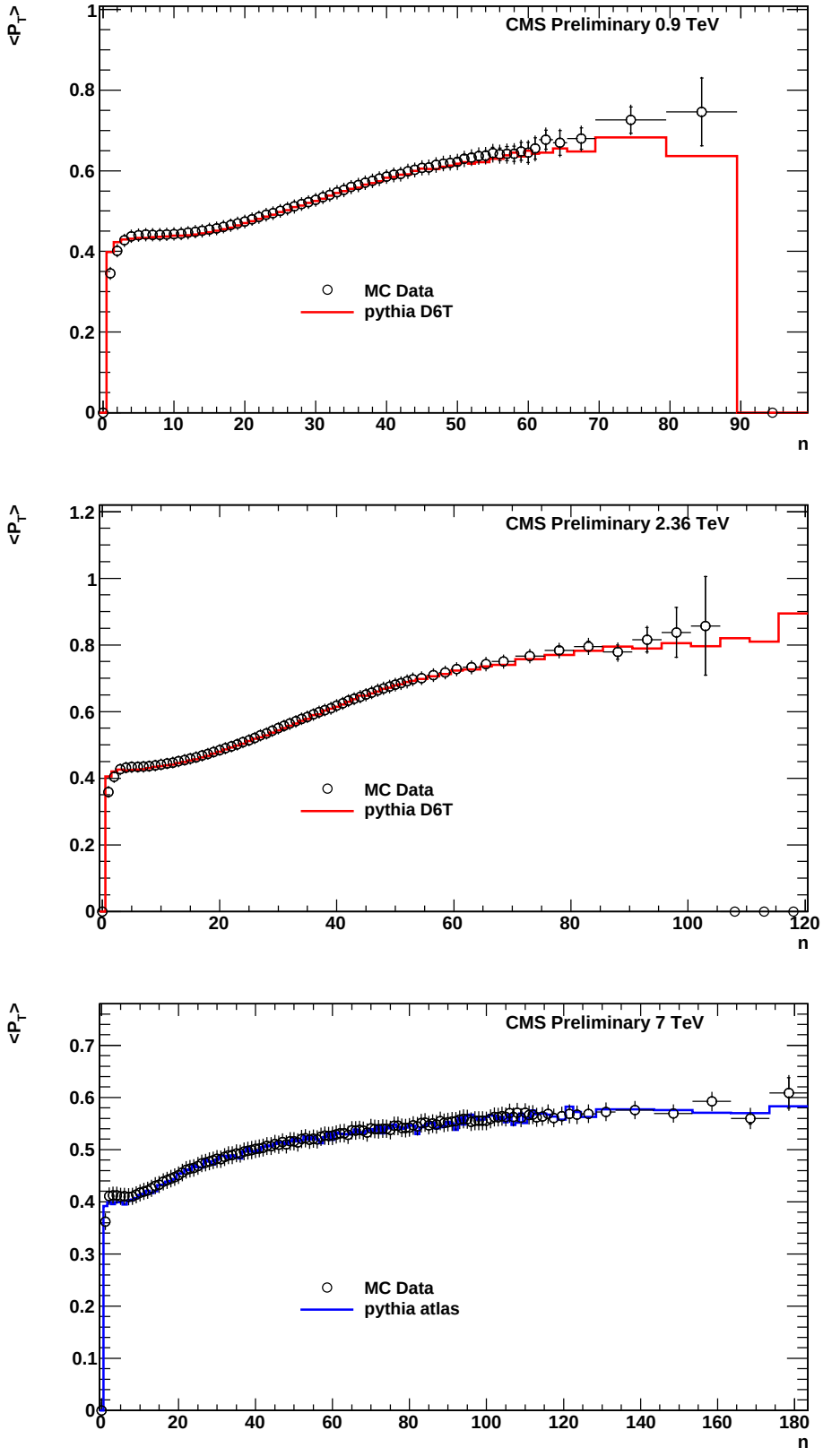

Figure 5.15: A sanity check of the unfolding procedure of the $< p_{\mathrm{T}} >$ vs n profile, correcting reconstructed tracks back to the hadron level.

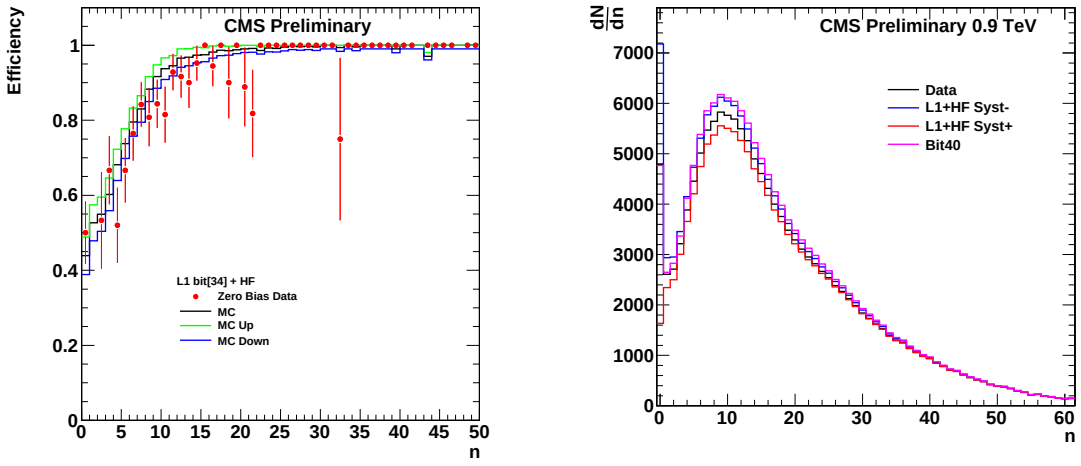

Figure 5.16: An independent measurement of the trigger efficiency with zero bias data compared with the up and down variations applied to the MC description to bracket the measured values from data (left). The effect of the systematic variation of the event and trigger selection efficiency is shown to be of the same size as the effect of replacing the HF offline selection criterium with an alternative tighter trigger requirement (right).

5.7 Systematic uncertainties

Three main sources of systematic uncertainties were identified in this analysis: the uncertainty on the trigger and event selection efficiency, the combined effect of tracking efficiency, fakes, secondaries, misalignment and vertexing, and the model dependent SD subtraction.

Trigger and selection efficiency

The corrections for the trigger and selection efficiency are entirely based on a MC simulation. Crosschecking the multiplicity dependent efficiency with zero bias events showed a good agreement between data and MC, as is shown in Figure 5.16. In order to contain most of the data efficiency measurements the MC efficiency was varied with 5% around its nominal value at low multiplicities and with 1% at high multiplicities. We performed an additional crosscheck that replaced the offline HF coincidence by a stronger trigger requirement demanding both BSC counters on both sides of CMS to fire simultaneously (trigger bit 40) instead of minimally one of the two counters showed a similar shift in the measured multiplicity distribution as can be seen in the right hand plot of Figure 5.16.

Figure C.1 in appendix shows the relative shift w.r.t. the nominal multiplicity distribution, P_n, observed in data due to the systematic increase or decrease of the efficiency correction factors. It is clear that this effect is largest for multiplicities below $n = 10$, ranging from 50% for $n = 0$ to less than 10% for $n = 10$.

Tracking reconstruction

A correct description of the tracking efficiency in the Monte Carlo simulation of the detector is essential in obtaining a correct unfolding matrix. At low transverse momenta,

Table 5.3: Summary of systematic uncertainties on the track reconstruction.

Source	tracking uncertainty (%)
Tracking efficiency	2.0
Acceptance	1.0
Pixel hit efficiency	0.3
Pixel cluster splitting	0.2
Correction for secondaries	1.0
Misalignment	0.1
Beam halo	0.1
Multiple track counting	0.1
Mis-reconstructed tracks	0.5
Total	2.5

the efficiency drops due to a loss of hits on the tracks that are effectively stopped within the tracking volume. A conservative estimate of the uncertainty of this efficiency of 2% is taken from [35]. The fake tracks are mostly secondaries originating from interactions with the material of the LHC beam pipe around the interaction point. This was estimated in [35] to be no more than 1%. These uncertainties, together with smaller contributions from the misalignment of the tracking detector and the extrapolation uncertainty from p_{T} =0.1 GeV/c to zero add up in quadrature to a total tracking uncertainty of 2.6%. As was discussed in section 5.4, the contamination from V^0 decays was negligible after associating tracks with the primary vertex. The effect on the multiplicity distribution, shown in Figure C.2 in appendix was estimated by decreasing and increasing the corrected multiplicity measurements with 2.6%. The systematic uncertainty becomes very small at a pivot point around $n = 20$ which is very close to the mean of the original multiplicity distribution. In general the systematic uncertainty remains below 10% for a large part of the multiplicity distribution. All Correction factors related to tracking and reconstruction are summarized in Table 5.3.

Model dependence

Finally, the modeling of single diffractive events was shown to be different between PYTHIA and PHOJET, both in relative fraction and in predicted multiplicity distribution. The largest variation that can be obtained is by subtracting the PHOJET multiplicity distribution for SD events, using the relative fraction predicted by PYTHIA and vice versa. The impact is largest for small multiplicities, as can be seen in Figure C.3 in appendix. The relative uncertainties are largest at small multiplicities, since this is where the diffractive events contribute most. A pivot point that corresponds to the crossing point of the two alternative SD distributions can also be clearly observed. The model dependent systematic for tracklets is identical to that for the tracking result.

All three systematic uncertainties are calculated separately for upward and downward effects on the multiplicity distribution and added separately in quadrature to yield a total up-down systematic uncertainty. The systematic uncertainties on the n vs. $< p_{\mathrm{T}} >$ distribution were calculated in an analogous way. The effect of the event selection and

tracking efficiency was observed to be very small in this case. The uncertainty of the transverse momenta of the charged particles was taken into account by lowering or raising their p_{T} values with their respective uncertainty, obtained from the track fit. This resulted in a systematic variation of $< p_{\mathrm{T}} >$ of 2

The total uncertainty on P_n remains below 10% for a large part of the multiplicity distribution and increases to 25% at low and high multiplicities, as can be seen in Figure 5.17.

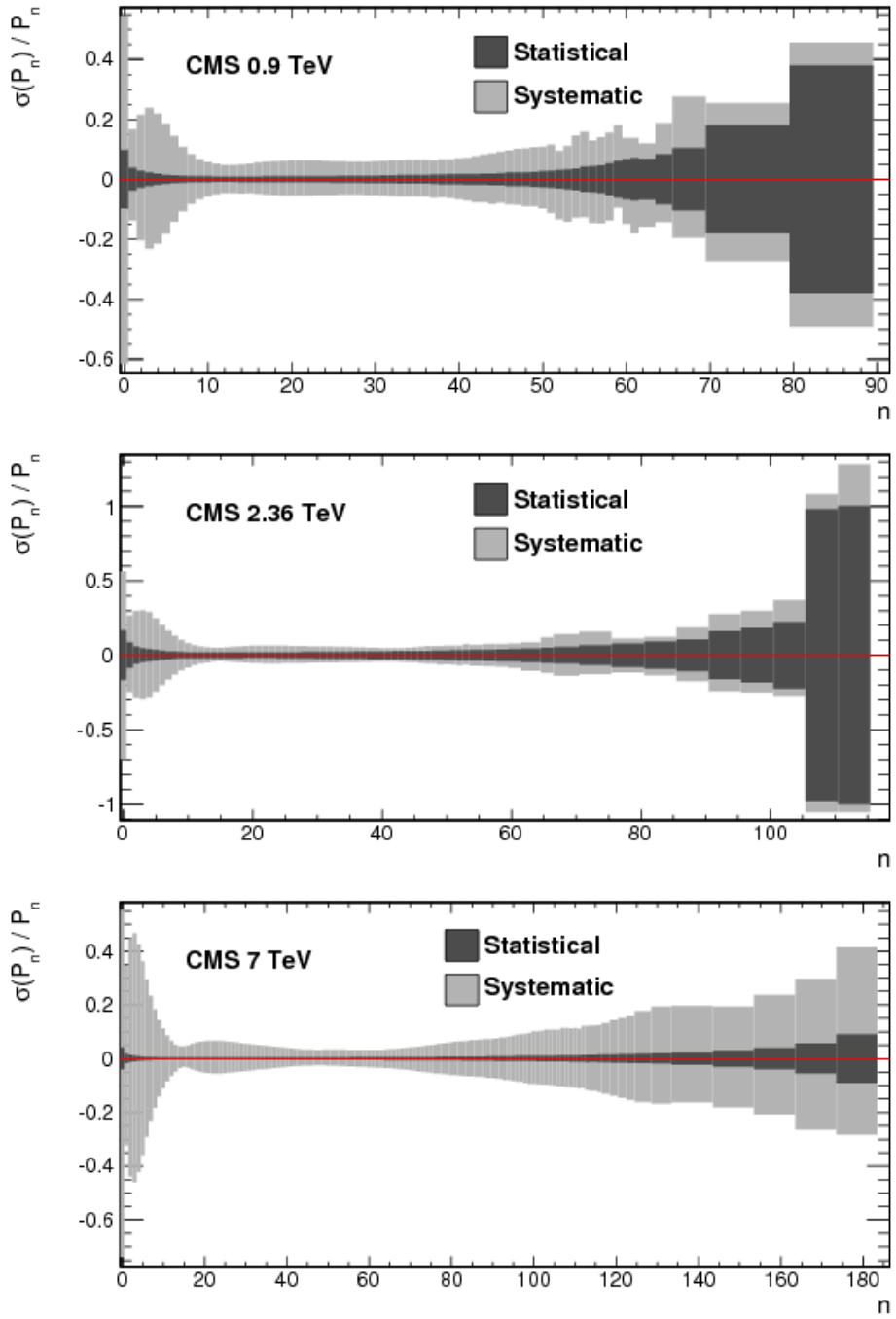

Figure 5.17: The relative size of the statistical and total uncertainty on P_n (in percents) for each multiplicity.

5.8 Results

5.8.1 Charged hadron multiplicity distributions

The NSD charged hadron multiplicity distributions are measured in increasing ranges of pseudorapidity from $|\eta| < 0.5$ to $|\eta| < 2.4$. The fully corrected results at $\sqrt{s} = 0.9$, 2.36, and 7 TeV are compared in Figure 5.18 with earlier measurements in the same pseudorapidity ranges performed by the UA5 [46, 65] and ALICE [81, 82] collaborations. Our measurements were also compared with results obtained with a CMS cross-check analysis (not shown) on data at $\sqrt{s} = 0.9$ and 7 TeV, using a tracklet-based tracking algorithm as in Reference [35]. With a reconstruction efficiency exceeding 90% for $p_{\mathrm{T}} > 50\,\mathrm{MeV}/c$, the latter provided a cross-check of the extrapolation for tracks below $p_{\mathrm{T}} < 100\,\mathrm{MeV}/c$, including the use of data without magnetic field at 7 TeV. The charged particle multiplicity was also measured for hadrons with transverse momentum higher than 500 MeV, and compared with ATLAS results [83] in Figure 5.19. All measurements agree well within their total uncertainties.

In the largest pseudorapidity interval of $|\eta| < 2.4$, there is a change of slope in P_n for $n > 20$, indicating a multicomponent structure as was discussed in [51] and [84] in terms of multiple-soft-Pomeron exchanges. This feature becomes more pronounced with increasing center-of-mass energies, notably at $\sqrt{s} = 7\,\mathrm{TeV}$.

As the CMS tracker is completely submerged inside a uniform magnetic field, it is easy to differentiate positively and negatively charged particles crossing the detector volume. Such a comparison is shown in Figure 5.20 for $|\eta| < 2.4$ and only with particles with $p_{\mathrm{T}} > 500\,\mathrm{MeV}/c$. There are no discernible differences the two distributions, as expected by the dictate of charge conservation.

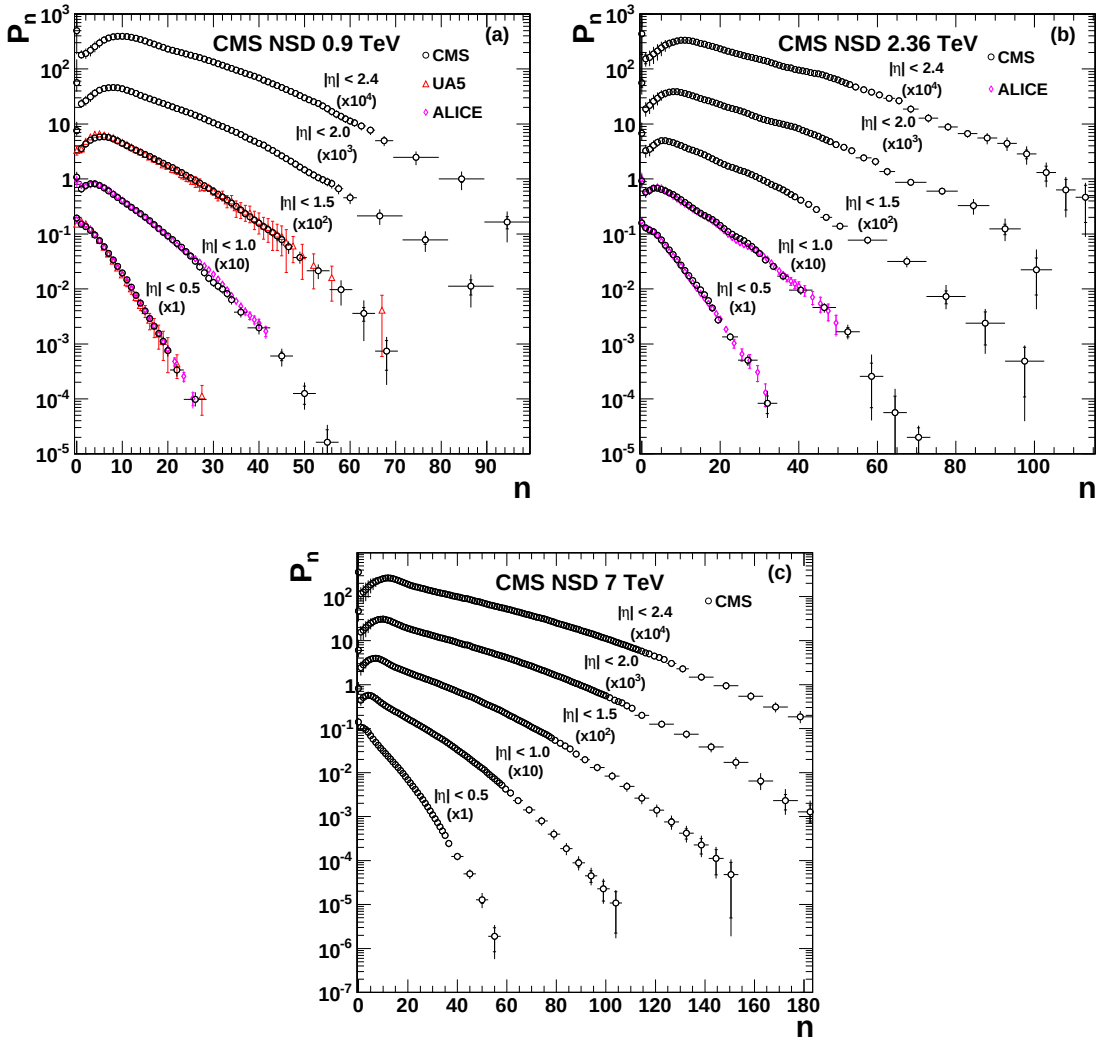

Figure 5.18: The fully corrected charged hadron multiplicity spectrum for $|\eta| < 0.5$, 1.0, 1.5, 2.0, and 2.4, (a) at $\sqrt{s} = 0.9\,\text{TeV}$, (b) 2.36 TeV, and (c) 7 TeV, compared with other measurements in the same η interval and at the same center-of-mass energy [46,65,81,82]. For clarity, results in different pseudorapidity intervals are scaled by powers of 10 as given in the plots. The error bars are the statistical and systematic uncertainties added in quadrature.

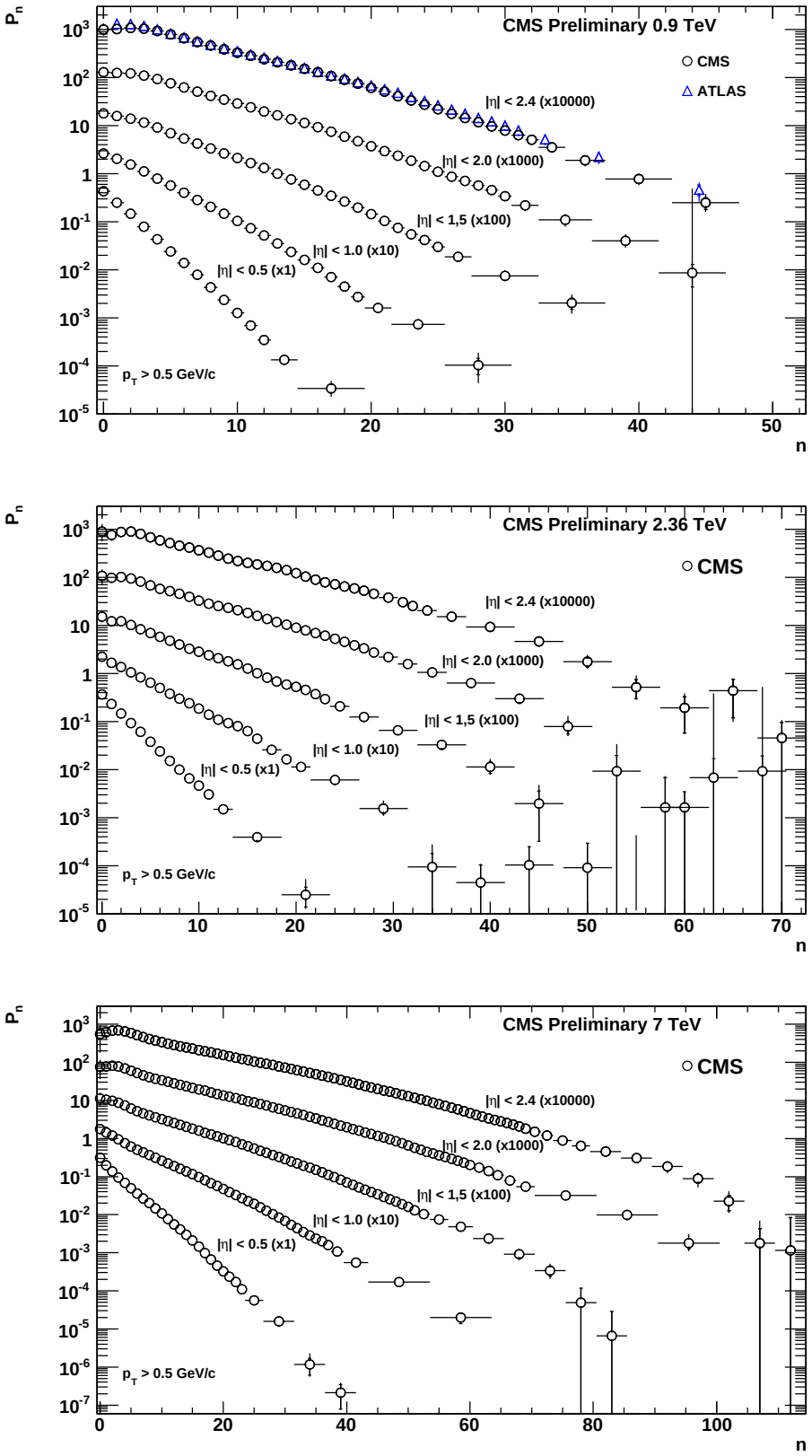

Figure 5.19: The fully corrected charged hadron multiplicity spectrum for $|\eta| < 0.5$, 1.0, 1.5, 2.0, and 2.4, with $p_{\mathrm{T}} > 500\,\mathrm{MeV}/c$ at $\sqrt{s}$=0.9, 2.36 and 7.0 TeV, compared with another measurement in the same η interval and at the same center-of-mass energy [83]. For clarity, results in different pseudorapidity intervals are scaled by powers of 10 as given in the plots. The error bars are the statistical and systematic uncertainties added in quadrature.

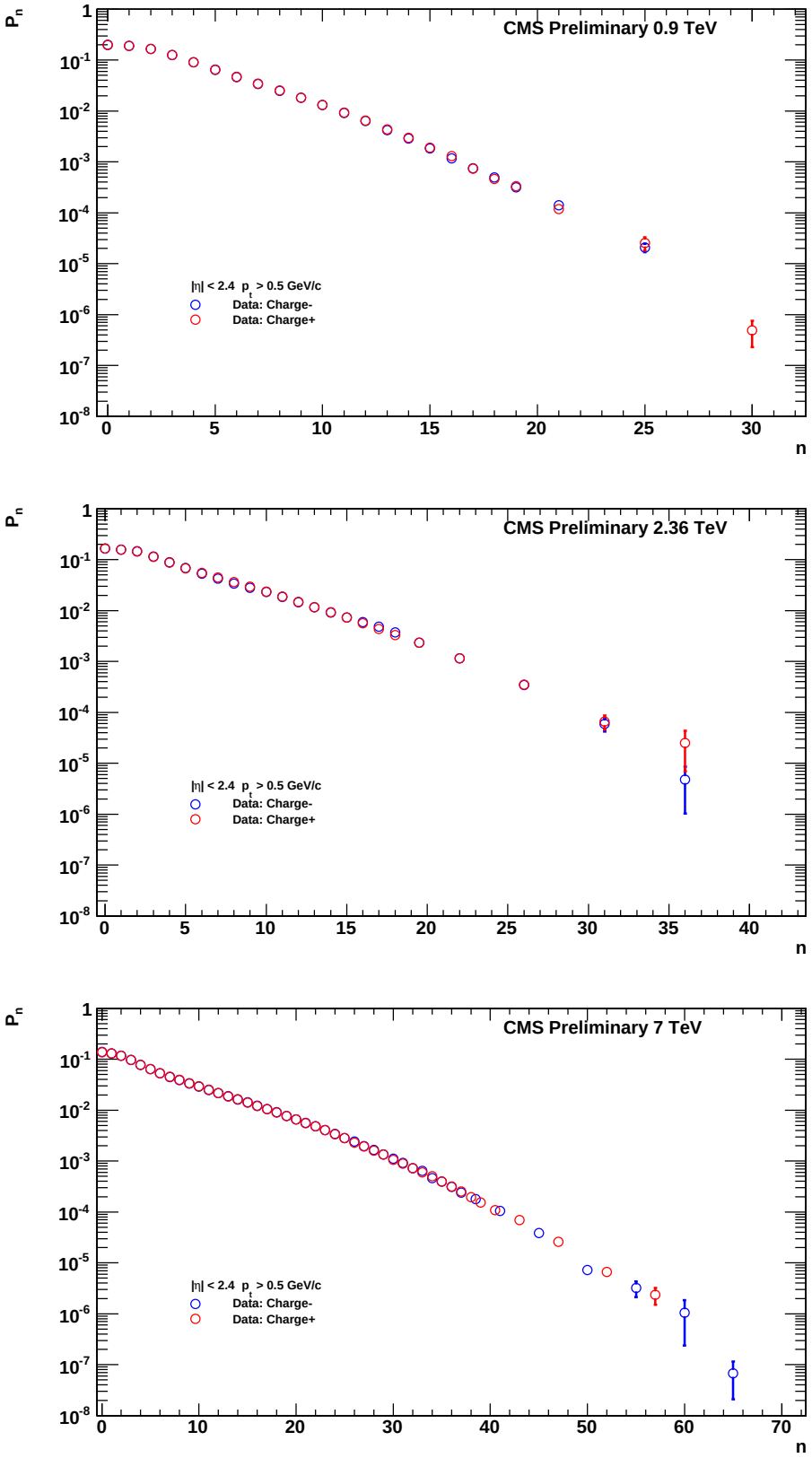

Figure 5.20: A comparison between the fully corrected multiplicity spectra for positively and negatively charged hadrons with $p_{\mathrm{T}} > 500\,\mathrm{MeV}/c$ in increasing pseudorapidity bins ranging from $|\eta| < 0.5$ to $|\eta| < 2.4$ at $\sqrt{s}$=0.9, 2.36 and 7.0 TeV.

5.8.2 Violation of KNO scaling

The multiplicity distributions are shown in KNO form in Figure 5.21 for a large pseudorapidity interval of $|\eta| < 2.4$, where we observe a strong violation of KNO scaling between $\sqrt{s} = 0.9\,\text{TeV}$ and $7\,\text{TeV}$, and for a small pseudorapidity interval of $|\eta| < 0.5$, where KNO scaling holds. Scaling is a characteristic property of the multiplicity distribution in cascade processes of a single jet with self-similar branchings and fixed coupling constant [85–92].

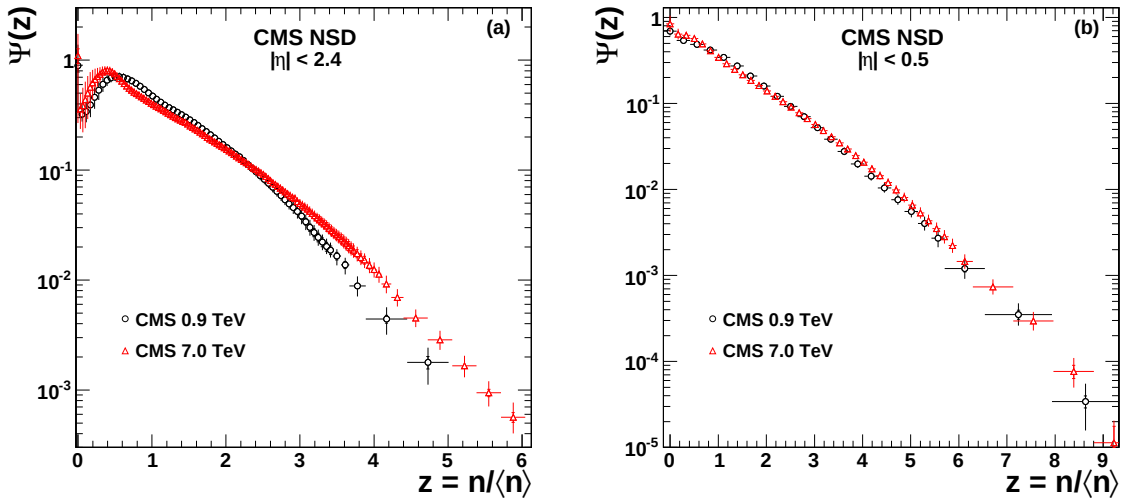

Figure 5.21: The charged hadron multiplicity distributions in KNO form at $\sqrt{s} = 0.9$ and $7\,\text{TeV}$ in two pseudorapidity intervals: $|\eta| < 2.4$ (left) and $|\eta| < 0.5$ (right).

