ULB FACULTÉ DES SCIENCES
UNIVERSITÉ LIBRE DE BRUXELLES

VUB VRIJE UNIVERSITEIT BRUSSEL

Search for new physics in the dark sector with the CMS detector

From invisible to low charge particles

Thesis presented by David VANNEROM
with a view to obtaining the Ph.D. Degree in Sciences (ULB - "Docteur en Sciences", VUB - "Doctor in de Wetenschappen")
Academic year 2018-2019

Supervisors : Professor Barbara CLERBAUX
(Université libre de Bruxelles)

Professor Steven LOWETTE
(Vrije Universiteit Brussel)

Thesis jury :

Professor Michel TYTGAT (Université libre de Bruxelles, Chair)
Professor Freya BLEKMAN (Vrije Universiteit Brussel, Secretary)
Professor Giacomo BRUNO (Université catholique de Louvain)
Professor Shahram RAHATLOU (Sapienza Università di Roma)

fnrs FREEDOM TO RESEARCH

Contents

List of Figures

List of acronyms

AD	Antiproton Decelerator
APD	Avalanche Photodiodes
ATLAS	A Toroidal LHC ApparatuS
BEH	Brout-Englert-Higgs
BG	Background
BSM	Beyond the Standard Model
CDM	Cold Dark Matter
CERN	European Laboratory for Particle Physics
CL	Confidence Level
CMB	Cosmic Microwave Background
CMS	Compact Muon Solenoid
CR	Control Region
CSC	Cathode Strips Chamber
CTF	Combinatorial Track Finder
DM	Dark Matter
DT	Drift Tube
DY	Drell-Yan
EB	ECAL Barrel
ECAL	Electromagnetic Calorimeter
EE	ECAL Endcaps
EFT	Effective Field Theory
EW	Electroweak
FCP	Fractionally Charged Particle
FED	Front End Driver
FPGA	Field Programmable Gate Array
GUT	Grand Unification Theory
HB	HCAL Barrel

HCAL	Hadronic Calorimeter
HE	HCAL Endcaps
HF	HCAL Forward
HLT	High-Level Trigger
HO	HCAL Outer
HPD	Hybrid Photodiodes
ISR	Initial State Radiation
L1/L1T	Level-1 Trigger
LHC	Large Hadron Collider
LINAC	Linear Accelerator
LO	Leading Order
MB	Muon Barrel
MC	Monte Carlo
ME	Muon Endcaps
MET	Missing E_T, Missing Transverse Energy
MIP	Minimum Ionizing Particle
NLO	Next-to-Leading Order
PDF	Parton Distribution Function
PF	Particle-Flow
PS	Proton Synchrotron
PUPPI	PileUp Per Interaction
PXB	Pixel Barrel
PXF	Pixel Forward
QED	Quantum Electrodynamics
QFT	Quantum Field Theory
RBX	Read-out Box
RF	Radiofrequency
ROC	Read-out Chip
RPC	Resistive Plate Chamber
SA	Standalone
SC	Supercluster
SL	Superlayer
SM	Standard Model

SPS	Super Proton Synchrotron
SSB	Spontaneous Symmetry Breaking
STEAM	Strategy for Trigger Evaluation And Monitoring
TEC	Tracker Endcaps
TIB	Tracker Inner Barrel
TID	Tracker Inner Disks
TOB	Tracker Outer Barrel
TP	Trigger Primitive
TT	Trigger Tower
VBF	Vector Boson Fusion
VPT	Vacuum Phototriodes
WIMP	Weakly Interacting Massive Particle

Abstract

The Standard Model of particle physics is the framework that describes all known phenomena and interactions between elementary particles. It has proven to give outstanding results over the years and was succesfully completed with the discovery of the Brout-Englert-Higgs boson in 2012 by the ATLAS and CMS collaborations at CERN. However, several observations escape its reach: the matter-antimatter asymmetry, the nature of Dark Matter or the quantization of the electric charge. These are all examples of measured facts not explained by the Standard Model formalism and that call for an extension to a *Beyond the Standard Model* (BSM) theory. In this thesis, we have looked for evidence of new physics using proton-proton collision data produced by CERN's Large Hadron Collider (LHC) at a center-of-mass energy of 13 TeV. Collected from 2016 to 2018 by the CMS detector, it corresponds to an integrated luminosity of 136 fb^{-1}. After an introduction to the theoretical context and the experimental tools, two analyses are presented. The first one is a search for Dark Matter particles recoiling against a jet and leaving the detector unnoticed. With this "monojet" analysis, we are able to exclude mediator masses up to 1.8 TeV, and masses of Dark Matter particles up to 700 GeV. The second analysis is a search for fractionally charged particles. Using the fact that their stopping power is lower than Standard Model particles, we are able to exclude their existence up to masses of 765 GeV for a charge of $2/3\ e$.

Abstract

Le Modèle Standard de la physique des particules est le cadre qui décrit l'ensemble des phénomènes et interactions entre les particules élémentaires. Au cours des années, il a donné d'excellents résultats et se vit compléter en 2012 lorsque les collaborations ATLAS et CMS découvrirent le boson de Brout-Englert-Higgs au CERN. Cependant, certaines observations expérimentales échappent à sa portée: l'asymétrie matière-antimatière, la nature de la matière noire ou la quantification de la charge électrique. Autant d'examples de faits expérimentaux que le Modèle Standard n'explique pas et qui appellent à élargir le cadre théorique actuel vers des théories au-délà de la physique connue. Dans cette thèse, nous avons cherché à mettre en évidence des signaux de nouvelle physique en utilisant les collisions de protons effectuées au Large Hadron Collider (LHC) du CERN à une énergie dans le centre de masse de 13 TeV. Collectées de 2016 à 2018 par le détecteur CMS, ces données correspondent à une luminosité integrée de 136 fb^{-1}. Après une introduction au contexte théorique et aux outils expérimentaux, deux analyses sont présentées. La première est une recherche de particules de matière noire produites avec jet et quittant le détecteur sans avoir été detectée. Cette analyse "monojet" nous a permis d'exclure l'existence du médiateur entre le Modèle Standard et la matière noire jusqu'à une masse de 1.8 TeV, et l'existence de particules de matière noire jusqu'à une masse de 700 GeV. La deuxième analyse est une recherche de particules à charges fractionnaires. En utilisant le fait que le pouvoir d'arrêt de ces particules est plus faible que celle des particules du Modèle Standard, nous avons pu exclure leur existence jusqu'à une masse de 765 GeV pour des particules de charge 2/3 e.

Abstract

Het Standaard Model van de deeltjesfysica is het kader dat alle gekende fenomenen en interacties tussen elementaire deeltjes beschrijft. Gedurende jaren heeft het bewezen uitmuntende resultaten te geven en het was succesvol vervolledigd met de ontdekking van het Brout-Englert-Higgs boson in 2012 door de ATLAS en CMS experimenten aan het CERN. Verschillende waarnemingen echter zijn niet beschreven: de materie-antimaterie asymmetrie, de natuur van Donkere Materie, of de quantisatie van de elektrische lading. Dit zijn allen voorbeelden van gemeten feiten die niet verklaard worden door het Standaard Model formalisme, en die een uitbreiding vereisen naar een theorie voorbij het Standaard Model. In deze thesis hebben we gezocht naar sporen van nieuwe fysica met proton-proton botsingsdata geproduceerd door CERNs Large Hadron Collider aan een massamiddelpuntsenergie van 13 TeV. Geregistreerd tussen 2016 en 2018 door de CMS detector, komt dit overeen met een geintegreerde luminositeit van 136 fb^{-1}. Na een inleiding tot de theoretische context en de experimentele tools worden twee analyses voorgesteld. De eerste is een zoektocht naar Donkere Materie deeltjes die terugstoten tegen een zogenaamde jet en die de detector ongehinderd verlaten. Met de "monojet" analyse kunnen we massas van het medierende deeltje uitsluiten tot en met 1.8 TeV, en massas van het Donkere Materie deeltje tot en met 700 GeV. De tweede analyse is een zoektocht naar fractioneel geladen deeltjes. Gebruik makend van het feit dat de zogenaamde stopping power lager is dan voor Standaard Model deeltjes, kunnen we hun bestaan uitsluiten tot en met massas van 765 GeV voor een lading van 2/3 e.

Introduction

For thousands of years, curious people have been asking themselves simple questions about the nature of things: what are we made of, what is our universe made of? Those questions brought the scientific community to take interest in increasingly smaller objects in a quest to unravel the nature of the elementary building blocks of matter. Following the intuition of ancient greek philosophers on these $\alpha\tau o\mu o\varsigma$ (atomos, unbreakable) objects, chemists in the 19th century were the firsts to rigorously classify the simplest components of matter known at that time [1]. They made the distinction between complex chemical compounds and simple elements which Dmitri Mendeleev classified in his famous table in 1869 [2].

However it was soon realized that the atom, despite its name, was not truly elementary. In 1897, physicist Joseph Thomson interpreted his experimental results on cathode rays as evidence for the existence of an electrically charged subatomic particle: the electron [3]. This was followed by his proposal of the commonly called *Plum Pudding* model for the structure of the atom, in 1904 [4]. In this model, a positively charged volume is filled with negatively charged electrons. Ernest Rutherford, directing the work of Geiger and Marsden, proved Thomson wrong in a paper published in 1911 [5] where he presented the results of a scattering experiment. Using α and β rays from radioactive sources, he proved that all the positive charge of the atom is actually located in a very small central region. This led to the discovery of the atomic nucleus and the development of the Rutherford–Bohr atomic model in 1913 [6]. In 1919, Rutherford demonstrated that the positive charge inside the nucleus was made of Hydrogen nuclei which he later named *protons* [7]. James Chadwick completed this work by discovering in 1932 the other nuclear subcomponent, the *neutron* [8].

From this series of events, it seems that the study of the elementary components of matter has been following two paths, closely intertwined. The first one is the continuous change of paradigm regarding the understanding of what matter is and how it behaves. It implies to develop models that describe new realities with the mathematical tools at hand. On this road, one of the most important milestones is probably Max Planck's 1900 hypothesis that the electromagnetic energy emitted from a black body in thermal equilibrium is quantized [9]. In his 1905 paper on the photoelectric effect [10], Albert Einstein showed that these quanta of energy are to be understood as particles of light (*lichtquanta*) that were later renamed photons. These ideas layed the foundations of a radically different conception of matter based on quantized energy states: quantum mechanics. The quantum theory rules the microscopic world at the scale of the molecules and smaller and has proven to give outstanding results.

The year 1905 was quite productive for Einstein as he also presented a new theory of electrodynamics in a moving referential [11]. His *theory of special relativity* corrected Galileo's approach that collapses at very high velocities, close to the speed of light c. Soon enough, physicists such as Oskar Klein, Walter Gordon and Paul Dirac attempted to come up with an equation that could describe the behaviour of microscopic fast moving object, i.e. a quantum and relativistic equation. The Klein-Gordon equation was presented in 1926 [12, 13] followed in 1928 by the Dirac equation [14]. This is the beginning of *Quantum Field Theory* (QFT), a theory suited for the characterization of microscopic objects at high kinetic energies.

We just presented one side of the coin but the need for new frameworks to describe our world was twofold. On one hand, it resulted from theoretical inconsistencies that were hints that

something was not fully understood. But such problematic theoretical situations were not the only motivations to go beyond classical physics. Starting from the mid-1800's, new observations and measurements enabled by technological innovations came challenging the old models. The extending experimental reach is the second path subatomic physics followed throughout the 20th century and until today, more than ever. Following Louis de Broglie's 1925 hypothesis that the momentum and wavelength of a quantum object are inversely proportional [15], physicists understood that in order to study the internal structure of very small objects they needed to somehow give them a large kinetic energy. The first accelerator was invented in 1930 by John Cockcroft and Ernest Walton and used two years later to break Lithium nuclei with accelerated protons [16]. In parallel to these endeavours, Enrico Fermi led the effort towards a deeper knowledge of the atomic nucleus. In 1942, he and his team initiated and controlled the first self-sustained chain nuclear reaction in the first nuclear reactor, the *Chicago Pile 1* [17]. This enabled the discovery of the neutrino by Clyde Cowan and Frederick Reines at the Hanford site reactor in 1956 [18]. These experimental breakthroughs opened the way to the high-energy physics era.

In this historical presentation, five subatomic objects have showed up: the proton and the neutron that compose the atomic nucleus, the electron that orbits around it, the photon, and the neutrino. The two *nucleons*, as the proton and the neutron are generically called, were later found to be composite objects but to that day the three others are believed to have no internal structure. They are called elementary particles. Along the 20th century, the field of particle physics developed exponentially. Many more composite and elementary particles were predicted and discovered, not always in that order (the discovery of the muon, for instance, was a surprise for the community). Eventually, a model for the decription of these elementary particles and the interactions between them was built. It was written in the language of Quantum Field Theory and was fed by the experimental observations. It is called the *Standard Model* (SM) of particle physics. Although challenged by indirect observations, the Standard Model is currently the most coherent complete theory able to describe all non-gravitational phenomena. This is where the quest for ever more elementary objects is now standing.

Today, particle physics is at a peculiar stage of its development. The situation is exciting as we know our standard theory shows cracks but we have not directly observed any new physics signal. The searche for experimental evidence of *Beyond the Standard Model* (BSM) phenomena has been at the core of the physics program of hundreds of experiments around the globe and so for decades. This work is a humble brick to this construction of the human understanding of nature. Using some of the current most advanced experimental tools that are CERN's *Large Hadron Collider* (LHC) and the *Compact Muon Solenoid* (CMS) detector, we have looked for evidence of new physics phenomena. In the LHC, proton beams are accelerated and collided at a center-of-mass energy of 13 TeV, the highest energy ever reached by a human-made accelerator. With this huge energy available, the hope is to produce never observed particles in rare events. A most famous example of such *terra incognita* is the nature of Dark Matter (DM). Many cosmological observations, such as galaxy rotation curves or temperature fluctuations in the Cosmic Microwave Background (CMB), point to the fact that there must exist something more than just ordinary matter of which we are made. Currently, Dark Matter has only been observed to interact through gravitational effects. This means that, should it interact with standard matter, it would do so very weakly. In a first analysis, we tried to find a signal of Dark Matter particles escaping the detector unnoticed. If produced inside CMS, they would leave without interacting with the matter of the detector and we would be left with *missing energy*. Another example of a fact the Standard Model formalism does not encompass is the observed quantization of the electromagnetic charge of the elementary particles. Nothing in the Standard Model imposes the charge to be quantized and many BSM theories (among which theories of Dark Matter) predict the existence of particles carrying a lower fractionally charge. In a second analysis, we searched for fractionally charged particles using a measurement of the particle ionization in the CMS tracker.

Outline

This thesis is divided into two main parts. The first four chapters serve as theoretical and experimental introductions. They allow to properly set the scene for the original research that is presented afterwards. In the second part, i.e. chapters 5 and 6, I present the two main physics searches in which I was involved during these past four years. Below, I briefly review the content of each chapter to help the reader through the thesis.

The first chapter of this work intends to lay out the status of particle physics in 2019. I start by showing how this field of physics was built and led to the construction of the Standard Model of particle physics. The known particles and interactions are reviewed and their main characteristics are given. To conclude, some of the main challenges of modern high-energy physics are presented such as the quest for Dark Matter or the puzzle of electromagnetic charge quantization. They serve as a motivation for the physics searches that follow in the two last chapters.

In the second and third chapters of this work, the experimental context of the thesis is established. First, the Large Hadron Collider (LHC) and the technical details of its operations are given in chapter 2. This is relevant to understand how exactly the proton-proton collisions are produced and how the center-of-mass energy and the luminosity are defined. This is followed by chapter 3 on the CMS detector. All the subdetectors are reviewed and their performance during the relevant time periods are given. I also go through the reconstruction procedure, linking the detector information to the physics objects used in physics analyses. Finally, I briefly go over the trigger system which is a crucial feature of high intensity collider experiments. In this context, I must introduce the *Experimental Physics Responsibilities*, or EPR. In CMS, physicists are assigned experimental tasks to help the collaboration and make sure the data used in analyses are well calibrated and understood. For more than two years, I was a leading member of the Strategy for Trigger Evaluation And Monitoring (STEAM) group in charge of the prediction and monitoring of the trigger rates and prescales. In communication with the whole Trigger Studies Group (TSG) and the analysis contacts, we were developing lists of trigger paths ("menus") and making sure the expected rates were not overshooting the CMS capacities.

Chapter 4 has a special place in this thesis. It is concerned with an important variable in the collision event description, the missing transverse energy (E_T^{miss} or "MET"), and could have found its place in the CMS chapter. In the context of the EPR, I was also an active member of the CMS group responsible for the management of this variable and the monitoring of its behaviour through the data taking periods. As a consequence, I decided to dedicate a whole chapter to it. One of my duties in this group was to be responsible for what we call the MET filters. They are used to reject events containing large MET and in which the source of the MET is suspected to come from instrumental issues. In this chapter, I define the MET, how it is measured and controlled and I go over the different MET filters. I finish by presenting an original personal research project I carried out with the conveners of the MET group. The goal was to investigate whether we could estimate the energy lost in dead regions of the electromagnetic calorimeter.

The search for Dark Matter particles in the "monojet" final state that is presented in chapter 5 is the first data analysis I was involved in. This was my first experience in the field and it taught me how to choose a good trigger to efficiently target the signal, how to properly do an

event selection with background rejection cuts and finally how to set upper limits on the signal cross-section. As the monojet signature is made of at least one high energy jet and missing transverse energy, the general strategy is to look for an excess of data in the MET spectrum. Therefore, my expertise with the MET filters is really where I contributed here. I was in charge of making sure the filters were working as expected and were not rejecting "good" events. In the second part of chapter 5, I present a related analysis where we looked for invisible decays of the Higgs boson. As the target final state is also made of jets and missing transverse energy, the procedure is similar to what is done in the monojet analysis. Although mainly estimated directly from data, a significant part of the background estimation method relies on Monte Carlo simulations. My contribution here was to compute "k-factors" that were used to correct the expected number of events for the two main backgrounds. This enhanced the sensitivity of the analysis.

Benefiting from the experience accumulated as a member of the monojet analysis team, I started to build a whole new search with my supervisors. For a little more than a year, we developed an analysis to look for particles carrying a low electromagnetic charge, i.e. lower than that of the electron. The existence of such particles is well motivated as they appear in supersymmetric theories as well as in string and Grand Unified Theories. Chapter 6 presents this work and it composes the bulk of the original research of this thesis. I start by reviewing the theoretical motivations for such a search and how fractionally charged particles are constrained by previous observations. Then, I go over the full analysis procedure from the model phenomenology to the results. Finally, I show that our results are competitive with respect to previous works.

Chapter 1

Particle physics in 2019

Motivated abstract

The late 19th century and the first half of the 20th century saw the development of subatomic physics. The atom was broken down to subpieces that were later called the proton, the neutron and the electron. The discovery of the free-propagating electron in 1897 opened the route to the particle physics era. Consequently, quantum mechanics was developed to describe this world of infinitely small objects. With the ever higher energies reached by the accelerators, the need of a relativistic quantum theory arose. The Standard Model (SM) of particles physics is the result of such a description of the elementary components of nature. It describes all observed particles and interactions except gravity. It is still today the most solid description of the quantum world we have, although we see some cracks in its structure. These deviations from the SM predictions have motivated the quest for new physics that has been on for decades. This thesis participates to this effort.

In this chapter, we present the current status and challenges of modern particle physics. We first start by discussing the main characteristics of the Standard Model of particle physics, the elementary particles and the fundamental interactions. We then give a taste of the challenges modern particle physicists have to address. This chapter motivates the searches for new physics presented in chapters 5 and 6 of this thesis.

1.1 The Standard Model

The Standard Model of particle physics is the present theory that describes the elementary constituents of matter and the forces between them. It is currently the most coherent structure we have to describe all observations, although some of them escape its reach. Following the successes of quantum mechanics in the first half of the 20th century, its construction really started in the 1960's and was successfully completed in 2012 when the Higgs boson was observed by the CMS and ATLAS experiments at CERN.

In quantum mechanics particles are pictured as fields $\psi(x)$, solutions of the Schrödinger equation. This formalism is applicable to low energy systems such as the Hydrogen atom. The SM is the result of the will to generalize the Schrödinger equation to the relativistic regime, i.e. to describe the high-energy quantum world. Energy is here meant as kinetic energy, i.e velocities close to the speed of light c. The history of quantum mechanics led to the quantization of the energy levels in a quantum system. Similarly, the effort to find a relativistic version of the Schrödinger equation resulted in the quantization of the solutions to these equations, the fields themselves. In the SM, matter particles are pictured as quantized fields that can be created and destroyed. Such a framework is called a quantum field theory (QFT) and the procedure that leads to it is referred to as the "second quantization". More information can be found in reference [19]

for all QFT-related details. The first attempt to describe the relativistic behaviour of quantum fields, i.e. to generalize the Schrödinger equation to special relativity, was done by Oskar Klein and Walter Gordon [12, 13]. However, the Klein-Gordon equation showed to be applicable to spinless particles only. Paul Dirac pushed the effort further with his equation which solutions proved to describe half-integer spin particles. For the first time, it also predicted the existence of antiparticles. All matter particles, introduced below, are spin 1/2 fields and therefore solutions of the Dirac equation. They are called Dirac fermions as they follow the Fermi-Dirac statistics. Integer-spin particles follow the Bose-Einstein statistics and are called bosons. We will see that interaction particles are bosons.

The mathematical description of the SM uses Lagrangian field theory. It is a generalization to quantum field theory of the Lagrangian formalism in classical mechanics. For a formal description of the Lagrangian formalism, we refer to Appendix III of [20]. The SM Lagrangian contains all the information of the theory: its particle content and the interactions allowed between them. It is built on the principle of local gauge invariance. Let us show this mechanism with a simple example on the following Lagrangian:

$$L = i\bar{\psi}\partial_\mu\gamma^\mu\psi - m\bar{\psi}\psi \tag{1.1}$$

where ψ is a Dirac fermion, and γ_μ are the Dirac matrices. The first term is called the kinetic term of the field ψ and the second term is the mass term. The bar symbol stands for $\psi^\dagger\gamma^0$ where γ^0 is the 0^{th} Dirac matrix. This Lagrangian is invariant under the global phase rotation $\psi(x) \to e^{i\alpha}\psi(x)$. However, we would like it to be invariant under a local phase rotation such as $\psi(x) \to e^{i\alpha(x)}\psi(x)$. The argument behind this is that in the absence of a way for two fields to communicate through space and time, a local change in their phase should not change the physics of the problem. Under the local gauge transformation, the Lagrangian becomes:

$$L = i\bar{\psi}\left(\partial_\mu\gamma^\mu + i\partial_\mu\alpha(x)\right)\psi - m\bar{\psi}\psi \tag{1.2}$$

The first term in the Lagrangian is not invariant anymore. To ensure invariance, one must add a new ingredient: the gauge field associated to the gauge transformation performed. In our case, this goes through the redefinition of the vector of partial derivatives ∂_μ into a covariant derivative $D_\mu = \partial_\mu + ieA_\mu$. The new vector field A_μ is required to transform as $A_\mu \to A_\mu - \frac{1}{e}\partial_\mu\alpha(x)$ to compensate for the partial derivative of the phase in equation (1.2). This gauge field is exactly what the theory is missing: a way to propagate the information from one matter field to another. One notices that the new field must have the same structure as the partial derivative, i.e. it is a vector. As such, it has a spin 1 and is therefore a boson. All gauge fields are bosons.

We should now add a kinetic term for the new field A_μ that keeps the Lagrangian invariant under the local phase rotation. Defining $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, a possible term is $F_{\mu\nu}F^{\mu\nu}$. With a conventional factor in front of it, the new Lagrangian is:

$$L = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}D_\mu\gamma^\mu\psi - m\bar{\psi}\psi \tag{1.3}$$

This is the Lagrangian of Quantum Electrodynamics (QED), the quantum field theory of electromagnetism. We shall see shortly that electromagnetic interactions are actually embedded in a more general theory. The transformation group of all n×n unitary matrices is called the U(n) group. The phase rotation presented above belongs to the U(1) group. The second term in equation (1.3) tells us that the ψ matter field can talk to the A_μ gauge field. We say that ψ is charged under the U(1) group, i.e. the gauge field A_μ allows ψ particles to interact with each other through it. The Lagrangian (1.3) is only a part of the full SM Lagrangian. The full Lagrangian is invariant under the combined SU(3)×SU(2)×U(1) groups, where the SU(n) groups are the groups of n×n unitary matrices of determinant 1. A charge is associated to each symmetry group: the hypercharge Y for U(1), the third component of the weak isospin T_3 for SU(2) and the color C for SU(3). The number of degrees of freedom of a charge is determined by the dimension of the corresponding group: n^2 for U(n) groups and $n^2 - 1$ for SU(n) due to the

additional constraint on the determinant. The dimension of the symmetry group also determines the number of associated gauge fields: this makes 1 gauge field for the hypercharge group, 3 gauge fields for the weak isospin group and 8 gauge fields for the color group. The SM symmetry group is often written as $SU(3)_C \times SU(2)_L \times U(1)_Y$ to account for the physics of the symmetry groups, where L stands for "left-handed" chirality. The chirality is related to the weak isospin of the particle. Left-handed particles have $T = 1/2$ and form SU(2) doublets with $T_3 = \pm 1/2$. Right-handed particles have $T = 0$ and are SU(2) singlet, i.e. they are not charged under SU(2).

Sheldon Glashow, Abdus Salam, and Steven Weinberg unified the $U(1)_Y$ and $SU(2)_L$ parts of the SM in a global description of both the electromagnetic and the weak interactions [21, 22, 23]. This is called the electroweak (EW) theory. In this theory, the hypercharge boson and the three weak isospin bosons mix to form the physical observed states: the photon γ that carries the electromagnetic force and the three weak bosons $W^{\pm}$ and Z^0 that carry the weak force. Unfortunately, this theory is unable to assign a non-zero mass to the resulting gauge bosons without breaking gauge invariance. This was solved in 1964 by Robert Brout, François Englert and Peter Higgs by adding a scalar field to the picture [24, 25]. They postulated that, although symmetric, the potential of this scalar field has a non-zero expectation value at its minimum. After Spontaneous Symmetry Breaking (SSB), the BEH mechanism allows the four massless gauge fields to transform into the three observable massive weak gauge bosons and the massless photon. Let us notice that the BEH mechanism gives an explanation for the way particles acquire mass but does not predict the value of the masses. It does not explain either why the SM particles have such mass discrepancies. The scalar field itself gains mass by self-interaction and can propagate: it is called the Higgs boson and has been discovered at CERN by the CMS [26] and ATLAS [27] Collaborations in 2012.

The interaction associated to the $SU(3)_C$ group is called the strong interaction and the corresponding theory is Quantum Chromodynamics (QCD). The eight associated gauge bosons are the gluons. As opposed to the photon that is electrically neutral, the weak bosons can interact between them and so can the gluons. This is a feature of non-abelian symmetry groups.

Figure 1.1 shows all matter and interaction particles of the SM and the link between them. To each particle is associated an antiparticle with all charges reversed. The photon is its own antiparticle. Matter comes in three generations, ordinary matter being made of first generation particles only. Particles of the next generations are heavier and unstable, eventually decaying back to first generation particles when produced. Matter particles are further divided in quarks and leptons. Quarks come in six different *flavours*: up, down, charm, strange, top and bottom. To each flavour is associated a quantum number conserved by the strong interaction. Quarks are color-charged particles, i.e. each quark flavour is a color triplet of the SU(3) group. With regard to the strong interaction, there exists three different versions of each quark flavour, treated differently. QCD tells us that colored objects cannot propagate freely, only "white" composite objects are observable. Those QCD bound states are called hadrons. Depending on their quark content, hadrons have a different name: three quarks of three different colors form a baryon while a quark and an antiquark of the anticolor is called a meson. It is worth noting that the top quark is too massive to be involved in QCD bound states. Its decay time to lightest states is shorter than the time necessary to form hadrons. Left-handed quarks are grouped in SU(2) doublets made of one up-type and one down-type quark. Since the SU(2) group treats equally the two members of a doublet, weak interactions do not conserve the flavour. All quarks carry a fractional electric charge, therefore quarks feel all interactions.

Leptons do not feel the strong interaction. They are grouped in SU(2) doublets containing a charged lepton and the associated neutrino. However, neutrinos only have left chirality. This means anti-neutrinos only have right chirality. Right-handed neutrinos do not exist in the SM and the right-handed charged leptons are SU(2) singlets. As a consequence, no gauge-invariant neu-

trino mass term is possible. Neutrinos are massless in the SM. Due to their zero electric charge, neutrinos only feel the weak interaction. The neutrino-nucleon cross-section is consequently very low at low center-of-mass energies, making them extremely hard to detect. All leptons carry a flavour quantum number according to their generation: L_e, L_μ and L_τ. These numbers are conserved in electroweak processes.

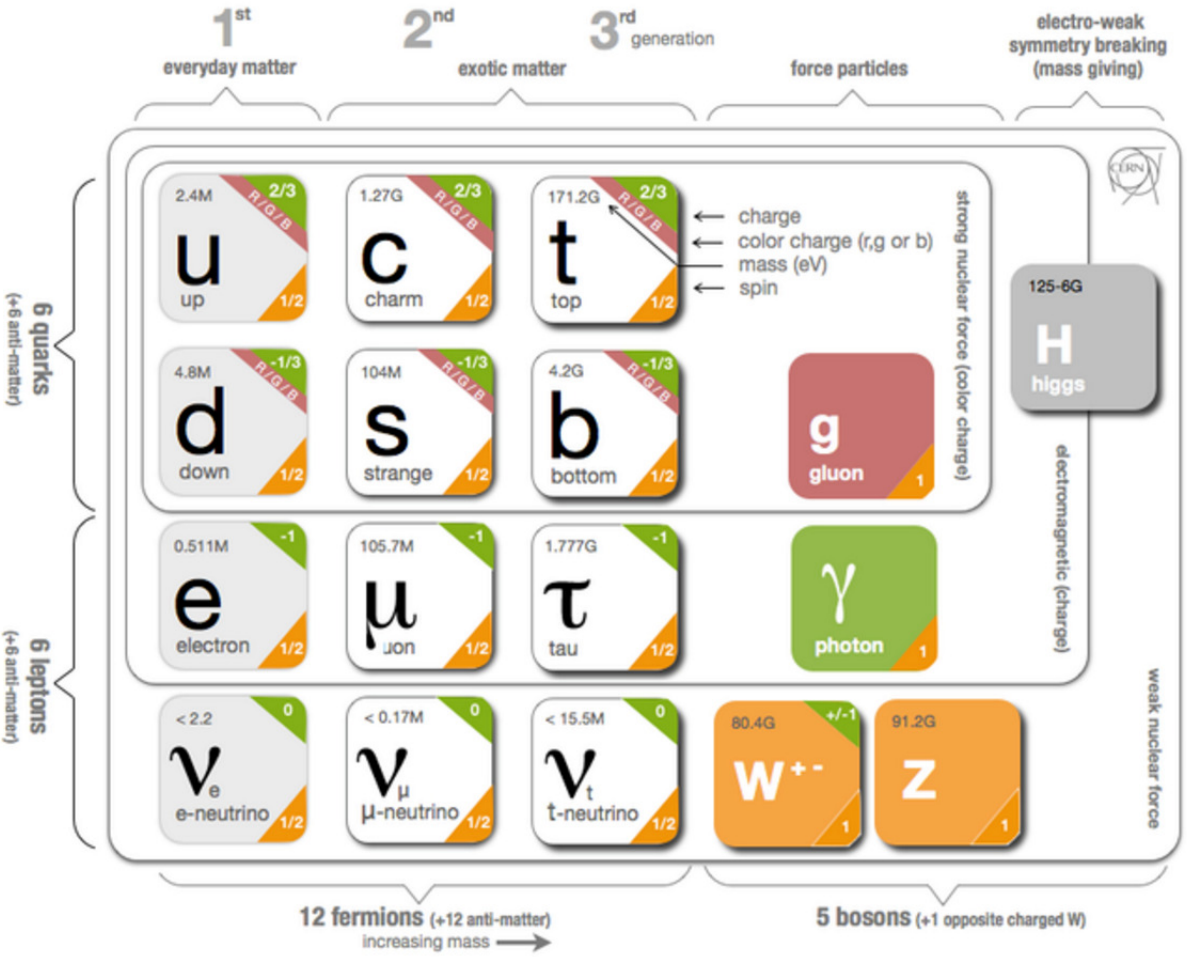


Figure 1.1: The elementary particles and the fundamental vector of interactions in the Standard Model (made by CERN).

1.2 Current challenges

Although most physics experiments show the observations to be in agreement with the SM predictions, many questions stay unanswered. They are briefly reviewed here with an emphasis on the dark matter sector.

1.2.1 Why is the universe made of matter?

Since Dirac's first suggestion and the subsequent discovery of the positron by Anderson in 1932, we know antimatter exists. However, it is still unkown why the universe is made of matter and not antimatter. This fact was ascertained not only by earth-based conclusions but also by the observation of cosmic rays from extra-galactic sources: they are made of baryons and not anti-baryons. The question of the asymmetry boils down to what happened in the very early moments of the universe: either more matter was initially created or equivalent amounts of matter and antimatter were produced but somehow antimatter disappeared. The latter is often preferred as

it leaves the asymmetry to a later phenomenon that we can describe and not to the out-of-reach initial conditions [28].

To understand the matter-antimatter asymmetry, we must discuss the symmetries that appear in the Standard Model. Parity P is a "mirror" symmetry under which each space coordinate is reversed, $\vec{x} \to -\vec{x}$. After the first observations of parity violating processes, such as in the muon decay, theorists built a stronger symmetry that was thought to be conserved in every process: the conjugation of parity with the charge conjugation C. The action of the latter symmetry is to replace particles with their antiparticles. If matter and antimatter behaved the same, physics would be invariant under the CP symmetry. The muon decay is symmetric under this change as illustrated in figure 1.2. However, the observation of CP violation in Kaon electroweak decay in 1964 [29] showed that not all processes are CP-symmetric. This means that we know there is an asymmetry in the way matter and antimatter are treated by the electroweak force. Answers to these questions are sought at many experiments among which the ones working with CERN's Antiproton Decelerator (AD) beam [30, 31, 32, 33, 34].

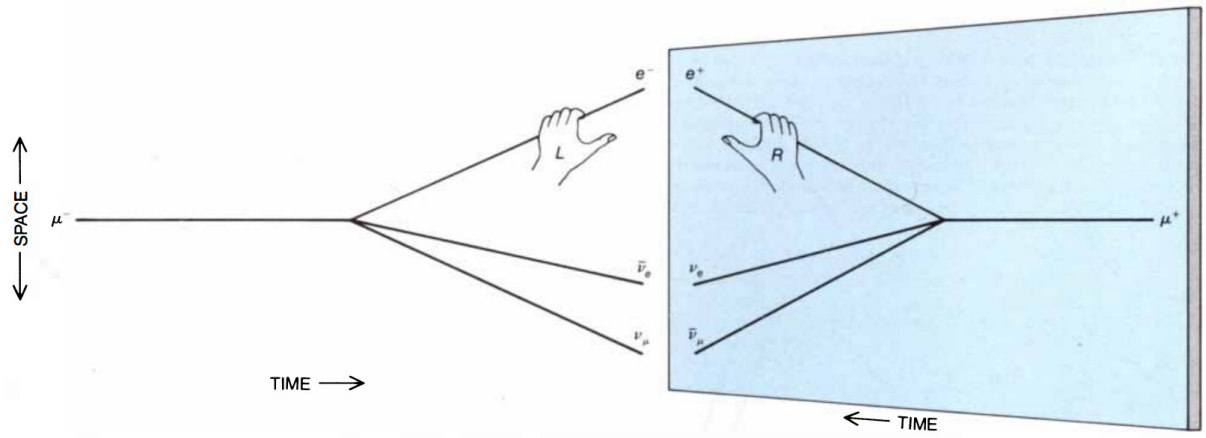


Figure 1.2: Muon decay and CP-symmetric process of the antimuon decay [28]. Here, CP is conserved meaning that both processes have the same rate or probability.

1.2.2 The neutrino puzzle

When looking into neutrino physics, many questions arise. The first peculiar observation is that only left-handed neutrinos (and right-handed antineutrinos) participate in weak interactions. Being the only neutral fermions of the SM, one might wonder whether neutrinos are their own antiparticles, i.e. whether they are Majorana or Dirac fermions. This question is tested with neutrinoless double beta decay experiments [35, 36]. The observation of a deficit in the solar neutrino flux led to the conclusion that neutrinos oscillate from one flavour to another [37]. This is due to the fact that that neutrinos are produced as flavour (and not mass) eigenstates in weak interactions. This phenomenon was further confirmed with the observation of neutrino oscillations in atmospheric and reactor neutrino fluxes. These oscillations depend on the mass difference between the neutrino mass states with the consequence that at least two of the three SM neutrinos carry a non-zero mass. As neutrinos are massless in the SM, the neutrino mass sector is another source of potential new physics [38]. Some experiments [39] try to directly measure the neutrino mass. Current limit show that it is much smaller than that of the other fermions. This could be another gate to new physics.

1.2.3 Where is the missing mass in the universe?

Numerous cosmological observations support the existence of a large amount of non-radiating matter in the universe which role cannot be endorsed by any SM particle. This non-visible mass is called Dark Matter (DM) and its nature is still unknown. Latest results show that DM accounts

for almost 85 % of the universe matter content.

The main experimental observations that led to the DM hypothesis are the following:

- When immersed in a gravitational field, astrophysical objects are observed to be moving faster than what is expected if gravity was only due to visible matter. This is observed in the rotation curves of stars in galaxies [40] and in the observation of galaxy clusters.
- The bullet cluster is a cluster of galaxies that entered in collision with another cluster. The comparison of the mass distribution from red-shift measurement and gravitational lensing showed two different results. This indicates that most of the mass is not in the form of baryonic gas but has a non-radiating nature.
- When the universe reached a temperature low enough for the atoms to form, it became electrically neutral and hence transparent to photons. The last light produced at that time is still propagating in the universe today and is a picture of this important transition. This photon gas is today much cooler than before due to the expansion of the universe and its frequency lies in the microwave range. It is called the Cosmic Microwave Background (CMB). The measurement of this radiation gives us a lot of information about the state of the universe at that time. In particular, this data can be fitted with predictions from the standard model of cosmology, namely the Λ-CDM model (Λ is the cosmological constant that appears in Einstein's equation and CDM stands for "Cold Dark Matter"). A measurement of the power spectrum of the temperature and polarisation has been done by the Planck [41] and WMAP experiments showing a striking agreement between data and the Λ-CDM model and a failure to fit in the absence of DM.

The most common interpretation of the nature of DM is that it is in the form of particles. Moreover, it is thought to exist as WIMPs, Weakly Interacting Massive Particles. Different approaches have been attempted to detect those WIMPs: direct detection, indirect detection and searches at colliders. In a direct detection experiment, physicists try to measure the nuclear recoil in a hypothetical DM-nucleon interaction. The de-excitation energy of the nuclei, typically around 10 keV, is measured through the heat elevation of the medium, through the emission of scintillation light or through ionization of neighbouring atoms by elastic scattering with the recoiling nucleus [42]. The current most advanced DM direct detection experiments are Xenon1T [43], Lux [44] and PandaX-II [45]. In indirect detection experiments, DM particles are assumed to annihilate to produce SM particles that could then be detected. This can be achieved with gamma-ray telescopes (such as Fermi-LAT [46]), neutrino telescopes (such as IceCube [47]) or cosmic-ray telescopes (such as PAMELA [48]) [49]. Finally, collider searches aim at producing DM particles out of proton-proton collisions through the decay of DM-SM mediators. In chapter 5 of this work, we present a search for DM particles with the CMS experiment. Figure 1.3 illustrates the difference between the three methods.

1.2.4 Is the electric charge quantized?

All electrically charged particles carry a charge proportional to that of the down-type quarks, i.e. a third of the electron charge. However, there is no a priori reason for the charge associated with an abelian gauge field to be quantized [50]. Still, the requirement that chiral anomalies cancel out in one-loop processes translates into the fact that all fermions in the same generation must carry a charge proportional to each other. A generic example of such diagram for different gauge bosons is given by figure 1.4.

The theory is anomaly free if, for this diagram, the following relation is met [52]:

$$\mathrm{tr}\left(t^a\left\{t^b, t^c\right\}\right) = 0 \tag{1.4}$$

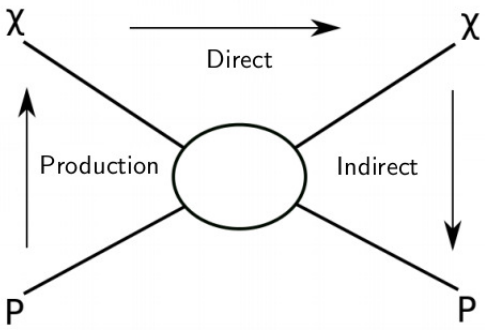


Figure 1.3: Illustration of the three methods used to look for DM particles. χ symbolises the DM particle and p is a proton or any SM particle depending on the context. In a direct detection experiment, one expect the DM particles to directly interact with a SM nucleon and measure its recoil. In an indirect detection experiment, SM products of a DM-DM annihilation process are expected to be detected by ground-based or satellite telescopes. At the LHC, we are trying to produce DM through pp collisions.

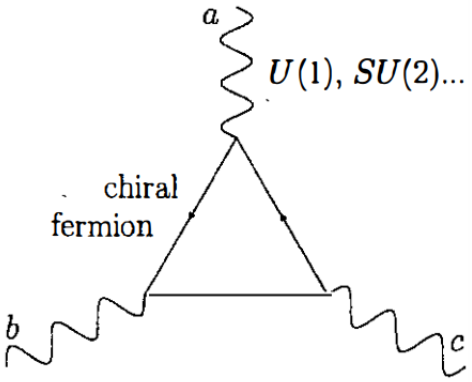


Figure 1.4: Fermion loop with the emission of three gauge bosons [51].

For the SU(2) group, equation 1.4 is rewritten as:

$$\mathrm{tr}\left(\sigma^a\left\{\sigma^b,\sigma^c\right\}\right)=0 \tag{1.5}$$

This equation is always true due to the properties of the sigma matrices. Now, applying this requirement to the $\mathrm{U}(1)_Y$ group again for the diagram of figure 1.4 yields:

$$\begin{aligned}\mathrm{tr}\left(Q\left\{t^b,t^c\right\}\right) &= \frac{1}{2}\mathrm{tr}\left[Q\right]\delta^{bc} &(1.6)\\ &= 0 &(1.7)\end{aligned}$$

The right-hand side of equation 1.6 is proportional to the sum of the fermion electromagnetic charges. Therefore, the requirement of a chiral anomaly free theory boils down to:

$$\begin{aligned}\Sigma_{1^{st}\text{generation fermions}}Q &= 3Q_u+3Q_d+Q_e\\ &= 3\left(\frac{2}{3}\right)+3\left(\frac{-1}{3}\right)+(-1)\\ &= 0 &(1.8)\end{aligned}$$

From this we understand that adding new chiral fermions to the theory is only possible if they carry a charge compatible with this anomaly cancelation.

Quantization of charge might appear naturally in models of new physics. In particular, Dirac showed that the existence of magnetic monopoles implied a quantization of the electric charge [50]. In the case the SM is the effective version of a UV complete theory, there are strong motivations to look for particles with charges lower than the electron charge. For instance, Grand Unification Theories with an SU(N) symmetry group breaking down to extra U(1) naturally lead to the existence of new particles with arbitrary charges [53]. This is developed in chapter 6 where a search for fractionally charged particles with the CMS experiment is presented.

Chapter conclusion

The Standard Model of particle physics was presented. Elementary particles and their properties were introduced. We defined quarks, charged leptons and neutrinos as forming what we call matter. Force carriers were also properly described: the photon for the electromagnetic interaction, the $W^{\pm}$ and Z^0 bosons for the weak interaction and the gluon for the strong interaction. The Higgs scalar field completes this picture by allowing some particles to acquire a non-zero mass. Then we picked a few challenges modern particle physics has to face: the matter-antimatter asymmetry, new physics in the neutrino sector, the quest for Dark Matter and the question of the electric charge quantization. The two last items are the subject of two distinct CMS searches presented in this thesis.

Chapter 2

The LHC

Motivated abstract

The Large Hadron Collider (LHC) is in 2019 the most powerful particle collider ever built. Located at CERN, it is a 26.7 km synchrotron able to accelerate protons up to a kinetic energy of 6.5 TeV. Since it started its high-energy operations in 2010, the LHC has been standing at both the luminosity and energy frontiers. Since 2015, it operates at a proton-proton (pp) center-of-mass energy of 13 TeV and in 2018 it reached a peak instantaneous luminosity of 20 nb^{-1}Hz. The searches carried on in this thesis use LHC proton-proton collision data collected by the CMS detector in 2016, 2017 and 2018 summing up to an integrated luminosity of 136 fb^{-1}.

This chapter is organized as follows. We start by giving a general overview of the LHC operations over the past few years. Then we continue with some details on how protons are accelerated, bunched and guided inside the accelerator ring. We finish with the description of the proton-proton interactions and the introduction of the concept of luminosity and pileup.

2.1 LHC operations

The LHC has been in operation at high energies since 2010, providing four experiments with collision data: CMS, ATLAS, ALICE and LHCb. The first three years are referred to as Run 1. During this period, the center-of-mass energy $\sqrt{s}$ was limited to 7 (2010-2011) and 8 (2012) TeV. During the Long Shutdown 1 (LS1) from 2013 to 2015, the machine was upgraded. Then the LHC Run 2 started for four more years of data taking at the center-of-mass energy of 13 TeV. Figure 2.1 shows the LHC timeline from mid-Run 1 until the end of Run 2. Results presented in this thesis exclusively use LHC Run 2 data.

The LHC is the last of a chain of accelerators as depicted in figure 2.2. Hydrogen atoms are stripped from their electrons before entering the Linear Accelerator (LINAC2) and accelerated up to an energy of 50 MeV. In the Booster, they reach an energy of 1.4 GeV before entering the Proton Synchrotron (PS) where they gain up to 25 GeV. The SPS is the last step before the LHC where protons are brought to an energy of 450 GeV. All beam lines are put down to very low pressure, i.e. a high vacuum, to avoid proton-gas interactions. The pressure inside the LHC vacuum tube when operating is of 10^{-6} mbar. All information regarding the LHC technical details can be found in references [55] and [56].

