Centro Brasileiro de Pesquisas Físicas

Gabriel Oliveira Corrêa

Exploring charged particle distributions in high-energy pp collisions with CMS Open Data

Rio de Janeiro - RJ

June 2022

"EXPLORING CHARGED PARTICLE DISTRIBUTIONS IN HIGH-ENERGY PP COLLISIONS WITH CMS OPEN DATA"

GABRIEL OLIVEIRA CORRÊA

Dissertação de Mestrado em Física apresentada no Centro Brasileiro de Pesquisas Físicas do Ministério da Ciência Tecnologia e Inovação. Fazendo parte da banca examinadora os seguintes professores:

Arthur Marques Moraes

Arthur Marques Moraes – Presidente/Orientador/CBPF

André Asevedo Nepomuceno - UFF

Gilvan Augusto Alves – CBPF

Rio de Janeiro, 23 de junho de 2022.

Gabriel Oliveira Corrêa

Exploring charged particle distributions in high-energy pp collisions with CMS Open Data

Dissertation presented to the Postgraduate Program of the Centro Brasileiro de Pesquisas Físicas as partial requirement for obtaining the title of Master of Physics.

Supervisor:
Arthur Marques Moraes

June 2022

Abstract

Measurements of transverse momentum (p_T), pseudorapidity (η), mean transverse momentum ($\langle p_T \rangle$), and multiplicity distributions contribute to our understanding of the hadron production in high-energy collisions and provide information on the event characterization, essential to the development of future analysis. The charged particle transverse momentum distribution also provides tests of the predictability of quantum chromodynamics. While we can describe the high-p_T distribution using perturbative quantum chromodynamics, we also need to use non-perturbative phenomenological methods to depict the whole spectrum. Non-extensive statistical mechanics supplies one of those methods, where it provides a parametrization that fits the power-law behavior for high-p_T as well as the exponential behavior for low-p_T. Questions arise as to why a non-extensive statistical mechanical distribution successfully describes the experimental data and what are its possible physical implications. Using proton-proton collision data collected by the CMS Experiment and made available through the CERN Open Data Portal, the p_T, η, $\langle p_T \rangle$, and multiplicity distributions of charged hadrons for proton-proton collisions with center-of-mass energies ($\sqrt{s}$) of 0.9, 2.76, and 7 TeV are reproduced in this dissertation. The spectra obtained were then compared to those published by the CMS and ATLAS collaborations. For the pseudorapidity and multiplicity distributions, their behavior measures how close the distributions from the open data are to the published results. For the p_T distributions, the Tsallis distribution is used to fit the data, and comparisons are drawn between the parameters obtained.

Keywords: charged particle production, nonextensive statistical mechanics, CMS Open Data

Resumo

Medidas de distribuições de momentum transverso (p_T), pseudorapidez (η), momentum transverso médio ($\langle p_T \rangle$), e multiplicidade contribuem para o nosso conhecimento sobre a produção de hádrons em colisões de altas energias e fornecem informações sobre a caracterização de eventos, essencial para o desenvolvimento de análises futuras. As distribuições de momentum transverso de partículas carregadas também fornecem testes sobre o poder de previsão da cromodinâmica quântica. Mesmo podendo utilizar a cromodinâmica quântica perturbativa para descrever as interações com alto-p_T, nós também precisamos utilizar métodos fenomenológicos não-perturbativos para reproduzir todo o spectrum de momentum transverso. A mecânica estatística não-extensiva fornece um desses métodos, onde ela oferece uma parametrização que faz o ajuste das distribuições de momentum transverso considerando o comportamento de lei de potência para baixo-p_T e o comportamento exponencial para alto-p_T. Questões surgem sobre o porquê da mecânica estatística não-extensiva descrever com sucesso os dados experimentais e quais são suas possíveis implicações físicas. Utilizando dados de colisões próton-próton coletadas pelo experimento CMS disponibilizadas pelo CMS Open Data Portal, nós reproduzimos distribuições de p_T, η, $\langle p_T \rangle$, e multiplicidade de hádrons carregados para colisões próton-próton com energia de centro de massa ($\sqrt{s}$) de 0.9, 2.76 e 7 TeV. As distribuições obtidas foram comparadas às distribuições publicadas pelas colaborações CMS e ATLAS. Para as distribuições de pseudorapidez e multiplicidade, seus comportamentos medem o quão perto as distribuições dos dados abertos estão dos dados publicados. Para as distribuições de momentum transverso, a distribuição de Tsallis é utilizada para ajustar os dados, e comparações são feitas entre os parâmetros obtidos.

Palavras chave: produção de partículas carregadas, mecânica estatística não-extensiva, CMS Open Data

Acknowledgments

Primeiramente, agradeço à minha família e amigos por todo o apoio, amor e compreensão nessa minha trajetória. Agradeço principalmente à minha mãe, Sandra, por estar sempre do meu lado durante esses dois últimos complicados anos.

Agradeço ao meu orientador, Prof. Arthur, por sua paciência e seus ensinamentos, especialmente por me considerar não apenas como um estudante mas também como um profissional. Seus conselhos me ajudaram a tomar diversas decisões a respeito da minha carreira e os carregarei sempre.

Agradeço aos profissionais do CBPF, professores e funcionários, pela paciência, disponibilidade, e conhecimentos passados durante esses anos de mestrado.

Por fim, agradeço a CAPES pelo apoio financeiro através do fornecimento de uma bolsa de mestrado para a realização dessa pesquisa.

Contents

List of Figures

List of Tables

List of Acronyms

APD Avalanche Photodiodes

ALICE A Large Ion Collider Experiment

AOD Analysis Object Data

ATLAS A Toroidal LHC Apparatus

BG statistics Boltzmann-Gibbs statistics

BPTX Beam Pick-up Timing Experiment

BSC Beam Scintillator Counter

BSM Beyond the Standard Model

CD Central-diffractive

CERN *Conseil Européen pour la Recherche Nucléaire*

CMS Compact Muon Solenoid

CMSSW CMS Software

CSC Cathode Strip Chamber

DD Double-diffractive

DT Drift Tube

EB Electromagnetic Calorimeter Barrel

ECAL Electromagnetic Calorimeter

EE Electromagnetic Calorimeter Endcaps

GEN step Generation step

HB Hadronic Calorimeter Barrel

HCAL Hadronic Calorimeter

HE Hadronic Calorimeter Endcaps

HF Forward Hadronic Calorimeter

HLT High-Level Trigger

HO Outer Hadronic Calorimeter

HPD Hybrid Photodiodes

L1 Level-1 Trigger

LEIR Low Energy Ion Ring

LEP Large Electron Positron Collider

LHC Large Hadron Collider

LHCb Large Hadron Collider Beauty Experiment

MB Minimum Bias

MC Monte Carlo

ND Non-diffractive

NSD Non-single-diffractive

PS Proton Synchrotron

PSB Proton Synchrotron Booster

PV Primary Vertex

pQCD perturbative Quantum Chromodymanics

QCD Quantum Chromodymanics

QED Quantum Electrodynamics

QFT Quantum Field Theory

QGP Quark-Gluon Plasma

RECO step Reconstruction step

RF system Radiofrequency system

RPC Resistive Plate Chamber

SBM Statistical Bootstrap Model

SD Single-diffractive

SIM step Simulation step

SL SuperLayer

SM Standard Model

SPS Super Proton Synchrotron

TEC Tracker EndCaps

TIB Tracker Inner Barrel

TID Tracker Inner Disks

TOB Tracker Outer Barrel

VPT Vacuum Phototriodes

1 | Introduction

After more than a century of development, the Standard Model (SM) of particle physics [1–3] represents the current knowledge about the elementary particles that make the universe and its interaction forces. The SM provides us with great detail the description and the prediction of various high-energy phenomena through quantum field theory (QFT). It also explains many of the current experimental data. Details about the SM in the QFT context are found in various textbooks [4–6].

Fermions (particles with half-integer spin) and bosons (particles with integer spin) make up the particles of the SM. The fermions compose the universe matter, while the bosons mediate their interactions. Exemplifying these particles, we can start with the atoms, which are bound states of fermions (electrons) orbiting a nucleus (made up of hadrons). The hadrons, in turn, are bound states of another kind of fermions, the quarks. Quantum electrodynamics (QED) describes the atom bound state, where virtual photons mediate electrically charged particle interaction. Quantum chromodynamics (QCD), in turn, depict the hadron bound state, where virtual gluons mediate the interaction between color-charged particles. The weak interaction completes the fundamental interactions of particle physics, where it is responsible for radioactive atomic decay. Massive bosons, W and Z, mediate the weak interaction. The weak force provides access for the electrically neutral part of the fermions of the SM, the neutrinos. For a complete interaction model, we need to introduce the gravitational force. However, gravity is not a part of the SM. Since the gravitation interaction is negligible compared to other interactions, we only need to consider it on a macroscopic scale. Additionally, there is no clear quantum picture of gravity. The Higgs boson, a spin-0 particle, completes the SM. It provides a symmetry-breaking mechanism for which all the other particles acquire mass. Table 1.1 presents the twelve fermions that make the SM, and Table 1.2 shows the four fundamental interactions and their respective force mediators.

Even though the SM is a successful model, it still is an incomplete theory. Among its problems, we can cite the lack of explanation of phenomena like the nature of dark matter, the particles' mass origin,

	Leptons				Quarks			
	Particle		$Q/\|e\|$	mass (GeV)	Particle		$Q/\|e\|$	mass (GeV)
First Generation	electron	e^-	-1	0.0005	down	d	$-1/3$	0.003
	electron neutrino	ν_e	0	$< 10^{-9}$	up	u	$+2/3$	0.003
Second Generation	muon	μ^-	-1	0.106	strange	s	$-1/3$	0.1
	muon neutrino	ν_μ	0	$< 10^{-9}$	charm	c	$+2/3$	1.3
Third Generation	tau	τ^-	-1	1.78	bottom	b	$-1/3$	4.5
	tau neutrino	ν_τ	0	$< 10^{-9}$	top	t	$+2/3$	174

Table 1.1: The SM fermions divided into leptons and quarks [4].

Interaction	Strength	Boson		Spin/Parity	Mass (GeV/c^2)
Strong (QCD)	1	Gluon	g	1^-	0
Electromagnetic (QED)	10^{-3}	photon	γ	1^-	0
Weak	10^{-8}	W Boson	$W^\pm$	1^-	80.4
		Z Boson	Z^0	1^-	91.2
Gravity	10^{-37}	?			

Table 1.2: The four known interaction forces [4].

the mechanism of generation of neutrino mass, among others. There are several models developed to circumvent these various problems within the SM. Among them are theories like Supersymmetry, the possibility of extra dimensions, grand unification theory, and string theory, among others.

Although the development of theories beyond the SM is essential, we still have to devote part of our attention to better describing the phenomena present within the SM, which requires precision testing. A detailed understanding of QCD is among these needs.

From an experimental viewpoint, we use particle accelerators as a source of information to develop high-energy physics theories. One issue is how to model the interaction between partons (elementary particles that compose hadrons: quarks, anti-quarks, and gluons) from the QCD. Therefore, we consider the scattering of hadrons in high-energy colliders such as the Large Hadron Collider (LHC) [7], where we aim to describe the hadron-hadron interaction from a partonic model [8]. Ideally, one would use QCD to explain all the processes that involve strong interaction. But that is not possible, as we cannot fully describe hadron-hadron scattering from the first principles of the interaction Lagrangian.

QCD can successfully describe the parton scattering involving high transverse momentum values [8–10]. However, a problem arises when describing the low-p_T part of the spectrum. We encounter this problem because on this scale, the strong coupling constant, $\alpha_s(Q^2)$ ($Q^2 \equiv -q^2 > 0$, where q represents the transferred momentum of the virtual gluon [8]), becomes too large for us to use the QCD perturbative models. As a result, we obtain divergent cross-sections for p_T approaching zero. Since α_s is inversely proportional to Q^2, we will have low values of α_s for high energies, and with that, we can use the perturbative QCD (pQCD). But for lower values of transferred momentum, this is no longer possible.

It is required to combine pQCD for the high-p_T region with phenomenological models for the low-p_T sector to describe the entire p_T spectrum in hadron collisions. Non-extensive statistical mechanics provides one of the phenomenological models, supplying the Tsallis distribution [11–13]. This distribution successfully describes the entire p_T spectrum for high energy collisions [14, 15]. Consequently, questions arise about what physical implications can be implied from the Tsallis fit. Other ways to phenomenologically analyze hadron collisions come from the formulation of Monte Carlo event generators that simulate high-energy proton-proton collisions. These event generators employ models that simulate the collision process on the parton level, using the current knowledge (from QCD theory to experimental data) to describe the hard scattering, the parton shower, and the final hadronization. Two examples of event generators are Pythia [16, 17], which uses the Lund Model [18], and Phojet [19–21], which utilizes the Dual Parton Model [22–24].

The idea of using a statistical approach to explain hadron production in collisions of high energy has been present since the mid-twentieth century [25]. Fermi proposed a method assuming that the formation probabilities of the different types of particles are determined by the statistical weights of these distinct possibilities. Over the next decade, Hagedorn proposed the statistical bootstrap model (SBM) [26]. The SBM follows the Boltzmann-Gibbs statistics and consists of the idea that hadrons make hadrons in an infinite chain. In this model, Hagedorn finds a critical temperature $T_0 \approx 160$ MeV that would later be interpreted as the temperature at which the bound states of hadrons break down and form a new phase of matter, the quark-gluon plasma (QGP) [27, 28]. Soon after, Hagedorn arrived at an empirical formula to describe the hadron production [28]. This formula accounted for the spectrum's exponential behavior for low-p_T and power-law behavior for high-p_T.

Tsallis' approach arises from a generalization of Boltzmann's statistics, called non-extensive statistical mechanics. The non-extensive parameter q characterizes this approach, giving us a measurement of how far the experimental data diverges from the Boltzmann theory. For $q \to 1$, we return to the Boltzmann statistics. This model has been efficient in the description of high-energy phenomena, as well as other fields. An extensive and detailed bibliography on the subject is found in [29]. As stated earlier, non-extensive statistical mechanics emerges as an efficient phenomenological model to describe the charged hadron production in particle accelerators. More recently, the ALICE [30] and CMS [31] experiments at the LHC have published results [32–37] using Tsallis statistics to fit the data for the p_T spectrum of hadrons at high energies.

Besides the transverse momentum distributions, other distributions, such as pseudorapidity, multiplicity, and $\langle p_T \rangle$ distributions, help in the collision characterization. This characterization is essential for the understanding of standalone collisions. These distributions provide the experimental behavior of collision properties necessary for the Monte Carlo model formulation. Thus, with experimental results guiding simulation models, these distributions can provide insights regarding partonic interactions. The data for these analyses are collected in a setting with low pile-up[1] and with triggers set to accept collisions with as with minimum bias as possible. These measurements are among the first made by experimental collaborations and are present extensively in the literature [34–40].

This dissertation uses public data from the CMS Collaboration made available through the CERN Open Data Portal [41] to explore these distributions. In addition to collision data (raw or reconstructed), the CERN Open Data Portal also provides several sets of simulated and derived data from high-energy collisions. These Monte Carlo datasets are essential for the corrections related to the CMS detector. The portal also supplies the CMS software (CMSSW) [42], turning accessible to the entire community the analysis of the data detected by the CMS.

The first objective of this dissertation is to reproduce the distributions of charged particles produced in proton-proton collisions using open data made available by the CMS Experiment for center-of-mass energies of 0.9, 2.76, and 7 TeV. Among them are the distributions of multiplicity, average transverse momentum, pseudorapidity, and transverse momentum. Additionally, this dissertation compares the distributions measured with open data and the results published by the CMS and ATLAS collaborations. This comparison provides the first indication of the possibility of reproducing publishable results using data from the CERN Open Data Portal. After analyzing the fidelity of the distributions from open data, this dissertation focuses on studying the transverse momentum spectrum, where the distributions are analyzed using parametrizations from non-extensive statistical mechanics. Finally, the measured distributions are compared to distributions from Monte Carlo tunings, where details about why this kind of analysis should continue to be carried out are outlined.

[1]Since the LHC collides bunches of protons instead of single protons, multiple collisions can occur in the interaction points. Thus, pile-up interactions are the interactions that are recorded simultaneously inside the CMS detector. As for the low pile-up setting, the experiment aspires to obtain a dataset with only one collision per event.

This dissertation is organized as follows: Chapter 2 highlights general aspects of QCD along with hadron-hadron interactions. Specifically, the scenario for high-pt collisions is displayed. Chapter 3 introduces the theory of non-extensive statistical mechanics and its formulation to describe the hadron production in particle accelerators. Chapter 4 presents the CMS detectors and all the subdetectors relevant to this dissertation. It also shows the CERN Open Data Portal, its datasets, and the software necessary for data analysis. Chapter 5 exhibits the simulated and collision datasets used in this dissertation. It also displays the selection cuts applied to these datasets. Chapter 6 indicates the corrections necessary to obtain the final distributions. Chapter 7 presents the results of this analysis. Finally, Chapter 8 draws the dissertation's conclusions. The calculations and equations introduced in this dissertation follow the kinematic variable definitions present in Appendix A.

2 | Hadron Interactions

2.1 Quantum Chromodynamics

We can separate the particles that make up the SM between fermions and interacting bosons. Specifically regarding QCD, we consider only quarks and gluons, as they are the only fundamental particles within the SM that carry the color charge. We group quarks and gluons into a new category of particles called partons.

At the beginning of the 20th century, there was still no knowledge about the parton existence, only hadrons, which are non-elementary particles composed of partons. Until the early 1930s, the known hadrons were just the proton and the neutron, discovered through experiments with the atomic nucleus [43, 44]. With the advancement of technology over the years, several experiments were carried out analyzing the interaction of (supposedly) elementary particles at increasingly higher energies, leading to several other hadron discoveries.

A new theory was formulated from the discovery of these new hadrons. This theory was composed of particles called baryons (among them are the proton and the neutron) that interacted through the exchange of other particles, later called mesons [45]. Baryons and mesons are two subcategories of hadrons. Even with the failure of this theory to describe the strong interaction, it provided a basis for the model that would follow, as it already used the concept of symmetries in group theory to formulate a model for the strong interaction.

The quark concept emerged in the early 1960s [46] from the need to correctly describe the strong interaction experimental data. The data indicated the conservation of two quantum numbers, labeled as the third component of isospin and the hypercharge. When the known hadrons were organized into a graph taking these two quantum numbers into account, the results looked like the SU(3) weight diagrams. We can obtain all SU(3) representations as subrepresentations of tensor products of representations 3 and 3^*. Thus, a proposal emerged stating that the fundamental representation of the

SU(3) should correspond to the fundamental particles in hadron theory. With this, we obtained a distinction between baryons and mesons. Baryons are composed of three quarks, while mesons are formed by a quark and an antiquark.

Since quarks are fermions, they must obey the Fermi exclusion principle. For the quark model to comply with this principle, it was necessary to introduce eigenstates that differentiate the quarks within a hadron. This new quantum number was labeled as color. A quark is described by specifying its flavor (u, d, s, c, b ou t) and its color (R, G, or B for red, green, or blue). Hadrons are postulated as colorless (the sum of the colors results in white), that is, they contain equal mixtures of R, G, and B. What in this context we call color is just a label for the orthogonal eigenstates in the color space given by SU(3). We describe a quark q as a triplet in this space given by

$$q = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} = \begin{pmatrix} q_R \\ q_G \\ q_B \end{pmatrix} \qquad (q = u,d,s,c,t,b). \tag{2.1}$$

Eight massless gluons mediate the strong interaction, corresponding to the eight SU(3) gauge symmetry generators. Unlike previous attempts to describe the strong interaction from an approximate SU(3) flavor symmetry, we are now working with a theory with an exact SU(3) color symmetry. More details can be found in several textbooks [5, 6, 47].

2.2 Perturbative QCD

An evolution of this theory is necessary so that the reproduction of experimental results is possible, where we need to go to the scenario in which particles can be scattered, created, or destroyed. The measurable experimental parameters of quantum mechanics are given by probabilities. In the Schrödinger representation, the measurements we want to predict are given by $|\left\langle f; t_f \middle| i; t_i \right\rangle|^2$, where $|i; t_i\rangle$ is the initial state at a time t_i, and $\left\langle f; t_f \right|$ is the final state we are interested in at a time t_f. We now move to the quantum field theory, where we interpret coordinates and conjugate moments as operators in the Heisenberg representation. For the case where we evolve momentum eigenstates from a time $t = -\infty$ to a time $t = +\infty$, we call the time evolution operator the S matrix (or scattering matrix). The S matrix is defined as [5]

$$\langle f|S|i\rangle_{\text{Heisenberg}} = \langle f;t_f|i;t_i\rangle_{\text{Schrödinger}} . \tag{2.2}$$

The S matrix has all the information about the evolution of the initial and final states. It is this matrix that must be calculated to compare with the experimental results. If we want to perform calculations that take interactions into account, we need to move to a scenario where we must compute the matrix S perturbatively.

In QFT, we use the time-ordered perturbation theory to formulate the interaction process between particles mediated through particle exchange. There are different temporal orderings of how the interaction between particles can take place in a given scattering. A Feynman diagram represents the sum over all possible arrangements.

Thus, Feynman diagrams appear as the ideal tools to calculate the S matrix. These diagrams also supply a set of Feynman rules, which provide us with a way to represent the perturbation expansion. It is the Feynman rules that establish the calculations of physical results in the QFT. There are different ways to derive Feynman's rules, which will not be covered here. These rules are found in [5, 6].

In short, it is necessary to use the S matrix to study collisions, and for it to be used successfully in QCD, we need to use perturbation theory. At first, it is not guaranteed that we can perform a perturbative expansion for QCD. As far as we know, the strong coupling constant (α_s) has a value approximately to a unit for the bound states, which is about two orders of magnitude larger than the fine structure constant $\alpha \approx 1/137$ of QED, which in turn allows us to perform a perturbative expansion.

The historical step was noting that α_s depends on the energy we are working on, where we have that its value decreases as energy increases. Thus, quarks behave almost as free particles in high-energy collisions when the approximation between hadrons is small enough.

This phenomenon is called asymptotic freedom, a property that allows us to use perturbative methods at high energies for QCD. The discovery of asymptotic freedom in non-abelian field theories was the main reason for the QCD establishment [48, 49]. Going to higher orders of α_s of a Feynman diagram, fermion and gluon loops related to vacuum polarization will be introduced (for gluons also carry color charge and therefore couple with each other). Figure 2.1 illustrates diagrams representing loops for α_s^2 correction orders.

Figure 2.1: Feynman diagrams for the α_s^2-order corrections of the quark-gluon coupling.

From diagrams like those in Figure 2.1, we obtain an expression for the strong coupling constant at high energies [8]

$$\begin{aligned}\alpha_s(Q^2) &\approx \alpha_s \left\{ 1 - \frac{\alpha_s b_0}{4\pi} \ln\left(\frac{Q^2}{\mu^2}\right) + \left[\frac{\alpha_s b_0}{4\pi} \ln\left(\frac{Q^2}{\mu^2}\right)\right]^2 + \cdots \right\} \\ &\approx \frac{\alpha_s}{1 + \dfrac{\alpha_s b_0}{4\pi} \ln\left(\dfrac{Q^2}{\mu^2}\right)} \equiv \frac{1}{\dfrac{b_0}{4\pi} \ln\left(\dfrac{Q^2}{\Lambda^2}\right)}\end{aligned} \tag{2.3}$$

where $\Lambda^2 = \mu^2 \exp\left(-4\pi/\alpha_s b_0\right)$ is the QCD scale parameter, Q^2 is the transferred momentum, $b_0 = \frac{11}{3}n_c - \frac{2}{3}n_f$, μ^2 is the value of Q^2 in which α_s is measured, n_c is the number of colors, and n_f is the number of quarks flavors.

For $n_c = 3$ and $n_f = 6$, we have $b_0 = 7 > 0$, and so we get the addition of a positive term in the denominator of Equation 2.3. Therefore, $\alpha_s(Q^2) \to 0$ when $Q^2 \to \infty$. Thus, this is why quarks and gluons appear as almost free particles when observed at high energies. Asymptotic freedom is an essential ingredient for us to study the partons present in the hadronic structure perturbatively.

We can also note that $\alpha_s(Q^2) \to \infty$ when $Q^2 \to \Lambda^2$, and therefore we cannot use perturbation theory for small values of Q^2. Leaving the momentum space to the coordinate space, we get that

$$\alpha_s(r) = \frac{1}{\dfrac{b_0}{2\pi} \ln\left(\dfrac{1}{\Lambda r}\right)}, \tag{2.4}$$

and thus we can see that the coupling becomes stronger as the separation between the quarks increases. This phenomenon is thought to be the effect that creates the confinement of quarks within hadrons, and consequently, we do not observe free quarks. As hadrons have a size of approximately 1 fm, it is defined that $\Lambda \approx 0.2$ GeV.

2.3 Hadron-hadron scattering

From a hadron collider standpoint, like the LHC, all processes are connected in some way to the interactions of quarks and gluons, whether they are proton-proton collisions or heavy-ion collisions. It is possible to study the mechanisms of QCD through collisions at high energies by thinking of the hadrons present in these collisions as parton clusters [8, 9]. We can separate hadron scatterings in high-energy experiments into two categories, elastic and inelastic scatterings.

Both hadrons present in the collision maintain their shape and are not fragmented to form new hadrons on elastic scatterings. The elastic collision of two hadrons A and B can be represented as $A + B \rightarrow A + B$. This process is only likely to happen if there is small momentum transferred in the interaction.

The fragmentation of one or both hadrons present in the collision characterizes an inelastic scattering. These events are categorized as diffractive (single or double diffraction) or non-diffractive processes. By diffraction, we can make an analogy with light diffraction, where it is possible to characterize the structure of obstacles through the diffraction pattern. In high-energy physics, we analyze hadrons through diffraction, making it possible to study their composition.

Through the excitation of the partonic structure of the hadrons, new colorless hadrons, called fragments, are formed. Fragments carry most of the momentum fraction of the primary particle and manifest as high-energy jets in the forward or backward (or both) directions along the hadron propagation axis. A jet is composed of few high-energy particles produced by the hadronization of a parton.

Single-diffractive scatterings are the processes in which a singular hadron is fragmented. In this type of process, the energy of the center of mass is much greater than the momentum transferred in the interaction and results in particle jets being emitted either forwards or backward. These processes are represented by $A + B \rightarrow A + X$ or $A + B \rightarrow X + B$, where X represents all other particles present in the final state.

Double-diffractive scatterings are the processes in which both hadrons are fragmented. In this process, jets are emitted in both directions of the collision axis. They are represented as $A + B \rightarrow X_1 + X_2$.

The rest of the possible events are classified as non-diffractive and are represented by the generic process $A+B \to X$. In most non-diffractive inelastic interactions, the partons of A and B decelerate and combine to produce new hadrons with low momentum that fill the central region of the collision. The remaining partons of A and B keep moving in the collision direction and eventually recombine into other hadrons. The momentum transferred in the parton interactions in the central region is small. The transverse momentum (p_T) of the final state is only a small fraction of the interacting hadrons' total momentum (p), not exceeding 2 GeV. This type of collision is classified as soft. Figure 2.2 illustrates the possible types of hadron scattering mediated by a Pomeron[1] $\mathbb{P}$.

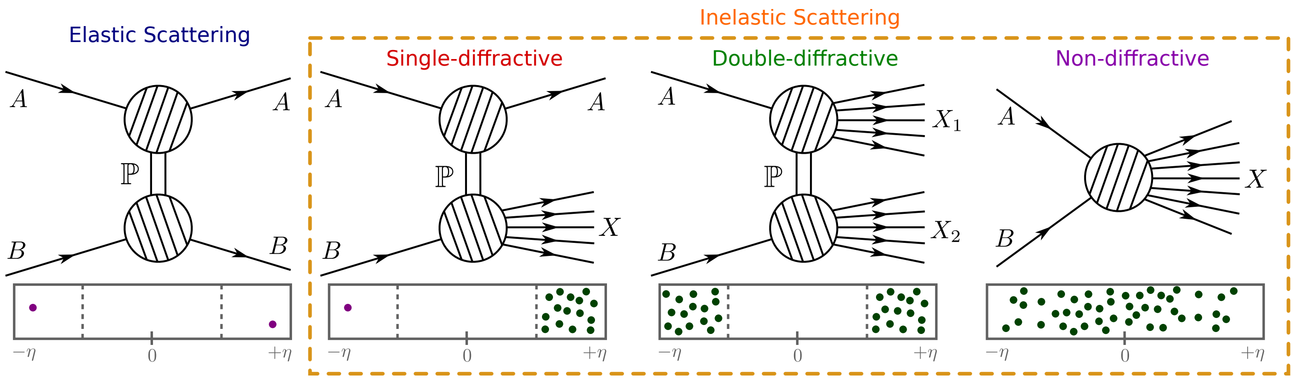


Figure 2.2: Schematic view of the possible scatterings between hadrons A and B mediated by a Pomeron $\mathbb{P}$. The lower side of the illustration shows a qualitative spatial distribution of the final state particles for each of the possible scatterings (the green dots represent the particles produced by the collision and the purple dots represent the colliding hadrons in the initial state).

Two partons can also pass very close to each other. A small impact parameter b classifies these collisions, with the partons scattering at large angles. The impact parameter is the variable conjugated to the transverse momentum. Thus, an interaction between partons with a small impact parameter has large transverse momentum. These high-p_T scatterings require very large transferred momentum in the process. These scattering processes are classified as hard.

Aiming to compare how much of the total hadronic cross-section is from each diffractive process, Table 2.1 provides the experimental measurements of the cross-sections of the possible types of proton-proton collisions at $\sqrt{s} = 7$ TeV. Since the total collision cross-section (σ_{tot}) is given by the sum of the elastic (σ_{el}) and inelastic (σ_{inel}) cross-sections, we have that, for a given center-of-mass energy $\sqrt{s}$,

$$\sigma_{\text{tot}}(s) = \sigma_{\text{el}}(s) + \sigma_{\text{inel}}(s) = \sigma_{\text{el}}(s) + \sigma_{\text{SD}}(s) + \sigma_{\text{DD}}(s) + \sigma_{\text{ND}}(s) \tag{2.5}$$

[1]A Pomeron is a color singlet object with the same quantum number as the vacuum. It is found in the context of the Regge theory, and it is the Pomeron that mediates the interaction between colliding hadrons. More about the Pomeron and Regge theory can be found in [50].

where σ_{SD}, σ_{DD}, and σ_{ND} are the single-diffractive, double-diffractive, and non-diffractive cross-sections, respectively. Thus, even though σ_{el} and σ_{ND} are not shown in Table 2.1, they can be calculated using Equation 2.5. Therefore, we can see that soft processes (non-diffractive) dominate the hadronic collisions, corresponding to approximately two-thirds of the inelastic events, which in turn correspond to about three-fourths of the total events.

Measurement	ALICE ($\sqrt{s} = 7$ TeV)	ATLAS ($\sqrt{s} = 7$ TeV)
σ_{tot} (mb)	–	95.35 ± 1.30
σ_{inel} (mb)	$73.2^{+2.0}_{-4.6} \pm 2.6$	$71.34 \pm 0.36 \pm 0.83$
σ_{SD} (mb)	$14.9^{+3.4}_{-5.9} \pm 0.5$	–
σ_{DD} (mb)	$9.0 \pm 2.6 \pm 0.3$	–

Table 2.1: Cross-section measurements for proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC. The results are provided by the ALICE and ATLAS Collaborations [51, 52]. The CMS Collaboration also carried out cross-section measurements [53, 54]. The CMS measurements are not included in this table since it has a different event selection than the results from ALICE and ATLAS.

QCD successfully describes the parton scattering of hard interactions, that is, for high values of transverse momentum ($p_T \gtrsim 2$ GeV). On the other hand, QCD calculations cannot be applied to interactions with small transverse momentum values, as mentioned above. In high-energy hadron colliders, such as the LHC, non-diffractive inelastic interactions are the most common type of hadron scattering. As such, soft partonic interactions dominate the produced collisions. It is important to remember that the separation of hadronic interactions between soft and hard is artificial and usually depends on some cut in transverse momentum.

It is necessary to combine perturbative QCD (pQCD) with phenomenological methods to explain a hadron collision at high energies, making it possible to study both the high p_T and the low p_T parts of the transverse momentum spectrum.

2.3.1 Scenario for processes with high transverse momentum

At first, hadron-hadron scatterings seem complicated to explore since both particles have an underlying structure. However, at high energies and with high values of momentum exchange, hadron interactions appear to be less complex. In this regime, the interactions seem to be due to a hard scattering of the hadron constituents. From that, we can draw some similarities between hadron-hadron and electron-hadron collisions.

A typical example of a process with high p_T is given by $A + B \to C + X$. In this process, the

C particle has a large transverse momentum with respect to the $A - B$ collision axis ($p_T \equiv |\boldsymbol{p}| \sin\theta$, where θ is the angle between the particle's momentum $\boldsymbol{p}$ and the collision axis), and X represents all other particles in the final state. The A and B colliding particles contain a and b scattering partons, producing partons labeled c and d that carry a high transverse momentum.

For processes involving two particles in the initial and final state, it is interesting to define the Mandelstam variables where we have that

$$
\begin{aligned}
s &\equiv (p_A + p_B)^2 \\
t &\equiv (p_B - p_C)^2 \\
u &\equiv (p_A - p_C)^2,
\end{aligned}
\tag{2.6}
$$

where $p_i = (E_i, \boldsymbol{p}_i)$ is the quadrimomentum of the i particle. It is important to remember that the final state does not exactly consist of two particles, but rather a particle (not necessarily elementary, with high p_T) plus the remainder particles labeled as X. Considering circular hadron colliders, such as the LHC, two hadron beams collide frontally with the particles A and B having the same momentum magnitude and opposite directions. Their four-momentum are given by

$$p_A = (E_A, \boldsymbol{p}) \quad \text{e} \quad p_B = (E_B, -\boldsymbol{p}). \tag{2.7}$$

Therefore, the invariant

$$s = (E_A + E_B, \boldsymbol{p} - \boldsymbol{p})^2 = (E_A + E_B)^2 \tag{2.8}$$

is the square of the center-of-mass energy.

If s, t, u, M^2, and p_T are all large (that is, larger than $m^2_{A,B,C}$, where $m_{A,B,C}$ is the mass of hadrons A, B, or C), where M is the missing mass in the final state due to the particles labeled as X, we can expect that no scales of intrinsic masses are governing the dynamics. Hence, Figure 2.3 can describe the scattering process.

Essentially, we hypothesize that there are soft fragmentations given by $A \to a + \alpha$, $B \to b + \beta$, $c \to C + \gamma$, $d \to X$, where the fragments carry a finite fraction x of the momentum of the primary particles. Furthermore, we use the hypothesis that all of the high p_T arises from the hard scattering $a + b \to c + d$.

Another hypothesis is that a, b, c, and d are formed by partons with momentum q_a, q_b, q_c, and q_d, respectively. If this is true, we can calculate the cross-sections of processes $A + B \to C + X$ if we

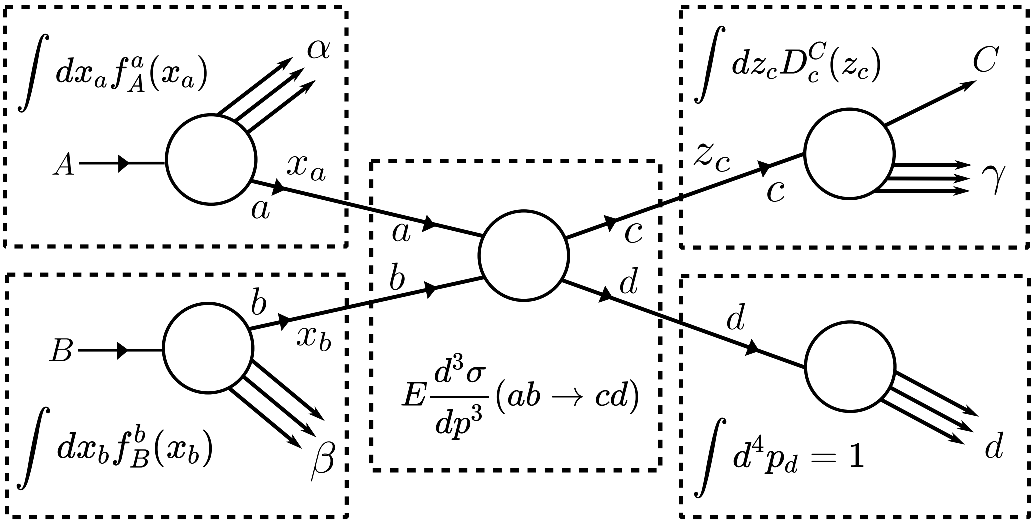


Figure 2.3: Illustration of a high-p_T $A + B \to C + X$ process generated from the $a + b \to c + d$ fragmentation.

know

i. the parton distribution in hadrons $f_{A,B}^{a,b}(x_{a,b})$ (also called structure functions), which is a function of the momentum fraction given by $x_{a,b} = q_{a,b}/p_{A,B}$. These functions describe the vertices a and b in Figure 2.3 and can be determined from electron scattering. If A is a proton, the function determination becomes possible;

ii. the fragmentation functions $D_c^C(z_c)$, which represent the probability that the resulting parton c will produce a hadron C that carries a fraction of the momentum given by $z_c = p_C/q_c$. This function describes vertex c in Figure 2.3. In principle, these functions can be determined from processes such as $e^+e^- = C + X$;

iii. the hard scattering subprocess $a + b \to c + d$.

Knowing the distribution functions of (i) and (ii), a model for $a + b \to c + d$ will completely specify the cross-section for the process $A + B \to C + X$. Inversely, if we have data about the scattering $A + B \to C + X$, we will be able to extract the behavior of $a + b \to c + d$ and thus understand the underlying parton dynamics and the strong interaction nature.

In this way, the invariant differential cross-section becomes [9]

$$E_C \frac{d^3\sigma}{dp_C^3}\bigg|_{AB\to CX} = \int dx_a f_A^a(x_a) \int dx_b f_B^b(x_b) \left[E\frac{d^3\sigma}{dp^3}(ab \to cd)\right] \int dz D_c^C(z_c). \tag{2.9}$$

Details on how to obtain this expression are found in [10, 55, 56].