The validity of KNO scaling is shown more quantitatively in Figure 5.22 by the normalized order-q moments C_q of the multiplicity distribution, complemented with measurements at lower energies [47, 93, 94]. For $|\eta| < 2.4$ the values of C_q increase linearly with $\log s$, which demonstrates the KNO scaling is violated. On the other hand, they remain constant up to $q = 4$ over the full center-of-mass energy range for the small pseudorapidity interval $|\eta| < 0.5$, which indicates the scaling is valid. This is well illustrated by the fits of the C_q moments in Figure 5.22.

Multiplicity distributions for e^+e^- annihilations up to the highest LEP energies show clear evidence for multiplicity scaling, both in small ranges ($\Delta\eta < 0.5$), in single hemispheres, and in full phase-space. However, at LEP energies, scaling is broken for intermediate-size ranges where, besides two-jet events, multi-jet events contribute most prominently [95–99].

For hadron-hadron collisions, approximate KNO scaling holds up to ISR energies [44, 45], but clear scaling violations become manifest above $\sqrt{s} \approx 200$ GeV both for the multiplicity distributions in full phase space and in central pseudorapidity ranges [51,52,93,100]. In $p\bar{p}$ collisions, and for large rapidity ranges, the UA5 experiment was the first to observe a larger than expected high-multiplicity tail and a change of slope [47, 51], which was interpreted as evidence for a multi-component structure of the final states [9, 64, 84]. Our observation of strong KNO scaling violations at $\sqrt{s} = 7\,\text{TeV}$, as well as a change of slope in P_n, confirm these earlier measurements.

All these observations, together with the sizable growth with energy of the non-diffractive inelastic cross-section, point to the increasing importance of multiple hard,

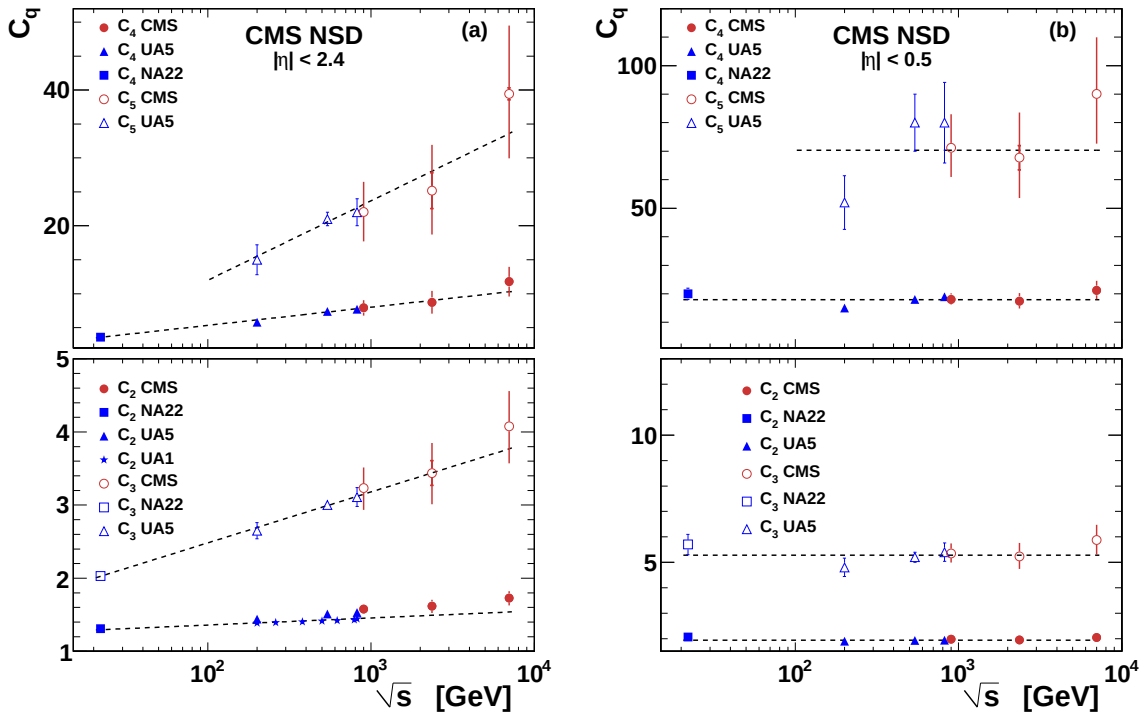

Figure 5.22: Fits of the $\log s$ dependence of the normalized moments C_q of the multiplicity distribution for $|\eta| < 2.4$, assuming linear dependence (left) and $|\eta| < 0.5$, assuming no dependence (right), including data from lower energy experiments [47, 93, 94]. For $\sqrt{s} = 0.9\,\text{TeV}$, data from experiments other than CMS were drawn shifted to lower $\sqrt{s}$ for clarity.

semi-hard, and soft partonic subprocesses in high energy hadron-hadron inelastic collisions [17, 51, 64, 101–103]. These effects are less prominent in central events, where also fewer long-range correlations are taken into account.

5.8.3 Comparison with Monte-Carlo models

An extensive range of tunes [41,55,57–60] based on the PYTHIA 6 fragmentation model have been developed. They differ mainly in their parametrization of the multiple-parton-interaction model. Some reproduce the charged hadron multiplicities better than others, but none is able to give a good description simultaneously at all the center-of-mass energies and in all pseudorapidity ranges. For clarity, only the baseline tune D6T is thus shown, in comparison with other models having a different physical description of soft-particle production such as PHOJET [61, 62] and the new fragmentation model of PYTHIA 8 [73].

A comparison of our measurements with these three classes of models is shown in Figure 5.23 for all charged hadrons (left) and for those with $p_\text{T} > 500\,\text{MeV}/c$ (right). PYTHIA D6T underestimates drastically the multiplicity at all measured energies but improves when $p_\text{T} > 500\,\text{MeV}/c$ is required. PYTHIA 8 is the only model that gives a reasonable description of the multiplicity distribution at all energies, but tends to overestimate the multiplicity at 7 TeV when $p_\text{T} > 500\,\text{MeV}/c$ is required. PHOJET produces too few charged hadrons overall but gives a good description of the average transverse momentum $\langle p_\text{T} \rangle$ at fixed multiplicity n, as illustrated in Figures 5.24 and 5.25. Among the three classes of

models, PYTHIA 8 gives the best overall description of the multiplicity distribution and the dependence of the average transverse momentum on n.

Inspired by [104] we fit in Figure 5.25 a first-degree polynomial in $\sqrt{n}$ to the multiplicity dependence of $\langle p_{\mathrm{T}} \rangle(n)$ for $n > 15$ at each energy, yielding a good description which is valid at all three energies. The ratios of the data obtained at 7 and 2.36 TeV with respect to the data at 0.9 TeV show that the rise of the average transverse momentum with the multiplicity is weakly depending on energy.

All previous observations seem to indicate that the Monte-Carlo models produce too few particles with low transverse momenta, especially at 7 TeV. The PYTHIA models tend to compensate for this by producing too many particles with high transverse momentum, which is related to the modeling of semi-hard multiple-parton interactions. The number of multiple-parton-interactions have clearly been underestimated by the Monte-Carlo generators, and they also appear to be softer than what can be predicted by the models, even when increasing artificially the production of MPI.

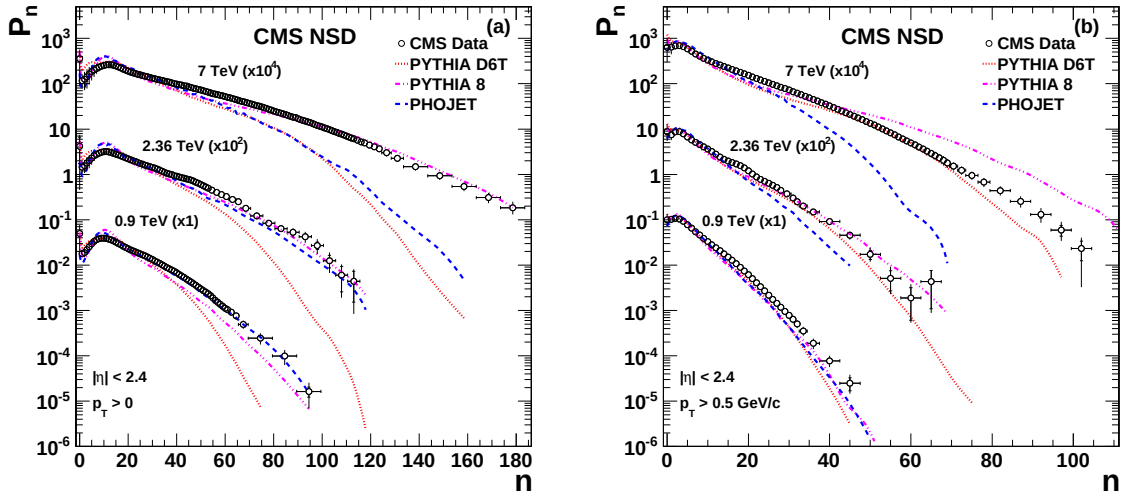

Figure 5.23: The charged hadron multiplicity distributions with $|\eta| < 2.4$ for (a) $p_{\mathrm{T}} > 0$ and (b) $p_{\mathrm{T}} > 500\,\mathrm{MeV}/c$ at $\sqrt{s} = 0.9$, 2.36, and 7 TeV, compared to two different PYTHIA models and the PHOJET model. For clarity, results for different center-of-mass energies are scaled by powers of 10 as given in the plots.

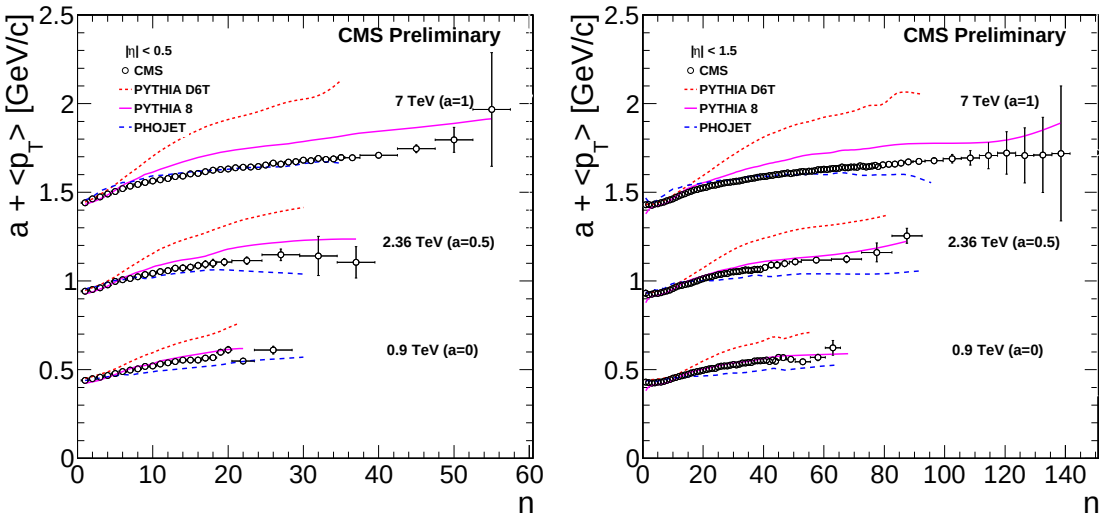

Figure 5.24: A comparison between our measurement of the average transverse momentum versus the charge multiplicity for $|\eta| < 2.4$ with several available PYTHIA tunes and the PHOJET model at $\sqrt{s}$=0.9, 2.36 and 7.0 TeV.

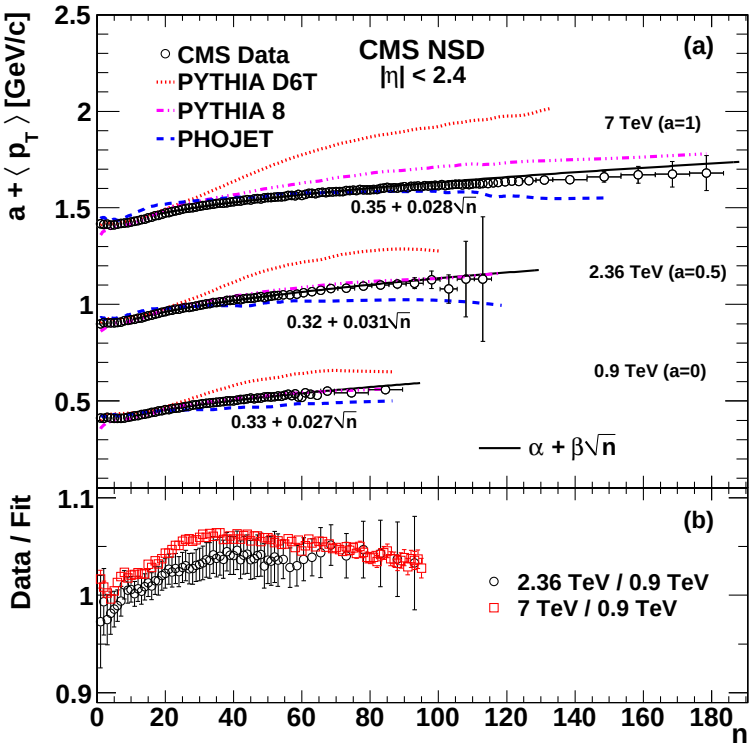

Figure 5.25: A comparison of $\langle p_{\mathrm{T}} \rangle$ versus n for $|\eta| < 2.4$ with two different PYTHIA models and the PHOJET model at $\sqrt{s} = 0.9$, 2.36, and 7 TeV. For clarity, results for different energies are shifted by the values of a shown in the plots. Fits to the high-multiplicity part ($n > 15$) with a linear form in $\sqrt{n}$ are superimposed. (b) The ratios of the higher-energy data to the fit at 0.9 TeV indicate the approximate energy independence of $\langle p_{\mathrm{T}} \rangle$ at fixed n.

5.8.4 Energy dependence of the mean multiplicity

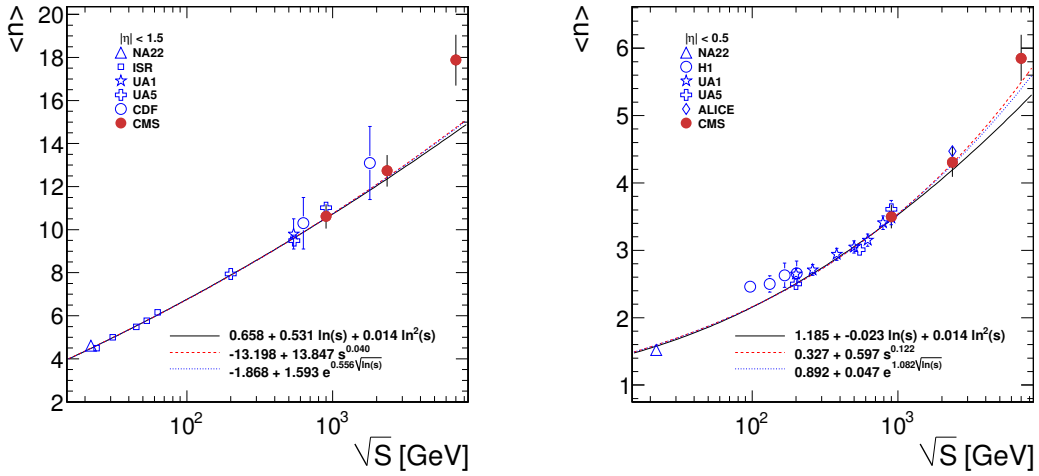

Figure 5.26: The evolution of the mean charge multiplicity with the center of mass energy for pseudo-rapidity acceptances of $|\eta| < 1.5$ (left) and $|\eta| < 0.5$ (right). Data from lower energy experiments have been added where available. [47, 50, 93, 94, 105, 106]. The data are compared with predictions from three analytical Regge-inspired models [69–71]

The mean multiplicity $\langle n \rangle$ is the first moment of the multiplicity distribution and is equal to the integral of the corresponding single-particle inclusive density in the η interval considered. The mean multiplicity $\langle n \rangle$ is observed to rise with increasing center-of-mass energy in hadron-hadron collisions [44–46, 49, 50, 65, 93, 105]. The same behavior is also observed in e^+e^- collisions, in deep-inelastic scattering [106], and in heavy ion collisions [53].

Our measured mean multiplicity is compared with experimental data obtained at lower energies [47,50,93,94,105] in Figures 5.27 and 5.26. Deep inelastic scattering data obtained by the H1 experiment at HERA in the pseudorapidity range $1 < \eta^* < 2$ [106] was added to the $|\eta| < 0.5$ bin and agrees well with other data. Recent Regge-inspired models [69–71] predict a power-like behavior among which only Reference [70] describes the highest energy data very well. Parton saturation models (such as [72]) predict a strong rise of the central rapidity plateau as well. Table 5.4 gives an overview of $\langle n \rangle$ for the data and for the PYTHIA D6T, PYTHIA 8, and PHOJET event generators. As in Section 5.8.3, PYTHIA D6T produces on average too few particles per event at all three energies. PHOJET is consistent with the data within uncertainties for $\sqrt{s} = 0.9\,\text{TeV}$, but is not able to predict properly the mean multiplicity at higher energies. PYTHIA 8 describes best the 7 TeV data, but underestimates $\langle n \rangle$ systematically at all energies.

Table 5.4: Mean multiplicity for data, PYTHIA D6T, PYTHIA 8, and PHOJET for $|\eta| < 2.4$ at each center-of-mass energy. For data, the quoted uncertainties are first statistical, then upward and downward systematic.

$\sqrt{s}$ (TeV)	$\langle n \rangle$			
	Data	PYTHIA D6T	PYTHIA 8	PHOJET
0.9	$17.9 \pm 0.1^{+1.1}_{-1.1}$	14.7	14.9	17.1
2.36	$22.9 \pm 0.5^{+1.6}_{-1.5}$	16.7	17.8	18.7
7	$30.4 \pm 0.2^{+2.2}_{-2.0}$	21.2	25.8	23.2

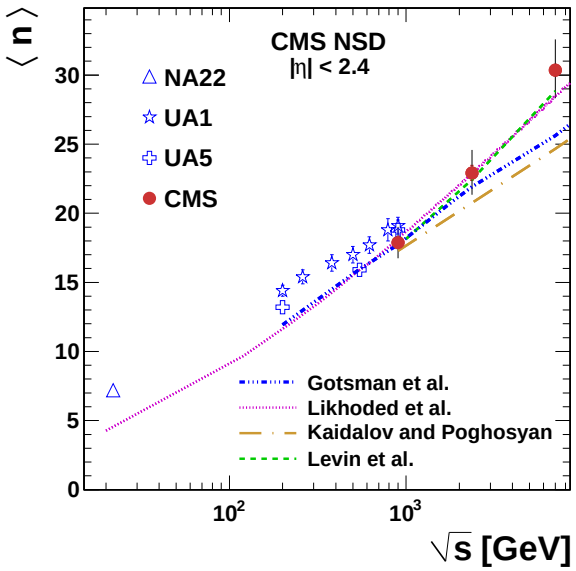

Figure 5.27: The evolution of the mean charge multiplicity with the center-of-mass energy for $|\eta| < 2.4$, including data from lower-energy experiments for $|\eta| < 2.5$ [47,50,93,94]. The data are compared with predictions from three analytical Regge-inspired models [69–71] and from a saturation model [72].

5.8.5 Energy evolution of the moments and cumulants of the multiplicity distribution

Alternatively, the shape of the multiplicity distribution can be analyzed in terms of its higher order moments. A complete review is given in [1, 53]. A summary of the mean $\langle n \rangle$, dispersion D_2, C-moments C_q, normalized Factorial moments F_q and Cumulants K_q (and their ratio H_q) of the multiplicity distribution is shown in Table B.1 to B.3 for three pseudorapidity domains.

The normalized moment C_q of order q is closely linked to the scaling properties of the multiplicity distribution: it should be independent of $\sqrt{s}$ if KNO scaling holds. As demonstrated previously in Figure 5.22, they increase linearly with the center-of-mass energy for $|\eta| < 2.4$, confirming that the KNO scaling does not hold, while on the contrary they are independent of $\sqrt{s}$ in the smallest pseudorapidity interval. The breaking of the scaling is largely attributed to the increase of MPI at 7 TeV, along with the cross-section

increase of other subprocesses. In the more central events, long-range correlations are less likely to be present inside the small acceptance window, and the C_q moments are more stable when increasing the energy compared to $|\eta| < 2.4$ where they are included. The long-range correlations hence seem to play a bigger role at high-energy.

The normalized factorial moments express the all inclusive correlations, with their evolution with $\log(\sqrt{s})$ shown in Figure 5.28 for extreme pseudorapidity intervals, along with previous UA_5 measurements when available. The moments increase steeply with the center-of-mass energy, which again hints to the growth of correlations at 7 TeV. The cumulants, which reflect only the genuine correlations, show a much more tamed increase with $\sqrt{s}$, even for higher orders. This is to be expected, as higher order genuine correlations are much less likely to occur. Figure 5.29 (left) presents the K_q moments for $|\eta| < 2.4$. Finally, the F_q and K_q moments are often described with their ratio H_q, to damper their sharp increase, especially when computing high q-orders. Their evolution with energy is represented in Figure 5.29 (right), although it is more common to plot them in function of their order, as it has been demonstrated that the H_q decrease with increasing q, down to zero around which they then oscillate [1]. Figure 5.30 shows such decrease, albeit up to $q = 5$ only.

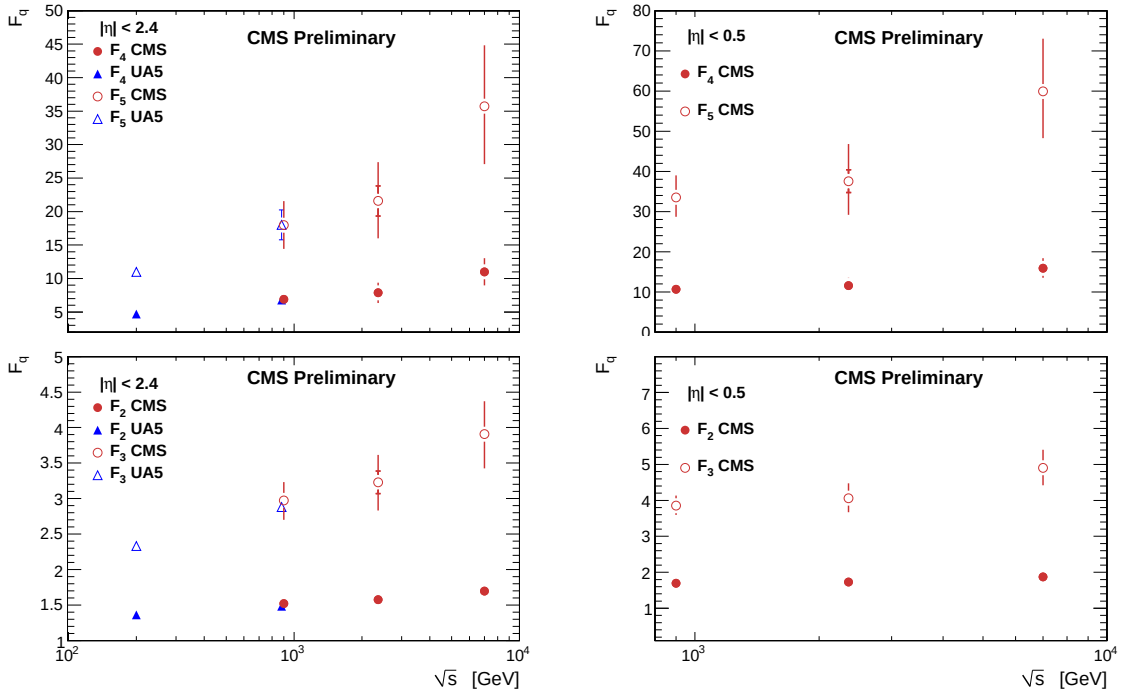

Figure 5.28: $\log s$ dependence of the normalized factorial moments F_q of the multiplicity distribution for $|\eta| < 2.4$ (left) and $|\eta| < 0.5$ (right), including data from lower energy experiments [47, 94]. For $\sqrt{s} = 0.9$ TeV, data from experiments other than CMS were drawn shifted to lower $\sqrt{s}$ for clarity.

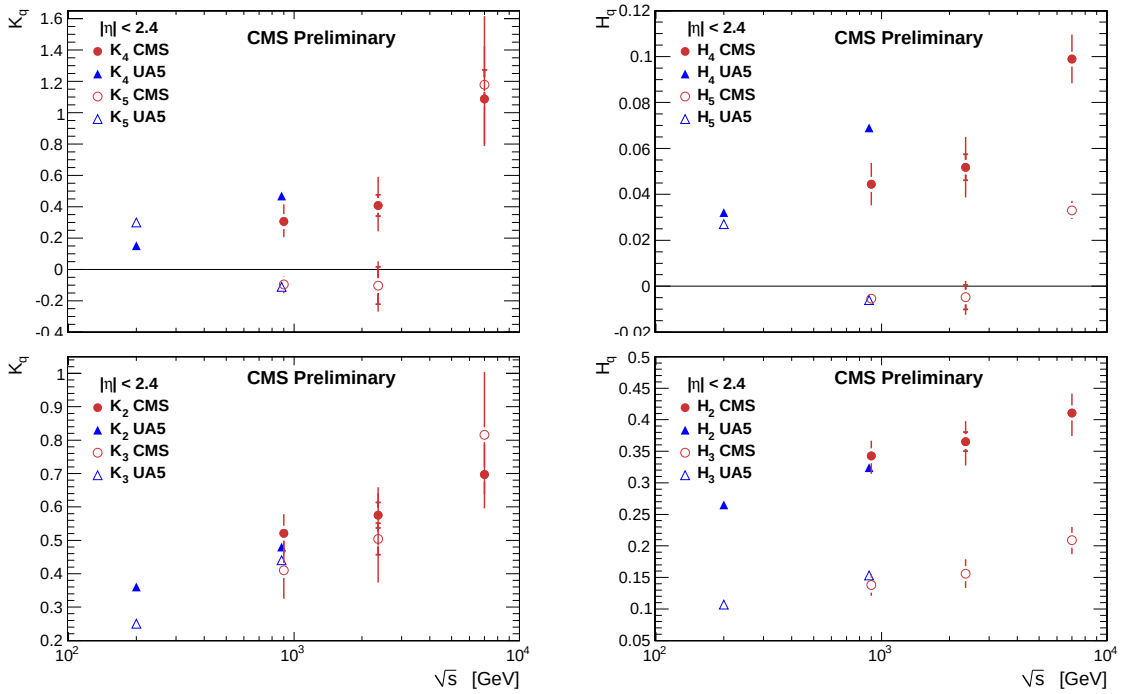

Figure 5.29: $\log s$ dependence of the normalized cumulants K_q (left) and the moments H_q (right) of the multiplicity distribution for $|\eta| < 2.4$, including data from lower energy experiments [47, 94]. For $\sqrt{s} = 0.9\,\text{TeV}$, data from experiments other than CMS were drawn shifted to lower $\sqrt{s}$ for clarity.

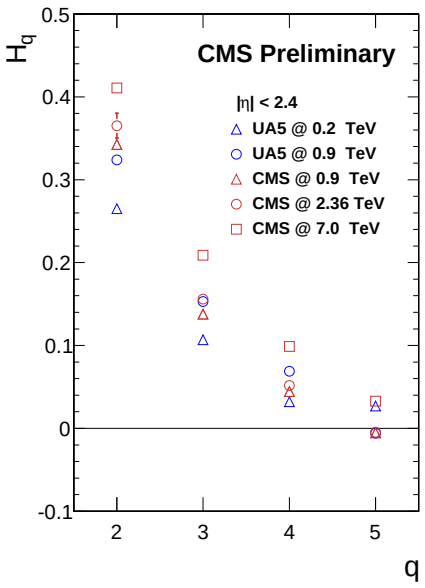

Figure 5.30: Evolution of the H_q moments of the multiplicity distribution with their q-order for $|\eta| < 2.4$, including data from lower energy experiments [47, 94].

5.9 Conclusion

The charged particle multiplicity distributions of non-single diffractive events were measured by the analysis of the minimum bias datasets collected by CMS at three center of mass energies: $\sqrt{s}$=0 .9, 2.36 and 7.0 TeV. The excellent tracking capabilities of the silicon pixel and strip detector of CMS, combined with an optimized tracking and vertexing algorithm, allow the reconstruction of charged tracks down to p_{T} =100 MeV/c with a high efficiency and low background contamination. A full correction of detector resolution and acceptance effects and an extrapolation to zero transverse momentum, allows a measurement of the charged particle multiplicity distribution in increasing pseudorapidity domains from $|\eta| < 0.5$ to $|\eta| < 2.4$ that can be compared with Monte-Carlo models for soft particle production and with experimental data at lower energies. At present, none of the PYTHIA tunes that provided an adequate description of the Tevatron and LEP data is able to describe the multiplicity distributions at higher energies. Also PHOJET has the same shortcomings. The remarkable large multiplicity tail observed at $\sqrt{s}$=7.0 TeV indicates a very strong violation of KNO scaling w.r.t. lower energies and merits deeper studies. The rise of the mean charged particle multiplicity with increasing center of mass energy is also steeper than extrapolations of fits at lower energies imply. Finally, the multiplicity distribution can be described by its moments, and they show a clear increase in correlations at 7 TeV.