2.2 Acceleration and bunching

The LHC uses radiofrequency (RF) cavities to accelerate the protons. In an RF cavity, the voltage oscillates such that the protons always see an accelerating field. This is illustrated in figure 2.3.

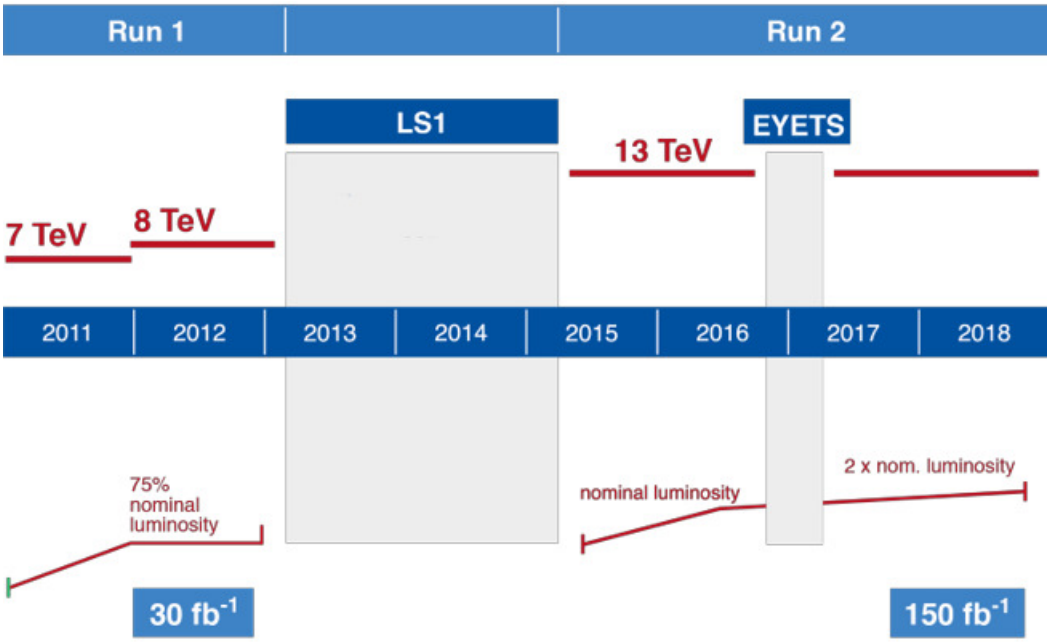


Figure 2.1: The LHC timeline from 2011 to 2018 (adapted from [54]). The integrated luminosities are approximative. LS stands for "Long Shutdown", EYETS for Extended Year-End Technical Stop.

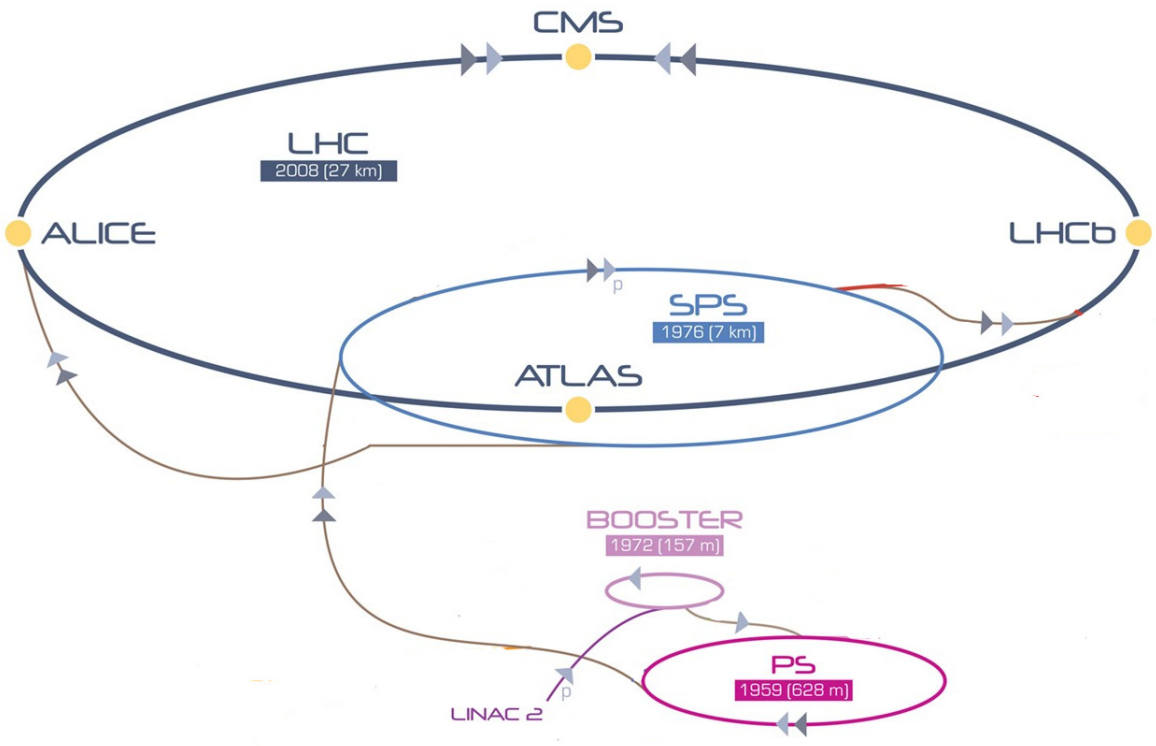


Figure 2.2: The LHC accelerator chain (credits to CERN).

Each beam is accelerated by 8 such cavities operating in superconducting conditions. This allows the cavities to provide a large accelerating electric field without much loss. The field experienced by the particles in the cavities changes amplitude while particles are passing by. This implies that depending on the particle arrival time, the field magnitude is different. Figure 2.4 illustrates what happens for particles with three different arrival times. Particle A is the reference particle for which the RF frequency is designed. It always feels the same field amplitude after each loop.

A particle having more energy such as particle B arrives a little earlier and sees a weaker field. It is then less accelerated then particle A and reduces its lead. For a slower particle with less energy, the exact opposite happens. This is the case for particle C that arrives later in the cavity and sees a stronger field. It is then given more energy and catches up the bulk of the particles. This effect leads to a longitudinal bunching of the particles. As opposed to cyclotrons that deliver continuous beams, synchrotrons produce bunched beams.

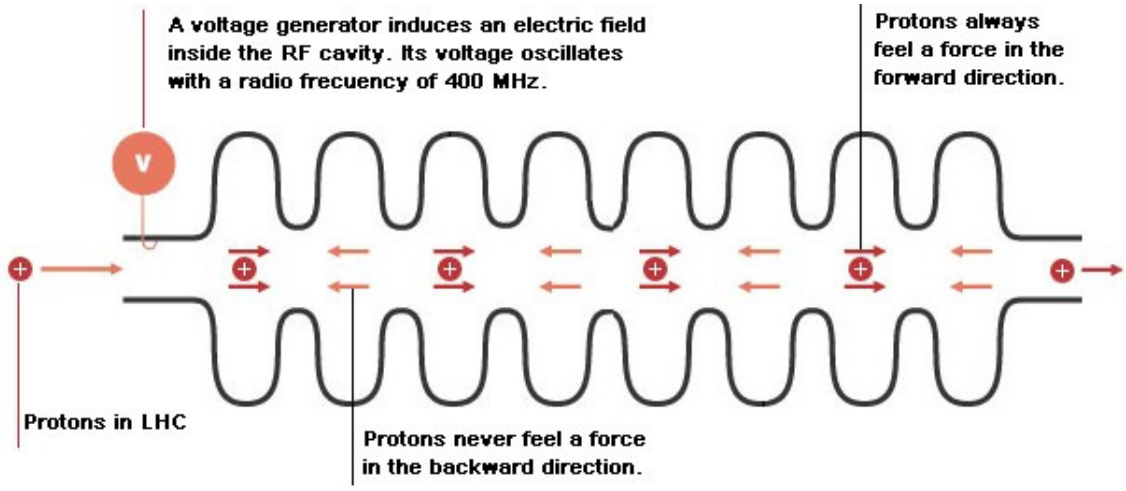


Figure 2.3: Schematic diagram of a radiofrequency cavity (credits to CERN). The charged particles always see an accelerating field.

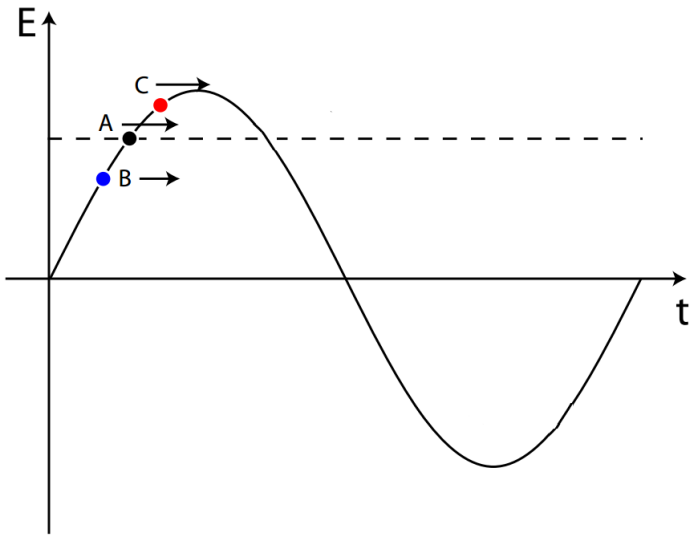


Figure 2.4: Bunching mechanism [57]. Particle A is the reference particle and is always accelerated by the same field amplitude at each passage. Particle B arrives earlier and is therefore accelerated by a weaker field which allows it to catch up. The opposite happens to particle C.

During the LHC Run 2, bunches were made of up to 1.1×10^{11} protons. They were separated by 25 ns, leading to a collision frequency of 40 MHz. This high collision rate is reduced by a trigger system to select events of interest, as is explained in section 3.4.

2.3 Magnets and beam optics

In synchrotrons, accelerated particles loop over a closed path while they are given more and more kinetic energy. In this configuration, the bending magnetic field must increase sychronously with the particles energy to keep the particles on the accelerator course. Dipole magnets fill

this task with a simple magnetic field topology. Proton beams pass through a vertically oriented field as illustrated in figure 2.5 (left). The resulting Lorentz force $\vec{F} = q\vec{v} \times \vec{B}$ curves the beam trajectory in the desired direction depending on the orientation of the dipole. The LHC contains 1,232 such dipole magnets operating at a magnetic field of 8.33 T. To reach such fields, Nb-Ti superconducting magnets are cooled down to 2 K using superfluid helium.

On top of the LHC dipoles, magnets of more involved topologies are used to focus the beams. This is needed to counteract the Coulomb repulsion between the protons that tends to increase the transverse size of the bunches. The simplest focusing magnet topology to describe is that of the 858 quadrupoles (figure 2.5 right). The field configuration is such that protons deviating from the center along one direction are brought back by the Lorentz force, while the deviation is aggravated in the other direction. This is therefore not useable as is. To properly focus charged particles beam in both directions, one must use what is referred to as FODO cells. They consist in two quadrupoles each focusing in one of the two directions and separated by empty space. The FODO name comes from the matrix formalism in optics where F stands for the focusing matrix, D for the defocusing matrix and O for the matrix of propagation in the vacuum.

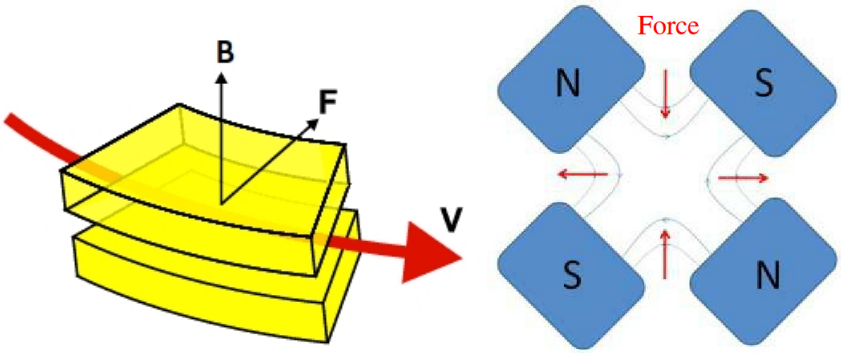


Figure 2.5: Sketch of a dipole (left [58]) and of a quadrupole (right [59]). On the left, **B** is the magnetic field, **V** the beam velocity and **F** the corresponding Lorentz force.

2.4 Proton-proton interactions

The amplitude of the proton-proton interaction is quantified by the corresponding cross-section that is made of an elastic and an inelastic part. The elastic cross-section corresponds to all processes where the protons exchange energy but are not destroyed. In inelastic collisions, the protons are broken to produce new hadrons. Figure 2.6 shows a compilation of different measurements of the total, elastic and inelastic pp cross-section as a function of the center-of-mass energy.

In addition to their small cross-section, elastic collisions lead to a very small deflection of the resulting protons with respect to the beam axis. The final state protons are often not even entering the detector volume. What is detected by CMS is the final state of pp inelastic collisions where the deflection of the final state particles with respect to the z axis is large enough. As for any inelastic QCD process, the nature of those interactions can be split into two categories: soft and hard interactions. A hard QCD interaction implies a large momentum transfer between the interacting partons. These are the events high-energy physicists are mostly interested in because a large energy is available for new physics such as the production of new heavy particles. In these interactions, two partons of the incoming protons interact in a way that can be described by perturbation theory. The rest of the protons will hadronize in soft processes: this is the underlying event described by non-perturbative QCD models.

When proton bunches collide inside the CMS detector, only a very small fraction of the $\sim 10^{11}$ protons interact. The number of extra pp interactions per bunch-crossing is called the pileup. However, in the CMS jargon we refer to the "pileup" as the total number of pp interactions per bunch-crossing. Unfortunately, the pp inelastic cross-section is dominated by the soft cross-

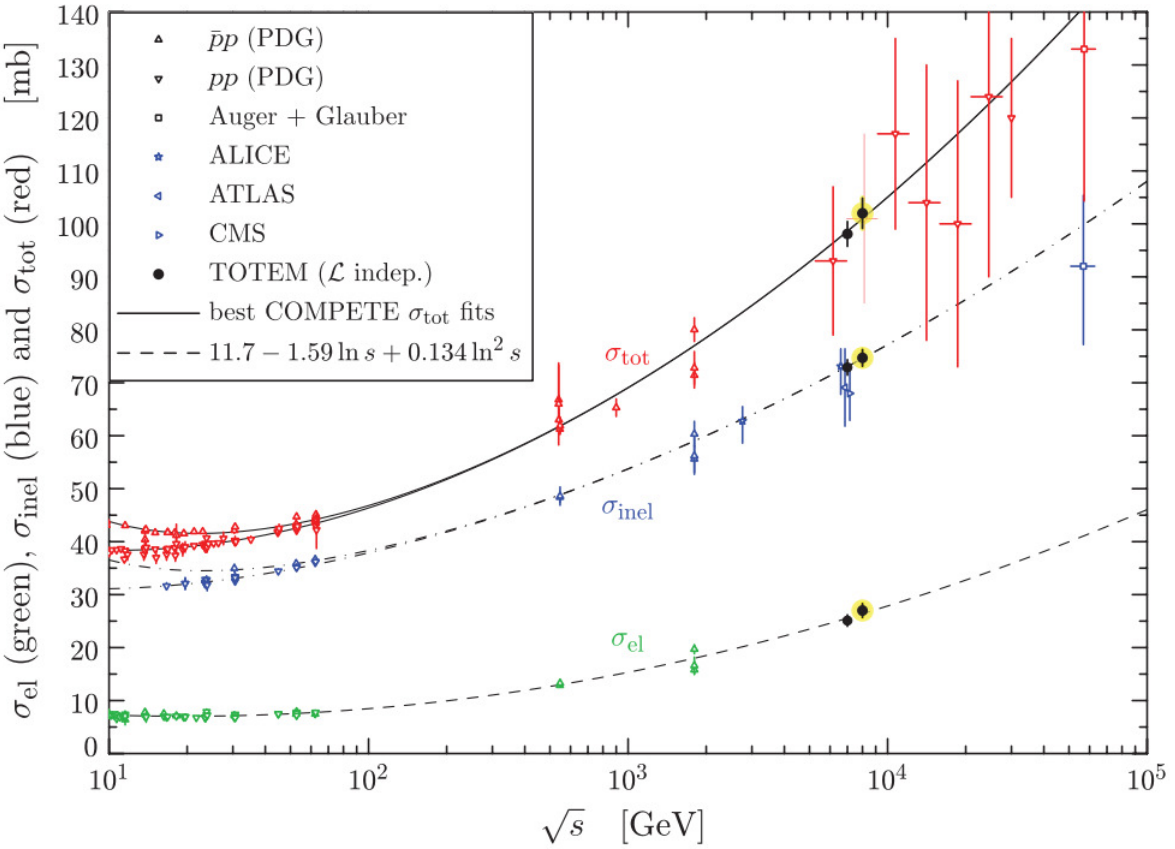


Figure 2.6: Measurement of the total, elastic and inelastic pp cross-sections at a center-of-mass $\sqrt{s} = 8$ TeV by the TOTEM collaboration and comparison with other results [60]. At the LHC energies, the inelastic cross-section dominates largely.

section [61] meaning that most of the pileup interactions are soft QCD processes. In these interactions, low momentum particles (mainly light hadrons) are produced [62]. If a hard QCD interaction occurs during a bunch-crossing, all the particles produced in the soft QCD pileup interactions will superimpose in the event. This has to be dealt with by the analyzers as this adds energy to the event description that does not come from the hard interaction. At the LHC nominal luminosity, about 25 pileup interactions are expected. Not all pileup interactions are identified by the detector. If only very low energy particles are produced, it may not be possible to reconstruct the original spatial position of the interaction. This position is called the primary vertex of the interaction. A display from an actual CMS event with 78 reconstructed vertices is shown in figure 2.7. In figure 2.8, the pileup distribution for the 2016, 2017 and 2018 data taking years is given.

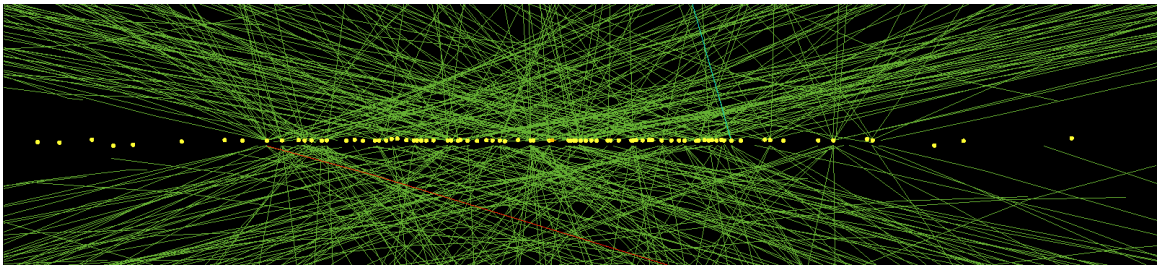


Figure 2.7: Display of a CMS event with 78 reconstructed vertices recorded in a high pileup run in 2012 [63].

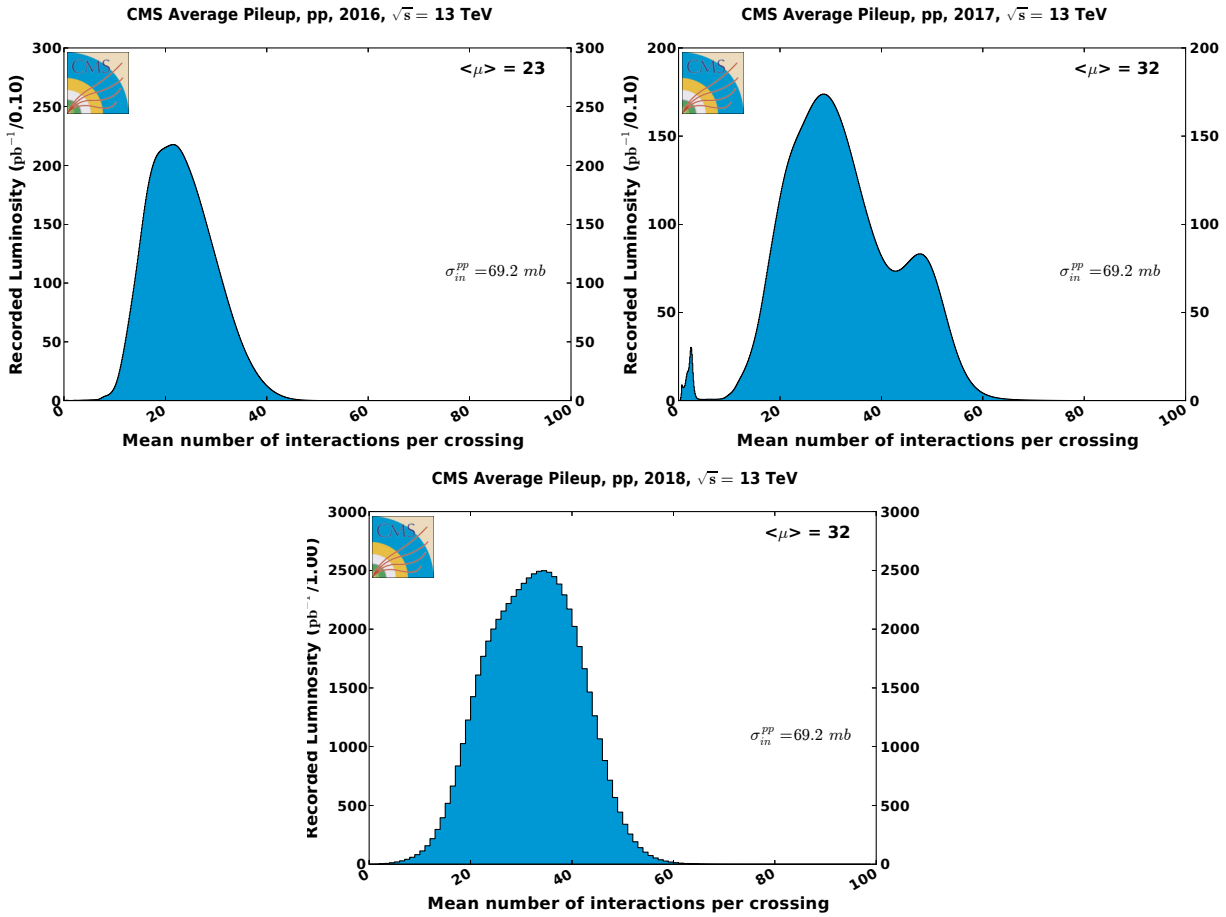


Figure 2.8: Mean number of pileup interactions per bunch-crossing in 2016 (left), 2017 (middle) and 2018 (right) data [64]. σ_{in}^{pp} is the *minimum bias* pp cross-section.

2.5 Luminosity

The instantaneous luminosity $\mathcal{L}$ is a measure of the number of collisions the LHC produces per second. It is a pure machine-dependent variable. Given a process $pp \to X$, it links the rate of observed events dN/dt to its cross-section $\sigma(pp \to X)$:

$$\frac{dN}{dt} = \mathcal{L} \times \sigma(pp \to X) \tag{2.1}$$

It is measured in inverse femtobarns per second $\mathrm{fb}^{-1}\mathrm{s}^{-1}$ or in inverse squared centimeters per second $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, 1 barn being equal to 10^{-24} cm^2. The integrated luminosity L is the integral of the instantaneous luminosity over an interval of time. It is measured in fb^{-1}. The instantaneous luminosity depends on a number of variables as given by:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma}{4\pi\epsilon_n \beta^*} F \tag{2.2}$$

The definition of the variables and their nominal values in 2016 [65, 66] is given here:

- $N_b = 1.1 \times 10^{11}$ is the number of protons per bunch.
- $n_b = 2076$ is the number of bunches per beam.
- $f_{\mathrm{rev}} = 11.245$ kHz is the revolution frequency.
- $\gamma = 6928$ is the Lorentz factor.

- $\epsilon_n = 3.4$ μm is the normalized transverse beam emittance. It describes the beam spread in the position-momentum phase space.
- $\beta^* = 40$ cm is the value of the beta function at the interaction point. The beta function describes the transverse size of the beam.
- $F = 1/\sqrt{1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2}$ is a luminosity reduction factor to cope for the non zero crossing angle between the beams. Beams must cross at a non-zero angle to avoid long-distance (typically elecromagnetic) interactions between the beams before and after the interaction point.
  - $\theta_c = 370$ μrad is the full crossing angle
  - $\sigma_z = 7.8$ cm is the RMS bunch length
  - $\sigma^* = 16.7$ μm is the transverse RMS beam size at the interaction point

Plugging in those values yields an instantaneous luminosity of $8.7{\times}10^{33}$ cm^{-2}s^{-1} or 8.7 nb^{-1}s^{-1} which is about the peak instantaneous luminosity of the LHC in 2016. The LHC operators can play with different parameters to increase this value: the number of bunches per beam, the crossing angle or the β^* are some examples. Peak and integrated luminosity for 2016, 2017 and 2018 are shown in figure 2.9. It can be noticed that not all the delivered luminosity is recorded by CMS. This is due to inefficiencies in the CMS operations while the LHC is in *stable beams* mode. Deadtimes can arise due to subdetectors failing or the data aqcuisition system being overloaded with data and imposing a time buffer. The total integrated luminosity recorded is of 37.8 fb^{-1} in 2016, 45.0 fb^{-1} in 2017 and 63.7 fb^{-1} in 2018 adding up to a total of 147 fb^{-1}. Not all this data will be useable for analysis as part of it needs to be discarded for quality reasons.

2.6 LHC motivations and design

One of the main motivation behind the LHC project was the will to discover the Higgs boson. Figure 2.10 gives the evolution of the Higgs production cross-section in pp collisions at $\sqrt{s} =$ 14 TeV as a function of the boson mass. The mass range in which the Higgs boson could have been looked for was bound from below at 114 GeV by previous accelerator results (see the LEP combined paper [67]) and from above at $\sim$ 650 GeV by theoretical constraints [68]. The largest contribution comes from the gluon-gluon fusion process, with a cross-section of $\sim 10^{-1}$ pb at $M_H = 1$ TeV which is a conservative scenario. Given a branching fraction of $\sim$ 10 % to the ZZ final state (H$\rightarrow$ZZ$\rightarrow 4\mu$ is one of the "golden channels"), about 1000 Higgs events were expected to be produced in CMS for an integrated luminosity of 10^5 pb^{-1} (100 fb^{-1}).

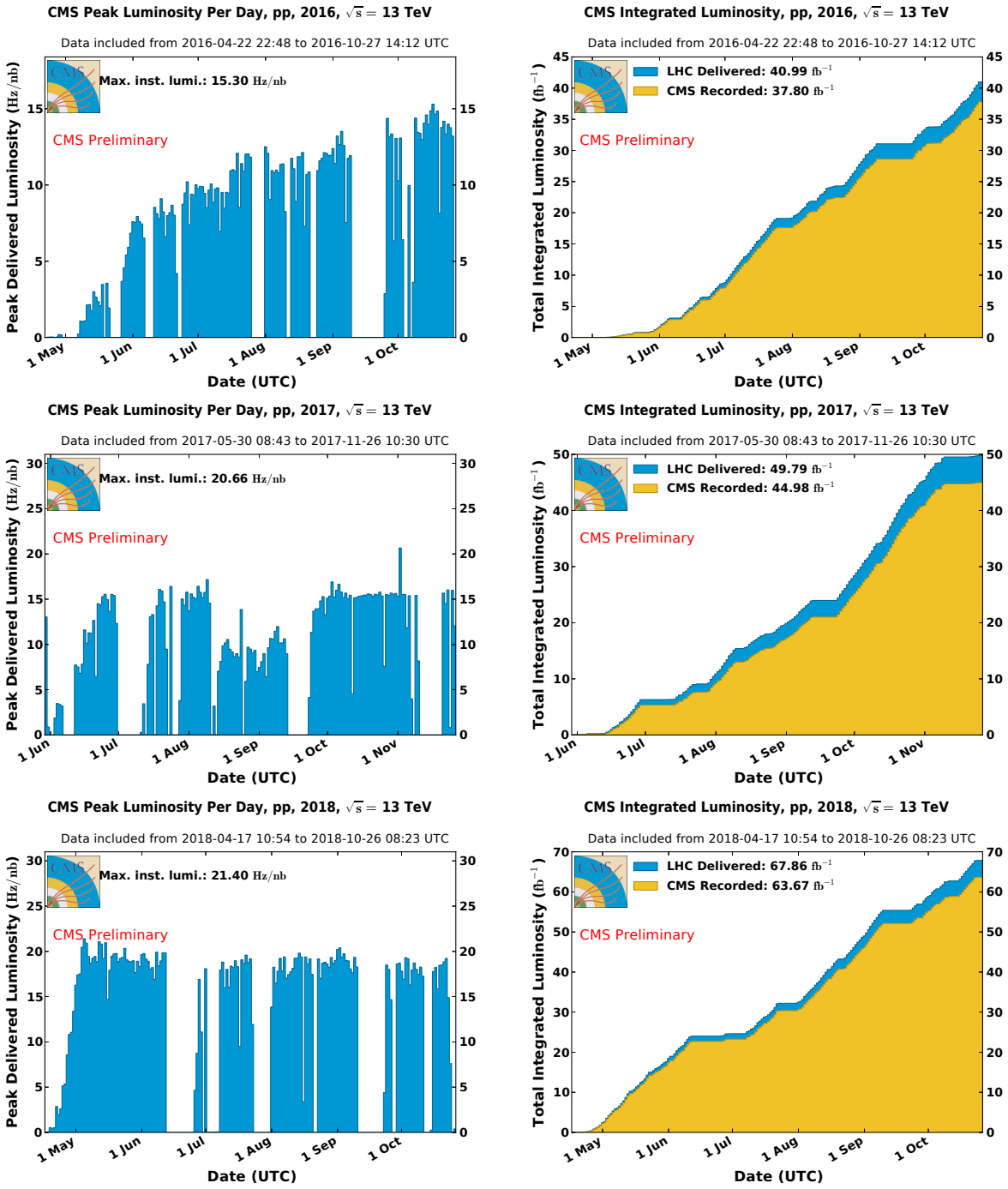


Figure 2.9: Peak instantaneous luminosity delivered per day by the LHC to CMS (left) and integrated luminosity delivered by the LHC and recorded by CMS per day (right) in 2016 (top), 2017 (middle) and 2018 (bottom) [64].

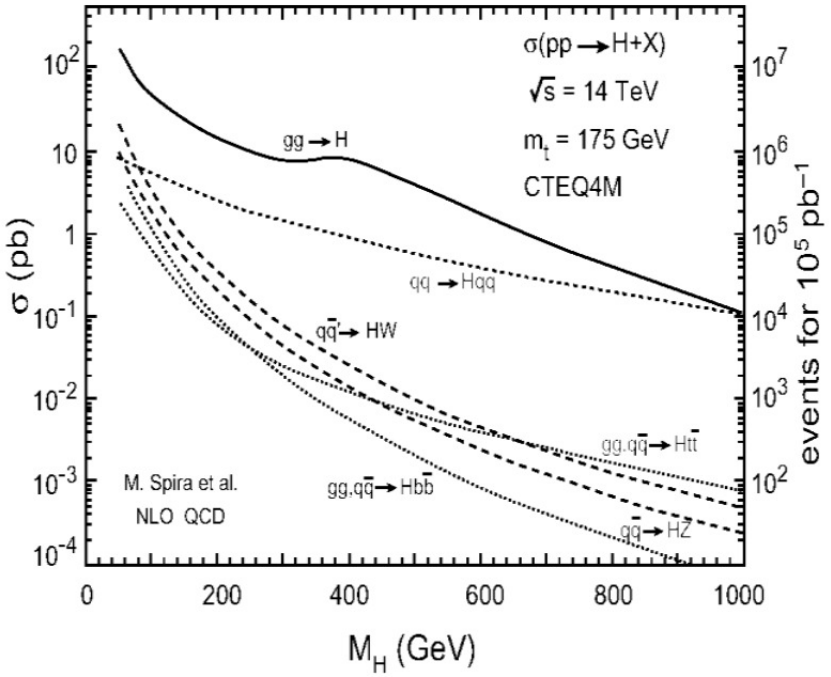


Figure 2.10: Evolution of the Higgs production cross-section as a function of its mass [69].

Chapter conclusion

The Large Hadron Collider was presented. We detailed its timeline since 2010 and the instantaneous luminosities and center-of-mass energies reached. We showed that it produces 6.5 TeV bunched proton beams that collide in four points among which is the CMS detector. The highest peak instantaneous luminosities reached are around 20 Hz/nb. We gave the basics of proton-proton interactions and we showed how inelastic interactions are predominant in this context. We also discussed the distinction between soft and hard QCD interaction that will be used later to understand how the trigger system functions. We also introduced the concept of pileup interactions that are to be taken into account along with the hard interaction of interest. The data analyzed in this thesis were collected during the Run 2 period of the LHC and correspond to an integrated luminosity of 147 fb^{-1}.

Chapter 3

The CMS experiment

Motivated abstract

The *Compact Muon Solenoid*, or CMS, is a multipurpose detector built at the LHC at CERN. Together with the ATLAS detector, it has been used to discover the Higgs boson in 2012. In operation since 2009, it will continue to collect data until the end of the LHC lifetime, i.e. around 2035 (although with major upgrades). CMS is made of several subsystems assembled in layers, each playing a well defined role. Technical features of some of those subsystems have a direct impact on the rest of the work and are worth underlining. In this chapter, we try to show the link between those features and the physics analyses.

The first analysis presented in this thesis is the search for Dark Matter particles in the monojet channel. It requires a good understanding of jet physics and of the missing transverse energy. Jets and missing transverse energy both are reconstructed using all CMS subsystems. Therefore, a general view of the whole detector is needed. We give more details on the jet reconstruction algorithms and performances in the coming sections.
The second and main analysis presented in this thesis is the search for fractionally charged particles. A critical feature of those particles is the fact that they deposit less energy via ionization processes when they interact with matter. The ionization energy is measured by the sensors of the CMS tracking system and will be used to discriminate background-like tracks from signal candidate tracks. A deep understanding of the physics behind the ionization energy measurement is therefore needed. The corresponding relevant information is presented in the tracking section of this chapter. Fractionally charged particles are reconstructed as muons in CMS. Detailed information on the performance of the muon system and the muon reconstruction algorithm is therefore also given.

We start this chapter with the geometry of the detector. We then go through each subsystem, describing the detection principle, the geometry and the performances with a level of detail relevant to the analyses presented afterwards. The hardware being presented, we give the details of the phyics object reconstruction, i.e. for instance how is an electron or a muon defined in CMS. We finish with the presentation of the CMS trigger system that is responsible for the selection of apriori interesting events. All details regarding the detector can be found in reference [70].

3.1 The CMS geometry

CMS has a standardized reference system. The cartesian system is centered on the center of the detector, the x axis horizontally pointing towards the center of the LHC, the y axis vertically upwards and the z axis along the beam axis according to the right-handed convention (i.e. pointing north-west). The *radius* is defined in the xy plane as $r = \sqrt{x^2 + y^2}$. The cartesian system is

often more conveniently replaced by the ϕ, η and r coordinates. The azimuthal angle ϕ is defined in the xy plane from the x axis while the pseudorapidity η is built from the 3-dimensional angle θ computed from the z axis:

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right). \tag{3.1}$$

We will often make use of the ΔR variable which is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. It is used to evaluate how close two objects (defined by their 4-vectors) are to each other. Figure 3.1 sketches the reference system of CMS. CMS is embedded in a 3.8 T magnetic field pointing in the z direction, therefore curving the tracks of charged particles in the xy plane.

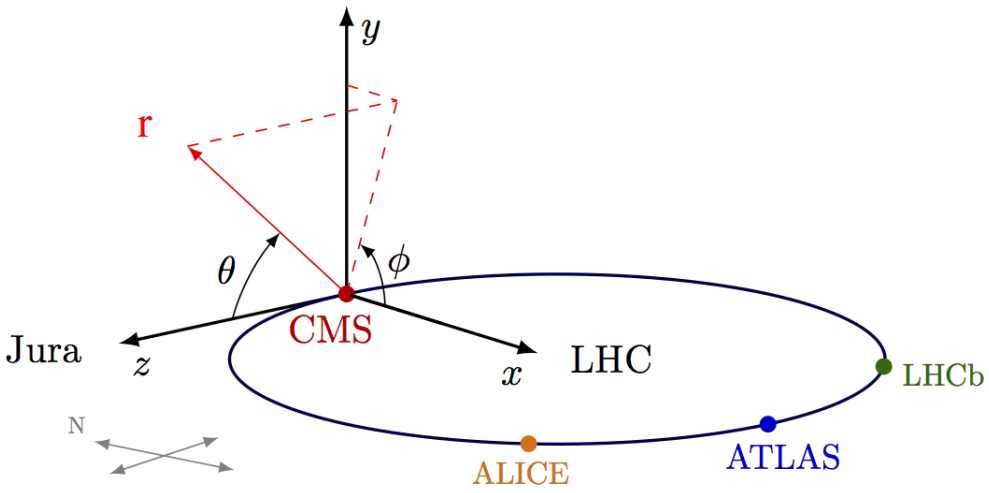


Figure 3.1: CMS coordinate system.

3.2 The detector subsystems

CMS is made out of several subdetectors (or subsystems), each fulfilling a well defined task. Those subdetectors are arranged in cylindrical layers, from the tracker at the very center to the muon system in the outmost part. Each detector is composed of a central part, the *barrel*, closed by forward and backward *endcaps*. In this chapter, we will describe all subdetectors, their roles and performances in a level of detail that is relevant to this work. In figure 3.2, a global view of the detector and its subsystems is shown.

3.2.1 The tracking system

A tracking system is a detector that carries out a series of non-destructive measurements, or hits, along the path of a crossing particle. Those measurements are done through the ionization of a medium, solid or gas, by the particle. The CMS tracking system, or simply *tracker*, is responsible for the tracking of the charged particles path in the innermost part of the detector. It is located at the very center of the detector in order to determine precisely the position of the primary and secondary vertices, i.e. where the particles emerge from. Being embedded in a solenoid whose magnetic field curves the particles' trajectory, it gives information on the transverse momentum and the charge of the particles. It also measures the stopping power dE/dx for each hit. This is defined as the energy loss per unit length. Together with the calorimeters, the tracker participates to the reconstruction of electrons and charged hadrons. Combined with the muon system, it is used to reconstruct the muons.

The crucial feature of the tracker is that it is the closest subdetector to the collision point. As a consequence, it must meet the following requirements:

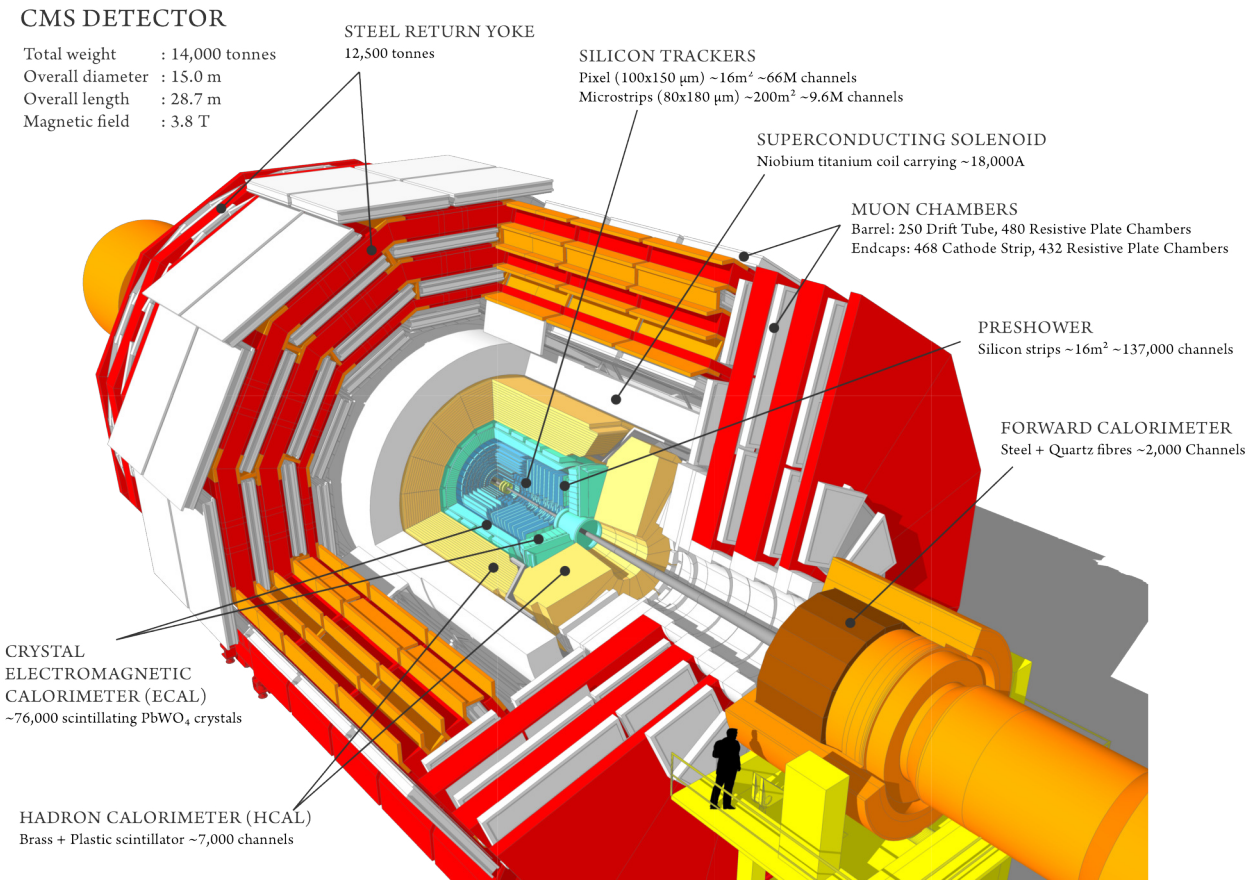


Figure 3.2: CMS and its subdetectors [71].

- high granularity, i.e. high spatial resolution,
- fast response, i.e. small readout time,
- the ability to handle high radiation doses, and
- the ability to handle high track density, i.e. a low track occupancy.

These requirements led to the need of two dedicated technologies to handle the high particle flux from the pp collisions: the pixel detector, closer to the interaction point, and the strip detector that surrounds it. To meet such requirements, the detector electronics needs high powers, hence a cooling system must be put in place. Cooling also reduces leakage current in the silicon sensors. The full tracker system is cooled at -20 °C by liquid C_6F_{14}.

As indicated by its name, the pixel detector is made of smaller sensors that guarantee the necessary high granularity in the innermost part of the tracking system. It should be clarified that in the CMS jargon, one often refers to the strip detector as the *tracker*. However, we will rigourously use *pixel* and *strip* for the innermost and outermost subsystems, respectively. In figure 3.3 the arrangement of the subparts of the tracker and the corresponding η ranges are sketched. The CMS tracker covers an η range of [-2.5,2.5]. The pixel detector was upgraded in 2017. We will first present the 2016 version and detail the changes at the end of the pixel section. The 2016 pixel detector is formed of 3 barrel layers (*Pixel Barrel*, PXB or BPix) closed by 2 endcaps disks (*Pixel Forward*, PXF or FPix) while the strip detector is composed of 4 parts: the *Tracker Inner Barrel* (TIB), the *Tracker Inner Disks* (TID), the *Tracker Outer Barrel* (TOB) and the *Tracker Endcaps* (TEC). One notices here the use of *tracker* to designate the strip detector.