2.3.2 Parton model predictions

An important outcome of perturbative QCD is obtaining the relationship between the amount of produced particles in hadron collisions and their transverse momentum values. Thus, the parton model provides a way to connect the perturbative theory with a measurable distribution from hadron collider experiments. Since pQCD only analyzes the high-p_T part of the spectrum, it only provides the cross-section for this regime. However, the p_T behavior of the cross-section is already enough to draw some questions around the parton model, which includes its prediction of the dominant partonic subprocess in hadronic collisions.

In short, for a theoretical prediction about the high-p_T region of a hadron-hadron collision, it is necessary to know the distribution of quarks within the initial hadrons, the probability of finding a hadron carrying a certain fraction of the momentum of a quark, and the differential cross-section for an elastic scattering involving partons.

We can isolate these three steps. Therefore, it is possible to use other experiments to determine these parameters. For example, we can use deep inelastic scattering processes from electron-proton collisions to establish the hadron structure functions and data involving hadron production to determine the fragmentation functions. It is necessary to adjust the experimental data to obtain the subprocess $a + b \to c + d$ cross-section. With that, we can guess as to which is the parton scattering predominant process.

Although we do not have a detailed knowledge of the structure and fragmentation functions, we use the property that they only provide dimensionless parameters. Thus, these parameters are not scale dependent [8]. Only the partonic scattering subprocess contains information dependent on the energy scale, allowing us to establish a relationship between the hadron production and the transverse momentum of the detected particles.

In the most basic parton model, the differential cross-section for the process $a + b \to c + d$ is given from a dimensional counting rule [57–59] (which can also be found in [55]). This process considers that n active fields (particles) participate in this reaction. The Lorentz invariant differential cross-section of a scattering process between hadrons A and B that results in $n - 2$ particles in the final state is given by [4]

$$d\sigma = \frac{(2\pi)^4}{2E_A 2E_B(v_A - v_B)} \left|\mathcal{M}_{fi}\right|^2 \delta^4 \left(p_A + p_B - \sum_{i=1}^{n-2} p_i \right) \prod_{i=1}^{n-2} \frac{d^3 p_i}{(2\pi)^3 2E_i}, \tag{2.10}$$

where $v_{A,B}$ is the velocity of the hadron A,B and $\mathcal{M}_{fi}$ is a Lorentz-invariant matrix element. The element $\mathcal{M}_{fi}$ is related to the transition matrix element T_{fi} between a initial state $|i\rangle$ to a final state $|f\rangle$, where

$$\mathcal{M}_{fi} = (2E_1 \cdots 2E_{n-2} 2E_A 2E_B)^{1/2} T_{fi}. \tag{2.11}$$

Therefore, the cross-section dimension is given by

$$[d\sigma] = \frac{1}{[E]^2}[|\mathcal{M}|^2]\left[\frac{d^3p}{2E}\right]^{n-2}[\delta^4(p)], \tag{2.12}$$

where $[y]$ represents the dimension of a generic parameter y. With $[p]$ representing the momentum dimension, we have that $[E] = [p]$, $[d\sigma] = [p]^{-2}$, and $\delta^4(p) = [p]^{-4}$, so the dimension of M is given by

$$\left[|\mathcal{M}|^2\right] = [p]^{-2(n-4)} = [\text{momentum}]^{-2(n-4)}. \tag{2.13}$$

In this case, where each of the elementary particles carries a fraction of the incident momentum, the relevant momentum scale is given by fractions of $\sqrt{s'}$, where $s' = (p_a + p_b)^2$. Thus, we have that

$$|\mathcal{M}|^2 \sim \frac{1}{(s')^{n-4}}. \tag{2.14}$$

We can perform the integration over the momenta of the fragments a, b, c, and d in the center-of-mass frame of each particle. Its result will not provide a center-of-mass energy s' dependence. The flux factor will scale with s' (or p^2), as well as the d^3p_c/E_c factor. Therefore the Lorentz invariant differential cross-section can be given as a function of s'

$$E_c \frac{d^3\sigma}{dp_c^3}\bigg|_{ab\to cd} \sim \frac{1}{(s')^2}|\mathcal{M}|^2 = \frac{1}{(s')^{n-2}} \tag{2.15}$$

Since we are making a dimensional-only argument, we will not have information about dimensionless quantities such as $x_T = 2p_{c,T}/\sqrt{s}$ and the center-of-mass scattering angle θ_{CM}. A more general expression for the cross-section is given by

$$E_c \frac{d^3\sigma}{dp_c^3}\bigg|_{ab\to cd} \sim \frac{1}{(s')^N} F(x_T,\theta_{CM}), \tag{2.16}$$

where $N = 2 - n = 2 -$ (number of active participants) and $F(x_T,\theta_{CM})$ is a scale-invariant function. For the case where the transverse momentum of the scattered particle is high, we change the relevant scale from s' to $p_{T,c}^2$, and hence

$$E_c \frac{d^3\sigma}{dp_c^3}\bigg|_{ab\to cd} \sim \frac{1}{(p_{T,c}^2)^N} F(x_T,\theta_{CM}). \tag{2.17}$$

If we disregard the hadron mass in our initial scenario of the process $A + B \to C + X$ (i.e., processes where $\sqrt{s} \gg m_{A,B}$), we have that the center-of-mass energy squared can be given by

$$s = (p_A + p_B)^2 \approx 2(p_A \cdot p_B). \tag{2.18}$$

And just as for the subprocess $a + b \to c + d$, we have that

$$s' = (p_a + p_b)^2 \approx 2(p_a \cdot p_b) = 2x_a x_b (p_A \cdot p_B) \approx x_a x_b s. \tag{2.19}$$

Using this relationship between s and s', we have, except for dimensionless factors, that

$$E_c \frac{d^3\sigma}{dp_c^3}\bigg|_{ab\to cd} \sim \frac{1}{(p_{T,C}^2)^N} F(x_T,\theta_{CM}). \tag{2.20}$$

We need to find which subprocess allows us to adjust the experimental data according to the p_T^{-2N} cross-section dependence. The dominant process in high-p_T parton-parton scatterings can be assumed to be given by $qq \to qq$ processes (or other processes $2 \to 2$, which may involve quarks of different flavors and also gluons). Thus, the counting rule leads us to believe that the dependence on p_T of the differential cross-section is given by $N = 4$. In [60], a table illustrates the Mandelstam variables relation for several $2 \to 2$ processes. However, the value of $n \approx 7$ was found for proton-proton collisions produced at the LHC [61]. It was also proposed that the fundamental process would consist of a parton-meson scattering to obtain a value closer to that found experimentally, which would yield a power index $N = 8$. In addition, there are also modified proposals that depict the fundamental process as a meson production, like the reaction $g + q \to q +$ meson, which would provide a power index $N = 6$.

The partonic model was quite successful when describing lepton scattering processes, such as DIS and electron-positron annihilation. However, even if we cannot analyze hadron-hadron collisions with the same level of detail as leptons scatterings, we still should not rule out this model. Even though there are several obstacles, it is still possible to explain the dependence of the cross-section concerning the transverse momentum. Thus we can find that [8]

$$E_C \frac{d^3\sigma}{dp_C^3}\bigg|_{AB\to CX} \sim p_T^{-N}(1-x_T)^m, \tag{2.21}$$

where m is an arbitrary factor such that, together with N, the power values of p_T and x_T manage to adjust the experimental data for hadron-hadron collisions. At least, this behavior indicates that it is feasible to carry out the separation of this process in stages. In addition, we obtain an indication that there is an underlying hard scattering process that must be described with fundamental particles.

In a way, the analysis for the p_T dependence of the $A + B \to C + X$ process is more of a phenomenological test for our choice of the cross-section of the partonic subprocess than a theory that allows us to make claims about the quark scattering model. Although it has not been addressed in this analysis, it is necessary to emphasize that there are characteristics of the data that we should successfully fit if this model is correct. However, these characteristics are weakly dependent on the partonic subprocess cross-section [56]. Although it seems that we disregard parameters to make explicit the p_T dependence of the hadron-hadron scattering differential cross-section, these parameters are still necessary if we want to perform a real test of the quark model.

Finally, we conclude that the partonic model is efficient enough so that it is possible to carry out a hadron collision analysis, allowing us to make a phenomenological study of the available high-p_T hadronic data. While it is not a theory that provides us with a more detailed picture of the collision sub-processes, it still is an idea to be considered.

More details on how the partonic model describes high-p_T processes in hadron collisions and their consequences are found in [8, 10, 56]. The current scenario that considers the perturbative QCD model and the phenomenological approach through non-extensive statistical mechanics can be found in [62].

3 | Statistical Hadronic Models

Since pQCD cannot provide an equation that describes the entire p_T spectrum, we need to look for phenomenological models to complement the QCD theory. Usually, these models are based on statistical mechanical approaches. Specifically considering the hadronic model, nonextensive statistical mechanics provide a framework that describes the entire p_T spectrum.

This chapter starts by outlining the description of the hadronic production if we decided to use the standard statistical mechanics, i.e., Boltzmann-Gibbs statistics. Next, the theory of nonextensive statistical mechanics is introduced, focusing on the description of the hadronic production in high-energy collisions.

3.1 Standard statistical mechanics

Statistical mechanics provides a macroscopic description of nature from a microscopic foundation [63]. Even though we can know the interaction between particles, it is impossible to use the equations of motion to describe the behavior of a system composed of a large number of components. Thus, we combine mechanics with the theory of probabilities to arrive at a macroscopic description of a physical system. Ultimately, the statistical approach should lead to the principles of classical thermodynamics.

The formulation of an entropic functional shortens the path to the macroscopic formulation of the theory, ignoring part of the microscopic information of the system [64]. In statistical mechanics, the entropy measures how the probability of a system being in a given state spreads through the possible microstates. Moreover, it is the entropy that is related to the thermodynamical quantities. Thus, it connects the systems' microscopic information with the macroscopic laws.

However, the statistical formulation was not the first context in which the entropy appeared. The term entropy was coined in 1865 by Clausius [65] when he associated the increase of entropy dS with the exchange of heat δQ into a closed system. For example, for a reversible process, we have that

$$dS = \frac{\delta Q}{T}, \tag{3.1}$$

where T is the system temperature (in equilibrium with its surroundings). Thus, the Clausius entropy has a thermodynamic definition and directly connects with a physical meaning.

As a consequence of Clausius's understanding of the second law of thermodynamics, we find the relation between the entropy, the internal energy U, and the work W done by a system. Therefore, we have that the first law of thermodynamics gives

$$TdS = dU + \delta W = TdS = dU + PdV, \tag{3.2}$$

where P and V are the pressure and the volume of the system, respectively. As a result, the entropy is formulated as an extensive thermodynamic variable.

By extensive, we mean that the entropy is proportional to the number N of components of a system in the thermodynamic limit, in other words,

$$0 < \lim_{N\to\infty} \frac{S(N)}{N} < \infty. \tag{3.3}$$

The statistical approach to mechanics was already present in Maxwell's work on the kinetic theory of gases [66, 67]. His work introduced concepts such as velocity distributions, mean free path, and ergodicity. But it was not only until Boltzmann that the probability distributions connected to the thermodynamic concept of entropy. Among his accomplishments are the formulation of entropy and the second law of thermodynamics on a microscopic level [68]. Subsequently, Gibbs further complemented Boltzmann's work by adding statistical ensembles to describe the thermodynamic equilibrium [69]. These ensembles directly connect with physical parameters through thermodynamic potentials. The collection of these formulations composes the foundation of standard statistical mechanics and is available in multiple textbooks [63, 70, 71]. We will refer to this formulation as Boltzmann-Gibbs (BG) statistics.

For the BG statistics, the connection between the entropy and the array of probabilities $\{p_i\}$ of the possible microstates $\{i\}$ for a set of W discrete states is given by

$$S_{BG} = -k_B \sum_{i=1}^{W} p_i \ln p_i, \tag{3.4}$$

with

$$\sum_{i=1}^{W} p_i = 1, \tag{3.5}$$

where k_B is a positive constant. Thus, for the specific scenario of equal probabilities ($p_i = 1/W$ for all microstates), the BG entropy becomes

$$S_{BG} = k_B \ln W. \tag{3.6}$$

As a consequence of the definition from Equation 3.4, we obtain the property of additivity for the BG entropy. By additive, we mean that the sum of the entropy of two probabilistically independent subsystems A and B equals the entropy of the composed system $A + B$. Thus, for a joint probability $p_{ij}^{A+B} = p_i^A p_j^B$, we find that

$$\begin{aligned} S_{BG}(A+B) &= -k_B \sum_{i=1}^{W_A} \sum_{j=1}^{W_B} p_{ij}^{A+B} \ln p_{ij}^{A+B} \\ &= -k_B \sum_{i=1}^{W_A} \sum_{j=1}^{W_B} p_i^A p_j^B \ln p_i^A p_j^B \\ &= -k_B \sum_{i=1}^{W_A} p_i^A \ln p_i^A - k_B \sum_{j=1}^{W_B} p_j^B \ln p_j^B \\ &= S(A) + S(B). \end{aligned} \tag{3.7}$$

The probability restriction from Equation 3.5, the energy constraint

$$U \equiv \langle E \rangle = \sum_{i=1}^{W} p_i E_i, \tag{3.8}$$

and the number of particles restraint

$$N \equiv \langle N \rangle = \sum_{i=1}^{W} p_i N_i \tag{3.9}$$

provide, through entropy optimization, the BG weight for a system in thermodynamical for a system in thermodynamical equilibrium (thermal and chemical) equilibrium at temperature T and chemical potential μ

$$p_i = \frac{1}{Z_{BG}} e^{-\beta E_i + \beta\mu N_i}. \tag{3.10}$$

Here, $\beta = 1/k_B T$ is the Lagrange multiplier related to the energy constraint, $\{E_i\}$ is the systems' energy spectrum, $\{N_i\}$ is the microsystems' number of particles, and Z_{BG} is the partition function given by

$$Z_{BG} = \sum_{j=1}^{W} e^{-\beta E_j + \beta\mu N_j}. \tag{3.11}$$

Further analyzing the entropy, we can formulate relations between the entropy and thermodynamic properties by using Lagrange multipliers. Thus, in the grand-canonical ensemble developed from the previous constraints, we obtain that [70]

$$\frac{1}{T} = \frac{\partial S_{BG}}{\partial U}, \tag{3.12}$$

that the internal energy is

$$U = -\frac{\partial}{\partial\beta} \ln Z_{BG} - \mu N, \tag{3.13}$$

that the expected number of particles is

$$N = \frac{1}{\beta}\frac{\partial}{\partial\mu} \ln Z_{BG}, \tag{3.14}$$

and that the grand thermodynamic potential is

$$A \equiv U - TS_{BG} - \mu N = -k_B T \ln Z_{BG}. \tag{3.15}$$

If we wish to describe hadrons statistically, we can begin with an ideal quantum gas in the BG framework. In this formulation, we can frame the partition function in terms of the energy states instead of the microsystems

$$Z_{BG} = \sum_{j=1}^{W} e^{-\beta E_j + \beta\mu N_j} = \sum_{\{n_k\}} e^{\beta \sum_k (\epsilon_k - \mu) n_k} = \prod_k \sum_{\{n_k\}} e^{-\beta(\epsilon_k - \mu) n_k} = \prod_k Z_{BG,k}, \tag{3.16}$$

where

$$Z_{BG,k} = \sum_{\{n_k\}} e^{-\beta(\epsilon_k - \mu) n_k}, \tag{3.17}$$

and n_k and ϵ_k are the number of particles and the energy of the particle in the state k, respectively. In Equation 3.16, we have used the energy and number of particles conservation

$$N_i = \sum_k n_k \tag{3.18}$$

$$E_i = \sum_k n_k \epsilon_k. \tag{3.19}$$

A quantum gas contains fermions and bosons, which obey the Bose-Einstein and Fermi-Dirac statistics, respectively, that is,

$$n_k = 0,1,2,3,\cdots \qquad \text{for bosons} \tag{3.20}$$

$$n_k = 0,1 \qquad \text{for fermions.} \tag{3.21}$$

Thus, the partition function in each energy state for bosons and fermions are, respectively,

$$Z^B_{BG,k} = \sum_{n_k=0}^{\infty} \left(e^{-\beta(\epsilon_k-\mu)}\right)^{n_k} = \frac{1}{1-e^{-\beta(\epsilon_k-\mu)}} \tag{3.22}$$

$$Z^F_{BG,k} = \sum_{n_k=0}^{1} e^{-\beta(\epsilon_k-\mu)n_k} = 1 + e^{-\beta(\epsilon_k-\mu)}. \tag{3.23}$$

Hence, the occupation factor in a given state k is

$$f^B_k \equiv \left\langle n^B_k \right\rangle = \frac{1}{\beta}\frac{\partial}{\partial\mu} \ln Z^B_{BG,k} = \frac{1}{e^{\beta(\epsilon_k-\mu)}-1} \tag{3.24}$$

$$f^F_k \equiv \left\langle n^F_k \right\rangle = \frac{1}{\beta}\frac{\partial}{\partial\mu} \ln Z^F_{BG,k} = \frac{1}{e^{\beta(\epsilon_k+\mu)}+1}. \tag{3.25}$$

The occupation factor can also be written as

$$f_k = \frac{1}{e^{\beta(\epsilon_k-\mu)}-\kappa} \qquad \text{with} \qquad \begin{cases} \kappa = +1 & \text{for bosons} \\ \kappa = -1 & \text{for fermions.} \end{cases} \tag{3.26}$$

As we have that the grand thermodynamical potential can be written as

$$\begin{aligned} A &= -k_B T \ln Z_{BG} = -k_B T \sum_k \ln Z_{BG,k} \\ &= \kappa k_B T \sum_k \ln \left[1 - \kappa e^{-\beta(\epsilon_k - \mu)}\right] = -\kappa k_B T \sum_k \ln\left(1 + \kappa f_k\right) \end{aligned} \tag{3.27}$$

and that Equations 3.8 and 3.9 can be given as

$$U = \sum_{i=1}^{W} p_i E_i = \sum_{i=1}^{W} \sum_k p_i n_k \epsilon_k = \sum_k \langle n_k \rangle \epsilon_k = \sum_k f_k \epsilon_k \tag{3.28}$$

$$N = \sum_{i=1}^{W} p_i N_i = \sum_{i=1}^{W} \sum_k p_i n_k = \sum_k \langle n_k \rangle = \sum_k f_k, \tag{3.29}$$

then the entropy according to Equation 3.15 is

$$S_{BG} = \frac{1}{T}\left(U - A - \mu N\right) = \frac{1}{T} \sum_k \left[f_k \epsilon_k + \kappa k_B T \ln\left(1 + \kappa f_k\right) - \mu f_k\right]. \tag{3.30}$$

Using Equation 3.26, we have that

$$\begin{aligned} & e^{\beta(\epsilon_k - \mu)} = \frac{1 + \kappa f_k}{f_k} \Rightarrow \\ & \beta(\epsilon_k - \mu) = \ln\left(1 + \kappa f_k\right) - \ln f_k \Rightarrow \\ & f_k(\epsilon_k - \mu) = k_B T \kappa^2 \ln\left(1 + \kappa f_k\right) - k_B T f_k \ln f_k, \end{aligned} \tag{3.31}$$

where we have used that $\kappa^2 = 1$. Thus, Equation 3.30 becomes

$$\begin{aligned} S_{BG} &= \frac{1}{T} \sum_k \left[-k_B T f_k \ln f_k + k_B T \kappa^2 \ln\left(1 + \kappa f_k\right) + k_B T \kappa \ln\left(1 + \kappa f_k\right)\right] \\ &= -k_B \sum_k \left[f_k \ln f_k - \kappa\left(1 + \kappa f_k\right) \ln\left(1 + \kappa f_k\right)\right]. \end{aligned} \tag{3.32}$$

Finally, we have that the entropy for bosons and fermions are, respectively,

$$S_{BG}^{B} = -k_B \sum_k \left[f_k \ln f_k - \left(1 + f_k\right) \ln\left(1 + f_k\right)\right] \tag{3.33}$$

$$S_{BG}^{F} = -k_B \sum_k \left[f_k \ln f_k + \left(1 - f_k\right) \ln\left(1 - f_k\right)\right]. \tag{3.34}$$

This formulation composes just one physical situation in the extensive array of systems corresponding to BG statistics. The main feature in these descriptions is that they all have the entropic functional S_{BG} at their core. However, even with its great success, the mechanical foundation of BG statistics is not formulated from first principles, i.e., it is not deducted only from microscopic dynamics. Thus, its formulation leaves room for the use of other entropic functionals, with the physical system imposing which one should be applied.

The analysis of systems with long-range interactions is one of the examples of the shortcomings of BG statistics. For this class of situations, the BG entropic functional is not suited for the study, with one possible alternative existing in nonextensive statistical mechanics, addressed in Section 3.3.

3.2 Early phenomenological models

Statistical models to describe the hadronization phenomena in high-energy processes have been present since Koppe [72] and Fermi [25] in the middle of the XX century. Fermi's method assumed that the particles produced in a collision evenly occupied the available phase space, with the possible number of particles generated being determined by statistical weights. This idea depicts strong interaction processes only qualitatively but provides an upper limit to particle production.

Hagedorn developed a model [26] after accounting for the multiple newly-discovered hadronic resonances in the following decades. In his statistical approach, he considered that higher and higher hadronic resonances occur and participate in the system's thermodynamics as if they were particles. This analysis implicates a critical temperature corresponding to the highest possible temperature for strong interactions. Eventually, this critical temperature would be interpreted as the temperature in which a phase transition occurs, leading to a new state of matter, known as the quark-gluon plasma (QGP) [27, 28].

In the attempt to develop a thermodynamical model for the strong interactions, we start by making a few assumptions:

i. the collective motion of the system is negligible;
ii. there are many degrees of freedom;
iii. there is some internal equilibrium;
iv. the strong interactions are internal.

Much of the challenges of statistical mechanics come from handling interacting gases. Thus, the last assumption is worth noticing, for it reduces the problem to a formalism similar to the ideal gas.

This statement is satisfied considering that the gas consists of infinity components representing the strong interaction. As a consequence of these assumptions, the structure of the theory results in what Hagedorn called a hadronic bootstrap. It introduces the name fireball for all hadrons with a circular definition, where it postulates that a fireball is a statistical equilibrium of an undetermined number of all kinds of fireballs, which in turn consists of fireballs, and so on.

One of Hagedorn's primary goals was arriving at the hadronic mass spectrum. He started by considering a macroscopic system in the canonical formalism of standard statistical mechanics, which gives, in natural units ($k = 1$), the partition function

$$Z_{BG} = \sum_i e^{\beta E_i} = \sum_i e^{E_i/T} = \int_0^\infty \sigma(E) e^{E_i/T} dE \tag{3.35}$$

where $\sigma(E)dE$ is the number of energy states between E and $E + dE$. Since Z and $\sigma(E)$ are each other's Laplace transforms, we can use the density $\sigma(E)$ to obtain the partition function. However, directly calculating $\sigma(E)$ is not an easy task. Thus, we turn to the expression with the sum over the possible energy states for assistance.

Considering particles of the type j with momentum p_i (and thus energy $\varepsilon_{ij} = \sqrt{p_i^2 + m_j^2}$, where m is the particle's mass), the partition function for an ideal quantum gas in a box of volume V is

$$Z(V,T) = \sum_{\{\nu\}} \exp\left(-\frac{1}{T} \sum_{i,j}^{\infty} \nu_{ij} \varepsilon_{ij} \right), \tag{3.36}$$

where ν_{ij} are the occupation numbers according to the Pauli exclusion principle:

- $\nu_{ij} = 0, 1$ for fermions;
- $\nu_{ij} = 0, 1, 2, \cdots, \infty$ for bosons.

Following the steps in reference [73], we find that

$$Z(V, T) = \exp\left[\frac{VT}{2\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \int_0^\infty \rho(m; n) m^2 K_2\left(\frac{nm}{T}\right) dm \right], \tag{3.37}$$

where $\rho(m)dm$ is the number of excited hadrons with its mass between m and $m+dm$, and $K_2(nm/T)$ is the Hankel function.

For Equation 3.35 to stay consistent with Equation 3.37, Hagedorn proposed the bootstrap condition that is essentially

$$\frac{\ln \rho(m)}{\ln \sigma(m)} \xrightarrow{m\to\infty} 1. \tag{3.38}$$

The analytical results for this condition are further discussed in references [74, 75]. The bootstrap condition establishes that all fireballs are on equal footing, i.e., the fireball entropy is the same function of its mass as the entropy of its fireball components, which for Hagedorn was given by

$$\rho(m) \xrightarrow{m\to\infty} \frac{C_H}{m^{5/2}} e^{m/T_0}, \tag{3.39}$$

where C_H is a constant and T_0 is the highest possible temperature of the system.

Even though it is impossible to prove or disprove the mass spectrum in Equation 3.39 by direct experiments [76], we can still fit the available part of the spectrum, which behaves exponentially. Hagedorn's analysis yields $T_0 \approx 160$ MeV [76, 77]. This value is obtained by extrapolating a experimental curve with the data of the mass spectrum (available in [78]) with the Equation 3.39. It is the temperature T_0 that operates as the boiling point of the hadronic matter, in which the particle creation is so intense that the temperature can not increase. In modern terms, it is the temperature at which the hadronic matter transitions into the QGP.

As a consequence of the thermodynamical approach through the standard statistical mechanics, we obtain a Boltzmann-type transverse momentum distribution ($\sim \exp\big(-p_T/T\big)$, with a temperature never larger than T_0). Thus, the transverse momentum distribution only accurately describes the low-p_T sector of the spectrum. Since the transverse momentum distribution and the mass spectrum are closely related, this is as far as we can go with the BG statistics.

In the following decades, researchers resorted to a semi-empirical formula to describe the p_T distribution [79, 80]. This approach combined the exponential behavior for low-p_T and the power-law behavior for high-p_T, previously encountered in BG statistics and pQCD calculations, respectively. Thus, we have that

$$E\frac{d^3N}{dp^3} = A\left(1 + \frac{p_T}{p_0}\right)^{-n}, \tag{3.40}$$

where A, p_0, and n are fitting parameters. For $p_T \to 0$ we have

$$E\frac{d^3N}{dp^3} \approx \exp\left(-\frac{np_T}{p_0}\right), \tag{3.41}$$

and for $p_T \to \infty$ we have

$$E\frac{d^3N}{dp^3} \approx \left(\frac{p_0}{p_T}\right)^n. \tag{3.42}$$

3.3 Nonextensive statistical mechanics

As previously stated, there is not a reason that prevents the generalization of the BG statistical mechanics since it does not come from microscopic dynamics alone. Thus, even successfully describing many situations, the BG statistical mechanics do not apply to various physical systems. For example, it is inefficient or nonapplicable in the case of complex systems.

Since the BG statistical mechanics does not seem to be universal, Tsallis proposed in 1988 [11] a generalization of the BG entropy in Equation 3.4, postulated as

$$S_q = k_B\frac{1-\sum_{i=1}^{W} p_i^q}{q-1}, \tag{3.43}$$

where q is an entropic index. It is worth noticing that just as the BG statistical mechanics does not have a complete mathematical formulation, the statistical mechanics generated by Equation 3.43 have even less justification. Thus, the Tsallis entropy operates as a way to broaden the validity domain of statistical mechanics.

As a generalization, the Tsallis entropy should return to the BG entropy for a specific case. Indeed, for $q = 1$, we have that

$$\begin{aligned} S_1 &= \lim_{q\to 1} S_q = \lim_{q\to 1} k_B\frac{1-\sum_{i=1}^{W} p_i^q}{q-1} = \lim_{q\to 1} k_B\frac{1-\sum_{i=1}^{W} p_i p_i^{q-1}}{q-1} \\ &= \lim_{q\to 1} k_B\frac{1-\sum_{i=1}^{W} p_i \exp\left[(q-1)\ln p_i\right]}{q-1} = \lim_{q\to 1} k_B\frac{\frac{d}{dq}\left\{-\sum_{i=1}^{W} p_i \exp\left[(q-1)\ln p_i\right]\right\}}{\frac{d}{dq}(q-1)} \\ &= \lim_{q\to 1} -k_B\sum_{i=1}^{W} p_i \ln(p_i)\exp\left[(q-1)\ln p_i\right] = -k_B\sum_{i=1}^{W} p_i\ln p_i = S_{BG}. \end{aligned} \tag{3.44}$$

Finding the maximum of S_q leads to the equiprobability situation, and thus, with $p_i = 1/W$, we have that

$$S_q = k_B\frac{1-W^{1-q}}{q-1}. \tag{3.45}$$

And with $q = 1$, we recover the Boltzmann entropy

$$S_1 = \lim_{q\to 1} k_B \frac{1 - W^{1-q}}{q-1} = \lim_{q\to 1} k_B \frac{1 - \exp\left[(q-1)\ln W\right]}{q-1} = \lim_{q\to 1} k_B \frac{\frac{d}{dq}\left\{1 - \exp\left[(q-1)\ln W\right]\right\}}{\frac{d}{dq}(q-1)}$$
$$= \lim_{q\to 1} k_B \ln W \exp\left[(q-1)\ln W\right] = k_B \ln W = S_{BG}. \tag{3.46}$$

For convenience, we express Equation 3.45 as

$$S_q = k_B \ln_q W, \tag{3.47}$$

where $\ln_q(x)$ is labeled as the q-logarithm, defined as

$$\ln_q(x) \equiv \frac{1 - x^{1-q}}{q-1} \quad (\text{for } x > 0; \ln_1(x) = \ln(x)). \tag{3.48}$$

An immediate property that follows from the Tsallis formulation is that S_q is said to be nonadditive for $q \neq 1$. Indeed, for two independent systems A and B, with joint probability $p_{AB,ij} = p_{A,i}p_{B,j}(\forall(ij))$, we have that

$$\frac{S_q(A+B)}{k_B} = \frac{1 - \sum_{i=1}^{W_A}\sum_{j=1}^{W_B} p_{AB,ij}^q}{q-1} = \frac{1 - \sum_{i=1}^{W_A}\sum_{j=1}^{W_B} p_{A,i}^q p_{B,j}^q}{q-1}$$
$$= \frac{2 - \sum_{i=1}^{W_A} p_{A,i}^q - \sum_{j=1}^{W_B} p_{B,j}^q - \left(1 - \sum_{i=1}^{W_A} p_{A,i}^q\right)\left(1 - \sum_{j=1}^{W_B} p_{B,j}^q\right)}{q-1}$$
$$= \frac{1 - \sum_{i=1}^{W_A} p_{A,i}^q}{q-1} + \frac{1 - \sum_{j=1}^{W_B} p_{B,j}^q}{q-1} - (q-1)\left(\frac{1 - \sum_{i=1}^{W_A} p_{A,i}^q}{q-1}\right)\left(\frac{1 - \sum_{j=1}^{W_B} p_{B,j}^q}{q-1}\right)$$
$$= \frac{S_q(A)}{k_B} + \frac{S_q(B)}{k_B} + (1-q)\frac{S_q(A)}{k_B}\frac{S_q(B)}{k_B}. \tag{3.49}$$

Thus, as we can observe, the additivity property of the entropy depends only on the entropic functional. The situation is quite different for the extensivity property, where it depends on both the entropic functional and the correlations within the system. Additivity and extensivity sometimes are used interchangeably, generating confusion when addressing the Tsallis statistical mechanics. Thus, it is important to emphasize that the defining difference between the Tsallis and BG statistical mechanics is that the BG entropy is additive, while the Tsallis entropy is not. That said, the name nonextensive statistical mechanics has been spread for many years and will be used when referred to the Tsallis formulation. However, the entropy refers to as nonadditive.

Both the S_{BG} and the S_q entropies appear as extensive or nonextensive, for they are dependent on the system. Indeed, for systems with components weakly correlated, the S_{BG} is extensive, whereas the entropy S_q is nonextensive for $q \neq 1$. Alternatively, a system with elements strongly correlated has an extensive entropy for a specific value of $q \neq 1$. Ultimately, the nonadditive entropy emerges as a way to preserve the system's extensivity, satisfying the macroscopic formulation of the entropy brought by Clausius. Thus, in principle, the entropic index q can be chosen so that the entropy is extensive.

Now, we turn to the thermodynamical results that emerge from the nonadditive entropy. Considering canonical ensembles, i.e., systems with one or more constraints, we have that formulating the constraints is not as straightforward as in BG statistical mechanics. For example, establishing the same restriction as Equation 3.8 yields unsatisfactory connections with thermodynamics. Further discussions are present in [13].

Following the analyses exhibited in [14, 81–83], we start our thermodynamical approach through S_q with the following constraints for the average energy U_q and the average number of particles N_q:

$$N_q = \sum_i p_i^q N_i \tag{3.50}$$

$$U_q = \sum_i p_i^q E_i \tag{3.51}$$

In this framework, p_i are the probabilities of the accessible microstates with energy E_i and number of particles N_i, pertaining to a grand canonical ensemble. The probability constraint is given by 3.5. Considering the principle of maximum entropy, we obtain

$$\frac{\partial}{\partial n_i}\left[\frac{S_q}{k_B} + \gamma\left(1 - \sum_i p_i\right) + \alpha\left(N - \sum_i p_i^q n_i\right) + \beta\left(U_q - \sum_i p_i^q E_i\right)\right] = 0, \quad \forall \text{i}, \tag{3.52}$$

where γ, α, and β are the Lagrange multipliers related to the constraints 3.5, 3.50, and 3.51, respectively. Thus, by differentiating Equation 3.52 and multiplying by $q-1$, we have that

$$\begin{aligned}
&qp_i^{q-1} + (q-1)\gamma + q(q-1)\alpha p_i^{q-1}N_i + q(q-1)\beta p_i^{q-1}E_i = 0 \Rightarrow \\
&\left(q + q(q-1)\alpha N_i + q(q-1)\beta E_i\right) p_i^{q-1} = -(q-1)\gamma \Rightarrow \\
&p_i = \left[\frac{-(q-1)\gamma}{q + q(q-1)\alpha N_i + q(q-1)\beta E_i}\right]^{\frac{1}{q-1}} = \left[\frac{q + q(q-1)\alpha N_i + q(q-1)\beta E_i}{-(q-1)\gamma}\right]^{\frac{1}{1-q}}.
\end{aligned} \tag{3.53}$$

Inserting Equation 3.53 into Equation 3.5, we obtain that

$$
\begin{aligned}
&\sum_i p_i = \sum_i \left[\frac{q + q(q-1)\alpha N_i + q(q-1)\beta E_i}{-(q-1)\gamma}\right]^{\frac{1}{1-q}} = 1 \Rightarrow \\
&\left[-(q-1)\gamma\right]^{1/(1-q)} = \sum_i \left[q + q(q-1)\alpha N_i + q(q-1)\beta E_i\right]^{1/(1-q)}.
\end{aligned}
\tag{3.54}
$$

Therefore, using Equation 3.54, eliminating $q^{1/(1-q)}$, and identifying the Lagrange multipliers α and β as (in natural units, $k_B = 1$)

$$
\alpha = -\beta\mu \quad \text{and} \quad \beta = 1/T, \tag{3.55}
$$

respectively, where μ is the chemical potential, we have that the probabilities are given by

$$
p_i = \frac{\left[1 + \beta(q-1)(E_i - \mu N_i)\right]^{1/(1-q)}}{Z_q}, \tag{3.56}
$$

where Z_q is the partition function given by

$$
Z_q = \sum_i \left[1 + \beta(q-1)(E_i - \mu N_i)\right]^{1/(1-q)}. \tag{3.57}
$$

Another way of representing the system through quantum numbers is through the sets of occupation numbers. They also determine the number of particles and the system's energy, where for a given microstate i, we have that

$$
N_i = \sum_k n_k \tag{3.58}
$$

$$
E_i = \sum_k n_k \epsilon_k, \tag{3.59}
$$

where n_k and ϵ_k are the number of particles and the energy of the particle in the state k, respectively. Thus, Equations 3.56 and 3.57 become

$$
p_{\{n_k\}} = \frac{\left[1 + \sum_k \beta(q-1)(\epsilon_k - \mu)n_k\right]^{1/(1-q)}}{Z_q}, \tag{3.60}
$$

where

$$
Z_q = \sum_{\{n_k\}} \left[1 + \sum_k \beta(q-1)(\epsilon_k - \mu)n_k\right]^{1/(1-q)}. \tag{3.61}
$$

The main difference between Equation 3.60 and the probabilities obtained through BG statistical mechanics is that Equation 3.60 cannot factorize into terms corresponding to each particle, for we do not have the usual exponential weights. Explicitly, we have that

$$\left[1+(q-1)(A+B)\right]^{1/(1-q)} \neq \left[1+(q-1)(A)\right]^{1/(1-q)} \left[1+(q-1)(B)\right]^{1/(1-q)} . \tag{3.62}$$

One possible approximation is given by [82] and expanded by [83]. They argue that the correlations between particles can be disregarded because we are dealing with a dilute gas, and thus the states of different particles are regarded as statistically independent. This approximation results in

$$Z_q = \prod_{k=1}^{\infty} \sum_{n_k=0}^{\infty} \left[1+\beta(q-1)(\epsilon_k-\mu)n_k\right]^{1/(1-q)} . \tag{3.63}$$

Thus, with Equation 3.63 and the mathematical arguments found in [82], the occupation factor in nonextensive statistical mechanics is

$$f_k = \frac{1}{\left[1+\beta(q-1)(\epsilon_k-\mu)\right]^{1/(q-1)} - \kappa} \qquad \text{with} \quad \begin{cases} \kappa = +1 & \text{for bosons} \\ \kappa = -1 & \text{for fermions.} \end{cases} \tag{3.64}$$

The statements to arrive at Equation 3.64 consist of analyzing the partition function considering different ranges of the index q and seeing how the summation over n_k behaves. For convenience, we formulate Equation 3.64 as

$$f_k = \frac{1}{e_q^{\beta(\epsilon_k-\mu)} - \kappa} \qquad \text{with} \quad \begin{cases} \kappa = +1 & \text{for bosons} \\ \kappa = -1 & \text{for fermions.} \end{cases} \tag{3.65}$$

where e_q^x is labeled as the q-exponential, defined as

$$e_q^x = \left[1+(q-1)x\right]^{1/(q-1)} \qquad (\text{for } x>0;\, e_1^x = e^x), \tag{3.66}$$

which is the inverse function of the q-logarithm.