Chapter 6

Multiplicity analysis with Inelastic events

6.1 Introduction

This analysis is a complement of the previous CMS multiplicity measurement performed for NSD events at center of mass energies of 0.9, 2.36 and 7 TeV. New data has been collected with much more recorded collisions at 0.9 and 7 TeV, while LHC provided a new energy point of 2.76 TeV to study. In order to be less model-dependent when correcting for diffractive events with the Monte-Carlo generators, requirements of central tracks with defined η and p_{T} acceptances was preferred to the non-single-diffractive event selection, allowing comparison to the ATLAS and ALICE experiments, as well as providing distributions which will be useful for tuning Monte-Carlos. As data at 7 TeV has been already published, new Monte-Carlo tunes are used here, and are expected to perform much better than pre-LHC tunes. The multiplicity distributions were obtained, as well as their moments and their KNO form. The evolution of the mean transverse momentum with the multiplicity was also measured.

The first section details the data sample and event selection used, followed by a section explaining how the track reconstruction works. The correction procedure is then reported, as well as the cross-checks and systematics that were performed. Finally, the results are presented and discussed.

6.2 Data Sample

The LHC delivered, during the year 2010, a few millions of low Pile-Up (PU) collisions to the experiments at 0.9 and 7 TeV, in-between the normal running period at higher luminosity. After the Pb-Pb collisions, a few runs of proton-proton collision at the same center of mass energy of 2.76 TeV were also recorded, to serve as baseline in the heavy ion analysis. The following reprocessing of those runs were used, in conjunction with the appropriate json file to select luminosities where the detector was in good shape:

```
/MinimumBias/Commissioning10-398patch2-v1/RECO
/MinimumBias/Commissioning10-398patch2_900GeV-v1/RECO
/ZeroBias/Commissioning10-Dec22ReReco_from_V4_v1/RECO
/AllPhysics2760/Run2011A-PromptReco-v2/RECO
```

For each center of mass energy, Monte-Carlo samples were produced by CMS, or privately when not available. PYTHIA 6 tune Z2 and PYTHIA 8 tune 4C were chosen for there relatively close predictions compared with data, as well as there very different internal descriptions of physics. As some samples were processed before data taking, the z position of their beamspot is a little different than observed in data, and therefore was re-weighted to mimic perfectly the data.

The trigger and offline event selection is nearly identical to other minimum bias event analysis published by CMS [107, 108]. As the 0.9, 2.76 and 7 TeV datasets were acquired during different periods, the minimum-bias events were recorded using different sets of triggers, combining a Level 1 trigger with an HLT one. For 900 GeV and 7 TeV, events were selected by a trigger signal in any of the forward or backward BSC scintillator counters, in coincidence with any of the two BPTX detectors indicating the presence of at least one proton bunch crossing from the IP. This corresponds to Technical trigger bits 0 and 34 at 0.9 TeV, which was later replaced by a physics trigger with bit 124 at 7 TeV. On HLT level, the `HLT_L1_BscMinBiasOR_BptxPlusORMinus` path was used. For the 2.76 TeV data taking, the Physics bit 124 was replaced by the 126, and the HLT path changed to `HLT_L1BscMinBiasORBptxPlusANDMinus_v1`.

Beam halo backgrounds are reduced by using the timing information from the BCS counters at opposite ends of the CMS detector. This is realized by a veto on technical trigger bits 36-39.

Beam induced backgrounds are further removed by requiring the cluster sizes in the pixel detector to be compatible with a single primary vertex, as is described in [35]. This selection is needed for minimum-bias tracking, as its computation degrades when far too many tracks are present in the event, which happens only in certain beam induced backgrounds events.

Events with only one vertex within 15 cm around the position of the nominal interaction point along the beamline were selected, in order to minimize the effect of Pile-Up. This does not impact significantly our statistic, as all the data analyzed here have low PU, thus less than 7% of our events actually have more than one vertex.

A total of 837k events were retained for final analysis at $\sqrt{s} = 0.9$ TeV, 24.2M at $\sqrt{s} = 2.76$ TeV and 14.6M at $\sqrt{s} = 7.0$ TeV.

Contrary to the previous result on multiplicity published by CMS, a selection independent on the MC definition of diffraction was preferred, in order to ease comparison between MCs with very different diffraction implementation. Those selection require a certain number of tracks in the central region, with a specific pt and within a defined η range. Therefore all results are corrected up to to this central selection, which can be easily defined at particle level. A summary of the different selection used in this paper can be seen in Table 6.1.

6.3 Track reconstruction and acceptance definition

Both the Pixel and SST detectors were used in the reconstruction of tracks within an acceptance of $|\eta| < 2.5$ and contain information both from the barrel and the end-cap sensors. The standard track reconstruction algorithm of CMS is based on a combinatorial track finder that performs multiple iterations, and is usually referred to as general tracking. This iterative procedure was further optimized for primary track reconstruction with transverse momenta down to 100 MeV/c in minimum bias events [35, 76], and was

Table 6.1: Different central requirements applied in the analysis. Those cuts are related to the event selection only, and not necessarily to the multiplicity acceptance.

# tracks min	$\|\eta\|$ max	p_{T} min [GeV/c]
1	0.8	0.5
1	1	0
1	2.4	0.5
2	2.4	0.1
6	2.4	0.5

used for the first CMS multiplicity paper. In order to increase the overall efficiency in this analysis, tracks from both trackings were merged into a single collection, enabling tracks not reconstructed by one of the two algorithms to be recovered. This mixed tracking has already been used in a previous analysis [107].

As with the NSD analysis, all reconstructed tracks are required to be associated to the primary vertex, in order to reduce the contamination by V^0 decay products and secondaries produced in interactions of charged particles with the material of the detector. Although the observables used to perform this association are identical to the first analysis ($\mathrm{d}p_{\mathrm{T}}/p_{\mathrm{T}}$, $\mathrm{d}_{xy}/\sigma_{xy}$ and $\mathrm{d}_{xy}/\sigma_{xy}$), the selection cuts applied are different and are the result of common discussions throughout the QCD analysis, to ensure uniformity of the objects used (here tracks) and thus a very deep knowledge of their characteristics. This is only the mere consequence of having more analysis using the low-p_{T} tracks and having gained experience with their usage. The association with the primary vertex is hereby commonly done by selecting tracks with a relative error on their measured transverse momentum smaller than 5% and by demanding that the impact parameter with respect to the primary vertex position both in the transverse plane and along the z-axis be smaller than 5 times its measured error:

- $\mathrm{d}p_{\mathrm{T}}/p_{\mathrm{T}} < 0.05$
- $\mathrm{d}_{xy}/\sigma_{xy} < 5$
- $\mathrm{d}_{z}/\sigma_{z} < 5$

A comparison between data and MC of the relative momentum error and of both impact parameter significances is given in Figures 6.1 to 6.3.After application of the aforementioned track quality and vertex association cuts, an overall good agreement between data and MC simulation is found in transverse momentum, pseudorapidity, number of tracker hits used in the track fit and the χ^2 of the track fit, as is shown in Figures 6.4- 6.6. Note that none of the MC models used is describing the pseudorapidity and the lower end of the transverse momentum distribution very well. This is not due to an inadequate simulation of the detector response, but rather a feature of the MC modeling of minimum bias data in the event generators. Reference [35] reported the same observations, implying the need for retuning of the MC generators.

The *raw* multiplicity spectra are obtained by counting all charged particles that are accepted by the track selection cuts and fall completely within the acceptance of the tracking detector, i.e. with $|\eta| < 2.4$. These spectra, compared with MC models are shown in Figure 6.7 at the three center of mass energies. They will serve as input for the unfolding and correction procedures described below.

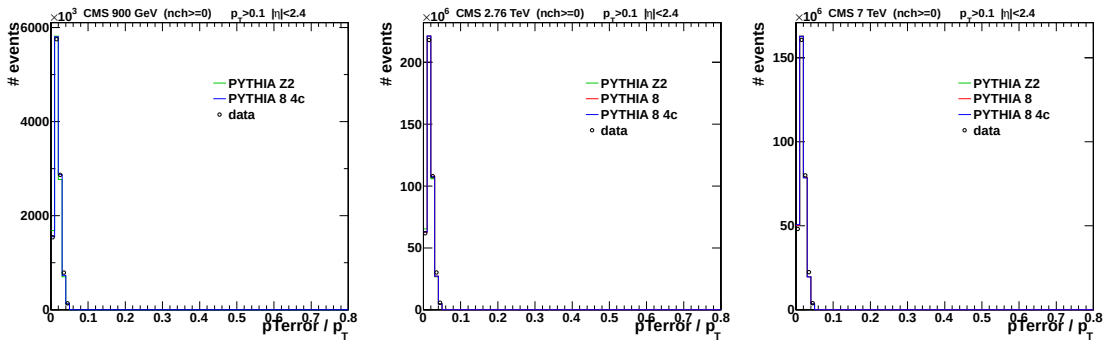

Figure 6.1: The relative error on the transverse momentum measurement at each center-of-mass energy.

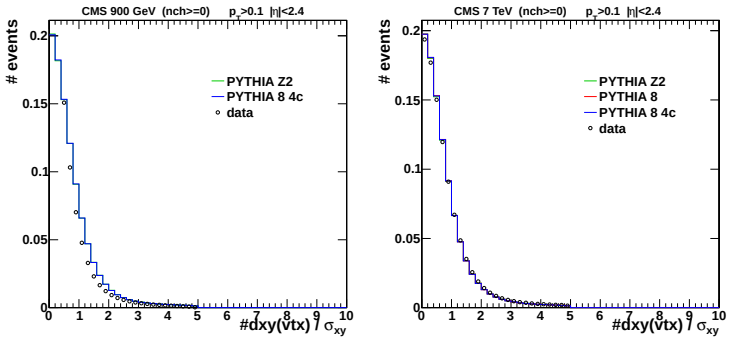

Figure 6.2: The impact parameter significance in the plane perpendicular to the beam pipe at two center-of-mass energy.

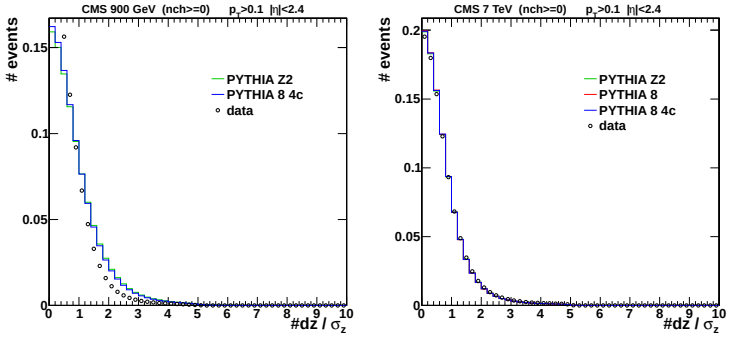

Figure 6.3: The impact parameter significance along the direction of the beam pipe at two center of mass energy.

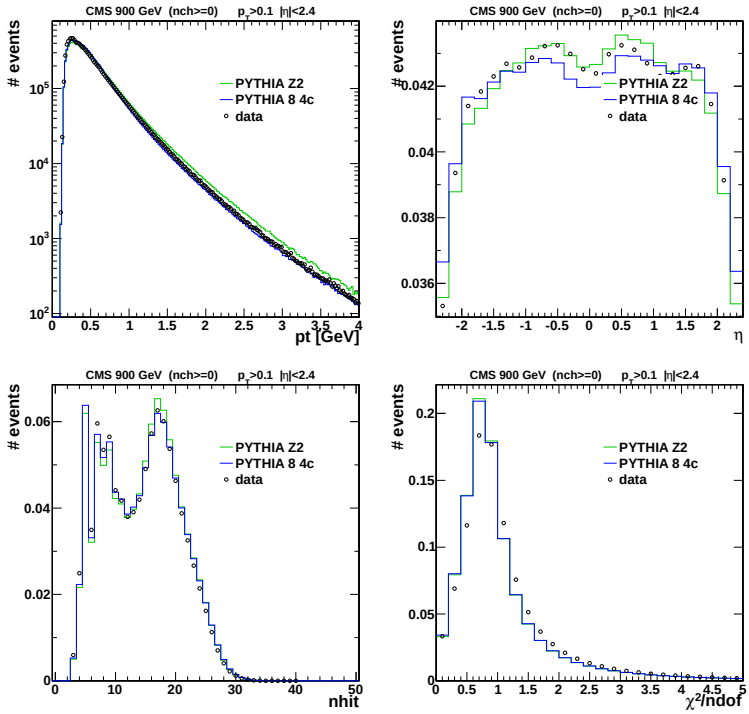

Figure 6.4: Comparison between data and MC at $\sqrt{s}$=0.9 TeV of the transverse momentum, pseudorapidity, number of track hits and the goodness of fit.

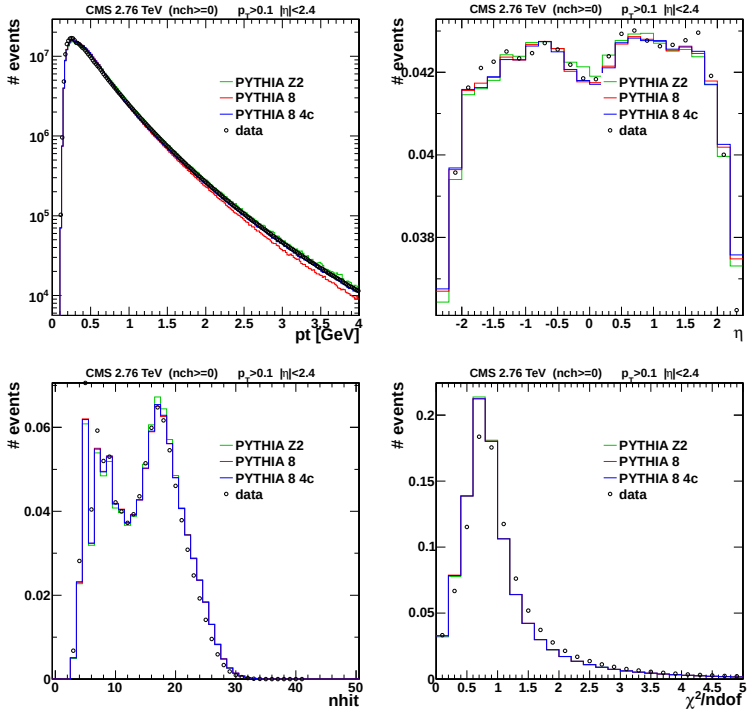

Figure 6.5: Comparison between data and MC at $\sqrt{s}$=2.76 TeV of the transverse momentum, pseudorapidity, number of track hits and the goodness of fit.

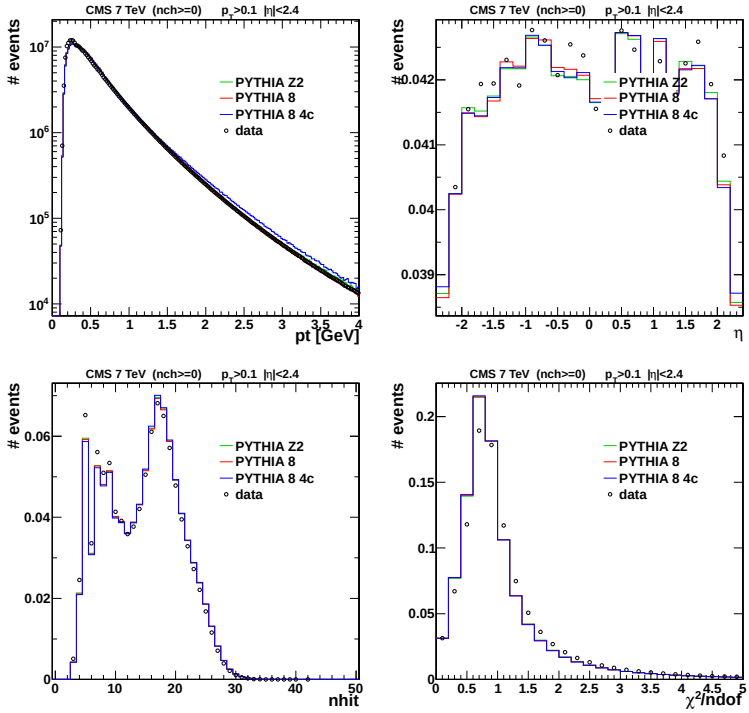

Figure 6.6: Comparison between data and MC at $\sqrt{s}$=7.0 TeV of the transverse momentum, pseudorapidity, number of track hits and the goodness of fit.

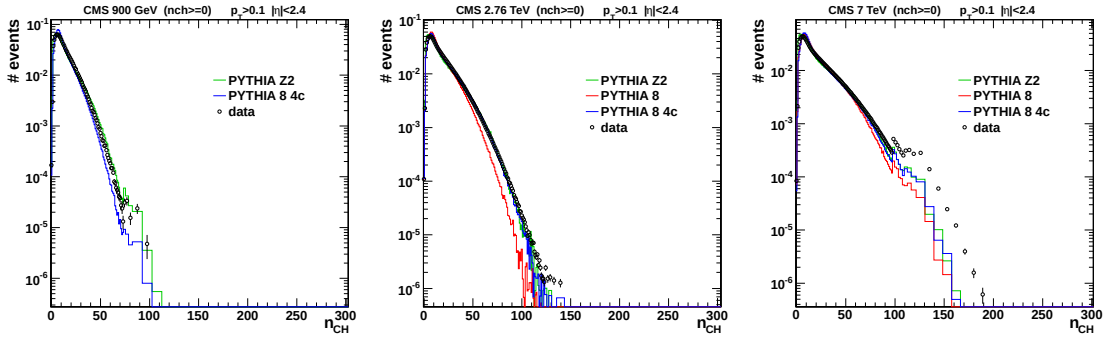

Figure 6.7: Comparison between data and MC models of the raw multiplicity distribution at $\sqrt{s}$= 0.9, 2.76 and 7.0 TeV. They have not been normalized by their binwidth as they serve as input to the correction.

6.4 Corrections

The minimum bias trigger will unavoidably introduce a bias in the measured charged particle multiplicity, as shown in Section 6.2. To correct for it, the trigger efficiency with central selection and vertex requirement was obtained using zerobias data for 0.9 and 7 TeV, while Monte-Carlo was used at 2.76 TeV due to lack of zerobias data at this energy.

A fraction of events will also be removed by offline requirements on the presence of a single good quality primary vertex. The efficiency of finding a single vertex in Monte-Carlo was computed, and is shown in Figure 6.8 for different central selections. Its decrease with multiplicity is easily explained by the increasing probability of reconstructing a second vertex when having more tracks, cases that we currently reject in our event selection.

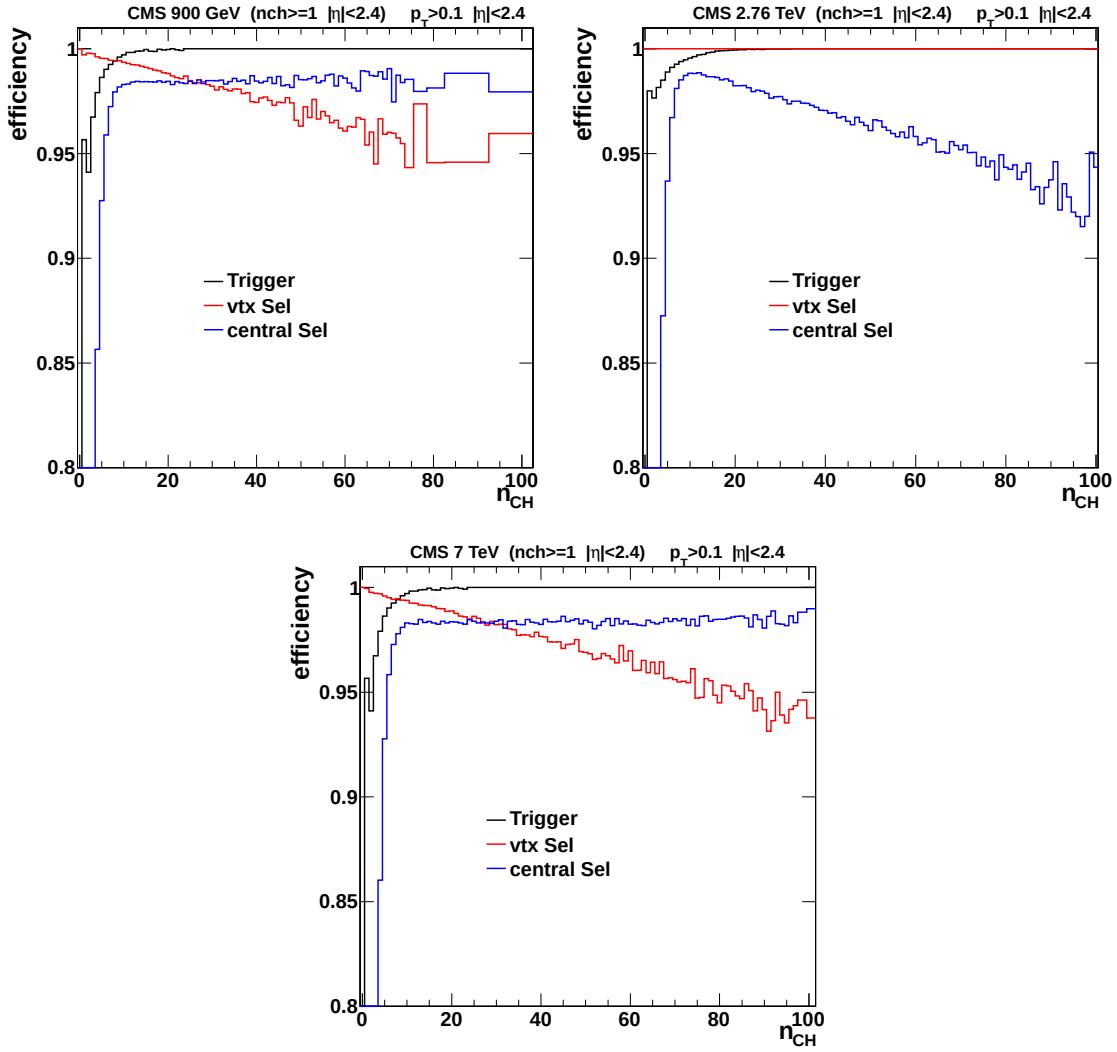

Figure 6.8: The efficiency of the trigger correction, the vertex requirement correction, and the central selection for the multiplicity distribution at each center of mass energy.

Due to inefficiencies in track reconstruction and acceptance, the creation of secondary particles by the interaction of primaries with the beam pipe and the detector material and the presence of decay products of long lived hadrons, one will in general not measure

the true accepted multiplicity, but a derived quantity. In order to recover the true multiplicity, a Bayesian unfolding [56] was used, as for the previous NSD measurement. The unfolding matrix, $P_{m,n}$, is shown as an example in Figure 6.9 for events with $|\eta| < 2.4$. As in the previous multiplicity paper, the number of iterations were tailored to each input distribution based on the χ^2 computation, while the statistical errors were obtained with resampling of both the input distribution and the unfolding matrix.

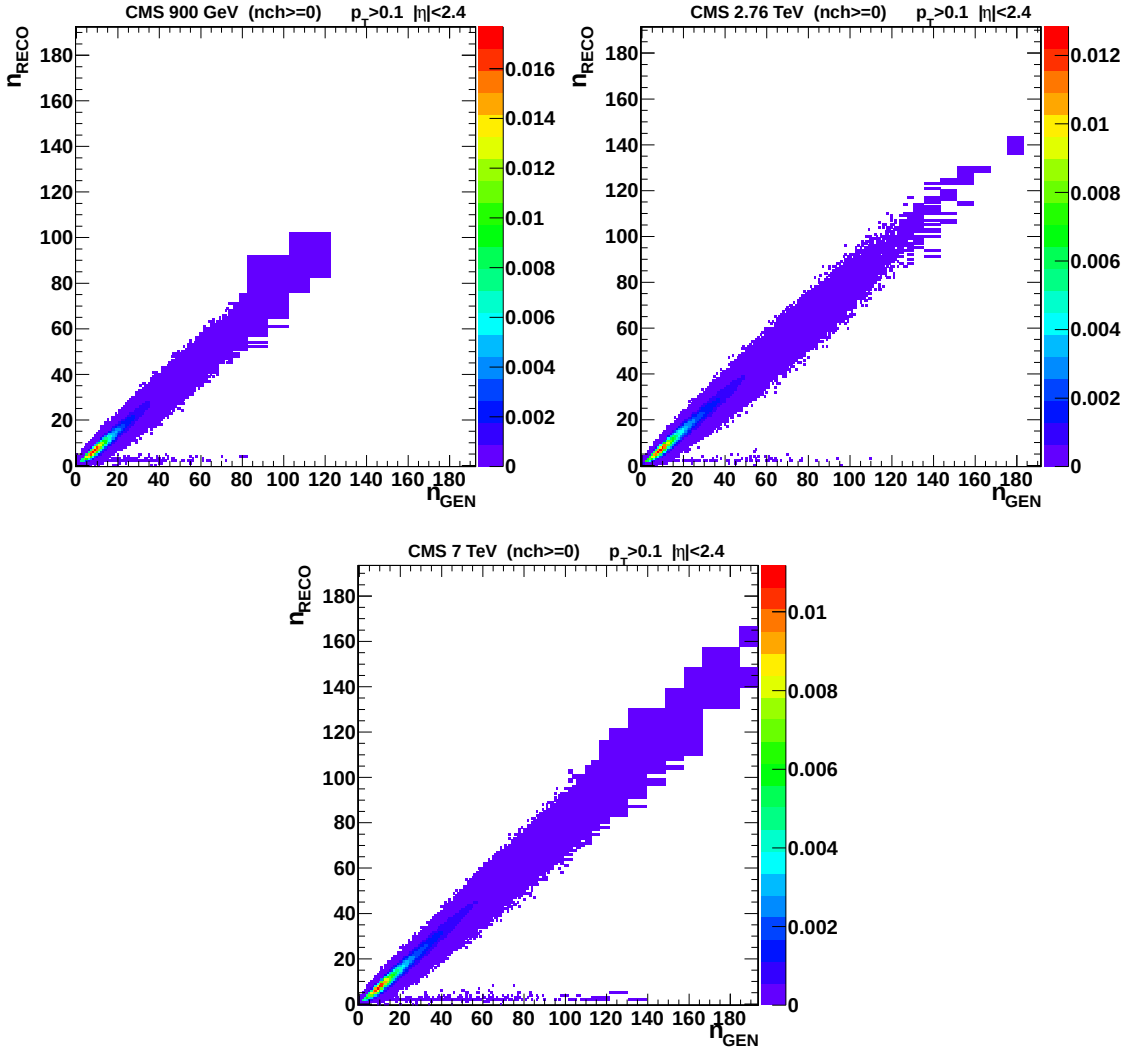

Figure 6.9: Matrices used for unfolding at different center of mass energies.

The average transverse momentum, $< p_{\mathrm{T}} >$ in each multiplicity bin was measured as well. The trigger bias was corrected for using the efficiency measurements as seen in Figure 6.10. For each bin in raw multiplicity, a Monte Carlo based correction factor $< p_{\mathrm{T}}^{gen} > / < p_{\mathrm{T}}^{rec} >$ was then applied to convert the measured $< p_{\mathrm{T}} >$ to the corresponding value for primary charged hadrons. Subsequently, the unfolding matrix $P_{m,n}$, was applied to correct the multiplicities and repopulate the $< p_{\mathrm{T}} >$ values.

The correctness of the unfolding procedure was verified using MC events where we compare the generated hadron multiplicity with the unfolded reconstructed tracks. At all energies, a perfect agreement with the generated hadrons is achieved after unfolding. The effect of these corrections is shown in Figure 6.11 for the p_{T} versus n profile.

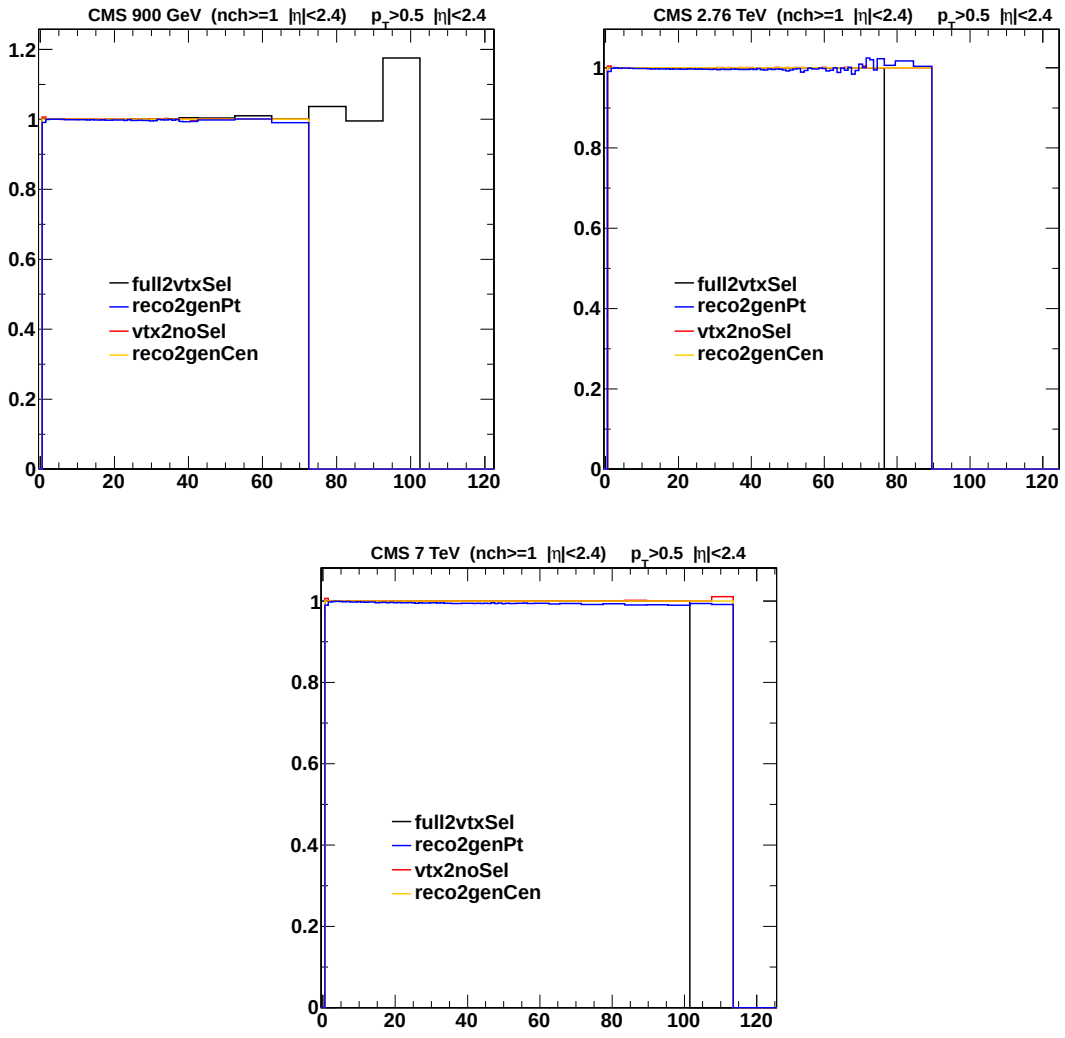

Figure 6.10: The efficiency of the trigger correction, the vertex requirement correction, the pt correction and the central selection for the mean p_{T} versus multiplicity distribution at each center of mass energy.