The four cited requirements that the tracker must fulfill have led to a tracker design fully based on solid state silicon. Before giving more details on how this technology is effectively

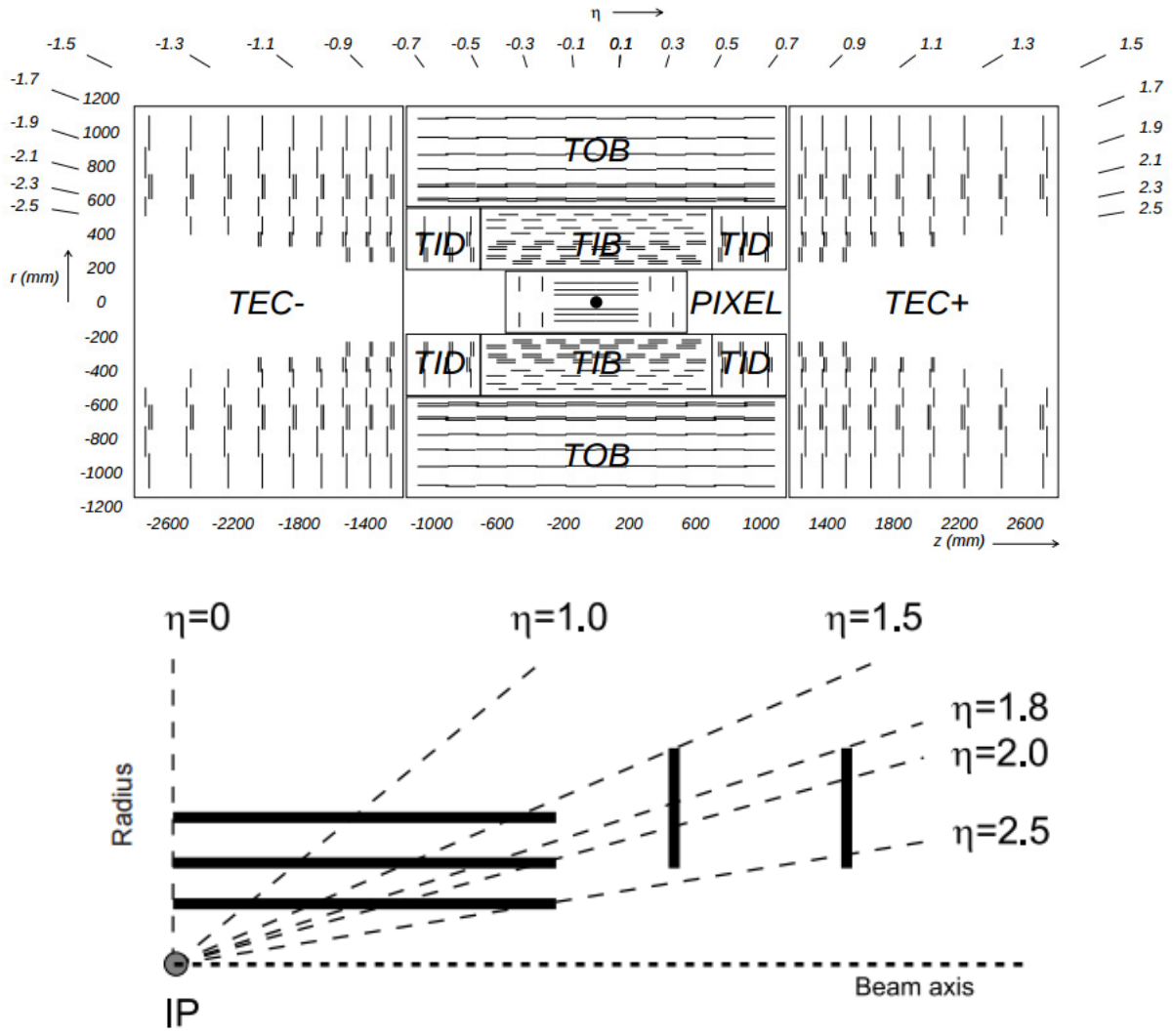
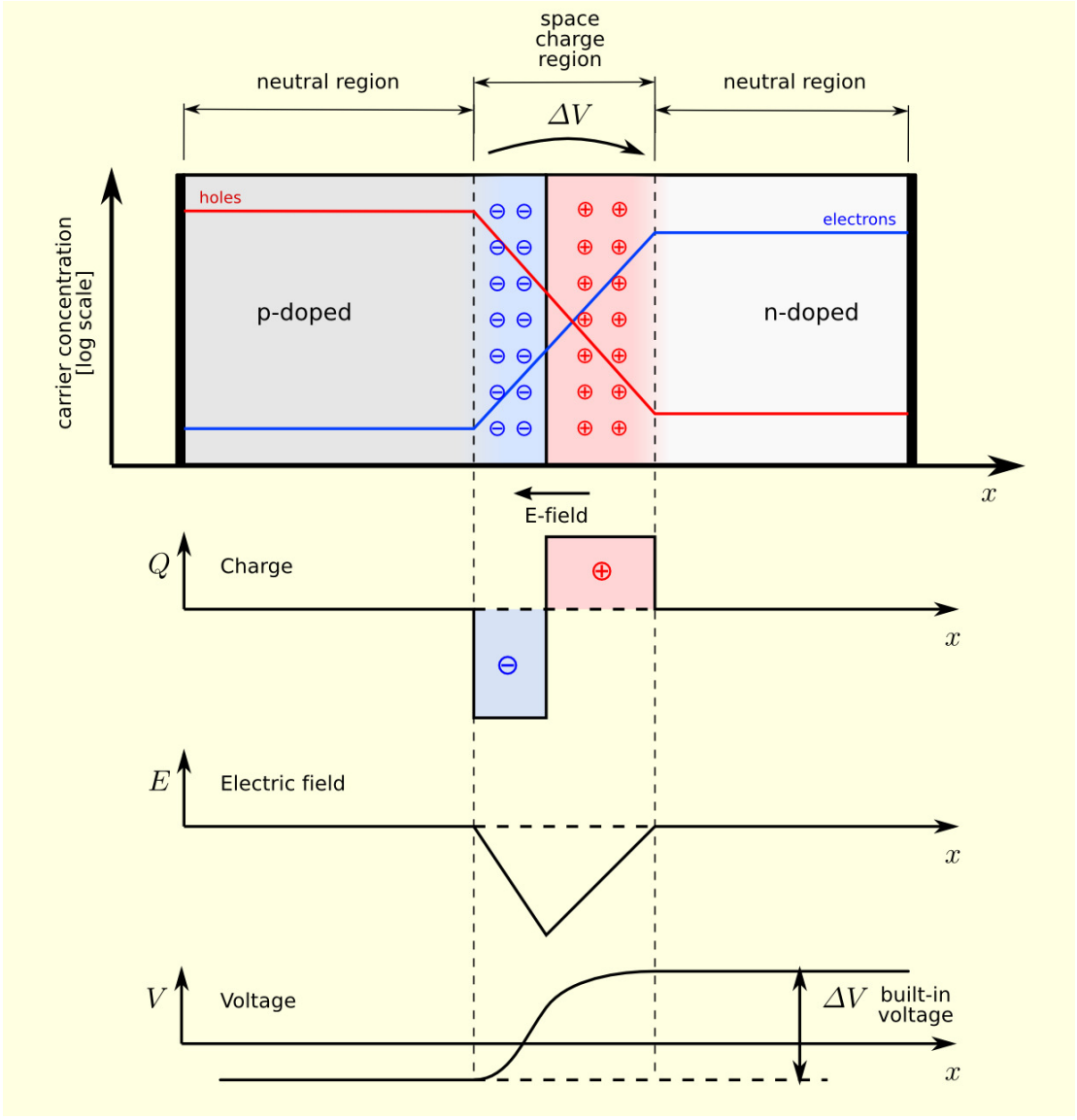


Figure 3.3: Schematic view of the tracking system (top) and a zoom on the 2016 pixel detector (bottom) [70].

applied in the pixel and the strip detectors, let us briefly explain the detection mechanism. The basic idea of a sensor in a tracking system is to have an active volume immersed in an electric field. When a charged particle traverses the medium it ionizes the atoms of the medium, creating free charges. Under the action of the external field, those charges drift towards electrodes to be collected and a signal is read out. In silicon however, the number of free carriers created by thermal excitations at room temperature is much higher (typically 4 orders of magnitude) than the number of free carriers created by the passage of a minimum ionizing particle. For a given material, a minimum ionizing particle (MIP) is a particle that has a momentum corresponding to the minimum plateau of the Bethe distribution: it will deposit a minimal amount of energy by ionization. The solution to this apparent issue is to dope the material with *donors* (elements with an extra electron with respect to silicon) or *acceptors* (elements with one electron less, or an extra "hole" with respect to silicon). Typical silicon sensors are then made of *p-n junctions*, combining a p-doped silicon bulk (with extra acceptors) with a n-doped silicon bulk (with extra donors). In such a device, electrons migrate to the p-doped region and holes move to the n-doped region, resulting in a depleted region in the middle. This region has no free carriers. Figure 3.4 illustrates these principles. On the left a p-n junction is sketched with the corresponding charge, electric field and voltage. On the right, a typical configuration for a silicon p-n junction is shown. The depletion zone is much larger to allow for a bigger active volume.



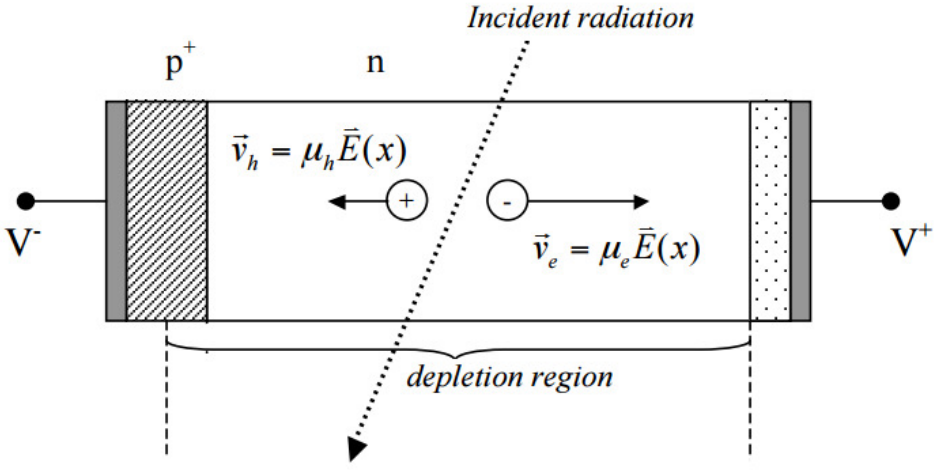


Figure 3.4: p-n junction (top) and typical configuration of a silicon-based sensor (bottom) [72]. On the right figure, $V^{\pm}$ is the voltage, $\vec{E}$ the corresponding electric field, $\vec{v}_h$ is the drift velocity for the holes, $\vec{v}_e$ is the drift velocity for the electrons and μ_h and μ_e are the corresponding mobilities.

The pixel detector

Each crossing of the proton bunches produces on average 1000 charged particles (at the design luminosity of 10^{34} cm^{-2}s^{-1}), leading to a significant track density below a 10 cm radius from the

collision point. In this region, in order to be able to spatially separate the passage of two closeby tracks, the active detection medium of the tracker is split into 100×150 μm^2 cells. This is called the pixel detector and it forms the core of the tracker. The small size of the cells quantifies the high granularity of the detector. This detector is made of 48 and 18 millions such silicon sensors that cover an area of 0.78 and 0.28 m^2 in the barrel and in the endcaps, respectively. The barrel pixel detector is formed of three layers at a radius of 4.4, 7.3 and 10.2 cm, while each endcap is made of two disks located at $z = \pm 34.5$ and ± 46.5 cm. A simple calculation shows that, given this geometry, the occupancy in the first barrel layer is of the order of 10^{-4} ($= 10^3$ tracks/10^7 pixels). This number quantifies the requirement of low occupancy.

The sensor readout is performed by an ASIC readout chip (ROC) that processes the signal of 52×80 cells. It amplifies and buffers the signal, communicates with higher level electronics and checks compatibility with L1 trigger information, i.e. only keeps the hit data if a L1 seed has triggered (see section 3.4). However, the most critical characteristics of the ROC (in the context of this work) is that it applies zero suppression on the readout signal. This means that only signals above a certain threshold are saved, this threshold being pixel-adjustable. Zero suppression helps rejecting noise arising from leakage current and capacitance effects. The former is due to thermal effects creating electron-hole pairs in the depleted zone while the latter is a consequence of the junction geometry and is the main source of noise. Another benefit of zero suppression is to reduce the number of readout hits per event and to allow the amount of data to stay reasonable. Let us add that, on top of zero suppression, the noise from capacitance effects is also reduced with the pixelization of the tracker detector, i.e the division of the active detection surface.

The main research project of this thesis is the search for low charge particles, which deposit less energy by ionization in matter than unit charge particles. Zero suppression is therefore something we have to take into account when working with signal simulations to make sure hits with a charge lower than the threshold are not counted. This will be developed in full details in chapter 6.

A hit is the result of multiple pixels clustering in order to collect all the ionization energy left by a particle in a given layer. A track is made of a collection of hits. The tracking algorithm allows the track to have missing hits along its path to cope for detector failure. Indeed if a cluster is missing (due to a damaged sensor, noisy channel or dead readout for instance), the track might still be reconstructed using information from further layers. The track is therefore composed of valid and missing hits. According to [73], the hit efficiency is "the probability to find a cluster in a given silicon sensor that has been traversed by a charged particle". It is defined by equation (3.2) where $N_{\text{missing hits with cluster}}$ is the number of missing hits found in a 500 μm radius from a cluster and $N_{\text{missing hits}} = N_{\text{missing hits with cluster}} + N_{\text{missing hits no cluster}}$ [73]. A hit with cluster is included in the numerator as the neighbouring cluster may be the result of ionization from the particle to which the track belongs but was not included in the tracking algorithm. This definition of the hit efficiency has been designed to estimate the pure sensor efficiency independently of the tracking efficiency. The number of hits appearing in this equation are computed for all tracks passing through the considered layer, with $p_T > 1$ GeV and a minimum of 11 hits. The two other pixel layers should contain a valid hit.

$$\text{Hit efficiency} = \frac{N_{\text{valid hits}} + N_{\text{missing hits with cluster}}}{N_{\text{valid hits}} + N_{\text{missing hits}}} \tag{3.2}$$

Figure 3.5 gives the average hit efficiency as a function of the instantaneous luminosity and pileup as measured in 2016 data. One should notice the strong inefficiency of the PXB layer 1 (PXBL1) hits at high luminosity or high pileup. Hits in this layer are the closest to the beam and so are the most affected by high radiation and large tracks density. The very large number of hits overloads the readout electronics, causing what is referred to as a "dynamic inefficiency". Another (smaller) proportion of the missed hits is made of those that dropped below the zero suppression threshold because of radiation damage. The hit resolution in the $r - \phi$ and z axes is shown in figure 3.6 for 2015 data. In the $r - \phi$ plane, the resolution is better (i.e. smaller) due to the track curvature in the magnetic field.

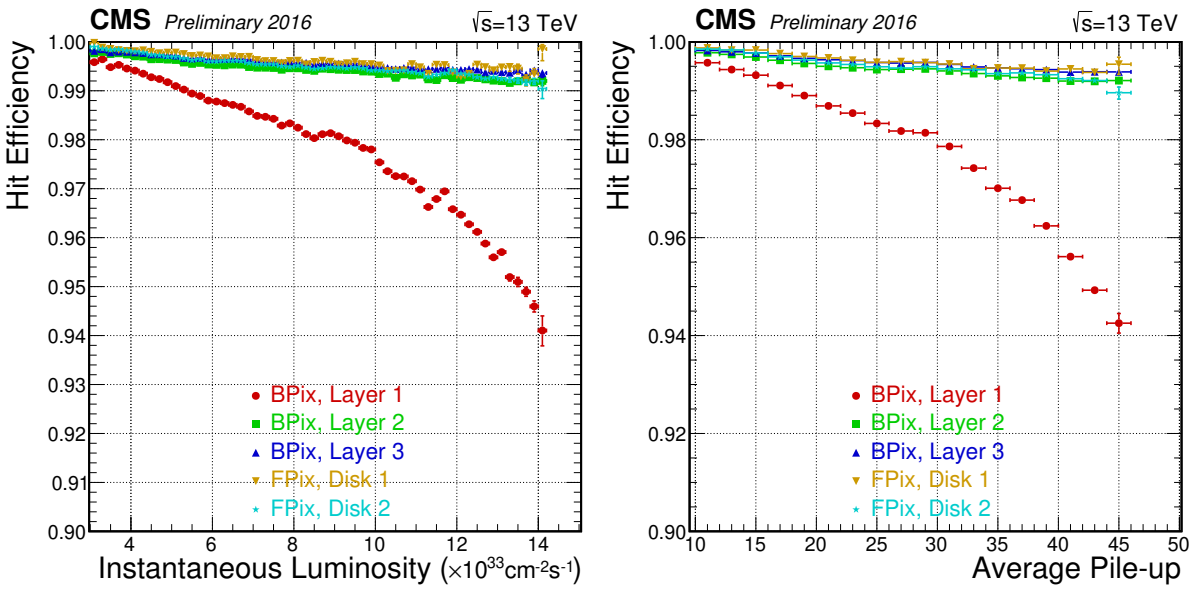


Figure 3.5: Average hit efficiency as a function of the instantaneous luminosity (left) and pileup (right) in 2016 data [74].

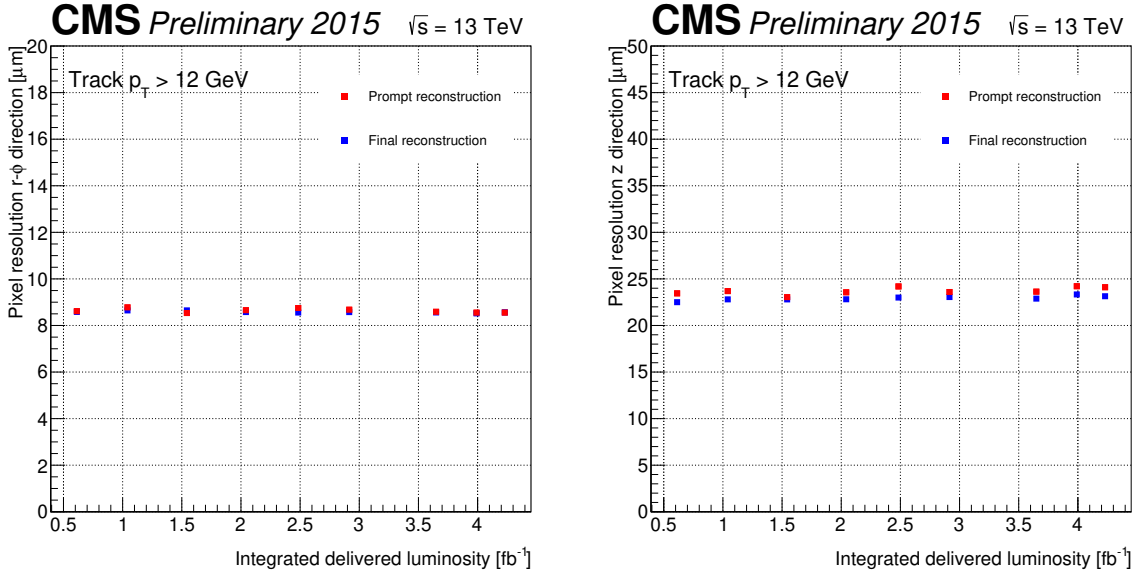


Figure 3.6: Pixel resolution as a function of the instantaneous luminosity in the $r-\phi$ (left) and z (right) directions for 2015 data [74]. Two reconstructions are compared: prompt (data is reconstructed right after recording) and final (more data is accumulated to allow for a better calibration and more precise correction factors).

In 2017, a new pixel detector has been installed inside CMS. The major change is the addition of a fourth layer closer to the beam with respect to the first layer of the 2016 detector. The new pixel is made of 80 million and 45 million pixels in the barrel and in the endcaps, respectively. Figure 3.7 compares the old versus the new pixel geometries. The hit efficiency for this new detector is shown in figure 3.8 with 2018 data. It is observed to be higher than in 2016.

The strip detector

Surrounding the pixel detector is the strip detector. Also made of silicon sensors, the technology used is nonetheless different. Further away from the beam, the need in granularity is lower. This led to the design of a detector made of silicon strips. The strip detector is itself split in several subsystems as described in the introduction and depicted in figure 3.3. The TIB extends the 10.2

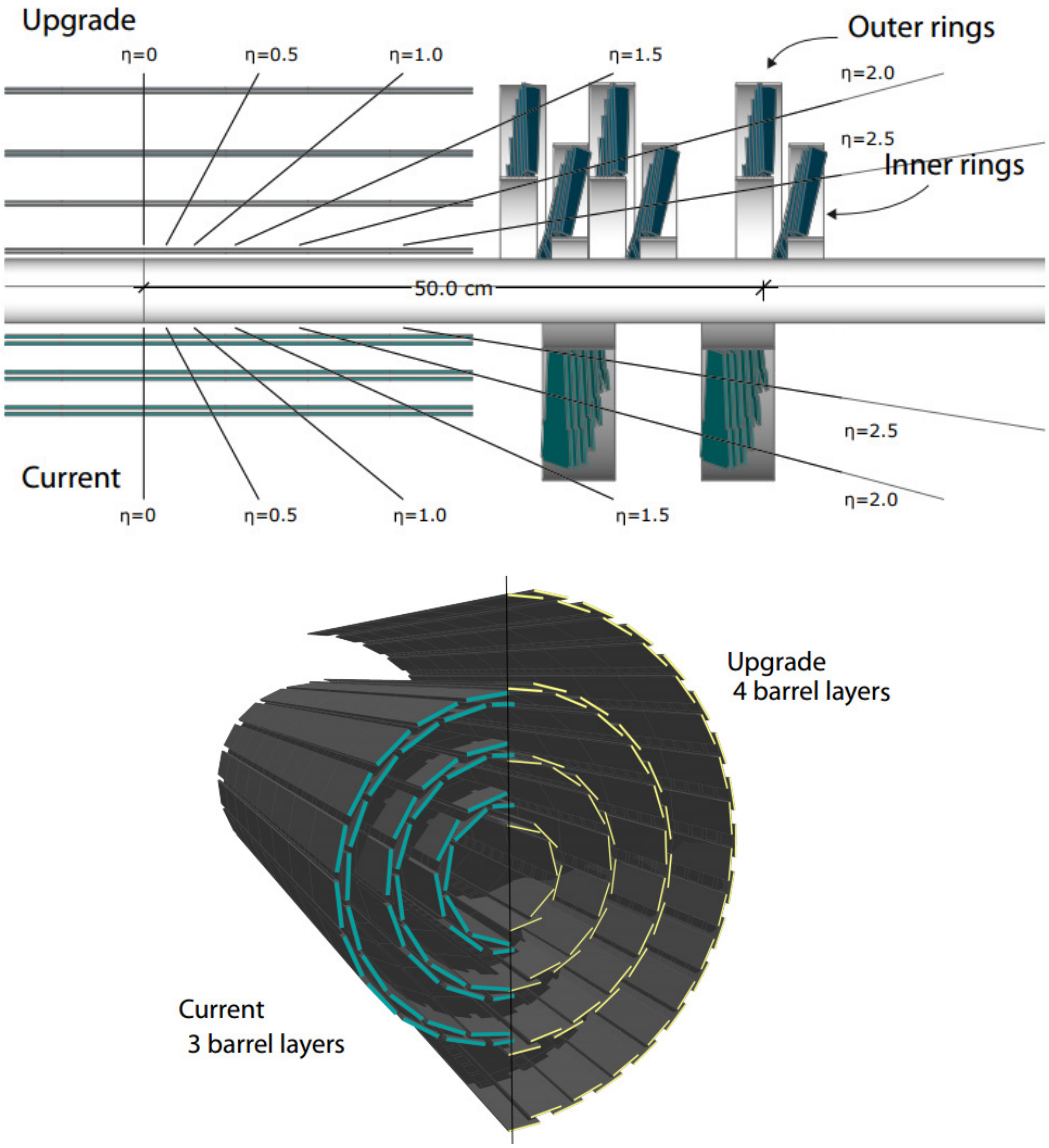


Figure 3.7: Top: sketch of the old (bottom) and the new (top) pixel detector. Bottom: sketch of the old (left) and the new (right) pixel detector. The position of the layers has changed, one layer is added in the barrel and one disk is added in the endcaps [75].

cm radius of the pixel detector range up to 55 cm. The full tracker coverage is then obtained with the TOB up to a radius of 1.2 m. The TID and the TEC provide forward coverage up to $|\eta|$ <2.5. The strip detector covers an area of 198 m^2.

The readout, amplification and shaping of the silicon strips signal is done by an integrated circuit called the APV25 chip. This chip reads the signal from 128 channels each being independently pre-amplified, shaped and sampled at the LHC frequency (40 MHz). Sensors in the strip detector are made of either 512 (4×128) or 768 (6×128) strips to match the input multiplicity of the readout chips. The strip detector is made of 9.3 million strips. The APV sends the signals via analog optical links to the Front End Driver (FED). Due to the lower number of input channels (66 million in the pixel versus 9 million in the strip detector), one can afford the zero suppression threshold to be applied in the FED and not in the ROC itself as it is the case for the pixels.

In early 2016 data, an inefficiency was observed in the track reconstruction due to missing hits. This was eventually understood as a saturation effect of the APV25 chip pre-amplifier consequent to the higher peak luminosity, the smaller inter-bunch crossing time and the change in detector temperature with respect to Run 1 (from 4 to -20 °C). Indeed, the behaviour of

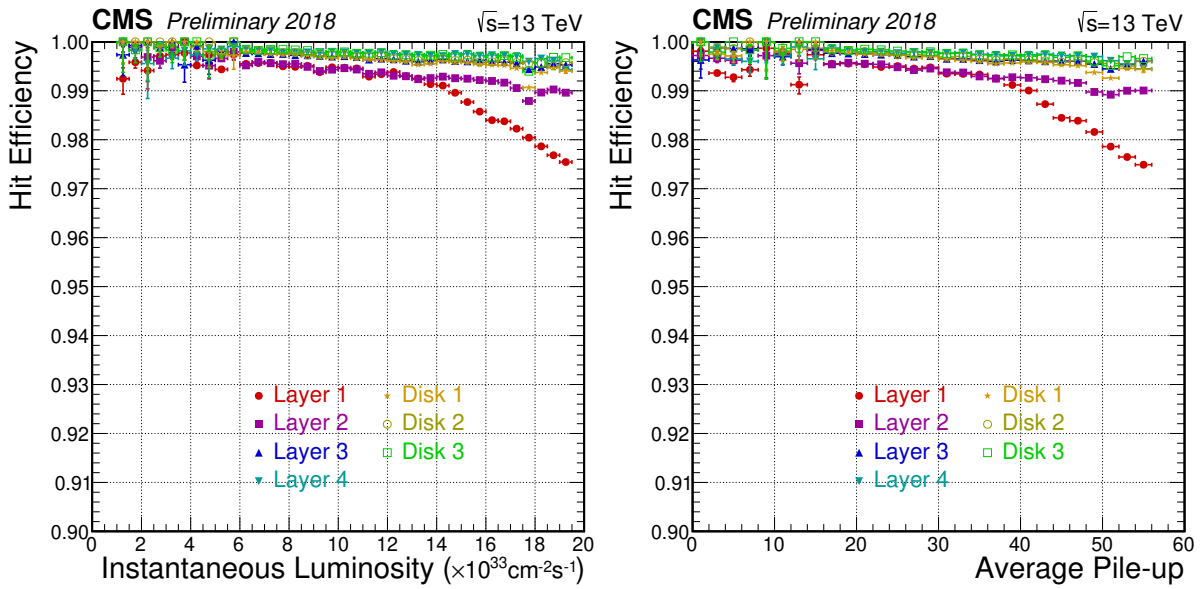


Figure 3.8: Average hit efficiency as a function of the instanteneous luminosity (left) and pileup (right) in 2018 data [74].

electronic elements are temperature-dependent but the chip settings were not changed.

Schematically, the output of the preamplifier is given by figure 3.9 (left), scaling with the integrated charge from 1 to 5 MIP going through the sensor. The larger the charge deposit, the higher the output voltage. After preamplification, the signal is processed by the shaper (figure 3.9 (right)) which is responsible for noise reduction and pileup mitigation. The larger the output of the preamplifier, the more negative the output of the shaper. The next step is to read the shaper output. In the CMS tracker electronics, this is done in deconvolution mode where the output is computed using a weighted sum of three samples (figure 3.10 left). The shaper electronic circuit is called a CR (differentiator)-RC (integrator) circuit, according to its composition in capacitance C and resitive R elements. Due to the further differentiation role of the deconvolution procedure the full circuit is of CR-RC-CR type, leading to what is called a bipolar pulse shaping (figure 3.10 (right)), i.e. a pulse with a negative undershoot of the voltage before settling back to the baseline voltage. When the preamplifier saturates, the output voltage of the preamplifier may not have enough time to discharge between two input signals and so would not return to its baseline voltage. If this happens, distinct pulses will overlap after deconvolution. Due to the undershoot feature of the APV25, the peak of the next pulse will be underestimated by the value of the undershoot, leading to an underestimation of the collected charge in the sensor.

This effect is present in all of the strips subsystems, though not with the same intensity. This is quantified on the left plot of figure 3.11. From late 2016 and up to 2018, this effect was mitigated after changes in the drain speed of the APV pre-amplifier were made. The efficiency is given before (empty circles) and after (red circles) the correction on the APV25 gain. The hit efficiency as a function of the instantaneous luminosity is illustrated in figure 3.11 (right) for the TOB layer 1. This very different behaviour in the hits collected charge during early 2016 and the rest of the data will force us to treat those two periods differently in the analysis presented in chapter 6.

Radiation damage in silicon-based detectors

Silicon crystals submitted to high hadron fluences see their properties change over time. This is due to non-ionizing interactions of the charged and neutral hadrons with the atomic nuclei of the crystal elements. The hadrons traversing the material can kick off atomic nuclei from their position in the crystal resulting in defects in the crystal periodic structure. Examples of such

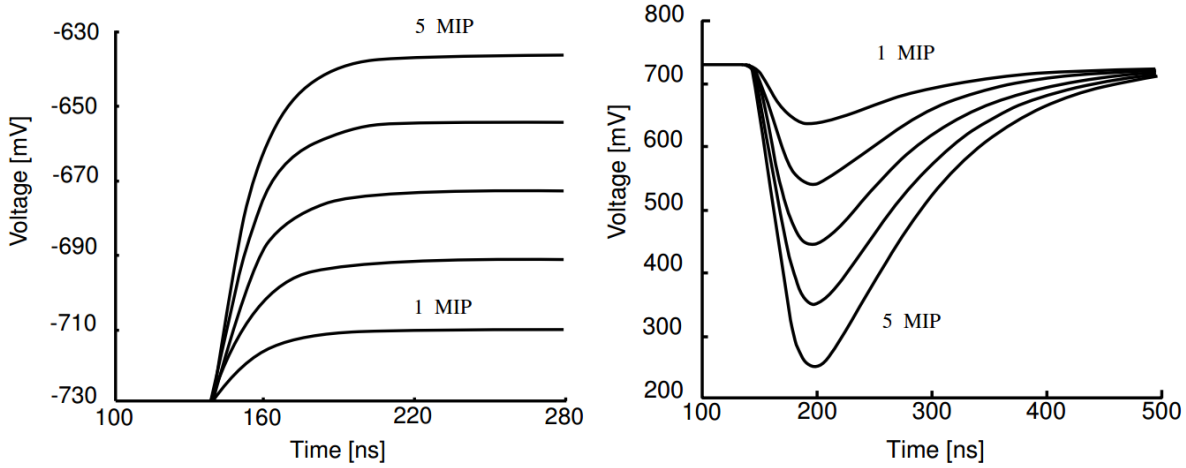


Figure 3.9: Output of the APV25 preamplifier for increasing collected charge (left) and corresponding output of the shaper (right) (adapted from [76]).

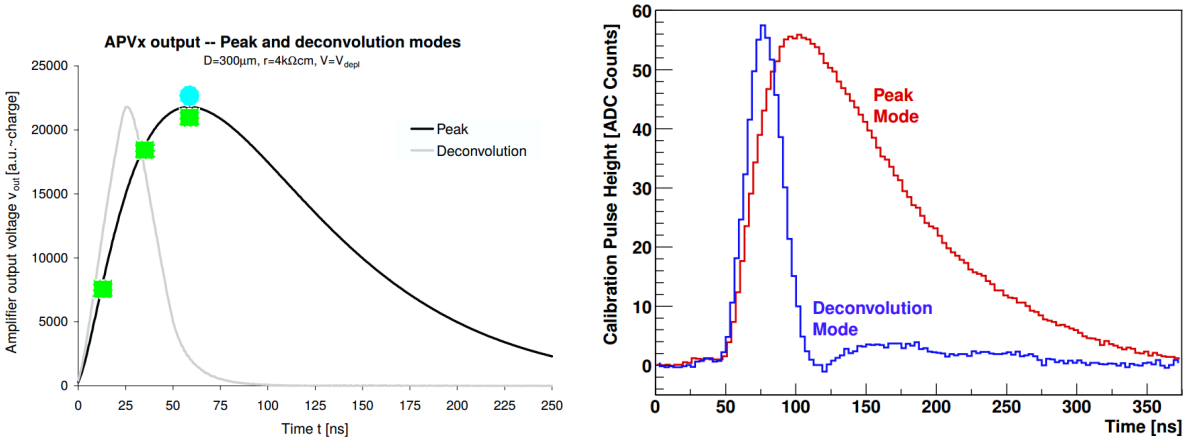


Figure 3.10: Peak and deconvolution sampling modes [77] (left) and realistic output of the deconvolution procedure [76] (right).

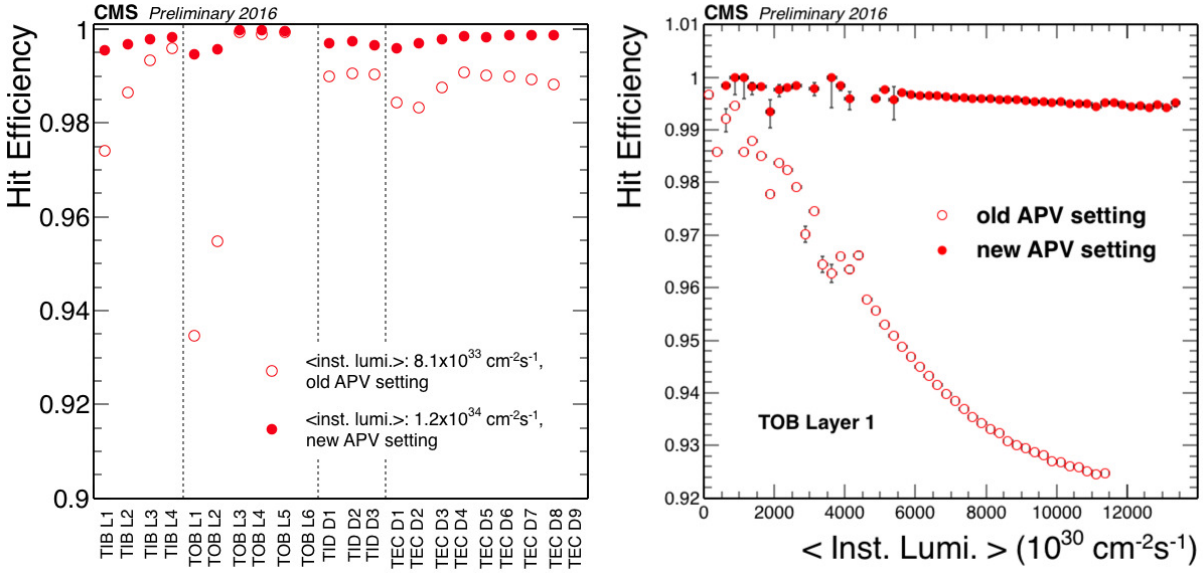


Figure 3.11: Hit efficiency as a function of the instantaneous luminosity (left) and inclusive for all layers of the strip detector (right) [74].

defects are depicted in figure 3.12.

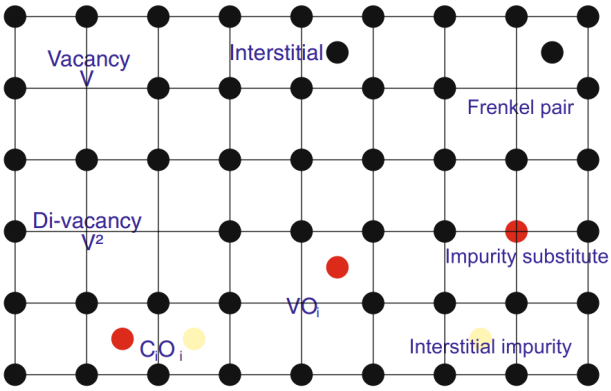


Figure 3.12: Sketch of possible defects in a silicon crystal [78]. A vancacy is a missing atom on a lattice site and the opposite is an interstitial site, i.e. is an extra off-lattice site. A Frenkel pair is created when an atom on a lattice site is kicked off to an interstitial site. Impurities are atoms of a different nature than the crystal bulk elements: C stands for Carbon, O for Oxygen. An impurity can substitute an atom of the bulk at its lattice site.

Scientific literature on radiation damage lists four main consequences of the defects on the properties of silicon sensors [78, 79]:

- Stronger leakage currents. In an ideal pn junction, no current should flow in the absence of the passage of an ionizing particle. However, leakage currents are observed in real devices.
- Modification of the depletion voltage.
- Charge trapping.
- Donor and acceptor removal.

The two last effects have direct implications on our physics search as it reduces the charge collected by the readout system. Charge trapping happens when impurities add energy levels with de-trapping time smaller than the peaking time of the signal read out by the chip. This is sketched in figure 3.13. The charge trapped during this time is "lost", meaning it is not collected by the electronics readout.

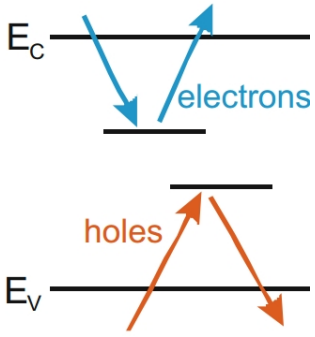


Figure 3.13: Sketch of the charge trapping mechanism. New levels added between the valence (E_V) and the conduction (E_C) bands by the presence of defects in the bulk can trap the charge during the integration time of the readout system. Adapted from [78].

The second important effect when studying the impact of radiations on silicon sensors is the phenomenon of donor and acceptor removal. The observation is that a doped silicon junction sees its concentration of donors and acceptors diminishing when irradiated. Figure 3.14 illustrates this effect by showing the evolution of the impurity concentration with neutron fluence in silicon

wafers for different resistivities. The physical phenomenon behind this change is the moving of the donor/acceptor from a substitutional site, where it acts a shallow donor/acceptor with an energy level close to the conduction/valence band, to an interstitial site, where it can no longer play its dopant role.

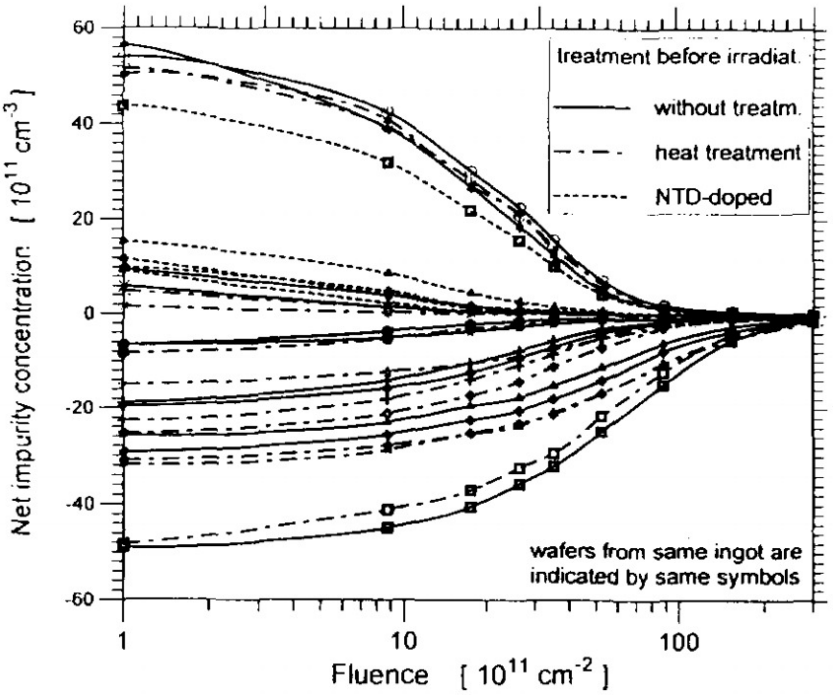


Figure 3.14: Evolution of the impurity concentration with neutron fluence in silicon wafers for different resistivity. Positive means n-type while negative stands for p-type [79]. The impurity concentration diminishes with the fluence, meaning that donors and acceptors are removed. Wafers have been submitted to different treatments that we don't detail here.

The fact that not all the ionization energy is collected and readout in radiation damaged sensors has a direct impact on the analysis presented in chapter 6. In this analysis, we use the stopping power dE/dx as a discriminating variable between the signal we are looking for and the background from muon tracks. The stopping power is obtained by dividing the collected ionization charge by the pathlength of the particle at a given hit. In a radiation damaged sensor, this measurement is biased and this must be understood and corrected for.

3.2.2 The electromagnetic calorimeter

A calorimeter is a destructive detector that measures the energy of a particle. The electromagnetic calorimeter, or ECAL, is responsible for the energy measurement of light electromagnetically-interacting objects namely the electron (and its antiparticle) and the photon. While photons are reconstructed using ECAL-only information, the combination of the ECAL with tracking information allows for electron and positron reconstruction. The energy of a jet is carried by hadrons, electrons and photons which may all interact in the ECAL. Therefore, it also enters the reconstruction of jets.

Describing the physics of electromagnetic calorimetry requires the introduction of electromagnetic *showers*. An electromagnetic shower forms when a high energy electron or photon enters a material. Due to electromagnetic interactions with the atoms and electrons of the medium, the primary particle generates secondary particles of lower energy and this process repeats itself until all the initial energy is absorbed in the material. At energies of interest here (i.e. above the GeV scale), the electron loses its energy by bremsstrahlung, i.e. by radiation of photons (see figure 3.15 left). In the vicinity of the large electric field of the atomic nuclei, a high-energy

photon entering matter will generate electron-positron pairs (see figure 3.15 right). Since electrons produce photons and photons produce electrons (and positrons), the development of the electromagnetic shower is exactly the same whathever the initial particle. Electron-induced and photon-induced electromagnetic showers show the exact same behaviour. The radiation length X_0 of the shower is defined differently for electrons/positrons and photons. For electrons and positrons, it is defined as "the mean distance over which a high-energy electron/positron loses all but 1/e of its energy by bremsstrahlungth" while for photons it is defined as "7/9 of the mean free path for pair production by a high-energy photon" [29]. The Molière radius R_M is defined as containing on average 90 % of the tranverse development of the electromagnetic shower [29]. They give the characteristic scales of the electromagnetic shower.

The ECAL is located radially just after the tracker. The ECAL Barrel (EB) starts at a radius of 1.29 m and covers the pseudorapidity range $|\eta|$ <1.479. The ECAL Endcaps (EE+ and EE-) start at a longitudinal position $|z| = 3.15$ m and extend the coverage up to $|\eta|$ =3. The ECAL Preshower (ES) is placed in front of the EE. Its role is to initiate the electromagnetic shower early on to help the distinction between high-energy photons and neutral pions. Figure 3.16 shows the ECAL geometry.

Similarly to the tracker requirements, the ECAL must have a high granularity, a fast response and stand high radiation doses. As pointed out in [70], the requirement on the spatial resolution was partly motivated by the possibility to observe a Higgs boson decay in the diphoton channel and the subsequent need to distinguish one photon from the other and to have a good diphoton mass resolution. Those requirements are met with an homogeneous calorimeter made of lead tungstate ($PbWO_4$) crystals. One sometimes refer to them as "ECAL cells". The term "homogeneous" means that the interaction medium, where incoming particles deposit their energy, is the same as the scintillator material that transforms this energy into visible light to be transferred to the electronic readout. Lead tungstate has the following material characteristics:

- Density $\rho = 8.28$ g/cm^3
- Radiation length $X_0 = 0.89$ cm
- Molière radius $R_M = 2.2$ cm

The scintillation material converts the particle energy into photons in the visible spectrum. Those photons must be able to propagate into the crystal to reach the photodetectors. A minimum requirement on the material transparency is therefore necessary. These two features are quantified in figure 3.17.

ECAL crystals are designed as 23 cm long parallelepipeds with a rectangular section increasing from 2.2×2.2 cm^2 in the front (closer to the beam) to 2.6×2.6 cm^2 at the back where the readout photodetectors are placed. In the barrel, pairs of Avalanche Photodiodes (APD) are used while Vacuum Phototriodes (VPT) have been chosen for the endcaps. The crystal length is chosen to be able to fully contain the electromagnetic shower. It represents 25.8 radiation lengths. After conversion of the particle energy into light, the photodetectors convert the scintillation photons into electrons through the photoelectric effect. Around 4.5 such photoelectrons are produced for each MeV.

The relative energy resolution of the ECAL crystals σ/E can be described by equation (3.3), where $S/\sqrt{E}$ describes the stochastic sources of uncertainties, N/E the uncertainty coming from noise, and C is a constant term. The resolution was measured in beam tests as shown in figure 3.18. The corresponding values for the three resolution terms are given in the figure.

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{3.3}$$

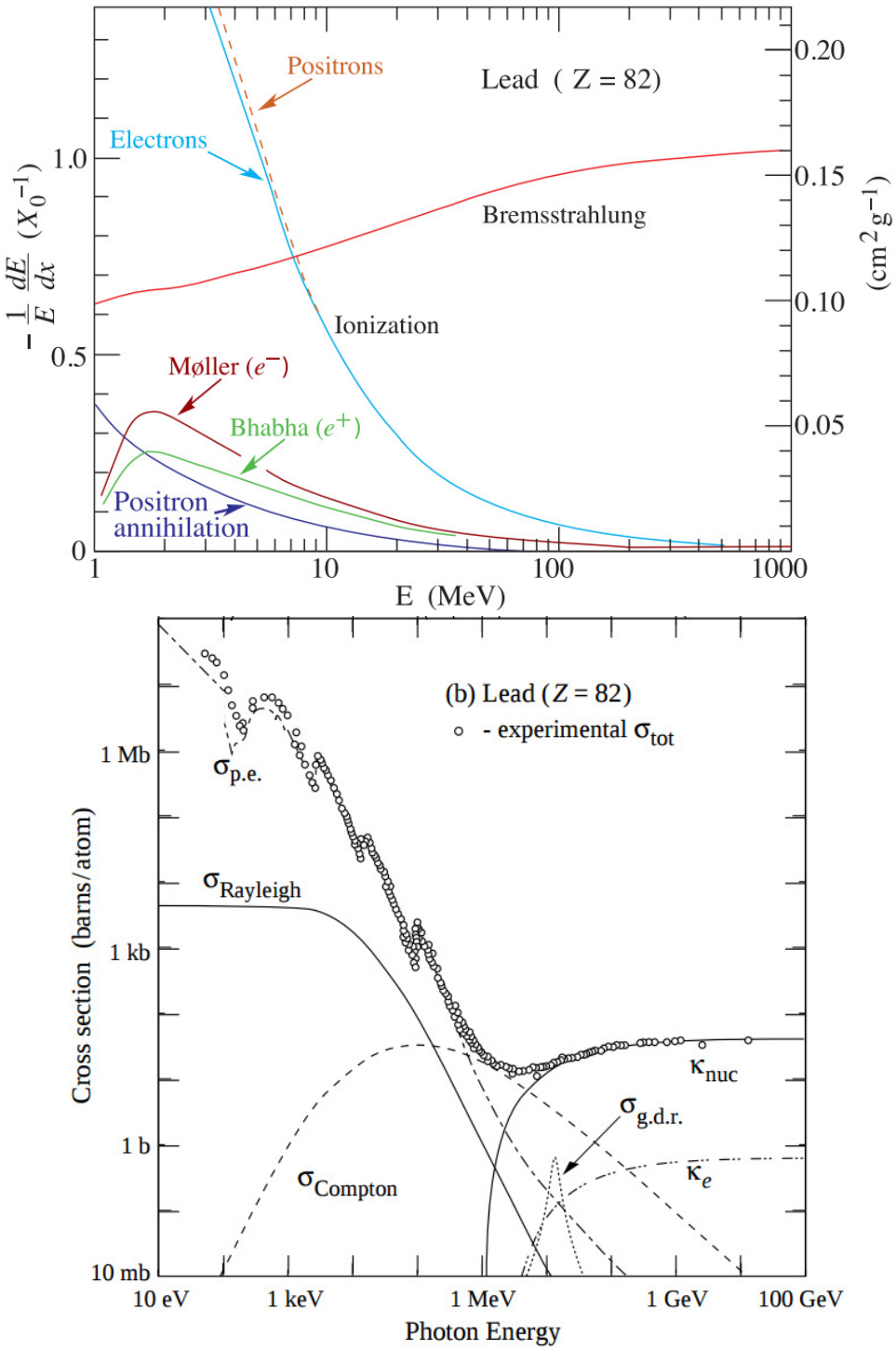


Figure 3.15: Electron stopping power in lead (top) and photon-lead cross-section (bottom) [29]. Above the GeV scale, the electron loses most of its energy by bremsstrahlung. At this energy, the largest photon-matter cross-section is the photon-nucleon cross-section κ_{nuc} describing the process where the photon produces a e^+e^- pair in the electromagnetic field of the nucleon.