Another way of arriving at Equation 3.64 is following the work in [14, 81, 84, 85]. We start by generalizing the entropy for fermions and bosons given by Equations 3.33 and 3.34, which in the

Tsallis formulation becomes

$$S_q^B = -g \sum_k \left[f_k^q \ln_q f_k - (1+f_k)^q \ln_q (1+f_k) \right] \tag{3.67}$$

$$S_q^F = -g \sum_k \left[f_k^q \ln_q f_k + (1-f_k)^q \ln_q (1-f_k) \right], \tag{3.68}$$

with $k_B = 1$. In this formulation, the constraints in Equations 3.28 and 3.29 are generalized to

$$U = \sum_k f_k^q \epsilon_k \tag{3.69}$$

and

$$N = \sum_k f_k^q. \tag{3.70}$$

Thus, the principle of maximum entropy leads to

$$\frac{\partial}{\partial f_k} \left[S_q + \alpha \left(N - \sum_i f_k^q \right) + \beta \left(U - f_k^q \epsilon_k \right) \right] = 0. \tag{3.71}$$

Inserting the entropy for fermions S_q^F given by Equation 3.68 into Equation 3.71, we have that

$$\begin{aligned}
\frac{\partial}{\partial f_k} S_q^F &= \frac{\partial}{\partial f_k} \left\{ \sum_k \left[-f_k^q \left(\frac{1 - f_k^{1-q}}{q-1} \right) - (1-f_k)^q \left(\frac{1 - (1-f_k)^{1-q}}{q-1} \right) \right] \right\} \\
&= \frac{1}{q-1} \left\{ -q f_k^{q-1} + 1 + q(1-f_k)^{q-1} \left[1 - (1-f_k)^{1-q} \right] - (1-f_k)^q (1-q)(1-f_k)^{-q} \right\} \\
&= \frac{1}{q-1} \left[-q f_k^{q-1} + 1 + q(1-f_k)^{q-1} - q - (1-q) \right] \\
&= \frac{q}{q-1} \left[\left(\frac{1-f_k}{f_k} \right)^{q-1} - 1 \right] f_k^{q-1}
\end{aligned} \tag{3.72}$$

Thus, replacing the S_q^F derivative in Equation 3.71 with Equation 3.72, we obtain that

$$\begin{aligned}
&\frac{q}{q-1}\left[\left(\frac{1-f_k}{f_k}\right)^{q-1}-1\right]f_k^{q-1}-\alpha q f_k^{q-1}-\beta q f_k^{q-1}\epsilon_k=0\Rightarrow\\
&\frac{1}{q-1}\left[\left(\frac{1-f_k}{f_k}\right)^{q-1}-1\right]=\alpha+\beta\epsilon_k\Rightarrow\\
&\frac{1-f_k}{f_k}=\left[1+(q-1)(\alpha+\beta\epsilon_k)\right]^{1/(q-1)}\Rightarrow\\
&f_k=\frac{1}{\left[1+(q-1)(\alpha+\beta\epsilon_k)\right]^{1/(q-1)}+1}
\end{aligned} \tag{3.73}$$

If we consider the entropy for bosons given by Equation 3.67 instead of the entropy for fermions, we obtain

$$\frac{\partial}{\partial f_k}S_q^B=\frac{q}{q-1}\left[\left(\frac{1+f_k}{f_k}\right)^{q-1}-1\right]f_k^{q-1} \tag{3.74}$$

instead of Equation 3.72.Therefore, the occupation factor for bosons becomes

$$f_k=\frac{1}{\left[1+(q-1)(\alpha+\beta\epsilon_k)\right]^{1/(q-1)}-1}. \tag{3.75}$$

Using the identifications $\alpha=-\mu/T$ and $\beta=1/T$ ($k_B=1$ in natural units), we have that the occupation factors for fermions and bosons are given by, respectively,

$$f_k^F=\frac{1}{\left[1+(q-1)(\epsilon_k-\mu)/T\right]^{1/(q-1)}+1} \tag{3.76}$$

$$f_k^B=\frac{1}{\left[1+(q-1)(\epsilon_k-\mu)/T\right]^{1/(q-1)}-1}. \tag{3.77}$$

Or, using the q-exponential defined in Equation 3.66, we have that

$$f_k^F=\frac{1}{e_q^{(\epsilon_k-\mu)/T}+1} \tag{3.78}$$

$$f_k^B=\frac{1}{e_q^{(\epsilon_k-\mu)/T}-1}, \tag{3.79}$$

which are the same expressions as the ones presented in Equation 3.65. It immediately follows that for $q\to1$, we return to the occupation factors given by BG statistical mechanics (Equation 3.26).

Generally, the constraints in Equations 3.69 and 3.70 provide non-invariant occupation factors, i.e., f_k changes with a transformation of the energies' zeros. Thus, it was proposed that the mean values used should be normalized, solving this problem [13, 86]. However, since f_k are not probabilities, the constraints used in this analysis are sufficient in the present context and do not need to be normalized [84].

In the classical limit, i.e., for $e^{(\epsilon_k-\mu)/T} \gg 1$ both Tsallis formulations for the Fermi-Dirac and Bose-Einstein distributions return to a distribution similar to Boltzmann's, generalized for q.

$$f_k^F = e_q^{-(\epsilon_k-\mu)/T}. \tag{3.80}$$

Thus, changing to a continuum formulation (in natural units, $\hbar = 1$)

$$\sum_k \to V \int \frac{d^3p}{(2\pi)^3}, \tag{3.81}$$

the number of particles given by Equation 3.70 becomes

$$N = \sum_k f_k^q = gV \int \frac{d^3p}{(2\pi)^3} \left[1 + (q-1)\frac{E-\mu}{T}\right]^{-q/(q-1)}, \tag{3.82}$$

where g is the parameter added considering the degeneracy factor.

If we wish to describe the hadron production in particle accelerators, we can start by finding the invariant differential yield, which according to Equation 3.82, can be given by

$$E\frac{d^3N}{dp^3} = gV\frac{E}{(2\pi)^3}\left[1 + (q-1)\frac{E-\mu}{T}\right]^{-q/(q-1)}. \tag{3.83}$$

In terms of experimental parameters, that is, in terms of the transverse momentum p_T and rapidity y, Equation 3.83 becomes

$$E\frac{d^3N}{dp^3} = gV\frac{m_T\cosh y}{(2\pi)^3}\left[1 + (q-1)\frac{m_T\cosh y-\mu}{T}\right]^{-q/(q-1)}, \tag{3.84}$$

where $m_T = \sqrt{m^2+p_T^2}$ is the transverse mass and m is the particles' rest mass. Thus, for mid-rapidity $y = 0$ and zero chemical potential, Equation 3.84 reduces to

$$\left.E\frac{d^3N}{dp^3}\right|_{y=0} = gV\frac{m_T}{(2\pi)^3}\left[1 + (q-1)\frac{m_T}{T}\right]^{-q/(q-1)}. \tag{3.85}$$

The $y = 0$ approximation is appropriate because experiments collect collider data in a small range around mid-rapidity, where p_T distributions are roughly independent of y. As for the zero chemical potential, we assume that our system is in chemical equilibrium in the hadronic phase, i.e., there is no formation of new hadrons (chemical freeze-out).

Ultimately, it is Equation 3.85 that is going to be used to analyze the differential yield of charged particles in proton-proton collisions.

4 | The CMS Experiment

4.1 The Large Hadron Collider and the CMS detector

Located on the Swiss-French border near Geneva, the Large Hadron Collider (LHC) [7] is currently the largest particle accelerator in operation. Built by the European Organization for Nuclear Research (*Conseil Européen pour la Recherche Nucléaire*, CERN), the LHC is installed in a tunnel with about 27 km in circumference and has already collided protons with center-of-mass energies as high as $\sqrt{s} = 13$ TeV. In addition to proton-proton (p-p) collisions, the LHC has also performed lead ion (Pb-Pb) collisions at a center-of-mass energy per nucleon pair as high as $\sqrt{s_{NN}} = 5.5$ TeV.

The LHC appears as the Super Proton Synchrotron (SPS) successor. The SPS was an accelerator that performed proton-proton and proton-anti-proton ($p-\overline{p}$) collisions. Operating from 1981 to 1991, the SPS had as its main achievement the discovery of the W and Z bosons, carried out by the UA1 and UA2 experiments [87–90]. They provided confirming evidence of the unification between the weak and electromagnetic interactions. As for the consideration of CERN particle accelerators, the LHC emerged as the Large Electron Positron Collider (LEP) successor. The LEP was an electron-positron accelerator and was the largest lepton accelerator ever built. Between 1989 and 2001, the same tunnel that houses the LHC accommodated the LEP.

The LHC beam generation process begins at CERN. The proton source comes from the hydrogen gas injection into a cylindrical metal called Duoplasmatron, where an electric field will separate the gas, making negative hydrogen ions (H^-). A radiofrequency system accelerates the particles and focuses the particle beam.

Shortly after, the RF system takes the particles to a linear accelerator (LINAC2). Linear accelerators use radiofrequency cavities to charge cylindrical conductors, alternating their charge positively or negatively. Then, the potential difference between each cylinder accelerates the particles. The linear accelerators produce stable bunches by synchronizing the particle sets arriving on the conductors with the radiofrequency period. When reaching the end of LINAC2, the protons reach an energy of about 50 MeV. After the LHC shutdown between 2019 and 2022, the LINAC4 will replace the LINAC2. The LINAC4 will have protons attaining energies of 160 MeV at its end.

After passing through the linear accelerator, the protons arrive at the Proton Synchrotron Booster (PSB). The PSB is a circular accelerator measuring 157 meters in circumference. It has radiofrequency cavities responsible for accelerating the protons to an energy of 1.4 GeV.

From the PSB, the protons leave to the Proton Synchrotron (PS), another circular accelerator, this time with 628 meters in circumference. The PS accelerates the protons from 1.4 GeV to 25 GeV. The PS is also responsible for providing the 25 ns spacing between the LHC proton bunches.

The protons leave the PS and go to the Super Proton Synchrotron (SPS), another circular accelerator, this time with a circumference of 6.9 km. The protons are accelerated to 450 GeV and then taken to the LHC. The time to fill the LHC is estimated to be 4 minutes and 20 seconds. Inside the LHC, each beam is accelerated to energies up to 6.5 TeV.

A process analogous to the one described above occurs for heavy-ion beams. Lead ion beams start from a pure lead sample of about 500 mg. This sample is heated to about 500 °C to vaporize a small number of atoms. Then, an electric current removes the electrons. In turn, the lead ions pass through the LINAC3 linear accelerator and gain energies of 4.5 MeV per nucleon. The next step takes place in the Low Energy Ion Ring (LEIR), which accelerates ions to energies of 72 MeV per nucleon. The following steps are similar to the proton beams, where they are accelerated to 5.9 GeV on the PS, 177 GeV on SPS, and finally, reach up to 1.38 TeV at the LHC. In the PS, the rest of the surviving electrons are removed. Figure 4.1 illustrates the CERN particle accelerator complex.

There are currently eight experiments at the LHC aimed at hadron collision studies, each with its detector distinguishing experiment. The biggest experiments are the ATLAS (A Toroidal LHC Apparatus) [91] and the CMS (Compact Muon Solenoid) [31], which are general-purpose detectors to study the widest possible range of physical phenomena. Among the main experiments are the ALICE (A Large Ion Collider Experiment) [30] and LHCb (The Large Hadron Collider Beauty Experiment) [92] experiments. These detectors focus on the specific phenomena analysis. ALICE analyses the Quark-Gluon Plasma (QGP) through a detector dedicated to heavy-ion physics, while the LHCb studies the matter-antimatter asymmetry in the universe through the study of mesons composed of the b and c quarks.

Speaking specifically of the CMS, it is a general-purpose detector that contains a program of study

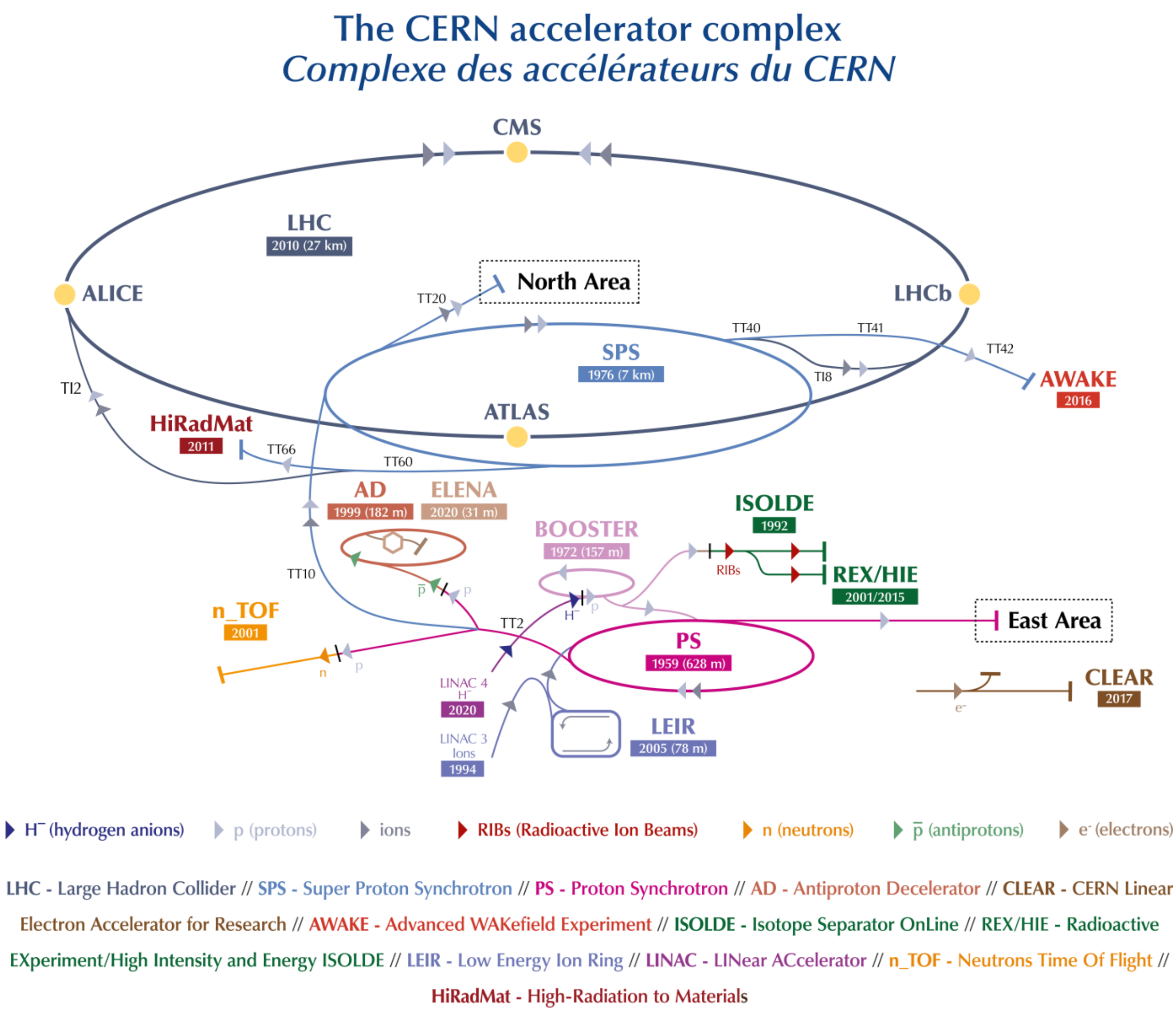


Figure 4.1: CERN's accelerator complex. Credits: CERN

that encompasses both the Standard Model (SM) and physics beyond the Standard Model (BSM). Among the objectives for the study of physics BSM are analyses such as the search for extra dimensions, the search for particles that can compose dark matter, and the examination of the matter-antimatter asymmetry in the universe. For physics within the SM, we can highlight the discovery of the Higgs boson [93] as the main achievement. The CMS also performs SM precision tests and analysis of the Higgs decay channels, further testing SM predictions.

The CMS has similar goals as the ATLAS experiment. The distinction arises from the detector differences. Using two different techniques to study physical phenomena allows a more precise validation of the results obtained. What sets the CMS apart from other experiments is its high-performance muon detection system, its ability to generate a high magnetic field through its solenoid magnet, and its compact detector. With a high-quality electromagnetic and hadronic calorimeter, and a reliable particle tracking system, the CMS detects muons, photons, and electrons with high precision.

4.2 CMS detector

The CMS comes from the necessity to use a detector with an optimized muon detection system to carry out proton collision analysis at high energies and high luminosities at the LHC. This choice aims to supply precise measurements of the momentum of the muon, which in turn can provide clean signatures of several physical processes (for example, the Higgs production). Thus, a magnetic configuration with a strong magnetic field was chosen, supplying the demand for a compact detector. The only practical way to generate a strong magnetic field under these conditions is through a solenoid. The choice made for the CMS detector was a solenoid about 13 meters long and 3 meters wide that generates a magnetic field of about 4 T. This choice ensures good resolution for highly energetic muons.

The solenoid was designed large enough to accommodate the particle tracking system and the calorimeters inside. The innermost part of the detector is composed of the particle tracking system, essential for mapping the curved paths of charged particles. This system surrounds the collision interaction vertex and has a length of 5.8 meters and a diameter of 2.5 meters. One method of measuring the particles' momentum is to measure the particles' trajectory curvature when considering the magnetic field. Therefore, that is another reason to have a high-quality tracking system. The tracker system, composed of pixels and microstrips, is entirely made of silicon. Each measurement of the particle position mapping has an accuracy of 10 μm. The tracking system is the part of the detector that detects the highest number of particles. Thus, its technology needs to have high granularity and a low response time.

The next step in designing the CMS detector was to build the best possible electromagnetic calorimeter (ECAL). A good ECAL is necessary for better resolution in photon and electron detection. The collaboration knew beforehand that an accurate ECAL would be needed to detect the decay of a Higgs into two photons, for example. The ECAL uses crystals with a $|\eta| < 3.0$ pseudorapidity range. The ECAL is a hermetic calorimeter, i.e., it was built to detect all particles emerging from the collisions. Silicon photodiodes in the barrel region and vacuum phototriodes in the endcap region detect the scintillations generated by the particles. Since the π^0 meson also decays into photons, a preshower system positions before ECAL to identify these particles.

The hadronic calorimeter (HCAL) surrounds the ECAL. It also has a pseudorapidity coverage of $|\eta| < 3.0$. The HCAL is responsible for measuring the energy of hadrons. In addition, it also provides indirect measurements of non-interacting particles, like neutrinos. HCAL is also a hermetic calorimeter. In this way, it is possible to identify if momentum and energy imbalance exist in a collision. This process consists of adding the properties of each particle and checking the momentum/energy conservation, making it possible to identify the production of "invisible" particles. HCAL is a sampling calorimeter, i.e., the material that produces the shower of particles is different from the material that measures the deposited energy. Additionally, a frontal hadronic calorimeter (HF) is positioned 11.2 meters from the interaction point. The HF extends the pseudorapidity coverage to $|\eta| < 5.2$.

The muon detector chambers complete the CMS detector system. Muons are the only charged particles that pass by both the calorimeters. Thus, it is ideal to designate specific detectors for these particles. We can accurately track the position of muons by combining the muon position detection in the multiple layers of each muon detection station with measurements of the muon position in the tracking system. By relating the curved trajectories to the magnetic field, it is possible to measure the momentum of particles.

The CMS experiment uses the right-handed coordinate system for particle accelerators, with the origin being at the nominal interaction point. The z-axis is the beam axis, the x-axis points towards the center of the LHC, and the y-axis points up (perpendicular to the LHC plane). Considering spherical coordinates for the momentum vector, θ is the polar angle (angle with respect to the z-axis), and ϕ is the azimuthal angle measured in the (x,y) plane, where $\phi = 0$ is the $+x$ and $\phi = \pi/2$ is the $+y$ direction. In this coordinate system, the pseudorapidity is given as $\eta = -\ln[\tan(\theta/2)]$, where $\eta = 0$ for $\theta = 0$ (direction of the y-axis), and the transverse momentum vector is given as $p_T = |\boldsymbol{p}| \sin\theta$, where $\boldsymbol{p}$ is the particle's momentum. More details concerning the CMS coordinate system is found in the Appendix A.

Figure 4.2 illustrates the complete sketch of the CMS detector. More details on the CMS detector lie in references [31, 94]. The technical notes [95–99] detail the specifications for each subdetector (note that the notes are outdated and upgrades were made over the years). The following subsections depict some of the characteristics of each part of the CMS detector.

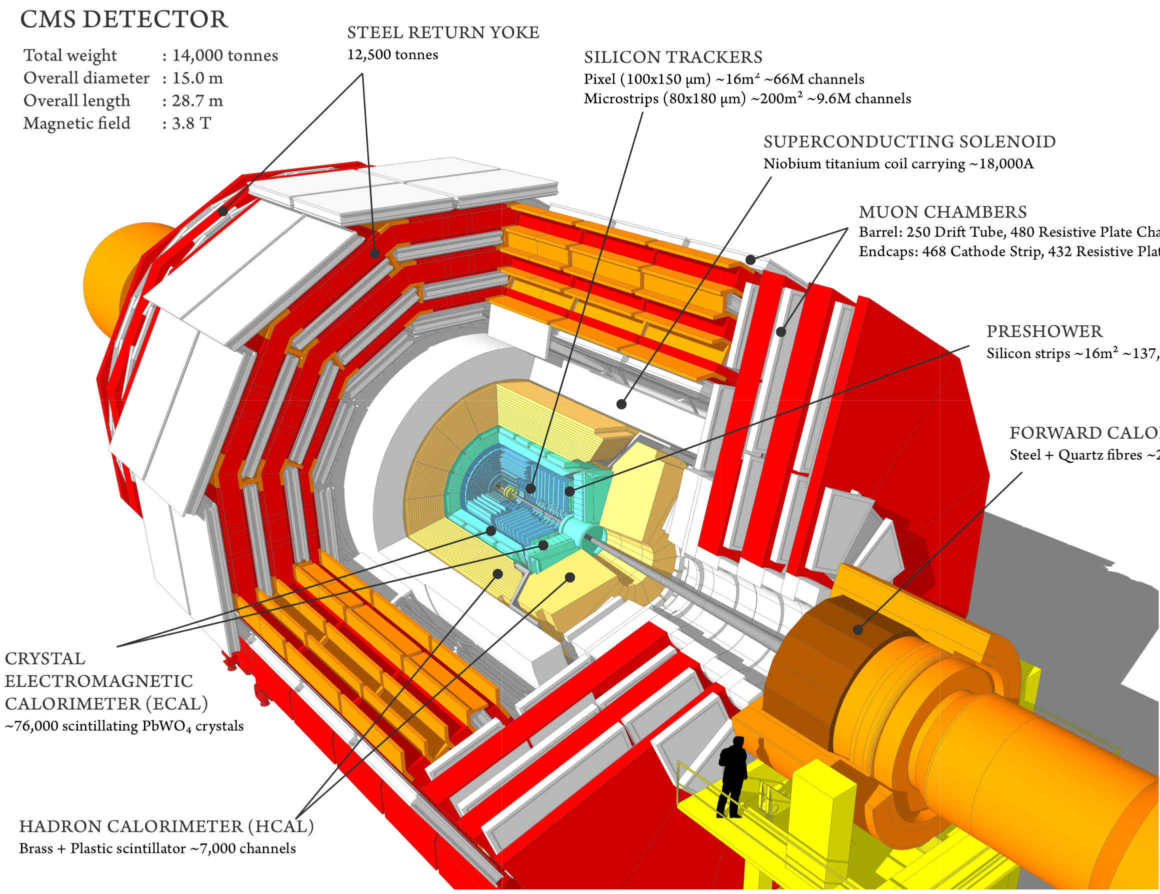


Figure 4.2: Schematic view of the CMS detector. Credits: CERN.

4.2.1 Superconducting solenoid

The CMS superconducting magnet achieves a uniform magnetic field of 4T inside the solenoid, whose length is 12.5 m and diameter is 5.9 m. The magnetic flux outside the solenoid returns through a 1.8 m thick iron yoke whose rings intersperse with the muon stations. The complete magnet system is designed as a 12-sided structure. The dimensions of the whole system provide a length of 21.6 m and an outside diameter of 14.6 m. The total mass of iron used is approximately 11600 tons.

Since it is a solenoid, it is made of coils of wire that produce a uniform magnetic field when electricity flows through them. As for the superconducting property, it allows the electricity to flow without resistance and thus, creates a strong magnetic field. The coil operates at liquid helium temperature to ensure the system's superconductivity. In turn, the CMS conductor consists of a superconducting cable, a pure aluminum stabilizer, and an aluminum alloy reinforcement.

The yoke consists of three rings that are part of the barrel and cover the superconducting solenoid and six discs that act as "lids" (three placed at each end). Each ring of the barrel part comprises three layers of iron, 30 cm thick for the first ring, 63 cm for the second, and 63 cm for the third. It is these layers that intersperse with the muon detectors. The two innermost disks positioned at the ends must withstand the high axial magnetic field. Thus, the two innermost discs are thicker (60 cm) than the outermost disc (30 cm). The superconducting coil is connected only to the barrel's center ring. The configuration of these subdivisions facilitates the assembly and maintenance of the muon detection stations.

The coil, cryogenic system, vacuum pumping station, and protection system against superconducting magnet quenching compose the superconducting coil system. The cryogenic system consists of an internal system with a winding circuit where liquid helium passes and an external system that consists of compressors, a cold box, helium gas containers, and cryogenic lines. Magnet quenching is the process in which there is an abrupt interruption in the magnet operation, where it changes from its superconducting state to a resistive state. The quenching detection system is composed of triggers capable of cutting the circuit and quickly discharging the current, working as a safety system. Figure 4.3 illustrates the magnet and its subsystems (with only one sector of the return yoke).

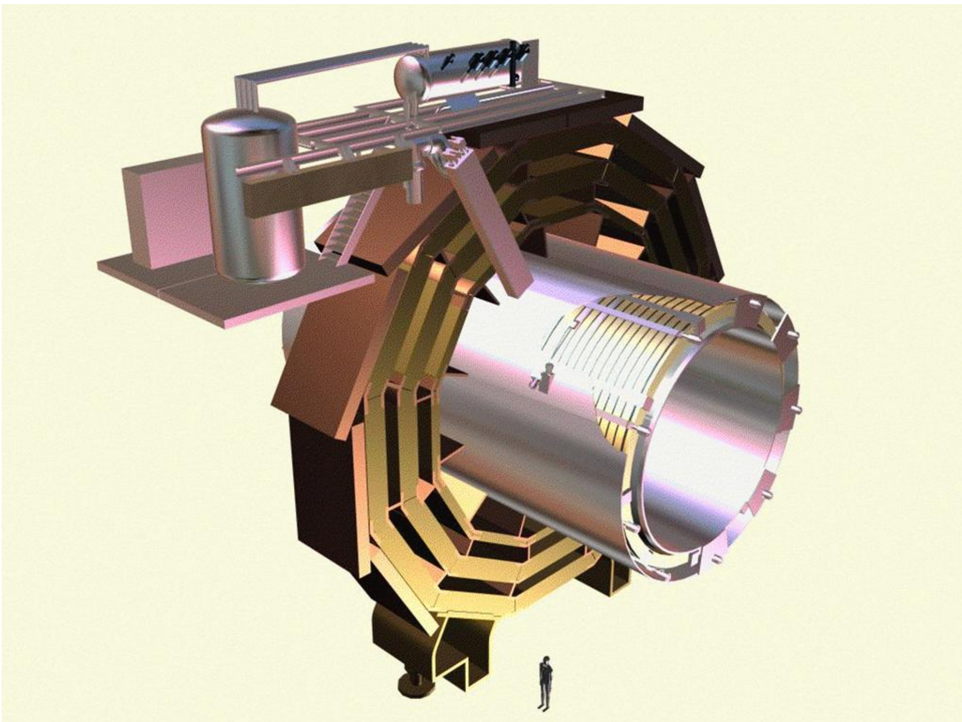


Figure 4.3: Three dimensional view of the superconducting solenoid system of the CMS detector. Credits: CMS, CEA.

4.2.2 CMS tracker

The CMS tracking system provides accurate measurements of the trajectories of charged particles produced in the LHC collisions. Considering the luminosity of about 10^{34} cm^{-2}s^{-1} from the LHC, there will be approximately a thousand particles emerging from more than 20 simultaneous proton-proton collisions passing through the tracking system every 25 ns. An all-silicon-based tracking system provides a detector with high granularity and low response time.

The CMS tracker is composed of a pixel detector and a silicon strip detector. The pixel detector contains three cylindrical layers with radii between 4.4 cm and 10.2 cm and the silicon strip detector with ten cylindrical layers extending the tracker to a radius of 1.1 m. Endcaps are placed at the extreme along the beam axis to complete the detection system, consisting of 2 layers at the ends of the pixel detector and 12 layers at the ends of the microstrip detector. These endcaps extend the detection range of charged particles to $|\eta| < 2.5$. Figure 4.4 illustrates the CMS tracking system.

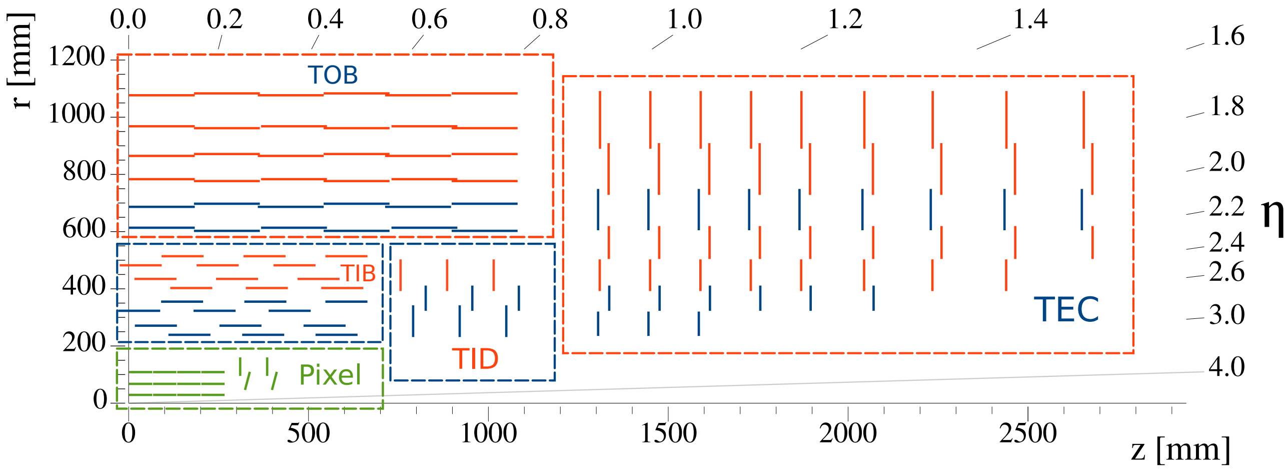


Figure 4.4: Layout of the CMS tracker, showing one quarter of the tracker system in the $r - z$ plane. Since the analysis in this dissertation only uses data from the LHC Run 1, the picture was modified to match the configuration of the Tracker before the upgrades. Credits: CERN (Modified).

Silicon pixel detector

Besides reconstructing the charged particles' trajectory, the pixel detector system is also necessary for reconstructing the collisions' primary and secondary vertices. Several LHC analyses are based on b-quark physics. They use the method of identifying jets originating from the b-quark decay (b-tagging). Hadrons that have a b-quark in their composition have relatively long lifetimes, of the order of 10^{-12} s. When produced in LHC collisions, the combination of this "long" lifetime with Lorentz's time dilation factor generates b-tagged hadrons that travel a few millimeters before decaying. In this context, the pixel detector is positioned as close as possible to the interaction point to identify the primary vertex (collision point) and the secondary vertex (the point at which the b-quark decays). The analysis of other particles with "long" lifetimes also fit this scenario, such as the c quark and the τ lepton.

The pixel detector system provides an optimal resolution by establishing the trajectory points in the plane (r,ϕ,z), using pixel cells with dimensions of 100×150 μm^2. The pixel detector covers the pseudorapidity range of $-2.5 < \eta < 2.5$. The positioning of the three barrel layers (BPix) and the two endcap disks (FPix) combine so that a charged particle leaves at least 3 points of its trajectory in almost the entire η range. In total, the BPix contains 48 million pixels covering 0.278 m^2. As for the FPix, it has 18 million pixels covering an area of 0.28 m^2. Figure 4.5 exposes the pixel detector.

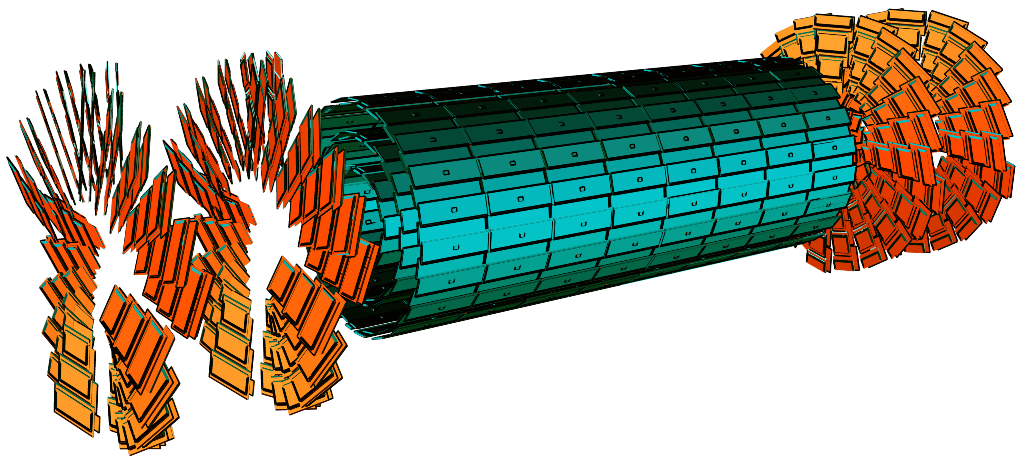


Figure 4.5: Illustration of the CMS pixel detector for the configuration of the CMS Tracker in the LHC Run 1. Credits: [100]

Silicon microstrip detector

The silicon microstrip detector occupies the region between 20 cm and 116 cm from the interaction point. It works with the pixel detectors to reconstruct trajectories and vertices (primary and secondary) and measure the produced charged particles' momentum in the LHC collisions. The tracking system divides into the Tracker Inner Barrel (TIB), the Tracker Inner Disks (TID), the Tracker Outer Barrel (TOB), and the Tracker EndCaps (TEC).

The TIB system consists of four concentric layers in the barrel structure. Right after that, six layers corresponding to the TOB system are positioned. The TID is composed of three discs at each end, each made up of three concentric rings. The TEC consists of nine discs on each side, with each disc being composed of 4 (outermost disc) to 7 (innermost disc) rings. The internal system (TIB and

TID) extends the system to a radius of 55 cm and provides up to 4 points in the $r - \phi$ plane using microstrips 320 μm thick. The TOB extends the trajectory system to 116 cm, consisting of microstrips 500 μm thick and providing at least 6 points in the $r - \phi$ plane. The TEC extend the detection system up to ±118 cm in the z-axis, using microstrips with thicknesses between 320 and 500 μm to provide up to 9 points in ϕ. This setting ensures that a trajectory within the range of $|\eta| < 2.4$ has between 8 and 14 points. The silicon microstrip detector has 9.3 million strips, containing an active area of 198 m^2. Figure 4.6 illustrates the microstrip detector substructures along with the pixel tracker.

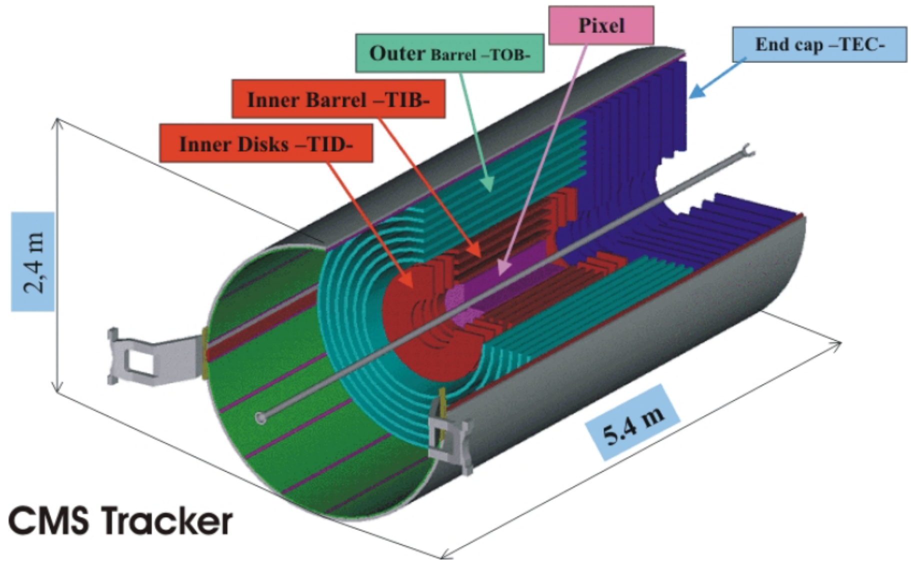


Figure 4.6: Sketch of the silicon microstrips along with the pixel detector in the CMS tracker. Credits: CMS.

4.2.3 Electromagnetic calorimeter

The CMS ECAL's functions include highly accurate photon, electron, and jet energy measurements. The ECAL also provides a hermetic cover to analyze the processes' missing transverse energy.

One of the main goals of the LHC before its construction was to study the symmetry breaking of the electroweak theory through the Higgs mechanism. One of the Higgs boson decay channels is given by $H \rightarrow \gamma\gamma$, where photons detection relies mainly on data collected by ECAL. Another objective of the experiment was the search for supersymmetric particles. The cascade decay of particles such as gluinos and squarks would result in the production of lepton pairs. These channels could provide information about the spectrum of supersymmetric particles.