To correct for the bias introduced by applying the central selection on tracks, where inefficiencies and fakes can occur, an efficiency factor derived from Monte-Carlo was used. This factor was computed making the efficiency in function of the generated multiplicity to pass the central selection requirement on reconstructed tracks w.r.t. on generated particles. It results in a small correction, shown in Figure 6.8.

Finally, a correction was applied for central event selections including tracks with $p_{\mathrm{T}} < 0.5\,\mathrm{GeV}$, in order to compensate for the fact data has more tracks at low-pt than predicted by the MC generators. Monte-Carlo generators produces the same p_{T} distribution above 500 MeV, but there are more tracks under 500 MeV in data than in Monte-Carlo. The detector inefficiencies on those tracks are corrected by the unfolding procedure, but it assumes a p_{T} distribution as in the Monte-Carlo, while in fact their probability is higher at low-p_{T}, where the inefficiencies are higher. To take this into account, the correction factors assuming track efficiencies, fake rates and p_{T} distributions to be independent from the event multiplicity were computed using the p_{T} distribution from data and MC, and

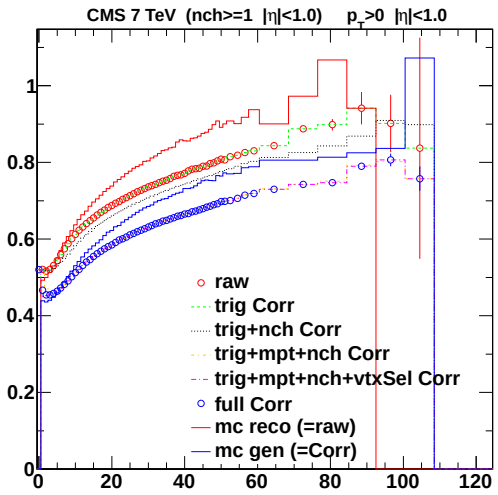

Figure 6.11: The different steps of data correction of the $< p_{\mathrm{T}} >$ vs n profile, correcting reconstructed tracks back to the hadron level.

their difference was taken as an additional correction on the multiplicity distribution:

$$N_{corr} = [\sum_{p_{\mathrm{T}}=0}^{\infty} \epsilon_{Tr}(p_{\mathrm{T}}) \cdot P(p_{\mathrm{T}})] \cdot N_{Observed}, \tag{6.1}$$

where ϵ_{Tr} is the p_{T}-dependent efficiency and P(p_{T}) is the p_{T}-dependent probability for a track to have a given p_{T}. The difference of this correction factor was computed for 0.9, 2.76 and 7 TeV using their own p_{T} distributions, and was found to be between 4% and 5.5%, depending on the energy and the Monte-Carlo tune used. When starting the sum for tracks with p_{T} above 500 MeV, the difference was smaller than 0.1%. When applied to the mean p_{T} versus multiplicity distribution, the correction factor variated from -3.8% to 1.4%.

6.5 Cross-Checks and Systematic uncertainties

All plots presented here represent on the y-axis the number of σ between the standard distribution and the cross-check/systematic one, ie each bin was calculated with: σ = (standard - cross-check) / standard. The corresponding statistical errors of the standard distribution divided by its bin content are represented in red bars, the systematic being in shaded grey.

6.5.1 Cross-Checks

We performed several cross-checks, to check the effect on the final results of some of the choices we made. As those were deemed negligible, they were not included in the computation of the systematic errors.

Vertex re-weighting:

As detailed in Section 2, since the Monte-Carlo produced did not have the exact same distribution of the vertex z position as in the data, each MC event were re-weighted accordingly to get an identical distribution to data. The effect after all corrections with and without applying those weights in the Monte-Carlo implies that even with a wrong vertex-z distribution, the outcome of the Monte-Carlo corrections would have been the same.

Number of iterations:

The number of iterations to use during the unfolding is not mathematically defined, and thus is up to the scientist to chose exactly when to stop the iterative procedure. Although the χ^2 estimation explained in Section 4 is a very good way to define the number of iterations to use, we decided to check what happens when adding two more iterations to the number used for the final results, which is shown in Figure C.4. The small wavy structure clearly indicates that adding more iterations would only start revealing the effect of the unfolding on small statistical variations in the multiplicity distributions, and are not related to any physics, thus strongly confirming the good choice of the number of iterations.

Tracking type:

In order to look at the bias that could be introduced using different tracking algorithms, we compared the corrected result obtained with the mixed tracking, with the one using general tracking. Figure C.5 shows that the effect for tracks above 500 MeV is negligible, while including smaller transverse momentum tracks demonstrates the improvements of mixing the tracks from minbias tracking and general tracking over the simple general tracks, which are specifically designed for high momentum particles.

Pile-Up:

Since the event selection we chose requires only one event in the vertex, the effect of the Pile-Up rests mainly in cases where two collisions happened in the event, but due to vertex mis-reconstruction, there is only one vertex correctly reconstructed. In that case, tracks belonging to the unreconstructed vertex can eventually be associated to the other vertex, and thus bias to higher multiplicity our analysis. To measure this effect, we looked at the 2.76 TeV data sample, were the Pile-up is the highest, therefore containing the higher number of events with more than one vertex, and selected events with two well reconstructed vertices, also applying our standard trigger selection. For each events, we assumed each tracks originated to its closest vertex, and then applied for each vertices the association cuts detailed in Section 6.3 on the tracks belonging to the other vertex. This is summarized in Figure 6.12, presenting in function of the vertex multiplicity, the mean number of tracks belonging to the other vertex but still associated to the studied vertex.

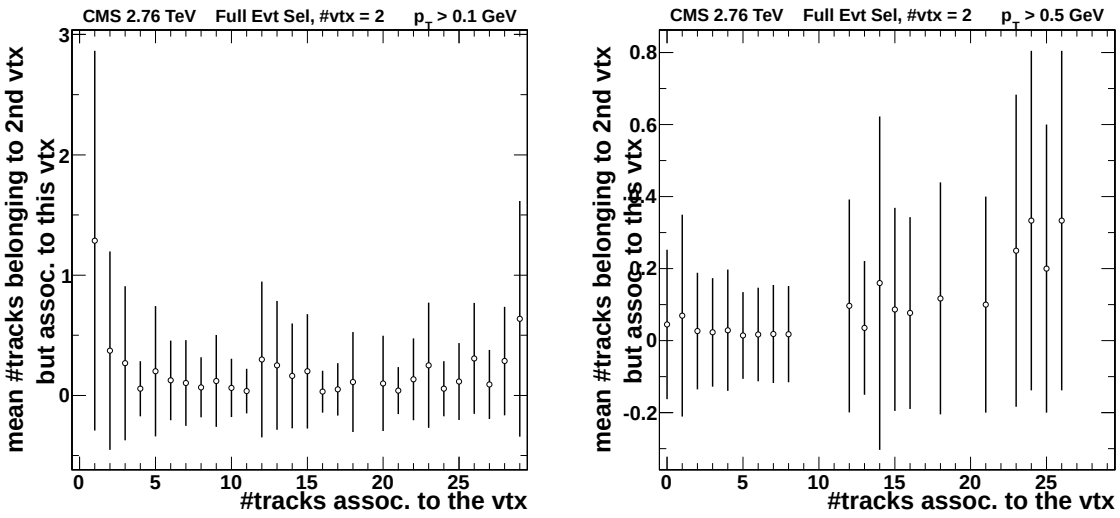

Figure 6.12: Mean number of tracks belonging to the second vertex, but that passed association cuts to the studied vertex, in function of its multiplicity, for fully selected events with exactly 2 reconstructed vertices, at 2.76TeV. Left: tracks were selected with $p_{\mathrm{T}} > 0.1\,\mathrm{GeV}$; Right: tracks were selected with $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$

6.5.2 Systematic uncertainties

Trigger systematics:

The correction of the trigger is using 0-bias data, and is limited by the statistic of the collected data sample. In order to take this into account, the trigger efficiency used during the correction of the multiplicity distribution was decreased/increased of its error for each bin in multiplicity, making sure it could not be more than 1. The systematic effect on the final result is fairly small, and can be viewed in Figure C.6.

Tracking inefficiencies in MC:

A correct description of the tracking efficiency in the Monte Carlo simulation of the detector is essential in obtaining a correct unfolding matrix. At low transverse momenta, the efficiency drops due to a loss of hits on the tracks that are effectively stopped within the tracking volume. A conservative estimate of the uncertainty of this efficiency of 2% is taken from [35]. The fake tracks are mostly secondaries originating from interactions with the material of the LHC beam pipe around the interaction point. This was estimated in [35] to be no more than 1%. These uncertainties, together with smaller contributions from the misalignment of the tracking detector and the extrapolation uncertainty from p_{T} =0.1 GeV/c to zero add up in quadrature to a total tracking uncertainty of 2.6%. As was discussed in Section 6.3, the contamination from V^0 decays was negligible after associating tracks with the primary vertex. The effect on the multiplicity distribution, shown in Figure C.7 was estimated by decreasing and increasing the corrected multiplicity measurements with 2.6%. The systematic uncertainty becomes very small at a pivot point around $n = 20$ which is very close to the mean of the original multiplicity distribution. In general the systematic uncertainty remains below 10% for a large part of the multiplicity

distribution. All Correction factors related to trackingreconstruction are summarized in Table 6.2.

Source	Related correction factor 7 TeV (%)
Algorithmic efficiency	1.6
P_T extrapolation	0.2
Pixel hit reconstruction efficiency	1.0
Pixel hit splitting	0.4
Acceptance uncertainty	1.0
Correction of secondary contribution	1.0
Tracklet selection and background subtraction	0.5
Misalignment	1.0
Random hit from beam halo	0.2
Total	2.6

Table 6.2: Summary of systematic uncertainties related to the efficiency and acceptance correction.

Monte-Carlo modeling:

Finally, as none of the current Monte-Carlo are able to describe perfectly the different observables even without any corrections, the effect of using the PYTHIA 8 tune 4C or the PYTHIA Z2 tunes was considered, and revealed small differences when using each Monte-Carlo for the unfolding, the vertex event selection correction, and the central event selection correction. Figure C.8 shows such systematic for different central selections. As no Monte-Carlo can be considered better than the other, the final results for each selections is taken as the mean of the two corrected distribution with both MC.

Total error:

All systematics have been considered independent, and were added in quadrature with the statistical errors, as displayed in Figure 6.13.

Mean p_{T} versus multiplicity systematic:

For the measurement of the mean p_{T} versus multiplicity, only the effect of changing the tracking algorithm (Figure C.9), and the mis-reconstruction of the inefficiencies in the MC (Figure C.10) were cross-checked as the others already have minor effects on the multiplicity, and were found negligible. The only systematic that has a sizable effect is the comparison between two Monte-Carlos, of the order of 1% and increasing when including tracks with $p_{\mathrm{T}} < 0.5\,\mathrm{MeV}$,as shown in Figure C.11.

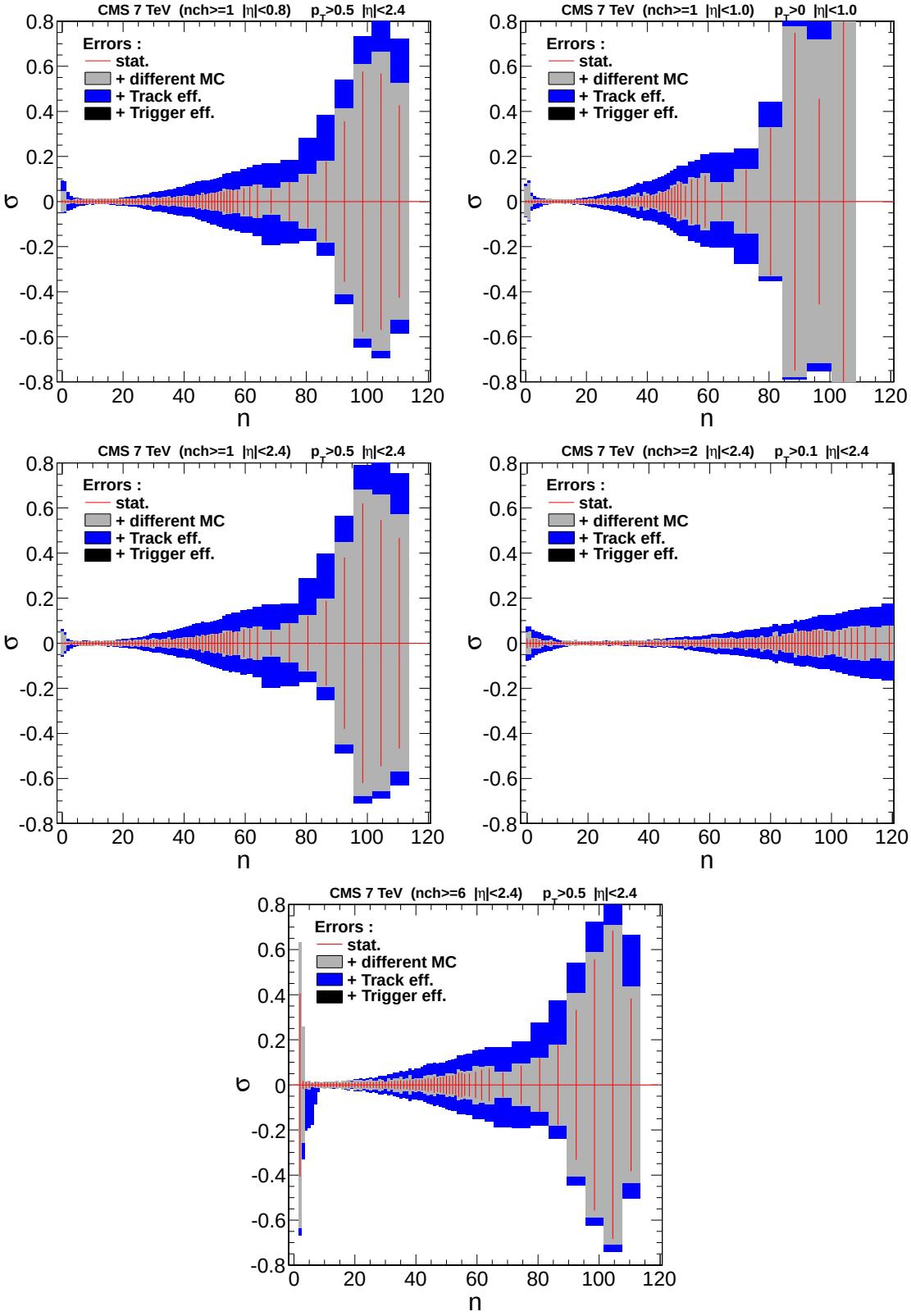

Figure 6.13: All systematic effects added one by one in quadrature to the statistical errors of the fully corrected multiplicity distribution for several central selections and track acceptances.

6.6 Results

6.6.1 Multiplicity distributions, $< p_{\mathrm{T}} >$ VS multiplicity

The inelastic charged hadron multiplicity distributions are measured with different energies, central acceptances, pseudorapidity intervals and transverse momentum requirements. Figure 6.15 shows the selections that can be compared with the ATLAS results [83, 109], while Figure 6.14 shows the central selection akin to be compared to ALICE [82]. The measurements from the experiments are in very good agreement, for all the energies and central selections considered. The high multiplicity tail as well as the shift in the slope observed previously are still present, and are especially visible when looking at the widest acceptances possible. It has to be noted that contrary to the previous results, there is a rather large quantity of diffractive events measured, be it single or double diffraction, as none are particularly favored or excluded by the triggering. Their inclusion does not change the observations made on multiplicity, so multiple-parton-interactions should be independent of such considerations.

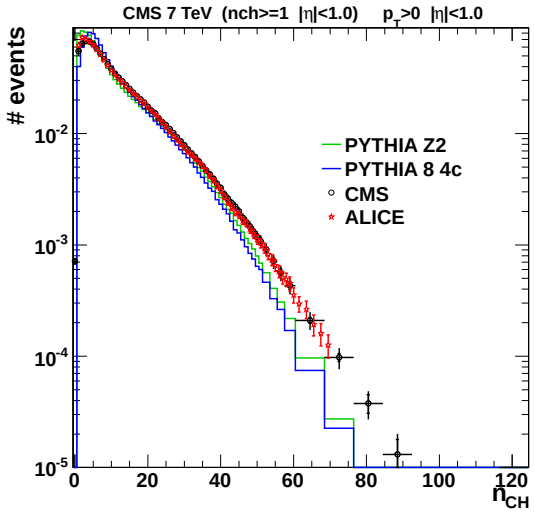

Figure 6.14: The fully corrected multiplicity distribution compared to ALICE results [82]. The generator prediction for PYTHIA 8 tune 4C and PYTHIA 6 tune Z2 were also added for comparison.

Two Monte-Carlo models are included with the comparisons to other experimental data. The PYTHIA Z2 tune is based on the PYTHIA 6.4 event generator, and as such has the same fragmentation model as the tunes D6T and ATLAS used in the NSD measurement. However, it took advantage of the results published by CMS, ATLAS and ALICE, as it was tuned manually using distributions from the Underlying Event analysis [15] published earlier as well as the multiplicity distributions presented in Chapter 5 and also publicly available prior to these new measurements. The other Monte-Carlo model included is based on the newer version of the fragmentation in PYTHIA 8. The tune 4C [110] is also tuned to newly available data, essentially that of ATLAS and ALICE.

Figure 6.16 presents the multiplicity distributions at all three center-of-mass energies, with the corresponding ATLAS and ALICE results, as well as the Monte-Carlo predictions using the two PYTHIA models tuned to new data aforementioned. Both Monte-Carlo tunes predict the same amount of high-multiplicity events at 7 TeV, as they were both specifically

tuned with data at this energy, although they still underestimate the distributions' tails, even more so when decreasing the p_{T} acceptance to 0.1 GeV. As the energy decreases, PYTHIA 8 tune 4C produces less particles per event than the Z2 tune, and though it describes correctly the multiplicity for 0.9 TeV with $p_{\mathrm{T}} > 0.5$ GeV, only Z2 manages to predict the correct low-pt amount of tracks at 0.9 TeV. As expected, the new tuned generators fare much better than their predecessors used for the NSD analysis, and tend to produce more high-multiplicity events than before. This has mainly been achieved by increasing the number of MPI per events, de facto increasing the number of particles likely to be produced.

The corrected $\langle p_{\mathrm{T}} \rangle$ versus multiplicity profile is also compared with other experiments and Monte-Carlo generators, at 0.9, 2.76 and 7 TeV in Figure 6.17. The comparison with ATLAS is excellent for particles with p_{T} higher than 0.5 GeV, while its measurement is lower than the CMS result when decreasing the p_{T} requirement, though staying within the errors. Albeit being tuned to 7 TeV data already, the Monte-Carlo generators tend again to produce particles with too strong p_{T}. It seems like the PYTHIA model cannot produce that many soft multiple-parton-interactions, and although it describes much better the physics in restricted acceptances, it shows the same inherent defects as with previous tunes, which indicates that the underlying model can still be further investigated.

6.6.2 KNO distributions, Moments

The scaling with energy of the multiplicity has been studied again using the KNO formalism detailed in Section 1.4, as was done with the NSD measurements. As shown in Figure 6.18 for tracks with $p_{\mathrm{T}} > 0.5$ GeV (left) and $p_{\mathrm{T}} > 0.1$ GeV (right), and within the full tracker acceptance ($|eta| < 2.4$), the multiplicity distribution is clearly dependent of the energy. This was already observed using a diffractive-dependent event selection, and it shows diffraction has no influence on the KNO scaling failing. A summary of the mean, dispersion, C-moments, normalized Factorial moments and Cumulants of the multiplicity distributions with the widest acceptance are shown in Table B.4 in appendix. The C-moments show an increase with energy far greater than their errors, which tends to confirm the measured scaling violation. Other moments show features similar to what was measured in Chapter 5, and confirm the overall increase of correlations at 7 TeV.

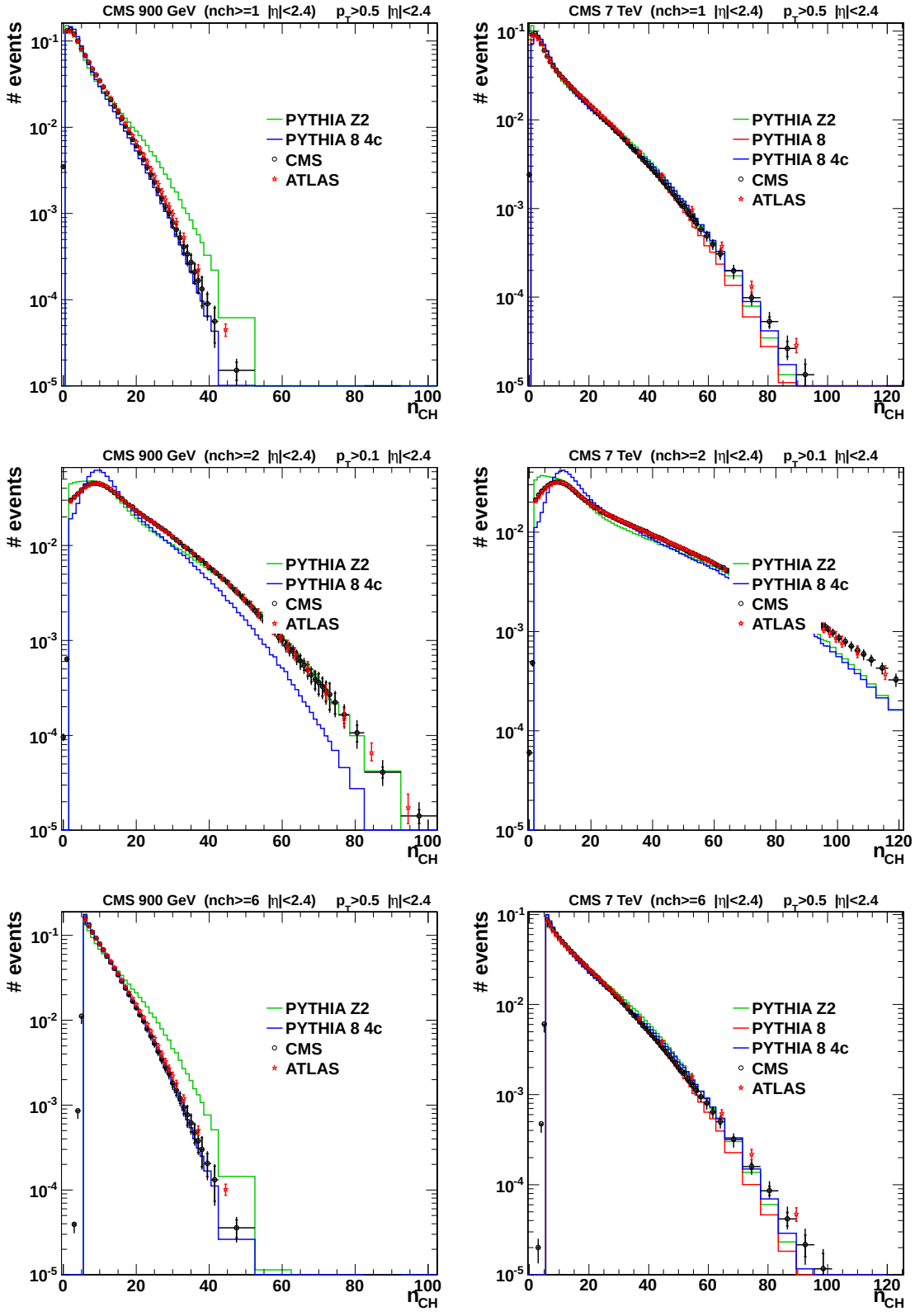

Figure 6.15: The fully corrected multiplicity distribution compared when available to ATLAS results [83, 109]. The generator predictions for PYTHIA 8 tune 4C and PYTHIA 6 tune Z2 were also added for comparison.

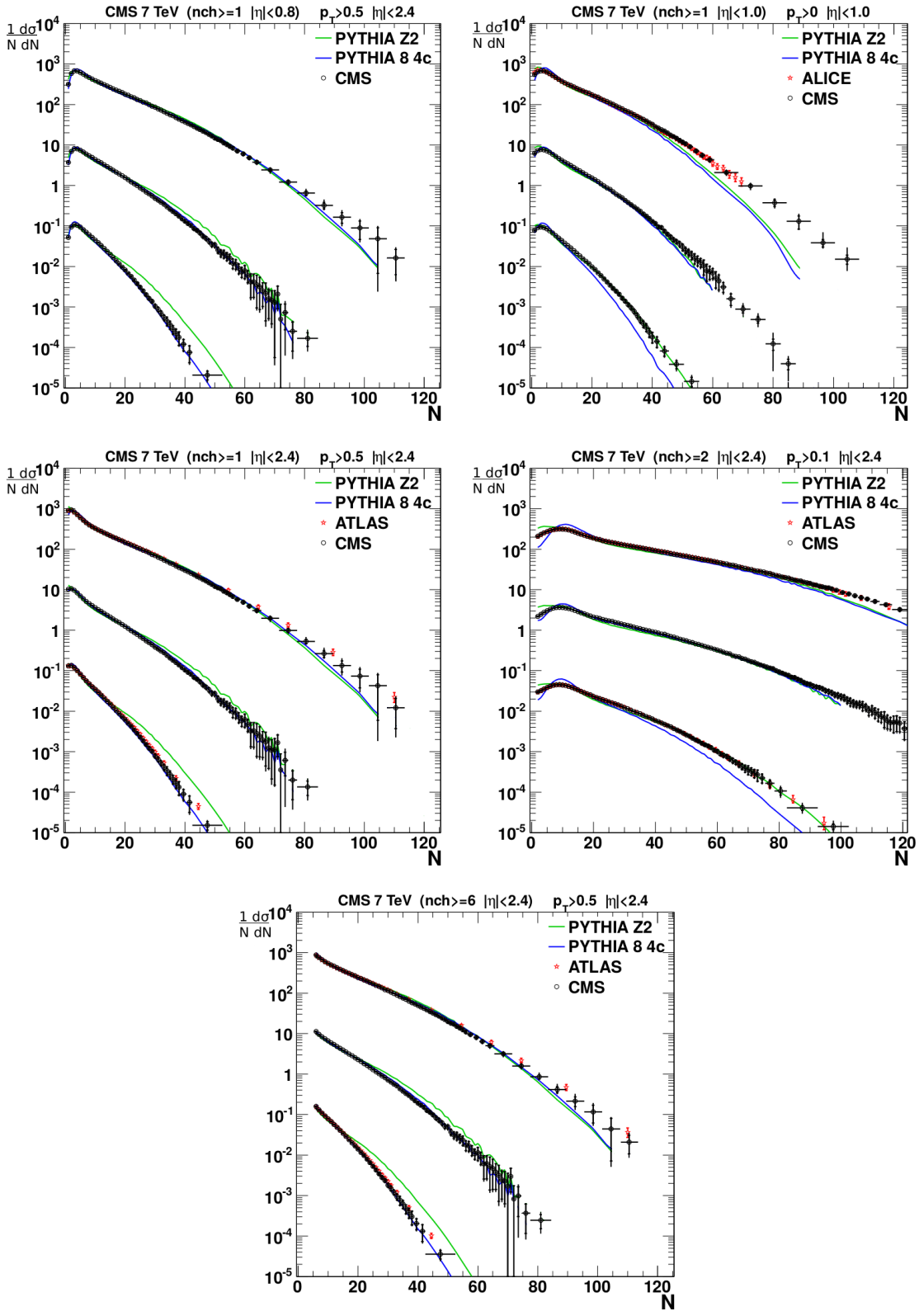

Figure 6.16: The fully corrected multiplicity distribution at 0.9, 2.76 and 7 TeV with different central selections. The generator prediction for PYTHIA 8 tune 4C and PYTHIA 6 tune Z2 were added for comparison, as well as results from ALICE and ATLAS when available. Distributions at 2.76 TeV (7 TeV) are scaled by a factor 100 (10000) to ease visibility.