Radiation damage in the ECAL

Similarly to the tracker, the ECAL suffers from radiation and sees its properties change. However, as opposed to what happens in silicon, those changes are not induced by additional defects created by inelastic interactions of particles with the $PbWO_4$ crystal [81]. It has been shown that a long exposure to radiation changes the charge-state of the existing defects in the crystal creating of so-called *color centers*. Those centers are light absorbers and they lead to a slow degradation (as a

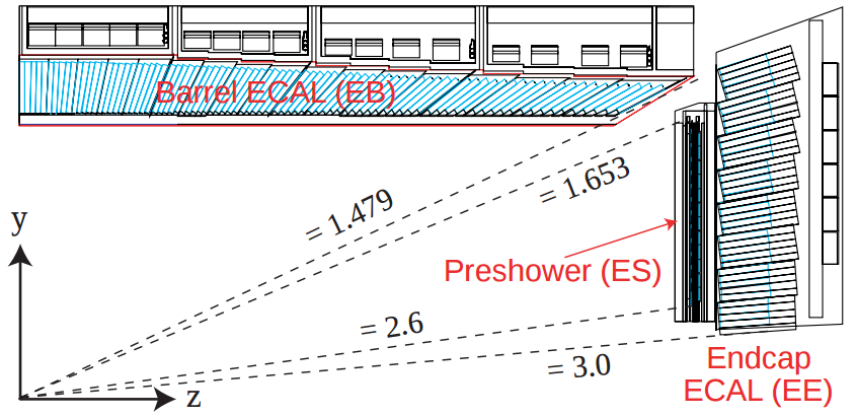


Figure 3.16: ECAL geometry [80].

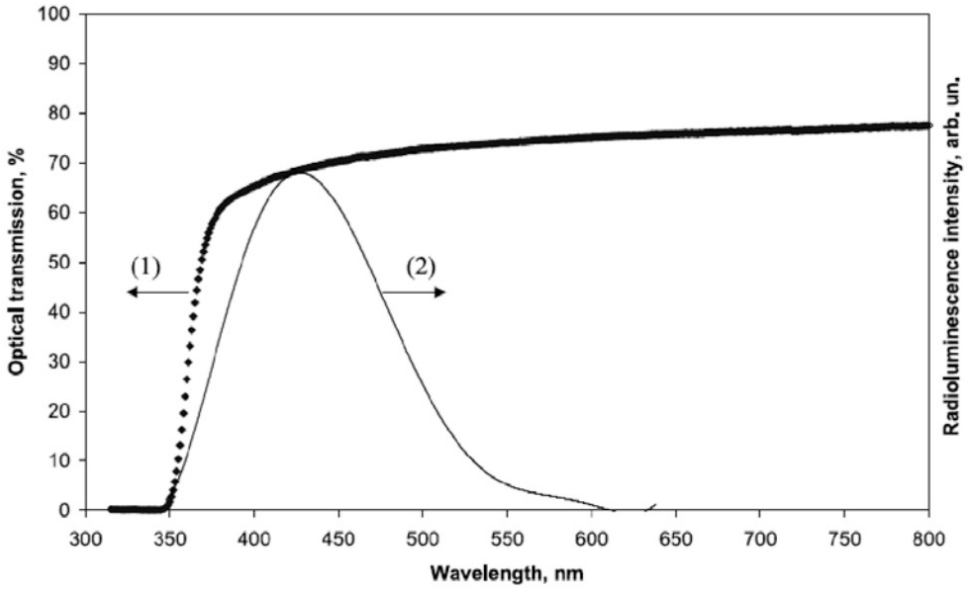


Figure 3.17: Longitudinal optical transmission (left scale) and radioluminescence intensity (right scale) in lead tungstate [70]. The longitudinal optical transmission is the ratio of the radiant energy entering on one side of the crystal to that exiting it on the other side. The scintillation light is emitted in the blue range of the visible spcterum which corresponds to a transmission efficiency of about 70 %.

function of the dose) of the crystal transmission intensity. The amount of collected light decreases with time as the crystal gets more and more opaque though annealing at room temperature (18 °C) helps recovering part of the transmission. The evolution of the opacity in the ECAL is monitored using a laser system sending light pulses through optical fibers.

3.2.3 The hadronic calorimeter

The hadronic calorimeter (HCAL) is responsible for the energy measurement of particles interacting predominantly via nuclear interactions, i.e. hadrons. Together with the tracker and the ECAL, it is used to reconstruct the jets.

Similarly to what happens in the ECAL, hadrons entering the HCAL develop hadronic showers. However, the nature and characteristics of these showers are very different. At LHC energies, interactions of hadrons with matter are driven by inelastic collisions through the nuclear force. Since the nuclear force has a much lower reach than the electromagnetic force, hadrons can travel long distances before the shower is initiated. The characteristic length associated to this distance is called the interaction length λ [82]. When the incident hadron interacts inelastically with a

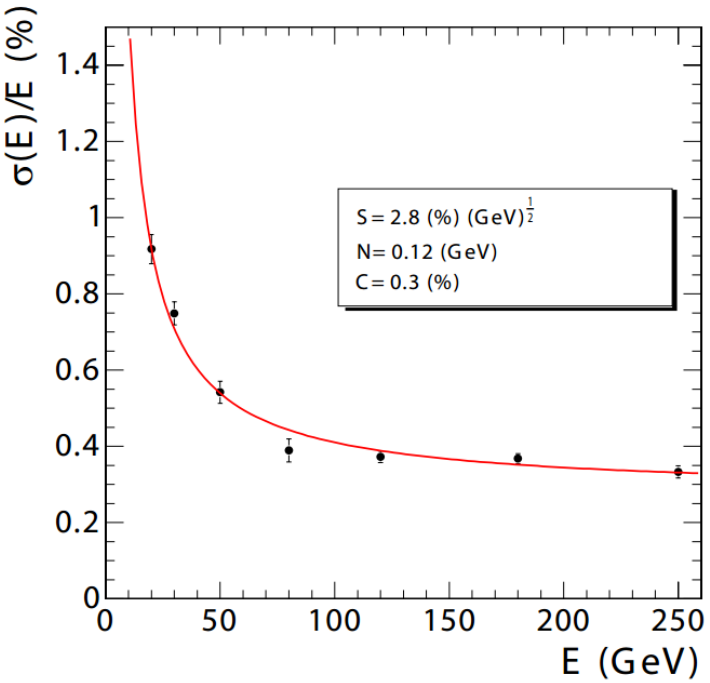


Figure 3.18: ECAL energy resolution with a test beam of 20 to 250 GeV momentum electrons [70].

nucleus of the calorimeter material, it generates many other hadrons of lower energies among which neutral pions π^0. While other hadrons continue to build up the hadronic shower through inelastic collisions with nuclei, neutral pions π^0 decay immediately into two photons inside the detector ($c\tau \sim 25$ nm in the pion referential), initiating an electromagnetic shower. The hadron initiated shower is therefore split into a hadronic and an electromagnetic sector (see figure 3.19).

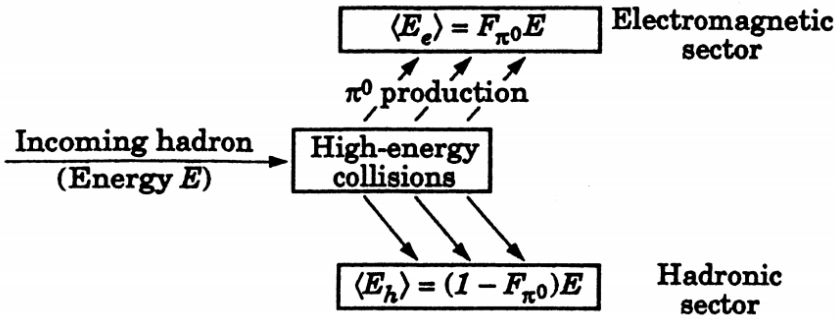


Figure 3.19: Hadronic and electromagnetic sectors of the hadron initiated shower. F_{π^0} is the total fraction of energy carried by the neutral pions [83].

The HCAL surrounds the ECAL and is similarly made of a barrel (HCAL Barrel, HB) covering $|\eta| < 1.3$ and two endcaps (HCAL Endcaps HE+ and HE-) extending this coverage up to $|\eta| = 3$. Being embedded in the CMS solenoid, the HCAL radial extension is limited between 1.77 m and 2.95 m and high-energy hadronic showers may not be fully contained. To cope with this, an additional HCAL subsystem called the HCAL Outer (HO) is installed at a larger radial distance from the beam in the barrel region. It is made of an extra scintillator layer (layer 1) installed at a radial distance of 4.07 m. Finally, covering the very forward region up to $|\eta| < 5$ is the HCAL Forward (HF). The geometry of the HCAL is illustrated in figure 3.20 (left). A zoom on the HO is displayed in figure 3.20 (right).

As opposed to the homogeneous ECAL, the HB and the HE are heterogeneous (or sampling)

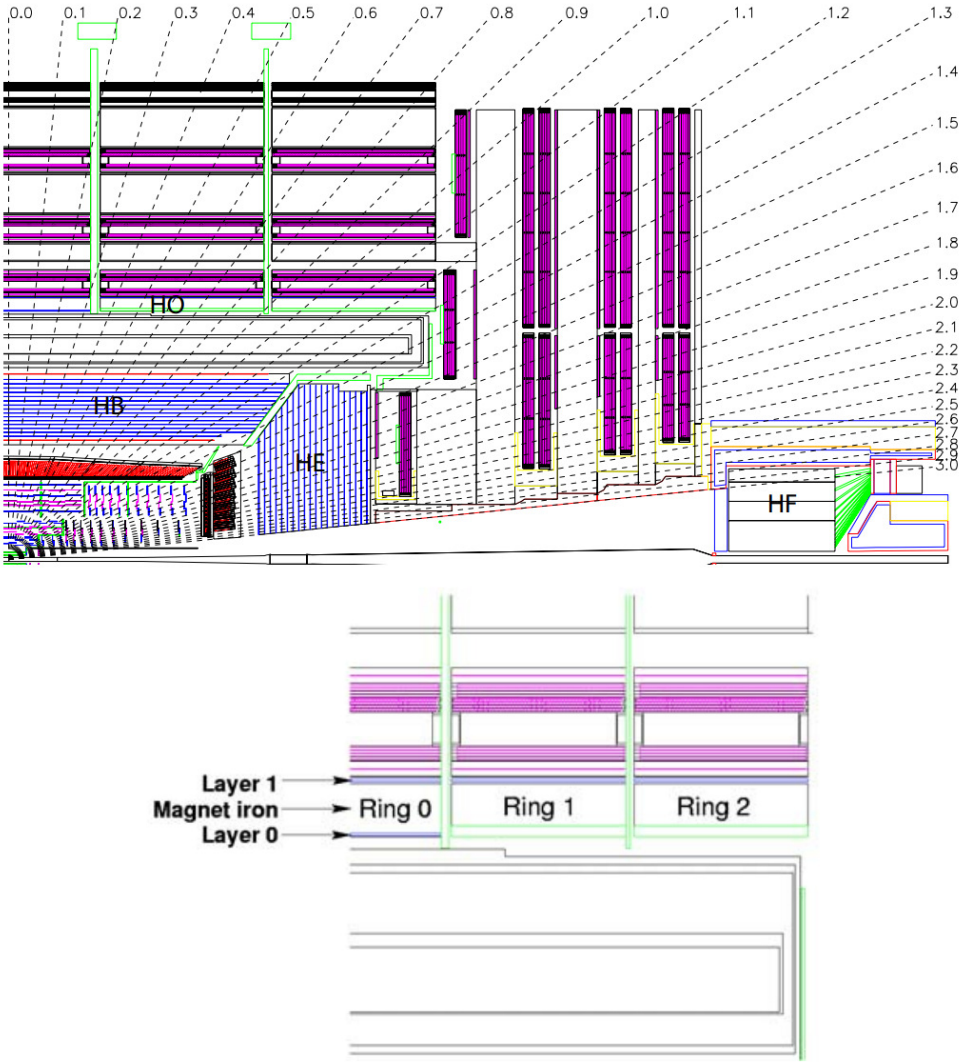


Figure 3.20: HCAL geometry (top) and zoom on the location of the HO scintillator surrounding the solenoid (bottom) [70]. Since the traversed volume depends on θ, particles in the very central part encounter less material. In this region, [-2.536,2.536] m, an additional HO scintillator layer (layer 0) is placed just before the solenoid at a radial distance of 3.82 m to help collect the energy.

calorimeters. This means that they are made of alternating layers of absorber and scintillator material. The absorber is made of brass (except for the innermost and outermost layers that are made of steel for structural purposes) with an interaction length $\lambda = 16.42$ cm and a radiation length $X_0 = 1.49$ cm. The thickness of the absorber depends on θ. In the transverse plane it corresponds to 5.82 interaction lengths, going up to 10.6 at $|\eta| = 1.3$. For both the HB and HE, the scintillator is plastic-made and is able to handle high radiation doses. Light produced in the scintillator active medium is collected by wavelength shifting fibres that generate lower energy photons from the scintillation light. Optical cables then bring out the light from the scintillator to the readout electronics. Hybrid Photodiodes (HPD) do the job of producing the electric signal from the collected photons.

The HO utilizes the solenoid volume itself as an absorber and collects the energy with scintillators located right after the solenoid (from a radial point of view).

Due to its very forward position, the HF undergoes very large particle fluence. This led to a different approach for the subsytem design. The hadronic shower is initiated in a steel absorber filled with quarz fibers and no scintillator is used. Instead, the Cherenkov light generated by the shower particles is collected by the fibers. Since in this region there is no ECAL coverage, one would like to be able to distinguish electromagnetic from hadronic particles in another way. Elec-

tromagnetic showers are initiated earlier than hadronic showers which allows for a discrimination along the z axis. Half of the fibers start 22 cm further away from the beggining of the HF along the z axis. Those so-called "short" fibers and the full coverage "long" fibers are read separately to allow for this discrimination.

3.2.4 The muon system

CMS was designed to have the best possible resolution on the measurement of muon four-momenta, given the limited budget and the need to be able to detect all particles but neutrinos. The ECAL was designed to be able to distinguish two close photons from a boosted Higgs boson decay. Similarly, one of the requirements on the muon system was to open the possibility for the Higgs boson detection in another "golden" channel, the H→ZZ→4μ decay mode. More generally, the muon system was designed to allow triggering on events with muons in the final state. It must enable the identification of events of interest within the trigger timing window (this is developed in section 3.4). Muons produced in high energy events at the LHC have a momentum typically larger than 1 GeV. This makes them minimum ionizing particles (MIP) going through the entire detector without being stopped and with a minimal ionization energy loss (see figure 3.21). The muon system is therefore placed in the outmost region of the detector where only muons can reach it. This allows for a cleaner detection and a better resolution on high-p_T muons.

The CMS muon system is a tracking detector responsible for the measurement of the muon p_T and charge. As opposed to the inner tracker, it is not made of silicon but of gaseous active mediums. Another distinction with the silicon tracking system is that dE/dx information is not read out. Together with the inner tracker, they are used for muon reconstruction. The muon system is made of three subsystems: the Drift Tubes (DT), the Cathode Strip Chambers (CSC) and the Resistive Plate Chambers (RPC). Figure 3.22 illustrates the geometry of the muon systems.

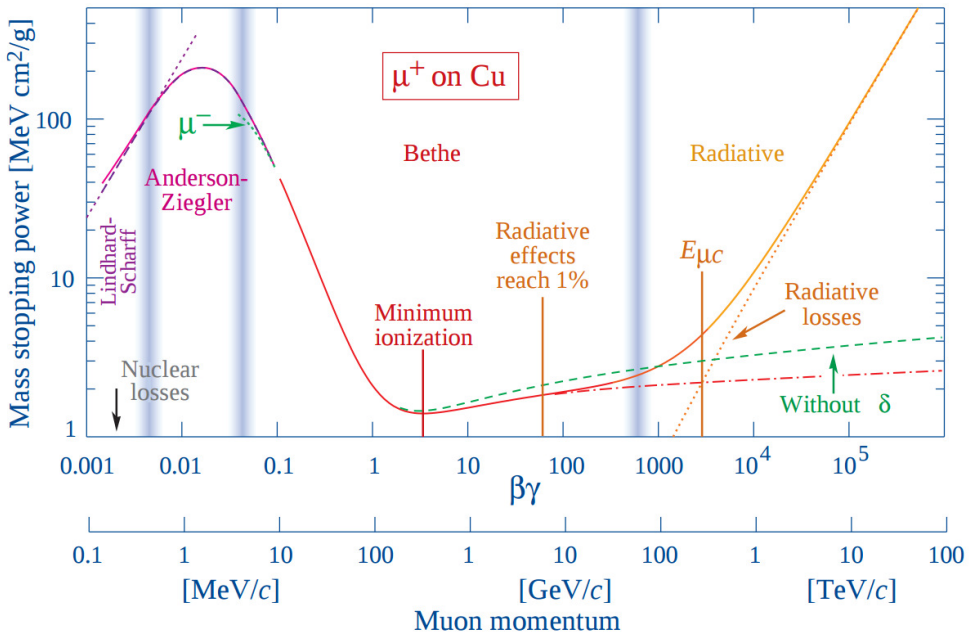


Figure 3.21: Muon stopping power in copper as a function of its momentum. Muons produced at the LHC have a typical momentum in the [1,1000] GeV range [29].

Drift Tubes

The particle flux in the central part of the detector is low enough to allow for the use of Drift Tubes technology. DTs are gas ionization chambers located in the central part of the detector

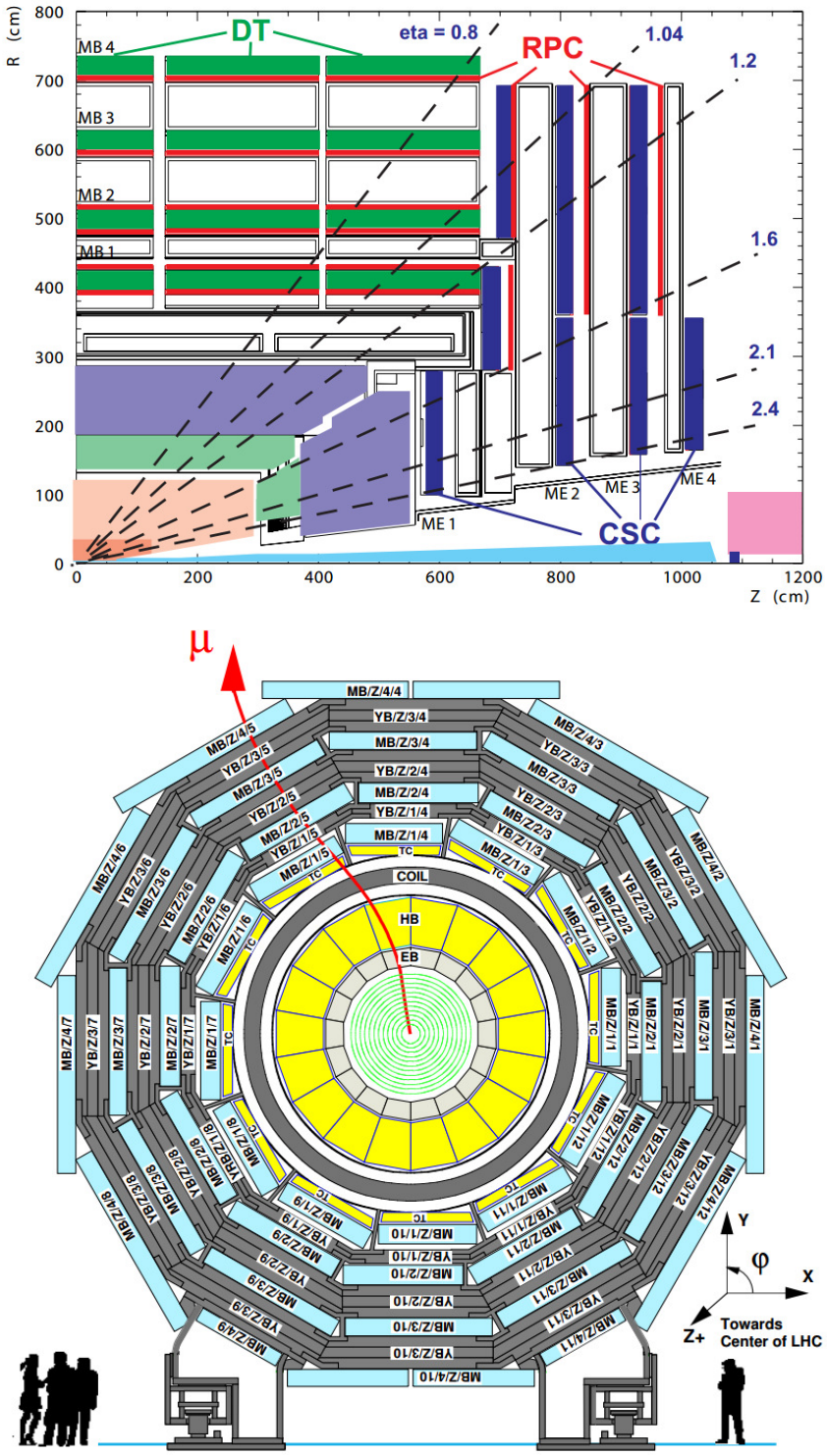


Figure 3.22: Geometry of the muon system (top) [80] and transverse view in the barrel (bottom) [70].

in the $|\eta| < 1.2$ range. DT cells are made of a steel wire anode and of two aluminium cathode strips. A muon passing through ionizes the gas, creating free electrons that are collected at the cathodes after drifting along the electric field lines. Figure 3.23 illustrates the design of a DT cell. The anode wire at the center is set at a voltage of 3600 V. On the left and right, the cathodes are set at a voltage of -1200 V. The electrode strips on the top and the bottom are set at a voltage of 1800 V and are used to flatten the field lines so that the electrons drift more efficiently towards the cathodes. The maximum transverse size of the cells (2.1 cm) is the result of a compromise between a fast enough drift time (380 ns) and a low enough number of input channels.

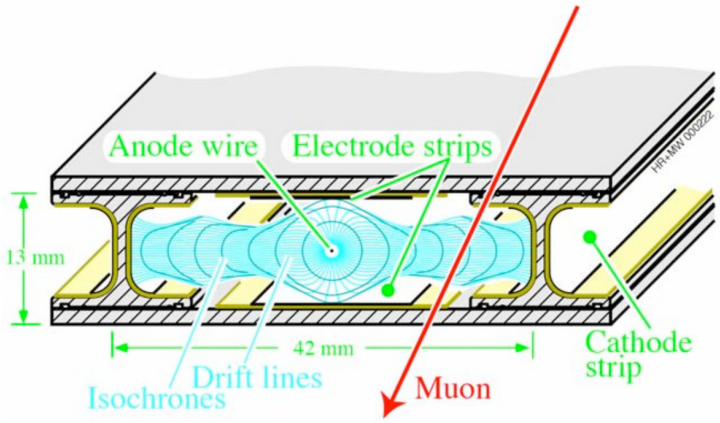


Figure 3.23: Illustration of a DT cell. [70].

DTs are arranged in four concentric layers, two of them being embedded in the magnet return yoke as depicted in figure 3.22. They are grouped by four to make superlayers (SL). The rectangular cells are placed either parallel or perpendicular to the beamline to measure the ϕ or the z coordinate, respectively. Each concentric layer has two ϕ-measuring SLs placed around a z-measuring SL except the fourth and outmost one that only has two ϕ-measuring SL.

The DT hit resolution is given in figure 3.24.

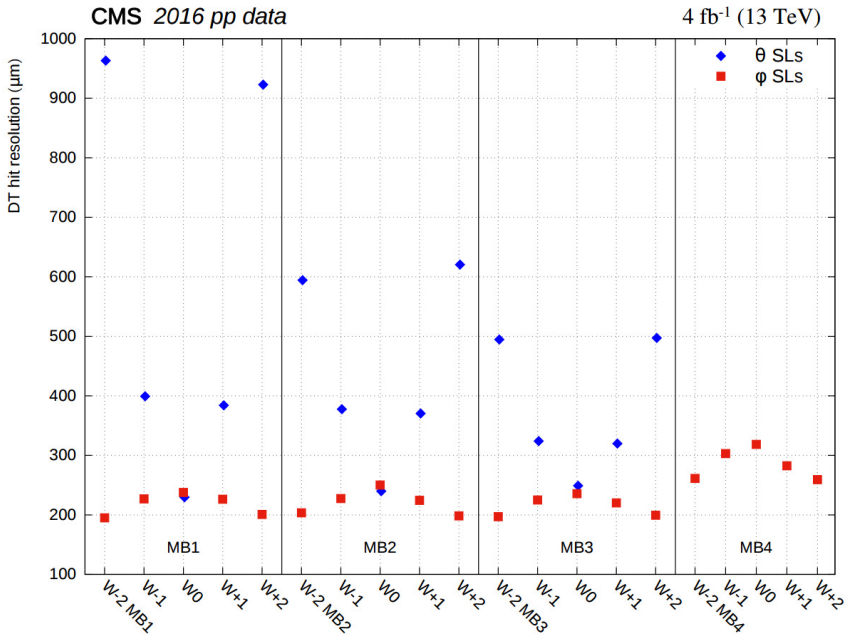


Figure 3.24: DT hit resolution measured in 2016 data for the z-measuring (here denoted as θ) superlayers (blue) and ϕ-measuring superlayers (red) in different DT stations [84].

Cathode Strip Chambers

In the endcaps, high particle rates impose the use of a different technology. A detector placed in such a harsh environment must have a fast response time as well as a good radiation hardness. Those conditions are met by gaseous proportional counters. The CMS Cathod Strip Chambers extend the η range of the muon spectrometer up to $|\eta| = 2.4$. They are made of "gaps" between two stripped cathode planes at the center of which a series of parallel anode wires are placed

perpendicular to the cathod strips. The high voltage between the anode wires and the cathode planes is of 3600 V. The perpendicularity between the strips and the wires allows for a precise measurement of both the η and the ϕ coordinates of the muon hit. As illustrated in figure 3.25, a muon crossing a CSC generates an avalanche on one of the wires. As wires are placed along the ϕ direction, this measurement gives the η coordinates. Due to the fact the cathode plane is stripped, each strip feels the positive charge of the drifting positively charged ions differently. A charge profile in the ϕ direction can therefore be extracted from the readout of neighbouring strips.

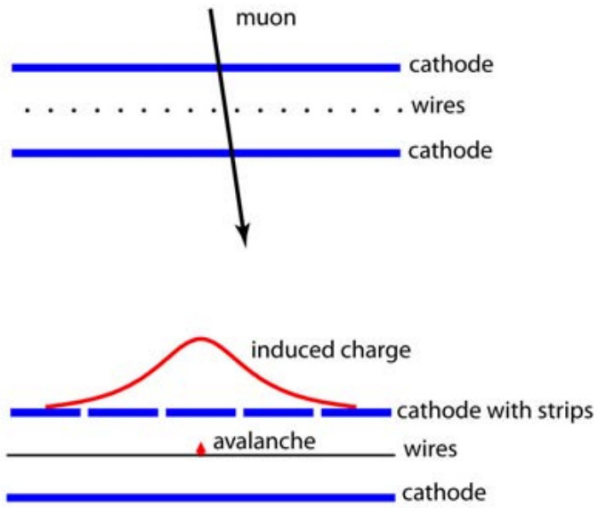


Figure 3.25: Sketch of the CSC gap and detection method. [70].

In figure 3.26 (left), the geometry of the CSC is shown. Similarly to the DTs, the CSCs are embedded in the forward and backward parts of the return yoke. On the right, a chamber is shown with the wires running perpendicularly to the cathode strips.

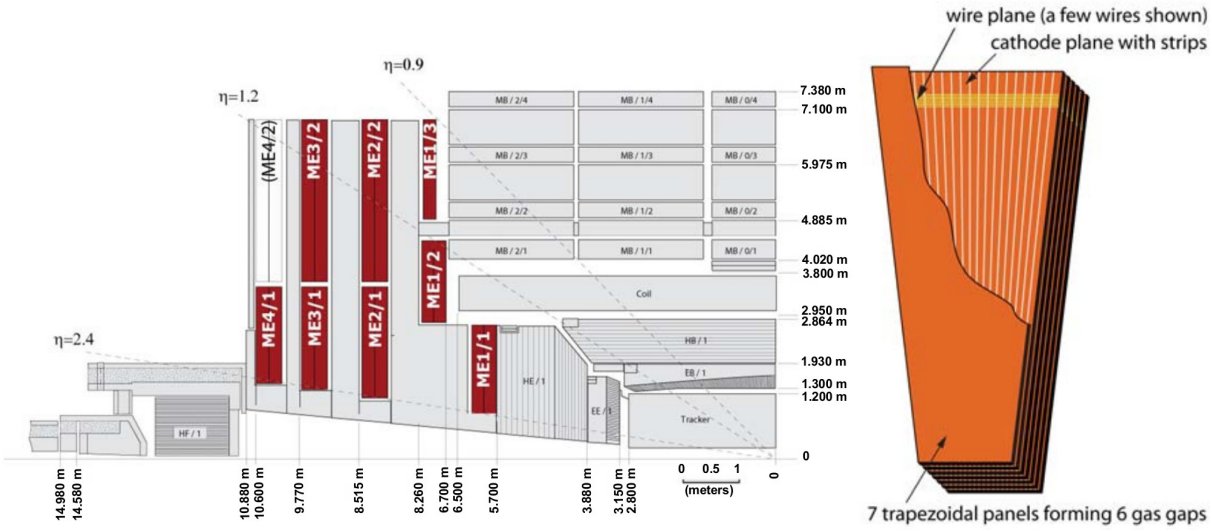


Figure 3.26: Geometry of the Cathode Strip Chambers (in red) and arrangement of a single chamber. The chamber opening angle increases with the radius in the transverse plane. [70].

The hit resolution in the CSCs runs from 45 μm (in the ME1/1a station) to 134 μm (in the ME4/2 station) with increasing distance from the interaction point [84].

Resistive Plate Chambers

To enhance the timing properties of the muon spectrometer, a third system is added both in the barrel and in the endcaps. Resistive Plate Chambers have a time resoluton of 2 ns [84] and so are able to distinguish two ionizing events in less than the 25 ns between two proton bunch crossings. This is important to avoid degeneracy in the muon identification and to ensure proper triggering (see section 3.4 on trigger). RPCs are gaseous detectors made of double gaps between electrodes. They operate in avalanche mode. The RPC hit resolution ranges from 0.78 to 1.27 cm in the barrel and from 0.89 to 1.38 cm in the endcap [84].

3.3 The reconstruction procedure

Reconstruction is the procedure according to which measurements from several detectors are assembled and interpreted as the passage of a particle. Except for neutrinos, CMS detects all long-lived, free-propagating particles (i.e. not quarks): the charged leptons, the photons and the hadrons. The tau lepton should be considered apart from the two other charged leptons as it decays before entering the tracking system ($c\tau \sim 87$ μm). It is therefore reconstructed via its decay products. The same is true for hadrons that decay promptly such as the neutral pion. We first introduce the particle-flow framework before going through all the objects detectable by CMS. Figure 3.27 sketches how different particles interact with the detector.

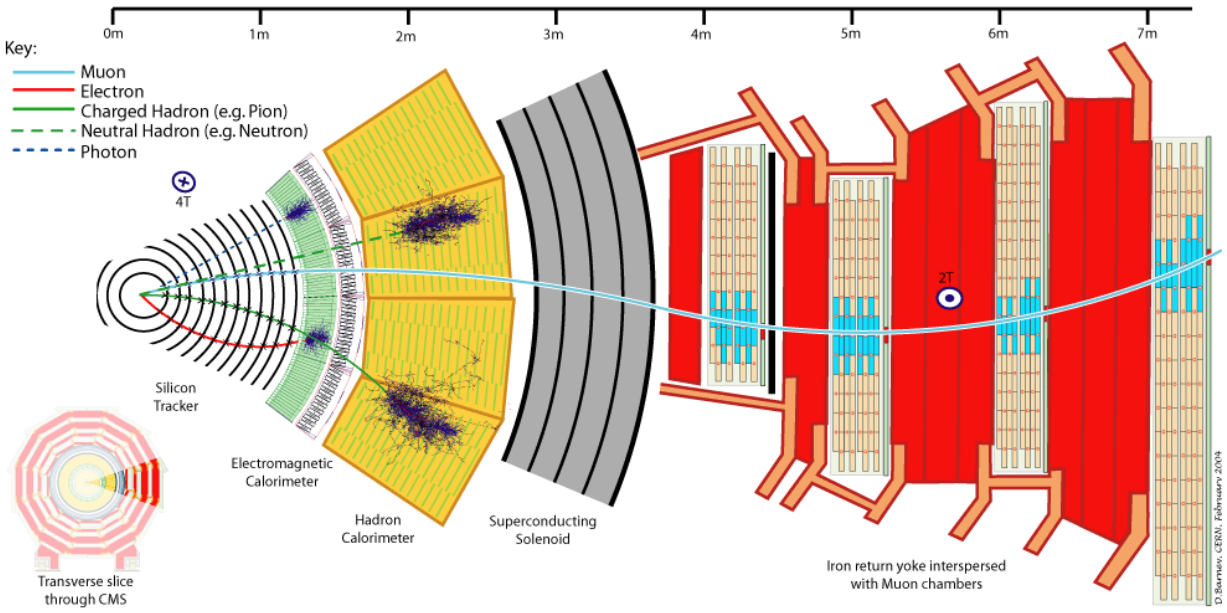


Figure 3.27: Transverse view of the CMS detector with particles-detector interactions. [85].

3.3.1 The particle-flow framework

The particle-flow (PF) framework is a reconstruction algorithm that uses information from all CMS subdetectors to build up a coherent description of the particles in the event. The reconstruction of each type of particle is not done in a standalone way but is part of a global method. The particle-flow algorithm helps reducing the ambiguities regarding particle identification and leads to higher reconstruction efficiencies. The PF algorithm relies on the definition of basic elements and of links between them. The elements are:

- Inner tracks
- Muon tracks

- ECAL clusters

- HCAL clusters

The next step is to define *blocks* by linking closeby elements. Those blocks are a rough collection of elements likely to belong to the same particle. The requirement of vicinity is defined in η and ϕ and allows to keep small the number of elements in the blocks (typically up to three). This keeps the computing time independent of the complexity of the event. The PF reconstruction logic goes as follows. Objects are reconstructed sequentially using available blocks which are removed from the list once used. The reconstruction continues until all blocks are clustered in PF *candidates*. The reconstruction order is the following:

- **Muons** Muons are reconstructed as described in section 3.3.3 and associated tracks are removed from the PF blocks.

- **Electrons** The electrons have their own tracking algorithm, as mentioned in section 3.3.4. The second step in the PF algorithm is to associate those candidate electron tracks to an ECAL energy cluster. The energy of the associated HCAL cluster must not be larger than 10 % of the considered ECAL cluster energy. This matching defines PF electrons. The associated calorimeter clusters and tracks are removed from the PF blocks

- **Photons** If no track is found, the ECAL energy is associated to a prompt PF photon. Once again, the ECAL information is removed from the blocks.

- **Hadrons and non-prompt photons** Following the PF procedure, the remaining tracks are matched to the remaining calorimeter clusters to form charged hadrons. The matching is only valid if the p_T sum of the tracks pointing to the calorimeter cluster are compatible with the cluster energy. If it is significantly lower, the cluster energy is attributed to a neutral particle. If the calorimetric energy is lower or equal to the ECAL energy, it is interpreted as a PF photon. Otherwise, a PF neutral hadron is created.

The PF algorithm has been used extensively during the Run 2 period. All information on this matter can be found in reference [86]. We give more details on PF performances for muons and jets as those objects are used in the searches presented in this work.

3.3.2 Tracks

Before reviewing the main physics objects, we go over the reconstruction procedure of the tracks in the inner silicon tracker. The information presented here is extracted mainly from reference [73].

First, we must briefly introduce what tracker hits are. Hits are clusters of signal from several pixel or strip channels and form the basic blocks for track reconstruction. A track is a collection of hits resulting from the reconstruction procedure. It is the interpretation of a collection of hit as the passage of a particle. In particular, the sign of the particle's charge and its transverse momentum are extracted from the tracking algorithm. In CMS, the absolute value of the particle's charge is always assumed to be equal to one (in units of the electron charge).

The track reconstruction algorithm used in CMS is called the Combinatorial Track Finder (CTF), which is derived from the Kalman filter method for tracking [87]. The CTF algorithm is applied iteratively (six times) to the collection of hits in the event to reduce the computation time. It starts by reconstructing cleaner and easier-identified tracks (high p_T, isolated) and then ignores the hits involved in those tracks in the further iterations. We now go through each step of the reconstruction procedure applied at each iteration.

Seed generation

The CTF algorithm starts with the reconstruction of *seeds*. Seeds are tracks made of two or three hits that serve as a starting point for the tracking algorithm. When only two hits are considered, extra information from the beam spot or the primary vertex is needed to define the full set of track parameters. These primary vertices are computed from a fast-reconstruction algorithm using pixel-information only. They are called *pixel vertices*. Tracking seeds are built from inner layers after which the tracking algorithm builds the full track outwards. At each of the six tracking iterations, the *seeding layer* (set of layers used to find seeding hits) is different. The main seeding layers are:

- **Pixel triplets:** these are made of three pixel hits. Most promptly produced particles are reconstructed from these high quality seeds.
- **Mixed pairs with vertex:** in this seeding layer, two hits (either from the pixels or the TEC) and a pixel vertex are used for seed reconstruction.
- **Mixed triplets:** to build such seed, one to three pixel hits are required and less than three strip hits. This seeding layer aims at finding seeds of displaced tracks and prompt tracks with missed pixel hits.

Track finding and fitting

Starting from the seeding track, the track finding algorithm takes over. It is a four-steps recursive algorithm that loops over the tracker layers to add hits to the current track and redefines the track parameters with a Kalman filter [87] until the full track is reconstructed. First, the algorithm determines what layers could be crossed by the track using an extrapolation of the current track. This step is referred to as the "navigation step". To do that, it assumes a perfect helical track, i.e. a perfectly homogeneous magnetic field and no Coulomb scattering of the particle with the detector material. The helical track is defined by five parameters. They depend on the coordinate system and are defined with respect to a reference point (typically the vertex). In the *perigee* system, they list:

- d_0: the perigee, or the impact parameter in the transverse plane
- z_0: the impact parameter along the z axis
- The angular coordinates η and ϕ
- The curvature $R = \frac{p_T}{qB}$, where p_T is the momentum in the plane perpendicular to the magnetic field B and q is the particle's charge.

In the second step, the algorithm looks for compatible modules in the layers selected by the navigation. Compatible means that the extrapolated track passes within three standard deviations of the module boundary. However, as some modules of the same layer overlap, two of them may be crossed by the same track. To deal with this, mutually exclusive modules in the same layer are grouped together. Mutually exclusive means that a track cannot physically cross two modules of the same group.

Then, the collection of hits is identified in each of the compatible groups. The compatibility of the hits with the extrapolated track is checked with a simple, loose, $\chi^2 < 30$ requirement. "Ghost" hits may be added to the hits collection to cope with the possibility of defective sensors.

Finally, the current track parameters are updated to form a collection of new track candidates, each with exactly one hit from each group. Every time a new hit is added to a track, the new collection of hit is fitted with a Kalman filter. The perigee parameters presented above are extracted from the track fitting and the hit collection is updated. A Kalman filter is a tool to analyze linear dynamical systems. Such systems describe the linear time evolution of a state vector. In our case, the dynamical system is the track which is described by a vector in space $\vec{x}$ dependent on five

parameters. Moreover, in the context of tracking with CMS, the evolution of the state vector is discrete, as it is evaluated at the tracker layers only. The Kalman filter actually runs in two main steps: the filtering and the smoothing. The filtering is the estimation of the track parameters (or the state vector) at the current layer with prior knowledge of hits in all previous layers. Then comes the smoothing step where the Kalman filter is applied to each previous vector backward to the first hit. The precision on the track parameters gets better as more hits are included in the Kalman filter fit. By applying the Kalman filter algorithm backward to finish with the seeding hits, the smoothing step then allows to further enhance the precision on the track parameters.

To limit the multiplication of track candidates at each layer, only the best five candidates are kept. The selection is done with the normalized χ^2 and a penalty on the number of ghost hits. The algorithm terminates when the end of the tracker is reached, or if quality requirements (e.g. on the minimal p_T) are not met anymore.

Track selection

Finally, a selection is made on the reconstructed tracks to reject fakes. The main criteria are:

- "A minimum number of layers in which the track has at least one associated hit" [73]
- "A minimum number of layers in which the track has an associated 3D hit" [73]
- "A maximum number of layers intercepted by the track containing no assigned hits" [73]
- A minimal $\chi^2/\mathrm{N_{d.o.f.}}$
- A minimal impact parameter to the beam spot
- A minimal z-distance to the primary vertex

A 3D hit is a pixel hit or a "matched hit" in the strips. A matched hit is a hit read by two modules titled with respect to each other hence providing a stereo view. These requirements depend on the layers crossed and are numerically given in reference [73].

The tracking efficiency is evaluated as a function of the tranverse momentum for muons and electrons in simulated events. This is shown in figure 3.28.

3.3.3 Muons

Muons deposit a small amount of energy by ionization in the inner tracker and the muon system. This energy is clustered in hits, distributed along their path. Therefore, muons should be thought of as tracks reconstructed from those hits. The two tracking detectors can be used in combination or each one alone, as we describe below. The information detailed in this section is extracted from references [88, 89].

Standalone muons

Standalone (SA) muons are tracks reconstructed in the muon system only. Hits from the same muon chamber are first assembled in straight-line tracks referred to as "segments". This is called "local reconstruction". A subset of these segments is then used as a seed for the track generation using a Kalman filtering technique [90]. This procedure is called "global reconstruction". Momentum resolution using the muon system only is given in figure 3.29.

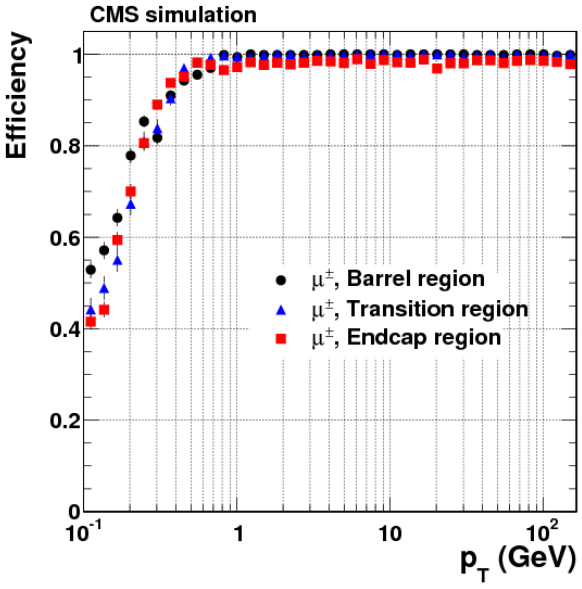


Figure 3.28: Tracking efficiency as a function of the tranvserse momentum for muons in simulated events.

Global muons

Global muons are built of SA muons matched to a tracker track. This strategy is referred to as "outside-in". Once a candidate tracker track is found, a global refit is done using all information from both tracks. Global muons have the best resolution on the muon transverse momentum as they use all the information available (see figure 3.29). The associated reconstruction efficiency is lower due to the more stringent requirements on the muon tracks.

Tracker muons

Tracker muons are the result of an "inside-out" strategy: the algorithm starts from inner tracks and looks for a matching signal in the muon chambers. As opposed to global muons for which the tracking is redone using information from both systems after track matching, tracker muons have the same momentum as the seeding track. Due to its higher granularity, the tracker has a better momentum resolution than the muon system alone at low p_T. This difference tends to even out at large p_T as the curvature of the tracks in the transverse plane gets very small.

The muon reconstruction and identification efficiency is given in figure 3.29 (right). It is very high and flat as a function of the pseudorapidity.

Performances of particle-flow

Thanks to their MIP behaviour, muons are the only detectable particles to reach the muon system. This gives them excellent tracking and identification efficiencies. For this reason, the PF does not modify the muon reconstruction algorithm. However, for a muon to be considered in the global PF description, it must fulfill several quality requirements:

- **Isolation** The sum of the p_T of all tracks and of the E_T of all calorimeter clusters in a $\Delta R \leq 0.3$ cone around the muon track should be smaller than 10 % of the muon p_T itself. The smaller of the tracker and global muon p_T is the one considered.

- **Non-isolated muons** Those muons are still considered but must pass the tight identification criteria defined in reference [88]. Most of these muons are real muons coming from the decay of heavy flavour hadrons.