For this, a scintillating crystal calorimeter was chosen, offering a high energy resolution, as most of the electron and photon energy is deposited in the homogeneous crystal. Lead tungstate crystals

($PbWO_4$) were picked as the scintillating material of the calorimeter, as it is the material that best meets the LHC operation condition. The ECAL consists of 61200 crystals in the barrel and 7324 crystals in each endcap. Silicon photodiodes (avalanche photodiodes, APDs) were used as photodetectors in the barrel part of the ECAL and vacuum phototriodes (VPTs) on the endcaps. APDs and VPTs have the function of converting the scintillations to an electrical signal and amplifying it.

The barrel part of the ECAL (EB) covers the pseudorapidity range given by $|\eta| < 1.479$. The barrel has 360 subdivisions in ϕ and 170 subdivisions in η, totaling 61200 crystals. In the EB, the crystals have dimensions of $2.2 \times 2.2 \times 23$ cm^3. The front of the EB crystals locates at a radius of 129 cm from the collision axis. The ECAL endcaps (EE) extend the coverage in pseudorapidity to $|\eta| < 3.0$. The EE crystals are arranged in an $x - y$ plane, with their curvature focusing 130 cm beyond the interaction point. Each endcap contains 138 subdivisions containing 5×5 crystals plus 18 different subdivisions in such a way as to establish the circular shape of the lids. In total, there are 3662 crystals at each end. Each crystal in the EE has $3 \times 3 \times 22$ cm^3 of dimension, with the distance between the interaction point and the EE being 315.4 cm. Figure 4.7 illustrates the electromagnetic calorimeter.

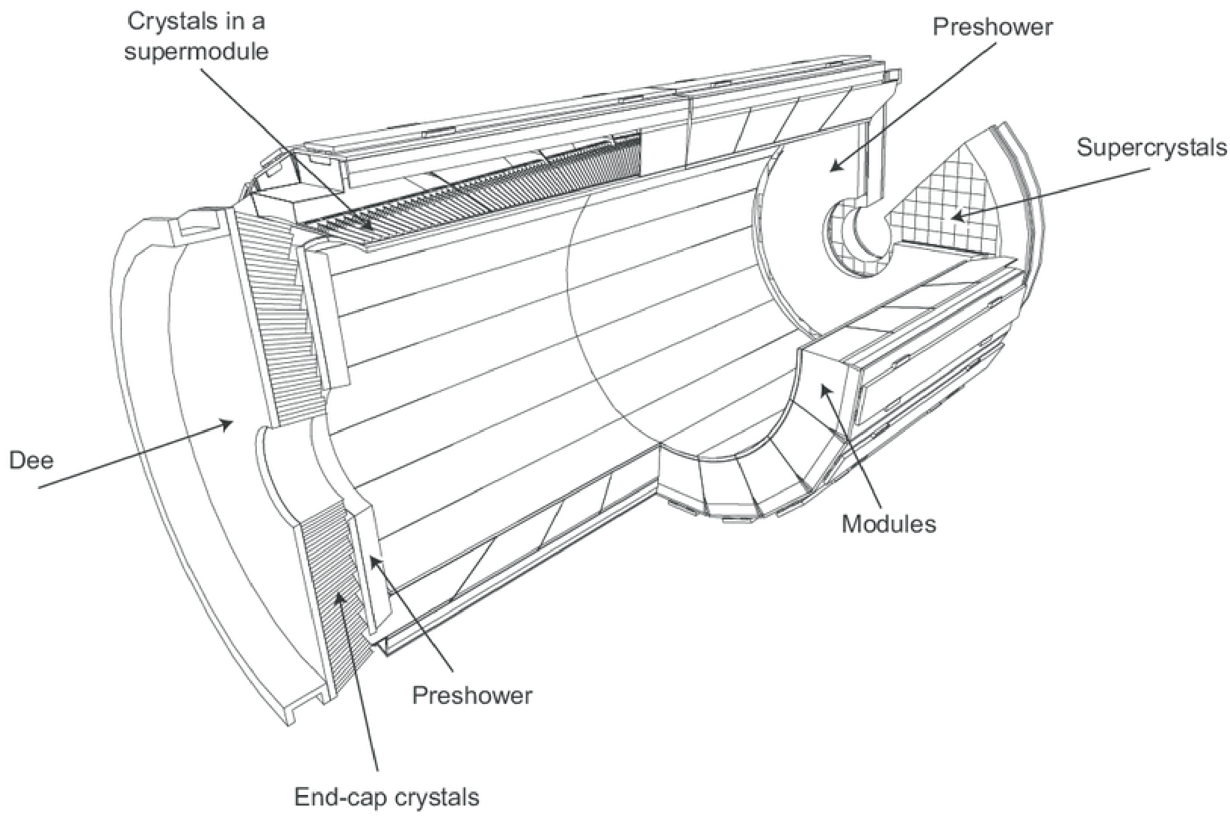


Figure 4.7: Sketch of the electromagnetic calorimeter component of the CMS detector. Credits: [101]

In the EE region, preshower detectors are positioned in front of the ECAL. Thus, the photons (with low energy and small separation) generated by the π^0 decay are not confused with the highly energetic photons generated by the Higgs boson decay. The preshower system has a finer grain than EE crystals to avoid the false signals, containing detectors only 2 mm long. The preshower consists of two layers of lead followed by silicon detectors. The lead layers cause the photons electromagnetic shower, and sensors detect them. By detecting highly energetic photons with ECAL, it is possible to extrapolate the path they traveled by identifying the hits in the preshower. Thus, it is possible to calculate their energy and deduce if they are highly energetic photons or just a photon pair.

4.2.4 Hadronic calorimeter

The HCAL has as its goal the precise measurement of the energies and the direction of particle jets and the events' missing transverse energy. Thus, the CMS also indirectly provides the measurements of quarks, gluons, and neutrinos. Determining the missing transverse energy is essential in searching for physics BSM (therefore, the calorimeter must be hermetic). The search for supersymmetric particles is one example. The HCAL also assists the photon, electron, and muon identification along with the ECAL and the muon detector.

As it is a sampling calorimeter, the HCAL is made of alternating layers of a dense absorber (in this case, brass or steel) and plastic scintillators. When a hadron hits an absorber plate, it produces a cascade of secondary particle showers. The showers travel through the plastic scintillators, emitting a blue-violet light. This light shifts to the green side of the spectrum through "wavelength-shifting" fibers. Then, the optical signals turn into electronic signals through hybrid photodiodes (HPDs). The signal combination from the cascade of particles results in calorimeter towers. The towers measure the particle's energy and can indicate the particle type.

The HCAL consists of a hadron calorimeter barrel (HB) and endcaps (HE) constrained between the ECAL and the superconducting magnet. They provide a pseudorapidity cover of $|\eta| < 3.0$. The spatial restriction limits on the amount of material that the HB can use. Thus, an outer hadronic calorimeter (HO) positions outside the superconducting solenoid, extending the barrel calorimeter. This extension enhances the HCAL ability to absorb the hadronic shower. A forward hadron calorimeter (HF) is placed at 11.2 m from the interaction point. The HF extends the pseudorapidity range to $|\eta| < 5.2$ and completes the HCAL.

The HB, a sampling calorimeter, covers the η range of 0 to approximately 1.4. It divides into two half-barrels, each containing 18 identical azimuthal wedges. In turn, each wedge splits into four azimuthal angle sectors. The HB scintillator also divides into 16 η subdivisions. Thus, the HB segmentation in η and ϕ is 0.087×0.087.

The HE covers the $1.3 < |\eta| < 3.0$ range and has the same subdivisions in ϕ as the HB. The HE contains 14 segmentations in $|\eta|$, with 0.087 η spacing in $|\eta| < 1.74$ and between 0.09 to 0.035 for $|\eta| > 1.74$.

The HO covers the $|\eta| < 1.2$ region. It is divided into 12 subdivisions in ϕ, each segmented into six 5 degrees sectors. The HO is composed of 5 rings in eta, following the solenoid and the muon chambers' positions. Their segmentation follows the HB granularity, ending with 0.087×0.087 subdivisions in $\Delta\eta \times \Delta\phi$.

The HF is a cylindrical detector placed at 11.15 m from the interaction point. It has an outer radius of 130 cm and an inner radius of 12.5 cm from the center of the beamline, extending the η cover to $|\eta| < 5.2$. The segmentation of the HF detector is such that its $\Delta\eta \times \Delta\phi$ granularity is 0.175×0.175 for $|\eta| < 4.7$ and 0.175×0.35 for $|\eta| > 4.7$. Figure 4.8 illustrates a schematic view of one-fourth of the HCAL.

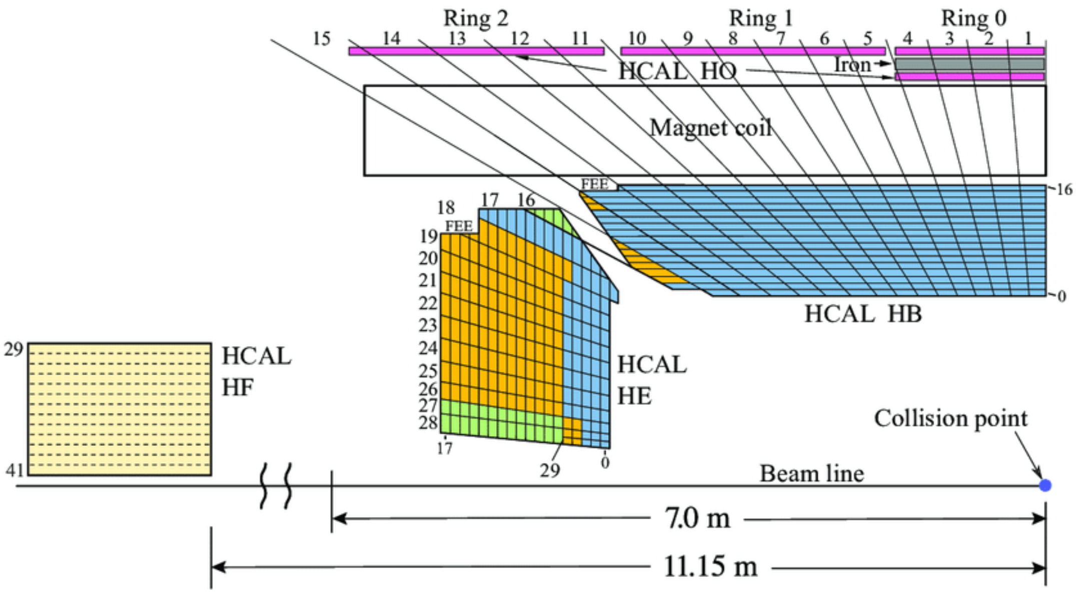


Figure 4.8: Layout of the hadronic calorimeter of the CMS detector, showing one quarter of the HCAL in the $r - z$ plane. Credits: CMS.

4.2.5 Muon detectors

The muon system is designed in such a way that it can identify muons, measure their momentum with good resolution, and assist in the triggering. These measurements are essential for the CMS physics program. For example, the discovery of the Higgs boson was the main goal of the experiment. In turn, the Higgs is predicted to decay through the channel $H \to ZZ \to 4l$. This channel is called "gold-plated" if the leptons are all muons because it has the best 4-particle mass resolution of the possible 4-lepton channels. This resolution is achieved because the muons are less affected by the tracker material than the electrons.

Following the superconducting magnet, the muon system also has a cylindrical shape. Thus, it divides into a barrel section and two planar endcap detectors. The barrel region covers the $|\eta| < 1.2$ region, and the endcaps cover the $0.9 < |\eta| < 2.4$ region. The drift tubes (DTs) and the cathode strip chambers (CSCs) track the particles in the barrel and endcap regions, respectively. The trigger capacities lie in the resistive plate chambers (RPCs) for both detectors. In total, there are 1400 muon chambers: 250 DTs, 540 CSCs, and 610 RPCs.

The DTs are organized into four concentric layers around the beamline interspersed with the return yoke. The three inner layers have 60 drift chambers (with five subdivisions in η and twelve subdivisions in ϕ), while the fourth cylinder has 70 drift chambers (with thirteen subdivisions in ϕ). Each DT chamber consists of 3 (or 2, for the outer layer) superlayers (SL), each made of four layers of rectangular drift cells. Two of the SLs have wires parallel to the beamline, providing the track measurement in the $r - \phi$ plane. The remaining SL has its wire perpendicular to the beamline and provides the z-axis position (this SL is not present in the outer layer). The DTs are made of aluminum, and their volume is filled with a mixture of 85% Ar $+15\%$ CO_2 (kept at atmospheric pressure). When a charged particle crosses a DT cell, it ionizes the gas that later will result in a digital signal.

The endcap muon sector consists of four detector stations containing the CSCs. The innermost station has three rings of detectors, each containing 72 CSCs with a trapezoidal shape. The other three stations have two rings, where the inner ring has 36 CSCs, and the outer ring has 72 CSCs. A charged particle in the $1.2 < |\eta| < 2.4$ range crosses three or four CSC. In the $0.9 < |\eta| < 1.2$ range, it crosses CSCs and DTs. Each CSC comprises layers of positively charged wires interleaved with negatively charged copper cathode strips within a gas volume. The wires run in the ϕ direction and provide the tracks' radial coordinate, while the strips run in the r direction and supply the ϕ coordinate. When a charged particle passes through the CSC, it ionizes the gas, which will send electrons to the wires, generating an avalanche of charged particles. This avalanche generates a digital signal. Alternatively, the positive ions move towards the strips and induce a charged pulse, producing an electrical signal. The gas composition, temperature, and pressure do not affect the CSC precision.

RPCs are fast gaseous detectors that consist of two high-resistive parallel plates (made of a phenolic resin), one charged positively and the other negatively, separated by a gas volume. Aluminum strips, placed on the outer surface of a resistive plate, perform the read-out. As a muon passes through the chamber, it knocks electrons from the gas and generates an electron avalanche. The metallic strips pick up the electrons, which are "invisible" to the resistive plates. The pattern of the hit strips provides a quick measurement of the muon momentum. Its good time and spatial resolution provide the fast particle tracking necessary for an efficient muon trigger. An RPC tags an ionizing event in about one nanosecond, a much shorter time than 25 ns, which is the period between two consecutive bunches crossing in the LHC. Thus, the RPC's fast response time provides a reliable decision of whether the CMS system stores an event.

Six layers of RPCs attach between the DT stations and the return yoke in the barrel sector. Two RPCs locates on both sides of each of the two innermost DT stations, and one RPC is placed on the front side of each of the two outermost DT stations. In the endcaps, four layers of RPCs localize among the return yoke and the CSCs, covering the region up to $|\eta| = 2.1$. Two RPCs couple on each side of the innermost yoke ring, while the third couples on the front side of the third yoke ring and the fourth RPC situates on the backside of the last CSC. Figure 4.9 outlines the muon system substations among the other CMS detectors. Figure 4.10 illustrates the CMS subdetectors and how each particle type interacts with the detector.

4.2.6 Trigger system

The CMS trigger system [103] is composed of a trigger at the hardware level (Level-1 Trigger, L1) and one in a computational module (High-Level Trigger, HLT). The triggers job is to select only the events of interest between the millions of events registered by the detector at each second, diminishing the computational work.

As the hardware system, the L1 trigger has to decide within 4 microseconds of a collision whether an event should be accepted or rejected. The ECAL, HCAL, and the muons chambers process information through several steps before the combined event information is evaluated at a global trigger. Since the L1 trigger system includes the calorimeters, it also has a coverage of $|\eta| < 5.2$. The global

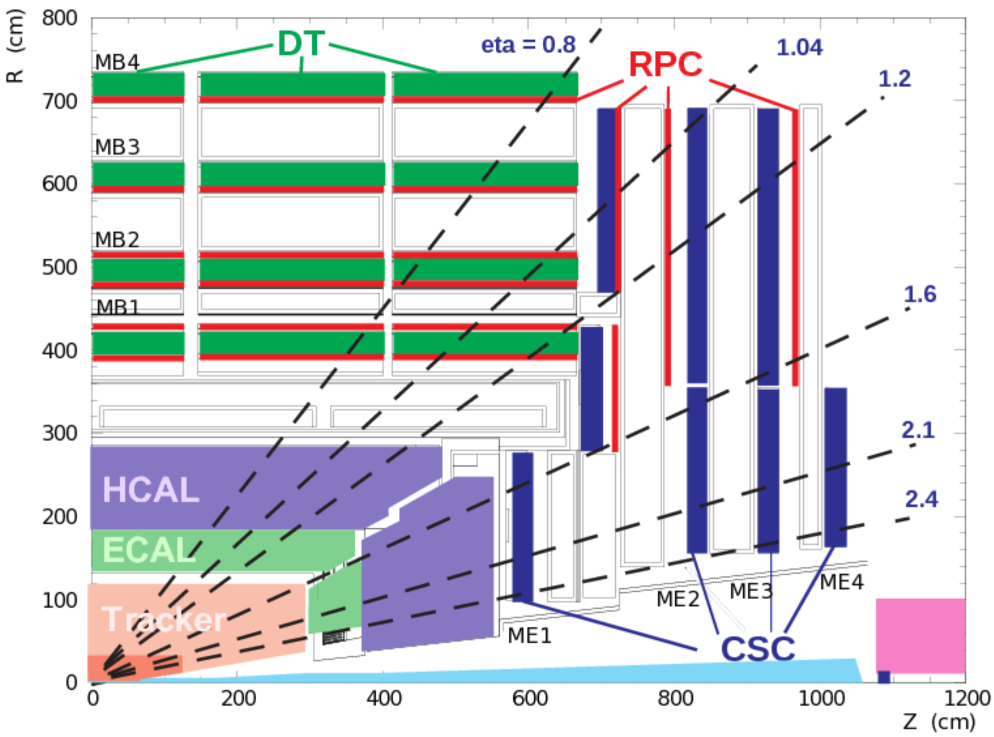


Figure 4.9: Layout of one quarter of the CMS detector, highlighting the stations of the muon detection system. Credits: [102]

trigger is the final step of the L1 trigger system and implements a variety of triggers, deciding if an event will pass for a future evaluation by the HLT.

The HLT uses reconstructed objects (such as electrons, muons, and jets) and identification criteria to retain only the events of interest. An HLT contains several trigger processes, each corresponding to a specific trigger. Each HLT is implemented as a sequence of algorithmic steps, where a set of these steps forms an `HLT path`.

Each trigger (L1 and HLT) has a prescale P related to it. Because the CMS detector detects only a fraction of the events produced, the prescale is necessary to correct the events detection rate. The prescale has a lower limit of 1, meaning that if $P = 1$, all events are stored. For $P > 1$, the CMS keeps 1 out of P events. The L1 and HLT have independent scale factors for each event, such that the total prescale is

$$P_{\text{tot}} = P_{\text{L1}} \times P_{\text{HLT}}, \tag{4.1}$$

where P_{L1} is the prescale of the L1 trigger and P_{HLT} is the prescale of the HLT trigger. The prescales can change within a run but cannot change within a luminosity section[1]. Thus, it becomes necessary to establish the prescales as the weights when filling a histogram.

[1]Luminosity sections are groups of temporally consecutive events with the same calibration configuration.

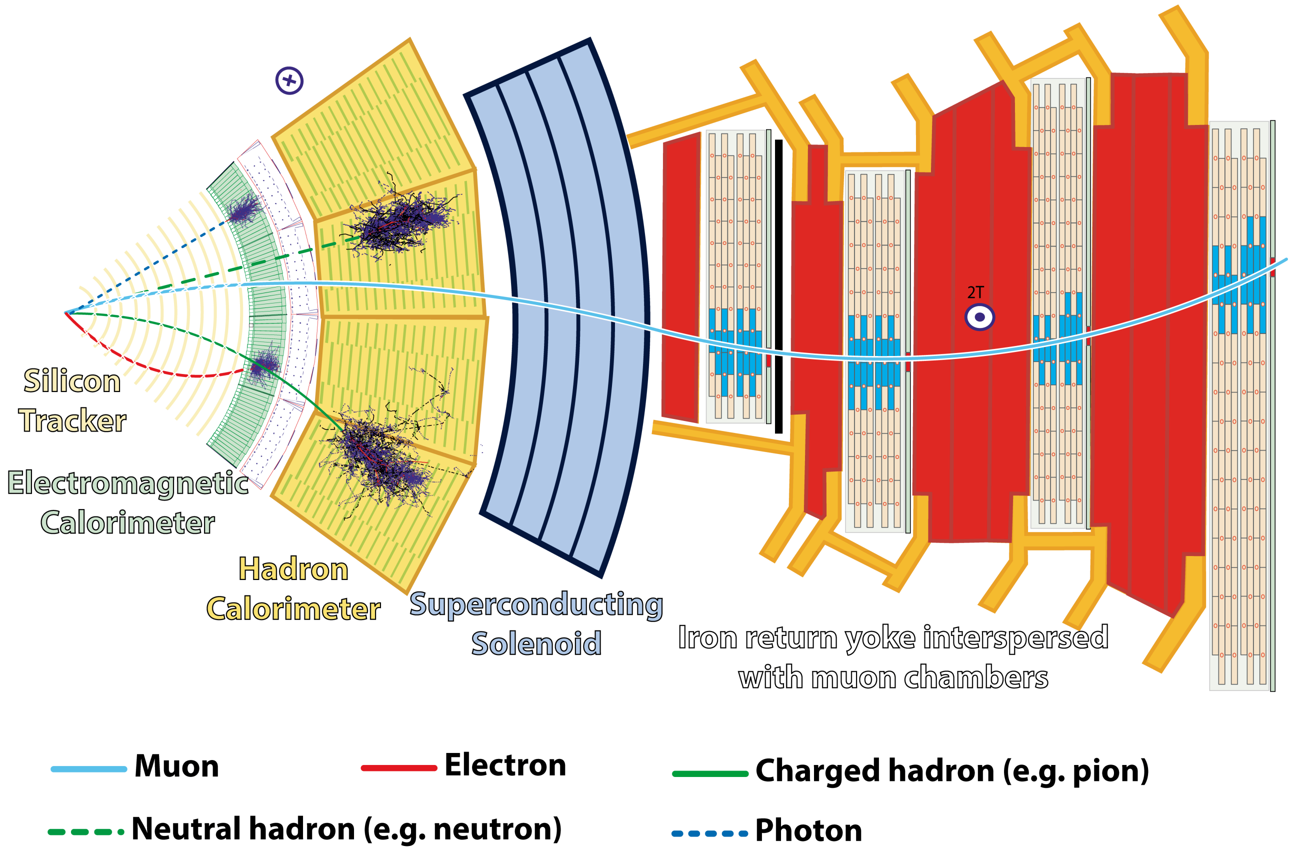


Figure 4.10: Transverse slice of the CMS sub-detector and how each kind of particle interacts with them. Credits: CMS.

4.2.7 Monitoring systems

The CMS detector uses several subsystems to monitor the collisions originating from the bunches crossing inside the LHC. Among the ones worth noting in this dissertation are the Beam Scintillator Counter (BSC) [104] and the Beam Pick-up Timing Experiment (BPTX) [105]. They are designed to supply the proton beam and collision production information. Thus, they provide information concerning the zero-bias and minimum-bias triggers.

BSCs are composed of multiple scintillator tiles that provide the hits and coincidence rates of the proton collisions. The scintillators and PMTs used for the BSC are the same used in the OPAL detector [106]. There are two BSC detectors on each side, one (BSC1) placed at 10.9 m from the interaction point (in front of the HF), while the second (BSC2) lies at 14.4 m from the interaction point. The BSC1 consists of two tile types: the disks and the "paddles". The disks are in the innermost part of the scintillator, with an inner radius of 22 cm and an outer radius of 45 cm, and provide the

[2] Beam halo muons are machine-induced muons traveling along the beam line. They are produced from processes such as collisions of beam particles with residual gas inside the LHC vacuum chamber.

rate information corresponding to the beam condition. The "paddles" are in the outermost part of the BSC, at a radial distance between 55 cm and 80 cm, and supply the coincidence information used to tag beam halo muons[2]. The BSC2 corresponds to two tiles at each side of the interaction point, with an inner radius of 5 cm and an outer radius of 29 cm. It distinguishes between the beamline ingoing and outgoing particles. The BSC configuration exposed in this section corresponds to the LHC Run 1 period, when the data used in this dissertation was collected.

The BPTX is a beam pickup device used everywhere around the LHC ring to monitor the beam position. There are two for the CMS, at 175 m at the right and left from the interaction point. They provide information on the timing and phase of each crossing bunch and its intensity. The experimental clocks' phases compare with the measured one with a precision better than 200 ps, also allowing the measurement of the interaction point z. The read-out of the BPTX signals is also sent as technical trigger inputs to the CMS trigger system. They provide information on if each or both of the bunches are occupied. If both beams are filled, it can indicate whether a collision is possible in this crossing bunch. Thus, it provides a zero-bias trigger for the commissioning of the trigger system.

4.3 CMS Open Data

In the end of 2009, the CMS experiment at CERN began its data taking with the first collisions produced by the LHC. In November 2014, the CERN collaborations solidified its openness commitment with the first release of LHC data to the open community. The first batch came from the CMS experiment, where the data from the first LHC run in 2010 was made available through the CERN Open Data Portal [41].

The CERN community released multiple batches of data since its beginning. The datasets include data from the ATLAS, CMS, LHCb, and ALICE collaborations. While some of the released data has only the purpose of education as its goal, the scientific community can use most of the data available to make research-level physical analyses. Focusing on the CMS Collaboration's released data, it is possible to access almost all of the collected data from proton-proton collisions from 2010 to 2012, with the center-of-mass energy of the collisions ranging from 0.9 to 8 TeV. Heavy-ion collision data was released as well, where it is possible to access data of Pb-Pb collisions with 2.76 TeV of center-of-mass energy per nucleon pair.

The CERN Open Data Portal has multiple datasets regarding the same data-taking run. Each of these different datasets has different selection cuts deciding what kind of events are stored. The triggers enforce the distinct criteria, where the thresholds on the hardware and software level select what events to retain. For instance, one can use a specific dataset with events that have at least two energetic photons to find the Higgs boson through the $H \to \gamma\gamma$ channel.

Besides the data collected from high-energy collisions, the CMS Open Data also provides simulated datasets. These datasets supply the necessary information to reach a physical analysis closer to the truth. For example, they have the necessary details to extract the CMS detector's track reconstruction efficiency required to correct the collision data.

Aside from the datasets, the CMS Open Data Portal contains the documentation necessary so that the whole community can carry out the physical analysis. The documentation includes information about how to access, extract, and analyze the data from the datasets. There are different guides and tutorials regarding how to obtain the content from the physical objects, triggers used, luminosity, etc.

4.3.1 Collision datasets

The first step in an analysis is finding what dataset to use. The CMS Open Data provides multiple primary datasets. Each primary dataset contains different selection cuts for the events stored. For example, datasets imposing the presence of at least two high-energy muons [107] or datasets enforcing the presence of at least one high-energy jet [108].

These datasets contain multiple files in the ROOT format (based on the ROOT framework [109]). The ROOT files store the physical objects made for analysis. Most of these files are in the analysis object data (AOD) format, containing high-level reconstructed objects (like muons, electrons, and photons, for example). They also store properties like the particles trajectory, the calorimeter hits, and the trigger information. Datasets in the raw format are also available, providing the CMS detectors output.

The main difference between the primary datasets is the triggers run in the data taking. There is a trigger set related to the selection cuts applied for each dataset. For example, datasets with events imposing the two energetic muons condition contain triggers related to the event's muon multiplicity and their energy value. The sequence of algorithms of the possible triggers are found in the runs' documentation in the CMS Open Data Portal.

4.3.2 CMS software framework

The CMS software framework (CMSSW) [42] is necessary to analyze the CMS open data. The CMSSW offers the tools needed to access, extract, and even reconstruct the data. There are also tools to simulate events and their detection by the CMS detector. The CMSSW is available through a Docker container or a Virtual Machine image. The executable `cmsRun` implements a job given by a python configuration file. This configuration file tells what data to access and what modules to execute. They ultimately generate ROOT files that are suitable for analysis.

The modules executed by the configuration file are the core of the analysis algorithm. Called by `EDAnalyzers`, they are C++ programs that extract and analyze the physical objects from the datasets. For example, in the Higgs reconstruction from the channel $H \to 4l$, they are the programs that access the lepton information from each event and execute the algorithm that reconstructs the Higgs boson. They also are the programs that store the accessed (and calculated) physical values into ROOT histograms. The CMS Open Data Portal provides several examples describing how to access and analyze the data.

The `EDAnalyzer` module is also capable of extracting the trigger information. If we wish to enforce a selection cut based on the trigger, it is possible to find if it was run, accepted, or produced an error. The trigger prescale is also acquired through the EDAnalyzer. The prescale value is not available for some runs, and in those cases, it is necessary to resort to the `brilcalc` tool [110].

4.3.3 Simulated datasets

Besides the simulated datasets made available by the CMS Open Data Portal, it is also possible to run your Monte Carlo (MC) simulations in the CMSSW. Monte Carlo refers to the computational algorithms based on random numerical samples. It generates particle collisions and simulates the particle interaction within the detector material. Monte Carlo data serves as a synonym for simulated data.

The `cmsDriver` executable generates configuration files using event generator fragments (like Pythia [16, 17] and Herwig [111]). These fragments will be the difference between the configuration files. Like the other configuration files, the `cmsRun` command executes the simulation file. It is possible to control each step of the simulation: starting with the generation of events (GEN), to the detector simulation (SIM), to the signal digitalization and trigger application, until the final event reconstruction (RECO). Figure 4.11 illustrates the different stages present in the CMS event simulation.

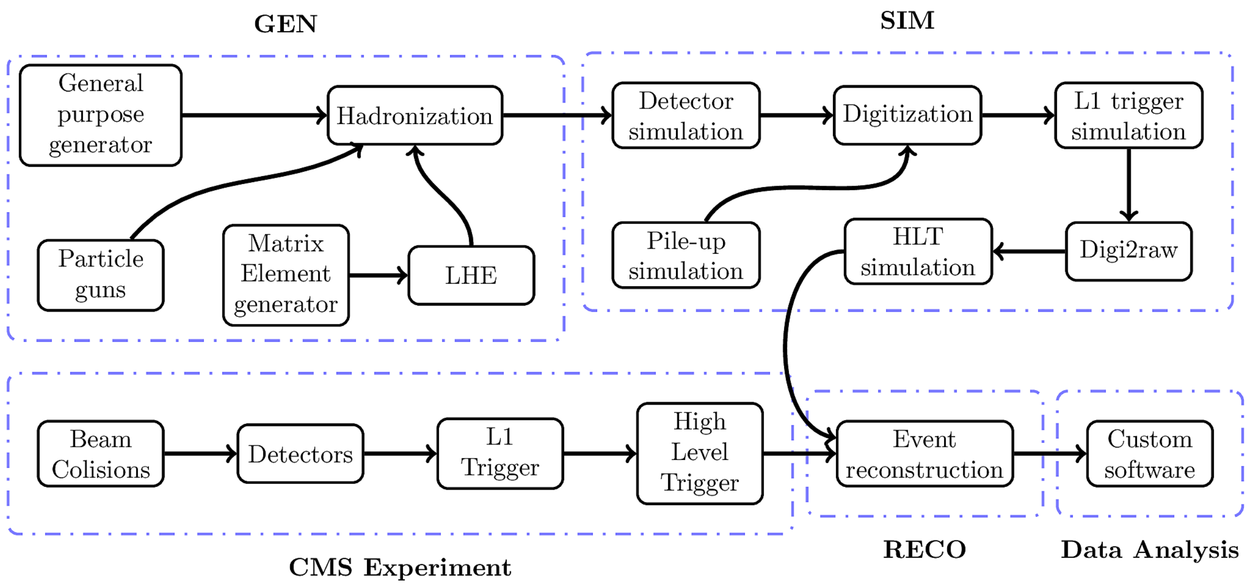


Figure 4.11: Diagram outlining the data manipulation steps for Monte Carlo and collision datasets and their equivalence. Credits: CERN Open Data Portal [41].

The GEN step uses one of the available event generators to simulate the beam collision. One way to produce GEN-level datasets is through a general-purpose generator, like Pythia and Herwig, which generates the event and the hadronization. Another way is using an array element generator (Matrix Element generator), which brings the event to the parton level and then goes through a general-purpose generator to hadronize the event. In this dissertation, the event generator produces the collisions and the hadronization.

The SIM (simulation) step simulates how the produced particles interact with the CMS detector. First, it digitalizes the signals produced by the generated particles going through the detector. When going through the digitization, the L1 trigger is applied, filtering which events are stored. The next step, DIGI2RAW, has the function of converting the digital signals into the raw format. The last SIM step is the high-level trigger (HLT) simulation. This step filters the data according to the required analysis. The CMS detector response in this step is based on GEANT 4 [112].

The last step, RECO, provides the reconstruction of the event. The RECO simulated files are in the AOD format and have the same physical objects as the collision datasets. The output also has information about the generated particles from the GEN step.

5 | Minimum Bias Datasets

5.1 Minimum bias measurements

As stated in Section 2.3, high-energy hadronic collisions fit two categories, elastic and inelastic collisions. Both hadrons maintain their form and do not generate new hadrons on elastic hadron scattering. Inelastic hadron scatterings characterize by the fragmentation of at least one hadron. These inelastic events are labeled as diffractive or nondiffractive (ND) processes. As for the diffractive events, they can be categorized as single-diffractive (SD) or double-diffractive (DD) events, depending on the number of fragmented colliding hadrons. In colliders like the LHC, where the center-of-mass energy ($\sqrt{s}$) is in the TeV scale, ND interactions are the most common type of hadron scattering. Some studies also contain central-diffractive (CD) processes in the analysis of inelastic events. However, this dissertation does not include this type of event in the analysis, since the cross-section of CD processes is much smaller than the SD and DD cross-sections [54].

The charged particle production study developed in this dissertation analyzes standalone collisions in a minimum bias (MB) setting. A MB dataset contains inelastic events with as small bias as possible, i.e., with trigger conditions as loose as can be. The MB datasets include ND, SD, and DD events. For comparison reasons, the results presented in this dissertation feature mainly non-single-diffractive (NSD) interactions. This comparison motive comes mostly because experimental collaborations, such as the UA5 in the SPS [113] and the CDF in the Tevatron [114], used to classify MB events as NSD interactions. This classification was a result of the experimental approach, being due to the trigger system that was inclined to detect NSD events.

The minimum bias event characterization is essential for describing hadronic collisions using Monte Carlo event generators. This experimental characterization enables combining the calculations from perturbative QCD with phenomenological models. Usually, to study soft interactions, experimental collaborations use observables such as

i. the charged-particle multiplicity as a function of the pseudorapidity, $dN/d\eta$ vs. η, also called pseudorapidity distribution (where N denotes the number of charged particles),

ii. the charged-particle multiplicity distributions, $dN_{\text{ev}}/dn_{\text{ch}}$ vs. n_{ch} (where N_{ev} is the number of events and n_{ch} is the detected multiplicity), and

iii. the mean transverse momentum as a function of the charged particle multiplicity, $\langle p_T \rangle$ vs. n_{ch}.

The charged-particle multiplicity distributions, either as a function of the multiplicity or the pseudorapidity, provide insights concerning the rate of partonic interactions. Since these distributions are sensible to the collision center-of-mass energy, they supply a prediction of how the rate of partonic interactions rises as a function of the center-of-mass energy. Thus, this type of study helps the event generator tuning in the new energy frontier of the LHC.

As for the mean transverse momentum distributions, they are helpful in the tuning of color reconnection parameters in Monte Carlo event generators. Experimental distributions show that tracks from high-multiplicity collisions have higher momentum on average than low-multiplicity events. This effect is explained by the color reconnection, where the color interference between partons introduces a correlation between the transverse momentum and the multiplicity [115].

5.2 CMS Open Data datasets

Thus, the first decision concerning the dataset choice is to use datasets containing MB events. In the CERN Open Data Portal, the MB datasets hold events after enforcing MB-like triggers during the data taking. For example, these datasets contain triggers that select events if two proton bunches are crossing in the detector (using the BPTX information) or if a collision occurred (through the BSC hits).

The next step is to choose the proper MB dataset that satisfies our need to understand standalone collisions. This criterion is implemented by using only datasets containing events collected in low pile-up runs. Considering only the datasets with $\sqrt{s} = 7$ TeV, we have several options, such as [116–120], collected in different runs. The dataset [116] is the most suitable for our analysis, for it contains events detected in the low pile-up run of 2010. For dataset [116], collected in a commissioning run, only about 6% of the selected events contain more than one reconstructed primary vertex. As for the other datasets, more than 50% of the events had more than one reconstructed primary vertex. The CERN Open Data Portal also has MB datasets with $\sqrt{s} = 0.9$ [121] and 8 TeV [122, 123]. The 0.9

TeV events were also collected in commissioning runs with low pile-up and therefore were suitable for this analysis, for only about 2% of the events had more than one reconstructed primary vertex. However, due to the high pile-up in the data taking, none of the 8 TeV datasets were used. The CMS Collaboration also provides datasets with $\sqrt{s} = 2.76$ TeV, but even though they are not MB datasets, they are still suitable for all physics and features MB-like triggers. In the 2.76 TeV dataset used [124], about 9% of the events contained more than one reconstructed primary vertex. Table 5.1 presents the datasets explored. Figure 5.1 shows the percentage of events with each number of reconstructed vertices.

$\sqrt{s}$ (TeV)	Valid Events	$\mathcal{L}_{\text{int}}$ (nb^{-1})	Reference
0.9	$\approx 1.07 \times 10^7$	> 0.120	[121]
2.76	$\approx 5.18 \times 10^7$	231.448	[124]
7	$\approx 7.39 \times 10^6$	4.927	[116]

Table 5.1: Collision datasets explored from the CMS Open Data Portal. The integrated luminosity $\mathcal{L}_{\text{int}}$ values were extracted with the `brilcalc` tool mentioned in Section 4.3.2. For the 0.9 TeV dataset, only one of the two collision runs had its $\mathcal{L}_{\text{int}}$ value available.