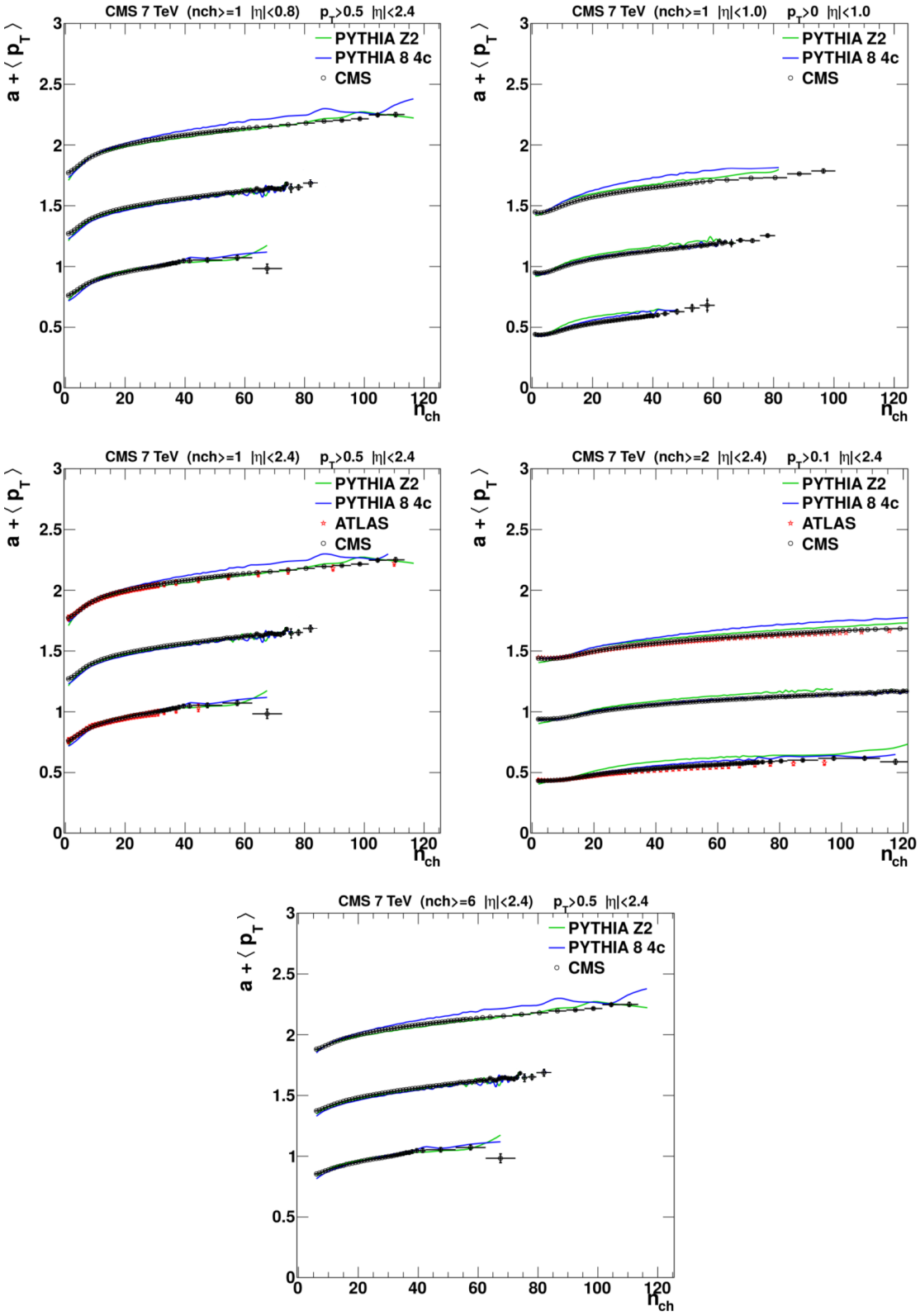

Figure 6.17: The fully corrected mean p_{T} versus multiplicity distribution at 0.9, 2.76 and 7 TeV with different central selections. The generator prediction for PYTHIA 8 tune 4C and PYTHIA 6 tune Z2 were added for comparison, as well as results from ALICE and ATLAS when available. Distributions at 2.76 TeV (7 TeV) were added a factor 0.5 (1) to ease visibility.

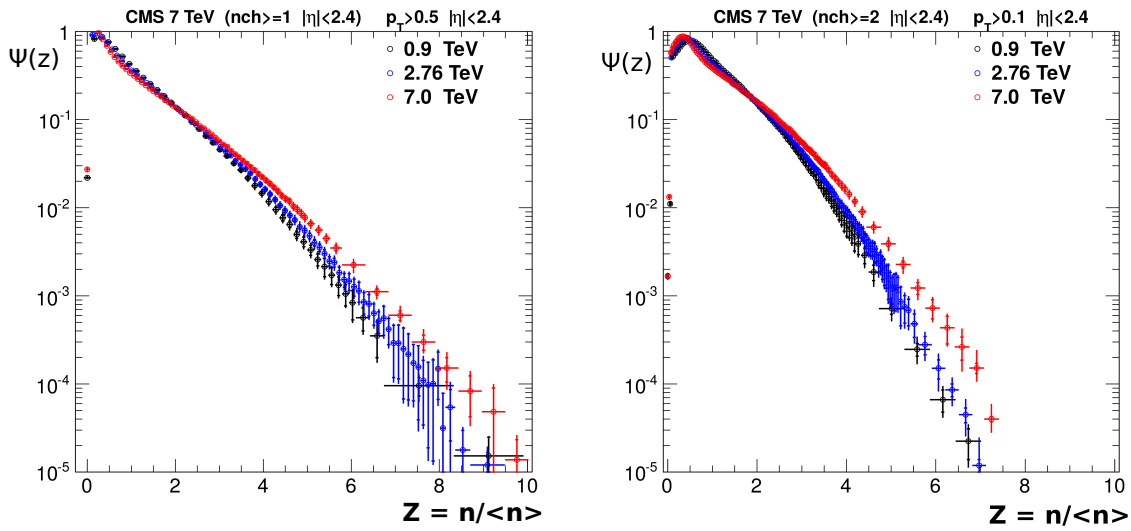

Figure 6.18: The multiplicity distribution in KNO form at 0.9, 2.76 and 7 TeV with different p_{T} cuts in $|\eta| < 2.4$. They clearly indicate KNO scaling violation.

6.6.3 Conclusion

This analysis uses the knowledge gained with the previous NSD results to measure the multiplicity with an event selection not based on diffraction. The multiplicity distributions were corrected using the same method as previous results. The measured charged hadron multiplicity were compared to ATLAS and ALICE results with a good agreement. New PYTHIA generators tuned to the available 7 TeV data were also compared to our measurements, and albeit matching much better the number of high-multiplicity events observed in data, they still tend to produce too few MPIs hence high-multiplicity events, with a harder transverse momentum than data. The shoulder in the multiplicity distributions were confirmed, as well as the breaking of KNO scaling in wide pseudorapidity domains.

Summary

The microscopic world has been the subject of many studies throughout the years, although the last century has proven to show many progress in the field. It is amusing that the study of the smallest particles known to mankind requires nowadays the largest machineries ever built to understand them. The LHC is such a machine, and currently the one which provides the highest colliding energy. This collider, built on the premises of its elder, the LEP, was designed to provide collisions with 14 TeV of center-of-mass energy, and four experiments are placed along the ring to detect the result of the two proton beams colliding. The CMS detector is one of those experiments, and has a wide physics program dedicated to advancing the scientific knowledge in the particle physics' field. As the last piece of the Standard Model that had yet to make its appearance, the Higgs Boson has been driving the development of the LHC and CMS from their conception, which proved commendable as it has finally been discovered in July 2012. It is obviously not the only physics goal of CMS, and many regions of the infinitesimally small world have been probed during the LHC Run I.

The LHC started providing the first collisions in September 2008, although some sections of the collider were damaged due to a malfunction shortly afterwards. The year spent repairing the LHC ring was albeit not lost for the experiments, as they used it to thoroughly commission their freshly assembled detector. This thesis described in great length the inner-workings of the subdetector closest to the interaction point and part of the tracking system, the pixel detector. Its 66 million channels are used to closely measure the trajectory of the particles, and deduct the exact location they originate from. Several calibrations procedures were explained, with an emphasis on the gain calibration. This calibration is needed to translate the charge collected by the pixel cells in ADC counts back to electrons, this for all of its channels. Along with the procedure to retrieve the calibration information on a pixel basis, results from several calibration runs were presented, with their evolution in time and thus the effects of the prolonged irradiation the detector suffers and the condition changes applied in order to keep it operating optimally. It was also the opportunity for the CMS experiment to test the data analysis flow, by collecting some Cosmic Rays, of which some were presented here. In particular, the pixel efficiency measurement using the result of the tracking was detailed in this thesis. An average $(97.55 \pm 0.05)\%$ efficiency was obtained for the barrel region, using data from about 95% good working modules, while the configuration of the forward section lacked enough statistics to provide a measurement, as cosmics and collisions are too different.

As the LHC started delivering pp collisions with increasing energy, from the 0.9 TeV provided by the SpS, to 2.26 TeV, then to its full 7 TeV that was deemed the highest while still safe to operate energy that it could accelerate the beams to, the experiments recorded their first data, from which physics results were quickly extracted. The analysis presented in this thesis was part of this initial phase, as it measured the charged particle multiplicity at the three energies for which data was acquired. Mainly for historical reasons, the

measurement was first performed for a selection favoring non-single diffractive events, to allow comparison with work done by the previous experiments at lower collision energies. Using tracking that allows to reconstruct the trajectories of particles with a momentum down to 100 MeV, the multiplicity distribution was first corrected to account for the trigger inefficiency, after which the remaining single-diffractive component was subtracted, using Monte-Carlo predictions. To infer the true multiplicity devoid of the effects of the tracking algorithm efficiency, the presence of secondaries, or that of decay products from long lived hadrons, an unfolding procedure was implemented based on the Bayes' theorem. The statistical uncertainties were obtained with resampling. Many effects were studied in order to see their impact on the resulting charged particle multiplicity measurement, but only three were significant enough to be taken into account as systematics: the model dependence of the single-diffractive subtraction, the trigger and selection efficiency, and the tracking efficiency description. Finally, a small correction accounted for particles with a momentum smaller than 100 MeV, momentum that made them hard to detect with the phase I design of CMS.

The fully corrected charged particle multiplicity distributions were presented in several pseudo-rapidity acceptance intervals from $|\eta| < 0.5$ to $|\eta| < 2.4$. They agree remarkably well with results preceding the LHC era, as well as with those from the other LHC experiments. One of the striking features of the distributions when energy increases, is the clear apparition of a shoulder at high multiplicity, which derives from a large proportion of events with a high multiplicity. The charged particle multiplicity distributions were also compared with expectancies from Monte-Carlo generators, where the lack of generated events at high multiplicity is flagrant, as the tail of the multiplicity in data cannot be reproduced by any generator. This demonstrates the rising importance of the multiple parton interaction at 7 TeV, with Monte-Carlo generators underestimating their amount, producing to few high multiplicity events. The study of the mean momentum with respect to the multiplicity, also presented here at the three energies, describes further the characteristics of high multiplicity events, which have softer momentum than expected by the Monte-Carlo generators, tending to produce too many jets. As expected, it also seemed to be energy independent, as the ratio of the distributions between two energies demonstrates. The increase of the mean multiplicity with energy is again steeper than what the Monte-Carlo generators predict, and few models follow the correct trend. Using the KNO formalism, the multiplicity distributions were shown to break the KNO scaling for wide pseudorapidity intervals, which was confirmed through the study of their ordinary statistical moments. The KNO scaling violation can be essentially attributed to changes of cross-sections of each processes that compose minimum-bias data with the collision energy, as each of them separately does follow such scalings, but the increase for instance of the multiple parton interactions at 7 TeV disturbs the measured multiplicity distribution such that KNO scaling breaks. On the other hand, the KNO scaling was found to hold for the smaller pseudorapidity intervals, perhaps due the presence of only central events devoid of most of the long-range correlations. Other moments showed the harsh increase of the inclusive correlations with energy, while the genuine correlations were more tamed, hinting to presence of more long-range correlations at high energy.

After consequently more data provided by the LHC, along with pp collisions at 2.76 TeV to provide comparison with the heavy ions collisions at the same energy, the charged particle multiplicity analysis was applied to the new data, albeit with a different event selection, and was presented here. The goal of the new selection was twofold: first erase the dependence on the definition of diffraction, which differs between generators; then provide a concerted set of selections between all LHC experiments. The results gained from the knowledge acquired from the first measurements, and showed an excellent agreement

between all experiments. The previous observations extracted from the smaller datasets were here confirmed. The Monte-Carlo generators offered a better agreement, as they had been tuned with the 7 TeV measurements provided by several previous analysis, but the underlying physics still needs to be better deciphered to comprehend all of the soft QCD mechanisms.

Acknowledgements

My years spent in Antwerp for my PhD and during the writing of this thesis have been the occasion to cross path with many people that have influenced my PhD life. I would like to spare a few lines of this thick piece a paper to thank them:

I am first and foremost grateful that Nick van Remortel chose me as one of his first PhD students. Always present when needed, I enjoyed the guidance he showed during all those years. I also appreciate all the opportunities that I got during those years, especially to travel, and am glad you kept following me even during all those -long- years of writing.

Although few in numbers, my discussions with Eddi de Wolf have always shown me what little I know about physics in general, and what wisdom looks like. There's no limit to what you can learn in a couple of hours.

A special thanks to Xavier Janssen, if not for you this analysis would never have been finished. I've enjoyed working with you in the stressful times and coding with you in the more relaxed periods. Although you've apparently been a bad influence concerning bash scripting :)

I would like to thank Luca Mucibello and Michele Selvaggi for sharing our first years of PhD and Antwerp life. They have been great companions during this journey, and have brought a little of their Italy with them to my great benefit.

I am grateful to Benoit Roland for his friendship and for having been my window to the Wallonian part of Belgium. B.t.w., you still have two of my comics books :)

Many thanks to all the colleagues I've encountered or shared an office with in Antwerp. I've managed through you to enjoy many countries and cultures while staying still.

I would also like to thank everyone in the pixel group. I've really enjoyed learning how it works, and helping with the technical work, and the long nights at CERN making sure it functioned properly. For our little pixel friend, Banzaï !!

Finally, last but not least, a special thanks goes to my family for, well, everything. I could not have done this without you. Literally :) Short of writing this thesis, you've helped me whenever I needed it. Especially when I was in dire shortage of quenelle and saucisson :p

Oh, and thank *you* for reading this too. You're quite courageous if you've managed to read up until here ! I hope you've found what you were looking for . If not, well see you next time ! Or not ...

Abbreviations

AOH Analog Optical Hybrid.

BPIX Barrel PIXel.

BSC Beam Scintillator Counter.

CCU Communications and Control Unit.

CERN Conseil Européen pour la Recherche Nucléaire.

CMS Compact Muon Solenoid.

CRAFT08 Cosmic Run At Four Tesla 2008.

CTF Combinatorial Track Finder.

DAC Digital to Analog Converter.

DAQ Data Acquisition System.

DD Double-Diffractive.

DOH Digital Optical Hybrids.

DQM Data Quality Monitoring.

DT Drift Tube.

EW ElectroWeak.

FEC Front-End Controller.

FED Front-End Driver.

FPIX Forward PIXel.

HDI High Density Interconnect.

HLT High Level Trigger.

IOV Interval Of Validity.

IP Interaction Point.

ISR Intersecting Storage Rings.

LEP Large Electron-Positron.

LHC Large Hadron Collider.

LS Long Shutdown.

MPI Multiple Parton Interaction.

NBD Negative Binomial Distribution.

ND Non-Diffractive.

NDOF Number of Degree Of Freedom.

NSD Non-Single Diffractive.

PLL Phase Locked Loop.

PS Proton Synchrotron.

PU Pile-Up.

PUC Pixel Unit Cell.

QCD Quantum ChromoDynamics.

QCD Quantum ElectroDynamics.

ROC Read-Out Chip.

SC SynchroCyclotron.

SD Single-Diffractive.

SLHC Super LHC.

SM Standard Model.

SPS Super Proton Synchrotron.

SST Silicon Strip Tracker.

TBM Token Bit Manager.

TDR Technical Design Report.

TEC Tracker Enc-Caps.

TIB Tracker Inner Barrel.

TID Tracker Inner Disks.

TIF Tracker Integration Facility.

TOB Tracker Outer Barrel.

UE Underlying Event.

List of Figures

Figure 1.1 Several examples of negative binomial distributions taken from [4] . 11
Figure 1.2 Schematic representation of two protons colliding and dissociating into a single-diffractive system (left) and a double-diffractive system (right). Taken from [18]. 14

Figure 2.1 Schematic view of the CERN accelerator chain and all experiments present on site. 16
Figure 2.2 Schematic view of the CMS experiment with all its subdetectors. . . 18
Figure 2.3 Slice view of the CMS detector presenting each subdetectors and the particles they detect. 19
Figure 2.4 Longitudinal sketch of the tracker subdetector. Each line represents a module, while double lines indicate back-to-back modules delivering stereo hits. 20
Figure 2.5 Longitudinal sketch of one quarter of the ECAL subdetector. 21
Figure 2.6 Longitudinal sketch of one quarter of the HCAL subdetector and its subcomponents, which are all explained in the text. 22
Figure 2.7 Longitudinal sketch of one quarter of the muon chambers and its subcomponents, which are all explained in the text. 23

Figure 3.1 Schematic view of the pixel detector (left), with the three barrel layers and two end-cap disks on each side, along with the pseudorapidity acceptance of each layers/disks (Right). 28
Figure 3.2 The flow diagram of the steps in the construction of the FPix detector. 30
Figure 3.3 Picture of a bpix half module (left) and full module (right). Center: The components of a pixel barrel detector module (from top to bottom): the Kapton signal cable, the power cable, the HDI, the silicon sensor, the 16 ROCs and the basestrips 31
Figure 3.4 Charge deposition in the sensor (left), and attached to a Pixel Unit Cell (right). 32
Figure 3.5 Diagram of a ROC with the 3 parts highlighted: the active part containing the double columns and all the PUCs, the double column periphery containing the data and timestamp buffers, and the chip periphery. A more detailed color-coded schematic view of the electronics in each section can be viewed in Figure 3.7. 33
Figure 3.6 Schematic view of a portion of the sensor on top of a ROC with its PUC, double column periphery and chip periphery. 33
Figure 3.7 Schematic overview of all the electronical components and their layout within a ROC. The different DAC settings available are also presented. 35
Figure 3.8 Schematics of the TBM roll in the readout chain. 36
Figure 3.9 The analogue readout of a module with only one hit, in the first ROC. 37

Figure 3.10 Layout of the updated pixel detector, with the four layers and three disks on each side, and their global pseudorapidity acceptance. 40

Figure 3.11 Result of the delay scan, with 2 ns steps. The best value is with 10 ns, where the efficiency is high and the cluster size the highest. 42

Figure 3.12 (Left) S-Curve for one pixel, with the fit giving the threshold and noise values. (Right) Distribution for both bpix and fpix of the thresholds. 44

Figure 3.13 The left plot presents the distributions for the six different ADC levels used to encode the pixels address within one ROC. The central plot presents the RMS in ADC of each peak for all the active ROCs. The peaks are all separated, as shown in the right plot, where the separation between the mean of two adjacent peaks is plotted against sigma, the quadrature sum of their RMS. 45

Figure 3.14 Example of a fit of the gain calibration curve for one pixel that was successful. 47

Figure 3.15 3D plot of several variables - obtained during offline analysis of a gain calibration run - for a module in one of the end-cap disks, with two rows of 4 ROCs. Each point is the value for one pixel of the module, where its x and y coordinates are the pixel's column and row number. The variables presented here are respectively: the gain, the pedestal, the χ^2/NDOF of the fit, the abcisse (in VCAL) of the lowest point used in the fit, the abcisse of the highest point used in the fit, the number of points used in the fit, the ADC range between the lowest and highest point in the fit, and the computed ADC value of the plateau of the calibration curve. 51

Figure 3.16 Different representations of the extracted gain value: the first plot shows the mean per subcomponent of the pixel detector (3 barrel layers and 4 end-cap disks), the second one is the distribution of the gain values for all pixels in the barrel, the third plot shows the correlation of the gain with the pedestal with one entry per column (one entry for the computed mean for each column of the gain (x-axis) and pedestal (y-axis)), the fourth plot shows the mean gain of the module in function of its detid for the whole pixel detector. 56

Figure 3.17 2D representation of the mean gain (first 2 rows) and mean pedestal (last 2 rows) per module, with the whole pixel detector module layout: respectively the layer 1,2 and 3 of the barrel, and disks 1 and 2 of $+z$ side of the end-caps followed by disks 1 and 2 on the -z side. The modules on the layers are represented in function of their ladder and their position within one, while the disks use the plaquette number and a combination of the blade number and the panel number to represent the more complex geometry of the end-caps. 57

Figure 3.18 Evolution over several years of the mean of the gain and pedestal for all 3 layers of bpix and 4 disks of fpix. Each abcisse number represents an internal CMS gain calibration run number. 58

Figure 3.19 2D occupancy map of the first (left), second (center) and third (right) barrel layers, with each point representing the number of hits associated to a track within a ROC. White bins represent the ROCs that where taken out of data taking. The plot origin corresponds to $\phi = 0$ and $z = -26.7\,\text{cm}$. 62

Figure 3.20 Distribution of the cluster charge for the barrel (left) and end-caps (right). The dots are the data points and the line represents the simulation. The clusters must have more than one pixel, no edge pixel, and their charge is corrected to the particle path length in the sensor, dependent on its incidence angle. 63

Figure 3.21 Distribution of the size of the clusters in the transverse plane - the module's local x direction - in function of the cotangent of the local α angle - defined as the angle between the track and the surface of the module in the xz plane- for the barrel (left) and the end-caps (right), with two sets of data: with and without the 3.8 T magnetic field. The data points are fitted in order to derive the minimum of the curve defining the Lorentz Angle. . . 65

Figure 3.22 Hit residuals along the x and y coordinates obtained through Gaussian fits of each bins in α (left), β (center) and charge (right). 66

Figure 3.23 Hit resolution along the local coordinates x (left) and y (right) for the overlapping modules with enough data. The average over all overlapping sites gives $18.6 \pm 1.9\,\mu$m along x and $30.8 \pm 3.2\,\mu$m along y. 67

Figure 3.24 Transversal (left) and longitudinal (right) track impact parameter residuals obtained through the splitting method [33]. Distributions are fitted with a Gaussian (solid line). 67

Figure 3.25 (left) Efficiency of finding tracks passing through the pixel subdetector. The points refer to CRAFT data, the change in the timing set-up of the pixel subdetector during data taking period is reported with two different set of points. The dashed line refers to Monte-Carlo simulation. (center) Efficiency ϵ_{hit} for different values of the edge cut (number of standard variations of the propagated state on the layer for the x axis) and of the muon time cut (track muon arrival time window for the y axis). All the other cuts are already applied. (right) Slice of central plot showing efficiency dependence on the muon arrival time window for a fixed edge cut choice of 1 standard variation. 69

Figure 3.26 Schematic view of the telescope cut. In track number 1 the green hit under analysis satisfies the telescopic requirement due to the presence of at least one hit in the upper region and at least one hit in the lower region. In track number 2 the red hit under analysis does not satisfy the telescope since no hit is found on the same track in the lower region of the detector. . 71

Figure 3.27 (Left) Pixel cluster distribution of charge versus dimension without applying any selection on hits nor on tracks. Cluster charge (center) and dimension (right) after applying the specified cut sequence, with each distribution normalized to unity. 73

Figure 3.28 Hit detection efficiency measured in each sensor of the first (left), second (center) and third (right) barrel layer. The white cells with diagonal strip correspond to dead/half-dead modules, the colored cells with a border correspond to misconfigured modules. The white modules have relative errors higher than 10% on the calculated efficiency. 73

Figure 3.29 Relative errors for the hit detection efficiency measured in each sensor of the first (left), second (center) and third (right) barrel layer. As for Figure3.28, the dead/half-dead modules correspond to white cells with diagonal strip. 74

Figure 3.30 Efficiency in function of local variables (a) x_{local}, y_{local}, α_{local} and β_{local}. 75

Figure 3.31 Efficiency distribution of α_{local} versus x_{local} for barrel modules with 2 rows of ROCs. The figure on the left is obtained applying all the selection cuts while the right one is obtained without applying any selection cut. . . 76

Figure 3.32 Efficiency $\epsilon_{hit}(\Delta R)$ in function of the window search as defined in method II. The line in black refers to the observation done on all the non-dead modules, the line in blue refers to the observation done excluding all the dead and misconfigured modules. The asymptotic values reported of respectively 96.93% and 97.49% of efficiency are obtained fitting the red part of the curve with an horizontal straight line. 76

Figure 4.1 Schematic view of the making of tracklets from hits on pixel layers and a vertex. 80

Figure 5.1 The number of reconstructed vertices in each event at each center of mass energy. 90

Figure 5.2 The efficiency of the trigger and event selection as function of the true charged particle multiplicity in simulated NSD events at each center of mass energy. 90

Figure 5.3 A comparison of the track reconstruction efficiency (left) and fake rate (right) between the standard CMS tracking and the minimum bias tracking algorithm as function of the transverse momentum the tracks. . . . 91

Figure 5.4 The relative error on the transverse momentum measurement at each center of mass energy. 93

Figure 5.5 The impact parameter significance in the plane perpendicular to the beam pipe at each center of mass energy. 93

Figure 5.6 The impact parameter significance along the direction of the beam pipe at each center of mass energy. 93

Figure 5.7 Comparison between data and MC at $\sqrt{s}$=0.9 TeV of the transverse momentum, pseudorapidity, number of track hits and the goodness of fit. . 94

Figure 5.8 Comparison between data and MC at $\sqrt{s}$=2.36 TeV of the transverse momentum, pseudorapidity, number of track hits and the goodness of fit. . 94

Figure 5.9 Comparison between data and MC at $\sqrt{s}$=7.0 TeV of the transverse momentum, pseudorapidity, number of track hits and the goodness of fit. . 95

Figure 5.10 Comparison between data and MC models of the raw multiplicity distribution at $\sqrt{s}$= 0.9, 2.36 and 7.0 TeV. 95

Figure 5.11 The diffractive and non-diffractive contributions to the multiplicity spectrum in $|\eta| < 2.4$ at $\sqrt{s}$=0.9, 2.36 and 7.0 TeV, as predicted by PYTHIA 97

Figure 5.12 The contribution of SD events that are subtracted from the raw data multiplicity before unfolding. 97

Figure 5.13 Matrices used for unfolding. 98

Figure 5.14 A sanity check of the unfolding procedure of the charged particle multiplicity distribution, P_n, correcting reconstructed tracks back to the hadron level. The efficiency corrections for trigger and event selection criteria is also applied. 99

Figure 5.15 A sanity check of the unfolding procedure of the $< p_{\mathrm{T}} >$ vs n profile, correcting reconstructed tracks back to the hadron level. 100

Figure 5.16 An independent measurement of the trigger efficiency with zero bias data compared with the up and down variations applied to the MC description to bracket the measured values from data (left). The effect of the systematic variation of the event and trigger selection efficiency is shown to be of the same size as the effect of replacing the HF offline selection criterium with an alternative tighter trigger requirement (right). 101

Figure 5.17 The relative size of the statistical and total uncertainty on P_n (in percents) for each multiplicity. 103

Figure 5.18 The fully corrected charged hadron multiplicity spectrum for $|\eta| <$ 0.5, 1.0, 1.5, 2.0, and 2.4, (a) at $\sqrt{s} = 0.9\,\text{TeV}$, (b) 2.36 TeV, and (c) 7 TeV, compared with other measurements in the same η interval and at the same center-of-mass energy [46, 65, 81, 82]. For clarity, results in different pseudorapidity intervals are scaled by powers of 10 as given in the plots. The error bars are the statistical and systematic uncertainties added in quadrature. 105

Figure 5.19 The fully corrected charged hadron multiplicity spectrum for $|\eta| <$ 0.5, 1.0, 1.5, 2.0, and 2.4, with $p_\text{T} > 500\,\text{MeV}/c$ at $\sqrt{s}$=0.9, 2.36 and 7.0 TeV, compared with another measurement in the same η interval and at the same center-of-mass energy [83]. For clarity, results in different pseudorapidity intervals are scaled by powers of 10 as given in the plots. The error bars are the statistical and systematic uncertainties added in quadrature. 106

Figure 5.20 A comparison between the fully corrected multiplicity spectra for positively and negatively charged hadrons with $p_\text{T} > 500\,\text{MeV}/c$ in increasing pseudorapidity bins ranging from $|\eta| < 0.5$ to $|\eta| < 2.4$ at $\sqrt{s}$=0.9, 2.36 and 7.0 TeV. 107

Figure 5.21 The charged hadron multiplicity distributions in KNO form at $\sqrt{s} =$ 0.9 and 7 TeV in two pseudorapidity intervals: $|\eta| < 2.4$ (left) and $|\eta| < 0.5$ (right). 108

Figure 5.22 Fits of the $\log s$ dependence of the normalized moments C_q of the multiplicity distribution for $|\eta| < 2.4$, assuming linear dependence (left) and $|\eta| < 0.5$, assuming no dependence (right), including data from lower energy experiments [47, 93, 94]. For $\sqrt{s} = 0.9\,\text{TeV}$, data from experiments other than CMS were drawn shifted to lower $\sqrt{s}$ for clarity. 109

Figure 5.23 The charged hadron multiplicity distributions with $|\eta| < 2.4$ for (a) $p_\text{T} > 0$ and (b) $p_\text{T} > 500\,\text{MeV}/c$ at $\sqrt{s} = 0.9$, 2.36, and 7 TeV, compared to two different PYTHIA models and the PHOJET model. For clarity, results for different center-of-mass energies are scaled by powers of 10 as given in the plots. 110

Figure 5.24 A comparison between our measurement of the average transverse momentum versus the charge multiplicity for $|\eta| < 2.4$ with several available PYTHIA tunes and the PHOJET model at $\sqrt{s}$=0.9, 2.36 and 7.0 TeV. 111

Figure 5.25 A comparison of $\langle p_\text{T} \rangle$ versus n for $|\eta| < 2.4$ with two different PYTHIA models and the PHOJET model at $\sqrt{s} = 0.9$, 2.36, and 7 TeV. For clarity, results for different energies are shifted by the values of a shown in the plots. Fits to the high-multiplicity part ($n > 15$) with a linear form in $\sqrt{n}$ are superimposed. (b) The ratios of the higher-energy data to the fit at 0.9 TeV indicate the approximate energy independence of $\langle p_\text{T} \rangle$ at fixed n. . 111