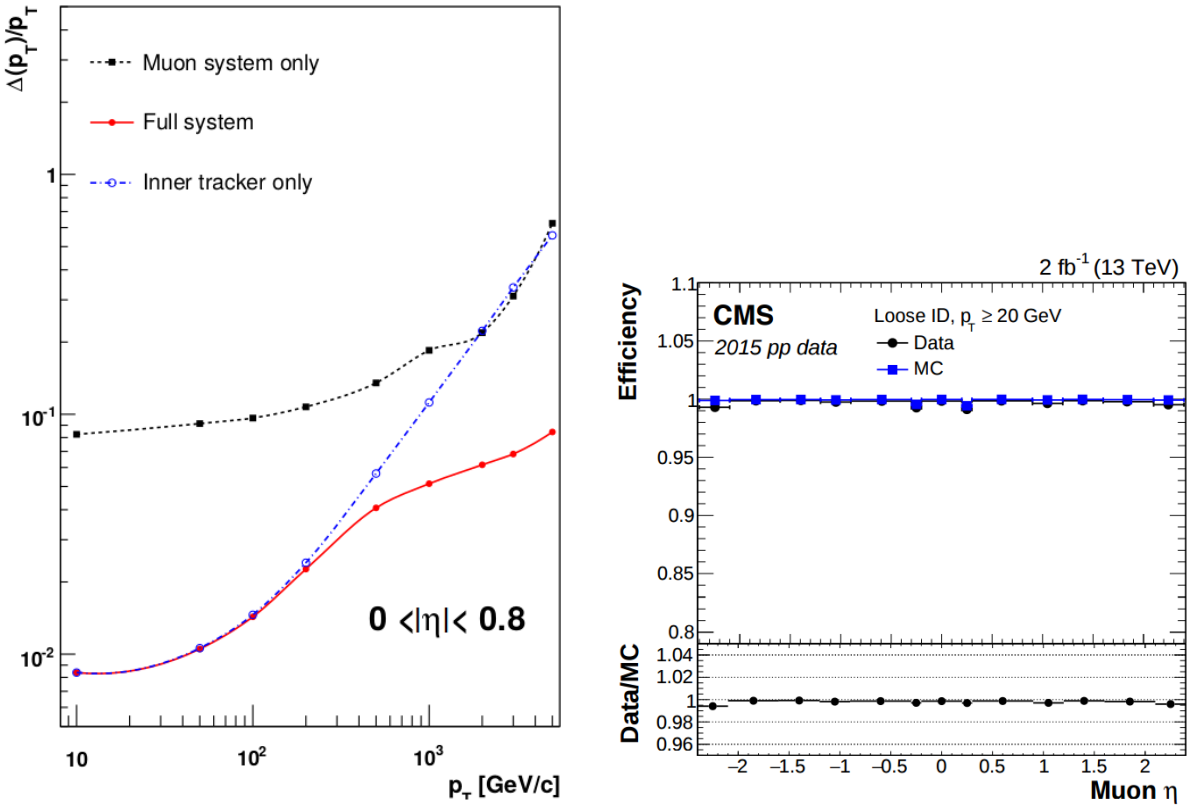


Figure 3.29: Resolution on muon track transverse momentum as a function of its transverse momentum [70] (left). This is an expectation from 2008 but is really close to what is observed. Reconstruction and identification efficiency measured in 2015 data with the tag-and-probe technique [89] (right). A "loose" muon is "a muon selected by the PF algorithm that is also either a tracker or a global muon" [89].

Muons passing either one of those criteria are ignored in the next PF reconstruction steps. This means their inner track and potential calorimeter deposits are removed from the PF blocks. Muon identification efficiency is compared for different algorithm in figure 3.30.

3.3.4 Electrons

As light charged particles, electrons interact both in the tracker and in the ECAL. We describe here how those two systems are used for electron reconstruction based on information from reference [91].

Electrons do not deposit all their energy in a single ECAL crystal. If an electron hits the center of a crystal in the absence of a magnetic field, around 97 % of its energy is contained in a 5×5 crystal array. This is however an ideal case. In CMS, electrons must first go through the tracker material leading to the emission of bremsstrahlung photons by interaction with the electric field of the silicon nuclei. This effect depends on the thickness of the crossed material, thus on the pseudorapidity. In this direction the electron track is not curved and so the radiated photons are emitted with the same pseudorapidity as that of the electron. In the transverse plane, however, electron tracks are curved by the large CMS magnetic field. Bremsstrahlung photons are emitted tangentially leading to a spread in ϕ of the inital electron energy. In both cases, bremsstrahlung photons must be included in the total energy of the initial electron. In the barrel, the "hybrid" clustering algorithm is used to collect the energy from neighbouring cells. First, a seed is defined as the crystal of local maximum of energy above a certain threshold. Then, arrays of 5×1 crystals in $\eta - \phi$ are built up. If their energy is above the array threshold, they are kept in the seed-array. This way, clusters of 5×1 crystals around a seed are built. The total energy

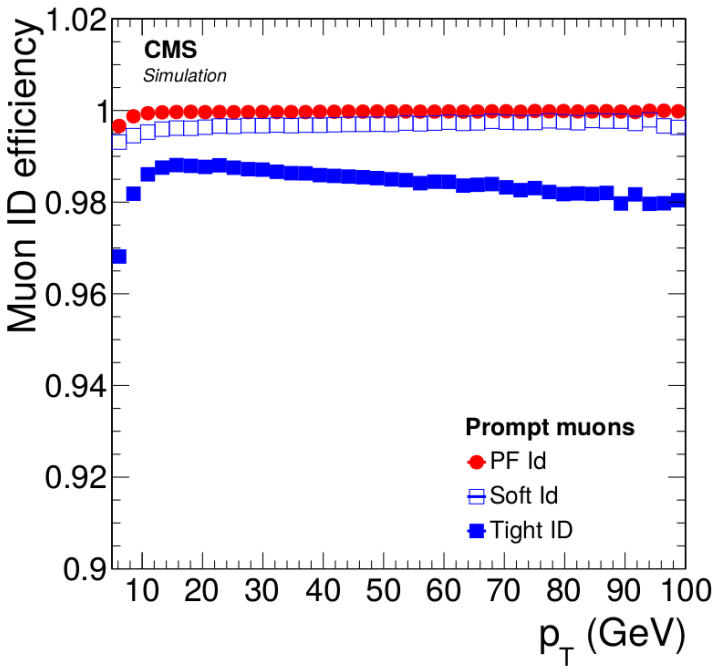


Figure 3.30: Muon identification efficiency for different algorithms [86] as a function of the reconstructed muon p_T. The efficiency is defined with respect to muons reconstructed either as global or tracker muons. Tight identification is the one referred to in the text and soft identification is optimized for low p_T muons [88].

is obtained by grouping clusters into superclusters (SC). In the endcaps, another algorithm called the "multi-5×5" is used due to the different crystal ordering. Seeds are defined as local maxima of energy with respect to the four closest neighbouring crystals. The clustering algorithm groups ECAL crystals in 5×5 clusters by looking for the second-to-highest energy crystal. SC are then built from those clusters.

As opposed to the tracker muons, electron tracks are not extracted from the tracker tracks collection. Due to their much smaller mass, electrons lose a large fraction of their energy via bremsstrahlung leading to a smaller curvature in the transverse plane. The standard tracking algorithm may therefore miss hits because it assumes the electron track to be straighter than it is. We will not develop here the details of the dedicated electron-tracking algorithm. It is called the GSF (Gaussian Sum Filter) method and we refer to [92] for any further details.

Electrons can be ECAL- or track-seeded. In the ECAL-seeded case, the track-SC matching is done through a geometrical matching between the track and the cluster itself. The requirement is looser in the ϕ direction to allow for large changes in the electron initial direction due to bremsstrahlung. It is required that $|\eta_{SC} - \eta_{\text{track}}| < 0.02$ and $|\phi_{SC} - \phi_{\text{track}}| < 0.15$, where the track coordinates are extrapolated from the innermost tracker layer to the ECAL. For track-seeded electrons, the track-SC matching is done through an MVA method.

As was shown by equation (3.3) and figure 3.18, the ECAL energy resolution improves with the energy. On the other hand, the momentum resolution on the track degrades with the momentum due to the straightening of the track. We observe the impact of those two effects on the electron energy resolution as given in figure 3.31 (left). On the right the electron reconstruction efficiency using an MVA-based method is shown.

3.3.5 Photons

As neutral electromagnetic particles, photons are reconstructed using ECAL information mainly but also inner tracker information is used to veto tracks. Photon reconstruction is described in detail in reference [93].

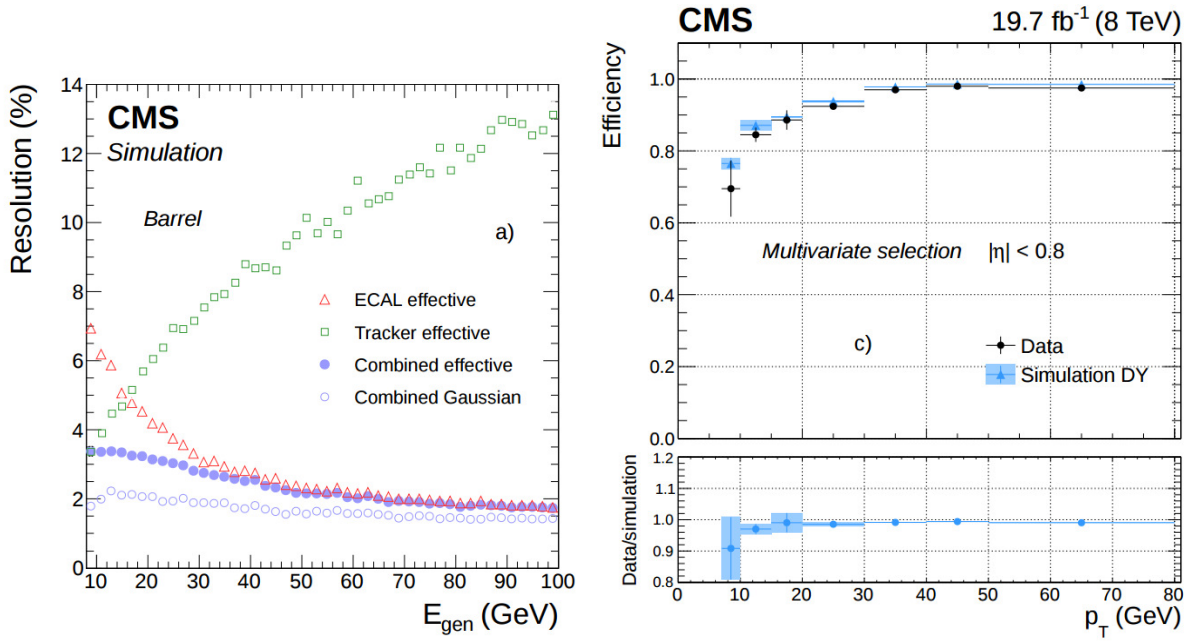


Figure 3.31: Resolution on electron momentum in simulated events as a function of its generated-level energy (left) and reconstruction efficiency as a function of its transverse momentum using an MVA-based method (right) [91].

Due to the the fact photon-induced and electron-induced electromagnetic showers have the same macroscopic shape and behaviour, photon energy in the ECAL is clustered using the same techniques as for electrons. This energy is corrected to take into account the impact of photon conversion in the tracker before reaching the ECAL. These corrections are done through a multivariate analysis taking as input the η and ϕ coordinates of the ECAL SC as well as shower shape variables among which the R_9 of the SC and the H/E ratio. The R_9 variable is defined as "the energy sum of the 3×3 crystals centred on the most energetic crystal in the supercluster divided by the energy of the supercluster" [93]. It peaks at unity for unconverted photons but has a broader distribution for converted photons. H/E is the ratio of the energy of the HCAL cells behind the SC to the SC energy. It is expected to be small for electromagnetic objects.

Photon energy resolution is given in figure 3.32. The resolution gets worse with η due to the higher photon conversion rate in regions with larger material budget.

3.3.6 Taus

Tau leptons most often decay in the vacuum tube and are therefore detected through the reconstruction of their decay products. Thanks to its large mass and as opposed to the two lighter leptons, the τ is able to decay hadronically. The branching ratio of its main decay channels are [29]:

- $\tau \to \pi^{\pm} + X$: 45.6 %
- $\tau \to l^{\pm} + X$: 35.2 %
- $\tau \to 3\pi^{\pm} + X$: 9.0 %

Tau reconstruction is described in details in reference [94].

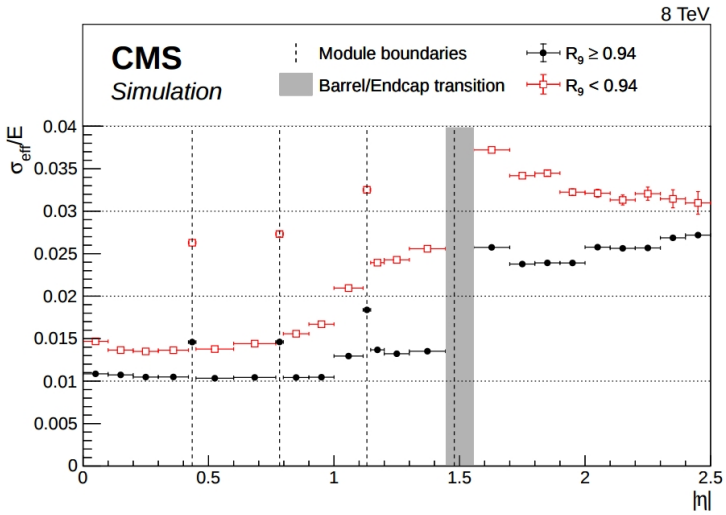


Figure 3.32: Photon energy resolution as a function of the photon pseudorapidity [93].

3.3.7 Jets

Jets are the result of the hadronization of quarks and gluons. According to QCD, those colored particles cannot freely propagate but must associate with other colored particles to form "white" objects called hadrons. This results in a collection of particles collimated in the direction of the initial parton (see figure 3.33).

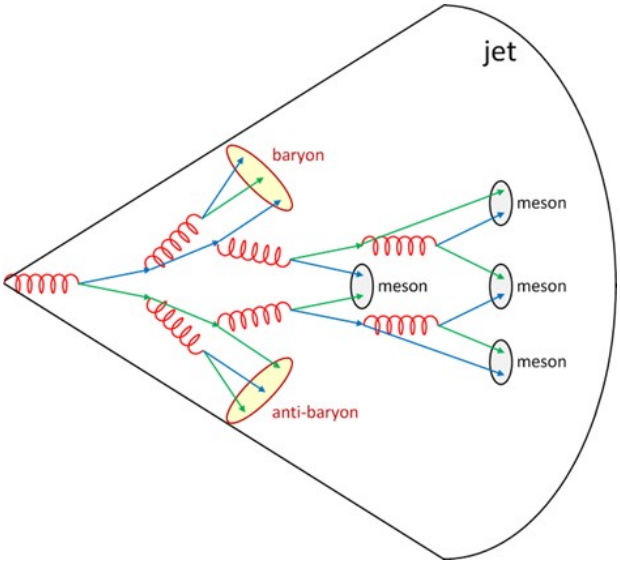


Figure 3.33: Illustration of the hadronization process [95].

Once the full event is PF-reconstructed, jets are formed by clustering charged and neutral hadrons. This is a particularity of the jets: jets are not simply reconstructed as are muons or electrons. Jet reconstruction implies to reconstruct each neutral and charged hadron as well as photons from neutral pion decays. Different clustering algorithms exist but the anti-k_T algorithm [96] is the one used in the searches presented in this work. To understand the anti-k_T algorithm it is more pedagogical to start from the k_T algorithm [97] that preceded it. This algorithm assumes the existence of what is referred to as "protojets", i.e. the smaller QCD objects in the detector be it individual particles or calorimeter clusters. They are the objects to be merged in jets. The k_T algorithm goes as follows:

- Define $d_i = E_{T,i}^2$ for each protojet i, where $\mathrm{E}_{T,i} = |\vec{p}_{T,i}|$.

- Similarly for all pairs i, j of protojets, define $d_{ij} = \min(E_{T,i}^2, E_{T,j}^2)\frac{\Delta R_{ij}^2}{R^2}$, where $\Delta R_{ij} = \sqrt{\Delta\eta_{ij}^2 - \Delta\phi_{ij}^2}$ and R is a parameter to be tuned. It should be chosen to be of order 1.
- Find the smallest of all d_i and d_{ij}. If it is a d_{ij}, cluster protojets i and j in a new protojet by summing their E_T and define a weighted sum of their angular coordinates. If it is a d_i, remove it from the list of further useable protojets: it is not mergeable.
- Repeat until no protojets are left.

This algorithm clusters soft protojets together to form protojets of higher and higher energies until the final jets are assembled. The difference of approach of the anti-k_T algorithm lies in the redefinition of d_i as $1/E_{T,i}^2$ and d_{ij} as $\min(1/E_{T,i}^2, 1/E_{T,j}^2)\frac{\Delta R_{ij}^2}{R^2}$. With this definition, the clustering happens the other way around. Hard particles accumulate all neighbouring soft particles around them to form a final jet of conical shape and radius R. The output of the two algorithms on simulated jets is displayed in figure 3.34.

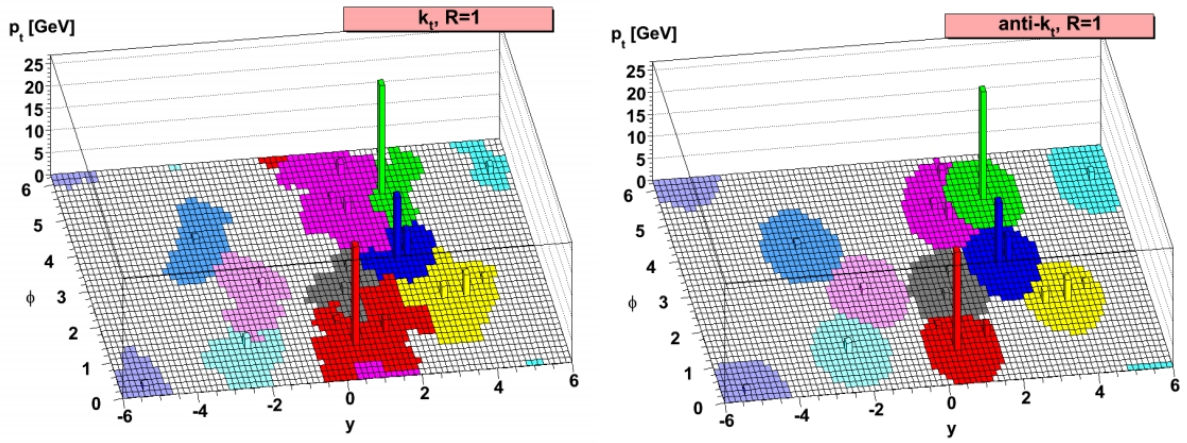


Figure 3.34: Results of the k_T (left) and anti-k_T (right) jet clustering algorithms on simulated jets with R =1 [96].

In CMS, the identification of jets is done partly with the help of the fraction of energy carried by charged and neutral PF candidates in the jet. More details on the performances of noise jet rejection in CMS data at 13 TeV can be found in reference [98]. Figure 3.35 shows the jet energy response and resolution on particle-flow and calo-jets. Calo-jets are jets built of HCAL and ECAL information only. The jet response is defined as the mean of the fitted Gaussian to the $(p_T^{\text{gen}} - p_T^{\text{rec}})/p_T^{\text{gen}}$ distribution.

3.3.8 Missing transverse energy

Since a significant part of this thesis was dedicated to the study of the missing transverse energy in the CMS detector, chapter 4 is dedicated to this topic.

3.4 The trigger system and data acquisition

As described in section 2.2, the LHC collision rate is 40 MHz. It corresponds to the inverse of the 25 ns separating two proton bunches in a beam. However, not all of those collisions can be stored for offline analysis. The first reason for this is that it is technically not possible to store and process all this information. Given that each collision corresponds to around 1 MB of data, one expects 40 MHz×20 pileup collisions×1 MB = 800 TB/s of data bandwith. The CMS recording rate must be limited to approximately 1 kHz so that the storage bandwith stays of the order of 1 GB/s, a value withstandable by realistic hardware. On top of this first limitation, it must be

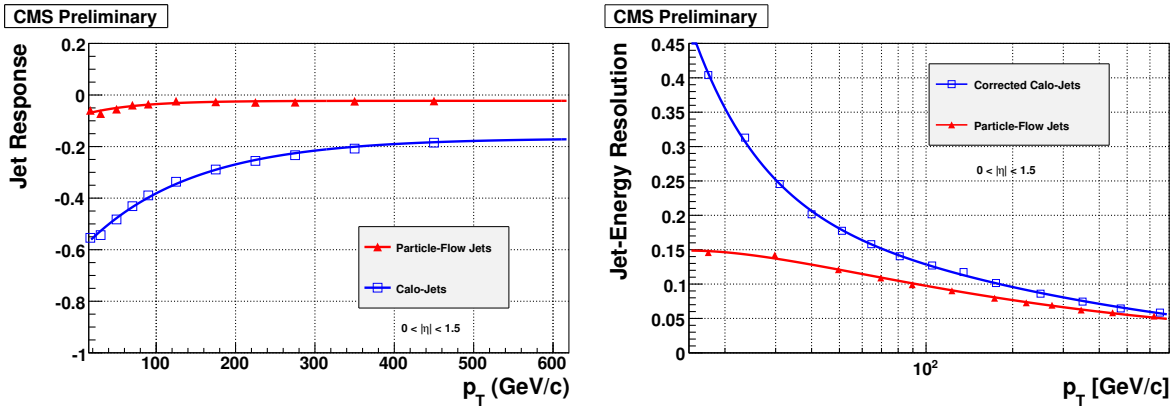


Figure 3.35: Jet response (left) and energy resolution (right) in the barrel [99]. On the right, calo-jets are corrected with correction factors derived such that simulated jet energy matches observations.

aknowledgrd that not all collisions are worth looking at. In figure 3.36, the pp total cross-section is shown to be orders of magnitude larger than that of rarer electroweak processes and high p_T jet production. The role of the CMS trigger system is to select events of interest within a decision time compatible with the LHC collision frequency. A buffer system in the readout electronics allows this time to effectively be larger than 25 ns. Selection can be done by different means: p_T thresholds, η range restriction, tighter identification criteria or the application of what we call "prescales". A prescaled trigger selects 1 event every p events that pass its selection criteria, p being the prescale factor (or simply the prescale).

The CMS trigger system design is made of two tiers: the Level-1 trigger (L1) and the High-Level trigger (HLT). The L1 trigger system reduces the input rate of 40 MHz to about 100 kHz to be processed by the HLT which reduces it further to 1-1.5 kHz.

3.4.1 Level-1 trigger

The L1 trigger is the first of the two trigger tiers. It is a hardware-based decision system that uses programmable electronics such as FPGAs (Field Programmable Gate Arrays). Due to the way the data pipeline is organized before being sent to the HLT, the L1 can only store information from 128 consecutive bunch-crossings. This represents a realistic processing time of about 1 μs. This time does not allow for a full, fine-grained, readout of the detector information: only coarse-grained information from the muon system and the two calorimeters participates to the L1 decision to accept an event or not. Each of the participating subdetectors generates local trigger information in the form of Trigger Primitives (TP). Let us list the TP for the main subdetectors:

ECAL ECAL cells are grouped 5×5 (in $\eta - \phi$) to form trigger towers (TT). The sum of the E_T of all crystals in a TT is the ECAL TP. An extra bit is used to store information regarding a shower shape variable.

HCAL The HCAL geometry is such that behind every ECAL tower is an HCAL tower which is the finest HCAL information element. It corresponds to a $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ angular coverage. The HCAL TP is the E_T of the corresponding HCAL tower.

Muon system Depending on the muon subsystem, TP are sent to the regional trigger in the form of local track segments or hit patterns. Hit patterns are sets of hits spatially close enough to form a candidate track but not reconstructed as such yet.

The Regional Calorimeter Trigger collects the information from ECAL and HCAL TPs. It sums the energy in each ECAL and HCAL tower and then combines ECAL and HCAL information of adjacent towers to generate isolated and non-isolated e/γ (tracking is not used) candidates. The

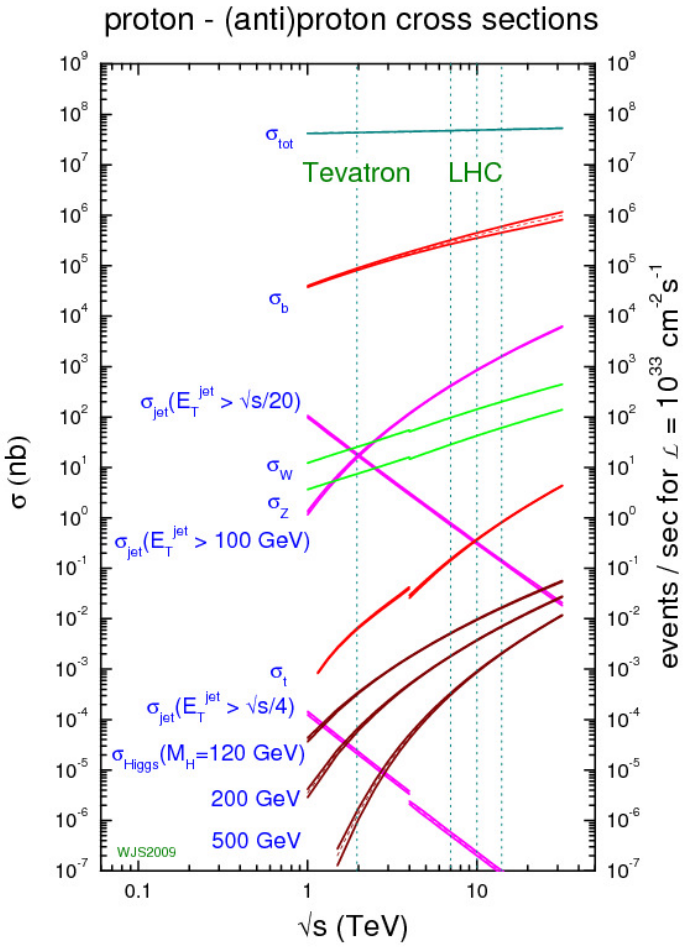


Figure 3.36: Proton-proton total cross-section as a function of the center-of-mass energy. Partial inelastic cross-sections for interesting processes are also given. The fact that the total cross-section is orders of magnitude larger simply reflects the fact that the pp cross-section is dominated by soft processes [100].

Global Calorimeter Trigger then creates jet candidates. The Regional Muon Trigger collects TP information from DTs and CSCs to build up candidate tracks with associated physical parameters. RPC TPs are used on their own to form candidate tracks. The Global Muon Trigger uses information from all three subdetectors to have the best possible resolution and efficiency on muon track reconstruction and identification. The Global Trigger makes the final decision using all CMS information. The L1 architecture is shown in figure 3.37.

A Level-1 algorithm dedicated to a given selection is often referred to as a "seed". An example of that is L1_SingleMu22 that selects muon candidates with a p_T threshold of 22 GeV. The word "seed" refers to the fact that the L1-selected events are then passed on to the HLT as input to its selection.

3.4.2 High-Level trigger

The High-Level trigger receives data from the L1 at a rate of about 100 kHz, i.e. on average every 10 μs. This allows for the trigger decision to use the full detector information. The event reconstruction at HLT level can take up to one second and is therefore less restricted in time as is the case at L1 level. The HLT reduces the output rate of the L1 trigger after "online" reconstruction of the events using a software-based decision framework. This reconstruction uses the same software as what is used "offline". "Offline" refers to the reconstruction used by the

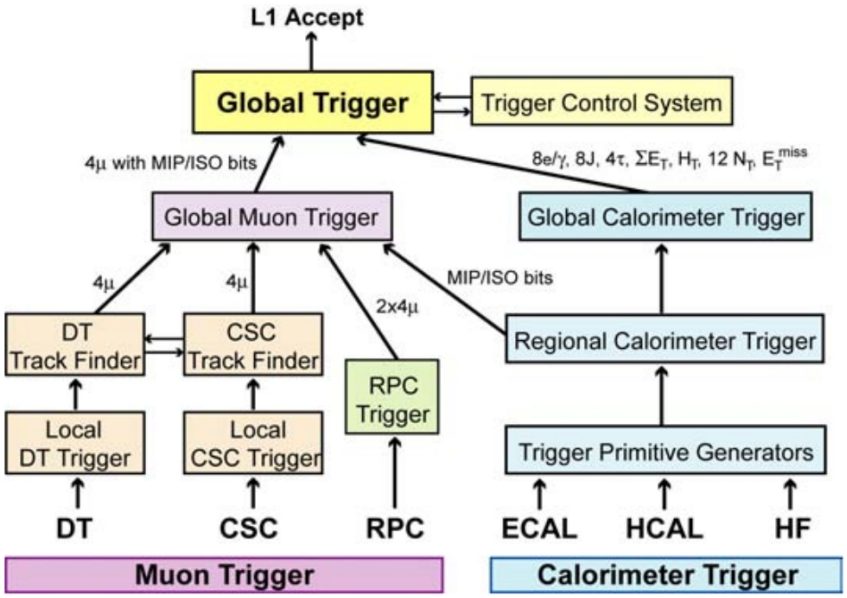


Figure 3.37: L1 trigger architecture [70].

analyzers while "online" refers to the HLT reconstruction applied before the decision to keep the event or not is made. Online selection criteria are applied to physics objects reconstructed in a way that resembles what is done offline and that we described in previous sections. Nonetheless, small differences exist to speed up the online process. We discuss here briefly the online reconstruction procedure for the main physics objects. More information on the HLT project can be found in reference [101].

Electrons/photons Electron and photon reconstruction and identification at HLT happens in three levels: first a reconstruction using calorimeter-only information. The energy is clustered in the ECAL, starting with local maximum of energy as seeds and agregating cells around them to form superclusters. An energy cut is then applied on the clustered objects. The following step is to use matching of hits in the pixel detector to the ECAL supercluster to discriminate electron from photon candidates. Finally, the full tracking information is used to reconstruct the electrons.

Muons The first step in the muon reconstruction consists in reconstructing standalone muon tracks in the muon system, refining the L1 p_T estimation. This is followed by the inclusion of tracker data to build global muons. The segment in the innermost muon chamber is matched to the hit in the outermost tracker layer taking into account possible energy loss of the muon through bremsstrahlung.

Jets Jet reconstruction at HLT uses a simplified version of the algorithms presented in section 3.3.7. Protojets are built out of calorimeter towers and are used as seeds to the jets energy clustering.

An HLT algorithm dedicated to a given selection is called a "path". An example of a path is "HLT_Mu50" that selects muon candidates with a p_T threshold of 50 GeV.

3.4.3 The STEAM group

For more than two years, I was involved in the *Strategy for Trigger Evaluation And Monitoring* group. It is a subgroup of the CMS *Trigger Studies Group* that is in charge of the rate monitoring of the HLT paths and the management of their prescales. Our duty was to make sure an HLT

menu (the full list of HLT paths) was sustainable and safe to be deployed online. The single-path rates as well as the global rate must be kept in a reasonable range (quantified in the introduction) for the continuously changing expected luminosity and pileup scenarios announced by the LHC. We were also responsible for rate predictions before data taking using simulations and previous data. These tasks were crucial at the beggining of Run 2, at a time when the LHC was ramping up in intensity and we did not know how the trigger rates would behave. We studied the evolution of the rates with the instantaneous luminosity and the pileup in simulations and in previous data to obtain predictions and tune the prescales accordingly. In parallel, I was regularly on shift in the CMS control room in charge of both the trigger and the data acquisition (DAQ) systems. These two activites were nicely linked and I could use my experience in STEAM when encountering issues as a trigger and a DAQ shifter.

Chapter conclusion

We presented the CMS detector, the reconstruction procedure and the trigger system. We started by going through the detector subsystems: the silicon tracker, the electromagnetic and hadronic calorimeters and the muon system. For each of them, we gave some details on the geometry, the detection principle and general performances. Based on this knowledge, we explained how physics objects are reconstructed using combined information from the subsystems. We defined what a muon, an electron, a photon, a tau lepton and a jet is for the CMS detector. Finally, the trigger system was introduced along with the challenges it has to address and the requirements it has to meet.

Chapter 4

The missing transverse energy

Motivated abstract

The missing transverse energy $\mathrm{E}_T^{\mathrm{miss}}$ is the measure of the momentum imbalance in the CMS transverse plane. As many theories of physics Beyond the Standard Model (BSM) predict the existence of experimentally invisible particles, missing transverse energy could be the key variable to invisible sectors of new physics. A good example of such invisible particles would be Dark Matter. Should DM exist in the form of particles, they would escape the detector unnoticed. This would translate into an excess of events in the $\mathrm{E}_T^{\mathrm{miss}}$ spectrum. To carry out this kind of searches, a good description and understanding of the expected amount of $\mathrm{E}_T^{\mathrm{miss}}$ from SM processes and instrumental effects is needed.

First, we define how the $\mathrm{E}_T^{\mathrm{miss}}$ variable is estimated in CMS. Then, we detail of missing transverse energy can arise in SM processes. We continue with the presentation of the experimental sources of $\mathrm{E}_T^{\mathrm{miss}}$ and we review the existing ways of dealing with them. We close this chapter with the suggestion of new ideas as to how one could recover lost energy.

4.1 Definition of $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ in the CMS detector

At the LHC, the energy of the protons is very well known. However, one cannot determine the fraction of it that is carried by their constituents (quarks and gluons) as these are not observable states. When a hard collision takes place inside the CMS detector, we know the centre-of-mass of the system is boosted along the z axis of the detector up to very small transverse fluctuations of the initial partons inside the protons. In any case, we have no a priori deterministic knowledge on the value of this boost. We only have information on the statistical distribution of the proton constituents energy through the measured PDFs. Thus, the only thing one can state is that by momentum conservation, the momentum sum in the transverse plane should balance to zero.

The missing transverse energy $\mathrm{E}_T^{\mathrm{miss}}$ (or more specifically, the PF $\mathrm{E}_T^{\mathrm{miss}}$) is defined as the amplitude of the missing transverse momentum vector $\vec{p}_T^{\,\mathrm{miss}}$. This vector is built as the opposite of the vectorial sum of the transverse momentum vector of all particle-flow reconstructed objects in the event:

$$\mathrm{E}_T^{\mathrm{miss}} = |\vec{p}_T^{\,\mathrm{miss}}| \tag{4.1}$$

$$= |-\sum_{\text{all PF-objects}} \vec{p}_T^{\,\text{PF-object}}| \tag{4.2}$$

Since energy is a scalar observable, the term "transverse energy" can be misleading. It is important to stress that the $\mathrm{E}_T^{\mathrm{miss}}$ is computed from momenta but is a scalar variable. To evaluate

the E_T^{miss} response and resolution, events with no genuine E_T^{miss} are used such as Z(ll) or photon + jets events. This is possible thanks to the fact that leptons and photons are very accurately measured in CMS. The principle is sketched in figure 4.1.

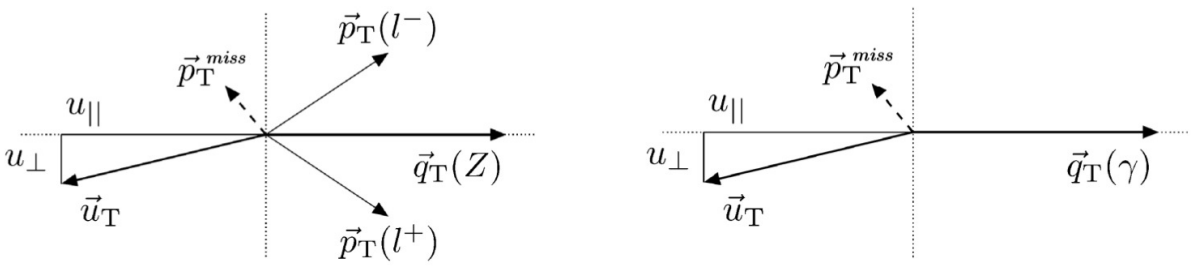


Figure 4.1: Z(ll) + jets (left) and γ + jets (right) events where $\vec{q}_T$ is the boson transverse momentum and $\vec{u}_T$ is the hadronic recoil, i.e. the transverse momentum of the jet system in the event [102]. The recoil is decomposed in the $u_{||}$ and $u_\perp$ components respectively parallel and perpendicular to the boson $\vec{q}_T$.

The boson p_T (here denoted as q_T) is compared with the hadronic recoil defined as the vectorial sum of all jets in the event. If no instrumental source of E_T^{miss} nor any energy mismeasurement enters the event description, the component of the recoil parallel to the boson $\vec{p}_T$ should be exactly opposite to the boson $\vec{p}_T$ with the same amplitude. This is why a good variable to estimate the E_T^{miss} response is the ratio of the means $\langle u_{||}\rangle/\langle q_T\rangle$. The corresponding resolution on each component is defined as the RMS on its distribution. The response as well as the resolution on the parallel component are given in figure 4.2.

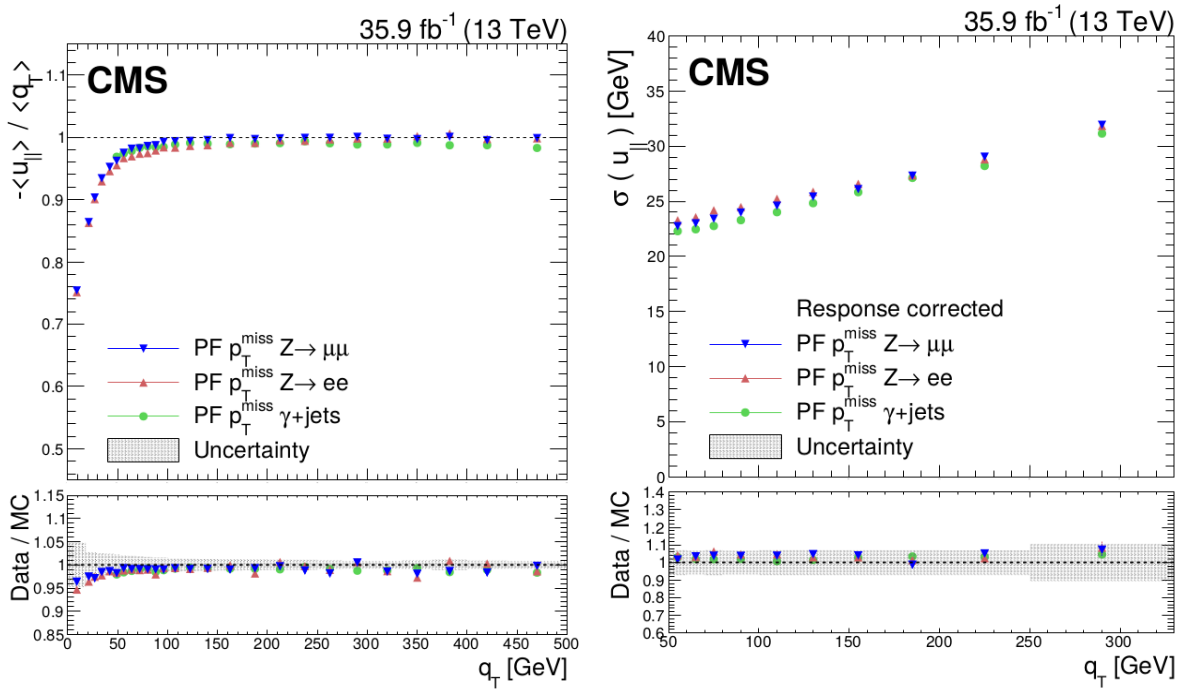


Figure 4.2: E_T^{miss} response (left) and resolution on the parallel component of the boson vector (right) for Z(ll) and γ + jets events in the full 2016 dataset [102].

The algorithms used to reconstruct the missing transverse energy are sensitive to the pileup. As the number of pileup collisions grows, it becomes increasingly harder to decouple their contribution to the total energy of the event from the one of the hard collisions of interest. Figure 4.3 shows how the parallel and the perpendicular components of the recoil evolve with the num-

ber of reconstructed vertices for two different algorithms. The PUPPI (PileUp Per Interaction) algorithm has been designed to mitigate the effect of pileup collisions on the computation of the E_T^{miss}. It is described in reference [102].

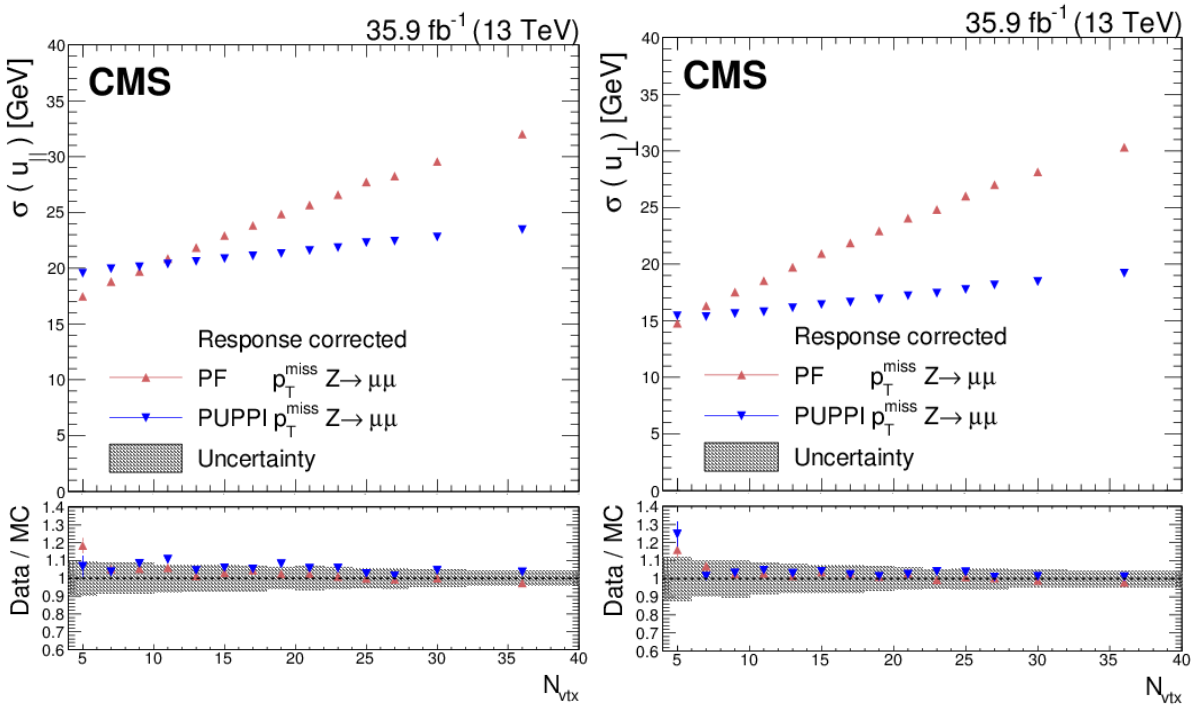


Figure 4.3: Evolution of the parallel (left) and perpendicular (right) components of the recoil as a function of the number of reconstructed vertices in the event for two different E_T^{miss} algorithms in the full 2016 dataset [102].

We did not mention that in the "reconstruction" section but the energy of the jets is corrected for several experimental effects such as pileup interactions. These Jet Energy Corrections (JECs) are propagated to the missing transverse momentum vector hence to the E_T^{miss} and are referred to as "type-I" corrections.

4.2 E_T^{miss} in Standard Model processes

The neutrino-nucleon cross section is so low that those particles escape the detector undetected. For a 100 GeV muon neutrino, a typical value for the charged current (CC) cross section is $\sigma_{CC}^{\nu_\mu - N} \approx 1$ fb [29], i.e. a path length of $\sim 10^{16}$ m in lead. This distance is over a light-year.

Neutrinos are produced in weak interactions through the decay of the weak vector bosons, as depicted in figure 4.4. These diagrams are not to be interpreted standalone: the bosons are either produced via the interaction of the proton's constituents (hard process) or they may appear later in decay processes. The main SM processes leading to a significant amount of E_T^{miss} are:

- Z($\nu\nu$)
- W($l\nu$)
- Diboson processes (WW, WZ, ZZ) where either one or both bosons decay to neutrinos
- $t\bar{t}$ where the W from the t(bW) decays leptonically

Neutrinos can also arise in the decay of heavy flavour mesons such as the $D^\pm - D^0$ and the $B^\pm - B^0$. The expected amount of E_T^{miss} coming from SM processes is modeled by MC simulations

or estimated from data-driven methods. Since the E_T^{miss} is particularly affected by a number of experimental effects, the latter is often prefered.

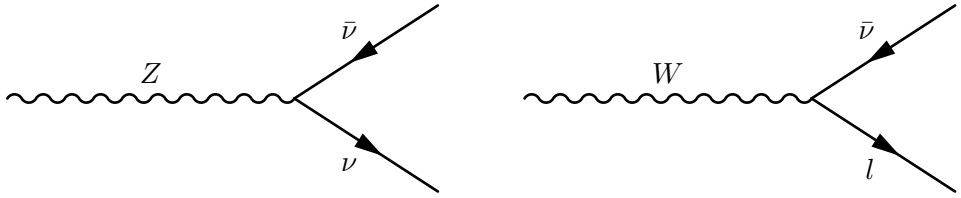


Figure 4.4: Production of neutrinos through the leptonic decay of the SM weak bosons.

4.3 Experimental sources of E_T^{miss}

On top of the expected missing transverse momentum from neutrinos escaping the detector, E_T^{miss} can arise from experimental sources. Particles produced outside the detector but leaving a signal in CMS (external, or non-collision events) is a first example of situation that must be dealt with. Instrumental issues such as electronics noise can also lead to unbalance in the transverse plane. Finally, a failure of the reconstruction algorithms can also enter the picture.

4.3.1 Non-collision events

CMS is subject to mainly two sources of E_T^{miss} from external particles: beam halo and cosmic muons.

Beam halo

Protons in the LHC beams undergo oscillations in the transverse plane. This means that the beam has an effective non-zero transverse section (see figure 4.5). The protons in the tails of the beam profile form what is called *primary beam halo*. To protect the beam pipe elements (typically the superconducting magnets) and the detectors from those peripheral particles, collimators are installed in the LHC ring. In their interaction with the collimators, primary beam halo particles produce secondary particles, such as pions and other light hadrons. Those particles then decay to produce what is called secondary beam halo, i.e. mainly muons. This is what analysts usually refer to as beam halo particles, or simply halo. The LHC collimation system is illustrated in figure 4.6.