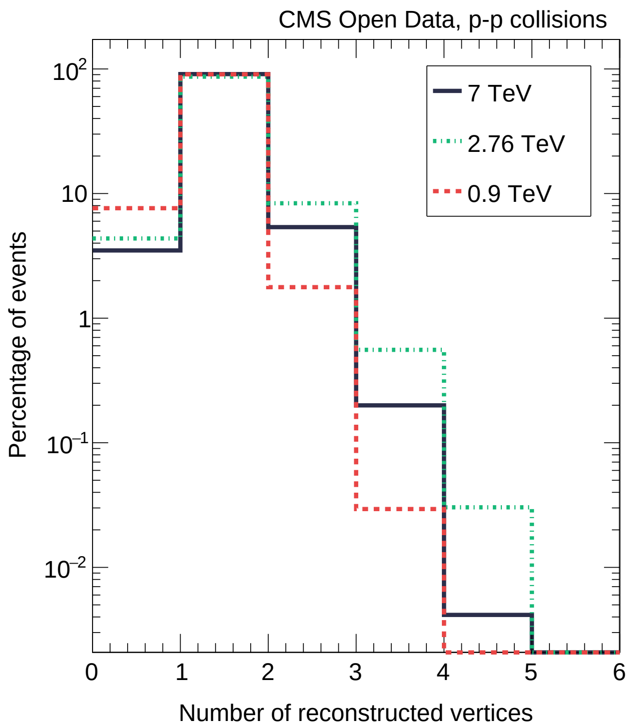


Figure 5.1: Number of reconstructed vertices for datasets from CMS Open Data with $\sqrt{s} = 0.9$, 2.76, and 7 TeV.

Considering the early measurements of the CMS Collaboration regarding the charged particle production, they published results with $\sqrt{s} = 0.9$, 2.36, and 7 TeV [34–36, 40]. The samples with

$\sqrt{s} = 0.9$ and 2.36 TeV were collected in the last quarter of 2009, with collision rates of 11 and 3 Hz, respectively. Additionally, the data with $\sqrt{s} = 7$ TeV was recorded in March 2010 with an interaction rate of about 50 Hz. At these collision rates, the chance that two or more MB collisions occur is minimum. However, the CMS open datasets correspond to 0.9 TeV runs at the beginning of February 2010 and 7 TeV runs in the middle of May 2010. Since all these samples were collected in the low pile-up condition, they all have similar collision rates. The CMS Collaboration also published charged particle distributions for $\sqrt{s} = 2.76$ TeV to compare with heavy-ion data [125]. This data sample was collected in the 2011 LHC run, corresponding to an integrated luminosity of 230 nb^{-1}. This dataset seems to match the 2.76 TeV sample from the CERN Open Data Portal, which was recorded in March 2011.

This thorough selection is carried out because the primary vertex selection, exposed in Section 5.3, is crucial in the analysis. By using datasets with a considerable fraction of multiple reconstructed primary vertices, we would need to either select the primary vertex with the highest multiplicity or discard these events entirely. Both options would result in an undesirable bias in our event selection, and thus this analysis only uses low pile-up data.

The CMS Open Data also provides simulated MB datasets corresponding to specific runs of data taking. For this analysis, the suitable simulated datasets are those matching the commissioning runs of 2010 (for the 0.9 [126, 127] and 7 TeV [128, 129] datasets) and run A of 2011 (for the 2.76 TeV dataset [130, 131]). One similar characteristic of the chosen datasets is that they were all simulated with the condition of no pile-up, meaning that there is only one collision per event. This aspect mimics the low pile-up condition present in the data taking. Unfortunately, all the simulated datasets available in the CMS Open Data Portal are from Pythia generators (Pythia 6 with tune Z2[1] and Pythia 8 with tune 4C [134]). Since we wish for model-independent results, the ideal scenario would be utilizing datasets from simulations containing different particle generation procedures, i.e., simulated from varying physical models. For example, considering the generators Pythia and Herwig, we have differences in the estimation of the multiple partonic interactions, the hadronization models, and the parton shower simulation.

This analysis was also carried out using natively simulated datasets to explore the effects of using simulated datasets from different event generators. Since event simulation is very time-consuming, this study was only done for collisions with $\sqrt{s} = 7$ TeV. Pythia and Herwig++ were the chosen event

[1]The tune Z2 is almost the same as the tune Z1 [132], the only difference is that the Z2 tune uses the `CTEQ6L` parton distribution function [133].

generators, for they were the only ones with MB fragments in the CMSSW repository. Appendix B describes the configuration of the natively simulated datasets, including the necessary simulation steps. All of the simulated datasets (available on the CERN Open Data Portal or simulated natively) contain the detector response information (based on GEANT 4) of each data-taking run and provide the simulation necessary to extract the detector's track reconstruction efficiency. Table 5.2 shows the simulated datasets used and their respective event generators.

$\sqrt{s}$ (TeV)	Generator	Tune	Events	DOI
0.9	Pythia 8	4C	2,270,000	[127]
0.9	Pythia 6	Z2	2,085,000	[126]
2.76	Pythia 8	4C	2,100,000	[131]
2.76	Pythia 6	Z2	2,005,000	[130]
7	Pythia 8	4C	1,995,000	[129]
7	Pythia 6	Z2	2,002,500	[128]
7	Pythia 8	4C	312,500	native
7	Herwig++	–	112,000	native

Table 5.2: Simulated datasets used for data correction.

As for the AOD physical objects used in this study, this analysis imposes the necessity of obtaining information about the tracks (simulated or from the detector), the vertices, and the calorimeter hits. Table 5.3 assigns the physical object to its data input tag. The CMS experiment reconstructs the tracks and vertices objects using information from the pixel and microstrip trackers. In this dissertation, the tracks represent the charged particles detected by the tracker. Section 5.3 explains in more detail the use of the vertices for event selection.

Physical object	Label	Handle
Simulated particles	`genParticles`	`GenParticleCollection`
Track	`generalTracks`	`reco::TrackCollection`
Vertex	`offlinePrimaryVerticesWithBS`	`reco::VertexCollection`
Calorimeter hit	`towerMaker`	`CaloTowerCollection`

Table 5.3: Physical objects used in the analysis of this dissertation. Label represents the name of the physical objects inside the AOD files. Handle designates the container that needs to be called to extract the physical objects of the AOD file.

The 4.2.8 version of the CMSSW was used to analyze the collision datasets with $\sqrt{s} = 0.9$ and 7 TeV, and the simulated datasets with $\sqrt{s} = 0.9$, 2.76, and 7 TeV. The 4.4.7 version examines the collision dataset with $\sqrt{s} = 2.76$ TeV.

5.3 Event selection

The first applied event selection condition is related to obtaining a MB sample, i.e., a sample containing standalone collisions with as minimum bias as possible. As stated in the previous section, a MB-like trigger available in the MB datasets enforces this criterion. In this case, the triggers contain thresholds related to the BPTX information and the BSC hits. The signal coincidence of both sides of the BPTX ensures that two proton bunches are crossing the CMS interaction point at the moment of a detected collision. The BPTX condition guarantees by itself a Zero-Bias dataset containing empty events, collision events, and beam gas background. Thus, to secure a higher efficiency in selecting collision events for a MB sample, the trigger also used the inclusive BSC condition that there is at least one hit in any BSC detector (BSC signal from either side of the detector). Table 5.4 presents the used trigger for each dataset. The triggers used in this dissertation are the same used by the CMS Collaboration to obtain a MB dataset [135]. Notice that for the dataset with $\sqrt{s} = 7$ TeV, there is an OR condition for the BPTX, meaning that the trigger requires at least one bunch crossing the interaction point.

$\sqrt{s}$ (TeV)	HLT trigger path	Average prescale
0.9	`HLT_MinBiasBSC_OR`	1.00
2.76	`HLT_L1BscMinBiasORBptxPlusANDMinus_v*`	424.82
7.0	`HLT_L1_BscMinBiasOR_BptxPlusORMinus`	33.64

Table 5.4: Triggers used in each CMS Open Data dataset to enforce the event selection cut of a BSC hit with proton bunches crossing.

The information about which triggers are available for each dataset is displayed on the dataset webpage in the CERN Open Data Portal. The portal also contains the configuration files used in the data-taking and HLT data processing step for each available run. For example, if we want to know what comprises a specific trigger from a 2010 run, we can go to [136] and look into the configuration files. In the case of the `HLT_MinBiasBSC_OR` trigger, we can see that its trigger path is

```
process.HLT_MinBiasBSC_OR = cms.Path( process.HLTBeginSequenceBPTX + process.
   hltL1sMinBiasBSCOR + process.hltPreMinBiasBSCOR + process.HLTEndSequence ) .
```

The `process.HLTBeginSequenceBPTX` condition enforces that bunches are crossing in the interaction point, and the `process.hltL1sMinBiasBSCOR` setting informs about the BSC hits. In turn, we can see that the `process.hltL1sMinBiasBSCOR` step configuration uses a technical trigger bit to provide the BSC information, given by the line

```
L1SeedsLogicalExpression = cms.string( "34" ), .
```

The 34 bit used in this dissertation corresponds to an inclusive BSC signal (hit on either side of the detector). The definition of each technical bit's signal coincidence can be found in the CMSSW repository [137]. For comparison, the `HLT_MinBiasBSC` trigger is also available and uses the technical trigger bits 32, 33, 40, and 41, each corresponding to a BSC coincidence (BSC signals on both sides of the detector).

Since the sample obtained through the first BSC trigger also contains background events, such as events with beam halo[2], other BSC-based triggers are used to provide information for the rejection of these events. For the datasets from the commissioning run of 2010, i.e., the datasets with $\sqrt{s}$ = 0.9 and 7 TeV, the trigger `HLT_L1Tech_BSC_halo_forPhysicsBackground` was used, corresponding to the technical trigger bits 36, 37, 38, and 39. For the dataset with $\sqrt{s} = 2.76$ TeV, the trigger `HLT_L1Tech_BSC_halo_v*` was used to reject beam halo events.

The final dataset contains only events with at least one 3 GeV hit at each end of the hadronic forward calorimeter (HF), providing a final sample populated mostly with NSD events. This condition is the same imposed by the CMS Collaboration to obtain a dataset with NSD events. The Pythia 8 simulated datasets check the efficiency of this selection cut, for they are the only ones with information about the type of event (if it is SD, DD, or ND). Table 5.5 exhibits the fraction of surviving events and the selection efficiencies for each event type after the HF cut.

$\sqrt{s}$	0.9 TeV		2.76 TeV		7 TeV	
Event type	Frac. (%)	Selec. Eff. (%)	Frac. (%)	Selec. Eff. (%)	Frac. (%)	Selec. Eff. (%)
SD	20.4	15.8	18.8	22.7	17.3	27.7
DD	11.1	41.8	11.4	37.1	11.2	37.0
ND	68.5	96.3	69.8	98.0	71.5	98.7
NSD	79.6	88.7	81.2	89.5	82.7	90.3

Table 5.5: Fraction of SD, DD, ND, and NSD processes before the event selection cuts and their respective selection efficiencies. All values were extracted from the Pythia 8 datasets.

Since the goal of this study is to analyze standalone collisions, events without a valid reconstructed primary vertex (PV) were rejected. For this selection, vertices from a position farther than 15 cm from the interaction point were rejected. PVs are reconstructed from tracks taken from the track collection, with the offline beam spot enforcing a constraint in the fit of the vertex position. The beam spot is the

[2]The beam halo is a low-density particle halo that surrounds the core of the beam. It indicates a beam loss of the LHC beam. More about the mechanisms that contribute to the beam halo can be found in [138].

luminous region where protons of both beams interact. After applying the PV condition, the analysis loops around the tracks that belong to a vertex instead of circling the track collection. For the case where there is more than one reconstructed PV, the vertex with the highest multiplicity is chosen. Figure 5.2 displays the flowchart of the event selection steps carried out in this dissertation. Table 5.6 presents the percentage of surviving events after each selection cut. The lower values for the surviving events after the BSC cut for $\sqrt{s} = 2.76$ TeV can be explained by the fact that the 2.76 TeV datasets contain events suitable for all physics and do not contain a majority of MB events (which occur for the datasets with $\sqrt{s} = 0.9$ and 7 TeV).

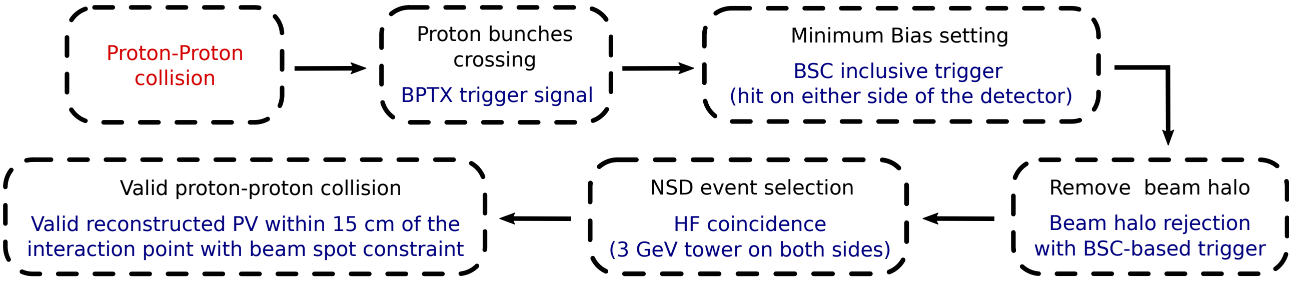


Figure 5.2: Flowchart illustrating the the event selection steps.

$\sqrt{s}$	0.9 TeV	2.76 TeV	7 TeV
Selection cut	% of surviving events		
BPTX + BSC signal	99.57	57.17	100.0
Beam halo rejection	97.97	57.17	99.99
HF coincidence	72.58	47.63	82.68
Valid reconstructed PV	65.67	40.98	74.77

Table 5.6: Percentage of surviving events after each selection cut for $\sqrt{s} = 0.9$, 2.76 and 7 TeV.

After all the selection cuts, 7,025,569, 22,718,660, and 5,525,148 events remained for the final analysis for the datasets with $\sqrt{s} = 0.9$, 2.76, and 7 TeV, respectively.

The selection cuts mentioned in this section provide a dataset suitable for comparison with the distributions published by the CMS Collaboration. However, the ATLAS Collaboration presents results without a selection cut that favors specific diffractive processes. Thus, the HF selection cut was not applied for the pseudorapidity and transverse momentum distributions in the ATLAS kinematic range.

6 | Corrections

The distributions obtained from the physical objects of the CMS Open Data datasets must be corrected considering that they provide only the reconstructed objects of the raw output of the CMS detector. Thus, it is necessary to consider the detector's efficiency in detecting and reconstructing particles and how effective is the event selection so that the proper correction can be applied. The corrections, extracted from the simulated datasets, are divided into three kinds: the track weights, the event weights, and the spectrum correction (for the multiplicity distributions).

6.1 Track detection efficiency

The raw number of charged particles $\Delta N(p_T,\eta)$ present in a given bin with transverse momentum width Δp_T and pseudorapidity width $\Delta\eta$ must be corrected considering a track weight, which depends on p_T and η, given by

$$w_{\text{tr}}(p_T,\eta) = \frac{1}{\varepsilon_{\text{tr}}(p_T,\eta)}(1 - f_{\text{sec}})(1 - f_{\text{fake}}). \tag{6.1}$$

The correction accounts for the detector's track reconstruction efficiency ε_{tr}, the fraction of secondary tracks f_{sec} (tracks that do not belong to the collision's primary vertex), and the fraction of misidentified ("fake") tracks that does not correspond to any charged particle f_{fake}.

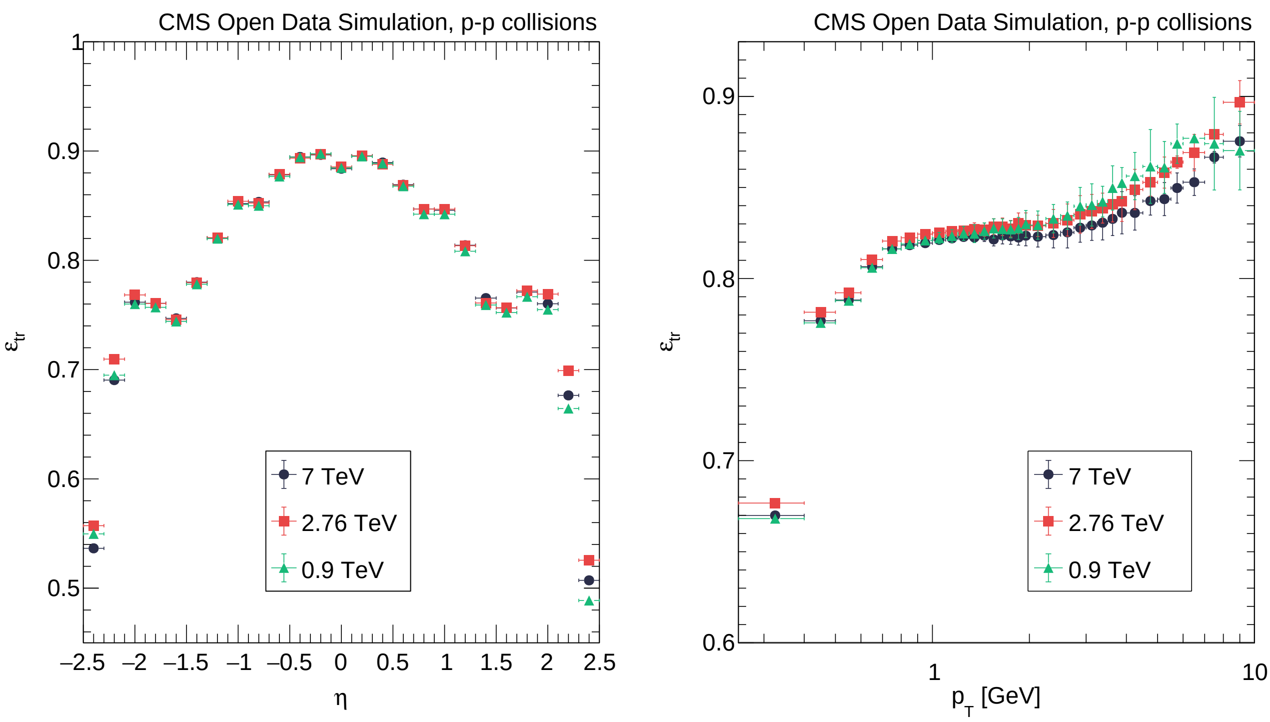
The simulated datasets provide the detector's track reconstruction efficiency, given by the ratio between the number of generated charged particles in the GEN step in a (p_T, η) bin N_{gen} and the number of reconstructed charged particles in a (p_T, η) bin matched to a generated particle after passing the GEANT4 simulation of the CMS detector $N_{\text{rec}}^{\text{matched}}$:

$$\varepsilon_{\text{tr}}(p_T,\eta) = \frac{N_{\text{rec}}^{\text{matched}}(p_T,\eta)}{N_{\text{gen}}(p_T,\eta)}. \tag{6.2}$$

The reconstructed particles are matched to the generated particle according to the ΔR parameter, given by

$$\Delta R = \sqrt{(\phi_{\text{gen}} - \phi_{\text{rec}})^2 + (\eta_{\text{gen}} - \eta_{\text{rec}})^2}, \tag{6.3}$$

where ϕ is the azimuthal angle, η is the pseudorapidity, and (ϕ_{gen}, η_{gen}) and (ϕ_{rec}, η_{rec}) are the (ϕ, η) coordinates of the generated and reconstructed particles, respectively. This parameter represents the radius of an angular cone around the reconstructed particle. The generated particle with the minimum ΔR value compared to the reconstructed particle is selected as its match. A reconstructed particle is labeled as misidentified if its ΔR matching value is greater than 0.05 (following the same value used by the ATLAS Collaboration [139]). Figure 6.1 illustrates the detector's track reconstruction efficiency behavior as a function of the pseudorapidity (Figure 6.1a) and the transverse momentum (Figure 6.1b) for $\sqrt{s} = 7$ TeV. The efficiencies are almost the same for all of the collision's center-of-mass energies. The reconstruction efficiencies in Figure 6.1 agree with the corrections published by the CMS Collaboration in [36].



(a) (b)

Figure 6.1: The detector's track reconstruction efficiency as a function of the a) pseudorapidity and the b) transverse momentum for $\sqrt{s} = 0.9$, 2.76, and 7 TeV.

The simulated datasets also estimate the fraction of contaminated tracks. The secondary track rate is obtained using the information on the decay status of the generated particles in the AOD datasets, which indicates if the particle comes from the inital proton-proton collision or from a particle decay.

The cone matching provides the fake track rate, given by the fraction of reconstructed particles with a ΔR parameter greater than 0.3. Figure 6.2 displays the pseudorapidity and transverse momentum dependency of the secondary and fake track rates.

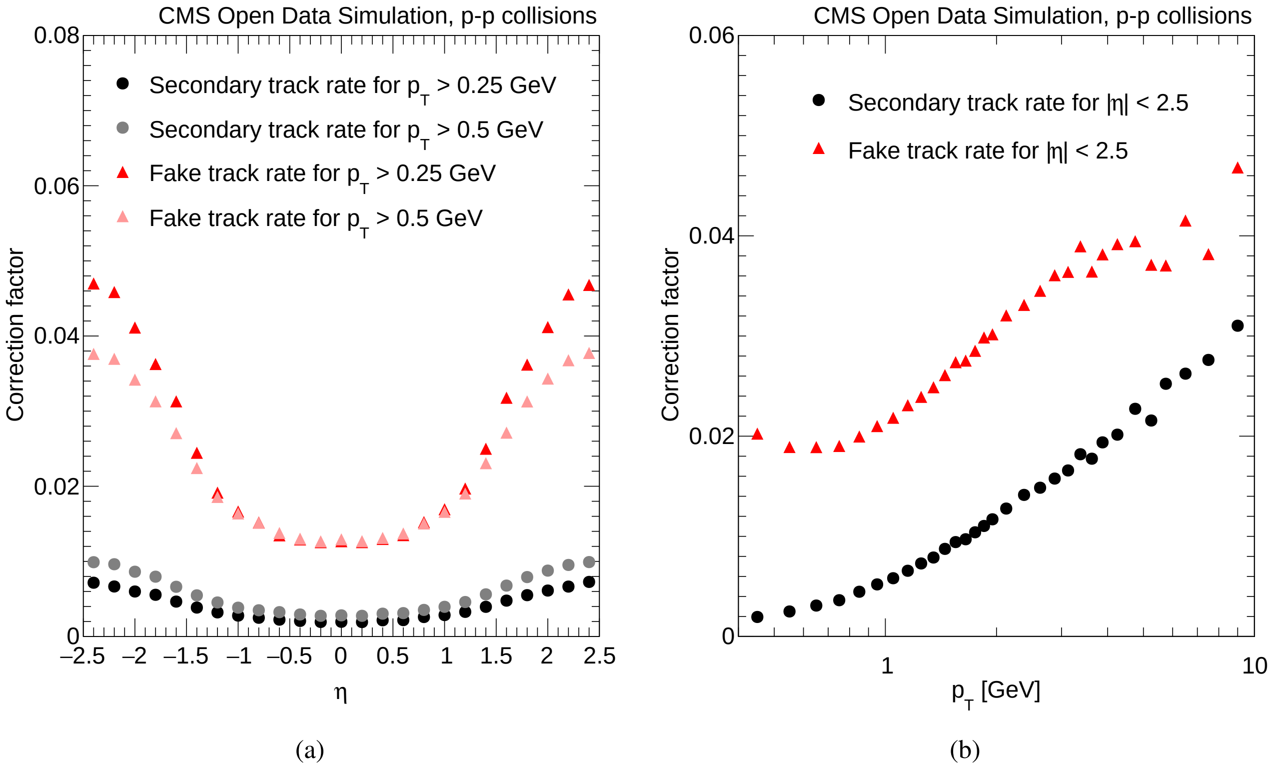


Figure 6.2: The secondary and fake track rates as a function of the a) pseudorapidity and the b) transverse momentum.

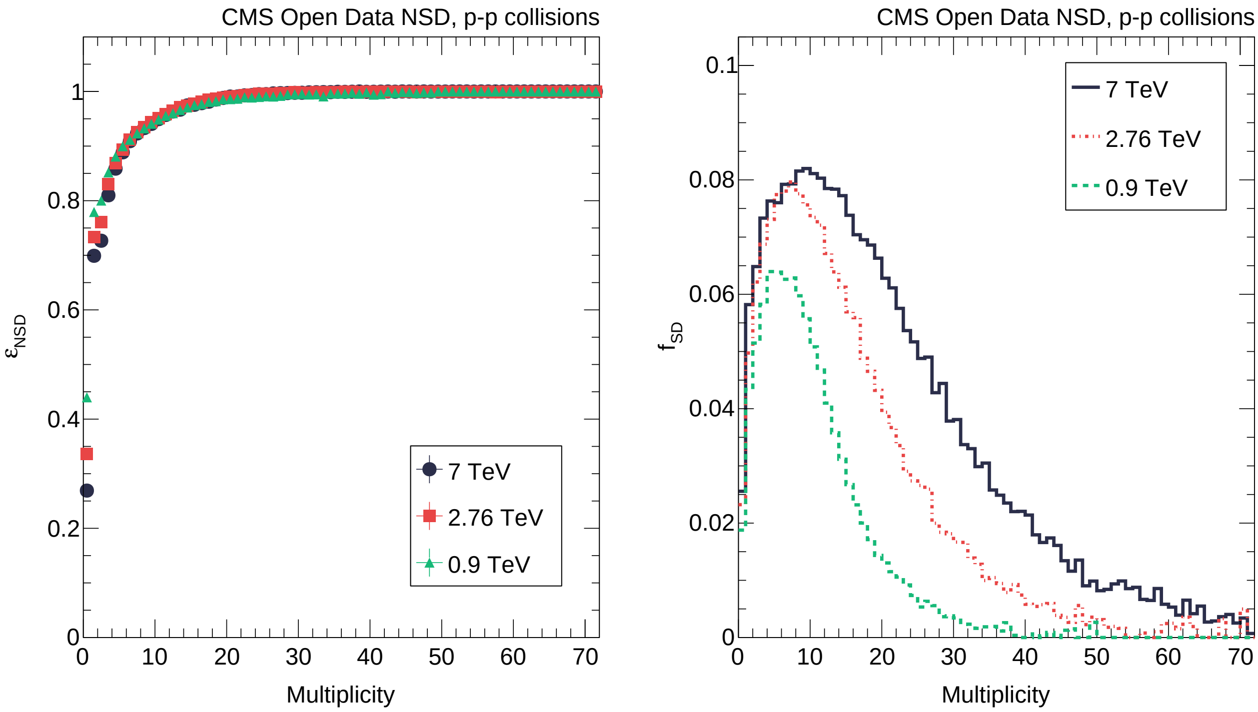
As for the mean transverse momentum distribution, a correction factor was applied to each raw multiplicity bin. The correction factor is given by the ratio $\langle p_T^{\text{gen}} \rangle / \langle p_T^{\text{rec}} \rangle$, where $\langle p_T^{\text{gen}} \rangle$ and $\langle p_T^{\text{rec}} \rangle$ are the mean transverse momentum values for each raw multiplicity bin of the generated particles before and after the detector reconstruction, respectively.

6.2 Event selection correction

The raw number of charged particles ΔN and the number of events N_{ev} must be corrected considering an event weight w_{ev}. This correction is a function of the detected multiplicity n_{ch}. Thus, w_{ev} is given by

$$w_{\text{ev}}(n_{\text{ch}}) = \frac{1}{\varepsilon_{\text{NSD}}(n_{\text{ch}})}(1 - f_{\text{SD}}(n_{\text{ch}})), \tag{6.4}$$

where $\varepsilon_{\mathrm{NSD}}$ is the NSD event selection efficiency and f_{SD} is the fraction of SD events in the final sample. The calculation of ϵ_{NSD} and f_{SD} was carried out using the generated events of the Pythia 8 datasets, for it is the only one with the collision signal ID. Figure 6.3a presents the NSD event selection efficiency as a function of the detected multiplicity. Figure 6.3b illustrates the SD event surviving fraction as a function of the detected multiplicity. Besides the corrections related to the event selection, the number of events has to be corrected considering the fraction of NSD events with no reconstructed tracks f^{0}_{NSD}, which was found to be approximately 6.5%.



(a) (b)

Figure 6.3: a) NSD selection efficiency and the b) fraction of surviving SD events in the final sample as a function of the detected charged particle multiplicity.

The NSD event selection corrections were not applied to the pseudorapidity and transverse momentum distributions in the ATLAS kinematic range, which did not contain this selection cut.

6.3 Multiplicity unfolding

For the multiplicity distributions, we have that the multiplicity spectrum obtained through the analysis of the AOD files is not the true multiplicity spectrum. Thus, the raw detected multiplicity spectrum must be unfolded to acquire the true multiplicity spectrum.

The general relation between the measured spectrum $\{M_m\}$ and the true spectrum $\{T_t\}$ is given as

$$M_m = \sum_t R_{mt} T_t, \tag{6.5}$$

where R is the response matrix that provides the probability of a measured value resulting from a given true multiplicity. Figure 6.4 illustrates the response matrices for the CMS detector (obtained with the simulated datasets) in the kinematic range of $|\eta| < 2.4$ and $p_T > 0.5$ GeV. However, to obtain $\{T_t\}$ from $\{M_m\}$, a simple inversion of Equation 6.5 is not enough because in the expression

$$T_t = \sum_m R^{-1}_{tm} M_m \tag{6.6}$$

the inverted matrix R^{-1} cannot be inferred in every case (the matrix R might be singular). Additionally, even for cases in which R is invertible, the true spectrum obtained in Equation 6.6 usually contains large statistical fluctuations.

A different method to obtain the unfolded distribution of a measured spectrum is given by [140], based on Bayes' theorem. This approach formulates that the conditional probability $P(T|M)$ of a cause T given an effect M is expressed as

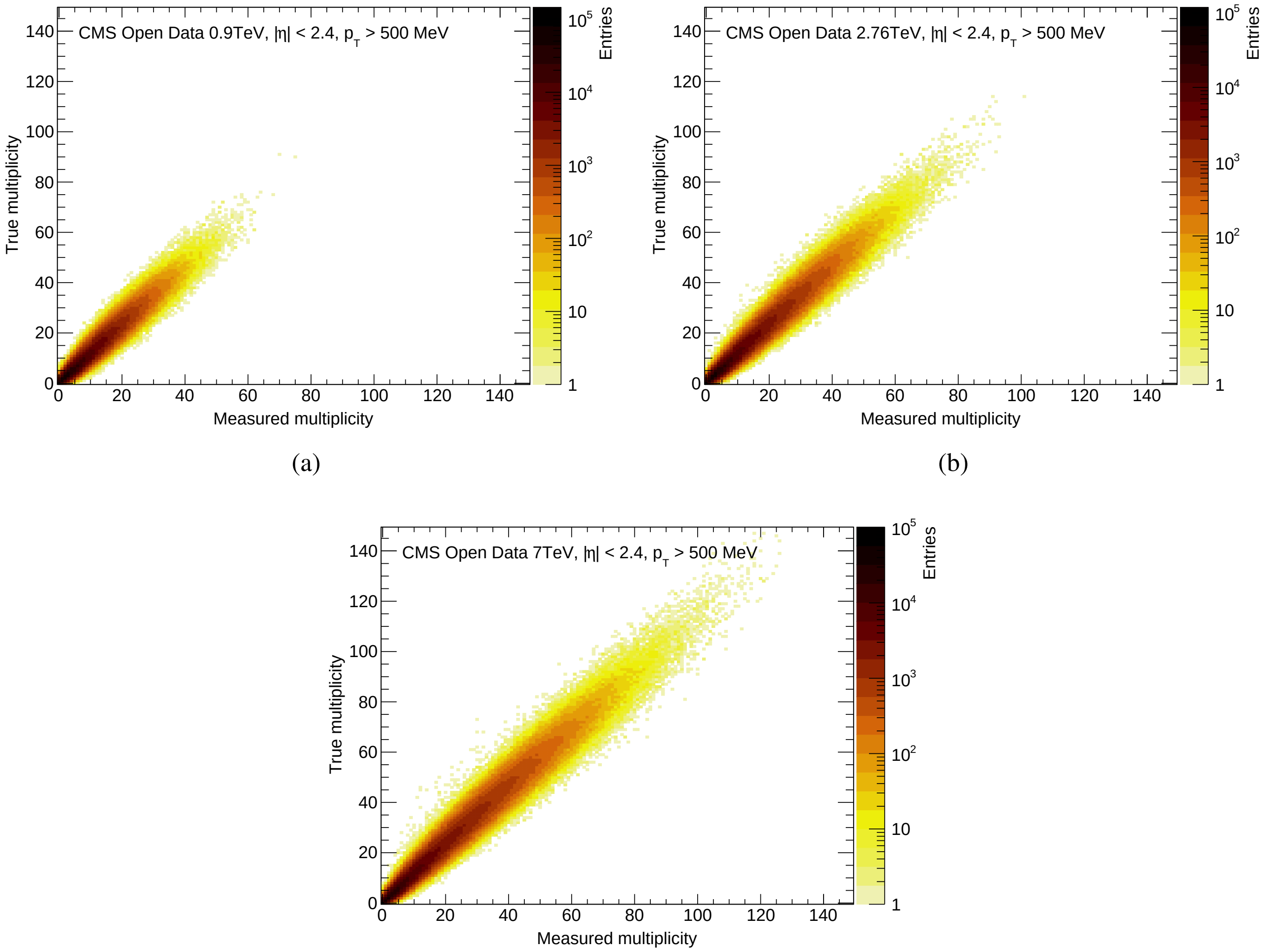
$$P(T|M) = \frac{P(M|T)P(T)}{P(M)}, \tag{6.7}$$

where $P(M)$ and $P(T)$ are the initial probabilities of M and T, respectively, and $P(M|T)$ is the conditional probability of a cause T to produce the effect M.

Considering the multiplicity distribution measurement, T is the true multiplicity of a collision within the detector, and M is the measured multiplicity containing the detector effects. The probabilities $P(T|M)$, known as elements of the smearing matrix, are the values of interest and can be determined from the components $P(M|T)$ of the response matrix. Thus, the probability spectra $\{M_m\}$ (measured distribution) and $\{T_t\}$ (true distribution) need to be previously known. However, T_t are the values that we aim to obtain. This contradiction is solved by the iterative method in [140], which expresses the elements $\widetilde{R}_{tm}$ of the smearing matrix as

$$\widetilde{R}_{tm} = \frac{R_{mt} P_t}{\sum_{t'} R_{mt'} P_{t'}}, \tag{6.8}$$

where P_t is the *a priori* distribution of the true spectrum. The next step is to calculate an unfolded



(a) (b) (c)

Figure 6.4: Unnormalized response matrix in the kinematic range of $|\eta| < 2.4$ and $p_T > 0.5$ GeV for a) $\sqrt{s} = 0.9$, b) 2.76, and c) 7 TeV.

distribution U_t using the smearing matrix and the measured multiplicity, given as

$$U_t = \sum_m \widetilde{R}_{tm} M_m. \tag{6.9}$$

For the next iteration, U_t replaces P_t as the *a priori* distribution, and the calculations in Equations 6.8 and 6.9 are carried out again. In this scenario, the first *a priori* distribution P_t is taken to be the true multiplicity obtained with the simulated datasets instead of a flat distribution as described by D'Agostini. This choice does not affect the result after a few iterations. The unfolded distribution converges after approximately four iterations. Figure 6.5 displays the smearing matrices after all iterations have been carried out.

Finally, after the convergence of U_t, the event selection weight (present in Section 6.2) is applied,

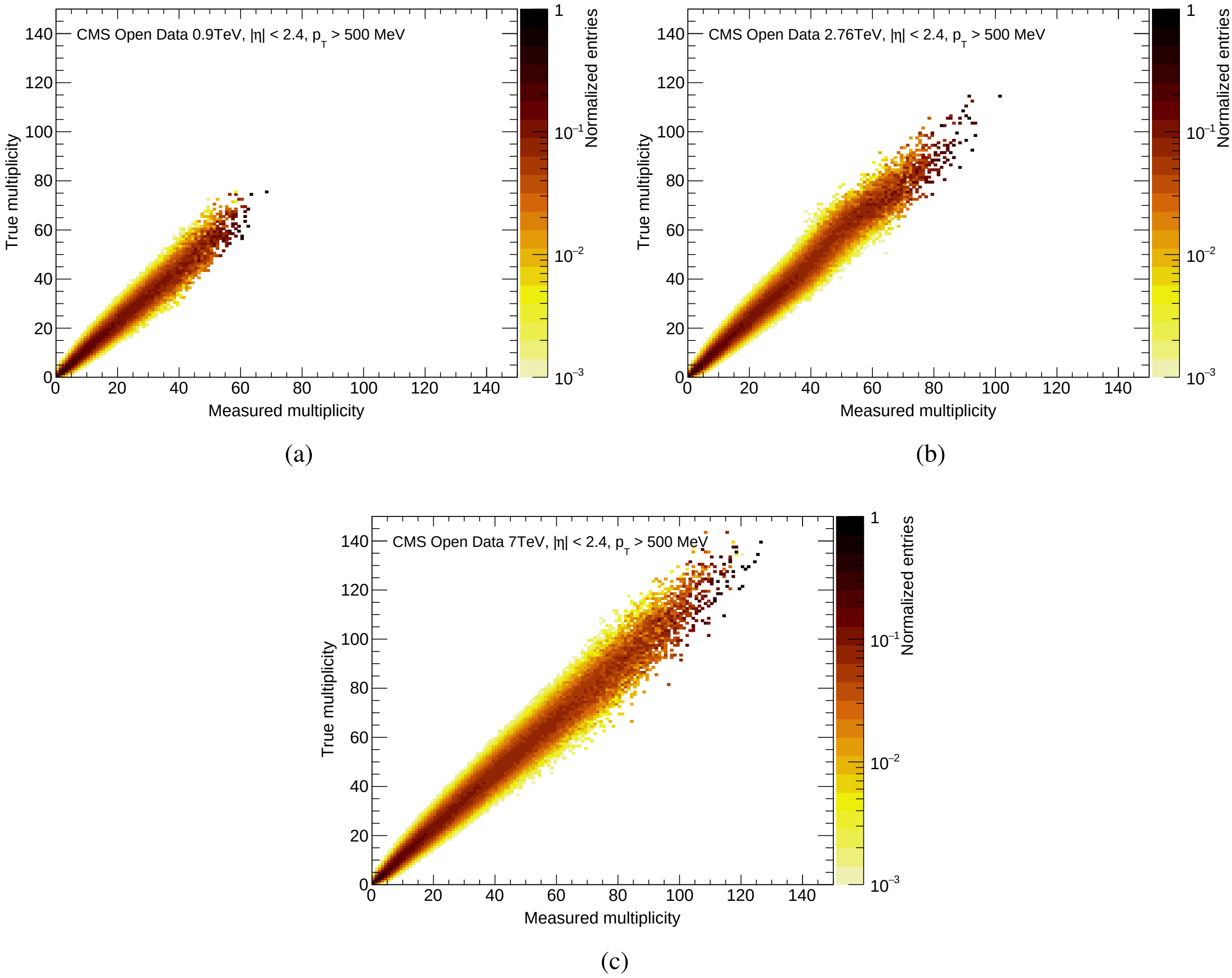


Figure 6.5: Normalized smearing matrix in the kinematic range of $|\eta| < 2.4$ and $p_T > 0.5$ GeV for a) $\sqrt{s} = 0.9$, b) 2.76, and c) 7 TeV.

and the final multiplicity distribution U'_t is given as

$$U'_t = \frac{(1 - f_{\text{SD}})}{\varepsilon_{\text{NSD}}} U_t. \tag{6.10}$$

The correct employment of the unfolding algorithm (applied with the RooUnfold library [141] in the ROOT Framework) was checked using only the simulated distributions. With the measured multiplicity set as the simulated multiplicity after the detector effects, the unfolded distribution agrees perfectly with the generated true multiplicity, as Figure 6.6 exhibits. The same sanity check was done for the mean transverse momentum distribution, as Figure 6.7 shows, where the multiplicity unfolding was applied after the correction factor $\left\langle p_T^{\text{gen}} \right\rangle / \left\langle p_T^{\text{rec}} \right\rangle$.