Figure 5.26 The evolution of the mean charge multiplicity with the center of mass energy for pseudo-rapidity acceptances of $|\eta| < 1.5$ (left) and $|\eta| < 0.5$ (right). Data from lower energy experiments have been added where available. [47, 50, 93, 94, 105, 106]. The data are compared with predictions from three analytical Regge-inspired models [69–71] 112
Figure 5.27 The evolution of the mean charge multiplicity with the center-of-mass energy for $|\eta| < 2.4$, including data from lower-energy experiments for $|\eta| < 2.5$ [47, 50, 93, 94]. The data are compared with predictions from three analytical Regge-inspired models [69–71] and from a saturation model [72]. 113
Figure 5.28 $\log s$ dependence of the normalized factorial moments F_q of the multiplicity distribution for $|\eta| < 2.4$ (left) and $|\eta| < 0.5$ (right), including data from lower energy experiments [47, 94]. For $\sqrt{s} = 0.9\,\text{TeV}$, data from experiments other than CMS were drawn shifted to lower $\sqrt{s}$ for clarity. . . 114
Figure 5.29 $\log s$ dependence of the normalized cumulants K_q (left) and the moments H_q (right) of the multiplicity distribution for $|\eta| < 2.4$, including data from lower energy experiments [47, 94]. For $\sqrt{s} = 0.9\,\text{TeV}$, data from experiments other than CMS were drawn shifted to lower $\sqrt{s}$ for clarity. . . 115
Figure 5.30 Evolution of the H_q moments of the multiplicity distribution with their q-order for $|\eta| < 2.4$, including data from lower energy experiments [47, 94]. 115

Figure 6.1 The relative error on the transverse momentum measurement at each center-of-mass energy. 120
Figure 6.2 The impact parameter significance in the plane perpendicular to the beam pipe at two center-of-mass energy. 120
Figure 6.3 The impact parameter significance along the direction of the beam pipe at two center of mass energy. 120
Figure 6.4 Comparison between data and MC at $\sqrt{s}$=0.9 TeV of the transverse momentum, pseudorapidity, number of track hits and the goodness of fit. . 121
Figure 6.5 Comparison between data and MC at $\sqrt{s}$=2.76 TeV of the transverse momentum, pseudorapidity, number of track hits and the goodness of fit. . 121
Figure 6.6 Comparison between data and MC at $\sqrt{s}$=7.0 TeV of the transverse momentum, pseudorapidity, number of track hits and the goodness of fit. . 122
Figure 6.7 Comparison between data and MC models of the raw multiplicity distribution at $\sqrt{s}$= 0.9, 2.76 and 7.0 TeV. They have not been normalized by their binwidth as they serve as input to the correction. 122
Figure 6.8 The efficiency of the trigger correction, the vertex requirement correction, and the central selection for the multiplicity distribution at each center of mass energy. 123
Figure 6.9 Matrices used for unfolding at different center of mass energies. . . . 124
Figure 6.10 The efficiency of the trigger correction, the vertex requirement correction, the pt correction and the central selection for the mean p_{T} versus multiplicity distribution at each center of mass energy. 125
Figure 6.11 The different steps of data correction of the $< p_{\text{T}} >$ vs n profile, correcting reconstructed tracks back to the hadron level. 126
Figure 6.12 Mean number of tracks belonging to the second vertex, but that passed association cuts to the studied vertex, in function of its multiplicity, for fully selected events with exactly 2 reconstructed vertices, at 2.76TeV. Left: tracks were selected with $p_{\text{T}} > 0.1\,\text{GeV}$; Right: tracks were selected with $p_{\text{T}} > 0.5\,\text{GeV}$ 128

Figure 6.13 All systematic effects added one by one in quadrature to the statistical errors of the fully corrected multiplicity distribution for several central selections and track acceptances. 130
Figure 6.14 The fully corrected multiplicity distribution compared to ALICE results [82]. The generator prediction for PYTHIA 8 tune 4C and PYTHIA 6 tune Z2 were also added for comparison. 131
Figure 6.15 The fully corrected multiplicity distribution compared when available to ATLAS results [83, 109]. The generator predictions for PYTHIA 8 tune 4C and PYTHIA 6 tune Z2 were also added for comparison. 133
Figure 6.16 The fully corrected multiplicity distribution at 0.9, 2.76 and 7 TeV with different central selections. The generator prediction for PYTHIA 8 tune 4C and PYTHIA 6 tune Z2 were added for comparison, as well as results from ALICE and ATLAS when available. Distributions at 2.76 TeV (7 TeV) are scaled by a factor 100 (10000) to ease visibility. 134
Figure 6.17 The fully corrected mean p_{T} versus multiplicity distribution at 0.9, 2.76 and 7 TeV with different central selections. The generator prediction for PYTHIA 8 tune 4C and PYTHIA 6 tune Z2 were added for comparison, as well as results from ALICE and ATLAS when available. Distributions at 2.76 TeV (7 TeV) were added a factor 0.5 (1) to ease visibility. 135
Figure 6.18 The multiplicity distribution in KNO form at 0.9, 2.76 and 7 TeV with different p_{T} cuts in $|\eta| < 2.4$. They clearly indicate KNO scaling violation. 136

Figure A.1 Examples of gain calibration curves for pixels where fits could not be performed: the left one represents a noisy pixel, the right one was measured for a pixel where an error occurred during the calibration run. 156

Figure C.1 The relative size of the systematic uncertainty on P_n due the trigger and event selection criteria for the measured multiplicities at $|\eta| < 2.4$ at $\sqrt{s}$=0.9, 2.36 and 7.0 TeV. The size of the grey band indicates the systematic error on P_n (in percents) for each multiplicity. 163
Figure C.2 The relative size of the systematic uncertainty due the mismodeling of tracking efficiencies and secondaries/fakes for the measured multiplicities at $|\eta| < 2.4$ at $\sqrt{s}$=0.9, 2.36 and 7.0 TeV. The size of the grey band indicates the systematic error on P_n (in percents) for each multiplicity. . . . 164
Figure C.3 The relative size of the systematic uncertainty due the mismodeling of the fraction and shape of the subtracted SD events for the measured multiplicities at $|\eta| < 2.4$ at $\sqrt{s}$=0.9, 2.36 and 7.0 TeV. The size of the grey band indicates the systematic error on P_n (in percents) for each multiplicity. 165
Figure C.4 Effect of adding 2 iterations to the unfolding on the fully corrected multiplicity distribution for several central selections and track acceptances. 166
Figure C.5 Effect of not taking the mixed tracking algorithms on the fully corrected multiplicity distribution for several central selections and track acceptances. 167
Figure C.6 Effect of 0-bias statistical errors when correcting for the trigger event selection on the fully corrected multiplicity distribution for several central selections and track acceptances. 168
Figure C.7 Effect of tracking inefficiencies misreconstruction in Monte-Carlo on the fully corrected multiplicity distribution for several central selections and track acceptances. 169

Figure C.8 Effect of different Monte-Carlo description on the fully corrected multiplicity distribution for several central selections and track acceptances. 170
Figure C.9 Effect of not taking the mixed tracking algorithms on the fully corrected mean p_{T} versus multiplicity distribution for several central selections and track acceptances. 171
Figure C.10 Effect of tracking inefficiencies misreconstruction in Monte-Carlo on the fully corrected mean p_{T} versus multiplicity distribution for several central selections and track acceptances. 172
Figure C.11 Effect of different Monte-Carlo description on the fully corrected mean p_{T} versus multiplicity distribution for several central selections and track acceptances. 173

Appendices

Appendix A

Gain Calibration Fitting Problems

As stated in the pixel chapter, there are many reasons a fit of the gain calibration curve for one pixel could go wrong: this is why their identification and categorization is important, to first inform on the quality of a specific calibration run, and then eventually understand and improve the extracted data. The status have been divided in two, with positive integers when the fit was good and resulted in usable values for the gain and the pedestal, and negative statuses when anything arose during the analysis for the studied pixel that lead to the constants being unusable. Beside their sign, the integers chosen for each feature have no special meaning, and were just selected in the order the features have been singled out:

- **0:** default value, stays at this value only if throughout the analysis problems arise meaning it is never set to something else.
- **1:** the fit was a success, the calibration curve is standard with no noticeable features.
- **3:** the fits was a success, but there is no plateau.
- **5:** the fit was a success, but its χ^2/NDOF is higher than 10 or its χ^2 probability is smaller than 0.05.
- **-2:** no fit because all points saturate the readout, the curve being a plateau.
- **-4:** no fit because the `ROOT` fit function returns failure, even after ten tries.
- **-6:** the fit returns *Not a Number (NaN)* for gain or slope.
- **-7:** no fit as there are only two valid data points to fit.
- **-8:** the fit has a problem as it returns a gain value higher than 25.

The ideal calibration would have only pixels with a status of 1. The status 3, where no plateau is detected, does not have any impact on the quality of the fit, on the contrary. It is also the main status for calibrations done with VCAL low settings, when the high VCAL values are not sampled, therefore injecting no charges that saturate the readout. The status identifying fits with a high χ^2 signifies only the data is not a perfect increasing line, something well known since the linear fit is only an approximation of the tanh function.

All the other statuses give meaningless gain and pedestal constants, if any. It is why they are skipped during the database object construction, and their value fixed to the mean

of either the column or the module if applicable: in case it is only during the calibration run that the pixel misbehaved, or that the dysfunction does not last in time, or even just has a problem specific to the gain calibration data but works fine when collecting data from charges drifting within the sensor, there is a sensible constant to operate the ADC/electron conversion during the reconstruction. Pixels are flagged with a -2 status when all the data from the calibration saturates the readout, meaning their is no rising portion to fit: this is the appanage of noisy pixels that are not yet masked. It is the reason a special flag has been reserved during the implementation of the database filling to store the information, as the gain calibration is one of the analysis that can identify noisy pixel channels. Figure A.1 (left) presents a calibration curve for a pixel always saturating its readout: it is evidently impossible to extract the gain and pedestal constants, thus the dismissal of the fitting procedure, but is rather easy to tag. The status -4 is there as a security, but the code happens to prevent fitting when it could completely fail, as I have never seen any pixels with this status. Likewise, the -6 status is very rare, and is implemented for completeness sake. In case not enough points are left to perform a fit, either because they were few to begin with, or the exclusion of points at small VCAL and the ones belonging to the plateau is too efficient, pixels are given a status of -7. Finally, a first limit on the gain value is set to 25, with the status -8: this is done to eliminate pixels for which the fit was correctly performed, but for whatever reason cannot give a sensible result. An example can be found in Figure A.1 (right), where it is clear that no calibration constant that would make sense can be extracted. The value of 25 is quite high, and lower gain values are usually still the mark of a problematic curve, but it can this way allow to limit the range of the histograms without cutting any tail of the gain distribution.

After the gross work of identifying evidently wrong configurations for the fit, some new maps presenting the fraction of bad fits per module and the fraction of gains higher than seven per module were put in place for all pixel subcomponents: it was easy this way to spot modules that had troubles, which was particularly helpful as a feedback to the online pixel experts.

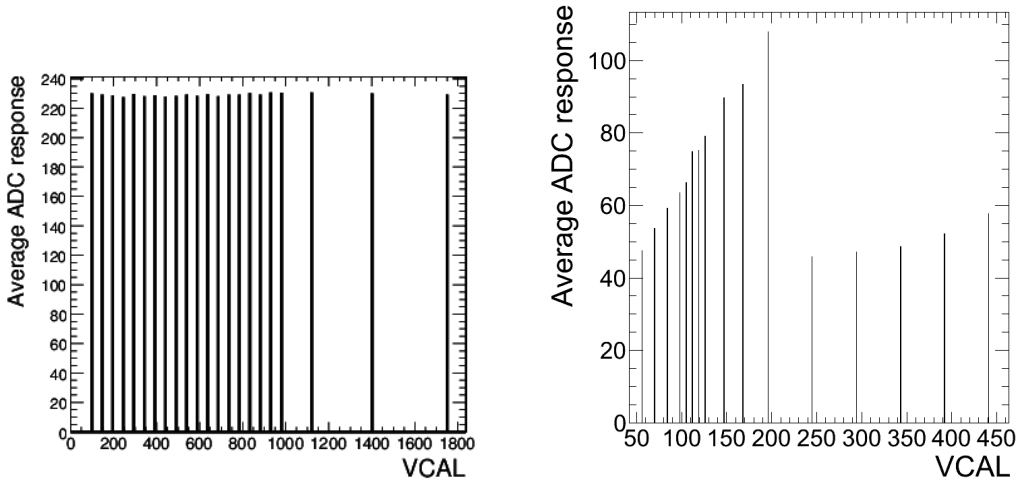

Figure A.1: Examples of gain calibration curves for pixels where fits could not be performed: the left one represents a noisy pixel, the right one was measured for a pixel where an error occurred during the calibration run.

Appendix B

Numerical Values for Moments of the Multiplicity Mistributions

This annex encompasses the numerical values of the mean $< n >$, the dispersion D_2 and orders 2 to 5 of the C, F, K and H moments (see Section 1.2). The first section comprises values for some of the results obtained using NSD events and detailed in Chapter 5, while the results for the analysis using inelastic events from Chapter 6 are compiled in the second section.

B.1 Non-Single-Diffractive Events

Table B.1: $< n >$ and moments for $|\eta| < 2.4$.

Center-of-mass Energy	0.9 TeV	2.36 TeV	7.0 TeV
$< n >$	$17.117 \pm 0.061^{+1.074}_{-1.007}$	$20.320 \pm 0.128^{+1.371}_{-1.207}$	$29.054 \pm 0.046^{+2.108}_{-1.902}$
D_2	$12.749 \pm 0.016^{+0.257}_{-0.226}$	$14.875 \pm 0.058^{+0.236}_{-0.271}$	$24.024 \pm 0.010^{+0.472}_{-0.515}$
C_2	$1.555 \pm 0.005^{+0.059}_{-0.066}$	$1.536 \pm 0.011^{+0.070}_{-0.079}$	$1.684 \pm 0.003^{+0.094}_{-0.102}$
C_3	$3.079 \pm 0.022^{+0.264}_{-0.281}$	$2.979 \pm 0.040^{+0.307}_{-0.327}$	$3.736 \pm 0.012^{+0.458}_{-0.467}$
C_4	$7.143 \pm 0.080^{+0.991}_{-1.013}$	$6.782 \pm 0.127^{+1.124}_{-1.139}$	$9.811 \pm 0.050^{+1.898}_{-1.834}$
C_5	$18.491 \pm 0.278^{+3.604}_{-3.546}$	$17.512 \pm 0.428^{+4.052}_{-3.919}$	$28.852 \pm 0.201^{+7.713}_{-7.082}$
F_2	$1.496 \pm 0.005^{+0.056}_{-0.063}$	$1.487 \pm 0.011^{+0.067}_{-0.076}$	$1.649 \pm 0.003^{+0.091}_{-0.099}$
F_3	$2.813 \pm 0.020^{+0.238}_{-0.255}$	$2.757 \pm 0.037^{+0.283}_{-0.302}$	$3.564 \pm 0.012^{+0.437}_{-0.446}$
F_4	$6.121 \pm 0.069^{+0.840}_{-0.865}$	$5.942 \pm 0.111^{+0.984}_{-0.998}$	$9.061 \pm 0.046^{+1.753}_{-1.694}$
F_5	$14.671 \pm 0.223^{+2.833}_{-2.804}$	$14.418 \pm 0.358^{+3.334}_{-3.231}$	$25.627 \pm 0.180^{+6.847}_{-6.289}$
K_2	$0.496 \pm 0.005^{+0.056}_{-0.063}$	$0.487 \pm 0.011^{+0.067}_{-0.076}$	$0.649 \pm 0.003^{+0.091}_{-0.099}$
K_3	$0.324 \pm 0.006^{+0.073}_{-0.069}$	$0.297 \pm 0.006^{+0.083}_{-0.076}$	$0.616 \pm 0.004^{+0.163}_{-0.148}$
K_4	$0.108 \pm 0.007^{+0.052}_{-0.045}$	$0.124 \pm 0.019^{+0.054}_{-0.041}$	$0.436 \pm 0.005^{+0.181}_{-0.144}$
K_5	$-0.377 \pm 0.021^{+0.108}_{-0.107}$	$-0.037 \pm 0.077^{+0.114}_{-0.105}$	$-0.532 \pm 0.006^{+0.193}_{-0.218}$
H_2	$0.332 \pm 0.002^{+0.024}_{-0.029}$	$0.327 \pm 0.005^{+0.029}_{-0.035}$	$0.394 \pm 0.001^{+0.032}_{-0.038}$
H_3	$0.115 \pm 0.001^{+0.015}_{-0.015}$	$0.108 \pm 0.001^{+0.018}_{-0.017}$	$0.173 \pm 0.000^{+0.022}_{-0.022}$
H_4	$0.018 \pm 0.001^{+0.006}_{-0.006}$	$0.021 \pm 0.003^{+0.006}_{-0.005}$	$0.048 \pm 0.000^{+0.010}_{-0.008}$
H_5	$-0.026 \pm 0.001^{+0.003}_{-0.003}$	$-0.003 \pm 0.005^{+0.008}_{-0.007}$	$-0.021 \pm 0.000^{+0.004}_{-0.003}$

Table B.2: $< n >$ and moments for $|\eta| < 1.5$.

Center-of-mass Energy	0.9 TeV	2.36 TeV	7.0 TeV
$< n >$	$10.622 \pm 0.027^{+0.555}_{-0.575}$	$12.734 \pm 0.052^{+0.735}_{-0.731}$	$17.888 \pm 0.022^{+1.170}_{-1.191}$
D_2	$8.667 \pm 0.029^{+0.204}_{-0.190}$	$10.154 \pm 0.008^{+0.153}_{-0.142}$	$15.801 \pm 0.009^{+0.335}_{-0.344}$
C_2	$1.666 \pm 0.006^{+0.057}_{-0.052}$	$1.636 \pm 0.006^{+0.074}_{-0.070}$	$1.780 \pm 0.002^{+0.103}_{-0.096}$
C_3	$3.633 \pm 0.029^{+0.284}_{-0.254}$	$3.416 \pm 0.021^{+0.346}_{-0.316}$	$4.249 \pm 0.008^{+0.536}_{-0.474}$
C_4	$9.466 \pm 0.121^{+1.198}_{-1.044}$	$8.352 \pm 0.068^{+1.355}_{-1.188}$	$12.158 \pm 0.036^{+2.413}_{-2.030}$
C_5	$27.928 \pm 0.483^{+4.965}_{-4.207}$	$22.960 \pm 0.256^{+5.195}_{-4.365}$	$39.391 \pm 0.151^{+10.790}_{-8.628}$
F_2	$1.572 \pm 0.006^{+0.052}_{-0.047}$	$1.557 \pm 0.005^{+0.069}_{-0.065}$	$1.724 \pm 0.002^{+0.099}_{-0.092}$
F_3	$3.180 \pm 0.027^{+0.245}_{-0.219}$	$3.043 \pm 0.018^{+0.307}_{-0.281}$	$3.957 \pm 0.008^{+0.499}_{-0.442}$
F_4	$7.571 \pm 0.103^{+0.948}_{-0.829}$	$6.850 \pm 0.055^{+1.110}_{-0.975}$	$10.793 \pm 0.032^{+2.142}_{-1.802}$
F_5	$20.076 \pm 0.370^{+3.532}_{-3.004}$	$17.100 \pm 0.209^{+3.867}_{-3.254}$	$33.044 \pm 0.125^{+9.047}_{-7.236}$
K_2	$0.572 \pm 0.006^{+0.052}_{-0.047}$	$0.557 \pm 0.005^{+0.069}_{-0.065}$	$0.724 \pm 0.002^{+0.099}_{-0.092}$
K_3	$0.465 \pm 0.010^{+0.089}_{-0.077}$	$0.371 \pm 0.004^{+0.100}_{-0.085}$	$0.784 \pm 0.003^{+0.202}_{-0.166}$
K_4	$0.300 \pm 0.012^{+0.098}_{-0.082}$	$0.091 \pm 0.021^{+0.060}_{-0.042}$	$0.738 \pm 0.004^{+0.289}_{-0.210}$
K_5	$-0.352 \pm 0.029^{+0.077}_{-0.084}$	$-0.363 \pm 0.094^{+0.179}_{-0.196}$	$-0.273 \pm 0.007^{+0.124}_{-0.114}$
H_2	$0.364 \pm 0.002^{+0.021}_{-0.020}$	$0.358 \pm 0.002^{+0.028}_{-0.028}$	$0.420 \pm 0.001^{+0.032}_{-0.032}$
H_3	$0.146 \pm 0.002^{+0.016}_{-0.015}$	$0.122 \pm 0.001^{+0.019}_{-0.018}$	$0.198 \pm 0.000^{+0.024}_{-0.022}$
H_4	$0.040 \pm 0.001^{+0.007}_{-0.007}$	$0.013 \pm 0.003^{+0.006}_{-0.005}$	$0.068 \pm 0.000^{+0.012}_{-0.009}$
H_5	$-0.018 \pm 0.001^{+0.003}_{-0.003}$	$-0.021 \pm 0.005^{+0.008}_{-0.007}$	$-0.008 \pm 0.000^{+0.004}_{-0.002}$

Table B.3: $< n >$ and moments for $|\eta| < 0.5$.

Center-of-mass Energy	0.9 TeV	2.36 TeV	7.0 TeV
$< n >$	$3.497 \pm 0.013^{+0.177}_{-0.173}$	$4.307 \pm 0.019^{+0.227}_{-0.215}$	$5.849 \pm 0.007^{+0.350}_{-0.333}$
D_2	$3.424 \pm 0.006^{+0.100}_{-0.080}$	$4.050 \pm 0.015^{+0.088}_{-0.074}$	$5.890 \pm 0.003^{+0.150}_{-0.136}$
C_2	$1.959 \pm 0.006^{+0.063}_{-0.060}$	$1.884 \pm 0.009^{+0.072}_{-0.072}$	$2.014 \pm 0.001^{+0.097}_{-0.097}$
C_3	$5.205 \pm 0.039^{+0.372}_{-0.351}$	$4.728 \pm 0.047^{+0.404}_{-0.390}$	$5.628 \pm 0.008^{+0.587}_{-0.560}$
C_4	$17.003 \pm 0.202^{+1.957}_{-1.789}$	$14.264 \pm 0.216^{+1.947}_{-1.811}$	$19.324 \pm 0.045^{+3.168}_{-2.886}$
C_5	$64.320 \pm 1.052^{+10.370}_{-9.193}$	$49.052 \pm 1.055^{+9.328}_{-8.342}$	$76.479 \pm 0.286^{+17.276}_{-15.038}$
F_2	$1.673 \pm 0.005^{+0.050}_{-0.048}$	$1.652 \pm 0.009^{+0.061}_{-0.062}$	$1.843 \pm 0.001^{+0.087}_{-0.088}$
F_3	$3.688 \pm 0.029^{+0.254}_{-0.244}$	$3.523 \pm 0.037^{+0.299}_{-0.290}$	$4.654 \pm 0.006^{+0.486}_{-0.465}$
F_4	$9.694 \pm 0.123^{+1.085}_{-1.011}$	$8.719 \pm 0.139^{+1.187}_{-1.112}$	$14.168 \pm 0.033^{+2.323}_{-2.121}$
F_5	$28.469 \pm 0.498^{+4.468}_{-4.038}$	$23.745 \pm 0.572^{+4.506}_{-4.064}$	$48.716 \pm 0.208^{+10.991}_{-9.582}$
K_2	$0.673 \pm 0.005^{+0.050}_{-0.048}$	$0.652 \pm 0.009^{+0.061}_{-0.062}$	$0.843 \pm 0.001^{+0.087}_{-0.088}$
K_3	$0.670 \pm 0.013^{+0.107}_{-0.098}$	$0.567 \pm 0.012^{+0.115}_{-0.106}$	$1.124 \pm 0.003^{+0.224}_{-0.202}$
K_4	$0.619 \pm 0.019^{+0.160}_{-0.142}$	$0.264 \pm 0.027^{+0.115}_{-0.094}$	$1.480 \pm 0.013^{+0.445}_{-0.368}$
K_5	$-0.353 \pm 0.035^{+0.092}_{-0.125}$	$-0.837 \pm 0.118^{+0.270}_{-0.255}$	$0.504 \pm 0.106^{+0.405}_{-0.288}$
H_2	$0.402 \pm 0.002^{+0.017}_{-0.018}$	$0.395 \pm 0.003^{+0.022}_{-0.023}$	$0.457 \pm 0.000^{+0.025}_{-0.027}$
H_3	$0.182 \pm 0.002^{+0.015}_{-0.016}$	$0.161 \pm 0.002^{+0.018}_{-0.018}$	$0.242 \pm 0.000^{+0.021}_{-0.021}$
H_4	$0.064 \pm 0.001^{+0.009}_{-0.009}$	$0.030 \pm 0.003^{+0.008}_{-0.008}$	$0.104 \pm 0.001^{+0.013}_{-0.012}$
H_5	$-0.012 \pm 0.001^{+0.004}_{-0.005}$	$-0.035 \pm 0.005^{+0.009}_{-0.006}$	$0.010 \pm 0.002^{+0.006}_{-0.005}$

B.2 Inelastic Events

Table B.4: $< n >$ and moments for $|\eta| < 2.4, p_{\mathrm{T}} > 0.1\,\mathrm{GeV}$ and nch>=2.

Center-of-mass energy	0.9 TeV	2.76 TeV	7.0 TeV
$< n >$	$17.6 \pm 0.0^{+0.5}_{-0.4}$	$23.1 \pm 0.0^{+0.6}_{-0.6}$	$28.5 \pm 0.0^{+0.7}_{-0.7}$
D_2	$12.8 \pm 0.0^{+0.3}_{-0.3}$	$18.3 \pm 0.0^{+0.5}_{-0.5}$	$24.3 \pm 0.0^{+0.6}_{-0.6}$
C_2	$1.534 \pm 0.001^{+0.003}_{-0.003}$	$1.626 \pm 0.000^{+0.001}_{-0.001}$	$1.728 \pm 0.000^{+0.001}_{-0.002}$
C_3	$3.144 \pm 0.006^{+0.012}_{-0.012}$	$3.597 \pm 0.002^{+0.008}_{-0.009}$	$4.121 \pm 0.003^{+0.008}_{-0.013}$
C_4	$7.893 \pm 0.029^{+0.043}_{-0.046}$	$9.774 \pm 0.014^{+0.053}_{-0.050}$	$12.072 \pm 0.022^{+0.043}_{-0.067}$
C_5	$23.021 \pm 0.149^{+0.153}_{-0.187}$	$30.835 \pm 0.084^{+0.328}_{-0.279}$	$40.945 \pm 0.140^{+0.243}_{-0.374}$
F_2	$1.477 \pm 0.001^{+0.002}_{-0.002}$	$1.582 \pm 0.000^{+0.001}_{-0.001}$	$1.693 \pm 0.000^{+0.001}_{-0.001}$
F_3	$2.889 \pm 0.006^{+0.010}_{-0.010}$	$3.390 \pm 0.002^{+0.008}_{-0.008}$	$3.942 \pm 0.003^{+0.007}_{-0.008}$
F_4	$6.873 \pm 0.027^{+0.034}_{-0.034}$	$8.874 \pm 0.013^{+0.053}_{-0.046}$	$11.227 \pm 0.021^{+0.034}_{-0.044}$
F_5	$18.870 \pm 0.133^{+0.110}_{-0.113}$	$26.840 \pm 0.078^{+0.312}_{-0.241}$	$36.881 \pm 0.133^{+0.174}_{-0.253}$
K_2	$0.477 \pm 0.001^{+0.002}_{-0.002}$	$0.582 \pm 0.000^{+0.001}_{-0.001}$	$0.693 \pm 0.000^{+0.001}_{-0.001}$
K_3	$0.458 \pm 0.003^{+0.003}_{-0.003}$	$0.642 \pm 0.001^{+0.006}_{-0.005}$	$0.862 \pm 0.002^{+0.003}_{-0.004}$
K_4	$0.497 \pm 0.009^{+0.005}_{-0.006}$	$0.792 \pm 0.005^{+0.023}_{-0.017}$	$1.177 \pm 0.009^{+0.009}_{-0.016}$
K_5	$0.440 \pm 0.032^{+0.036}_{-0.025}$	$0.801 \pm 0.023^{+0.083}_{-0.051}$	$1.256 \pm 0.040^{+0.057}_{-0.066}$
H_2	$0.323 \pm 0.000^{+0.001}_{-0.001}$	$0.368 \pm 0.000^{+0.000}_{-0.000}$	$0.409 \pm 0.000^{+0.000}_{-0.000}$
H_3	$0.158 \pm 0.001^{+0.001}_{-0.001}$	$0.190 \pm 0.000^{+0.001}_{-0.001}$	$0.219 \pm 0.000^{+0.000}_{-0.001}$
H_4	$0.072 \pm 0.001^{+0.001}_{-0.001}$	$0.089 \pm 0.000^{+0.002}_{-0.001}$	$0.105 \pm 0.001^{+0.001}_{-0.001}$
H_5	$0.023 \pm 0.002^{+0.002}_{-0.001}$	$0.030 \pm 0.001^{+0.003}_{-0.002}$	$0.034 \pm 0.001^{+0.002}_{-0.002}$

Appendix C

Distributions of the different systematic studies

This appendix shows the relative size of several systematic uncertainties on P_n, which are detailed in section 5.7 for C.1 and in section 6.5.2 for C.2 .

C.1 For NSD events

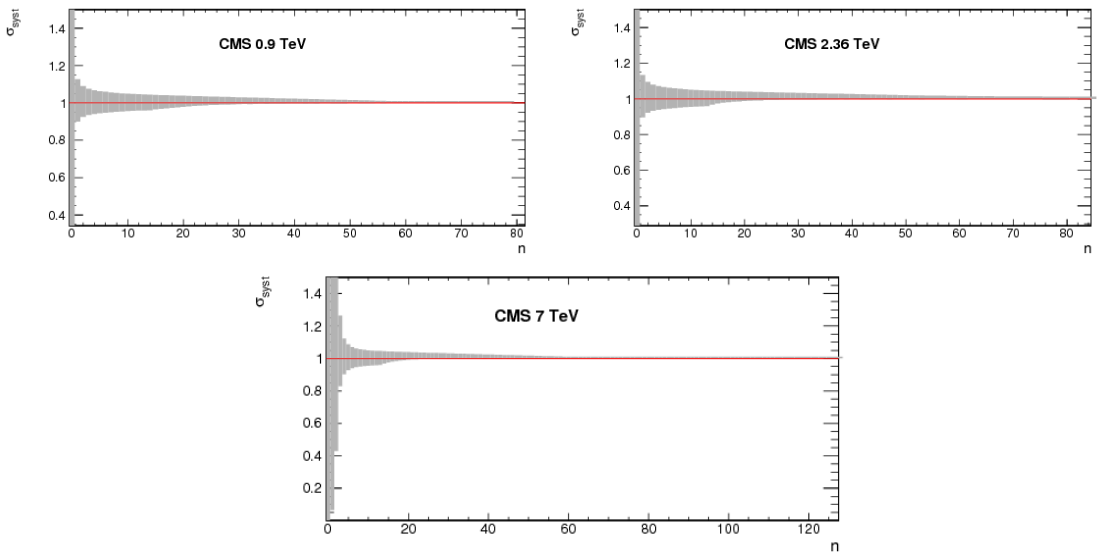

Figure C.1: The relative size of the systematic uncertainty on P_n due the trigger and event selection criteria for the measured multiplicities at $|\eta| < 2.4$ at $\sqrt{s}$=0.9, 2.36 and 7.0 TeV. The size of the grey band indicates the systematic error on P_n (in percents) for each multiplicity.