A beam halo event detected by CMS is shown in figure 4.7 (left). From this figure, one can understand how beam halo can be a source of E_T^{miss}: the signal left by the muon in the calorimeter mimics a jet, unbalanced in the transverse plane. The measured energy can be due to ionization (at low energy) or bremsstrahlung (at high energy) of the muon in the material but can also be the consequence of the muon hitting directly the electronics of the calorimeter readout. In the latter case, the cluster energy can reach very high (sometimes unphysical) values.

Since beam halo muons are minimum ionizing particles, the rate of beam halo events (i.e. of a muon losing energy by ionization in the calorimeters) is much lower than the collision rate. Furthermore, even when a muon leave a signal large enough to be detected in the calorimeter, it may not trigger due to quality requirements (e.g. on timing and shape) already at the trigger level. In the same way, muons produced in collisions do not often leave energy in the calorimeters but they are reconstructed using the tracker and the muon chambers. Moreover, halo muons being produced in the beam pipe (i.e. potentially further away from the detector as compared with the center of the detector), their signal in the detector tends to be out-of-time with respect to the bunch crossings and timing requirements are one of the ways to get rid of those undesired events.

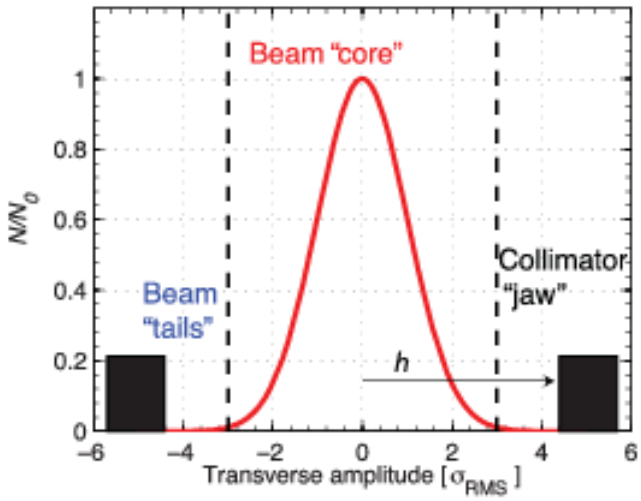


Figure 4.5: LHC beams section can typically be described with a Gaussian profile. Collimators (black rectangles on the figure) are used to intercept off-axis particles [103]. As a consequence, high-energy muons are produced parallel to the beam that can cross the detector. Due to the collimator geometry, this is preferentially observed at azimuthal angles $\phi = 0$ and $\phi = \pi$.

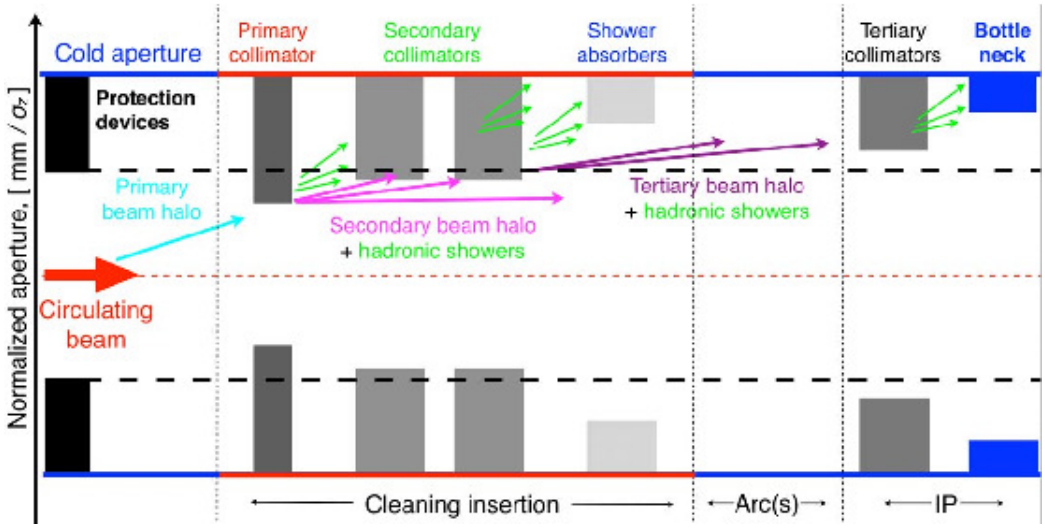


Figure 4.6: Sketch of the LHC multistage collimation system [103].

Cosmic rays

High energy protons from extraterrestrial sources interact in the high atmosphere, produce intermediate hadrons that decay to stable particles (electrons, muons, neutrinos). The only particles that live long enough, and are able to cross the hundred meters of earth above CMS are muons and neutrinos. Of course, neutrinos continue their way while muons can interact in our detector. Similarly to the effect of a beam halo muon, a cosmic muon can leave a signal in the calorimeters that can be misunderstood as a jet, as shown in figure 4.7 (right). The time at which they arrive in the detector is random and is used to discriminate them against the time of the pp bunch-crossings.

4.3.2 Instrumental issues

Electronics noise

In both the electromagnetic and hadronic calorimeters, large anomalous signals due to noise can mimic physical signals due to the passage of a particle. As a consequence, this generates

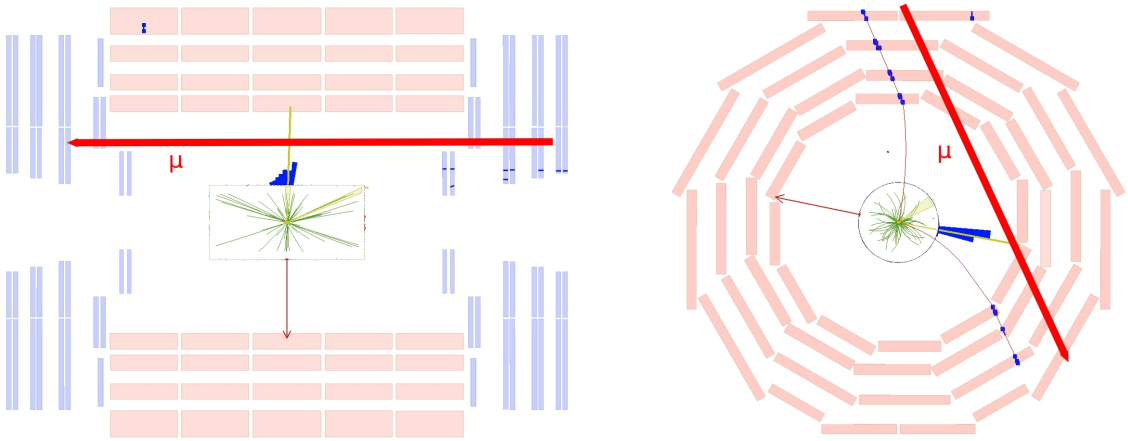


Figure 4.7: Left: beam halo muon passing through the muon chambers in the endcaps (CSC, in light blue on the right) and the hadronic calorimeter (dark blue blocks in the centre). Right: cosmic muon passing through the muon chambers in the barrel (DT, in red) and the hadronic calorimeter (dark blue blocks on the right side of the figure). The thin arrow indicates the orientation of the missing transverse momentum. The thick red arrow indicates the direction of the muon. These events were recorded in 2016.

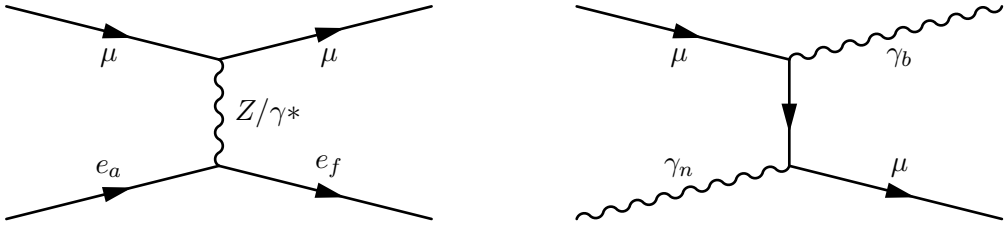


Figure 4.8: Left: ionization of matter by a high energy muon. e_a stands for an *atomic* electron and e_f for a *free* electron. Right: emission of a bremsstrahlung photon (γ_b) by a high energy muon in the electromagnetic field of an atomic nuclei (γ_n).

unphysical large missing transverse energy in the other direction. We will give a general description of the nature of this noise and its consequences. Details can be found in references [104, 105].

In the ECAL, this phenomenon is observed in the Barrel region where avalanche photodiodes (APD) are used as photomultipliers at the rear of the scintillating crystals. An unphysical energy can be recorded when a particle hits directly the APD instead of converting its energy to photons in the $PbWO_4$ crystal. This high, local signal is often referred to as a "spike". This leads to two major issues: a high contribution to the photon and electron trigger rates (online issue) and a misdescription of the events passing those triggers (offline issue). Unlike actual photons and electrons leaving their energy in several crystals, in this anomalous situations the signal is located in a single crystal. A dedicated algorithm is therefore able to make use of this information to reduce > 95 % of the fake rate coming from this source. Offline, the timing of the signal pulse is also used as a discriminating criterion. Indeed, since a spike does not produce scintillation light but hits directly the APD, it appears "earlier".

Noise in the HCAL is of different nature depending on the subdetector. Let us first briefly describe the sources of noise in the HB and HE:

- "Ion feedback": a thermally emitted electron generates an avalanche in one channel of the photomultipliers (Hybrid Photodiodes, HPD).
- "HPD noise": A misalignment between the CMS magnetic field of the detector and the electric field of the HPD changes the HPD voltage and can create large signals in multiple

channels of the HPD.

- "RBX noise": Electronics noise in a read-out box (RBX, front-end electronics that reads the signal from 4 HPDs) can generate a signal in multiple HPDs.

Depending on its impact of the physical objects in the event, noise in the HCAL is either cleaned from the event, or the entire event is filtered out.

Holes

In the ECAL, the information from a crystal can be degraded for several reasons. Either the crystal itself is damaged (by radiation for instance), or the readout electronics does not work as desired. The degradation can be partial or complete and the information can sometimes be recovered. Let us first remind that in the L1 trigger, ECAL crystals are grouped in 5×5 blocks to form Trigger Towers (TT, see section 3.4) as depicted in figure 4.9. Two distinct links are coming out of the readout: the full data information and coarser information to be used for trigger.

Different situations leading to missing energy in a given crystal exist that may or may not be related to trigger:

- Saturated channel: the energy deposited in a crystal is higher than what the electronics can record. As a consequence, the output energy is set at the highest possible value the data link an handle, i.e. 1.5 TeV.

- Saturated TP: the TT corresponding to a crystal is saturated. Similarly to what happens with the data link, the output energy is set at the highest possible value the trigger link can handle, i.e. 127.5 GeV.

- Dead: a single crystal channel is unreadable.

- Dead VFE: a very-front-end card (1×5 crystals) is unreadable.

- Dead FE: a front-end card (5×5 crystals) is unreadable.

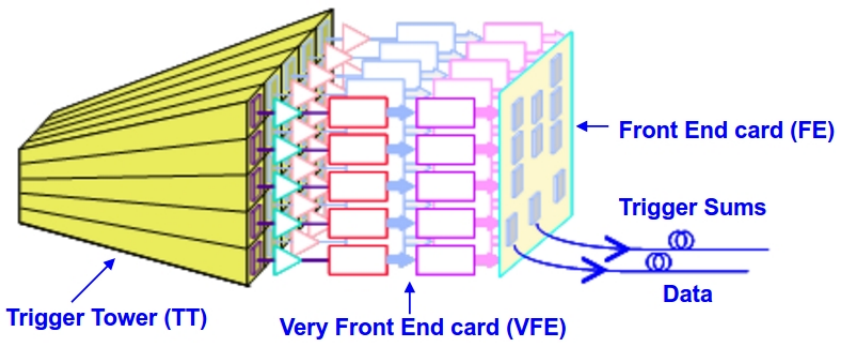


Figure 4.9: Front-end electronics in the ECAL (see reference [106] for example). Crystals are grouped 5×5 to form trigger towers. The output of the readout is twofold: one link for the full data and another for trigger.

Currently in CMS, the only used recovery method is called "TowerRecovered". It is applied in the case where the data link is dead but the trigger information of the TT is available. Each crystal is then assigned 1/25 of the TP energy. If the TP is saturated (see above), the event is filtered out (see section 4.4.1).

4.3.3 Reconstruction issues

On top of non-collision events and electronics noise, reconstruction algorithms can be a source of missing transverse energy. The transverse momentum of a muon track can be badly measured leading to both a high value and a high uncertainty on it. As a consequence, either the muon passes the PF requirements and enters the event picture as a muon, or it fails the PF requirements and the track is re-assigned to a charged hadron. In both cases, the unhealthy track adds large instrumental E_T^{miss} and the event must be filtered out (see 4.4.1).

4.4 Dealing with undesired E_T^{miss}

As many analyses make use of the missing transverse energy, being able to identify and deal with the undesired sources of E_T^{miss} is of major importance. Depending on the way an event is polluted by experimental E_T^{miss}, it can be cleaned or filtered.

4.4.1 Removing unhealthy events

As already mentioned in the previous section, there are ways to identify fake E_T^{miss}. In CMS, the E_T^{miss}-scanning working group is dedicated to the identification of events containing such spurious E_T^{miss}. It is responsible for the coordination between the analyzers and the experts who design the algorithms to reject those events. We call these E_T^{miss}-rejecting algorithms "E_T^{miss}-filters", or simply filters. We now go through each of them to explain their purpose.

"Beam halo" filter

This filter is constructed as a logical "AND" of two conditions:

- The calorimeter cluster has a halo-like pattern, i.e. out-of-time hits or a large η distribution in the barrel.
- This cluster is matched to segments in the CSC system.

The efficiency of the beam halo filter as a function of the E_T^{miss} is shown in figure 4.10. It is defined as the number of "good" events (i.e. not tagged as halo) divided by the total number of events. The data used has been recorded in 2017 and is enriched in jets, i.e. it contains a lot of calorimeter deposits which is useful to study the effect of the halo filter.

"HBHE noise" filter

This filter removes HCAL noise based on information from [105]:

- The channel multiplicity.
  - Too many channels recording energy within an HPD is a sign of "ion feedback" or "HPD noise".
- The number of zero ADC counts (Analog-to-Digital Converter that digitizes the signal from the RBX).
- The pulse shape in the RBX.

The efficiency of the filter is displayed in figure 4.11 (left).

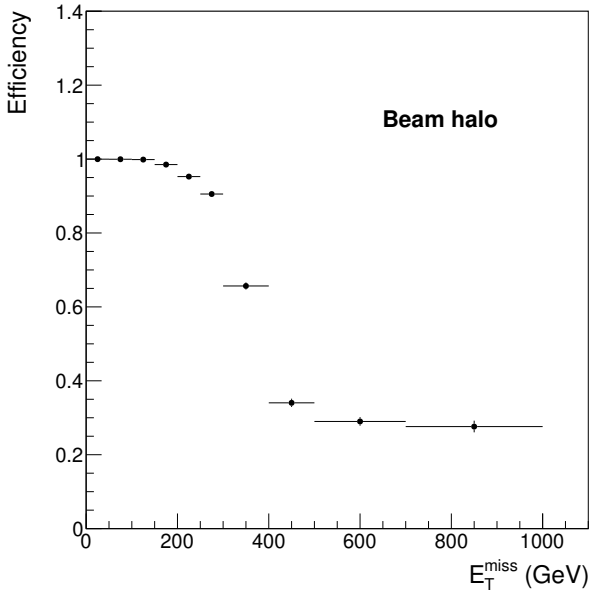


Figure 4.10: Beam halo filter efficiency as a function of the E_T^{miss}. Data has been recorded in 2017 and is enriched in jets.

"HBHE iso" filter

This filter targets the same events as the "HBHE noise" filter and is based on the same algorithm. However, it adds information from the ECAL and the tracker system as well as isolation criteria on the topology of the cluster in the HCAL to identify the problem. It increases the tagging efficiency of the "noise" filter from 98 % to more than 99% on jets-enriched data while reducing the mistag rate ($\sim$ 3 %). The mistag rate is the proportion of good events tagged as bad by the filter. This can be seen as the tagging efficiency of the filter at low E_T^{miss}. Indeed, the filter should not remove events with E_T^{miss} close to zero. Figure 4.11 shows the efficiency of the two HCAL filters. The better performance of the "iso" version is clear from the comparison of the two histograms.

"ECAL TP" filter

Currently in CMS, only one recovery method exists to deal with an event known to contain an undesired source of E_T^{miss}. This recovery procedure is used whenever the data link of an ECAL crystal is broken but trigger information (the trigger primitive TP) is available. The energy assigned to the crystal is then simply the energy of the TP divided by 25, i.e. the average energy per crystal. As a consequence, when the energy of the TP is recorded at the maximum possible value (127.5 GeV) it is impossible to know how strongly the energy is underestimated. Such an event is discarded by this filter. The efficiency of this filter is given in figure 4.12 (left).

"ECAL bad calibration" filter

This filter is based on a list of 17 ECAL crystals in the endcaps. Those crystals have large IC (intercalibration constants used to correct for the loss of transparency) that sometimes lead to anomalously large transverse energy. The algorithm is quite simple: if the transverse energy of a crystal in the list reaches 50 GeV, the event is rejected. The list is expected to grow with time as transparency gets worse in other crystals. The efficiency of this filter is given in figure 4.12 (right).

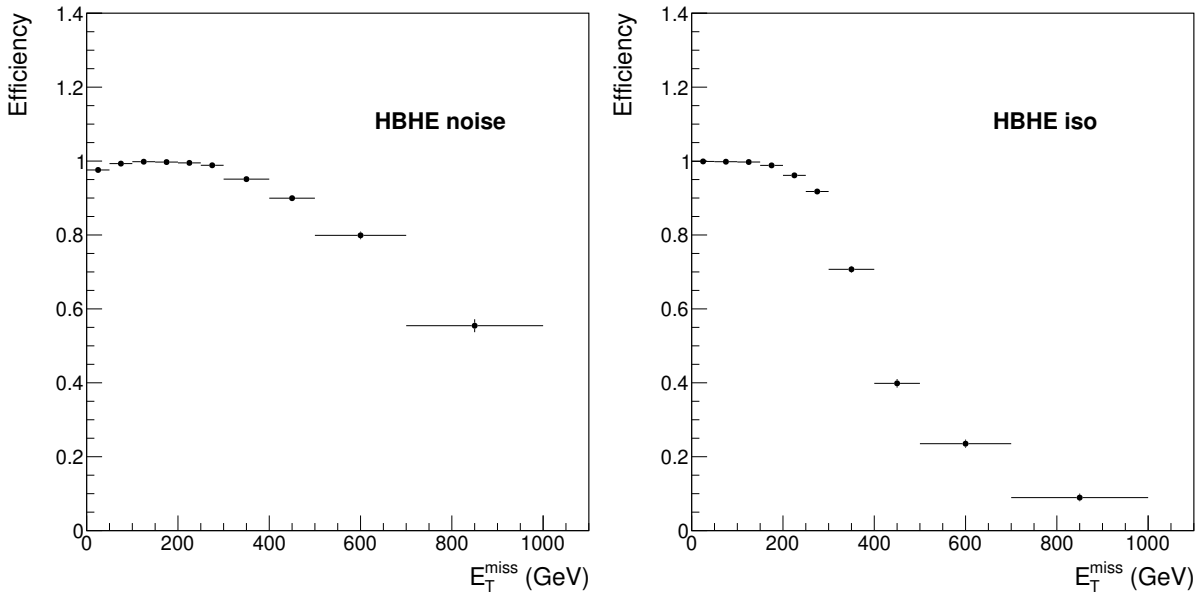


Figure 4.11: HBHE noise (left) and iso (right) filters efficiency as a function of the E_T^{miss}. Data has been recorded in 2017 and is enriched in jets. The better performance of the "iso" version of the filter is easy to observe by comparing the two histograms: the "iso" filter tags many more events at large E_T^{miss} while reducing the mistag rate (higher efficiency at low E_T^{miss}).

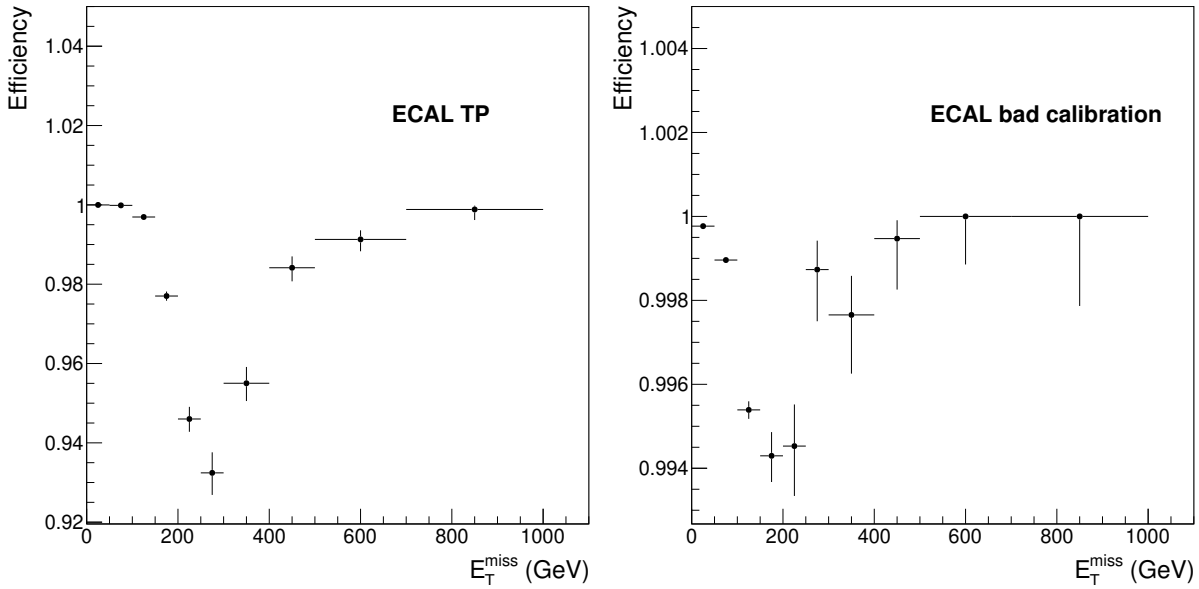


Figure 4.12: ECAL TP (left) and ECAL bad calibration (right) filters efficiency as a function of the E_T^{miss}. The dips in the efficiency curves are caused by the thresholds in the filters algorithms. The dip starts around 127.5 GeV for the TP filter and around 50 GeV for the calibration filter until it reaches a region where the E_T^{miss} is too large and is not caused by the target phenomena. Data has been recorded in 2017 and is enriched in jets.

"Bad PF muon" filter

This filter aims at ignoring events where a high p_T muon of low quality passes the PF requirements and enters the computation of the PF E_T^{miss} as a PF muon. Low quality is here understood as a large uncertainty on its transverse momentum. The filtering algorithm goes as:

- Look for muons with $p_T > 100$ GeV and large $\sigma(p_T)/p_T$
- If ΔR(muon,PF muon) < 0.00001: reject the event

The efficiency of this filter is given in figure 4.13 (left).

"Bad charged candidate" filter

In the PF reconstruction procedure, a track from a low quality muon not identified as a PF muon can be re-assigned to a PF charged hadron. Similarly to the previous situation, the potentially large p_T of this charged hadron can lead to large $\mathrm{E}_T^{\mathrm{miss}}$.

- Look for muons with $p_T > 100$ GeV and large $\sigma(p_T)/p_T$
- If ΔR(muon,PF charged hadron) < 0.00001: reject the event

The efficiency of this filter is given in figure 4.13 (right).

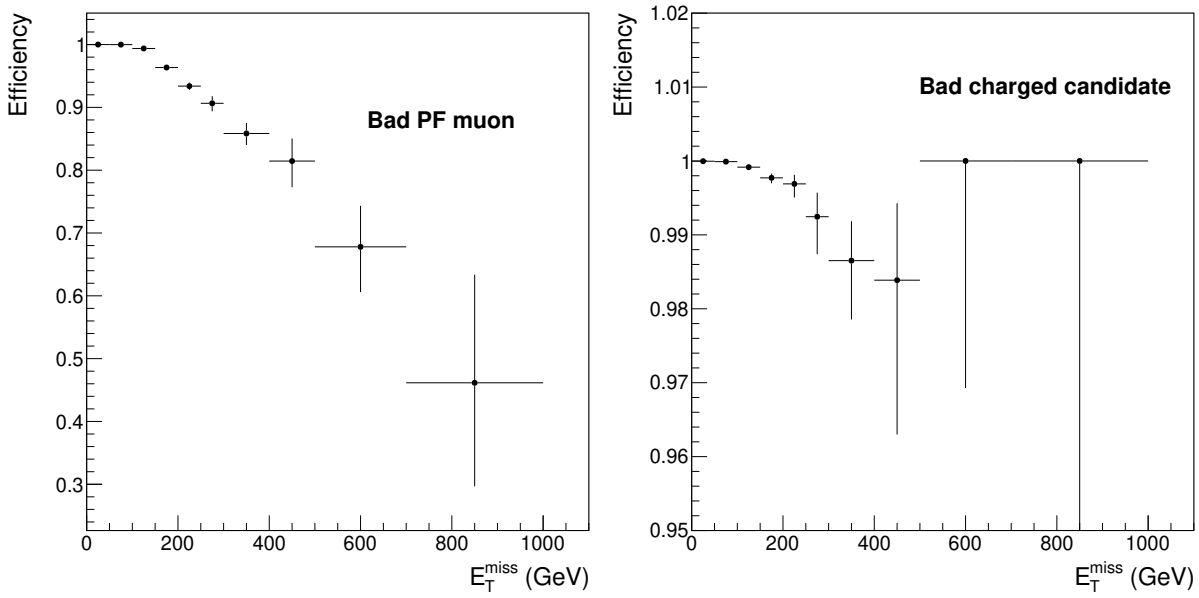


Figure 4.13: Bad PF muon (left) and Bad charged candidate (right) filters efficiency as a function of the $\mathrm{E}_T^{\mathrm{miss}}$. Data has been recorded in 2017 and is enriched in muons.

Figure 4.14 shows the $\mathrm{E}_T^{\mathrm{miss}}$ distribution before and after the application of the filters described above. From this figure, one understands the importance of the $\mathrm{E}_T^{\mathrm{miss}}$ cleaning when looking for new physics. Without the filters, the tail of the distribution is polluted by spurious sources of $\mathrm{E}_T^{\mathrm{miss}}$ that can potentially mask an excess of events from a signal of new physics.

4.4.2 Fixing unhealthy events

As explained in the section about the ECAL TP filter, one recovery method exists to estimate the energy of an ECAL crystal when the information from the data link is not available. However, it may be possible to estimate missing information in a given subdetector from the measurement in the surrounding subsystems. This would allow the analysts to fix the events in some way instead of rejecting them. In the context of my involvement in the MET group of CMS, I have carried out a project whose goal was to assign an energy to a non-responding ECAL crystal using the information from surrounding subdetectors. This had never been done before and was a project supported centrally by the CMS collaboration to improve the description of the tail of the $\mathrm{E}_T^{\mathrm{miss}}$ distribution. I will show here that this fixing procedure is actually not possible (in the limits of my work) on an event-by-event basisi as it leads to an enhanced $\mathrm{E}_T^{\mathrm{miss}}$ tail. To perform this study, QCD events simulated with the 2016 conditions are used. They contain simulated jets reconstructed with a realistic CMS, i.e. including the defective ECAL cells.

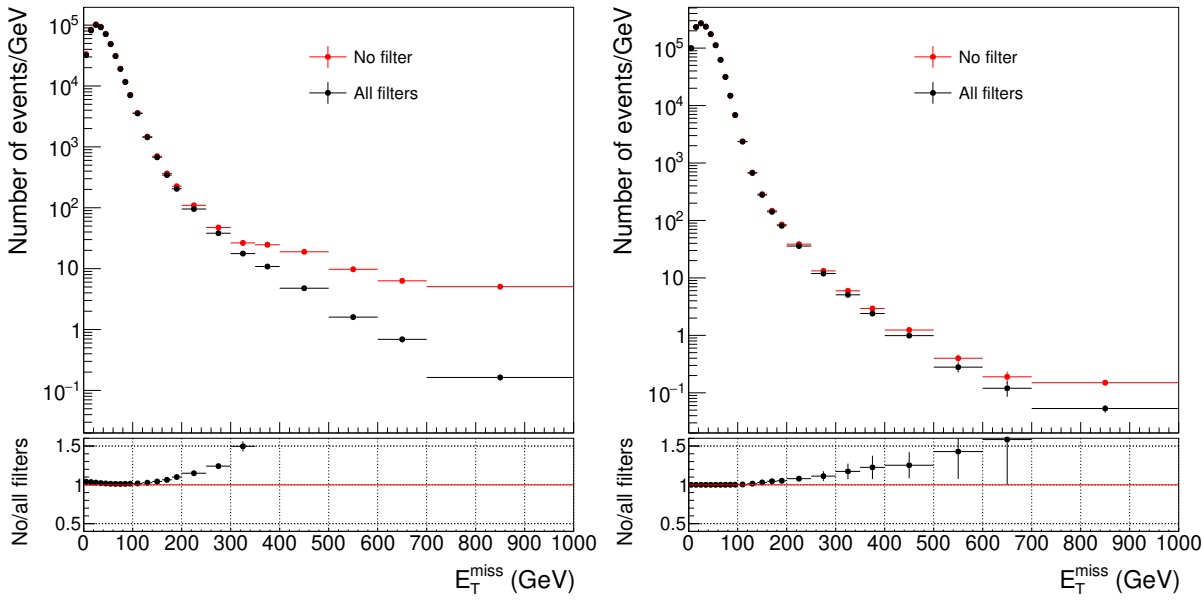


Figure 4.14: E_T^{miss} distribution before and after the application of the filters in a jet-enriched (left) and a muon enriched (right) dataset recorded in 2017. These are the same as the ones used for filters study. From these two histograms, one can observe that calorimeter-related instrumental issues lead to more instrumental E_T^{miss} than muon-related noise.

First of all, let us present the status of the non-responsive cells (or crystals) in the CMS ECAL. In figure 4.15 is shown the $\eta - \phi$ map of the crystals flagged as "TP recovered" (left) and "dead" (right). "TP recovered" crystals have a dead data link for the whole 5×5 region but an estimation of their energy is possible using trigger information (as described in the ECAL TP filter subsection above). Single "dead" crystals simply have zero energy assigned to them.

Figure 4.16 (left) shows the proportion of E_T^{miss} coming from energy lost in those non-responsive ECAL crystals. In green, are the events in which a jet hits a crystal of which the energy could be estimated from its TP and in light blue are the events in which a jet hits a crystal that is dead, i.e. which energy is set to zero and there is no recovery. The dark blue part of the stack accounts for all other sources of E_T^{miss}, i.e. mainly SM processes, HCAL noise or beam halo. On this figure, we see that the fraction of E_T^{miss} concerned by the problems of holes in the ECAL is around 10 % at high E_T^{miss}. However, when rid of events flagged by the ECAL TP filter, energy lost in dead ECAL cells becomes the main issue with a number of events reaching up to almost 20 % of the E_T^{miss} spectrum (figure 4.16 (right)). CMS analyses currently work with a E_T^{miss} distribution similar to that shown in figure 4.16 (right). This motivates further work on this matter.

Regarding the estimation methods, the basic idea is to investigate whether one of the two following sources of information can be used to determine the energy of a dead ECAL crystal:

- Tracks pointing towards the considered crystal
- Energy in the HCAL tower located behind the lost ECAL cell
- Neighbouring ECAL crystals (8 first neighbouring crystals, a priori)

Most of the jet electromagnetic energy is deposited by electrons and photons, the former being the ones with an associated track. By looking at figure 4.17, we see that a very small fraction of the jet energy is carried by electrons. Using tracks pointing at a dead ECAL cell to estimate the missing energy deposit is therefore hopeless.

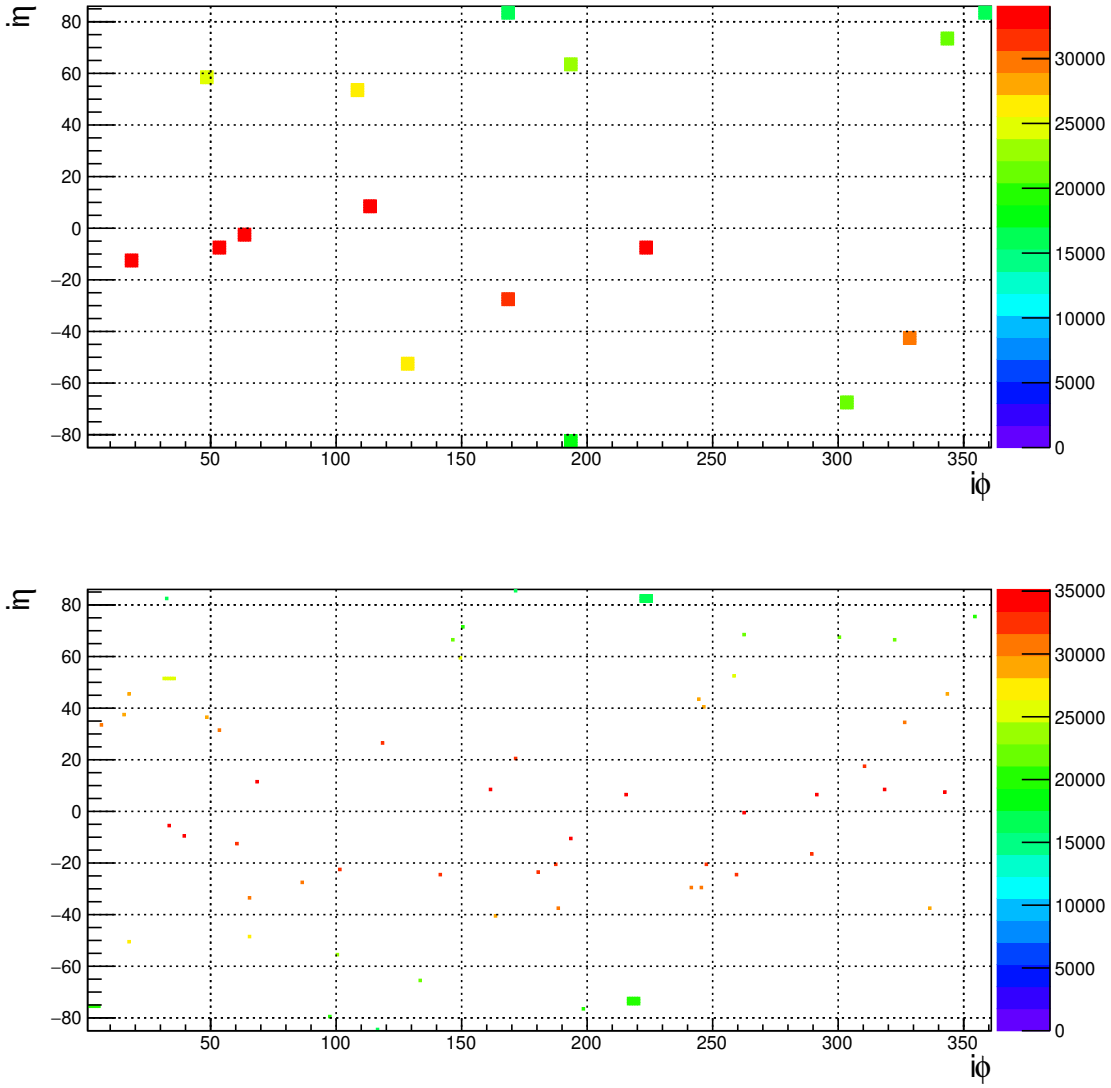


Figure 4.15: $\eta - \phi$ map of the ECAl crystals flagged as "TP recovered" (left) and "dead" (right). The coordinates of the crystals are not given in η and ϕ but as a function of their index iη and iϕ in these directions.

Another idea could be to use the energy of the HCAL tower located behind the lost ECAL cell in the jet. It may be that the energy of the highest-energy ECAL cell and the energy in the HCAL tower are correlated in some way. For instance, the average energy of the HCAL could scale with that in the ECAL in a jet. Figure 4.18 tells us that this is not the case.

We then try to use the neighbouring ECAL cells. We have good knowledge of the jet energy on average (as shown by figure 4.17 for instance) but the electromagnetic shower can develop with different shapes. This means using neighbouring cells is not trivial as the shape of the electromagnetic energy in the ECAL shows large fluctutations from jet to jet. This is illustrated in figure 4.19 where the energy pattern around the highest-energy ECAL cell is shown for four different jets.

Figure 4.20 shows the same map but in a 3D view for another jet. Each cell corresponds to one ECAL crystal. One can see that most of the energy is located in one cell. We have looked at the shape of the energy deposit but no simple pattern (such as a gaussian shape) could be found. This makes the recovery really hard: if most of the energy is located in a malfunctioning cell, the neighbouring cells do not bring much more information. Moreover, the energy difference between the highest energy cell and its neighbouring cells is very large. This means that the uncertainty

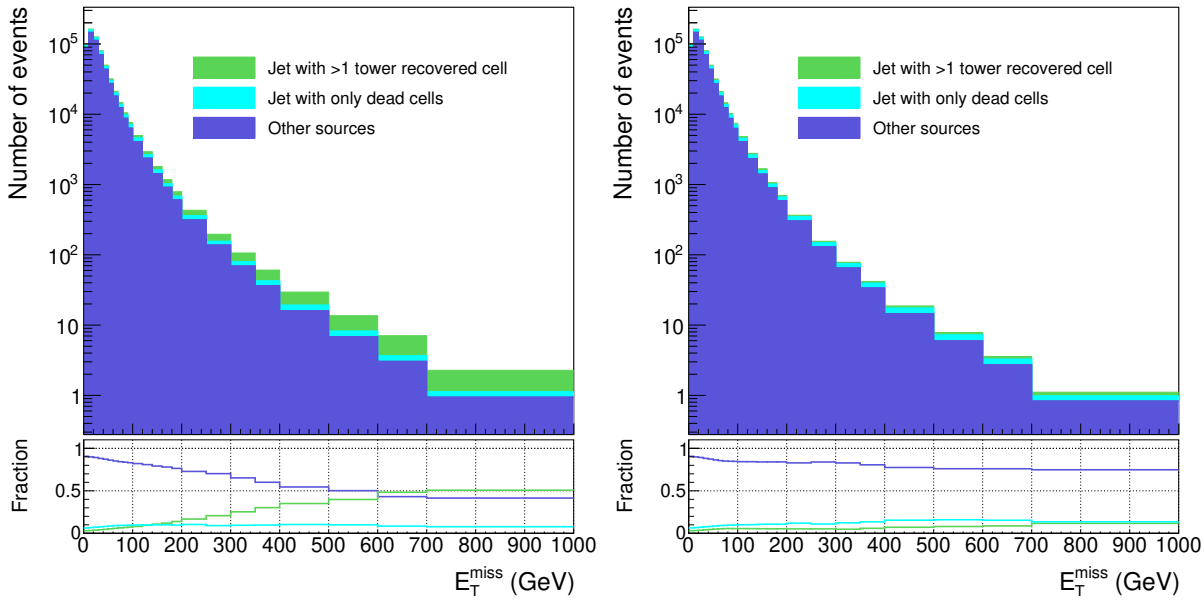


Figure 4.16: E_T^{miss} distribution in events containing at least one jet hitting a TP recovered region (gree), events containing only one or more dead cells (light blue) and events in which the E_T^{miss} comes from other sources. Those can be genuine E_T^{miss} from SM processes, HCAL noise, etc. The fraction of each source is given in the bottom pad. The distribution is shown before (left) and after (right) application of the ECAL TP filter. From the right figure, one can see that the single dead ECAL cells account for almost 20 % of the E_T^{miss} in the tail.

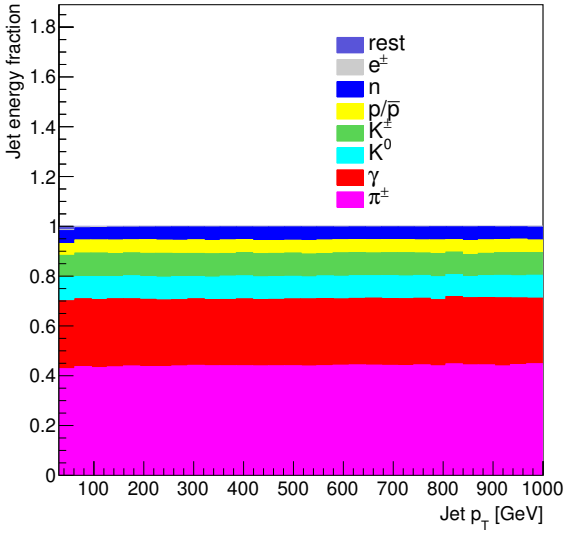


Figure 4.17: Distribution of the jet energy among its constituents as a function of the jet p_T in simulated jet-enriched events.

on the energy one would assign to the lost cell could potentially be very large as well. Therefore, there is no guarantee one would not be adding more energy than necessary. The whole point of the method, i.e. to reduce the E_T^{miss} in the event, would then not be met. On the other hand, if it is a peripherical cell that cannot be readout, the missed energy is small and its loss contributes very little to the E_T^{miss} tail.

After several presentations in working groups, including a workshop organized by the CMS E_T^{miss}-group with all relevant experts, we decided to stop the project. In agreement with the CMS working group conveners, we concluded that there was no obvious solution to this matter and

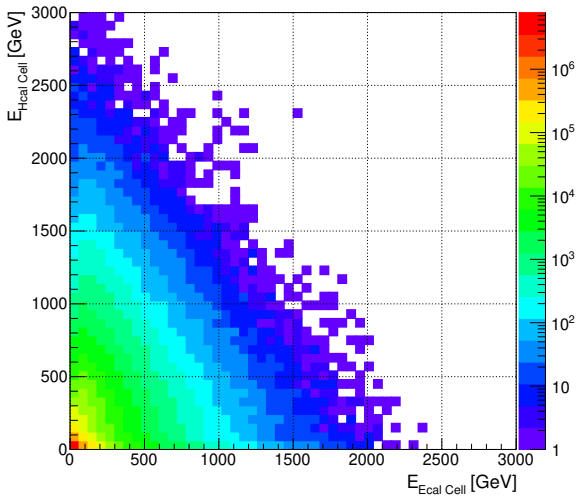


Figure 4.18: Energy of the HCAL tower (here denoted as "cell") located behind the highest-energy ECAL cell inside a jet as a function of this highest-energy cell. No clear useable correlation is observed.

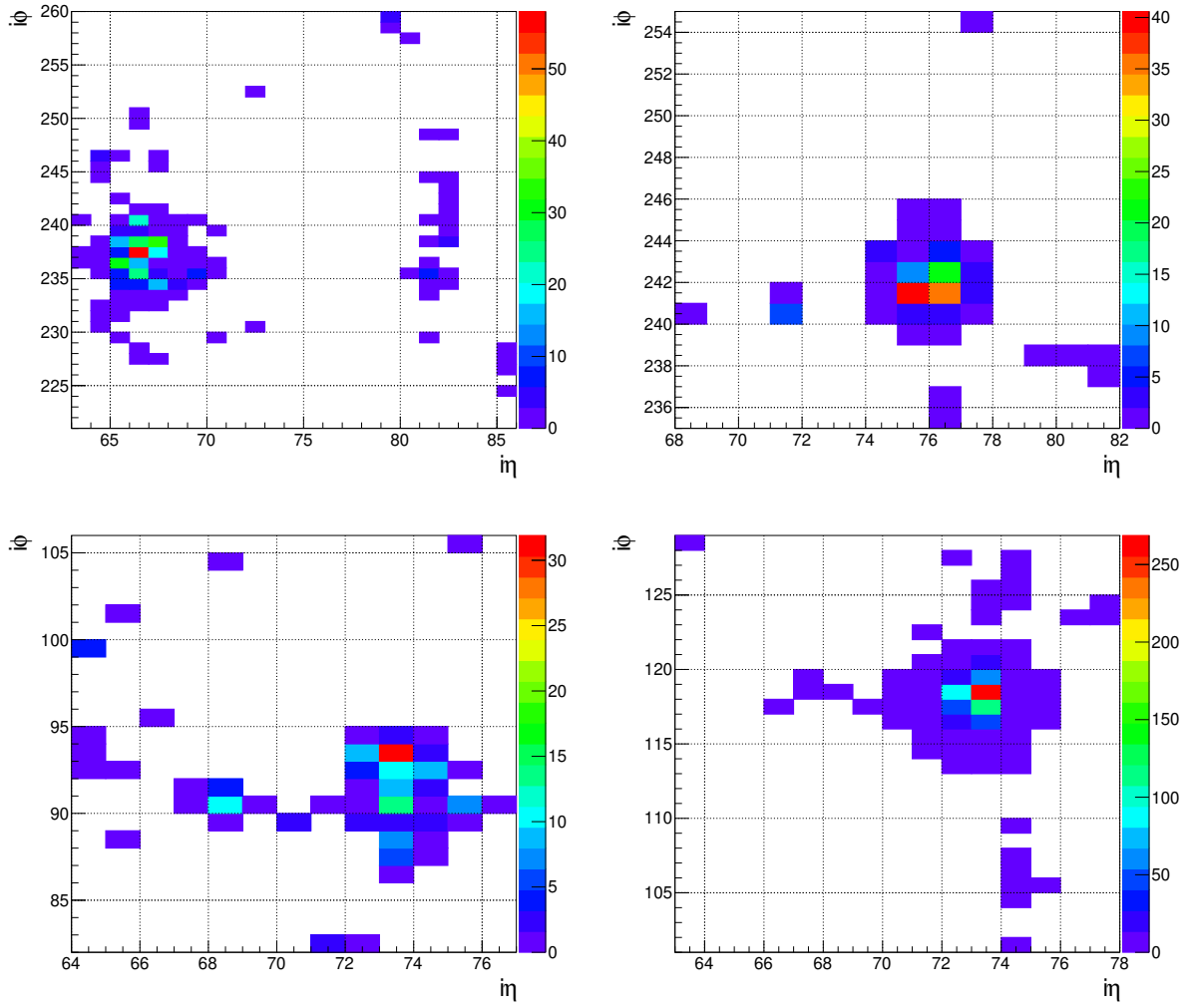


Figure 4.19: Jet energy deposit in the ECAL in the iη-iφ plane for four different jets with $p_T >$ 200 GeV. The colored scale for the z-axis is the energy in GeV. No clear pattern is observed. The electromagnetic shower in a jet is subject to large fluctuations both in energy and in shape.