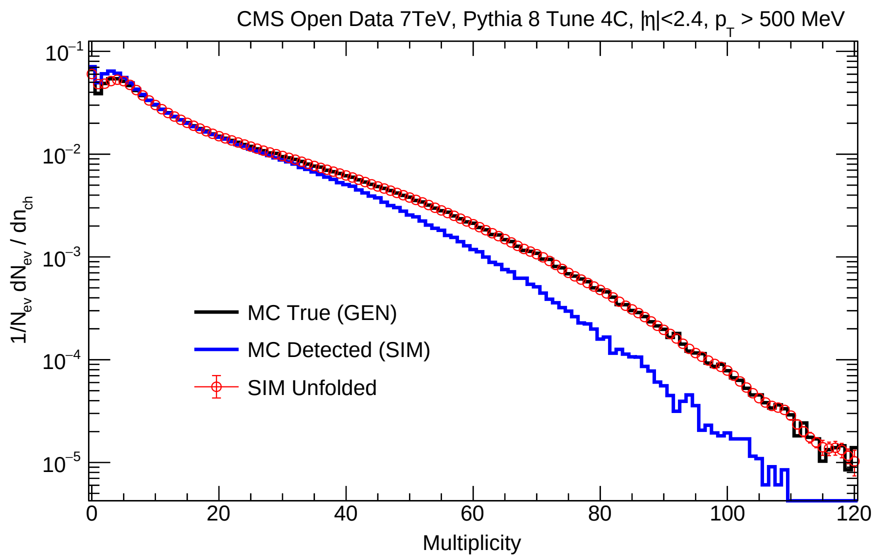


Figure 6.6: Sanity check of the unfolding algorithm for the charged particle multiplicity distribution.

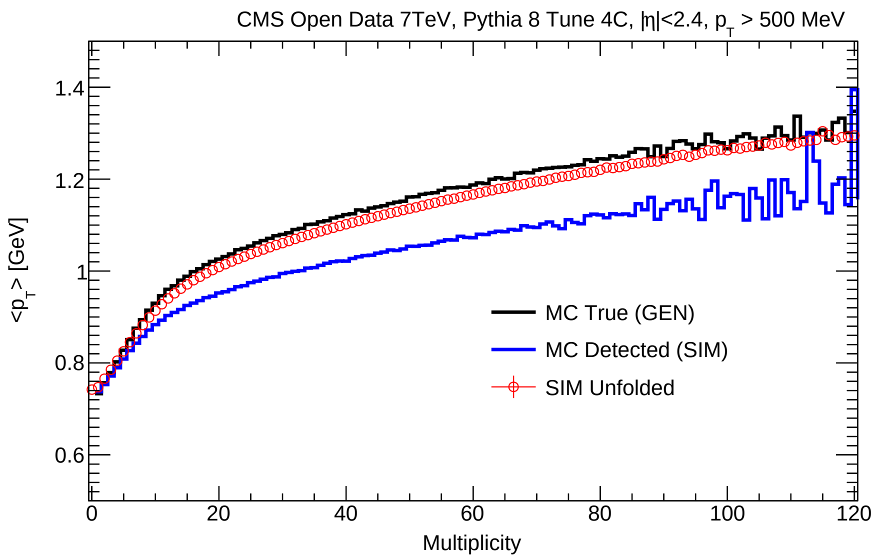


Figure 6.7: Sanity check of the unfolding algorithm for the mean transverse momentum distribution.

6.4 Transverse momentum extrapolation

Naturally, the CMS detector has a lower limit in transverse momentum for which tracks can be reconstructed. The p_T threshold for particle tracking depends on the track reconstruction method and lies between 30 MeV to approximately 100 MeV [34]. This inefficiency is displayed in Figure 6.8a, where a cut in transverse momentum can be seen for the general tracks in the CMS Open Data datasets. Another void can be observed while looking into the p_T spectrum of the reconstructed tracks that belong to a primary vertex, depicted in Figure 6.8b. This distribution suggests the existence of a threshold going up to 210 MeV for the primary vertices' particles in the range $|\eta| < 1.0$.

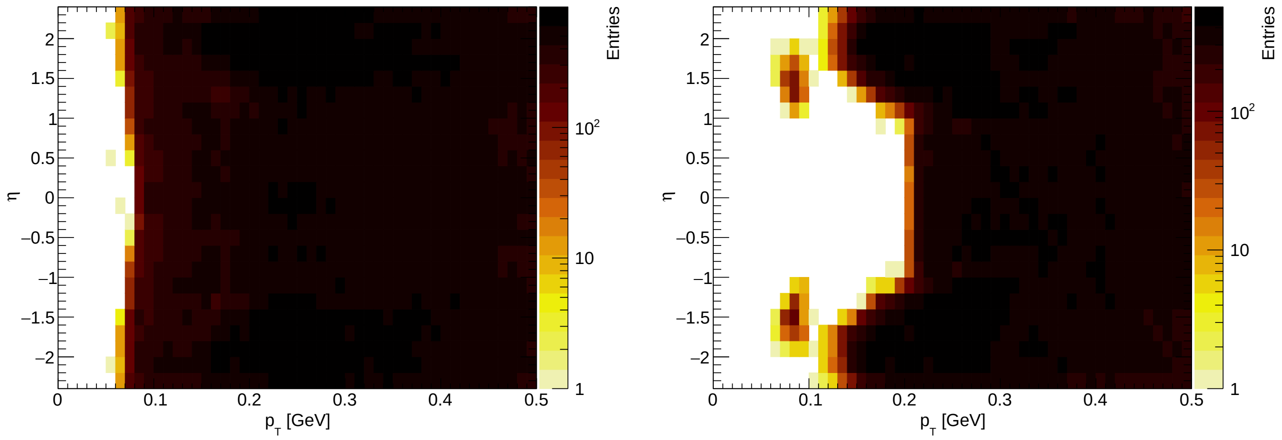


Figure 6.8: Distribution of a) tracks from the TrackCollection and b) tracks belonging to a primary vertex in the $p_T - \eta$ phase space for the dataset with $\sqrt{s} = 7$ TeV. The distributions were obtained without selection cuts.

Therefore, only particles with $p_T > 250$ MeV were selected for the transverse momentum and pseudorapidity distributions. For comparison reasons (see [34–36]), an extrapolation for $p_T \to 0$ has to be applied to the pseudorapidity distributions. This extrapolation is carried out using the distribution from nonextensive statistical mechanics in Equation 3.85, with the correction being dependent on η with bins ranging from -2.5 to 2.5 in 0.5 widths. This extrapolation accounts for approximately 25% of the charged particle pseudorapidity distribution.

As for the multiplicity distributions, a selection cut selecting only charged particles with $p_T > 500$ MeV was applied. This threshold was also chosen for comparison reasons, which repeats the kinematic range of distributions published by the ATLAS and CMS collaborations [39, 40]. The comparison with this specific kinematic range was decided because it is model-independent. Thus, an extrapolation in transverse momentum was not necessary.

6.5 Systematic uncertainties

As for systematic uncertainties, we need to consider errors related to the reconstruction algorithms, the event selection, the detector's track reconstruction efficiency, and the models used to correct the data.

Since this analysis uses datasets with reconstructed physical objects instead of the raw output of the CMS detector, the uncertainties related to the reconstruction algorithms and the track contamination rate were extracted using the CMS published papers. Additionally, the error associated with the event selection was also estimated using the CMS articles. The estimation of the event selection uncertainty was chosen because of the time consumption of calculating the error, even though its computation is possible with the CMS Open Data datasets. For example, one would need to use auxiliary triggers to extract the uncertainty related to the BSC and BPTX triggers. One such analysis uses the `HLT_MinBiasBSC_NoBPTX` trigger to analyze the BSC activity without crossing bunches, which provides how often we could get false positive signals in the scintillator.

The articles [34, 35, 125] provide an upper estimation for the systematic uncertainty of around 4-5%. These articles use the same event selection to obtain a minimum bias sample, which requires signals from the BPTX, BSC, and HF. They provide a event selection uncertainty of 3.0 [34], 3.5 [125], and 3.5 [35] percent for $\sqrt{s}$ = 0.9, 2.76, and 7 TeV, respectively. We added one additional percent to account for other sources, such as pile-up and track contamination. This percentage comes from a conservative estimation of the remaining uncertainties using the values presented in the CMS articles. It is worth reminding that not all of these articles have the same data-taking period as the data from the CMS Open Data, as exposed in Section 5.2. Nonetheless, these articles still provide a rough estimation of the systematic uncertainty.

The estimation of the remaining uncertainties related to the data corrections was carried out in this analysis. The uncertainty is estimated by calculating the distributions using the multiple simulated datasets in Table 5.2 for each center-of-mass energy. An uncertainty between 0.5 and 7 percent was obtained for the track reconstruction efficiency, which is p_T-dependent. The high values, especially for $\sqrt{s}$ = 0.9 TeV, correspond to the high-p_T part of the p_T distributions due to the lower number of tracks.

As for the extrapolation for $p_T < 0.25$ GeV, this correction is model-dependent since this analysis opted for using the Tsallis distribution. Even though this parametrization is known for efficiently describing the transverse momentum distribution, it still has some modifications in which equations are employed. For example, we could use the Equation 3.85, which is thermodynamically consistent, or an equation that only follows the non-extensive ansatz, such as [15]

$$E\frac{d^3N}{dp^3} = A\left[1-(1-q)\frac{m_T}{T}\right]^{1/(1-q)}, \tag{6.11}$$

and still fit the transverse momentum spectrum. These different descriptions provide an uncertainty of 1-2% for the extrapolation for $p_T \to 0$. This uncertainty is higher for lower values of $\sqrt{s}$ because

p_T distributions with lower center-of-mass energies have a lower percentage of charged particles in the $p_T < 250$ MeV region compared to collisions with higher $\sqrt{s}$.

The uncertainty related to the unfolding procedure was obtained by unfolding the multiplicity distribution with response matrices calculated from different simulated datasets and by varying the number of iterations. The calculation with multiple response matrices was carried out as a sanity check, for the Bayesian unfolding procedure is model-independent. The error is n_{ch}-dependent and was found to be 0.1-2.2% and 0.1-10.0% for $\sqrt{s} = 0.9$ and 7 TeV, respectively. The high upper limit for $\sqrt{s} = 0.9$ TeV is due to the low number of events with high multiplicity. Table 6.1 shows the systematic uncertainty value from the different sources for the datasets with $\sqrt{s} = 0.9$, 2.76, and 7 TeV.

$\sqrt{s}$	0.9 TeV	2.76 TeV	7 TeV
Source	Systematic uncertainty (%)		
Track reconstruction	0.5-7.0	0.5-3.5	0.5-3.5
Extrapolation for $p_T < 250$ MeV	1.8	1.6	1.5
Multiplicity unfolding	0.1-10.0	-	0.1-2.2
Remaining uncertainties (estimated from CMS papers)	4.0	4.5	4.5

Table 6.1: Summary of the systematic uncertainties affecting the charged particle distributions for $\sqrt{s} = 0.9$, 2.76, and 7 TeV. Uncertainties were also estimated from CMS papers for $\sqrt{s} = 0.9$ [34], 2.76 [125], and 7 TeV [35].

7 | Results

This chapter resumes the final charged particle distributions measured with the open data from the CMS Collaboration. Measurements of the transverse momentum, pseudorapidity, multiplicity, and $\langle p_T \rangle$ spectrum for $\sqrt{s}$ = 0.9, 2.76, and 7 TeV are presented, and a comparison with papers published by experimental collaborations is discussed.

7.1 Multiplicity and $\langle p_T \rangle$ distributions

Figure 7.1 shows the charged particle multiplicity distributions for $\sqrt{s} = 0.9$ and 7 TeV and their comparison with the published data from the ATLAS and CMS collaborations [39, 40]. The ratio of the published charged particle multiplicity distributions with respect to our results is shown in the lower panel. Only tracks in the kinematic range of $|\eta| < 2.4$ and $p_T > 500$ MeV were considered.

Figure 7.2 illustrates the mean transverse momentum spectrum as a function of the multiplicity for $\sqrt{s} = 0.9$ and 7 TeV and its comparison with the distributions from the ATLAS Collaboration [39]. The ratio of the published charged particle $\langle p_T \rangle$ distributions with respect to our results is displayed in the lower panel.

The multiplicity and $\langle p_T \rangle$ distributions provide the first evidence that the plots obtained using the open data from the CMS Collaboration are in reasonable agreement with the published results. The distributions obtained with the open data and those published by the CMS Collaboration both contain NSD events. However, the distributions from the ATLAS Collaboration do not include a selection cut favoring a specific diffractive process. This difference is visible in the low multiplicity part of the spectrum, which contains more SD events. Thus, the multiplicity distributions obtained in this dissertation are closer to those published by the CMS Collaboration. As for the $\langle p_T \rangle$ results, the open data distributions do not deviate by more than 4% of the ATLAS distributions, even though it has a slightly different event selection.

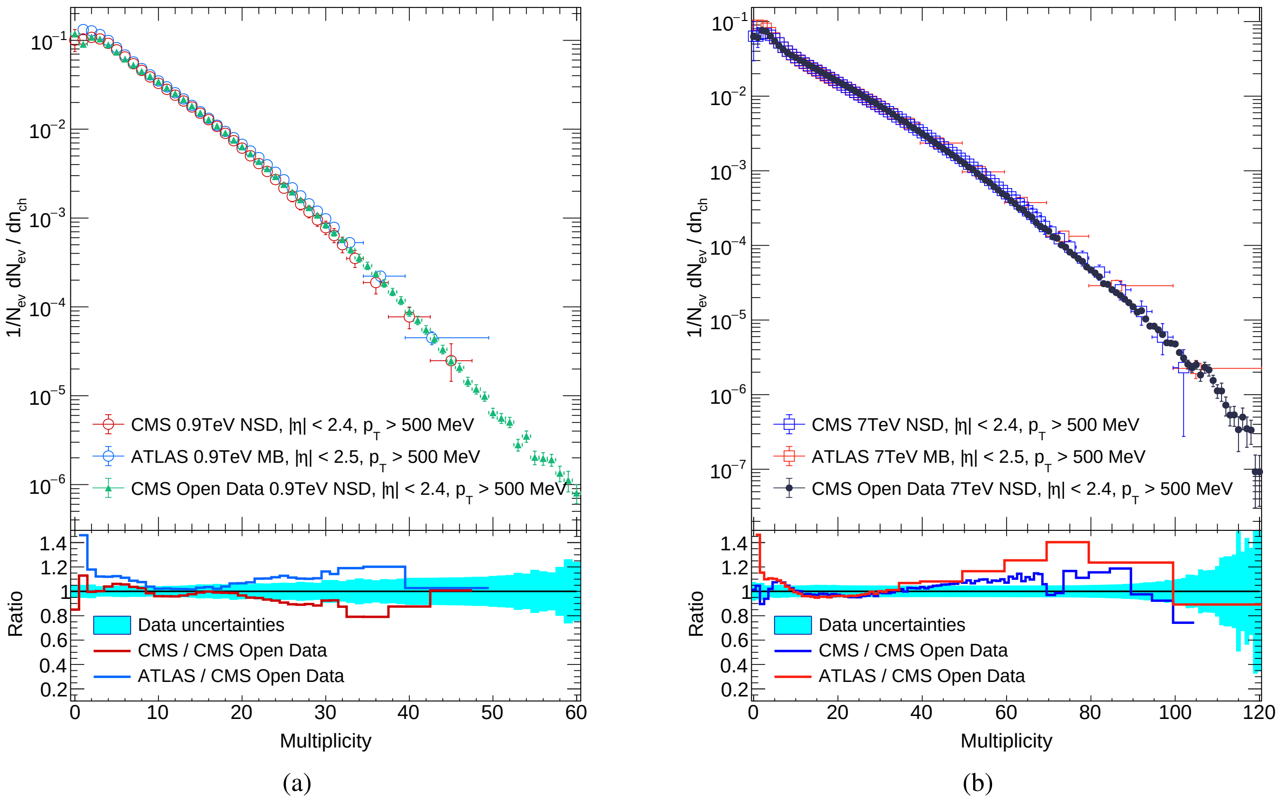


Figure 7.1: Charged particle multiplicity distributions in the $|\eta| < 2.4$ and $p_T > 500$ MeV kinematic range for a) 0.9 TeV and b) 7 TeV. Distributions from the ATLAS [39] and CMS [40] collaborations are also presented. The ratio of the published distributions with respect to our results is shown in the lower panel. The vertical bars represent the statistical and systematic uncertainties added in quadrature.

7.2 Pseudorapidity distributions

The pseudorapidity distributions were calculated using that

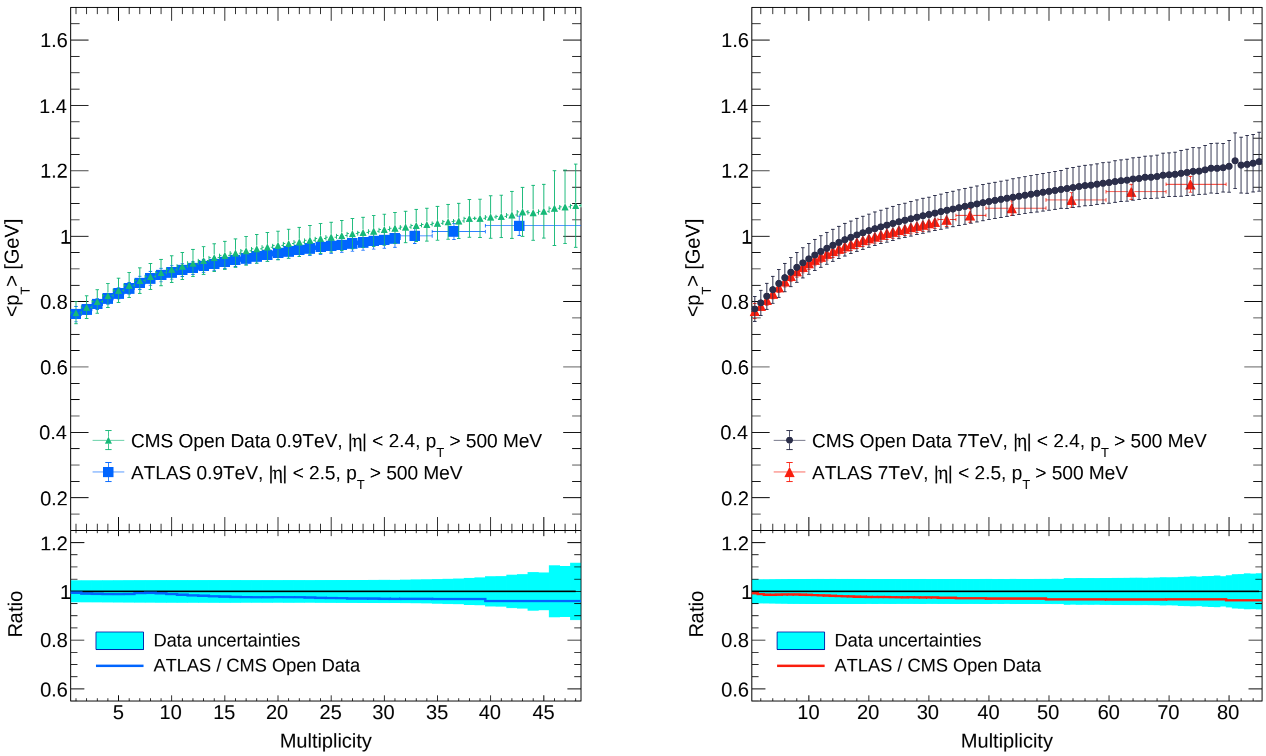
$$\frac{1}{N_{\text{ev}}}\frac{dN}{d\eta} = \frac{1}{N_{\text{ev}}}\frac{N_{\text{tr}}}{\Delta\eta}, \tag{7.1}$$

where $\Delta\eta$ is the width of the η bin, ΔN_{tr} is the corrected number of tracks in a (p_T,η) bin, and N_{ev} is the corrected number of events that survived the selection cuts. The corrected number of tracks and events are given by

$$N_{\text{tr}} = \sum_{n_{\text{ch}}} N_{\text{tr}}^{\text{raw}}(p_T, \eta, n_{\text{ch}}) \cdot w_{\text{tr}}(p_T, \eta) \cdot w_{\text{ev}}(n_{\text{ch}}) \tag{7.2}$$

and

$$N_{\text{ev}} = (1 + f_{\text{NSD}}^0) \sum_{n_{\text{ch}}} N_{\text{ev}}^{\text{selected}}(n_{\text{ch}}) \cdot w_{\text{ev}}(n_{\text{ch}}). \tag{7.3}$$



(a) (b)

Figure 7.2: Charged particle $\langle p_T \rangle$ distributions as a function of the charged particle multiplicity in the $|\eta| < 2.4$ and $p_T > 500$ MeV kinematic range for a) 0.9 TeV and b) 7 TeV. Distributions from the ATLAS Collaboration [39] are also present. The ratio of the ATLAS distributions with respect to our results is shown in the lower panel. The vertical bars represent the statistical and systematic uncertainties added in quadrature.

respectively, where n_{ch} is the event multiplicity (number of charged particles in an event), and w_{tr} and w_{ev} are the track and event weights outlined in Chapter 6. As mentioned in Section 6.2, the corrections w_{ev} and f^0_{NSD} were not applied for the pseudorapidity and transverse momentum distributions in the ATLAS kinematic range.

For comparison reasons, the pseudorapidity distributions in this dissertation are in two kinematic regions. The first one contains only tracks with $|\eta| < 2.5$ and $p_T > 500$ MeV and events with $n_{ch} \geq 1$ for the comparison with the distributions published by the ATLAS Collaboration. Figure 7.3 exhibits the pseudorapidity distributions in this phase space for $\sqrt{s} = 0.9$, 2.76, and 7 TeV and its comparisons with the results from the ATLAS Collaboration. The ratio of the ATLAS charged particle pseudorapidity distributions with respect to our results is shown in the lower panel.

The second kinematic region includes only tracks with $|\eta| < 2.5$ and $p_T > 0$ MeV and events with $n_{ch} \geq 0$ to compare to the distributions published by the CMS Collaboration. The CMS Collaboration has track reconstruction algorithms with different p_T thresholds, going from as low as 30 MeV

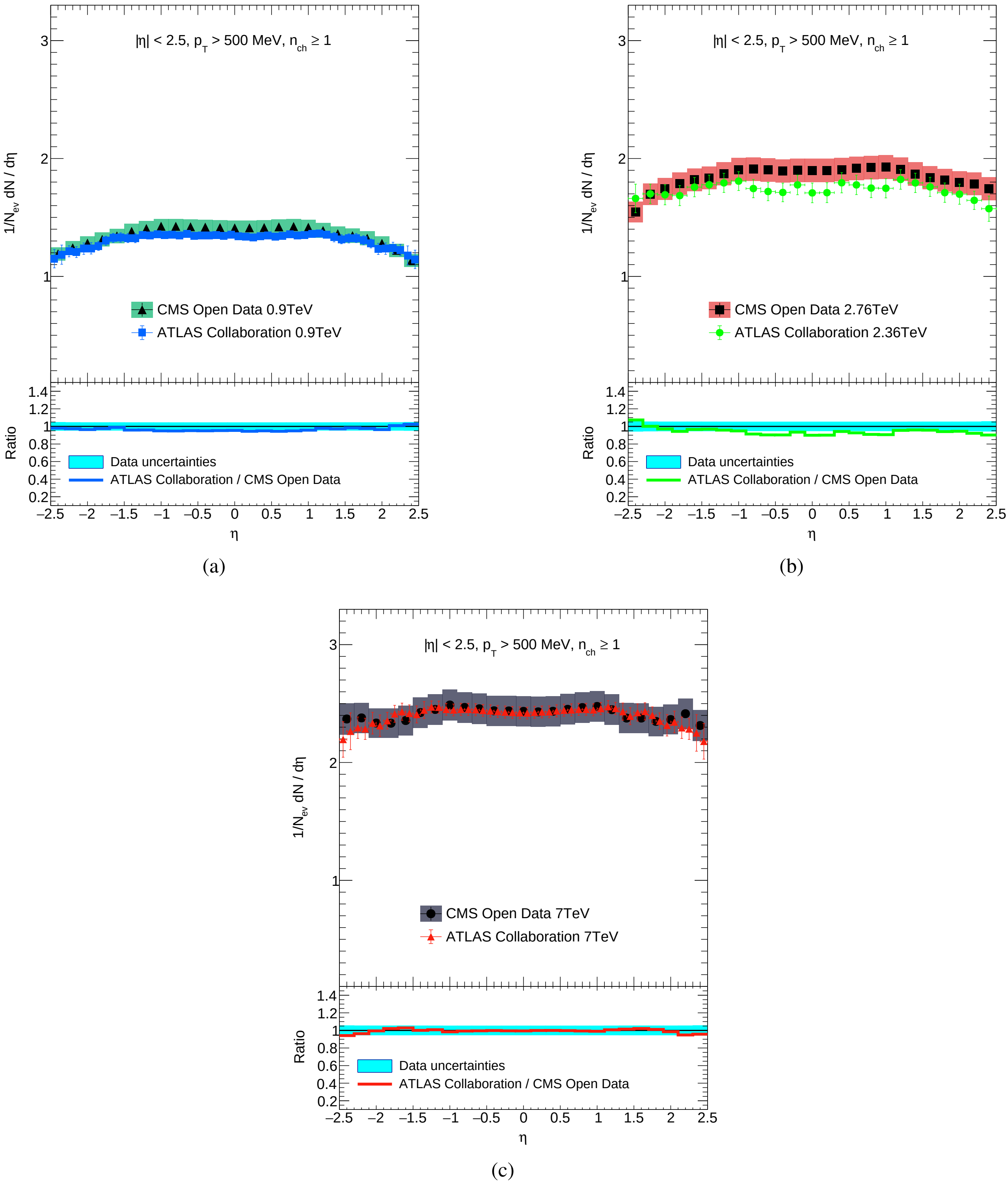


Figure 7.3: Charged particle pseudorapidity distributions with $p_T > 500$ MeV and $n_{ch} \geq 1$ for a) 0.9 TeV, b) 2.76 TeV, and c) 7 TeV. Charged particle pseudorapidity distributions from the ATLAS Collaboration [39] are also presented. The ratio of the ATLAS distributions with respect to our results is shown in the lower panel. The vertical bars represent the total uncertainty of the ATLAS data points. The shaded bands represent the statistical and systematic uncertainties of the distributions from this dissertation added in quadrature.

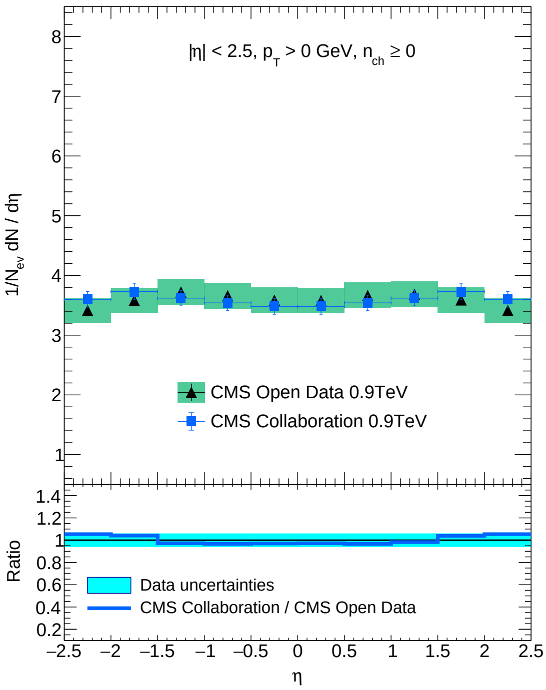
for one method (pixel counting method [34]) to 100 MeV for the one with the largest cutoff value (tracking method [34]). Thus, they extrapolated their pseudorapidity distributions to $p_T \to 0$ for all reconstruction methods. As mentioned in Section 6.4, the cutoff value for the distributions in this dissertation is 250 MeV. Therefore, an extrapolation using the Tsallis distribution is applied to each η bin to consider charged particles with p_T between 0 and 250 MeV. Figure 7.4 displays the pseudorapidity distributions in the CMS phase space for $\sqrt{s}$ = 0.9, 2.76, and 7 TeV and its comparisons with the results from the CMS Collaboration. The ratio of the CMS charged pseudorapidity distributions with respect to our results is displayed in the lower panel.

The ATLAS and CMS experiments did not publish pseudorapidity distributions for $\sqrt{s}$ = 2.76 TeV. However, they published pseudorapidity distribution with $\sqrt{s}$ = 2.36 TeV, close enough to draw some comparisons. The LHC 2.36 TeV run was carried out in 2010, during the first months of operation, and it does not have datasets available in the CMS Open Data. Table 7.1 presents the values obtained for the average charged particle multiplicity density $dN/d\eta|_{\eta\approx 0}$ and its comparison with the values obtained by the ATLAS and CMS collaborations for their specific kinematic ranges. The CMS Collaboration published charged hadron pseudorapidity distributions instead of using a broader selection including all charged particles. However, the comparison between this dissertation's distributions and the CMS distributions is still valid since leptons correspond to less than 1% of the charged particle distribution [34, 35]. All the other distributions from the ATLAS and CMS experiments present in this dissertation correspond to charged particle distributions.

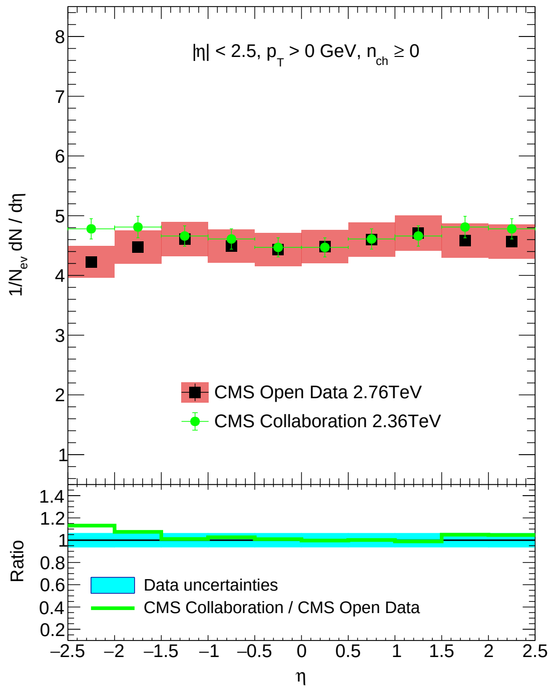
$\sqrt{s}$ (TeV)	$dN/d\eta\|_{\eta\approx 0}(n_{ch} \geq 0, p_T > 0$ MeV)		$dN/d\eta\|_{\eta\approx 0}(n_{ch} \geq 1, p_T > 500$ MeV)	
	This dissertation	CMS pub.	This dissertation	ATLAS pub.
0.9	3.6 ± 0.2	$3.48 \pm 0.02^{+0.13}_{-0.13}$	1.41 ± 0.06	$1.343 \pm 0.004^{+0.027}_{-0.027}$
2.36	–	$4.47 \pm 0.04^{+0.16}_{-0.16}$	–	$1.74 \pm 0.019^{+0.058}_{-0.058}$
2.76	4.5 ± 0.2	–	1.90 ± 0.09	–
7.0	5.5 ± 0.3	$5.78 \pm 0.01^{+0.23}_{-0.23}$	2.4 ± 0.1	$2.423 \pm 0.001^{+0.050}_{-0.050}$

Table 7.1: Comparison between the average charged multiplicity density values for $\eta \approx 0$ obtained in this paper with those published by the CMS and ATLAS collaborations [34, 35, 39]. The values from this dissertation and from the ATLAS Collaboration are derived from charged particle distributions. The values from the CMS Collaboration are obtained from charged hadron distributions.

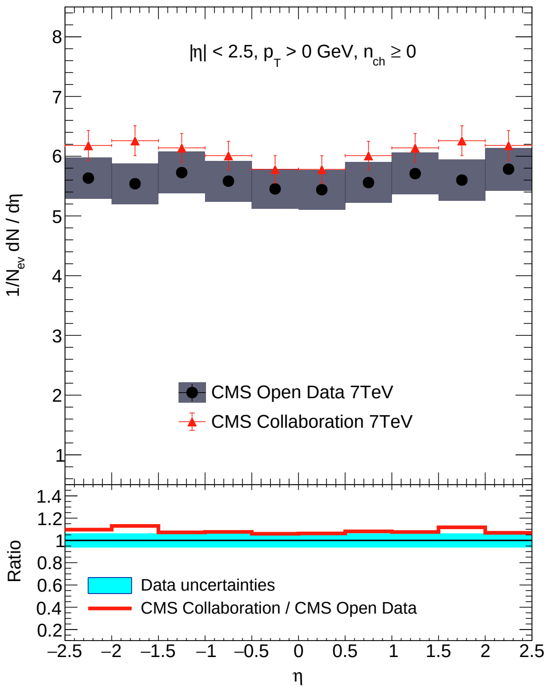
The pseudorapidity distributions provide the second test of whether the open data distributions are replicating the published results. Figures 7.3 and 7.4 show that the ratio between this dissertation's



(a)



(b)



(c)

Figure 7.4: Charged particle pseudorapidity distributions with $p_T > 0$ MeV and $n_{ch} \geq 0$ for a) 0.9 TeV, b) 2.76 TeV, and c) 7 TeV. Charged hadron pseudorapidity distributions from the CMS Collaboration [34, 35] are also present. The ratio of the CMS distributions with respect to our results is shown in the lower panel. The vertical bars represent the total uncertainty of the CMS data points. The shaded bands represent the statistical and systematic uncertainties of the distributions from this dissertation added in quadrature.

distributions and the data from the experimental collaborations lies within the experimental error. The spectrum ratios have an upper deviation of around 5% of the obtained value in the mid-rapidity region ($\eta \approx 0$).

The comparison of the pseudorapidity distributions also provides the first application of the Tsallis parametrization to describe the transverse momentum distributions. Since the results of the ATLAS Experiment had a p_T cutoff of 500 MeV, the plots for this kinematic range are model-independent (the data corrections relate only to the detector effects and not the event generators). But that is not the case for the distributions in the kinematic range of the CMS detector. As mentioned earlier, the CMS Experiment carries out an extrapolation for charged hadrons with transverse momentum below 100 MeV, which account for an increase of 5% in the estimated number of particles [34]. However, they do not mention how they calculate the extrapolation. As for our distributions, they rely on a larger extrapolation provided by the Tsallis parametrization. This correction supplies a result in agreement with the model-dependent spectrum published by the CMS Collaboration. It is worth noting that this correction belongs to the part of the spectrum described by nonperturbative QCD. Thus, by comparing our distributions with the ATLAS and CMS experiments, we have results agreeing with measurements containing both model-dependent and model-independent corrections.

7.3 Transverse momentum distributions

The corrected charged particle transverse momentum distributions were calculated using that

$$\frac{1}{N_{\text{ev}}}\frac{1}{2\pi p_T}\frac{d^2N}{d\eta dp_T}(p_T,\eta) = \frac{1}{N_{\text{ev}}}\frac{1}{2\pi p_T}\frac{\Delta N_{\text{tr}}}{\Delta\eta\Delta p_T} \tag{7.4}$$

where Δp_T is the width of the p_T bin. The yield given by Equation 7.4 was used to compare the distributions obtained with the ones from both the ATLAS and CMS collaborations. Notice that the differential yield Ed^3N/dp^3 from the CMS Collaboration was calculated similar to the way presented in Equation 7.4 (aside from the pileup correction), as exposed in reference [36]. Therefore, for comparison reasons, the additional term in Appendix A.2 was not applied to the distributions.

Figure 7.5 shows the charged particle differential yield as a function of the transverse momentum in the ATLAS phase space for $\sqrt{s}$ = 0.9, 2.76, and 7 TeV and its comparisons with the results from the ATLAS Collaboration. The ratio of the ATLAS charged particle transverse momentum distributions with respect to our results is shown in the lower panel. Since the ATLAS Collaboration did not publish transverse momentum distributions with $\sqrt{s} = 2.76$ TeV, this dissertation compares the distribution with $\sqrt{s} = 2.76$ TeV to a distribution for collisions with $\sqrt{s} = 2.36$ TeV. Figure 7.6 illustrates the charged particle differential yield as a function of the transverse momentum in the CMS kinematic range for $\sqrt{s} = 0.9$, 2.76, and 7 TeV and its comparisons with the results from the CMS Collaboration. The ratio of the CMS charged particle transverse momentum distributions with respect to our results is displayed in the lower panel. The histograms that provide the distributions in Figures 7.5 and 7.6 contain the same bin arrangement as the plots published by the experimental collaborations.