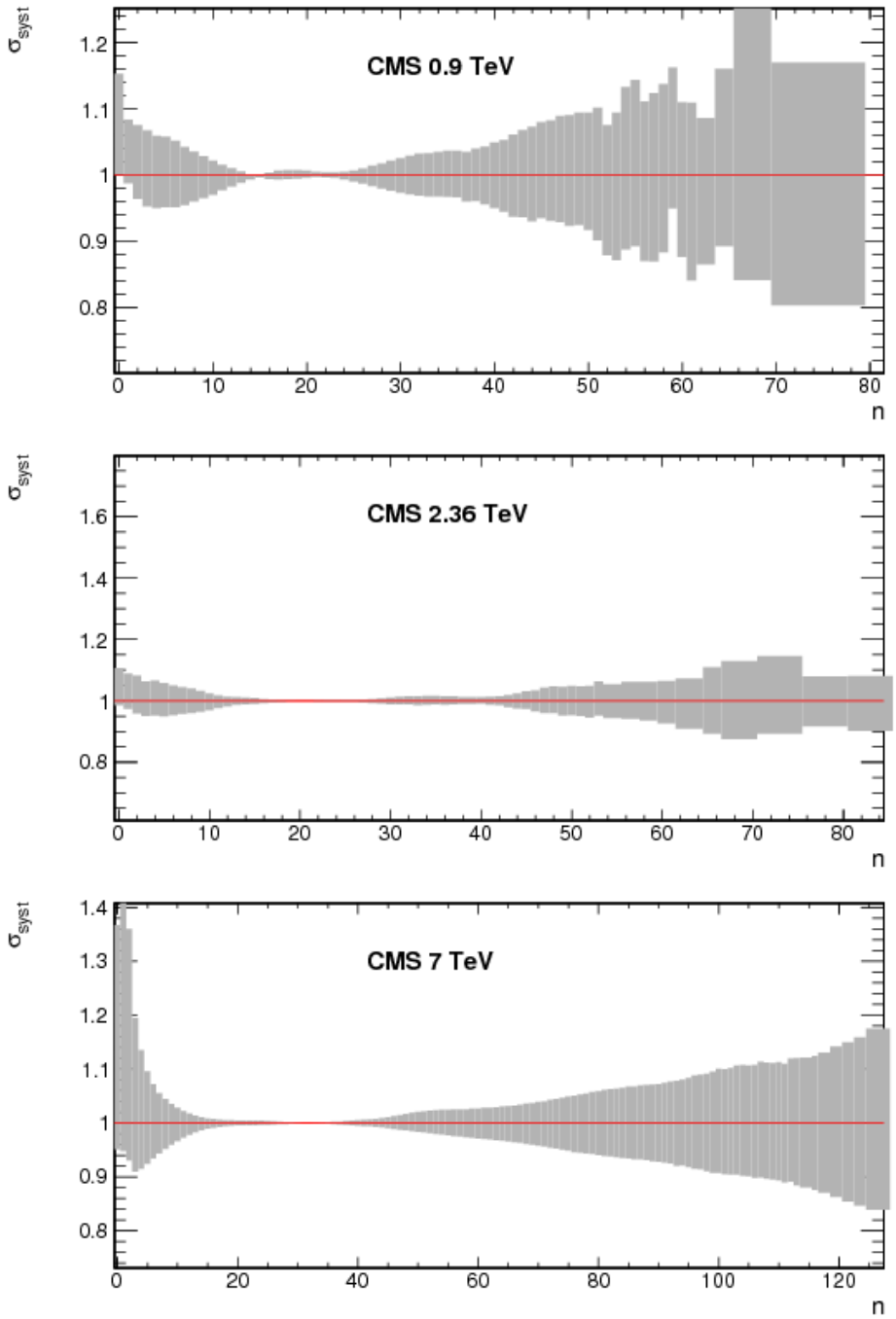

Figure C.2: The relative size of the systematic uncertainty due the mismodeling of tracking efficiencies and secondaries/fakes for the measured multiplicities at $|\eta| < 2.4$ at $\sqrt{s}$=0.9, 2.36 and 7.0 TeV. The size of the grey band indicates the systematic error on P_n (in percents) for each multiplicity.

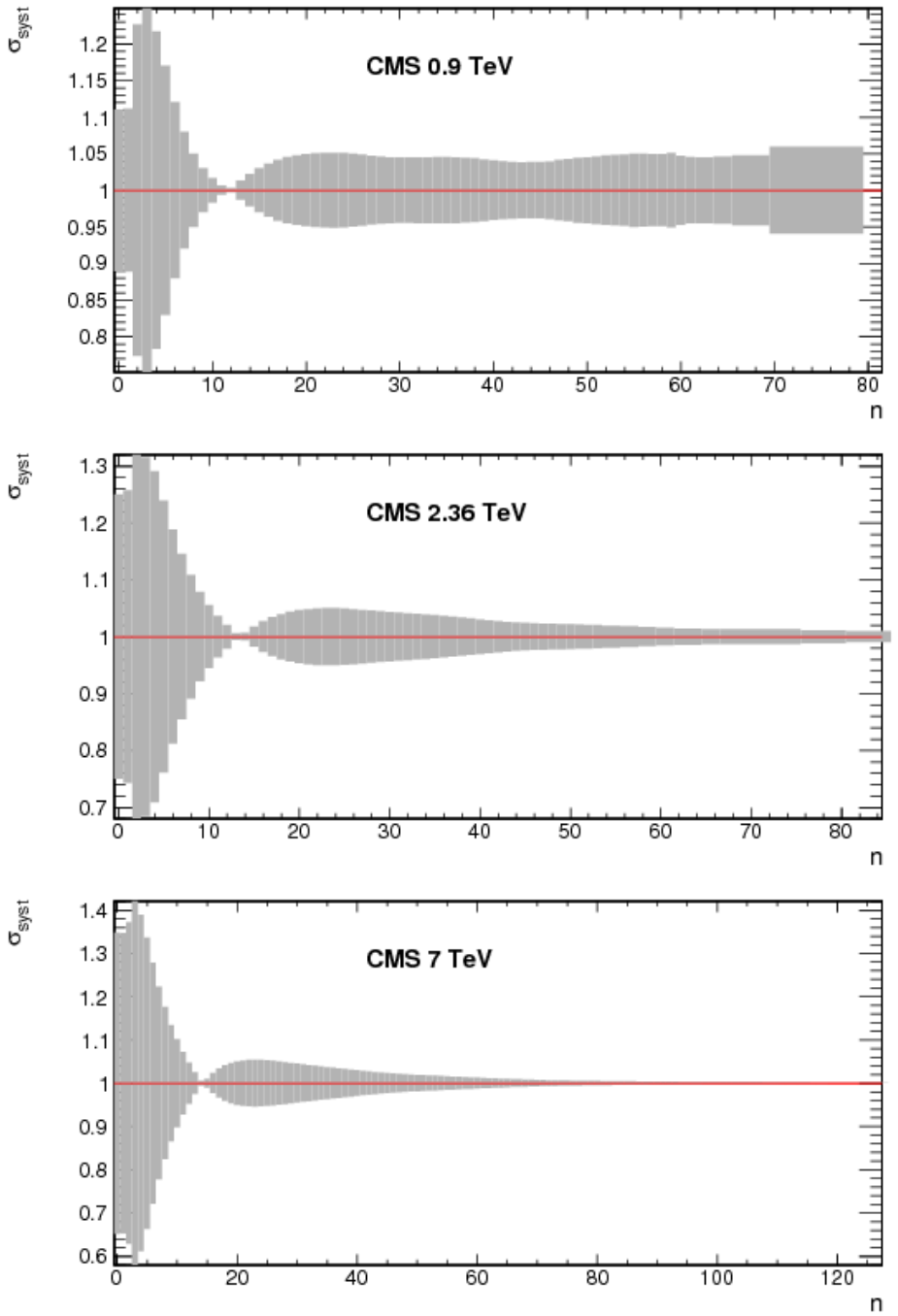

Figure C.3: The relative size of the systematic uncertainty due the mismodeling of the fraction and shape of the subtracted SD events for the measured multiplicities at $|\eta| < 2.4$ at $\sqrt{s}$=0.9, 2.36 and 7.0 TeV. The size of the grey band indicates the systematic error on P_n (in percents) for each multiplicity.

C.2 For Inelastic events

Cross-checks and systematics of the multiplicity distribution

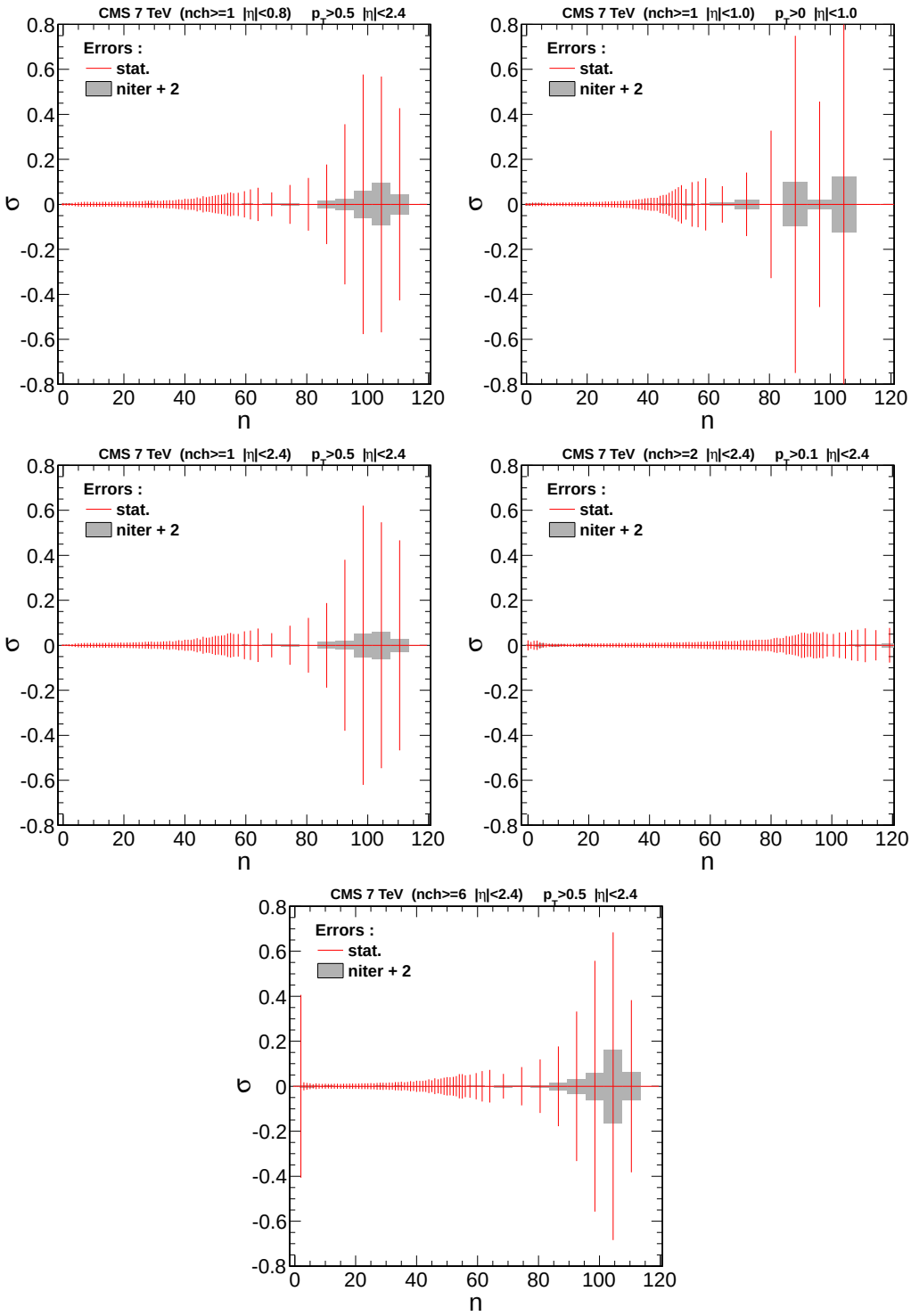

Figure C.4: Effect of adding 2 iterations to the unfolding on the fully corrected multiplicity distribution for several central selections and track acceptances.

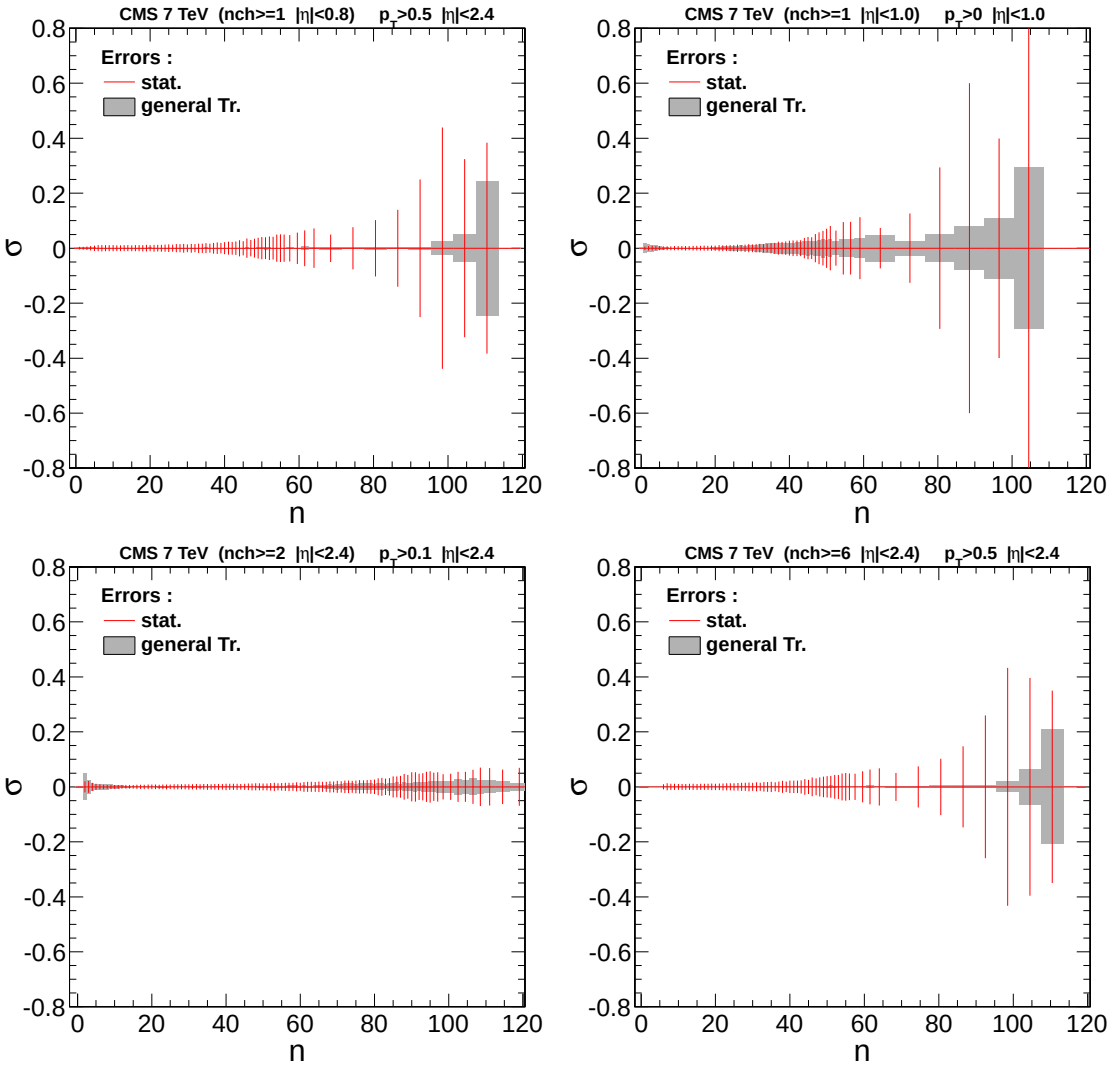

Figure C.5: Effect of not taking the mixed tracking algorithms on the fully corrected multiplicity distribution for several central selections and track acceptances.

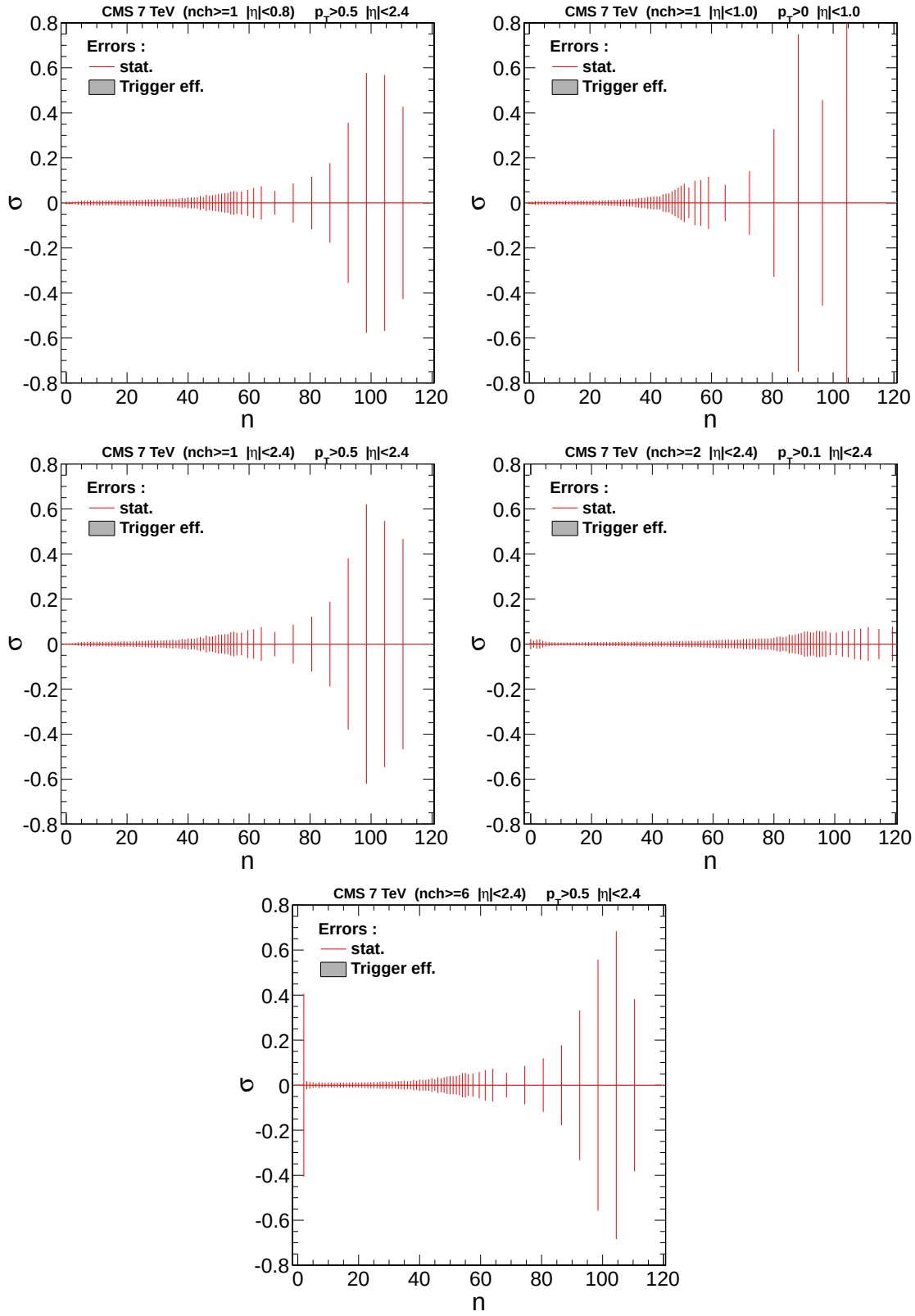

Figure C.6: Effect of 0-bias statistical errors when correcting for the trigger event selection on the fully corrected multiplicity distribution for several central selections and track acceptances.

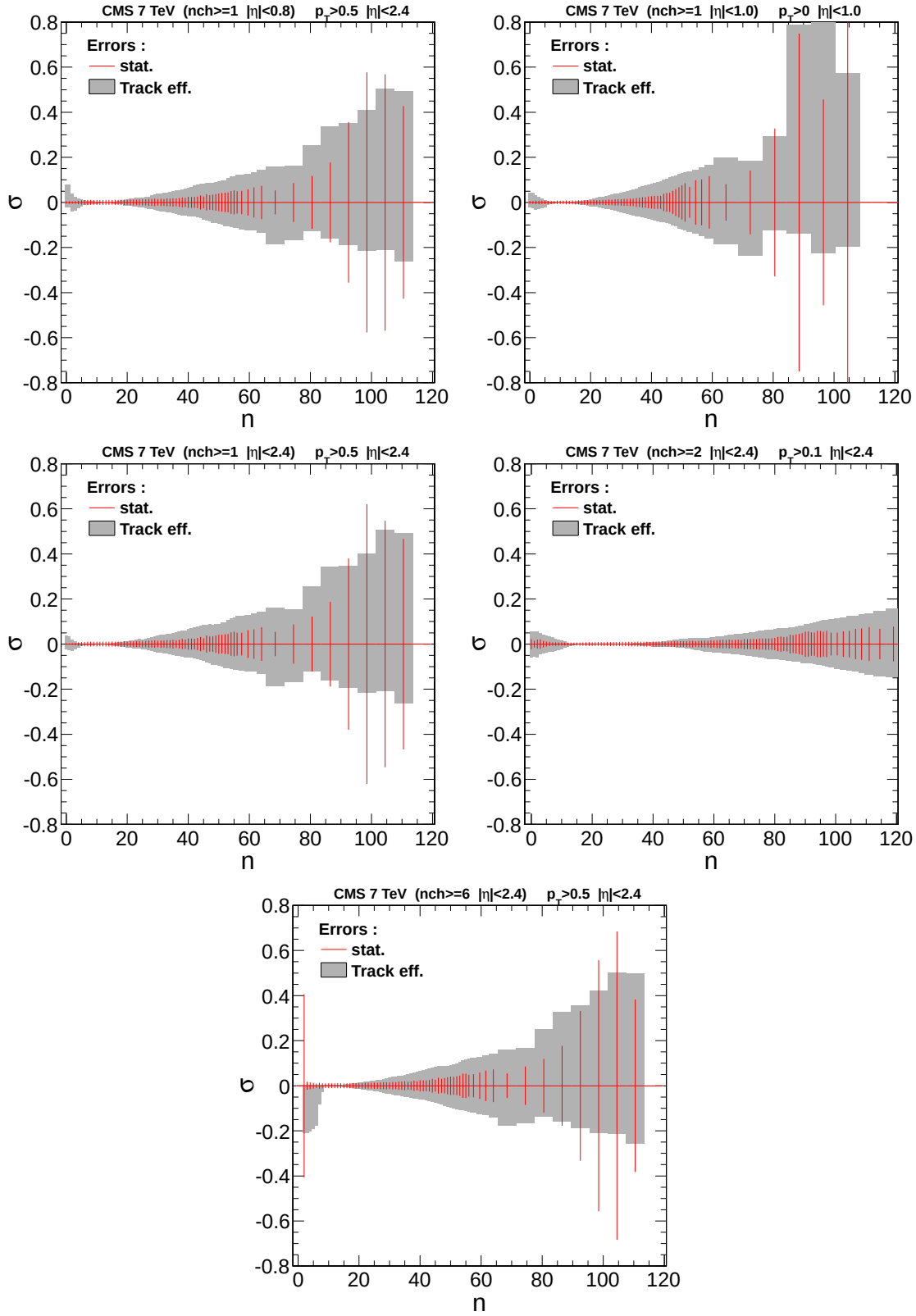

Figure C.7: Effect of tracking inefficiencies misreconstruction in Monte-Carlo on the fully corrected multiplicity distribution for several central selections and track acceptances.

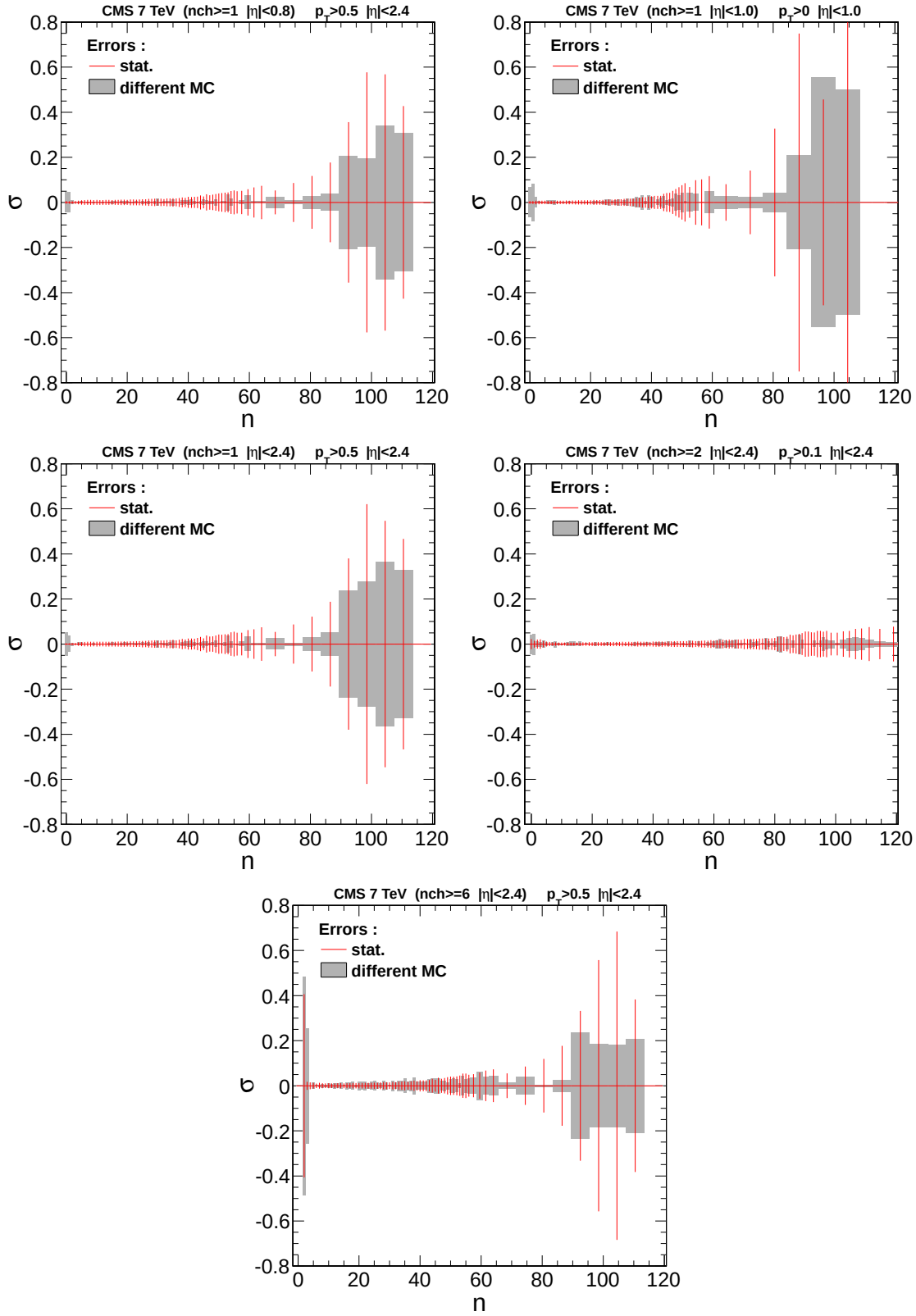

Figure C.8: Effect of different Monte-Carlo description on the fully corrected multiplicity distribution for several central selections and track acceptances.