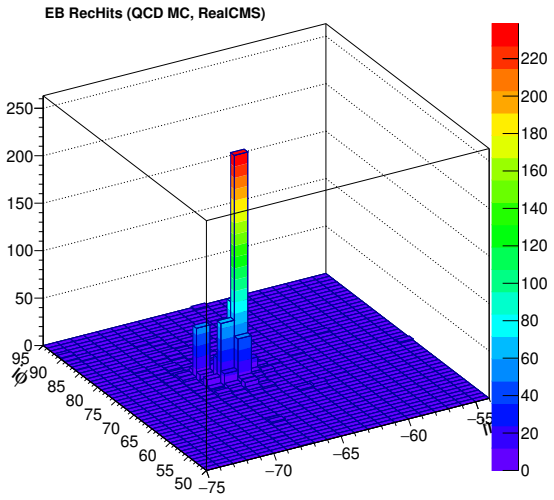


Figure 4.20: Jet energy deposit in the ECAL. The colored scale for the z-axis is the energy in GeV. Most of the energy is in one single cell.

that the goal was impossible to achieve with a reasonable amount of time and effort.

Chapter conclusion

In this chapter, we first defined the missing transverse energy E_T^{miss} in CMS as the norm of the vectorial sum of the transverse momentum of all reconstructed objects in the detector. It is a measure of the imbalance in the transverse plane that is useful to identify processes with experimentally invisible particles. Then, we explained that missing energy can arise from SM processes with neutrinos in the final state via the decay of the electroweak bosons. We continued by giving more details about the E_T^{miss} measurement techniques in CMS and how the response and resolution are evaluated using dilepton or photon + jets events. We showed that experimental effects such as non-collision events or instrumental and reconstruction issues can add E_T^{miss} to the overall event description. Finally, new ideas were presented to try to estimate the energy loss due to dead regions in the ECAL subdetector. We show that it is not a trivial endeavour but that with more time and effort, something might be tried using information from neighbouring ECAL crystals.

Chapter 5

Search for dark matter particles

Motivated abstract

As introduced in section 1.2.3, the quest for Dark Matter (DM) is one of the main challenges of physics in the 21^{st} century. Through gravitational effects, we see evidence that luminous matter cannot be the only type of matter in the universe. Another type of matter must exist that interacts only weakly with standard matter. Should it exist in the form of particles, it may be possible to produce DM at the LHC. Finding such particles could explain many of the deviations between SM predictions and cosmological observations. Now that the Higgs boson has been discovered, Dark Matter is one of the main targets of the physics program of the CMS collaboration.

In this chapter, we start by introducing the framework of simplified models. Simplified models are used at the LHC to allow for a clear interpretation of the results in terms of couplings and mass scales via the introduction of mediators. We then present two CMS analyses looking for Dark Matter. The so-called "Monojet" analysis assumes the recoil of a jet against a pair of DM particles, hence its name. We give the results of this analysis in terms of upper limits on the production cross-section of DM particles in CMS and we compare it to other experiments. In the second analysis, we look at invisible decays of the Higgs boson and interpret the results in terms of DM particles production. The focus is set on the computation of k-factors to constrain the background estimation procedure.

5.1 Simplified models

Searches for Dark Matter can be approached from two ends in terms of model simplicity. On one end are effective field theories (EFT) that consider the SM-DM interaction as a contact interaction. The details of the physics mediating this interaction is not described in full. This is equivalent to saying that the energy scale of the intermediate states is not reachable by the LHC. Those states are integrated out. On the other end of the spectrum are complete theories. They describe the new physics in full, often involving many parameters and new particles. Supersymmetrical models and Grand Unified Theories belong to that category. Between those two, a third path is possible using simplified models. In this case, a few new parameters and particles are added to the SM to model the phenomenology we are interested in for a certain specific experimental signature. This keeps the description simple while allowing for a general and rich physics representation. Figure 5.1 illustrates these three approaches. Effective field theories were commonly used in DM searches in the early years of the LHC when the center-of-mass energy was at 7 or 8 TeV. Already at that time, this framework was known to be oversimplified and to yield overly good limits. With the center-of-mass energy reaching 13 TeV, it was pointed out that the intermediate states between SM and DM physics might not be unreachable anymore. A more complex description of DM particles production would render a more proper image of the

underlying physics.

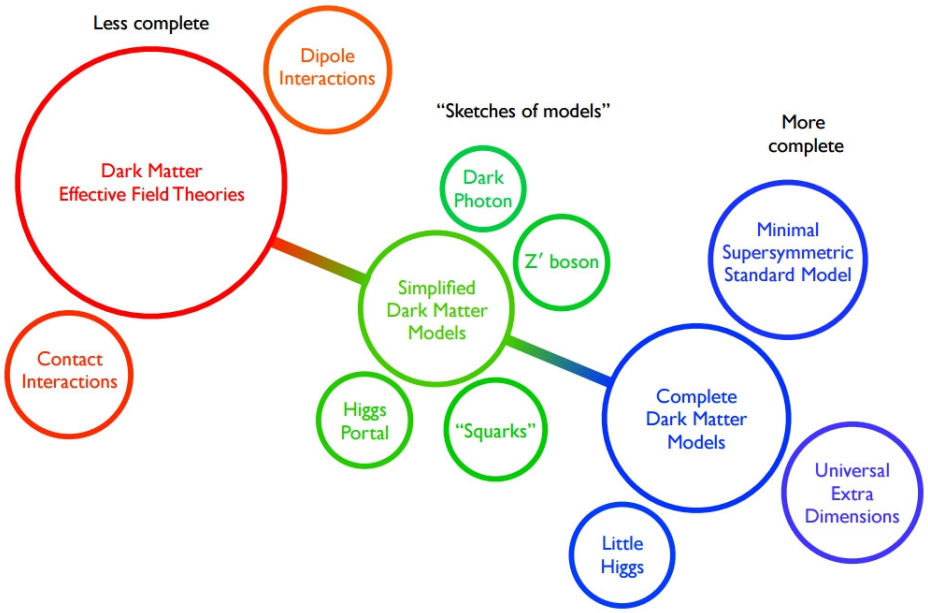


Figure 5.1: Spectrum of the DM theories in the "simplicity" space [107]. Effective field theories integrate intermediate states between the SM and the dark sector while complete theories describe the full physics. Simplified models are the bridge between the two.

In the analyses presented in this chapter, we intepret the results in the context of simplified models. A detailed description of the effort to use simplified models at the LHC is described in references [108, 107]. Let us give here some important aspects of these papers. First of all, in the framework of the simplified models used in CMS, the coupling of DM particles to SM particles is assumed to follow the same structure as in the SM. This condition is met if we impose Minimal Flavour Violation (MFV). In the SM, the Yukawa terms that give a mass to the quarks after Spontaneous Symmetry Breaking (SSB) of the Higgs field violate the flavour symmetry [109]. MFV requires these terms be the only source of the flavour symmetry breaking. An important consequence of this is that in a model with MFV, all scalar resonances couple to particles proportionally to their Yukawa coupling, i.e. to their mass. Now let us describe the typical particle content of the simplified models used in CMS to describe the dark sector. DM is described as a single type of particles: no generations or flavours as the ones distinguishing the different SM particles. Moreover, the DM particles are considered to be Dirac fermions. The impact of the the spin on the phenomenology in collider searches is small enough for this choice to be valid in most situations. On top of the new DM particles, a new mediator of spin 0 or 1 connecting the SM to the dark sector is added to the theory. Spin 0 mediators can be chosen as real or complex scalar fields, with only complex scalars yielding the possibility for both scalar and pseudoscalar particles. A spin 1 mediator is called an axial-vector ($g^V = 0$) or a vector ($g^A = 0$). The mediator has a mass m_{med} and couples to the SM quarks and the DM particles with a coupling g_q and g_{DM}, respectively (we do not cover the coupling to leptons in this work). The value of these couplings has to be chosen carefully such that the model agrees with current limits on the existence of DM particles. The choice of the couplings must also match the desired phenomenology. For instance, it directly impacts the decay width of the mediator. Due to a color enhancement factor and the fact that there are three families of quarks and only one family of DM fermions (under the hypotheses of this simplified models framework), the width of the mediator is largely driven by the coupling to the quarks. Considering searches for narrow resonances in the dijet final state ($q\bar{q}' \to Z' \to q\bar{q}'$), the coupling to quarks must stay small with respect to the coupling to DM particles. Moreover, a large coupling of the new mediator to the quarks would make the searches in other final states irrelevant as their sensitivity would be much smaller. The ATLAS/CMS Dark Matter Forum [108] recommends the use of $g_q = 0.25$ so that the dijet and the invisible

($q\bar{q}' \to Z' \to \chi\bar{\chi}$) final states stay quite independent. With a choice of $g_{DM} = 1$, this leads to a ratio of the mediator width to its mass $\Gamma/m_{\text{med}} = 0.06$, fulfilling the requirement of a narrow resonance. In the case of scalar mediators, MFV makes the coupling to fermions proportional to their mass and the dominant diagram for the dijet final state is through top quark loops as depicted by the diagram in figure 5.2. As this leads to a much smaller cross-section than that in the vector mediator scenario, there is no need to further require a suppression of the dijet channel with respect to the E_T^{miss} channel by setting g_q to a smaller value than g_{DM}. The requirement of a narrow resonance is met for $g_q = g_{DM} = 1$ ($\Gamma/m_{\text{med}} = 0.06$) which are the recommended values. These choices are further motivated in [107].

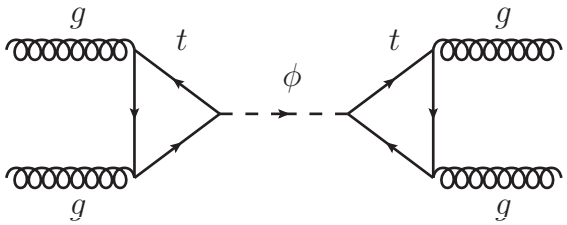


Figure 5.2: Dominant channel for dijet production through a scalar boson. This process is suppressed by the two top quark loops.

In figure 5.3, the main DM production diagrams are illustrated. On the left, the diagram for the production of DM particles through the decay of a new (pseudo)scalar is shown. The top quark loop diagram is the dominant contribution to this process. DM particles can also couple directly to the SM higgs boson in a similar way in what is often referred to as a "higgs portal". The middle diagram illustrates what happens when DM particles couple to a real scalar mediator that mixes with the SM higgs field. The physical states observable after mixing are denoted h_1 and h_2. Finally, the right diagram shows how DM can be produced via a vector mediator. From an experimental point of view, the left and right diagrams lead to jets and missing energy in the final state. This final state is referred to as the "monojet" signature. In the situation illustrated by the middle diagram, the final state is made of an electroweak boson and missing energy. If the boson decays to quarks, the final state is similarly found to be made of jets and missing energy. This is called the "mono-V" process. In the next section, we present a dedicated analysis to search for DM production along with jets, i.e. a combination of the monojet and the mono-V categories.

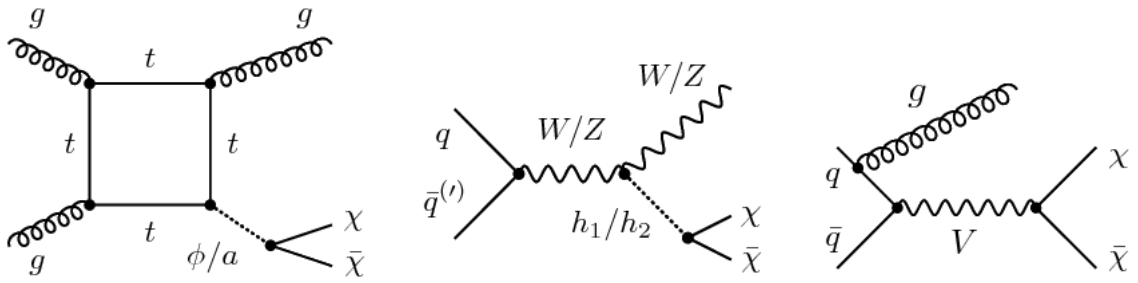


Figure 5.3: Diagrams for the production of fermionic DM χ through a scalar/pseudoscalar ϕ/a (left), through a the SM higgs boson mixing with a new scalar mediator (middle) and through a vector mediator (right) [107].

5.2 The monojet analysis

Should DM particles be produced inside the CMS detector, they would leave it unnoticed just as neutrinos do. The event would then look empty and no analysis would be possible. This

simple process is depicted by the top diagram in figure 5.4. The way-around this is to ask for the co-production of the pair of DM particles along with a detectable object. Given the initial state particles are protons and the large QCD couplings, a natural choice is to ask for the initial state radiation (ISR) of a gluon or a quark. We show this on the bottom-left diagram of figure 5.4 in the case of a gluon ISR. The final state is not empty anymore but contains one or more jets as the result of the gluon's hadronization. Although the signature is not exactly the same, the ISR of a vector boson decaying hadronically also gives jets in the final state. Such an event will be referred to as a mono-V event. The CMS monojet analysis combines these signatures. Both diagrams lead to a final state with one or more jets recoiling against missing transverse energy. The monojet analysis presented here was published and can be found under reference [110]. The data analyzed correspond to an integrated luminosity of 35.9 fb^{-1} recorded in 2016.

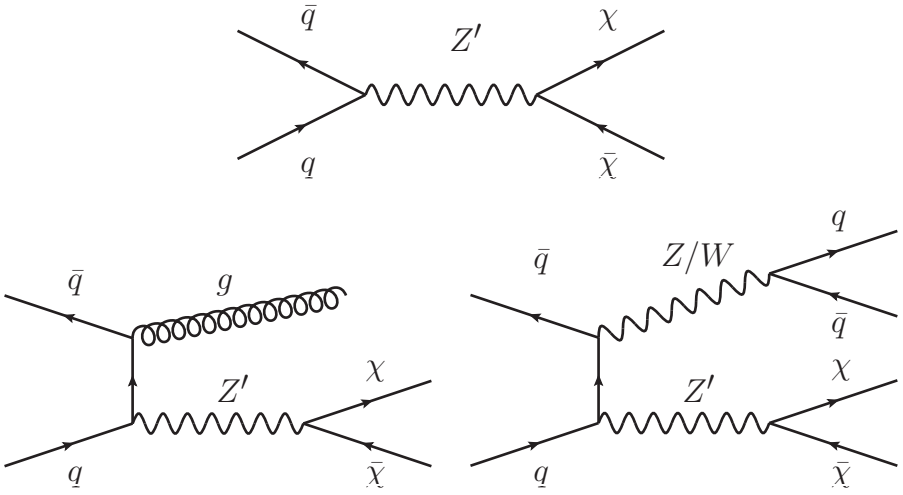


Figure 5.4: Production of DM particles through a new gauge boson Z' without (top) initial state radiation (ISR), with a gluon ISR (bottom-left) and with a vector boson ISR (bottom-right).

5.2.1 Event selection

Requiring an extra object to be produced alongside the DM system means that this object is triggered on. The triggers used in this analysis require both a large $\mathrm{E}_T^{\mathrm{miss}}$ and a large H_T^{miss}, where H_T^{miss} is the scalar sum of the p_T of all jets with $p_T > 20$ GeV and $|\eta| < 5$ in the event at trigger level. This trigger ensures the selection of events with a large hadronic activity but that are not QCD multijet events which do not show large $\mathrm{E}_T^{\mathrm{miss}}$. An important feature of this trigger is that both the $\mathrm{E}_T^{\mathrm{miss}}$ and the H_T^{miss} are computed from PF candidates but excluding the muons. This extends the selection at trigger level to events containing muons that will be used to estimate the main backgrounds (see section 5.2.2). The trigger thresholds are 110 or 120 GeV on both variables depending on the data taking year.

The offline event selection is driven by the monojet kinematics, i.e. jets back-to-back with $\mathrm{E}_T^{\mathrm{miss}}$ and no other high p_T particle in the event. Table 5.1 lists all selection criteria. Charged leptons and heavy flavour jets are vetoed to reject vector boson and top production. The cut on $\Delta\phi(\vec{p}_T^{\,\mathrm{jet}}, \vec{p}_T^{\,\mathrm{miss}})$ helps rejecting QCD multijet events in which the energy of a jet is mismeasured leading to $\mathrm{E}_T^{\mathrm{miss}}$. It applies to the four highest p_T jets that have a p_T larger than 30 GeV.

The selected events are then assigned to the monojet or the mono-V category depending on the topology of the jets. Mono-V events must contain at least one AK8 clustered jet (see section 3.3.7) fulfilling:

- $p_T > 250$ GeV and $|\eta| < 2.4$
- $\tau_2/\tau_1 < 0.6$
- Mass of the jet $\in$ [65,105] GeV

Variable	Selection	Target background
E_T^{miss}	> 250 GeV	QCD, top, Z(ll)+jets
E_T^{miss} filters	Passing the filters	Instrumental E_T^{miss}
Leading AK4 jet	$p_T > 100$ GeV, $\|\eta\| < 2.4$	All
Muon (electron) veto	$p_T > 10$ GeV, $\|\eta\| < 2.4$ (2.5)	Z(ll)+jets, W(lν)+jets, top
τ veto	$p_T > 18$ GeV, $\|\eta\| < 2.3$	Z(ll)+jets, W(lν)+jets, top
γ veto	$p_T > 15$ GeV, $\|\eta\| < 2.5$	γ+jets
Bottom jet veto	CSVv2 < 0.8484, $p_T > 15$ GeV, $\|\eta\| < 2.4$	Top
$\Delta\phi(\vec{p}_T^{\,jet}, \vec{p}_T^{\,miss})$	> 0.5 rad	QCD

Table 5.1: Event selection criteria for both the monojet and mono-V categories (adapted from [110]). For each cut the target background is mentioned. CSVv2 stands for the "Combined Secondary Vertex" version 2 algorithm for b-jet tagging [111].

If no such jet is found, the event is assigned to the monojet category. The increased p_T cut is justified by the fact that mono-V events typically have only one high p_T jet, coming from the hadronic decay of the vector boson. This jet alone must balance the large E_T^{miss}. On the other hand, jets in monojet events are the results of ISR gluons and an event is more likely to contain several of them. The p_T of the jet system can therefore be shared among the jets, allowing for a lower p_T cut for each jet. The τ_n variable is called the jet n-subjettiness [112] and is defined as:

$$\tau_N = \frac{1}{d_0}\Sigma_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, ..., \Delta R_{N,k}\} \tag{5.1}$$

where:

- 1, 2, ..., N, are the subjet indices
- k is the index of the subjet constituent particles
- $\Delta R_{i,k} = \sqrt{\Delta\eta_{ik}^2 + \Delta\phi_{ik}^2}$
- $d_0 = \Sigma_k p_{T,k} R_0$ and R_0 is the jet radius

The n-subjettiness is an efficient variable to distinguish QCD jets from the hadronic decay of boosted objects such as top quarks or vector bosons. The ratio of τ_2 to τ_1 is smaller for boosted objects decaying to two quarks than for QCD jets.

$\mathbf{E_T^{miss}}$ filters

The key variable of the monojet search is the missing transverse energy: signal extraction is done on the E_T^{miss} spectrum. This implies that events containing suspiciously high E_T^{miss} are to be removed from the event collection. This is done by applying the E_T^{miss} filters introduced in section 4.4.1. This is really where I contributed by studying the effect of the filters on the E_T^{miss} distribution. We realized that the filters do not clean all undesired events, as is shown in figure 5.5. The two variables displayed are the signal (highest p_T) jet charged and neutral energy fraction, i.e. the fraction of the jet energy coming from charged or neutral hadrons. A basic signal selection close to that of the analysis is applied:

- Leading jet $p_T > 100$ GeV, $|\eta| < 2.5$
- b-jet, leptons and photon vetos
- $\Delta\phi(\vec{p}_T^{\,jet}, \vec{p}_T^{\,miss}) > 0.5$

One can see that even after application of the E_T^{miss} filters, data shows an excess at high neutral energy fraction and low charged energy fraction. These are events in which calorimeter noise or beam halo generates energy in the calorimeters with no associated track. In the monojet

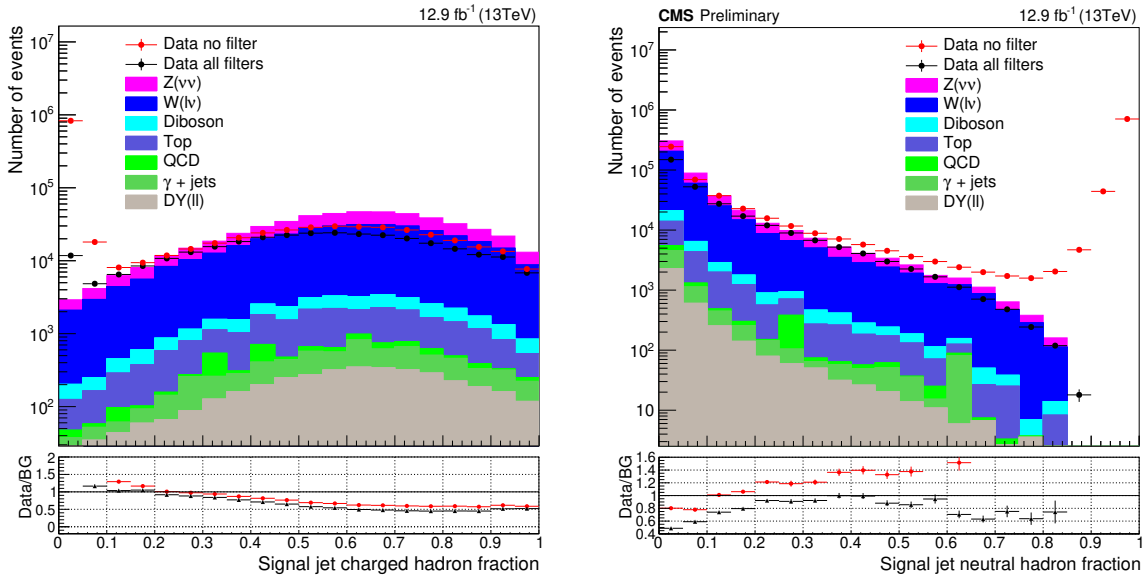


Figure 5.5: Signal jet charged (left) and neutral (right) energy fractions. Even after application of all filters (including the cut on the other energy fraction), not all noisy events are cleaned out.

analysis, we decided to apply a cut on each of these variables to reject events in which such jets are found. We remove events with the signal jet having either one of these two cuts:

- Charged energy fraction < 0.1
- Neutral energy fraction > 0.8

In figure 5.5, "all filters" includes the cut on the other energy fraction. Figure 5.6 shows the E_T^{miss} spectrum in 2016 data before and after application of all filters. Here, "all filters" includes both extra cuts on the energy fractions.

On top of this, my daily job was to make sure a minimal number of events was rejected in the process by analyzing the impact of each filter on the distribution. This work was coherent with my activities in the CMS E_T^{miss}-scanning group where I was studying the filters in details on the latest recorded data.

5.2.2 Background estimation

Main backgrounds

The two main backgrounds of this search are the Z($\nu\nu$) + jets and W($l\nu$) + jets processes, Z($\nu\nu$) being irreducible as it yields the exact same signature. Those two backgrounds are estimated with a data-driven method. The idea is to count the number of events passing the selection in control regions and to transfer this number to the signal region to be interpreted as the expected number of background events. The transfer is done with dedicated factors derived from simulation as is described below. Control regions are datasets orthogonal to each other and to the search dataset and that we understand contain clean SM events, i.e. with no new physics event. They ensure the description of the SM backgrounds is realistic in our experimental context. The five control regions used here are:

- Dimuon and dielectron to estimate Z($\nu\nu$)
- γ + jets to estimate Z($\nu\nu$)
- Single-muon and single-electron to estimate W($l\nu$) and Z($\nu\nu$)

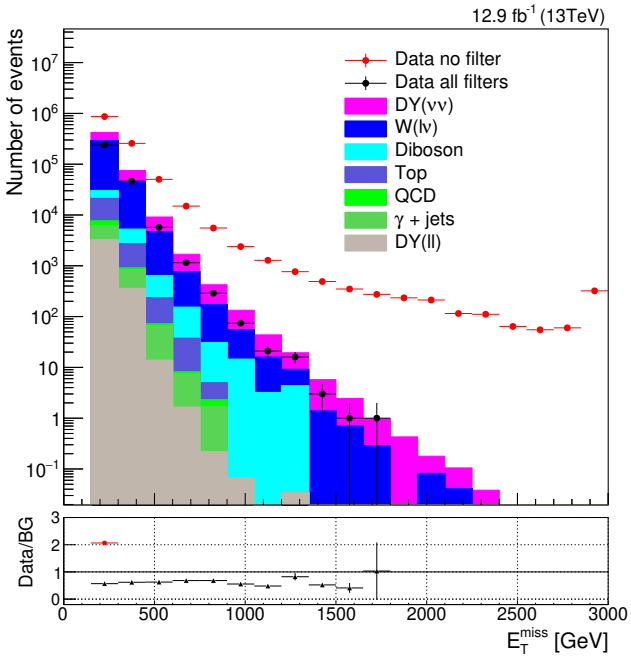


Figure 5.6: E_T^{miss} spectrum in a subset of 2016 data before (black) and after (red) E_T^{miss} filters are applied. The distribution in data is compared to the background expectation where less spurious E_T^{miss} is expected. Backgrounds are evaluated using simulation only and normalized to the luminosity without any correction factor (such as scale factors or pileup reweighting).

The dimuon and dielectron CRs contain events that have a kinematic description as close as possible to that of the Z($\nu\nu$) background. The photon + jets CR is added to reduce the statistical uncertainties for this process, especially at large E_T^{miss}. The larger number of events in this CR is explained by the fact that the photon does not decay while the two weak gauge bosons do. Therefore, the photon + jets cross-section is not reduced by the corresponding branching ratio. Similarly, the single-lepton CRs contain events that are as close as possible as the W($l\nu$) background. The number of events in these regions is large enough to be used without the addition of an extra control sample to estimate the W($l\nu$) background. To transfer the number of events estimated in the control regions to the signal region, E_T^{miss}-bin dependent transfer factors are derived using MC simulations. These are produced at leading order (LO) with their cross-section corrected to NLO(QCD+EW) by k-factors. As an example, the transfer factor for the Z($\mu\mu$) process in the bin i is defined as:

$$R_i^{\mu\mu} = \frac{N_{i,\mathrm{MC}}^{\mathrm{Z}(\mu\mu)}}{N_{i,\mathrm{MC}}^{\mathrm{Z}(\nu\nu)}} \tag{5.2}$$

where $N_{i,\mathrm{MC}}^{\mathrm{Z}(\mu\mu)}$ is the number of events in simulated Z($\mu\mu$) events passing the event selection for the dimuon control region and $N_{i,\mathrm{MC}}^{\mathrm{Z}(\nu\nu)}$ is the number of events in simulated Z($\nu\nu$) events passing the event selection for the signal region. Given the large statistical content of the single-lepton CRs, transfer factors connecting the single and di-lepton CRs are derived to allow the use of the single-lepton CR to constrain the Z($\nu\nu$) process. As already said, events in the single- and di-muon control regions are collected using the same triggers as for the signal region. Single- and di-electron events are collected using a single-electron trigger with a p_T threshold of 27 GeV. Photon + jets events are collected with the help of a single-photon trigger with a p_T threshold of 165 or 175 GeV depending on the year. The event selection described in the previous section is applied to each of these regions with the hadronic recoil p_T playing the role of the E_T^{miss} and without the corresponding veto on the leptons or the photon. The hadronic recoil is defined as the E_T^{miss} computed without the leptons or the photon. In an event with no E_T^{miss}, it corresponds to the vectorial sum of the p_T of the jets, hence its name.

The expected number of events in the signal region from the Z($\nu\nu$) and the W(lν) backgrounds are extracted from combined likelihood fits to all bins of the E_T^{miss} distribution in every control region and in the signal region simultaneously. A likelihood should be understood as an a posteriori probability to have observed a certain number of events given a known statistical distribution of unknown parameter. The goal is then to vary the unknown parameter to find the value that yields the maximum likelihood, i.e. the value that is the most likely to have given the observed outcome. Here, the statistical distribution is the one describing a counting experiment, i.e. the Poisson distribution, and the unkown parameter is the number of background events. For the Z($\nu\nu$) background, it reads:

$$\begin{aligned}
\mathcal{L}(\mu^{\mathrm{Z}(\nu\nu)}, \mu, \theta) &= \prod_i \mathrm{Poisson}\left(d_i^{\gamma} | B_i^{\gamma}(\theta) + \frac{\mu_i^{\mathrm{Z}(\nu\nu)}}{R_i^{\gamma}(\theta)}\right) \\
&\times \prod_i \mathrm{Poisson}\left(d_i^{\mu\mu} | B_i^{\mu\mu}(\theta) + \frac{\mu_i^{\mathrm{Z}(\nu\nu)}}{R_i^{\mu\mu}(\theta)}\right) \\
&\times \prod_i \mathrm{Poisson}\left(d_i^{ee} | B_i^{ee}(\theta) + \frac{\mu_i^{\mathrm{Z}(\nu\nu)}}{R_i^{ee}(\theta)}\right) \\
&\times \prod_i \mathrm{Poisson}\left(d_i^{\mu} | B_i^{\mu}(\theta) + \frac{f_i(\theta)\mu_i^{\mathrm{Z}(\nu\nu)}}{R_i^{\mu}(\theta)}\right) \\
&\times \prod_i \mathrm{Poisson}\left(d_i^{e} | B_i^{e}(\theta) + \frac{f_i(\theta)\mu_i^{\mathrm{Z}(\nu\nu)}}{R_i^{e}(\theta)}\right) \\
&\times \prod_i \mathrm{Poisson}\left(d_i | B_i(\theta) + (1 + f_i(\theta))\mu_i^{\mathrm{Z}(\nu\nu)} + \mu S_i(\theta)\right) \qquad (5.3)
\end{aligned}$$

where d_i^X is the observed number of events in bin i in the region X, B_i^X is the number of background events for the corresponding control region, θ represents the systematic uncertainties treated as a nuisance parameter in the fit, f_i is a transfer factor between the W and Z control regions and acts as an additional constraint, μ is the signal strength, S_i is the expected number of signal events and $\mu_i^{\mathrm{Z}(\nu\nu)}$ is the expected number of events from the Z($\nu\nu$) background. This latter parameter as well as the signal strength are left free in the fit. The likelihood is here built as a product of several Poisson distributions, one for each bin of each CR and for each bin of the signal region as well (in the last line). The number of events from the Z($\nu\nu$) background ($\mu_i^{\mathrm{Z}(\nu\nu)}$) is considered a free parameter in the fit and the maximization procedure will make it such that the likelihood is as large as possible. In other words, the maximum likelihood procedure finds the number of events from BG processes that maximizes the agreement with the observed data. A similar likelihood is built for the W($l\nu$) background but without the contribution from the dilepton and the γ + jets CRs.

Subdominant backgrounds

Here are the subdominant backgrounds for this search:

- Top
  - Top quarks decay at almost 100 % to a W boson and a b quark. If the W boson decays leptonically, the final state contains neutrinos and so E_T^{miss}. Additionally, if a charged lepton from the decay of the W boson is missed by the detector (being of low quality or outside the detector acceptance), E_T^{miss} will arise in the event description.
- Diboson

- Neutrinos in W leptonic decays are a source of E_T^{miss}. Leptons from Z or W decays can also be missed which adds even more E_T^{miss}.

- Z(ll) (with both leptons lost)
  - If the leptons are lost, the final state becomes similar to that of the Z($\nu\nu$) process.

- QCD multijets
  - Mismeasured jets are a source of E_T^{miss} if part of the calorimeter energy is missed.

With the exception of QCD multijet events, all other backgrounds are estimated using simulations only. As jet reconstruction often show a significantly different behaviour in data and in simulations, the QCD multijet background prediction makes also use of data [110].

5.2.3 Results

Signal extraction is done by comparing data and expectations in the E_T^{miss} spectrum. Figure 5.7 compares the E_T^{miss} spectrum in data with the background expectation before and after the combined fit in the monojet and the mono-V categories. From the lower panel, one can see that the agreement between data and predictions improves after the maximum likelihood fit (blue points to be compared to the red points before the fit is performed). No significant deviation is observed in data with respect to the SM expectations.

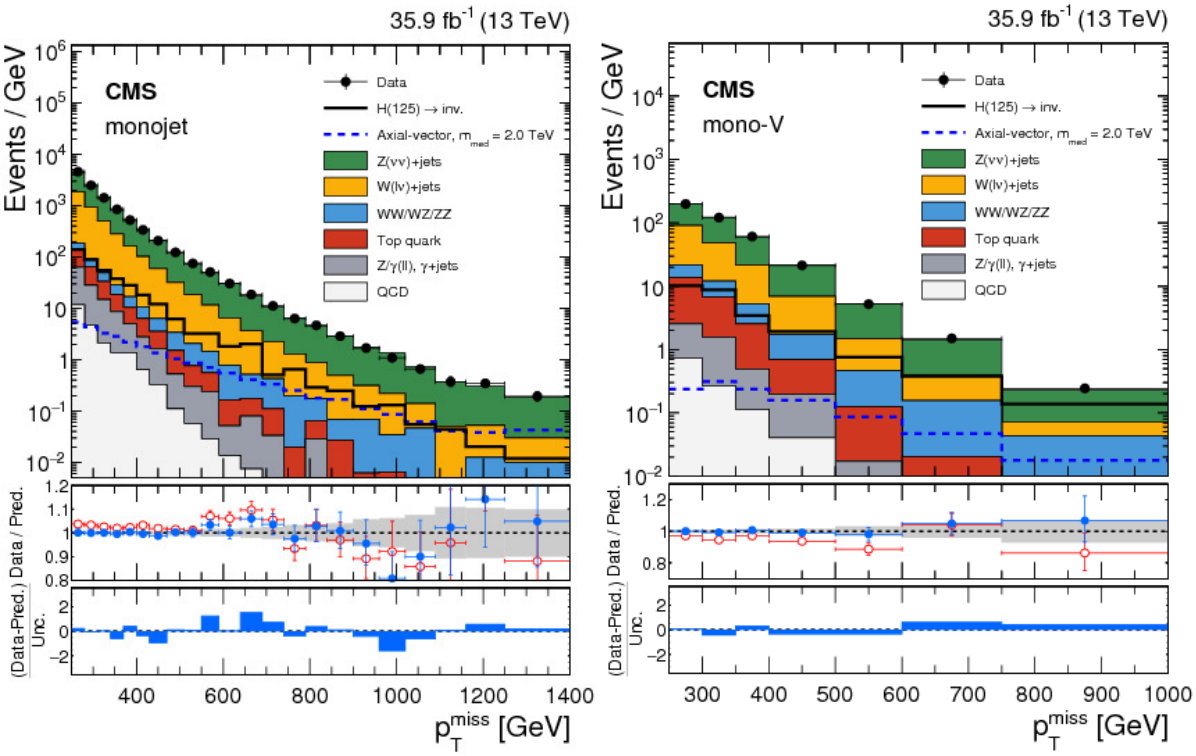


Figure 5.7: E_T^{miss} spectrum in the monojet (left) and mono-V (right) categories [110]. Data is compared to the background expectation before (blue) and after (red) the combined likelihood fit.

Upper limits are set on the DM production cross-section in the context of simplified models with the choice of couplings motivated in section 5.1. Figures 5.9 and 5.10 show the exclusion regions in the m_{med}-m_{DM} plane for the vector/axial-vector and scalar/pseudoscalar, respectively. In the scalar case, no exclusion is possible for masses as low as 1 GeV for the considered values of the model parameters. For a spin-1 mediator, masses up to 1.8 TeV are excluded (again for

the considered parameters). On these plots, the constraints from cosmological observations by the Planck collaboration are superimposed [41]. To understand those limits, one must briefly discuss the history of DM particles in the early universe [113]. In the framework of the Λ-CDM model, the course of DM starts in the very hot early moments of the universe when temperatures were larger than the mass of the DM particles $T > m_{\mathrm{DM}}$. At that times, DM particles were approximately as abundant as photons and their abundance was constant over time. With the expansion of the universe, the temperature decreased until eventually $T \leq m_{\mathrm{DM}}$. From that moment on, DM particles left their relativistic behaviour to become "cold". As a consequence, the energy available to the other particles to produce DM was too low ($T < m_{\mathrm{DM}}$) and the DM abundance dropped according to the annihilation rate (see figure 5.8). However, this annihilation stopped when DM went out of thermal equilibrium: this is called the "freeze-out". The expansion of the universe made the DM particles rarer and rarer and this prevented further annihilation. Another way to see this is that the expansion speed exceeded the annihilation rate. As both production and annihilation stopped, the DM abundance remained constant since then. Such a species that was in thermal equilibrium in the early universe but that froze out and still exists today is called a cold thermal relic. Kolb tells us that the current DM abundance Ω_{DM} depends on the model parameters (DM and mediator masses) only through the annihilation cross-section $\Omega_{DM} \propto \sigma_A^{-1}$ [113]. As $\sigma_A \propto m_{\mathrm{med}}^{-4}$, this means that larger m_{med} leads to a smaller annihilation cross-section and a larger abundance. For large enough mediator masses, DM is predicted to be overabundant with respect to Planck data and one can exclude its existence. The mass of the DM itself determines the time at which DM particles go out of the relativistic regime and start to disappear. The lighter the particle, the later this happens and the more DM is present today. As a consequence, Planck data allows to set a lower limit on the DM particles mass for a given mediator mass.

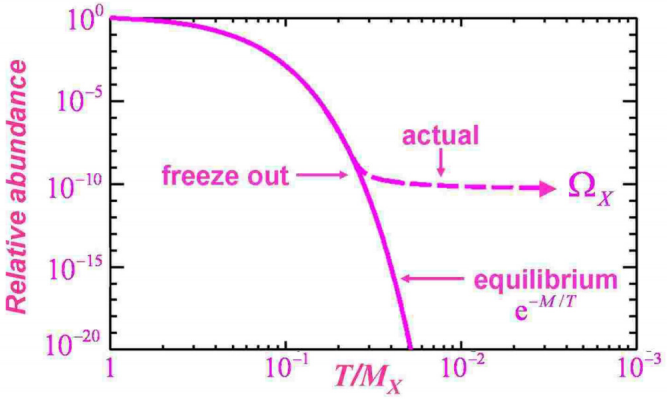


Figure 5.8: Relative abundance for a cold thermal relic and freeze-out mechanism [113]. On the X axis, T/M_X is the ratio of temperature to the mass of the particle. It begins when DM particles leave the relativistic regime and annihilation takes over production. Their abundance still follows the thermal equilibrium function until annihilation stops and they freeze out.

As was illustrated in figure 1.3, the SM-DM connection can be interpreted in three different ways: DM-DM annihilation to produce SM particles (indirect detection approach), DM-SM interaction (direct detection approach) and SM-SM annihilation to produce DM particles (collider approach). In the context of simplified models, those three approaches involve the same theoretical description of the SM-DM interplay and involve the same parameters: the DM and mediator masses as well as the couplings of the mediator to the dark sector and to the SM. We show here how CMS results can be interpreted in the vocabulary of direct detection experiments. In such experiments, the DM particle is expected to interact elastically with an atomic nucleus which recoil (typically in the 1-100 keV range depending on the DM and the nucleus masses) is then measured. The choice of a vector or a scalar mediator leads to spin-independent DM-nucleon cross-section while axial-vector and pseudoscalar mediators imply a spin-dependent cross-section.

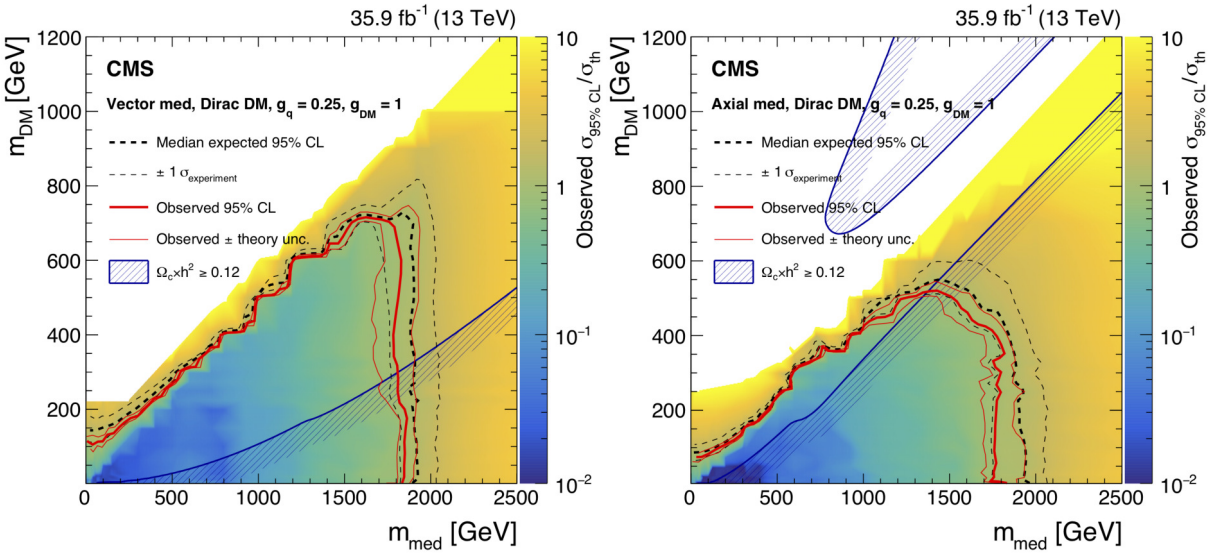


Figure 5.9: Exclusion limits at 95 % CL on the signal cross-section in the vector (left) and axial-vector (right) cases [110]. The dark blue contour is the limit from the Planck experiment.

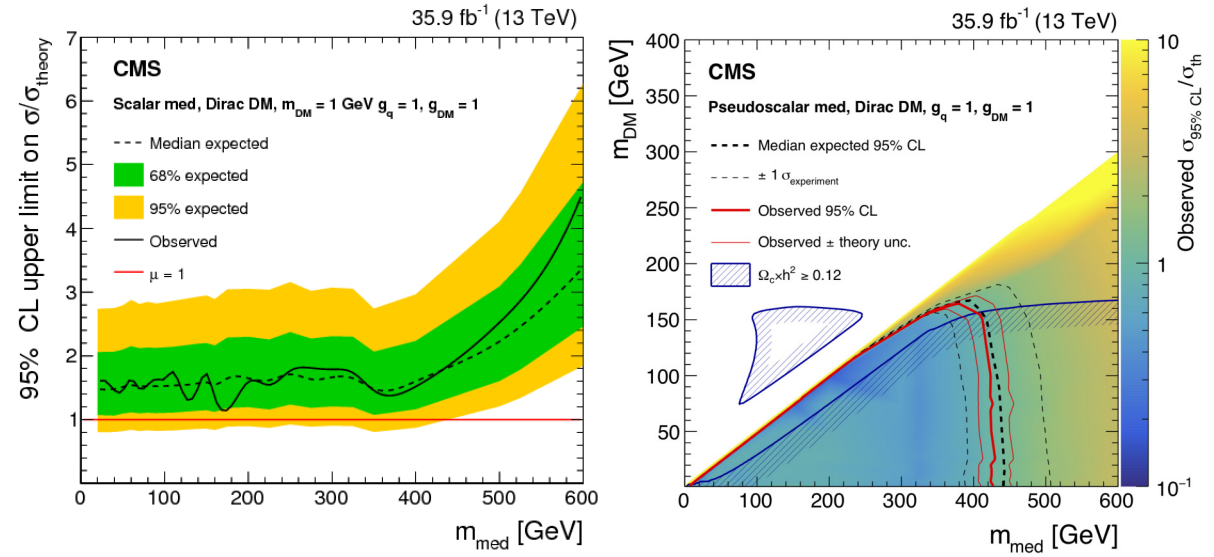


Figure 5.10: Exclusion limits at 95 % CL on the signal cross-section in the scalar (left) and pseudoscalar (right) cases [110]. Scalar limits are given for a fixed DM mass of 1 GeV as the expected and observed cross-sections do not reach below the predicted cross-section. The monojet analysis is not sensitive to this scenario with the amount of data analyzed and no exclusion is therefore possible. The dark blue contour is the limit from the Planck experiment.