One important aspect when considering dealing with the CERN Open Data datasets is how far on the p_T spectrum we could have meaningful statistics. In this dissertation, we present p_T spectra going up to around 20, 30, and 50 GeV for $\sqrt{s} = 0.9$, 2.76, and 7 TeV, respectively. This threshold was the highest possible cutoff value using the open data datasets. The main setback for higher p_T values came from the detector correction estimation, where there was not a sufficient number of events to estimate the corrections. Nevertheless, the fraction of tracks with $p_T > 20$ GeV is of the order of 10^{-6}, which results in a large statistical uncertainty since our datasets have only 10^7 events. Thus, only about $10^1 - 10^2$ charged particles have $p_T > 20$ GeV in our datasets.

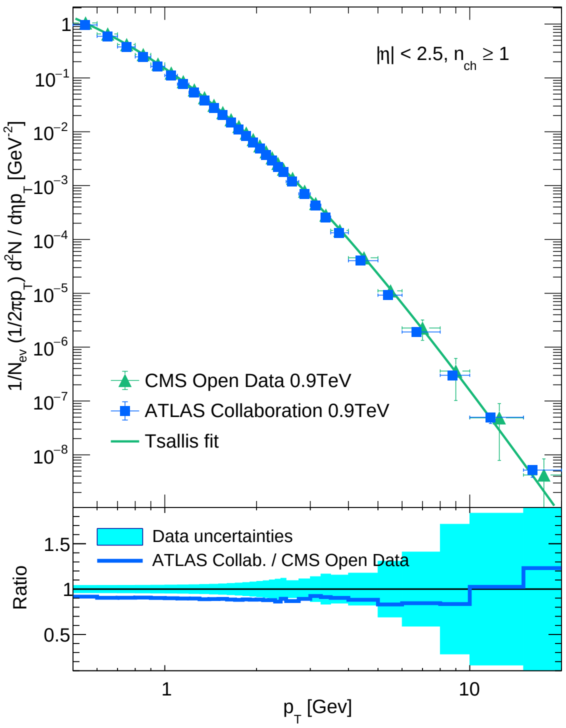
However, the lack of events is not a problem for the experimental collaborations, where results are published with a spectrum going up to 200 GeV [142]. This difference arises because we do not have access to all data taking runs through the CERN Open Data Portal. Nevertheless, this problem might be specific for this analysis since we rely only on low pile-up runs, usually carried out during the commissioning step of the experiment. If we could use general Minimum Bias datasets, we could obtain distributions with data from the 2010A, 2010B, 2011A, and 2011B runs (as an example for collisions with $\sqrt{s} = 7$ TeV), and that should be enough.

Unlike the pseudorapidity distributions, the transverse momentum distribution corrections are all model-independent and only reflect the detector effects. Figures 7.5 and 7.6 display that the open data distributions and the ones published by the experimental collaboration exhibit roughly the same behavior. The similarity in the distribution shape is confirmed by the fit using the Tsallis parametrization, where the charged particle differential yield was fitted using the equations

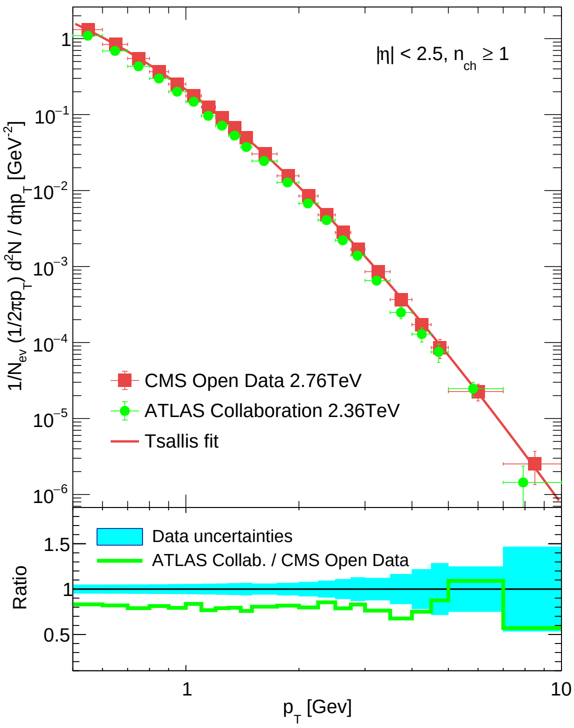
$$\frac{1}{N_{\text{ev}}}\frac{1}{2\pi p_T}\frac{d^2N}{d\eta dp_T} = Am_T\left[1 + (q-1)\frac{m_T}{T}\right]^{-\frac{q}{(q-1)}} \tag{7.5}$$

and

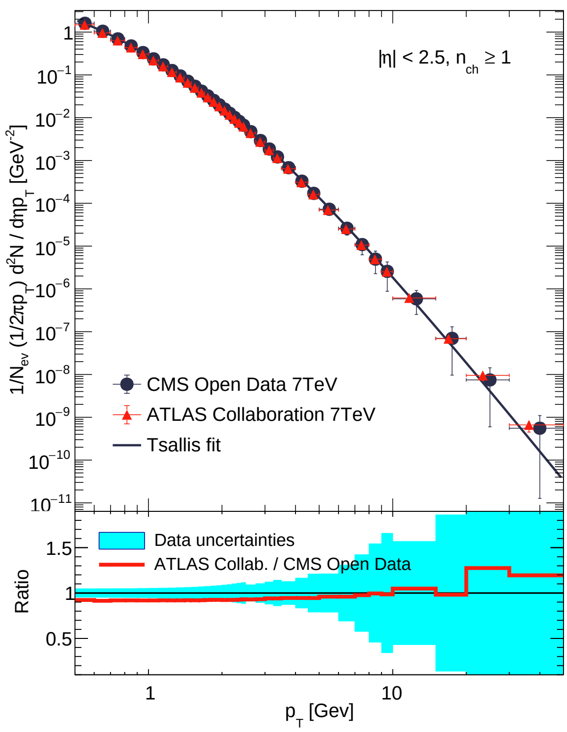
$$\frac{1}{N_{\text{ev}}}\frac{1}{2\pi p_T}\frac{d^2N}{d\eta dp_T} = A\left[1 + (q-1)\frac{m_T}{T}\right]^{-\frac{1}{(q-1)}}. \tag{7.6}$$



(a)

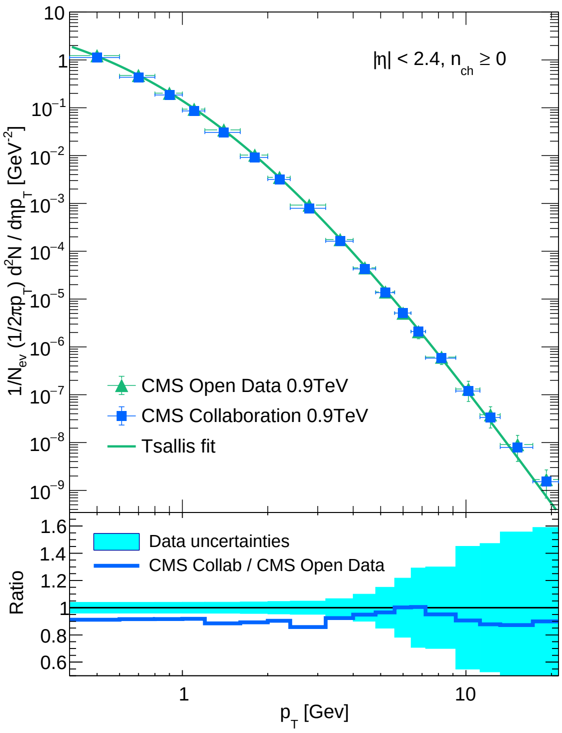


(b)

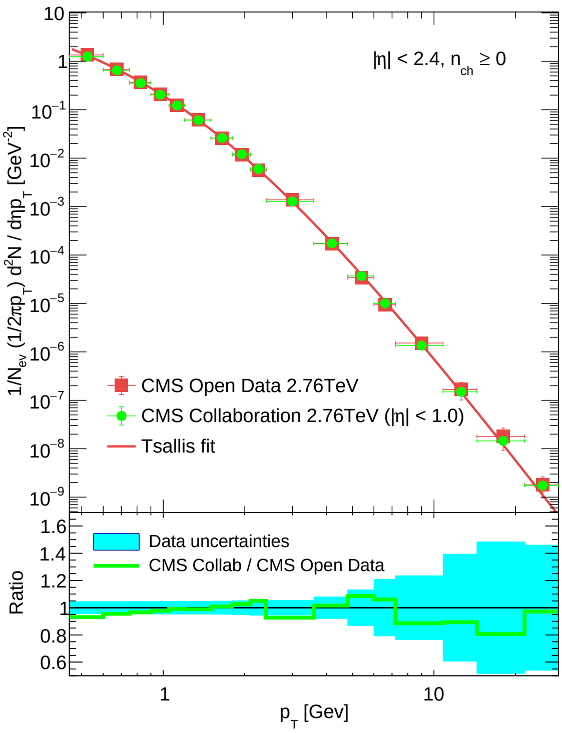


(c)

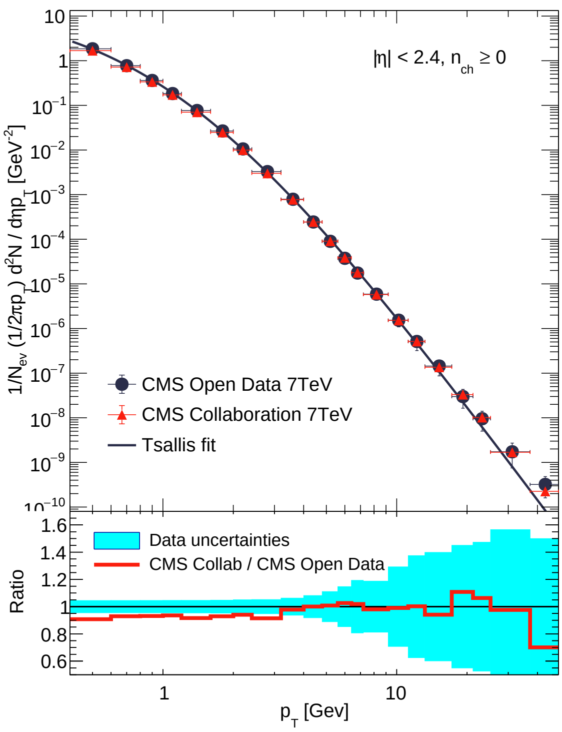
Figure 7.5: Charged particle differential yield as a function of the transverse momentum with $|\eta| < 2.5$ and $n_{ch} \geq 1$ for a) 0.9 TeV, b) 2.76 TeV, and c) 7 TeV. Distributions from the ATLAS Collaboration [39] are also present. The ratio of the ATLAS distributions with respect to our results is shown in the lower panel. The vertical bars represent the statistical and systematic uncertainties added in quadrature.



(a)



(b)



(c)

Figure 7.6: Charged particle differential yield in NSD events as a function of the transverse momentum with $|\eta| < 2.4$ and $n_{ch} \geq 0$ for a) 0.9 TeV, b) 2.76 TeV, and c) 7 TeV. Distributions from the CMS Collaboration [36, 125] are also present. The ratio of the CMS distributions with respect to our results is shown in the lower panel. The vertical bars represent the statistical and systematic uncertainties added in quadrature.

Equation 7.5 follows the thermodynamically consistent approach to nonextensive statistical mechanics, described in Chapter 3. Equation 7.6 is the standard distribution of the nonextensive ansatz used in other articles [15, 143] and is related to the one used by experimental collaborations.

The distributions used for the fit comparison are displayed in Figure 7.7 and use a different bin arrangement than the plots in Figures 7.5 and 7.6. The different bin array was chosen to obtain fit parameters with smaller uncertainties. Table 7.2 displays the fit parameters obtained using the Tsallis fit (Equation 7.5) for the plots in Figure 7.7, along with the parameter values for the ATLAS and CMS distributions. Figure 7.8 presents the values of the parameters q and T using both Equations 7.5 and 7.6. The fit of all datasets was performed in the transverse momentum range between 500 MeV and 10 GeV.

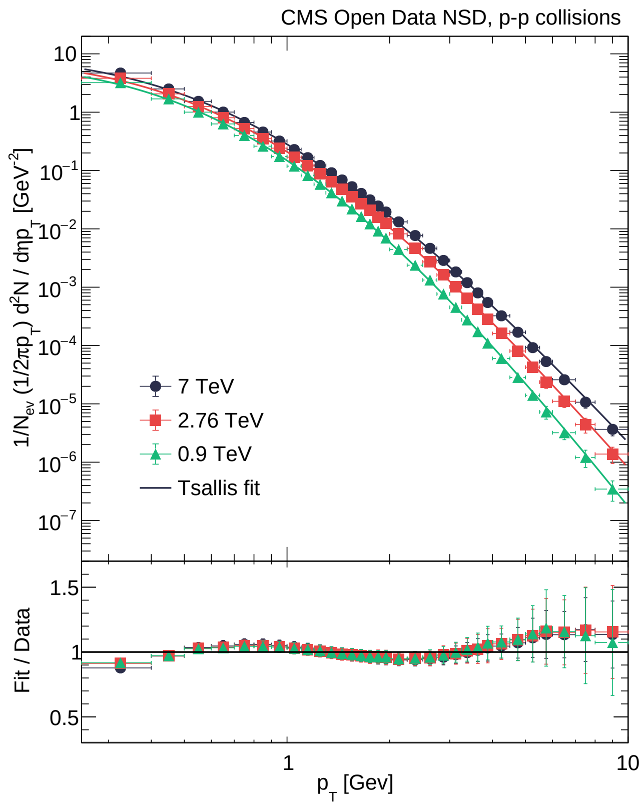


Figure 7.7: Invariant differential charged particle yield in NSD events in the $|\eta| < 2.4$ range as a function of the transverse momentum for $\sqrt{s} = 0.9$, 2,76, 7 TeV. The data was fitted to the Tsallis distribution. The vertical bars represent the statistical and systematic uncertainties added in quadrature.

Figure 7.8a shows that the q-index rises slightly with the center-of-mass energy of the collision, which was something already present in the literature [15, 144]. The obtained values of the q-index correspond to a power-law index n in the $8 - 10$ range when using the thermodynamically consistent approach. This range extends to $6 - 10$ when using Equation 7.6. This result allows comparing the

$\sqrt{s}$	A (GeV^{-2})	q	$n \rightarrow q/(q-1)$	T (GeV)
CMS Open Data 7 TeV	166 ± 19	1.143 ± 0.004	8.0 ± 0.2	0.114 ± 0.004
CMS 7 TeV	169 ± 175	1.144 ± 0.010	7.9 ± 0.5	0.113 ± 0.019
ATLAS 7 TeV	162 ± 56	1.143 ± 0.006	8.0 ± 0.3	0.117 ± 0.011
CMS Open Data 2.76 TeV	162 ± 20	1.135 ± 0.004	8.4 ± 0.2	0.109 ± 0.005
CMS 2.76 TeV	146 ± 127	1.135 ± 0.012	8.4 ± 0.7	0.113 ± 0.017
ATLAS 2.36 TeV	108 ± 130	1.139 ± 0.015	8.2 ± 0.8	0.105 ± 0.019
CMS Open Data 0.9 TeV	148 ± 17	1.123 ± 0.003	9.1 ± 0.2	0.106 ± 0.005
CMS 0.9 TeV	157 ± 181	1.123 ± 0.009	9.1 ± 0.6	0.105 ± 0.019
ATLAS 0.9 TeV	154 ± 79	1.121 ± 0.008	9.3 ± 0.5	0.107 ± 0.010

Table 7.2: Parameters obtained with the Tsallis fit for the charged particle differential yield distributions. The CMS and ATLAS parameters were obtained by fitting the data from the references [36, 39, 125] available in the HEPData portal. The results obtained in this dissertation are highlighted in orange.

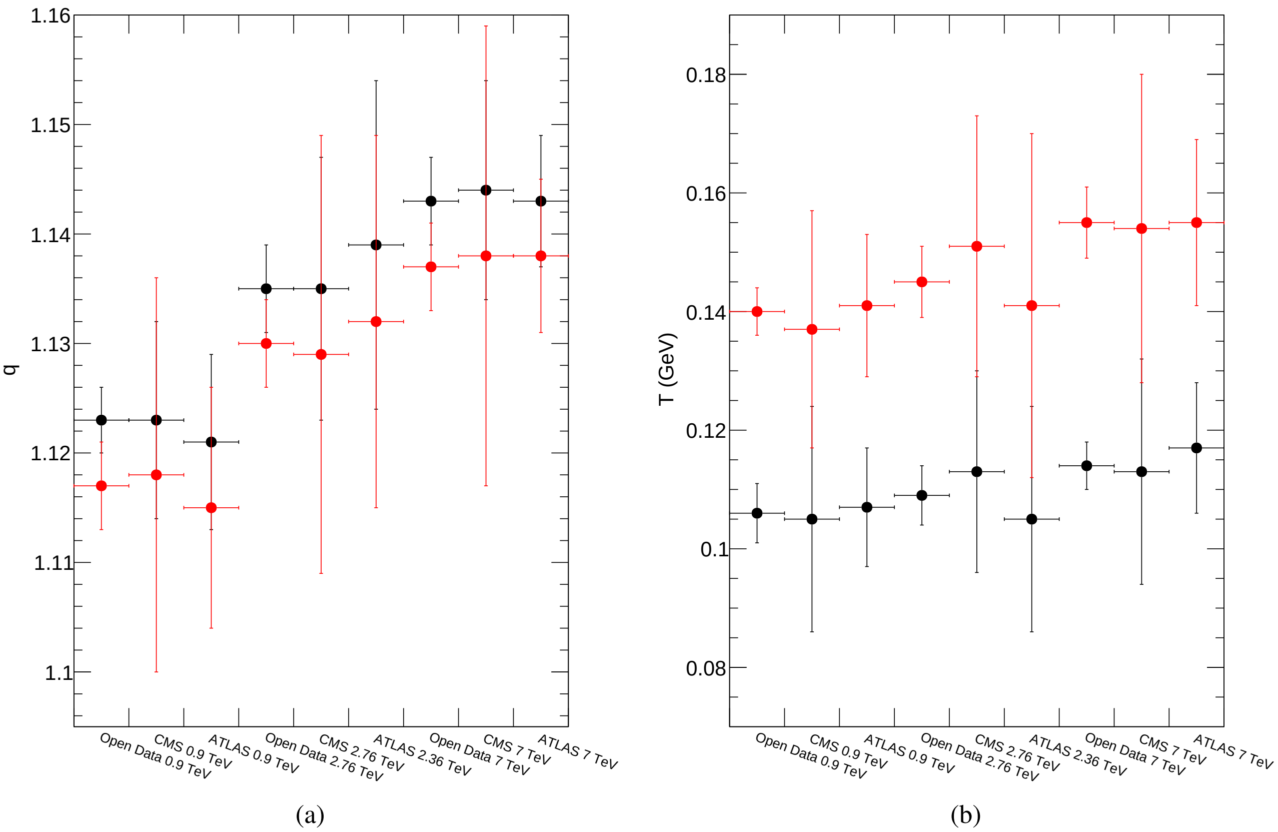


Figure 7.8: Fit parameters a) q and b) T obtained with the Tsallis fit (Equation 7.5 for the black data points and Equation 7.6 for the red data points) for the charged particle differential yield distributions. The CMS and ATLAS parameters were obtained by fitting the data from the references [36, 39, 125] available in the HEPData portal. The vertical bars represent the statistical uncertainty.

phenomenological approach using nonextensive statistical mechanics and perturbative QCD, where the parameter n relates to the number of active participants in the collision process [8], given by $n = 2 \times \left[(\text{number of active participants in the process}) - 2\right]$. Both plots in Figure 7.8 also show a difference in magnitude between the fit parameters obtained with Equations 7.5 and 7.6. This difference does not have a physical meaning since the only p_T distribution obtained from nonextensive statistical mechanics using thermodynamical relations is Equation 7.5. At this moment, Equation 7.6 can only be interpreted as a fit equation.

From Figure 7.8b, we can not affirm that the parameter T is anything other than constant when increasing the center-of-mass energy of the proton-proton collisions for the fits from both Equations 7.5 and 7.6. Studies present in the literature [145] explain that the main reason for using the thermodynamically consistent approach is because it predicts that the T parameter is a constant [146]. This result was made explicit by analyzing the distributions of various hadrons and seeing that only Equation 7.5 provides a constant T for the different systems. However, hadron identification is absent in this analysis, mainly because of the difficulty in developing identification algorithms for hadrons in the CMS detector based only on the information from the tracker. Thus, we can not confirm that only the thermodynamically consistent approach provides a constant T when analyzing just the charged particle distributions, which can be interpreted as a summation over several hadrons. The Tsallis fit provides a value of (110 ± 5) MeV for the T parameter if it is supposed to be invariant. This parameter is known as an effective temperature.

The analysis with the Tsallis fit also turns explicit the advantage of measuring distributions using open data. Usually, a researcher from outside of an experimental collaboration can only use the distributions available in the HEPData Portal. These distributions are similar to those used in this dissertation for the comparing data. The fit of these distributions results in fit parameters with large values of statistical errors, mostly due to the restricted bin arrangement provided. These large error values prevent us from drawing more precise conclusions concerning the measured parameters. However, using the open data, we can measure distributions with more suitable bin arrangements for the analysis at hand. Thus, we can get more precise measurements for the fit parameters.

7.4 Comparison to Monte Carlo models

Figure 7.9 displays a comparison between the distributions of the observables obtained in this dissertation and the same kind of distributions from Monte Carlo models for $\sqrt{s} = 7$ TeV. The ratio of the Monte Carlo charged particle transverse momentum distributions with respect to our results is shown in the lower panel. The comparison was done only for distributions with $\sqrt{s} = 7$ TeV since it is the only center-of-mass energy with simulated datasets from different event generators. The distributions in Figure 7.9 are from inelastic collisions, not containing the selection cut to obtain a sample with mostly NSD events.

The Pythia distributions, especially the Tune 4C, are the ones that best describe the LHC data. This result is expected since the Tune 4C was modeled following some input from early LHC data. The pseudorapidity (Figure 7.9c) and transverse momentum (Figure 7.9d) distributions from the Monte Carlo models fall within 20% of the distributions obtained in this dissertation for most of the spectra. The Herwig++ tuning estimates a pseudorapidity density about 20% higher than the results in this dissertation, while the Pythia Tune Z2 provides a pseudorapidity density about 20% lower. The Pythia Tune 4C provides the best description for the particles produced in the LHC collisions, containing a pseudorapidity density about 10% lower than the measured value. As for the transverse momentum distributions, the main difference between Herwig and Pythia tunings comes from the low-p_T part of the spectrum, with Herwig++ tuning overestimating the measured distribution by as much as 50%.

The advantage of the Pythia tunings comes from the description of the multiplicity distributions, especially the distribution of the charged particle average transverse momentum as a function of the multiplicity. Newer models simulate a color reconnection in the partonic interactions to describe the region with high-n_{ch}, which has significant contributions from multiple partonic interactions. Thus, Figure 7.9b makes visible the difference between the event generators as the multiplicity rises, where the ratio between the Herwig++ distribution with respect to our results is increasingly lower. The proportion between the Pythia distributions and our results is roughly constant with the multiplicity, staying around 5%. A similar analysis goes for the multiplicity distribution (Figure 7.9a) where the Pythia tunings offer a fair description within 20% of the measured distribution and the Herwig tuning overestimates the high-multiplicity region of the spectrum by one order of magnitude.

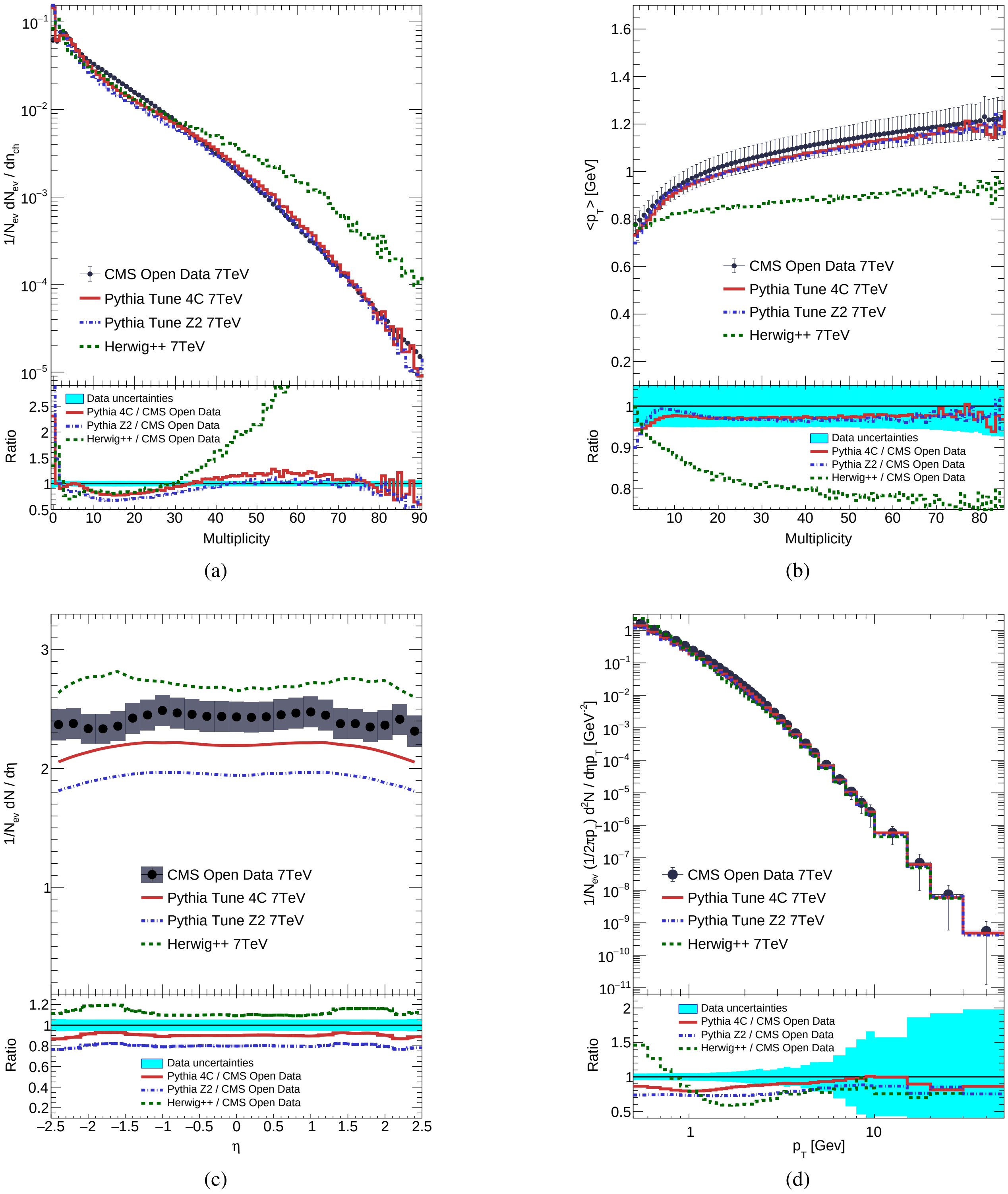


Figure 7.9: a) Charged particle multiplicity, b) charged particle $\langle p_T \rangle$, c) charged particle pseudorapidity and d) charged particle transverse momentum distributions with $n_{ch} \geq 1$, $|\eta| < 2.5$, and $p_T > 500$ MeV for $\sqrt{s} = 7$ TeV. The same distributions for different Monte Carlo models are also displayed. The ratio of the distributions from Monte Carlo datasets with respect to our results is shown in the lower panel. The vertical bars and shaded bands represent the statistical and systematic uncertainties added in quadrature.

8 | Conclusions

This dissertation presented transverse momentum, pseudorapidity, mean transverse momentum, and multiplicity distributions of charged particles produced in proton-proton collision in the LHC using open data made available by the CMS Collaboration with center-of-mass energies of 0.9, 2.76, and 7 TeV. The distributions followed the selection cuts and kinematic ranges of published results from the ATLAS and CMS experiments [34–36, 39, 40, 125].

The analysis carried out in this dissertation outlined the necessary process which we have to follow to produce results similar to the ones from experimental collaborations. This study includes the dataset selection, using triggers for the selection cut of suitable events, and applying data corrections using simulated datasets. The calculation of systematic errors is also discussed, even though it was not performed in its entirety.

The comparison between this dissertation's distributions and the CMS and ATLAS experiments results provides validity of using datasets from the CERN Open Data Portal for physical analysis. The ratios between the distributions were found to be in reasonable agreement. Specifically, the multiplicity and pseudorapidity distributions indicate that the study in this dissertation is measuring the same charged particle output from the LHC collisions as the CMS and ATLAS experiments.

The confidence that the results obtained in this dissertation agree with the data published by the experimental collaborations allowed us to carry out a physical analysis of our own. This dissertation's study analyzes the transverse momentum spectrum using a parametrization from nonextensive statistical mechanics. From this fit, we encounter a q-index that rises with center-of-mass energy, and consequently, we find the power-law index n to decrease with the center-of-mass energy of the collision. The q-index lies between 1.12 and 1.15, while the power-law index n exists between 6 and 10. As for the T parameter, known as an effective temperature, it is found to be roughly constant with the center-of-mass energy, with its value being approximately 110 MeV. The fit parameters are also found to agree with those obtained using the distributions from experimental collaborations.

One of the shortcomings of this analysis is that we can not confirm how nonextensive statistical mechanics relates to the underlying partonic process in hadronic collisions. However, it is still a remarkable result that it can provide a distribution capable of fitting the transverse momentum spectrum over several orders of magnitude using only three free parameters.

Besides the results from CMS and ATLAS presented in this dissertation, several other results from experiments surrounding the LHC have been published concerning soft interactions. For example, the ATLAS and CMS Collaborations have published the same observables present in this dissertation for $\sqrt{s} = 8$ and 13 TeV [147–150]. Since this analysis is essential for tuning Monte Carlo models, this same analysis will be made in the Run 3 of the LHC, currently in commissioning, where they aspire to reach a collision center-of-mass energy of 13.6 TeV. Besides studying Minimum Bias collisions, the analysis of the underlying event is also important for tuning Monte Carlo models and characterizing soft interactions. For example, studies are being made to improve color reconnection tunes based on underlying-event data [151].

This dissertation illustrates the challenges of carrying out these analyses of soft interactions, highlighting the importance of having the data prepared in a certain condition (low pile-up collisions and MB-like triggers) and the need to aspire for precision (a p_T threshold as low as possible is desired). Besides the methodology, this dissertation also presents an analysis that could gain from being carried out using reconstructed LHC data (instead of published distributions), either inside the collaboration or through open data. The study of the transverse momentum spectrum using nonextensive statistical mechanics needs more precise measurements than the ones available in the HEPData repositories to aspire for bolder descriptions of physical consequences.

A | Kinematic Variables

A.1 CMS coordinates

The four-momentum P of a particle with rest mass m and Lorentz factor $\gamma = 1/\sqrt{1-\beta^2}$, in natural units ($\hbar = c = 1$), is given by

$$P = (E,p_x,p_y,p_z), \tag{A.1}$$

where E is the particle's energy and $\overrightarrow{p} = (p_x,p_y,p_z)$ is the particle's momentum.

This dissertation uses the standard coordinate system for particle accelerators. The z-axis is the beam axis, the x-axis points towards the center of the accelerator, and the y-axis points up. Considering spherical coordinates for the momentum vector, θ is the polar angle (angle with respect to the z-axis), and ϕ is the azimuthal angle (angle between the x-axis and the transverse momentum vector $\overrightarrow{p_T} = (p_x,p_y)$).

However, experimental collaborations usually map a particle using the rapidity parameter y given by

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right). \tag{A.2}$$

instead of the polar angle. The advantage of using the rapidity of a particle comes from its additivity under Lorentz transformations where

$$\begin{aligned} y' &= \frac{1}{2}\ln\left(\frac{E'+p'_z}{E'-p'_z}\right) = \frac{1}{2}\ln\left(\frac{\gamma(1-\beta)(E+p_z)}{\gamma(1+\beta)(E-p_z)}\right) \\ &= y + \frac{1}{2}\ln\left(\frac{1-\beta}{1+\beta}\right) = y - \frac{1}{2}\ln\left(\frac{1+\beta}{1-\beta}\right). \end{aligned} \tag{A.3}$$

As a result, the rapidity differences are invariant under boosts along the beam direction, that is, $\Delta y = \Delta y'$. Since the experimental measurement of the rapidity is not straightforward, a new parameter called pseudorapidity is introduced, given by

$$\eta \equiv \frac{1}{2}\ln\left(\frac{p+p_z}{p-p_z}\right) = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]. \tag{A.4}$$

When the particles' masses can be neglected ($p >> m$), we have that

$$\begin{aligned} y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right) &= \frac{1}{2}\ln\left(\frac{\sqrt{p^2+m^2}+p\cos\theta}{\sqrt{p^2+m^2}-p\cos\theta}\right) \approx \frac{1}{2}\ln\left(\frac{p+p\cos\theta}{p-p\cos\theta}\right) \\ = \frac{1}{2}\ln\left(\frac{1+\cos\theta}{1-\cos\theta}\right) &= \frac{1}{2}\ln\left[\frac{\cos^2(\theta/2)}{\sin^2(\theta/2)}\right] = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \equiv \eta. \end{aligned} \tag{A.5}$$

and the pseudorapidity can be used in place of the rapidity. Figure A.1 illustrates the coordinate system used.

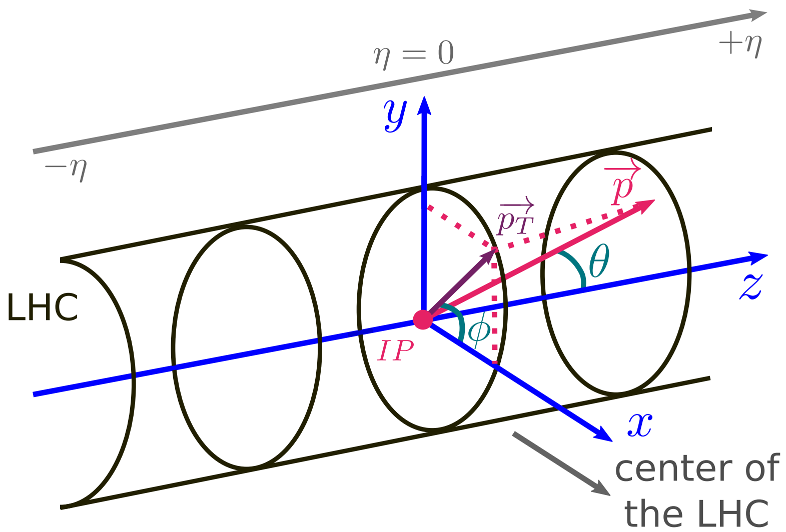


Figure A.1: The CMS coordinate system.

In this new coordinate system, we have that

$$\begin{aligned} p &= p_T\cosh\eta, \\ p_x &= p_T\cos\phi, \\ p_y &= p_T\sin\phi, \\ p_z &= p_T\sinh\eta = m_T\sinh y, \quad \text{and} \\ E &= m_T\cosh y, \end{aligned} \tag{A.6}$$

where $m_T \equiv \sqrt{m^2+p_T^2} = \sqrt{E^2-p_z^2}$.