Systematics of the mean-p_{T} vs multiplicity profile

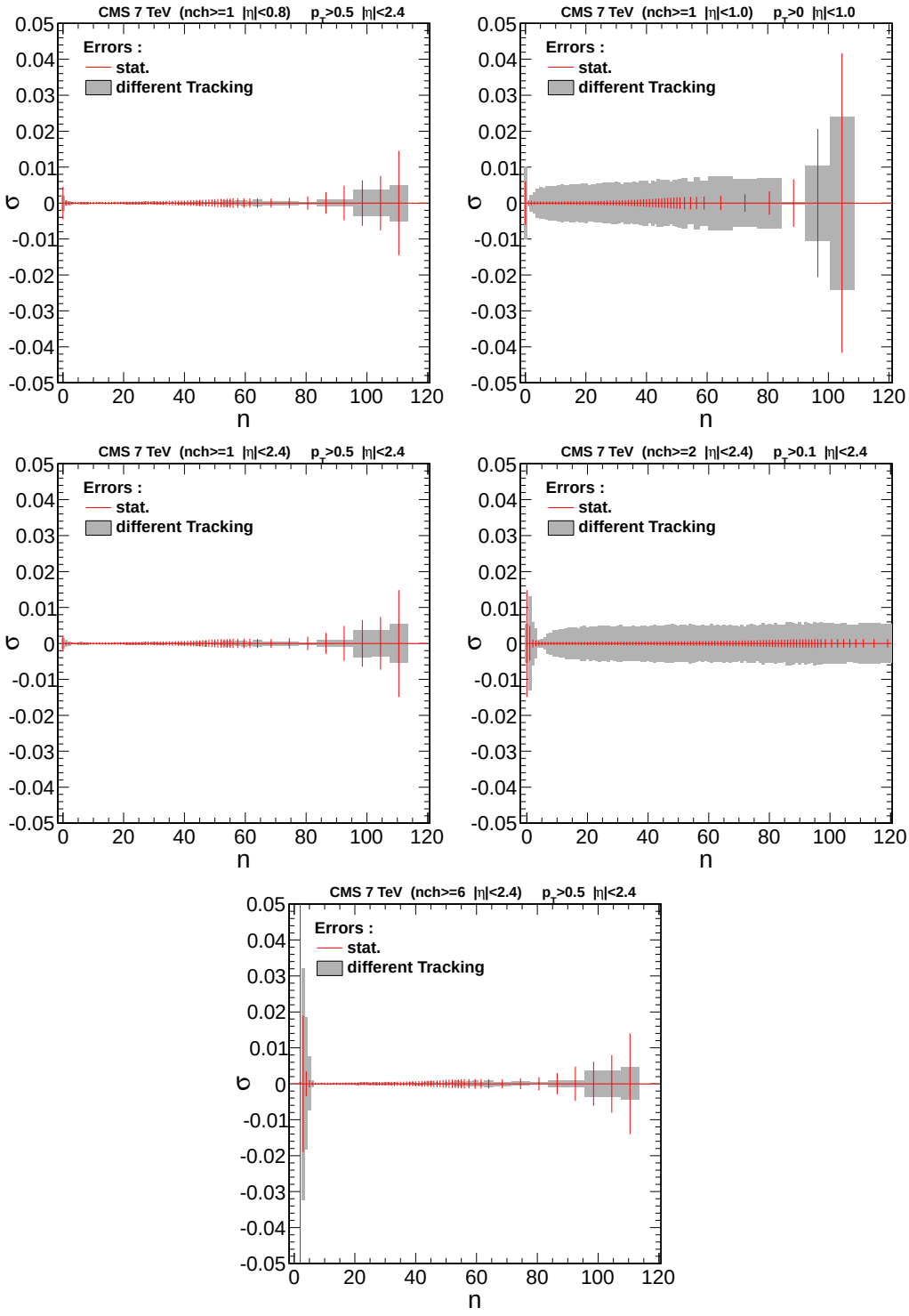

Figure C.9: Effect of not taking the mixed tracking algorithms on the fully corrected mean p_{T} versus multiplicity distribution for several central selections and track acceptances.

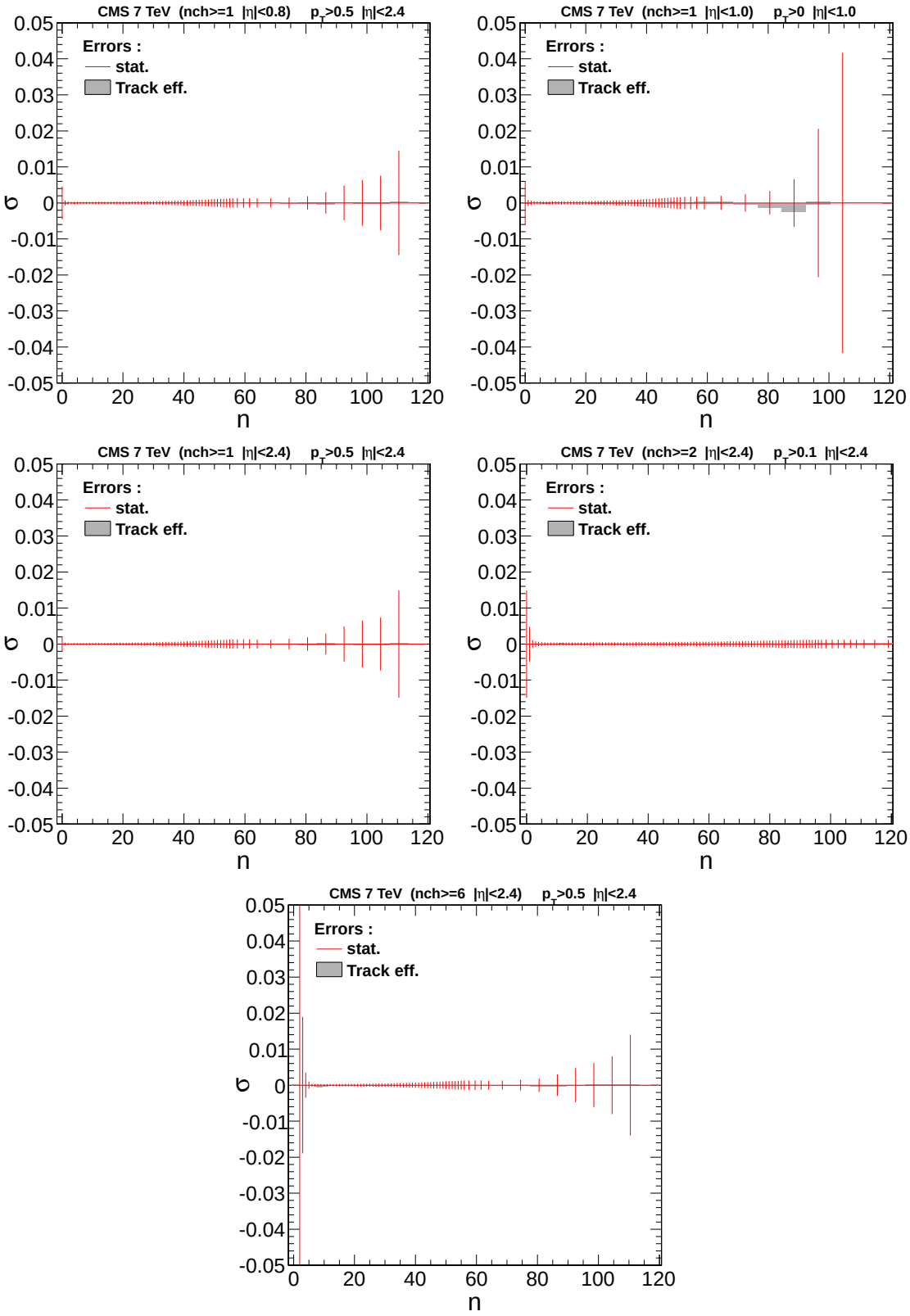

Figure C.10: Effect of tracking inefficiencies misreconstruction in Monte-Carlo on the fully corrected mean p_{T} versus multiplicity distribution for several central selections and track acceptances.

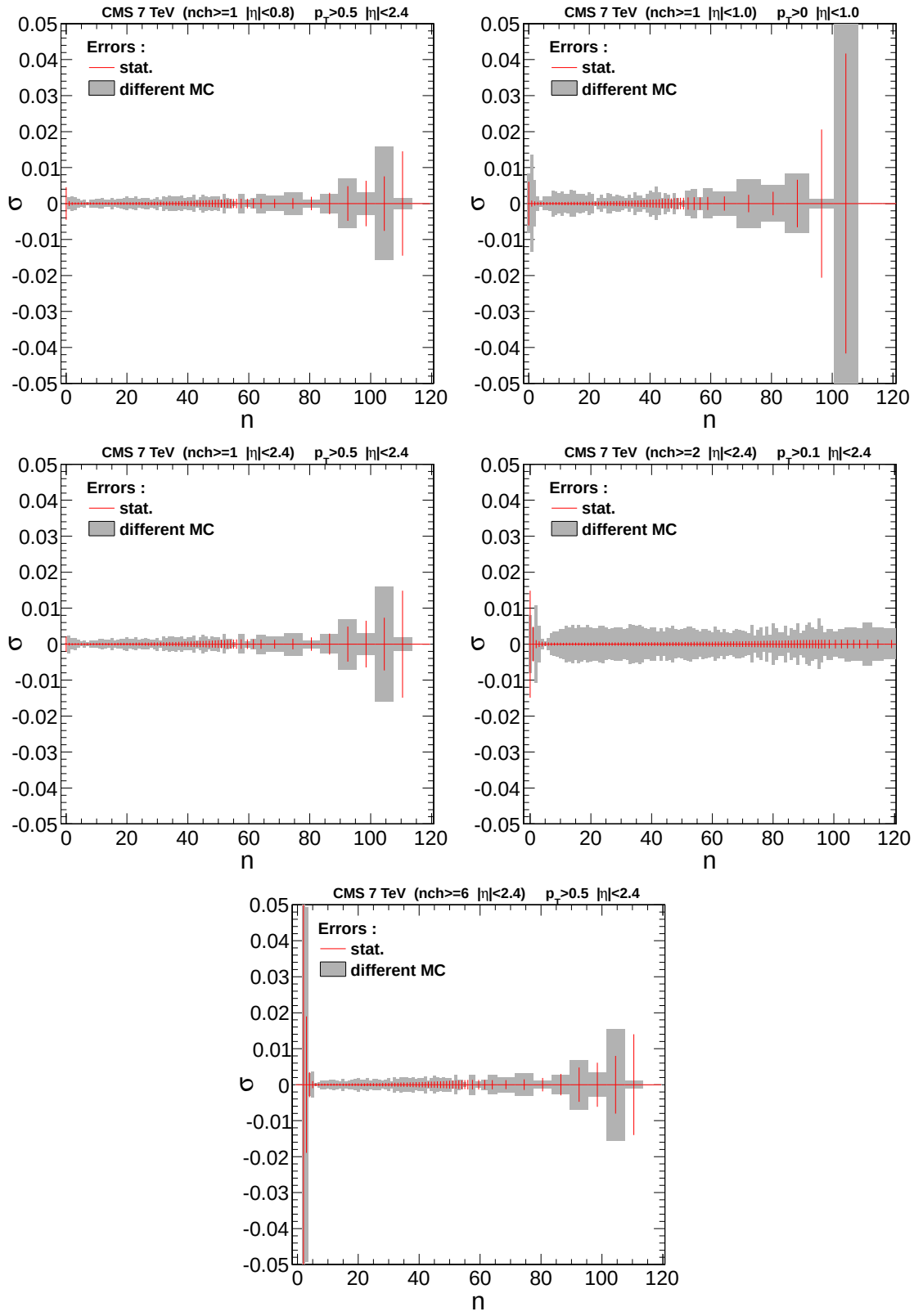

Figure C.11: Effect of different Monte-Carlo description on the fully corrected mean p_{T} versus multiplicity distribution for several central selections and track acceptances.

Bibliography

[1] E. A. De Wolf, I. M. Dremin, and W. Kittel, "Scaling laws for density correlations and fluctuations in multiparticle dynamics," *Phys. Rept.*, vol. 270, p. 1, 1996.

[2] I. Dremin, "Cumulant and factorial moments in perturbative gluodynamics," *Phys.Lett.*, vol. B313, pp. 209–212, 1993.

[3] R. P. Feynman, "Very high-energy collisions of hadrons," *Phys.Rev.Lett.*, vol. 23, pp. 1415–1417, 1969.

[4] J. F. Grosse-Oetringhaus and K. Reygers, "Charged-particle multiplicity in proton-proton collisions," *J. Phys.*, vol. G37, p. 083001, 2010.

[5] Z. Koba, H. B. Nielsen, and P. Olesen, "Scaling of multiplicity distributions in high-energy hadron collisions," *Nucl. Phys.*, vol. B40, p. 317, 1972.

[6] G. Ekspong, "On Scale Breaking in Multiplicities and a New Empirical Rule," *Conf.Proc.*, vol. C850609, pp. 309–320, 1985.

[7] A. Giovannini and L. Van Hove, "Negative Binomial Multiplicity Distributions in High-Energy Hadron Collisions," *Z.Phys.*, vol. C30, p. 391, 1986.

[8] A. Giovannini and L. Van Hove, "Negative Binomial Properties and Clan Structure in Multiplicity Distributions," *Acta Phys.Polon.*, vol. B19, p. 495, 1988.

[9] C. Fuglesang, "UA5 multiplicity distributions and fits of various functions," in *Multiparticle Dynamics: Festschrift for Léon Van Hove* (A. Giovannini and W. Kittel, eds.), p. 193, Singapore: World Scientific, 1990.

[10] A. Giovannini, S. Lupia, and R. Ugoccioni, "Common origin of the shoulder structure and of the oscillations of moments in multiplicity distributions in e+ e- annihilations," 1996.

[11] A. Giovannini and R. Ugoccioni, "Possible scenarios for soft and semihard components structure in central hadron hadron collisions in the TeV region: Pseudorapidity intervals," *Phys.Rev.*, vol. D60, p. 074027, 1999.

[12] T. Gaisser, F. Halzen, and A. D. Martin, "Multiplicities in a QCD motivated description of very high-energy particle interactions.," *Phys.Lett.*, vol. B166, p. 219, 1986.

[13] T. Sjostrand and M. van Zijl, "A Multiple Interaction Model for the Event Structure in Hadron Collisions," *Phys.Rev.*, vol. D36, p. 2019, 1987.

[14] X.-N. Wang, "Role of multiple mini - jets in high-energy hadronic reactions," *Phys.Rev.*, vol. D43, pp. 104–112, 1991.

[15] L. Mucibello, "Measurement of the Underlying Event activity with the CMS detector at the LHC," 2012.

[16] I. Dremin and V. Nechitailo, "IPPI model of hadron interactions," *Phys.Rev.*, vol. D70, p. 034005, 2004.

[17] A. Capella *et al.*, "Dual parton model," *Phys. Rept.*, vol. 236, p. 225, 1994.

[18] Y.-J. Lee, "Measurement of the Charged-Hadron Multiplicity in Proton-Proton Collisions at LHC with the CMS Detector," 2011.

[19] Y. Allkofer, C. Amsler, D. Bortoletto, V. Chiochia, L. Cremaldi, *et al.*, "Design and performance of the silicon sensors for the CMS barrel pixel detector," *Nucl.Instrum.Meth.*, vol. A584, pp. 25–41, 2008.

[20] G. Bolla, D. Bortoletto, C. Rott, A. Roy, S. Kwan, *et al.*, "Design and test of pixel sensors for the CMS experiment," *Nucl.Instrum.Meth.*, vol. A461, pp. 182–184, 2001.

[21] H. C. Kastli, M. Barbero, W. Erdmann, C. Hormann, R. Horisberger, *et al.*, "Design and performance of the CMS pixel detector readout chip," *Nucl.Instrum.Meth.*, vol. A565, pp. 188–194, 2006.

[22] A. Dominguez *et al.*, "CMS Technical Design Report for the Pixel Detector Upgrade," 2012.

[23] D. Kotlinski, "Status of the CMS Pixel detector," *JINST*, vol. 4, p. P03019, 2009.

[24] M. Swartz, D. Fehling, G. Giurgiu, P. Maksimovic, and V. Chiochia, "A new technique for the reconstruction, validation, and simulation of hits in the CMS pixel detector," *PoS*, vol. VERTEX2007, p. 035, 2007.

[25] S. Chatrchyan *et al.*, "Commissioning and Performance of the CMS Pixel Tracker with Cosmic Ray Muons," *JINST*, vol. 5, p. T03007, 2010.

[26] B. Henrich and R. Kaufmann, "Lorentz-angle in irradiated silicon," *Nucl.Instrum.Meth.*, vol. A477, pp. 304–307, 2002.

[27] U. Langenegger, "Offline calibrations and performance of the CMS pixel detector," *Nucl.Instrum.Meth.*, vol. A650, pp. 25–29, 2011.

[28] M. Swartz, "A Detailed Simulation of the CMS Pixel Sensor," 2002.

[29] M. Swartz, "CMS pixel simulations," *Nucl.Instrum.Meth.*, vol. A511, pp. 88–91, 2003.

[30] M. Swartz, V. Chiochia, Y. Allkofer, D. Bortoletto, L. Cremaldi, *et al.*, "Observation, modeling, and temperature dependence of doubly peaked electric fields in irradiated silicon pixel sensors," *Nucl.Instrum.Meth.*, vol. A565, pp. 212–220, 2006.

[31] V. Chiochia, M. Swartz, D. Bortoletto, L. Cremaldi, S. Cucciarelli, *et al.*, "Simulation of heavily irradiated silicon pixel sensors and comparison with test beam measurements," *IEEE Trans.Nucl.Sci.*, vol. 52, pp. 1067–1075, 2005.

[32] W. Adam *et al.*, "Stand-alone Cosmic Muon Reconstruction Before Installation of the CMS Silicon Strip Tracker," *JINST*, vol. 4, p. P05004, 2009.

[33] S. Chatrchyan *et al.*, "Alignment of the CMS Silicon Tracker during Commissioning with Cosmic Rays," *JINST*, vol. 5, p. T03009, 2010.

[34] W. Adam, B. Mangano, T. Speer, and T. Todorov, "Track Reconstruction in the CMS tracker," Tech. Rep. CMS-NOTE-2006-041, CERN, Geneva, Dec 2006.

[35] V. Khachatryan *et al.*, "Transverse momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s} = 0.9$ and 2.36 TeV," *JHEP*, vol. 02, p. 041, 2010.

[36] W. Adam, B. Mangano, T. Speer, and T. Todorov, "Track reconstruction in the CMS tracker," Dec 2006.

[37] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, "Adaptive vertex fitting," *J.Phys.*, vol. G34, p. N343, 2007.

[38] T. Miao, N. Leioatts, H. Wenzel, and F. Yumiceva, "Beam Position Determination using Tracks," Tech. Rep. CMS-NOTE-2007-021, CERN, Geneva, Aug 2007.

[39] "Tracking and Primary Vertex Results in First 7 TeV Collisions," Tech. Rep. CMS-PAS-TRK-10-005, CERN, 2010. Geneva, 2010.

[40] V. Khachatryan *et al.*, "CMS Tracking Performance Results from early LHC Operation," *Eur.Phys.J.*, vol. C70, pp. 1165–1192, 2010.

[41] A. A. Affolder *et al.*, "Charged jet evolution and the underlying event in $p\overline{p}$ collisions at 1.8 TeV," *Phys. Rev.*, vol. D65, p. 092002, 2002.

[42] D. Acosta *et al.*, "The underlying event in hard interactions at the Tevatron $\bar{p}p$ collider," *Phys.Rev.*, vol. D70, p. 072002, 2004.

[43] T. Sjostrand and M. van Zijl, "Multiple Parton-parton Interactions in an Impact Parameter Picture," *Phys.Lett.*, vol. B188, p. 149, 1987.

[44] W. Thome *et al.*, "Charged-particle multiplicity distributions in pp collisions at ISR energies," *Nucl. Phys.*, vol. B129, p. 365, 1977.

[45] A. Breakstone *et al.*, "Charged multiplicity distribution in pp interactions at ISR energies," *Phys. Rev.*, vol. D30, p. 528, 1984.

[46] G. J. Alner *et al.*, "Scaling violation favoring high-multiplicity events at 540 GeV center-of-mass energy," *Phys. Lett.*, vol. B138, p. 304, 1984.

[47] R. E. Ansorge *et al.*, "Charged-particle multiplicity distributions at 200 GeV and 900 GeV center-of-mass energy," *Z. Phys.*, vol. C43, p. 357, 1989.

[48] G. Alner *et al.*, "UA5: A general study of proton-antiproton physics at $\sqrt{s} = 546$ GeV," *Phys.Rept.*, vol. 154, pp. 247–383, 1987.

[49] G. Arnison *et al.*, "Charged-particle multiplicity distributions in proton anti-proton collisions at 540 GeV center-of-mass energy," *Phys. Lett.*, vol. B123, p. 108, 1983.

[50] C. Albajar *et al.*, "A study of the general characteristics of $p\overline{p}$ collisions at $\sqrt{s} =$ 0.2 TeV to 0.9 TeV," *Nucl. Phys.*, vol. B335, p. 261, 1990.

[51] T. Alexopoulos *et al.*, "The role of double parton collisions in soft hadron interactions," *Phys. Lett.*, vol. B435, p. 453, 1998.

[52] D. E. Acosta *et al.*, "Soft and hard interactions in $p\overline{p}$ collisions at $\sqrt{s} = 1800\,\text{GeV}$ and 630 GeV," *Phys. Rev.*, vol. D65, p. 072005, 2002.

[53] W. Kittel and E. A. De Wolf, *Soft multihadron dynamics*. Hackensack, USA: World Scientific, 2005.

[54] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP*, vol. 05, p. 026, 2006.

[55] R. Field, "Studying the 'underlying event' at CDF and the LHC," pp. 12–31, 2009.

[56] G. D'Agostini, "A multidimensional unfolding method based on Bayes' theorem," *Nucl. Instrum. Meth.*, vol. A362, p. 487, 1995.

[57] P. Bartalini *et al.*, "Multi-parton interactions and underlying events from Tevatron to LHC," *Proceedings of the 38th International Symposium on Multiparticle Dynamics*, 2008.

[58] A. Moraes, C. Buttar, and I. Dawson, "Prediction for minimum bias and the underlying event at LHC energies," *Eur. Phys. J.*, vol. C50, p. 435, 2007.

[59] A. Buckley *et al.*, "Systematic event generator tuning for the LHC," *Eur. Phys. J.*, vol. C65, p. 331, 2010.

[60] P. Z. Skands, "The Perugia Tunes," 2010.

[61] R. Engel, "Photoproduction within the two component dual parton model: amplitudes and cross-sections," *Z. Phys.*, vol. C66, p. 203, 1995.

[62] R. Engel and J. Ranft, "Hadronic photon-photon interactions at high-energies," *Phys. Rev.*, vol. D54, p. 4244, 1996.

[63] W. D. Walker, "Multiparton interactions and hadron structure," *Phys. Rev.*, vol. D69, p. 034007, 2004.

[64] S. G. Matinyan and W. D. Walker, "Multiplicity distribution and mechanisms of the high- energy hadron collisions," *Phys. Rev.*, vol. D59, p. 034022, 1999.

[65] G. J. Alner *et al.*, "Scaling violations in multiplicity distributions at 200 GeV and 900 GeV," *Phys. Lett.*, vol. B167, p. 476, 1986.

[66] T. Alexopoulos *et al.*, "Multiplicity dependence of the transverse momentum spectrum for centrally produced hadrons in $p\overline{p}$ collisions at $\sqrt{s} = 1.8\,\text{TeV}$," *Phys. Rev. Lett.*, vol. 60, p. 1622, 1988.

[67] A. Capella and A. Krzywicki, "On the correlation between transverse momentum and multiplicity of secondaries at collider energy," *Phys. Rev.*, vol. D29, p. 1007, 1984.

[68] R. E. Ansorge *et al.*, "Charged particle correlations in $p\overline{p}$ collisions at c.m. energies of 200 GeV, 546 GeV and 900 GeV," *Z. Phys.*, vol. C37, p. 191, 1988.

[69] E. Gotsman *et al.*, "QCD motivated approach to soft interactions at high energies: nucleus-nucleus and hadron-nucleus collisions," *Nucl. Phys.*, vol. A842, p. 82, 2010.

[70] A. K. Likhoded, A. V. Luchinsky, and A. A. Novoselov, "Light hadron production in semi-inclusive pp-scattering at LHC," 2010.

[71] A. B. Kaidalov and M. G. Poghosyan, "Predictions of quark-gluon string model for pp at LHC," *Eur. Phys. J.*, vol. C67, p. 397, 2010.

[72] E. Levin and A. H. Rezaeian, "Gluon saturation and inclusive hadron production at LHC," *Phys. Rev.*, vol. D82, p. 014022, 2010.

[73] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A brief introduction to PYTHIA 8.1," *Comput. Phys. Commun.*, vol. 178, p. 852, 2008.

[74] S. Agostinelli *et al.*, "GEANT4: A simulation toolkit," *Nucl. Instrum. Meth.*, vol. A506, p. 250, 2003.

[75] V. Khachatryan *et al.*, "Transverse-momentum and pseudorapidity distributions of charged hadrons in *pp* collisions at $\sqrt{s} = 7$ TeV," *Phys.Rev.Lett.*, vol. 105, p. 022002, 2010.

[76] F. Sikler, "Reconstruction of low p_{T} charged particles with the pixel detector," *CMS NOTE AN-2006/100*, 2006.

[77] F. Sikler, "Study of clustering methods to improve primary vertex finding for collider detectors," *Nucl. Instrum. Meth.*, vol. A621, p. 526, 2010.

[78] F. Becattini, P. Castorina, A. Milov, and H. Satz, "Predictions of hadron abundances in pp collisions at the LHC," *J.Phys.*, vol. G38, p. 025002, 2011.

[79] K. Alpgard *et al.*, "Particle multiplicities in $\mathrm{p\overline{p}}$ Interactions at $\sqrt{s} = 540$ GeV," *Phys. Lett.*, vol. B121, p. 209, 1983.

[80] K. Alpgard *et al.*, "Strange particle production at the CERN SPS collider," *Phys. Lett.*, vol. B115, p. 65, 1982.

[81] 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.*, vol. C68, p. 89, 2010.

[82] K. Aamodt *et al.*, "Charged-particle multiplicity measurement in proton-proton collisions at $\sqrt{s} = 7$ TeV with ALICE at LHC," *Eur. Phys. J.*, vol. C68, p. 345, 2010.

[83] 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.*, vol. B688, pp. 21–42, 2010.

[84] A. Giovannini and R. Ugoccioni, "Soft and semi-hard components structure in multiparticle production in high energy collisions," *Nucl. Phys. Proc. Suppl.*, vol. 71, p. 201, 1999.

[85] A. M. Polyakov, "A similarity hypothesis in the strong interactions. I. Multiple-hadron production in $\mathrm{e^+e^-}$ annihilation," *Sov. Phys. JETP*, vol. 32, p. 296, 1971.

[86] A. M. Polyakov, "A similarity hypothesis in the strong interactions. II. Cascade production of hadrons and their energy distribution associated with $\mathrm{e^+e^-}$ annihilation," *Sov. Phys. JETP*, vol. 33, p. 850, 1971.

[87] S. J. Orfanidis and V. Rittenberg, "On the connection between branching processes and scale invariant field theory for $\mathrm{e^+e^-}$ annihilation," *Phys. Rev.*, vol. D10, p. 2892, 1974.

[88] G. Cohen-Tannoudji and W. Ochs, "Jet model with exclusive parton hadron duality," *Z. Phys.*, vol. C39, p. 513, 1988.

[89] Y. L. Dokshitzer, "Improved QCD treatment of the KNO phenomenon," *Phys. Lett.*, vol. B305, p. 295, 1993.

[90] S. Carius and G. Ingelman, "The Log normal distribution for cascade multiplicities in hadron collisions," *Phys. Lett.*, vol. B252, p. 647, 1990.

[91] R. Szwed and G. Wrochna, "Scaling predictions for multiplicity distributions at lep," *Z. Phys.*, vol. C47, p. 449, 1990.

[92] M. Gazdzicki *et al.*, "Scaling of multiplicity distributions and collision dynamics in e^+e^- and pp interactions," *Mod. Phys. Lett.*, vol. A6, p. 981, 1991.

[93] M. Adamus *et al.*, "Phase space dependence of the multiplicity distribution in π^+p and pp collisions at 250 GeV/c," *Z. Phys.*, vol. C37, p. 215, 1988.

[94] G. J. Alner *et al.*, "An investigation of multiplicity distributions in different pseudorapidity intervals in $p\overline{p}$ reactions at a cms energy of 540 GeV," *Phys. Lett.*, vol. B160, p. 193, 1985.

[95] P. Abreu *et al.*, "Charged particle multiplicity distributions for fixed number of jets in Z^0 hadronic decays," *Z. Phys.*, vol. C56, p. 63, 1992.

[96] P. Abreu *et al.*, "Charged particle multiplicity distributions in restricted rapidity intervals in Z^0 hadronic decays," *Z. Phys.*, vol. C52, p. 271, 1991.

[97] D. Decamp *et al.*, "Measurement of the charged particle multiplicity distribution in hadronic Z decays," *Phys. Lett.*, vol. B273, p. 181, 1991.

[98] D. Buskulic *et al.*, "Measurements of the charged particle multiplicity distribution in restricted rapidity intervals," *Z. Phys.*, vol. C69, p. 15, 1995.

[99] P. D. Acton *et al.*, "A Study of charged-particle multiplicities in hadronic decays of the Z^0," *Z. Phys.*, vol. C53, p. 539, 1992.

[100] M. Adamus *et al.*, "Rapidity dependence of negative and all charged multiplicities in non-diffractive π^+p and pp collisions at 250 GeV/c," *Phys. Lett.*, vol. B177, p. 239, 1986.

[101] V. A. Abramovskii and O. V. Kancheli, "Regge branching and distribution of hadron multiplicity at high energies," *Pisma Zh. Eksp. Teor. Fiz.*, vol. 15, p. 559, 1972.

[102] A. Capella and A. Krzywicki, "Unitarity corrections to short range order: long range rapidity correlations," *Phys. Rev.*, vol. D18, p. 4120, 1978.

[103] P. Aurenche *et al.*, "Multiparticle production in a two-component dual parton model," *Phys. Rev.*, vol. D45, p. 92, 1992.

[104] L. McLerran and M. Praszalowicz, "Saturation and scaling of multiplicity, mean p_{T} and p_{T} distributions from 200 GeV $< \sqrt{s} <$ 7 TeV," *Acta Phys. Polon.*, vol. B41, p. 1917, 2010.

[105] F. Rimondi, "Multiplicity distributions in $\bar{p}p$ interactions at $\sqrt{s} = 1800$ GeV," 1993.

[106] S. Aid *et al.*, "Charged particle multiplicities in deep inelastic scattering at HERA," *Z.Phys.*, vol. C72, pp. 573–592, 1996.

[107] "Pseudorapidity distributions of charged particles in pp collisions at sqrt(s)= 7 TeV with at least one central charged particles," 2011.

[108] S. Chatrchyan *et al.*, "Study of the inclusive production of charged pions, kaons, and protons in *pp* collisions at $\sqrt{s} = 0.9$, 2.76, and 7 TeV," *Eur.Phys.J.*, vol. C72, p. 2164, 2012.

[109] G. Aad *et al.*, "Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC," *New J.Phys.*, vol. 13, p. 053033, 2011.

[110] R. Corke and T. Sjostrand, "Interleaved Parton Showers and Tuning Prospects," *JHEP*, vol. 1103, p. 032, 2011.