A.2 Invariant differential particle yield

When changing from cartesian coordinates to the (p_T,ϕ,y) coordinate system, we add a jacobian determinant J, given by

$$
\begin{aligned}
J &= \begin{vmatrix} \frac{\partial p_x}{\partial p_T} & \frac{\partial p_x}{\partial \phi} & \frac{\partial p_x}{\partial y} \\ \frac{\partial p_y}{\partial p_T} & \frac{\partial p_y}{\partial \phi} & \frac{\partial p_y}{\partial y} \\ \frac{\partial p_z}{\partial p_T} & \frac{\partial p_z}{\partial \phi} & \frac{\partial p_z}{\partial y} \end{vmatrix} \\
J &= (\cos^2\phi + \sin^2\phi) p_T m_T \cosh y = p_T m_T \cosh y = p_T E.
\end{aligned}
\tag{A.7}
$$

Thus, the invariant differential particle yield becomes

$$
E\frac{d^3N}{dp^3} = E\frac{d^3N}{dp_x dp_y dp_z} = \frac{E}{J}\frac{d^3N}{d\phi dy dp_T} = \frac{1}{p_T}\frac{d^3N}{d\phi dy dp_T} = \frac{1}{2\pi p_T}\frac{d^2N}{dy dp_T}. \tag{A.8}
$$

When considering the CMS coordinate system (p_T,η,ϕ), we have that

$$
\begin{aligned}
J &= \begin{vmatrix} \frac{\partial p_x}{\partial p_T} & \frac{\partial p_x}{\partial \phi} & \frac{\partial p_x}{\partial \eta} \\ \frac{\partial p_y}{\partial p_T} & \frac{\partial p_y}{\partial \phi} & \frac{\partial p_y}{\partial \eta} \\ \frac{\partial p_z}{\partial p_T} & \frac{\partial p_z}{\partial \phi} & \frac{\partial p_z}{\partial \eta} \end{vmatrix} \\
J &= (\cos^2\phi + \sin^2\phi) p_T^2 \cosh\eta = p_T p,
\end{aligned}
\tag{A.9}
$$

and therefore the invariant differential particle yield is given by

$$
\begin{aligned}
E\frac{d^3N}{dp^3} &= E\frac{d^3N}{dp_x dp_y dp_z} = \frac{E}{J}\frac{d^3N}{d\phi d\eta dp_T} = \frac{E}{pp_T}\frac{d^3N}{d\phi d\eta dp_T} = \frac{\sqrt{p^2+m^2}}{p}\frac{1}{2\pi p_T}\frac{d^2N}{d\eta dp_T} \\
&= \sqrt{1+\frac{m^2}{p^2}}\frac{1}{2\pi p_T}\frac{d^2N}{d\eta dp_T} = \sqrt{1+\frac{m^2}{p_T^2\cosh^2\eta}}\frac{1}{2\pi p_T}\frac{d^2N}{d\eta dp_T}.
\end{aligned}
\tag{A.10}
$$

B Natively simulated datasets

The Pythia 8 dataset with $\sqrt{s} = 7$ TeV was simulated using the MinBias_7TeV_pythia8_cff.py fragment present in the CMSSW repository. For comparisons reason, the Tune::pp was set to 5 (the same tuning as the Pythia 8 datasets available in the CERN Open Data Portal). The Tune 5 corresponds to the Tune 4C [134], containing modified multiparton interaction parameters to provide a better agreement with the early LHC data. The fragment read as follows

```
process.generator = cms.EDFilter("Pythia8GeneratorFilter",
    pythiaPylistVerbosity = cms.untracked.int32(1),
    filterEfficiency = cms.untracked.double(1.0),
    pythiaHepMCVerbosity = cms.untracked.bool(False),
    comEnergy = cms.double(7000.0),
    crossSection = cms.untracked.double(71390000000.0),
    maxEventsToPrint = cms.untracked.int32(0),
    PythiaParameters = cms.PSet(
        processParameters = cms.vstring('Main:timesAllowErrors    = 10000',
            'ParticleDecays:limitTau0 = on',
            'ParticleDecays:tauMax = 10',
            'SoftQCD:minBias = on',
            'SoftQCD:singleDiffractive = on',
            'SoftQCD:doubleDiffractive = on',
            'Tune:pp 5',
            'Tune:ee 3'),
        parameterSets = cms.vstring('processParameters')
    )
)
```

The configuration file for the event generation and detector simulation of Y thousands of events was obtained with the following command

```
cmsDriver.py MinBias_7TeV_pythia8_cff.py --mc --eventcontent=RAWSIM --datatier=
   GEN-SIM --conditions=START42_V17B::All --step=GEN,SIM --python_filename=
```

```
gensim_MB_seedX_Yk.py --no_exec --number=Y000 --fileout=gensim_MB_seedX_Yk.
  root
```

Since multiple simulated files were needed, each contained a different seed number X for random number generation, implemented with the following code:

```
process.RandomNumberGeneratorService = cms.Service("RandomNumberGeneratorService
   ",
  generator = cms.PSet(
    initialSeed = cms.untracked.uint32(X),
    engineName = cms.untracked.string('TRandom3')
  ),
  VtxSmeared = cms.PSet(
    initialSeed = cms.untracked.uint32(X),
    engineName = cms.untracked.string('TRandom3')
  ),
  g4SimHits = cms.PSet(
    initialSeed = cms.untracked.uint32(X),
    engineName = cms.untracked.string('TRandom3')
  )
)
```

For the digitalization step, the configuration file was obtained with

```
cmsDriver.py step1 --filein file:gensim_MB_seedX_Yk.root --step=DIGI,L1,DIGI2RAW
    --datatier GEN-SIM-RAW --conditions=START42_V17B::All --fileout=
   digi_MB_seedX_Yk.root --eventcontent RAWSIM --python_filename
   digi_MB_seedX_Yk.py --number=Y000 --mc --no_exec
```

Finally, for the reconstruction step, the configuration file was obtained with the command

```
cmsDriver.py step2 --filein file:digi_MB_seedX_Yk.root --fileout
   reco_MB_seedX_Yk.root --mc --eventcontent AODSIM --pileup NoPileUp --
   customise Configuration/GlobalRuns/reco_TLR_42X.customisePPMC,Configuration/
   DataProcessing/Utils.addMonitoring --datatier AODSIM --conditions
   START42_V14B::All --step RAW2DIGI,L1Reco,RECO --python_filename
   reco_MB_seedX_Yk.py --no_exec -n Y000
```

As for the Herwig++ dataset with $\sqrt{s} = 7$ TeV, the same steps as the Pythia 8 dataset were followed. The only difference was the CMSSW code fragment, MinBias_7TeV_herwigpp_cff.py, which read as

```
process.generator = cms.EDFilter("ThePEGGeneratorFilter",
    herwigDefaultsBlock,
```

```
    cm7TeV = cms.vstring('set /Herwig/Generators/LHCGenerator:EventHandler:
   LuminosityFunction:Energy 7000.0',
        'set /Herwig/Shower/Evolver:IntrinsicPtGaussian 2.0*GeV'),
    configFiles = cms.vstring(),
    crossSection = cms.untracked.double(101900000000.0),
    parameterSets = cms.vstring('cm7TeV',
        'pdfMRST2001',
        'Summer09QCDParameters',
        'basicSetup',
        'setParticlesStableForDetector'),
    filterEfficiency = cms.untracked.double(1.0),
    Summer09QCDParameters = cms.vstring('cd /Herwig/MatrixElements/',
        'insert SimpleQCD:MatrixElements[0] MEMinBias',
        'cd /',
        'cd /Herwig/Cuts',
        'set JetKtCut:MinKT 0.0*GeV',
        'set QCDCuts:MHatMin 0.0*GeV',
        'set QCDCuts:X1Min 0.01',
        'set QCDCuts:X2Min 0.01',
        'set /Herwig/UnderlyingEvent/MPIHandler:IdenticalToUE 0')
)
```

Bibliography

[1] S. L. Glashow, Partial-symmetries of weak interactions. *Nuclear Physics* **22**(4) (1961), 579, Available: https://doi.org/10.1016/0029-5582(61)90469-2.

[2] S. Weinberg, A model of leptons. *Physical Review Letters* **19**(21) (1967), 1264, Available: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[3] A. Salam, Elementary particle theory. In *Prog. Of the Nobel Symposium, 1968, Stockholm, Sweden*, volume 367.

[4] M. Thomson, *Modern Particle Physics*. Cambridge University Press, Cambridge (2013).

[5] M. D. Schwartz, *Quantum Field Theory and the Standard Model*. Cambridge University Press, Cambridge (2014).

[6] F. Mandl and G. Shaw, *Quantum Field Theory*. Wiley, Chichester (2010).

[7] L. Evans and P. Bryant, LHC Machine. *Journal of Instrumentation* **3**(8) (2008), S08001, Available: https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08001.

[8] P. D. Collins and A. D. Martin, Hadron reaction mechanisms. *Reports on Progress in Physics* **45**(4) (1982), 335, Available: https://iopscience.iop.org/article/10.1088/0034-4885/45/4/001.

[9] F. E. Close, *An Introduction to Quarks and Partons*. Academic Press, New York (1979).

[10] D. Sivers, S. J. Brodsky, and R. Blankenbecler, Large transverse momentum processes. *Physics Reports* **23**(1) (1976), 1, Available: https://doi.org/10.1016/0370-1573(76)90015-6.

[11] C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics* **52**(1-2) (1988), 479, Available: https://link.springer.com/article/10.1007/BF01016429.

[12] C. Tsallis, Nonadditive entropy: The concept and its use. *The European Physical Journal A*

40(3) (2009), 257, Available: https://doi.org/10.1140/epja/i2009-10799-0.

[13] C. Tsallis, *Introduction to nonextensive statistical mechanics: Approaching a complex world*. Springer, New York (2009).

[14] J. Cleymans and D. Worku, The Tsallis distribution in protonproton collisions at $\sqrt{s} = 0.9$ TeV at the LHC. *Journal of Physics G: Nuclear and Particle Physics* **39**(2) (2012), 025006, Available: http://arxiv.org/abs/1110.5526.

[15] C. Y. Wong, G. Wilk, L. J. Cirto, and C. Tsallis, From QCD-based hard-scattering to nonextensive statistical mechanical descriptions of transverse momentum spectra in high-energy pp and $p\bar{p}$ collisions. *Physical Review D - Particles, Fields, Gravitation and Cosmology* **91**(11) (2015), 114027, Available: https://journals.aps.org/prd/abstract/10.1103/PhysRevD.91.114027.

[16] T. Sjöstrand, S. Mrenna, and P. Skands, PYTHIA 6.4 physics and manual. *Journal of High Energy Physics* **2006**(5) (2006), 026, Available: https://iopscience.iop.org/article/10.1088/1126-6708/2006/05/026.

[17] T. Sjöstrand, S. Mrenna, and P. Skands, A brief introduction to PYTHIA 8.1. *Computer Physics Communications* **178**(11) (2008), 852, Available: https://doi.org/10.1016/j.cpc.2008.01.036.

[18] T. Sjöstrand and M. Van Zijl, A multiple-interaction model for the event structure in hadron collisions. *Physical Review D* **36**(7) (1987), 2019, Available: https://journals.aps.org/prd/abstract/10.1103/PhysRevD.36.2019.

[19] R. Engel, Photoproduction within the two-component Dual Parton Model: Amplitudes and cross sections. *Zeitschrift für Physik C Particles and Fields* **66**(1-2) (1995), 203, Available: https://link.springer.com/article/10.1007/BF01496594.

[20] R. Engel and J. Ranft, Hadronic photon-photon interactions at high energies. *Physical Review D - Particles, Fields, Gravitation and Cosmology* **54**(7) (1996), 4244, Available: https://journals.aps.org/prd/abstract/10.1103/PhysRevD.54.4244.

[21] R. Engel, PHOJET manual (program version 1.05 c, June 96). *University of Siegen preprint* (1995), 5.

[22] A. Capella and J. T. Thanh Van, A new parton model description of soft hadron-nucleus collisions. *Physics Letters B* **93**(1-2) (1980), 146, Available: https://doi.org/10.1016/0370-2693(80)90114-8.

[23] A. Capella and J. Tran Than Van, Hadron-nucleus interactions and the leading particle effect in a dual-parton model. *Zeitschrift für Physik C Particles and Fields* **10**(3) (1981), 249, Available: https://link.springer.com/article/10.1007/BF01549733.

[24] A. Capella, U. Sukhatme, C. I. Tan, and J. Tran Thanh Van, Dual parton model. *Physics Reports* **236**(4-5) (1994), 225, Available: https://doi.org/10.1016/0370-1573(94)90064-7.

[25] E. Fermi, High Energy Nuclear Events. *Progress of Theoretical Physics* **5**(4) (1950), 570, Available: https://doi.org/10.1143/ptp/5.4.570.

[26] R. Hagedorn, Statistical thermodynamics of strong interactions at high energies. *Nuovo Cimento* **3** (1965), 147, Available: https://cds.cern.ch/record/346206/.

[27] N. Cabibbo and G. Parisi, Exponential hadronic spectrum and quark liberation. *Physics Letters B* **59**(1) (1975), 67, Available: https://www.sciencedirect.com/science/article/pii/0370269375901586.

[28] R. Hagedorn and J. Rafelski, Hot hadronic matter and nuclear collisions. *Physics Letters B* **97**(1) (1980), 136, Available: https://www.sciencedirect.com/science/article/pii/0370269380905663.

[29] CBPF Group of Statistical Physics - Bibliography. Available: http://tsallis.cat.cbpf.br/biblio.htm. Accessed on: 2022-04-28

[30] ALICE Collaboration, The ALICE experiment at the CERN LHC. *Journal of Instrumentation* **3**(08) (2008), S08002, Available: https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08002.

[31] CMS Collaboration, The CMS experiment at the CERN LHC. *Journal of Instrumentation* **3**(8) (2008), S08004, Available: iopscience.iop.org/article/10.1088/1748-0221/3/08/S08004/.

[32] ALICE Collaboration, Production of pions, kaons and protons in pp collisions $\sqrt{s} = 900$ GeV with ALICE at the LHC. *European Physical Journal C* **71**(6) (2011), 1, Available: https://arxiv.org/abs/1101.4110.

[33] ALICE Collaboration, Neutral pion and η meson production in proton-proton collisions at $\sqrt{s} = 0.9$ TeV and $\sqrt{s} = 7$ TeV. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics* **717**(1-3) (2012), 162, Available: https://doi.org/10.1016/j.physletb.2012.

09.015.

[34] CMS Collaboration, Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s} = 0.9$ and 2.36 TeV. *Journal of High Energy Physics* **2010**(2) (2010), 041, Available: http://dx.doi.org/10.1007/JHEP02(2010)041.

[35] CMS Collaboration, Transverse-Momentum and Pseudorapidity Distributions of Charged Hadrons in pp Collisions at $\sqrt{s} = 7$ TeV. *Physical Review Letters* **105**(2) (2010), 022002, Available: https://link.aps.org/doi/10.1103/PhysRevLett.105.022002.

[36] CMS Collaboration, Charged particle transverse momentum spectra in pp collisions at $\sqrt{s} =$ 0.9 and 7 TeV. *Journal of High Energy Physics* **2011**(8) (2011), 086, Available: http://dx.doi.org/10.1007/JHEP08(2011)086.

[37] CMS Collaboration, Study of the inclusive production of charged pions, kaons, and protons in pp collisions at $\sqrt{s} = 0.9$, 2.76 and 7 TeV. *European Physical Journal C* **72**(10) (2012), 1, Available: https://arxiv.org/abs/1207.4724.

[38] ALICE Collaboration, Charged-particle multiplicity measurement in proton-proton collisions at $\sqrt{s} = 7$ TeV with ALICE at LHC. *European Physical Journal C* **68**(3) (2010), 345, Available: http://arxiv.org/abs/1004.3514.

[39] ATLAS Collaboration, Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC. *New Journal of Physics* **13**(5) (2011), 053033, Available: https://doi.org/10.1088/1367-2630/13/5/053033.

[40] CMS Collaboration, Charged particle multiplicities in pp interactions at $\sqrt{s} = 0.9$, 2.36, and 7 TeV. *Journal of High Energy Physics* **2011**(1) (2011), 79, Available: https://link.springer.com/article/10.1007/JHEP01(2011)079.

[41] CERN Open Data Portal. Available: http://opendata.cern.ch/. Accessed on: 2022-05-02

[42] CMS Offline Software. Available: https://github.com/cms-sw/cmssw. Accessed on: 2022-04-29

[43] E. Rutherford, LXXIX. The scattering of α and β particles by matter and the structure of the atom. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **21**(125) (1911), 669, Available: https://www.tandfonline.com/doi/full/10.1080/

14786440508637080.

[44] J. Chadwick, Possible Existence of a Neutron. *Nature* **129**(3252) (1932), 312, Available: https://doi.org/10.1038/129312a0.

[45] H. Yukawa, On the Interaction of Elementary Particles. I. *Proceedings of the Physico-Mathematical Society of Japan. 3rd Series* **17** (1935), 48, Available: https://doi.org/10.11429/ppmsj1919.17.0_48.

[46] M. Gell-Mann, A schematic model of baryons and mesons. *Physics Letters* **8**(3) (1964), 214, Available: http://www.sciencedirect.com/science/article/pii/S0031916364920013.

[47] F. Caruso, M. Anselmino, J. Mahon, and V. Oguri, *Introdução à QCD Perturbativa*. LTC, Rio de Janeiro (2013).

[48] D. J. Gross and F. Wilczek, Ultraviolet behavior of non-abelian gauge theories. *Physical Review Letters* **30**(26) (1973), 1343, Available: https://link.aps.org/doi/10.1103/PhysRevLett.30.1343.

[49] H. D. Politzer, Reliable perturbative results for strong interactions? *Physical Review Letters* **30**(26) (1973), 1346, Available: https://link.aps.org/doi/10.1103/PhysRevLett.30.1346.

[50] J. R. Forshaw and D. A. Ross, *Quantum Chromodynamics and the Pomeron*. Cambridge University Press, Cambridge (1997).

[51] ALICE Collaboration, Measurement of inelastic, single- and double-diffraction cross sections in proton-proton collisions at the LHC with ALICE. *European Physical Journal C* **73**(6) (2013), 1, Available: http://dx.doi.org/10.1140/epjc/s10052-013-2456-0.

[52] ATLAS Collaboration, Measurement of the total cross section from elastic scattering in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Nuclear Physics B* **889** (2014), 486, Available: http://dx.doi.org/10.1016/j.nuclphysb.2014.10.019.

[53] CMS Collaboration, Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 7$ TeV. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics* **722**(1) (2013), 5, Available: https://doi.org/10.1016/j.physletb.2013.03.024.

[54] CMS Collaboration, Measurement of diffractive dissociation cross sections in pp collisions at

$\sqrt{s} = 7$ TeV. *Physical Review D - Particles, Fields, Gravitation and Cosmology* **92**(1) (2015), 012003, Available: http://dx.doi.org/10.1103/PhysRevD.92.012003.

[55] C.-Y. Wong, *Introduction to High-Energy Heavy-Ion Collisions*. World Scientific, Singapore (1994), Available: https://cds.cern.ch/record/241251.

[56] R. D. Field and R. P. Feynman, Quark elastic scattering as a source of high-transverse-momentum mesons. *Physical Review D* **15**(9) (1977), 2590, Available: https://journals.aps.org/prd/abstract/10.1103/PhysRevD.15.2590.

[57] S. J. Brodsky and G. R. Farrar, Scaling Laws at Large Transverse Momentum. *Physical Review Letters* **31**(18) (1973), 1153, Available: https://link.aps.org/doi/10.1103/PhysRevLett.31.1153.

[58] S. J. Brodsky and G. R. Farrar, Scaling laws for large-momentum-transfer processes. *Physical Review D* **11**(5) (1975), 1309, Available: https://link.aps.org/doi/10.1103/PhysRevD.11.1309.

[59] V. A. Matveev, R. M. Muradyan, and A. N. Tavkhelidze, Automodellism in the large-angle elastic scattering and structure of hadrons. *Lettere al Nuovo Cimento (1971-1985)* **7**(15) (1973), 719, Available: https://doi.org/10.1007/BF02728133.

[60] B. L. Combridge, J. Kripfganz, and J. Ranft, Hadron production at large transverse momentum and QCD. *Physics Letters B* **70**(2) (1977), 234, Available: https://www.sciencedirect.com/science/article/pii/0370269377905287.

[61] C. Y. Wong and G. Wilk, Tsallis fits to p_T spectra for pp collisions at the LHC. *Acta Physica Polonica B* **43**(11) (2012), 2047, Available: http://arxiv.org/abs/1210.3661.

[62] C. Y. Wong and G. Wilk, Tsallis fits to p_T spectra and multiple hard scattering in pp collisions at the LHC. *Physical Review D - Particles, Fields, Gravitation and Cosmology* **87**(11) (2013), 114007, Available: http://arxiv.org/abs/1305.2627.

[63] R. Balian, *From Microphysics to Macrophysics: Methods and Applications of Statistical Physics*. Springer-Verlag, Berlin (1991).

[64] M. Gell-Mann and C. Tsallis, *Nonextensive Entropy: Interdisciplinary Applications*. Oxford University Press, New York (2004).

[65] R. Clausius, *The Mechanical Theory of Heat with its Applications to the Steam Engine and to Physical Properties of Bodies*. J. Van Voorst, London (1867).

[66] J. C. Maxwell, V. Illustrations of the dynamical theory of gases. Part I. On the motions and collisions of perfectly elastic spheres. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **19**(124) (1860), 19, Available: https://www.tandfonline.com/doi/abs/10.1080/14786446008642818.

[67] J. C. Maxwell, II. Illustrations of the dynamical theory of gases. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **20**(130) (1860), 21, Available: https://www.tandfonline.com/doi/abs/10.1080/14786446008642902.

[68] L. Boltzmann, *Vorlesungen über Gastheorie*. J.A. Barth, Leipzig (1896). *Lectures on Gas Theory*, transl. S. Brush (Univ. California Press, Berkeley, 1964).

[69] J. W. Gibbs, *Elementary principles in statistical mechanics*. Charles Scribner's Sons, New York (1902).

[70] S. R. A. Salinas, *Introdução à Física Estatística*. EDUSP, São Paulo (1997).

[71] K. Huang, *Statistical Mechanics*. Wiley (1987).

[72] H. Koppe, Die Mesonenausbeute beim Beschuß von leichten Kernen mit x-Teilchen:. *Zeitschrift für Naturforschung A* **3**(4) (1948), 251, Available: https://doi.org/10.1515/zna-1948-0411.

[73] R. Hagedorn, Thermodynamics of strong interactions. In *CERN Lectures*, CERN, Geneva, Available: https://cds.cern.ch/record/318595.

[74] S. Frautschi, Statistical bootstrap model of hadrons. *Physical Review D* **3**(11) (1971), 2821, Available: https://journals.aps.org/prd/abstract/10.1103/PhysRevD.3.2821.

[75] W. Nahm, Analytical solution of the statistical bootstrap model. *Nuclear Physics, Section B* **45**(2) (1972), 525, Available: https://doi.org/10.1016/0550-3213(72)90257-X.

[76] R. Hagedorn, On the hadronic mass spectrum. *Il Nuovo Cimento A Series 10* **52**(4) (1967), 1336, Available: https://link.springer.com/article/10.1007/BF02755235.

[77] R. Hagedorn, Hadronic Matter Near the Boiling Point. *Il Nuovo Cimento A* **56**(4) (1968), 1027,

Available: https://link.springer.com/article/10.1007/BF02751614.

[78] A. H. Rosenfeld, A. Barbaro-Galtieri, W. J. Podolsky, L. R. Price, P. Soding, *et al.*, Data on particles and resonant states. *Reviews of Modern Physics* **39**(1) (1967), 1, Available: https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.39.1.

[79] UA1 Collaboration, Transverse momentum spectra for charged particles at the CERN proton-antiproton collider. *Physics Letters B* **118**(1-3) (1982), 167, Available: https://doi.org/10.1016/0370-2693(82)90623-2.

[80] R. Hagedorn, Multiplicities, p_T distributions and the expected hadron $\rightarrow$ quark-gluon phase transition. *La Rivista del Nuovo Cimento (1978-1999)* **6**(10) (1983), 1, Available: https://doi.org/10.1007/BF02740917.

[81] F. Büyükkiliç and D. Demirham, A fractal approach to entropy and distribution functions. *Physics Letters A* **181**(1) (1993), 24, Available: https://doi.org/10.1016/0375-9601(93)91118-O.

[82] F. Büyükkiliç, D. Demirham, and A. Güleç, A statistical mechanical approach to generalized statistics of quantum and classical gases. *Physics Letters A* **197**(3) (1995), 209, Available: https://doi.org/10.1016/0375-9601(94)00941-H.

[83] F. Pennini, A. Plastino, and A. R. Plastino, Tsallis entropy and quantal distribution functions. *Physics Letters A* **208**(4-6) (1995), 309, Available: https://doi.org/10.1016/0375-9601(95)00720-1.

[84] A. R. Plastino, A. Plastino, H. G. Miller, and H. Uys, Foundations of nonextensive statistical mechanics and its cosmological applications. *Astrophysics and Space Science* **290**(3-4) (2004), 275, Available: https://link.springer.com/article/10.1023/B:ASTR.0000032529.67037.21.

[85] A. M. Teweldeberhan, A. R. Plastino, and H. G. Miller, On the cut-off prescriptions associated with power-law generalized thermostatistics. *Physics Letters, Section A: General, Atomic and Solid State Physics* **343**(1-3) (2005), 71, Available: https://doi.org/10.1016/j.physleta.2005.06.026.

[86] C. Tsallis, R. S. Mendes, and A. R. Plastino, The role of constraints within generalized nonextensive statistics. *Physica A: Statistical Mechanics and its Applications* **261**(3-4) (1998), 534, Available: https://doi.org/10.1016/S0378-4371(98)00437-3.

[87] UA1 Collaboration, Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV. *Physics Letters B* **122**(1) (1983), 103, Available: https://doi.org/10.1016/0370-2693(83)91177-2.

[88] UA1 Collaboration, Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider. *Physics Letters B* **126**(5) (1983), 398, Available: https://doi.org/10.1016/0370-2693(83)90188-0.

[89] UA2 Collaboration, Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider. *Physics Letters B* **122**(5-6) (1983), 476, Available: https://doi.org/10.1016/0370-2693(83)91605-2.

[90] UA2 Collaboration, Evidence for $Z^0 \rightarrow e^+e^-$ at the CERN pp collider. *Physics Letters B* **129**(1-2) (1983), 130, Available: https://doi.org/10.1016/0370-2693(83)90744-X.

[91] ATLAS Collaboration, The ATLAS experiment at the CERN large hadron collider. *Journal of Instrumentation* **3**(8) (2008), S08003, Available: iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003/.

[92] LHCb Collaboration, The LHCb detector at the LHC. *Journal of Instrumentation* **3**(8) (2008), S08005, Available: iopscience.iop.org/article/10.1088/1748-0221/3/08/S08005/.

[93] CMS Collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics* **716**(1) (2012), 30, Available: https://doi.org/10.1016/j.physletb.2012.08.021.

[94] CMS Collaboration, CMS, the Compact Muon Solenoid: Technical proposal. Technical report (1994), Available: https://cds.cern.ch/record/290969.

[95] CMS Collaboration, CMS, the magnet project: Technical design report. Technical report (1997), Available: http://cds.cern.ch/record/331056.

[96] CMS Collaboration, The CMS tracker system project: Technical Design Report Available: https://cds.cern.ch/record/368412.

[97] CMS Collaboration, The CMS electromagnetic calorimeter project: Technical Design Report Available: https://cds.cern.ch/record/349375.

[98] CMS Collaboration, The CMS hadron calorimeter project: Technical Design Report Available: https://cds.cern.ch/record/357153.

[99] CMS Collaboration, The CMS muon project: Technical Design Report Available: https://cds.cern.ch/record/343814.

[100] CMS Collaboration, Commissioning and performance of the CMS pixel tracker with cosmic ray muons. *Journal of Instrumentation* **5**(3) (2010), T03007, Available: https://iopscience.iop.org/article/10.1088/1748-0221/5/03/T03007.

[101] CMS Collaboration, Performance and operation of the CMS electromagnetic calorimeter. *Journal of Instrumentation* **5**(3) (2010), T03010, Available: https://doi.org/10.1088/1748-0221/5/03/T03010.

[102] CMS Collaboration, Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} =$ 7 TeV. *Journal of Instrumentation* **7**(10) (2012), P10002, Available: http://dx.doi.org/10.1088/1748-0221/7/10/P10002.

[103] CMS Collaboration, The CMS trigger system. *Journal of Instrumentation* **12**(1) (2016), P01020, Available: https://doi.org/10.1088/1748-0221/12/01/P01020.

[104] A. J. Bell, Design and Construction of the Beam Scintillation Counter for CMS. Master's thesis, University of Canterbury, New Zealand (2008), Available: https://cds.cern.ch/record/1204301.

[105] T. Aumeyr, Beam Phase and Intensity Monitoring for the Compact Muon Solenoid Experiment. Master's thesis, Vienna University of Technology, Austria (2008), Available: https://resolver.obvsg.at/urn:nbn:at:at-ubtuw:1-29643.

[106] OPAL Collaboration, The OPAL detector at LEP. *Nuclear Inst. and Methods in Physics Research, A* **305**(2) (1991), 275, Available: https://doi.org/10.1016/0168-9002(91)90547-4.

[107] CMS Collaboration, DoubleMu primary dataset in AOD format from RunA of 2011 (/DoubleMu/Run2011A-12Oct2013-v1/AOD). *CERN Open Data Portal* Available: http://doi.org/10.7483/OPENDATA.CMS.RZ34.QR6N.

[108] CMS Collaboration, Jet primary dataset in AOD format from RunA of 2011 (/Jet/Run2011A-12Oct2013-v1/AOD). *CERN Open Data Portal* Available: http://doi.org/10.7483/OPENDATA.CMS.UP77.P6PQ.

[109] R. Brun and F. Rademakers, ROOT - An object oriented data analysis framework. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **389**(1-2) (1997), 81, Available: http://dx.doi.org/10.1016/S0168-9002(97)00048-X.

[110] CERN Open Data Portal - CMS Guide on how to calculate luminosity. Available: http://opendata.cern.ch/docs/cms-guide-luminosity-calculation. Accessed on: 2021-07-14

[111] M. Bähr, S. Gieseke, M. A. Gigg, D. Grellscheid, K. Hamilton, *et al.*, Herwig++ physics and manual. *European Physical Journal C* **58**(4) (2008), 639, Available: https://link.springer.com/article/10.1140/epjc/s10052-008-0798-9.

[112] GEANT4 Collaboration, GEANT4 - A simulation toolkit. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506**(3) (2003), 250, Available: https://doi.org/10.1016/S0168-9002(03)01368-8.

[113] UA5 Collaboration, UA5: A general study of proton-antiproton physics at $\sqrt{s} = 546$ GeV. *Physics Reports* **154**(5-6) (1987), 247, Available: https://doi.org/10.1016/0370-1573(87)90130-X.

[114] CDF Collaboration, Pseudorapidity distributions of charged particles produced in $\overline{p}p$ interactions as $\sqrt{s} = 630$ and 1800 GeV. *Physical Review D* **41**(7) (1990), 2330, Available: https://journals.aps.org/prd/abstract/10.1103/PhysRevD.41.2330.

[115] J. R. Christiansen and P. Z. Skands, String formation beyond leading colour. *Journal of High Energy Physics* **2015**(8) (2015), 3, Available: http://dx.doi.org/10.1007/JHEP08(2015)003.

[116] CMS Collaboration, MinimumBias primary dataset in RECO format from the 7 TeV Commissioning run of 2010 (/MinimumBias/Commissioning10- May19ReReco-v1/RECO). *CERN Open Data Portal* Available: http://doi.org/10.7483/OPENDATA.CMS.HK0O.ZF9Q.

[117] CMS Collaboration, MinimumBias primary dataset in AOD format from RunA of 2010 (/MinimumBias/Run2010A-Apr21ReReco-v1/AOD). *CERN Open Data Portal* Available: http://doi.org/10.7483/OPENDATA.CMS.6B3H.TR6Z.

[118] CMS Collaboration, MinimumBias primary dataset in AOD format from RunB of 2010 (/MinimumBias/Run2010B-Apr21ReReco-v1/AOD). *CERN Open Data Portal* Available: http://doi.org/10.7483/OPENDATA.CMS.6BPY.XFRQ.

[119] CMS Collaboration, MinimumBias primary dataset in AOD format from RunA of 2011 (/MinimumBias/Run2011A-12Oct2013-v1/AOD). *CERN Open Data Portal* Available: http://doi.org/10.7483/OPENDATA.CMS.QSPG.CCQ4.

[120] CMS Collaboration, MinimumBias primary dataset in AOD format from RunB of 2011 (/MinimumBias/Run2011B-12Oct2013-v1/AOD). *CERN Open Data Portal* Available: http://doi.org/10.7483/OPENDATA.CMS.4W3B.GO96.

[121] CMS Collaboration, MinimumBias primary dataset in RECO format from the 0.9 TeV Commissioning run of 2010 (/MinimumBias/Commissioning10-07JunReReco-900GeV/RECO). *CERN Open Data Portal* Available: http://doi.org/10.7483/OPENDATA.CMS.1R58.OMBD.

[122] CMS Collaboration, MinimumBias primary dataset in AOD format from Run of 2012 (/MinimumBias/Run2012B-22Jan2013-v1/AOD). *CERN Open Data Portal* Available: http://doi.org/10.7483/OPENDATA.CMS.HU6U.DRLD.

[123] CMS Collaboration, MinimumBias primary dataset in AOD format from Run of 2012 (/MinimumBias/Run2012C-22Jan2013-v1/AOD). *CERN Open Data Portal* Available: http://doi.org/10.7483/OPENDATA.CMS.2KF0.OSQZ.

[124] CMS Collaboration, AllPhysics2760 primary dataset in RECO format from the 2.76 TeV proton-proton run of 2011 (/AllPhysics2760/Run2011A-16Jul2011-v1/RECO). *CERN Open Data Portal* Available: http://doi.org/10.7483/OPENDATA.V439.NXZC.

[125] CMS Collaboration, Study of high-p_T charged particle suppression in PbPb compared to pp collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *European Physical Journal C* **72**(3) (2012), 1, Available: https://link.springer.com/article/10.1140/epjc/s10052-012-1945-x.

[126] CMS Collaboration, Simulated dataset MinBias-TuneZ2-900GeV-pythia6-cff-py-GEN-SIM-START311-V2-Dec11-v2 in GEN-SIM-RECO format for 2010 commissioning data. *CERN Open Data Portal* Available: http://doi.org/10.7483/OPENDATA.CMS.JPB5.X7CN.

[127] CMS Collaboration, Simulated dataset MinBias-Tune4C-900GeV-pythia8-cff-py-GEN-SIM-START311-V2-Dec11-v2 in GEN-SIM-RECO format for 2010 commissioning data. *CERN Open Data Portal* Available: http://doi.org/10.7483/OPENDATA.CMS.UX3S.WLOG.

[128] CMS Collaboration, Simulated dataset MinBias-TuneZ2-7TeV-pythia6-cff-py-GEN-SIM-

START311-V2-Dec11-v2 in GEN-SIM-RECO format for 2010 commissioning data. *CERN Open Data Portal* Available: http://doi.org/10.7483/OPENDATA.CMS.K009.X3LA.

[129] CMS Collaboration, Simulated dataset MinBias-Tune4C-7TeV-pythia8-cff-py-GEN-SIM-START311-V2-Dec11-v1 in GEN-SIM-RECO format for 2010 commissioning data. *CERN Open Data Portal* Available: http://doi.org/10.7483/OPENDATA.CMS.642M.G24P.

[130] CMS Collaboration, Simulated dataset MinBias-TuneZ2-2760GeV-pythia6-cff-py-GEN-SIM-START311-V2-Dec11-v2 in GEN-SIM-RECO for- mat for Run2011A. *CERN Open Data Portal* Available: http://doi.org/10.7483/OPENDATA.CMS.6UPI.ZXM3.

[131] CMS Collaboration, Simulated dataset MinBias-Tune4C-2760GeV-pythia8-cff-py-GEN-SIM-START311-V2-Dec11-v1 in GEN-SIM-RECO format for Run2011A. *CERN Open Data Portal* Available: http://doi.org/10.7483/OPENDATA.CMS.P7WD.NL5I.

[132] R. Field, Early LHC Underlying Event Data-Findings and Surprises. *arXiv preprint.* Available: https://arxiv.org/abs/1010.3558v1.

[133] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. Nadolsky, *et al.*, New generation of parton distributions with uncertainties from global QCD analysis. *Journal of High Energy Physics* **6**(7) (2002), 325, Available: http://dx.doi.org/10.1088/1126-6708/2002/07/012.

[134] R. Corke and T. Sjöstrand, Interleaved parton showers and tuning prospects. *Journal of High Energy Physics* **2011**(3) (2011), 32, Available: https://link.springer.com/article/10.1007/JHEP03(2011)032.

[135] R. Rougny, Commissioning of the CMS pixel tracker and measurement of the charged particle muliplicity at the LHC. Ph.D. thesis, University of Antwerp, Belgium (2015), Available: http://cds.cern.ch/record/2023893.

[136] CMS Collaboration, High-Level Trigger information for 2010 CMS open data. Available: http://opendata.cern.ch/record/1699.

[137] cmssw/BSCTrigger.cc at `CMSSW_4_2_X`. Available: https://github.com/cms-sw/cmssw/blob/CMSSW_4_2_X/L1TriggerOffline/L1Analyzer/src/BSCTrigger.cc. Accessed on: 2022-04-27

[138] A. V. Fedotov, Beam halo formation in high-intensity beams. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **557**(1) (2006), 216, Available: https://doi.org/10.1016/j.nima.2005.10.073.

[139] ATLAS Collaboration, Charged-particle multiplicities in pp interactions at $\sqrt{s} = 900$ GeV measured with the ATLAS detector at the LHC. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics* **688**(1) (2010), 21, Available: http://dx.doi.org/10.1016/j.physletb.2010.03.064.

[140] G. D'Agostini, A multidimensional unfolding method based on Bayes' theorem. *Nuclear Inst. and Methods in Physics Research, A* **362**(2-3) (1995), 487, Available: https://doi.org/10.1016/0168-9002(95)00274-X.

[141] T. Adye, Unfolding algorithms and tests using RooUnfold. In *Proceedings of the PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*, 313–318, Available: https://doi.org/10.5170/CERN-2011-006.313.

[142] CMS Collaboration, Search for microscopic black hole signatures at the Large Hadron Collider. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics* **697**(5) (2011), 434, Available: https://www.sciencedirect.com/science/article/pii/S0370269311001778?via%3Dihub.

[143] L. J. L. Cirto, C. Tsallis, C.-Y. Wong, and G. Wilk, The transverse-momenta distributions in high-energy pp collisions – A statistical-mechanical approach. *arXiv preprint.* Available: http://arxiv.org/abs/1409.3278.

[144] A. S. Parvan, O. V. Teryaev, and J. Cleymans, Systematic comparison of Tsallis statistics for charged pions produced in pp collisions. *European Physical Journal A* **53**(5) (2017), 1, Available: https://link.springer.com/article/10.1140/epja/i2017-12301-y.

[145] L. Marques, E. Andrade-Ii, and A. Deppman, Nonextensivity of hadronic systems. *Physical Review D - Particles, Fields, Gravitation and Cosmology* **87**(11) (2013), 114022, Available: https://journals.aps.org/prd/abstract/10.1103/PhysRevD.87.114022.

[146] A. Deppman, Self-consistency in non-extensive thermodynamics of highly excited hadronic states. *Physica A: Statistical Mechanics and its Applications* **391**(24) (2012), 6380, Available: http://dx.doi.org/10.1016/j.physa.2012.07.071.

[147] ATLAS Collaboration, Charged-particle distributions in pp interactions at $\sqrt{s} = 8$ TeV measured with the ATLAS detector. *European Physical Journal C* **76**(7) (2016), 403, Available: http://dx.doi.org/10.1140/epjc/s10052-016-4203-9.

[148] ATLAS Collaboration, Charged-particle distributions at low transverse momentum in $\sqrt{s}$ =13 TeV pp interactions measured with the ATLAS detector at the LHC. *European Physical Journal C* **76**(9) (2016), 1, Available: https://link.springer.com/article/10.1140/epjc/s10052-016-4335-y.

[149] ATLAS Collaboration, Charged-particle distributions in $\sqrt{s}$ = 13 TeV pp interactions measured with the ATLAS detector at the LHC. *Physics Letters B* **758** (2016), 67, Available: https://doi.org/10.1016/J.PHYSLETB.2016.04.050.

[150] CMS Collaboration, Pseudorapidity distribution of charged hadrons in proton-proton collisions at $\sqrt{s}$ = 13 TeV. *Physics Letters B* **751** (2015), 143, Available: https://doi.org/10.1016/J.PHYSLETB.2015.10.004.

[151] CMS Collaboration, CMS PYTHIA 8 colour reconnection tunes based on underlying-event data. *arXiv preprint.* Available: https://arxiv.org/abs/2205.02905v1.