In the spin-independent scenario, the DM-nucleus interaction is assumed to happen similarly over all the nucleons independently on their proton or neutron nature [114]. This gives the spin-independent searches a much higher sensitivity as it scales with the square of the number of nucleons in the nuclei of the chosen element. On the other hand, spin-dependent results must be taken relative to the number of protons and neutrons in the nuclei, therefore yielding a sensitivity orders of magnitudes below the spin-independent results. Moreover, another effect plays a role in the fact that the upper limit on the spin-dependent cross-section is higher. We know from reference [115] that the spin dependency of the cross-section has an impact on its amplitude. The spin-independent cross-section is given by:

$$\sigma_{SI} = \frac{f^2(g_q) g_{\mathrm{DM}}^2 \mu_{n_\chi}^2}{\pi M_{\mathrm{med}}^4}, \tag{5.4}$$

where the DM-nucleon coupling $f^2(g_q)$ depends on the DM-quark coupling as $f(g_q) = 3g_q$. The spin-dependent cross-section reads:

$$\sigma_{SD} = \frac{3f^2(g_q)g_{\mathrm{DM}}^2\mu_{n_\chi}^2}{\pi M_{\mathrm{med}}^4}, \tag{5.5}$$

where $f(g_q) = 0.32g_q$. The different dependency of the DM-nucleon coupling to the DM-quark coupling comes from the way the mediator interacts with protons and neutrons. In the vector and scalar cases, the difference between a proton and a neutron is negligible. However, in the axial-vector and pseudoscalar cases, this difference must be taken into account and it yields a smaller DM-nucleon coupling. Plugging in the value of $f(g_q)$, the cross-sections read:

$$\sigma_{SI} = 6.9 \times 10^{-41}\mathrm{cm}^2 . \left(\frac{g_q g_{\mathrm{DM}}}{0.25}\right)^2 \left(\frac{1\mathrm{TeV}}{M_{\mathrm{med}}}\right)^2 \left(\frac{\mu_{n_\chi}}{1\mathrm{GeV}}\right)^2 \tag{5.6}$$

$$\sigma_{SD} = 2.4 \times 10^{-42}\mathrm{cm}^2 . \left(\frac{g_q g_{\mathrm{DM}}}{0.25}\right)^2 \left(\frac{1\mathrm{TeV}}{M_{\mathrm{med}}}\right)^2 \left(\frac{\mu_{n_\chi}}{1\mathrm{GeV}}\right)^2 \tag{5.7}$$

The upper limits on the spin-independent and spin-dependent DM-nucleon cross-sections are shown in figure 5.11 for the vector and axial-vector cases. Due to the fact that direct detection experiments measure the recoil of the nuclei, they have a better sensitivity at large DM masses. CMS is competitive in the low mass range, below $m_{\mathrm{DM}} = 5$ GeV in the vector mediator case. In the axial-vector scenario, one can see that the upper limit on the spin-dependent cross-section is much higher as expected from the arguments given above. In this region, CMS has the best limits up to 500 GeV. These results strongly depend on the model parameters.

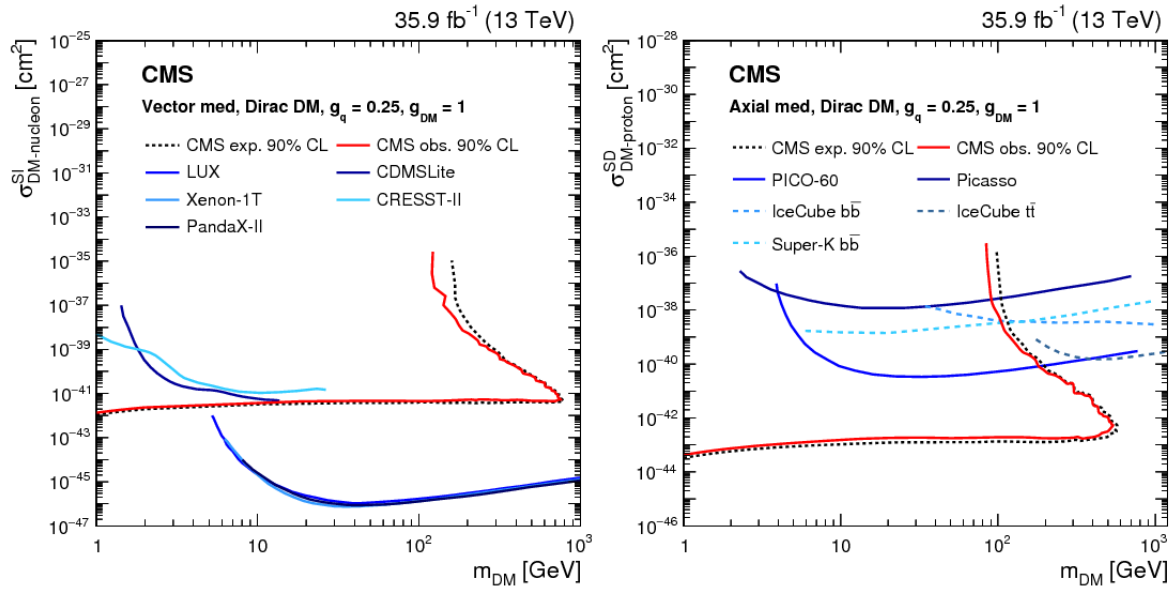


Figure 5.11: Exclusion limits at 95 % CL on the spin-independent (left) and spin-dependent (right) DM-nucleon interaction cross-section in the vector (left) and axial-vector (right) cases [110]. CMS has the best limits in the low mass region as direct detection experiment measure the nucleon recoil which decreases with lower DM mass. In the axial-vector scenario, CMS shows better results on a much broader range. This is due to the fact that the DM-nucleon cross section is spin-suppressed.

5.3 Search for invisible decays of the Higgs boson

The monojet analysis is closely related to the search for invisible decays of the Higgs boson in the Vector Boson Fusion (VBF) channel. The main difference with respect to the monojet analysis

comes from the nature of the process. In the monojet analysis, the mediator is produced through QCD interactions while in this analysis, the Higgs boson is produced through a pure EW process. The diagram for the VBF production of a Higgs boson decaying to a pair of DM particles is given in figure 5.12. The final state of a VBF process is made of two jets and the decay product of the particle produced in the boson fusion. If the vector bosons fuse to produce a Higgs that decay to invisible states, the final state is made of two jets and missing energy i.e. a signature similar to that of the monojet process. However, the kinematics show some significant differences. As opposed to the QCD monojet with multiple ISR quarks and gluons, the two VBF jets have a particular status. Due to the fact that they are not QCD-connected, they tend to have a large pseudorapidity difference and a large invariant mass. This is shown in figure 5.13 for simulations.

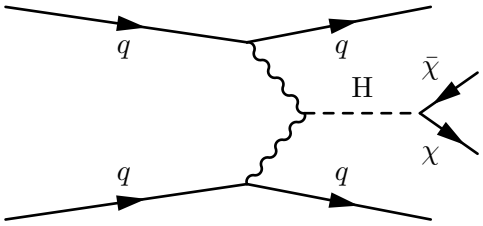


Figure 5.12: VBF production of a Higgs boson and subsequent decay to a pair of DM particles. The vector bosons can either be a W^+ and a W^- or two Z.

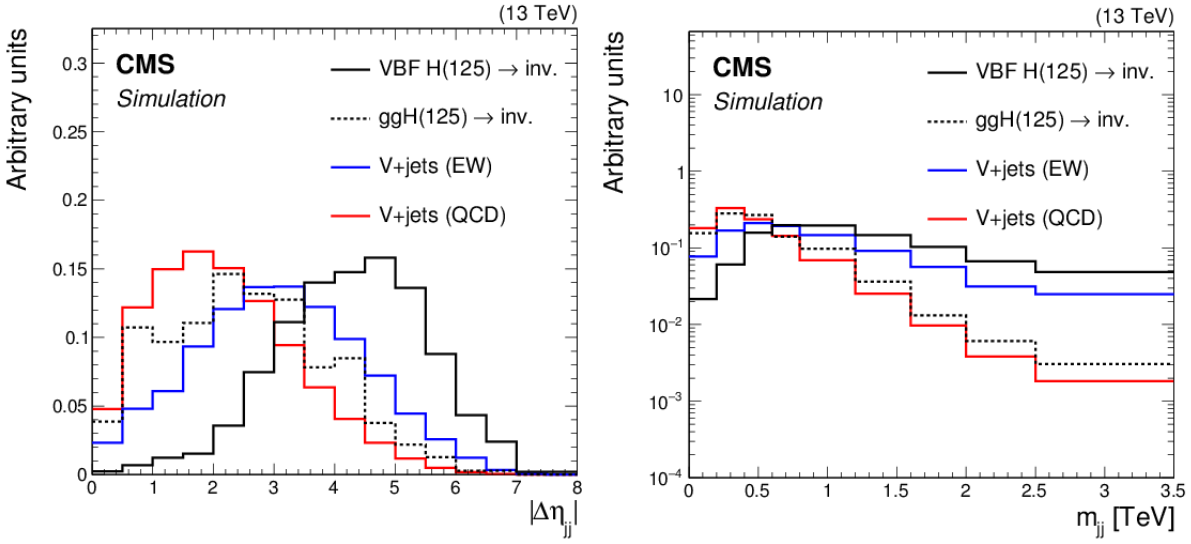


Figure 5.13: Pseudorapidity difference and invariant mass of the two VBF jets in VBF and non-VBF simulated events [116]. The jets in VBF processes show a larger η difference and invariant mass.

The analysis presented in this section has been published under reference [116]. Let us present briefly the analysis strategy and then focus on the aspect I have been working on directly, namely the computation of k-factors to be used in the combined likelihood for the background estimation. Events are collected using the combined E_T^{miss} and H_T^{miss} triggers with a threshold at 110 or 120 GeV depending on the data taking year. Offline, $E_T^{miss} > 250$ GeV is required. The main selection criterion is then to select events containing at least two jets, the leading jet having a $p_T > 80$ GeV and the subleading jets having a $p_T > 40$ GeV. The product of the jet pseudorapidities must be negative (enforcing opposite CMS hemispheres, i.e. a large $|\Delta\eta|$) and at least one of the two jets must have $|\eta| < 3$. Then, the strategy is similar to what is done in the monojet analysis to the notable exception that the signal extraction variable is the invariant mass of the two VBF jets, M_{jj}. A combined likelihood is used to determine the most likely number of background events in each M_{jj} bin. It is constructed using the same CRs as in the monojet analysis (except

for γ+jets) but the transfer factors are derived separately for the QCD and the EW background processes in each region. For the VBF Z($\nu\nu$) BG, it is defined as:

$$
\begin{aligned}
\mathcal{L}(\mu^{\mathrm{Z}(\nu\nu)}, \mu, \theta) &= \prod_i \mathrm{Poisson}\left(d_i | B_i(\theta) + (1 + f_i(\theta)_{\mathrm{QCD}})\mu_i^{\mathrm{Z}(\nu\nu)} + R_i^Z(1 + f_i(\theta)_{\mathrm{EW}})\mu_i^{\mathrm{Z}(\nu\nu)} + \mu S_i(\theta)\right) \\
&\times \prod_i \mathrm{Poisson}\left(d_i^{\mu\mu} | B_i^{\mu\mu}(\theta) + \frac{\mu_i^{\mathrm{Z}(\nu\nu)}}{R_i^{\mu\mu}(\theta)_{\mathrm{QCD}}} + \frac{\mu_i^{\mathrm{Z}(\nu\nu)}}{R_i^{\mu\mu}(\theta)_{\mathrm{EW}}}\right) \\
&\times \prod_i \mathrm{Poisson}\left(d_i^{ee} | B_i^{ee}(\theta) + \frac{\mu_i^{\mathrm{Z}(\nu\nu)}}{R_i^{ee}(\theta)_{\mathrm{QCD}}} + \frac{\mu_i^{\mathrm{Z}(\nu\nu)}}{R_i^{ee}(\theta)_{\mathrm{EW}}}\right) \\
&\times \prod_i \mathrm{Poisson}\left(d_i^{\mu} | B_i^{\mu}(\theta) + \frac{f_i(\theta)_{\mathrm{QCD}}\mu_i^{\mathrm{Z}(\nu\nu)}}{R_i^{\mu}(\theta)_{\mathrm{QCD}}} + \frac{f_i(\theta)_{\mathrm{EW}}\mu_i^{\mathrm{Z}(\nu\nu)}}{R_i^{\mu}(\theta)_{\mathrm{EW}}}\right) \\
&\times \prod_i \mathrm{Poisson}\left(d_i^{e} | B_i^{e}(\theta) + \frac{f_i(\theta)_{\mathrm{QCD}}\mu_i^{\mathrm{Z}(\nu\nu)}}{R_i^{e}(\theta)_{\mathrm{QCD}}} + \frac{f_i(\theta)_{\mathrm{EW}}\mu_i^{\mathrm{Z}(\nu\nu)}}{R_i^{e}(\theta)_{\mathrm{EW}}}\right) \qquad (5.8)
\end{aligned}
$$

The control-to-signal region transfer factors $R_i^X(\theta)$ as well as the Z to W transfer factors f_i^θ here depend on the nature of the process. In the signal region, a transfer factor R_i^Z connecting the Z_{EW} and the Z_{QCD} expectations is used as an extra constraint in the fit.
This likelihood resembles that of equation (5.3) except that two different contributions are considered for the corrections to the main background: the EW and the QCD corrections.

As in the monojet analysis, the transfer factors must be corrected for next-to-leading order corrections. My involvement in this analysis was to compute those corrections to be used in the likelihood. This work is presented in the next section.

5.3.1 k-factor computation

In a regime where the coupling constants are smaller than 4π, the matrix elements of a given process can be expanded perturbatively. The cross-section calculations can then be done at different orders of precision in the perturbative series. These are commonly referred to as leading order (LO), next-to-leading order (NLO), etc. The ratio of NLO/LO cross-sections, is called a k-factor. In all generality, a k-factor can relate any two orders in the computation, i.e. not only NLO/LO but also NNLO/LO or NNLO/NLO for instance. Typically, differential k-factors are used in analyses to correct leading order distributions for next-to-leading order effects. In the context of this analysis, we compute the QCD NLO corrections to the two main background at leading order, i.e. the NLO/LO k-factors. We detail the procedure of the computation using the VBFNLO tool [117, 118, 119]. All presented results are at generated level, i.e. no CMS reconstruction procedure is involved.

Configuration and use of the VBFNLO tool

The VBFNLO program is "a parton level Monte Carlo for processes with electroweak bosons" [117, 118, 119]. It can be used to compute cross-sections and QCD corrections at next-to-leading order for many VBF-related processes. In this section we derive the cross-sections and the k-factors in a "VBF phase-space". It is defined by the selection cuts listed in table 5.2.

A choice must be made regarding the value of the factorization μ_F and the renormalization μ_R scales as well as for the parton distribution functions (PDF). We choose to fix the scales to the momentum transfer of the exchanged boson. The PDF set is `NNPDF30_lo_as_0130` at LO and `NNPDF30_nlo_nf_5_pdfas` at NLO.

Results

Higher order corrections to the cross-sections evolve primarily with the transverse momentum of the final state boson. However, in a VBF analysis the invariant mass between the two jets is also

Parameter	Cut	Description
p_{T,j_1}	> 70 GeV	Minimum p_T of the leading VBF jet
p_{T,j_2}	> 40 GeV	Minimum p_T of the subleading VBF jet
$\Delta\eta_{jj}$	> 1.5	Minimum η difference between the 2 VBF jets
$\Delta\phi_{jj}$	< 1.5	Maximum ϕ difference between the 2 VBF jets
$\eta_{j_1} \times \eta_{j_2}$	< 0	VBF jets in two different CMS hemispheres
M_{jj}	> 250 GeV	Minimum invariant mass of the two VBF jets

Table 5.2: Definition of the VBF phase-space.

a good variable to discriminate the EW-induced signal from QCD-induced processes. With this in mind, and the fact that it is the variable used for signal extraction, we show the result of the cross-section computation as a function of M_{jj}. In figure 5.14, the LO and NLO cross-sections as well as the corresponding k-factor are given for the VBF Z($\nu\nu$) process. Figure 5.15 gives the equivalent for the W(lν) process. From these figures, one can see that the expected number of events is lowered by 5 to 10 % for both the Z($\nu\nu$) and the W(lν) processes with respect to the leading order expectations. These k-factors are used in the analysis to correct the control-to-signal transfer factors $R_i^X(\theta)_{QCD}$, leading to a better decription of the main backgrounds.

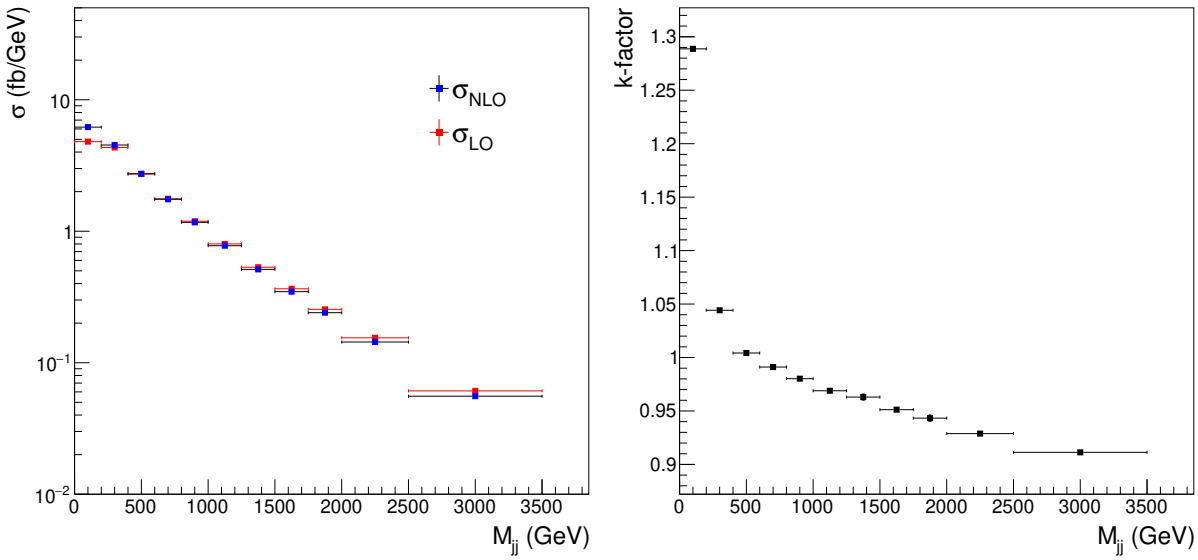


Figure 5.14: LO and NLO cross-sections (left) and corresponding k-factor (right) as a function of the boson p_T for the Z($\nu\nu$) process.

To fully understand the impact of those calculations, one must look at how they enter in the likelihood. We see that in equation (5.8), the yields of the QCD W and Z processes are connected through the $f_i(\theta)_{\text{QCD}}$ transfer factor. If the Z and the W k-factors correct these processes in the same way, the application of the k-factors will be canceled out in the likelihood. In figure 5.16, the ratio of the Z to the W k-factors is displayed in the $p_T^{\text{boson}} - M_{jj}$ plane. From this map, one can conclude that the effect of the correction is different for the two processes at 2-3 %. As a consequence, the application of the k-factors will effectively correct the expected number of background for the two processes.

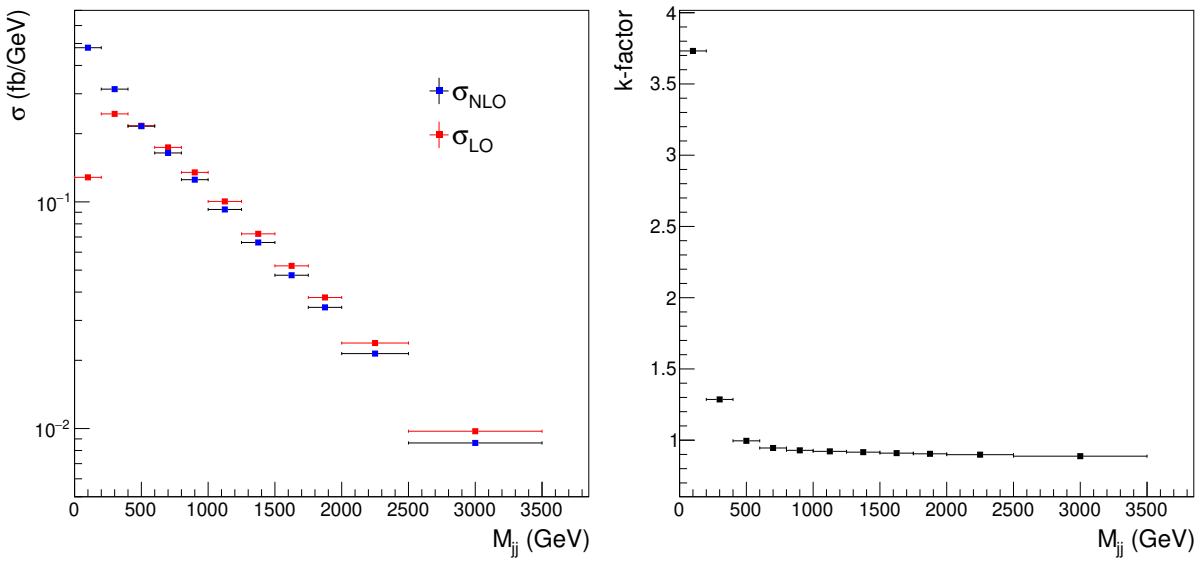


Figure 5.15: LO and NLO cross-sections (left) and corresponding k-factor (right) as a function of the boson p_T for the W(lν) process.

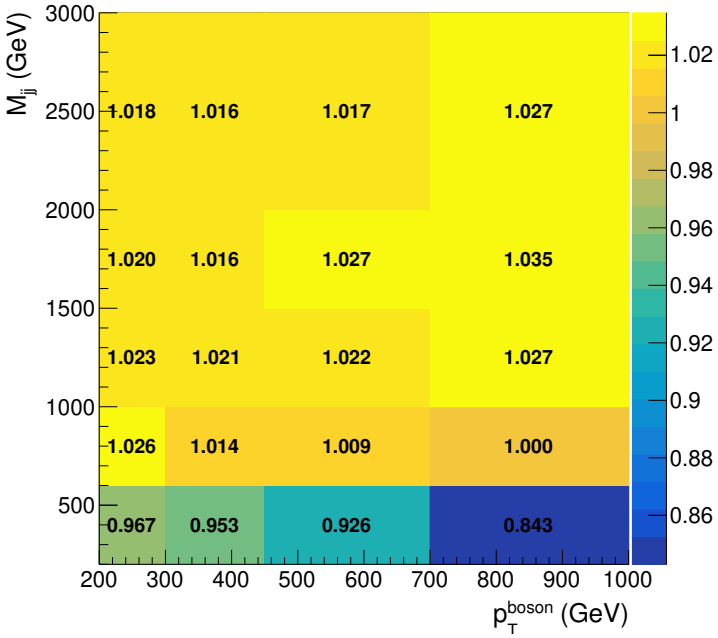


Figure 5.16: Ratio of Z to W k-factors as a function of the boson p_T and the VBF jets invariant mass. The deviation from 1 shows that the two processes are corrected differently by the k-factors.

Chapter conclusion

Searches for Dark Matter particles with the CMS detector were presented. We started by introducting the framework of simplified models. These are used to help the interpretation of DM searches by introducing boson mediators between the dark and the SM sectors. Then, we gave the results of the monojet analysis that looks for DM pair production recoiling against jets. Vector mediator masses are excluded up to 1.8 TeV with a corresponding DM particle mass of around 700 GeV. The search for invisible decays of the Higgs boson in the Vector Boson Fusion channel is closely related to the monojet analysis as it leads to a similar "jets + E_T^{miss}" final state. The general strategy of this second analysis is given. Finally, we present the results of a k-factor computation with the VBFNLO tool. These k-factors are applied to constrain likelihood fits used to estimate the background in the search for Higgs to invisible final states.

Chapter 6

Search for fractionally charged particles

Motivated abstract

While never observed, free-propagating particles with fractional electromagnetic charge are not forbidden from first principles as introduced in section 1.2.4. In the Standard Model, quarks carry a fractional charge but they are not observable states as they stay confined in hadrons. New questions arise from these observations: is the electron charge the lowest possible charge a free-propagating particle is allowed to carry and if so, why is the electromagnetic charge quantized in terms of the electron charge and what complete theory can explain this phenomenology? Do unknown particles carry an electromagnetic charge so small they have not been detected by any previous experiment?

In this chapter, we present the search for fractionally charged particles (FCP) using 136 fb^{-1} of proton-proton collision data collected during the LHC Run 2 by the CMS detector at $\sqrt{s} = 13$ TeV. We first motivate this search through several theoretical frameworks that could explain the existence of such particles. Then, we detail a particular simple model that we use as a guideline for phenomenological interpretations. We continue by presenting the details of the analysis: how we simulate signal events, what CMS data we analyze, how the instrumental background is dealt with, how we select signal events and estimate the corresponding background. No evidence of new physics is observed, thus we present the results in terms of upper limits on the signal cross-section. This analysis has not been published yet and is currently under review by the CMS Collaboration. This being said, these are the world best limits on this type of signature for the charge and mass ranges considered.

6.1 Theoretical motivations and constraints

In section 1.2.4, it was pointed out that the observed quantization of the electromagnetic charge is not explained by the Standard Model. Naturally, searches for particles with a charge lower than that of the electron have been carried out by many experiments. The theoretical motivations for the existence of such particles are closely linked to the physics of dark sectors. We call dark sectors hidden symmetries and fields weakly or not connected to the Standard Model sector. As we will develop in the next section, the existence of a hidden U(1) symmetry and of new fermions charged under this group naturally leads to these fermions carrying an effective low electromagnetic charge. This idea was first suggested in 1986 by Holdom [53]. Such situations where particles carrying only the charge of the hidden group effectively couple to SM gauge bosons are called "portals". Typically, they enable to access part of the dark sector through mixing between the SM and the new gauge fields. The existence of new U(1) symmetries is extremely well motivated as it appears in supersymmetric theories, string theories [120] and Grand Unified Theories as

remnants of large groups broken down to the SM symmetry group and extra U(1)'s [121, 122]. Breaking of supersymmetric hidden symmetries is another well established way to generate such new abelian groups [123]. The bosons associated to these hidden U(1) are often called "dark photons". For a review of Dark Photon phenomenologies, see reference [124].

The existing constraints on the existence of low charge particles are typically expressed in the charge-mass plane, as shown in figure 6.1. On this plot, the electromagnetic charge is given as a fraction ϵ of the electron charge. We will use Q instead and omit the electron charge e after the fractional value. For instance, a particle of charge 2/3 e will be said to have a charge Q = 2/3. Many experiments have placed limits on the existence of low charge particles in large ranges of mass M and charge Q. We only discuss here the ones that are in the region of phase-space accessible at the LHC, i.e. mainly M below a few TeV and Q $\geq$ 0.1. As we will see in the analysis presented below, the real criterion (besides the available center-of-mass energy) is actually the stopping power dE/dx that depends on both the charge and the mass according to the Bethe function. In CMS, we are sensitive to the dE/dx region above a certain threshold. Below, the dE/dx is too low for our tracker to record the hit.

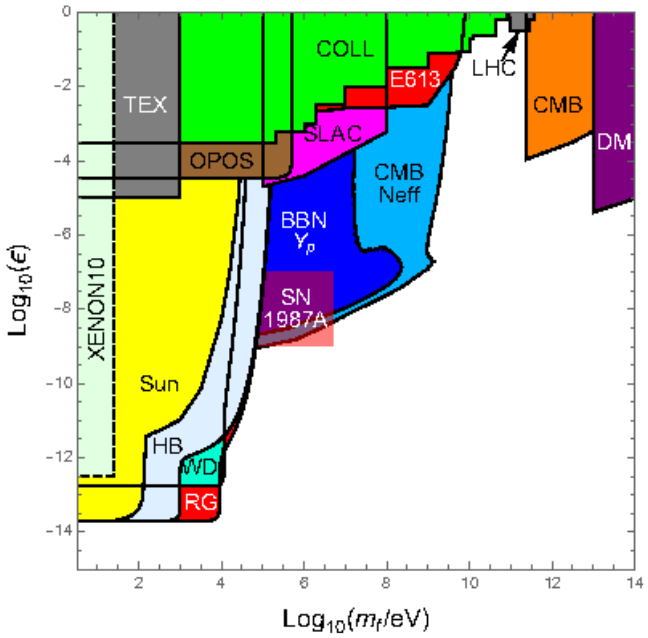


Figure 6.1: Constraints on the existence of low charge particles in the charge-mass plane [125] for a low charge particles-dark photon coupling of 0.1. ϵ stands for the effective electric charge of the low charge particles in units of the electron charge and m_f for the mass of the particles.

Let us now review the existing limits relevant in the phase-space of our search. At low masses, the main constraints in the phase-space of LHC searches come from previous collider experiments such as LEP. These exclude the existence of particles with charge Q = 2/3 at masses below 84 GeV [126]. At high masses, the most stringent limits are given by the CMB anisotropy and requirements on the relic abundance of the low charge particles. The CMB anisotropy is partly due to the acoustic oscillations of the baryon-photon plasma in the early moments of the universe. As pointed out in [125], it may be that low charge particles play a role similar to that of the baryons in this very hot environment. If this were the case, it would modify the anisotropy. Using WMAP data [127], this permits to set a limit on their abundance and hence on their mass and charge. This limit is shown in orange in figure 6.1.

The new low charge particles targetted in this section would interact with ordinary matter only through feeble electromagnetic interactions. Being massive, this makes them a good DM candidate in the WIMP scenario. This nicely connects this search to the ones presented in the previous chapter. In this context, their existence is constrained by cosmological observations as

their abundance must not exceed that of DM in the universe. In the hot early universe, the low charge particle number density just after the freeze-out depends on their annihilation cross-section which in turn depends on their charge and mass. In [128], it is shown that the lower limit on the cross-section translates into an upper limit on the mass: M $>$ 10^4 GeV for a coupling to the dark boson of 0.1. For higher masses, higher couplings are allowed leading to the purple triangle-shaped excluded region in figure 6.1.

At the LHC, CMS has already excluded part of the phase-space with two analyses published in 2013 at 7 and 8 TeV. Their results are given at the end of the next section, after having presented a minimal model used to help the interpretation of the results. Our goal is to push these limits further with the new data and a revised methodology.

On top of these limits, it is also possible to constrain the existence of these particles from the observation of cosmic rays. Indeed, being electrically charged, these new particles should be accelerated in cosmic accelerators just as charged SM particles are. Therefore, their corresponding "cosmic ray" energy spectrum should be detectable and have a predictable shape. In reference [129], data from the Super-Kamiokande experiment is used to constrain the existence of low charge particles from the observation of cosmic rays. The results are given in figure 6.2 in the same mass-charge plane (zoomed).

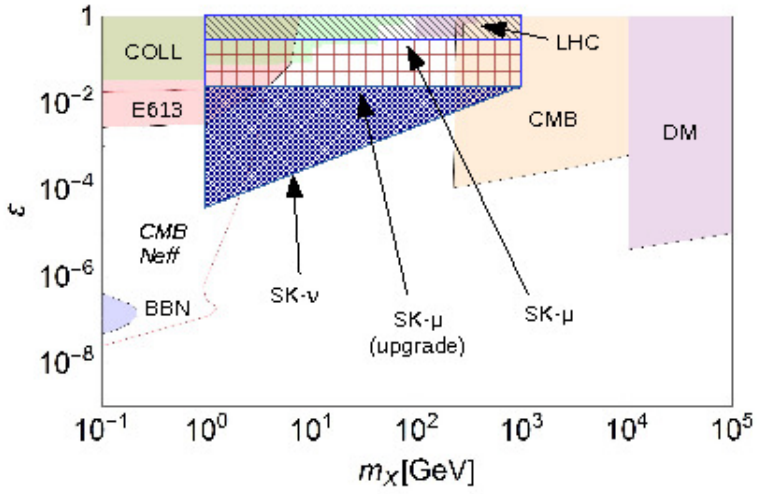


Figure 6.2: Constraints on the existence of low charge particles in the charge-mass plane [129] for a low charge particles-dark photon coupling of 0.1. ϵ stands for the effective electric charge of the low charge particles and m_X for their mass. Two methods are used in this work, one with a neutrino signature (SK-ν) and the other with a muon signature (SK-μ). The expected limits from the upgraded experiment with a muon signature are also shown.

6.2 A minimal model

In this section, we give more details regarding the kinetic mixing between the SM U(1)$_Y$ and the new U(1) gauge fields and how it yields fractionally charged particles. From reference [130], we see that adding a new U(1) gauge field A' coupling to a new Dirac fermion ψ to the SM Lagrangian:

$$\mathcal{L} = \mathcal{L}_{SM} - \frac{1}{4}A'_{\mu\nu}A'^{\mu\nu} + i\bar{\psi}(\not{\partial} + ie'\not{A}' + iM)\psi - \frac{\kappa}{2}A'_{\mu\nu}B^{\mu\nu}, \tag{6.1}$$

and redefining it:

$$A'_\mu \to A'_\mu - \kappa B_\mu, \tag{6.2}$$

a mixing term appears:

$$\mathcal{L} = \mathcal{L}_{SM} - \frac{1}{4} A'_{\mu\nu} A'^{\mu\nu} + i\bar{\psi}(\not{\partial} + ie'\not{A}' - ie'\kappa\not{B} + iM)\psi. \tag{6.3}$$

In these equations, B is the gauge field of the hypercharge U(1) symmetry group. This mixing is illustrated by the diagram in figure 6.3. SM particles (here quarks) interact through the SM U(1) gauge field that mixes with the new U(1) eventually decaying to a pair of new fermions. Through this kinetic mixing, the ψ field acquires a non-zero weak hypercharge $e'\kappa$, leading to its coupling to the SM photon and Z boson after electroweak symmetry breaking. The ψ electromagnetic charge $Q_\psi = e'\kappa\cos\theta_W = \epsilon e$ can take any arbitrary value coherent with the presented limits. A complementary way of understanding this is to use the equations of motion $\partial_\mu B^{\mu\nu} = eJ^\nu_{EM}$ to rewrite the mixing term as $-\frac{\kappa}{2} A'_{\mu\nu} B^{\mu\nu} = A'_\mu \times (\kappa e) J^\mu_{EM}$, where J_{EM} stands for the electromagnetic current [131]. Let us make the distinction between κ and ϵ more clear. The former defines the amplitude of the mixing between the two gauge groups as shown by the Lagrangian in equation (6.1). The latter is derived from the former and gives the effective electromagnetic charge of the new fermion ψ in units of the electron charge.

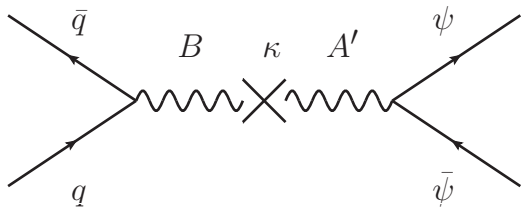


Figure 6.3: Feynman diagram illustrating the mixing between the SM U(1) gauge field B and the new U(1) gauge field A'. The parameter κ defines the amplitude of the mixing. SM particles (here quarks) interact through the hypercharge field B while the new fermions ψ interact through the new U(1) gauge group.

The tree-level production diagram with the observable states is drawn in figure 6.4. From now on, the L_Q symbol will be used to refer to the massive long-lived lepton-like with charge Q we are looking for. This is the model we use to simulate signal events and to design the analysis strategy. With only two parameters (the mass and the charge of the new particle), it can be used to target a large class of models containing massive low charge particles. In this work, we develop an analysis to look for particles carrying a charge in the range [$\frac{1}{3}$,1] in units of the electron charge.

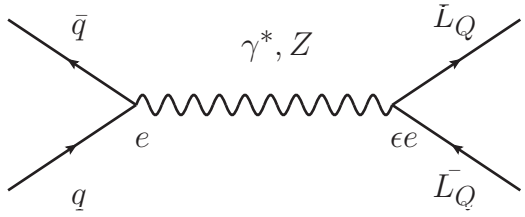


Figure 6.4: Feynman diagram for the production of a pair of fractionally charged particles. The parameter ϵ is to be understood as the effective electromagnetic charge of the new particles L_Q in units of the electron charge.

The strategy adopted is to use the low ionization power of the FCP to discriminate signal from background. In particular, we are looking for muon objects with a large multiplicity of low dE/dx hits on their inner tracks, where dE/dx is the ionization energy loss per unit length of the

hit in the tracker system. A CMS analysis using 2011 data at 7 TeV [132] and a similar procedure excluded the existence of such particles at 95% CL for masses below 310 GeV for a charge of 2/3 and for masses below 140 GeV for a charge of 1/3. Another CMS analysis with a different approach [133] excluded the existence of FCP at masses below 460 GeV for a charge of 2/3 and masses below 175 GeV for a charge of 1/3.

6.3 Signal simulation

We use the PYTHIA8 Monte Carlo (MC) event simulator [134, 135] to generate signal events according to the model introduced in the previous section. In this context, the FCP is called a τ' which is simply PYTHIA8 generic name for new leptons, influenced by fourth-generation vocabulary. We must enforce a pure vector coupling of the gauge bosons to the FCP due to the U(1) nature of the new gauge group. Unfortunately, PYTHIA8 does not allow the user to modify the coupling of SM gauge bosons to either SM or fourth-generation leptons. To overcome this issue, we choose to introduce a new gauge boson, called Z', that exactly replicates the behaviour of a Z and a γ with interference terms. We choose all its parameters to match the ones from the SM gauge bosons except a non-zero coupling to the τ' which is chosen to be vector only. This is equivalent to saying that the new leptons are not charged under SU(2) and therefore have no weak isospin, as shown by equations (6.4) and (6.5):

$$\begin{aligned} g_V &= I_3 - 2Q\sin^2\theta_W && (6.4)\\ &= 0.92 \\ g_A &= I_3 && (6.5)\\ &= 0 \end{aligned}$$

The full list of parameters is given here:

```
'NewGaugeBoson:ffbar2gmZZprime = on',
'Zprime:gmZmode = 5', # only Z'/gamma contribution
'32:m0 = 91.2',
'32:onMode = off', # switches off all Z' decay channels
'32:onIfAny = 17', # switches on Z'->tau'tau'
'Zprime:coup2gen4 = on',
'Zprime:universality = off',
'Zprime:vtauPrime = 0.92',
'Zprime:atauPrime = 0',
'17:m0 = 100.0',
'17:chargeType = -3',
'17:mayDecay = off',
'PhaseSpace:mHatMin = 200.0' # ensures convergence of the generation
```

`m0` is the mass of the FCP and `chargeType` is equal to three time the FCP charge in units of the electron charge. The last parameter enforces the invariant mass of the hard collision to be larger than twice the FCP mass. As we consider the mass of the FCP to be much larger than the Z boson pole, this helps convergence of the generator. Without this parameter, too few events have enough energy in the center-of-mass to produce two on-shell FCPs and the cross-section calculation fails.

Table 6.1 gives the signal cross-sections for a $Q = 1$ FCP in different mass scenarios with the presented PYTHIA8 configuration for 50000 events. These cross-sections are obtained after restricting the η range of the FCP tracks to the CMS tracker acceptance $[-2.6, 2.6]$. The corresponding cross-section for a FCP of charge Q is obtained by multiplying the $Q = 1$ cross-section by Q^2.

Mass (GeV)	Cross-section (pb)
100	7.381e-01 ± 2.463e-03
300	1.601e-02 ± 4.631e-05
500	1.941e-03 ± 5.521e-06
1000	5.713e-05 ± 1.583e-07

Table 6.1: $Q = 1$ signal cross-section for different mass scenarios. The uncertainty comes from the statistically limited number of generated events.

We validate our signal simulation against the one centrally produced by the CMS Collaboration. This is shown in appendix A.

6.4 Signal phenomenology

A deep understanding of the FCP-matter interaction is needed. Indeed, the way a particle interacts with the detector depends on its charge and mass and all cases must be rigorously distinguished and understood. A first general comment to be made is that due to the "Q =1" assumption for all reconstructed tracks, FCPs see their measured transverse momentum artificially enhanced by the inverse of their charge:

$$p_T^{\text{reco}} = \frac{1}{Q} p_T^{\text{true}} \tag{6.6}$$

Due to this feature, FCPs are "pushed" towards higher reconstructed p_T as is shown in figure 6.5 for a signal of mass 100 GeV. The p_T spectrum is observed to be stronger for lower L_{Q} charges (μ stands for the mean of the distribution). The fact that the $Q = 1/3$ spectrum gets suppressed at high p_T is due to a bad reconstruction for this very low charge scenario, as is developed right after when the Bethe curve is discussed. Nonetheless, the peak of the distribution is clearly shifted to higher transverse momentum. While this has no impact on the way particles actually interact with the detector, it has consequences at reconstruction level. For instance, it improves trigger efficiency as is discussed in section 6.6.

Now, let us start by describing the interaction between those particles and the matter of the detector. The energy deposited per unit length, i.e. the stopping power dE/dx, of a particle of charge Q is given by the Bethe function:

$$-\left\langle \frac{dE}{dx} \right\rangle = KQ^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \left(\frac{2m_e (c\beta\gamma)^2 T_{max}}{I^2} \right) - \beta^2 \right], \tag{6.7}$$

where A and Z are the mass atomic numbers of the atoms composing the material where the particle propagates, K is a constant, I is the medium mean excitation energy and T_{max} is the maximum energy transfer possible in a single collision [29]. This formula is only valid at intermediate momenta, i.e. $\beta\gamma$ between 0.1 and 1000, and does not take density effects into account. At higher energies (several hundreds of GeV), one must correct for radiative effects such as pair production and bremsstrahlung that increase the energy deposit per unit length. Figure 6.6 shows the evolution of the stopping power in copper as a function of the particle momentum for different mass and charge scenarios. The muon curve is superimposed for comparison purposes.

Several things can be deduced from those two graphs. First of all, the heavier the particle, the larger the energy it deposits by ionization in the detector at typical momenta in CMS analyses. On the other hand, the smaller its charge, the smaller the energy deposit. This is due to the Q^2 factor in the Bethe function that supresses signals from low-charge particles. Another thing that should be stated is that the Bethe curve displays two regimes in the momentum range considered: a steeply falling regime at low momentum and the minimum-ionizing plateau at high momentum. All those considerations should be taken into account when thinking about the way

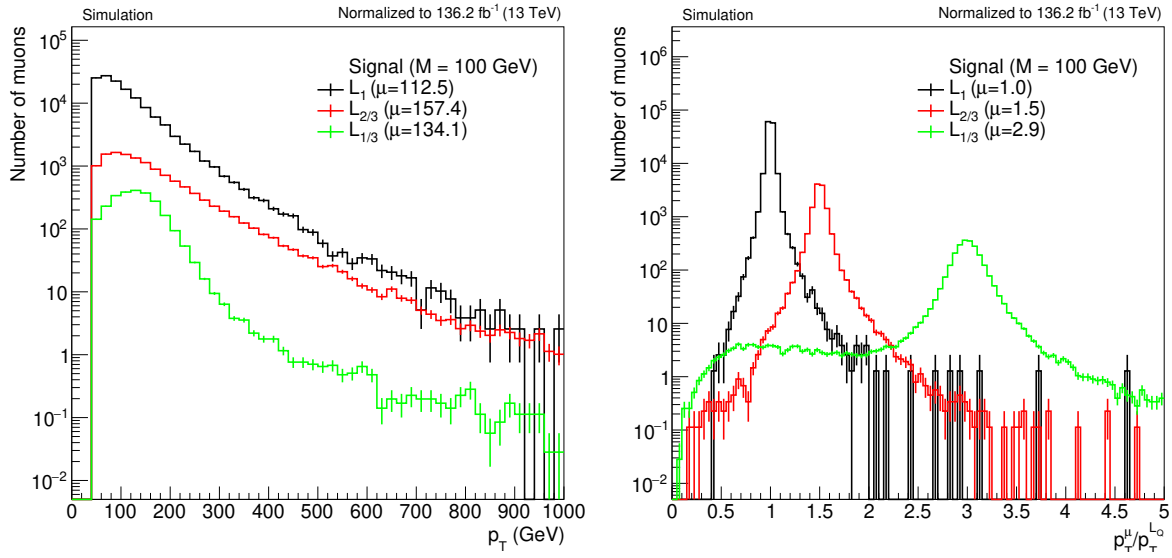


Figure 6.5: Transverse momentum distribution for muons geometrically matched to a generated-level L_Q particle (left) and ratio of the muon to the L_Q momentum (right) for a signal of mass 100 GeV. μ stands for the mean of the distribution. The events have been generated under the 2016 CMS data-taking conditions and the number of events is normalized to the full Run 2 integrated luminosity.

an FCP deposits energy in both the tracker and the muon chambers.

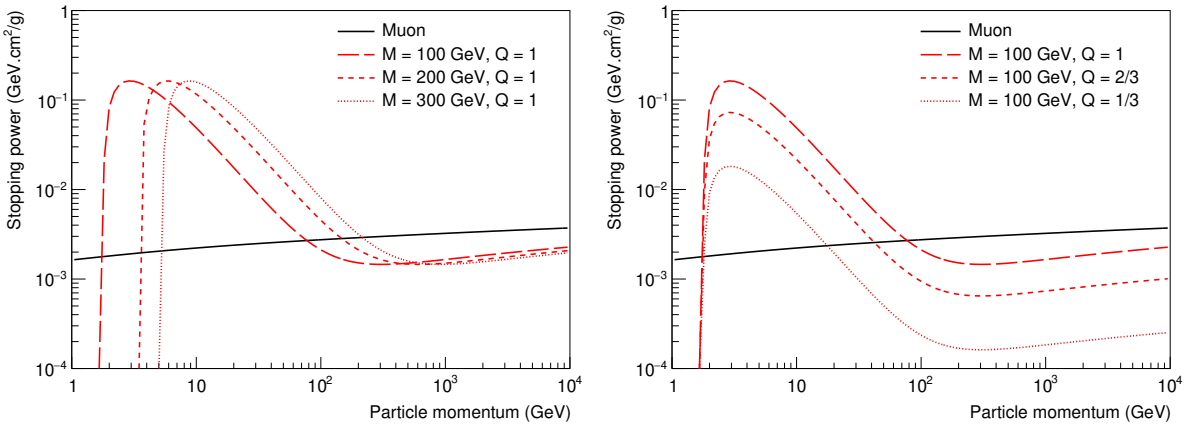


Figure 6.6: Bethe function for L_1 in different mass scenarios (left) and for L_Q with mass M_{L_Q} = 100 GeV in different charge scenarios (right). The curve for the muon is also displayed.

The first consequence of the previous discussion is that the reconstruction efficiency of an FCP as a given object is very different depending on the object (inner track or muon) and the FCP mass and charge. In figure 6.7 the differential efficiency for an FCP of mass 100 GeV to be reconstructed as a track or a muon (3 different muon types are shown) is displayed. From these graphs, one can conclude that FCPs with a too low ionization power are not being properly reconstructed by the CMS tracker. The reason for this is that the tracker modules work in a zero-suppression mode meaning that hits with a charge lower than a given threshold (chosen with respect to the detector's noise level) are not read out [70]. When the number of reconstructed hits gets too low, the tracking algorithm fails and the track is lost. On the other hand, it is worth noting that the FCP is still very well reconstructed as a stand-alone muon. The reason for this may be manyfold. First of all, no zero-suppression is applied in the muon chambers. Second, the fact that the muon chambers are located further away from the beamline makes the track density

much lower at this radius and therefore the tracking algorithm has an easier job.

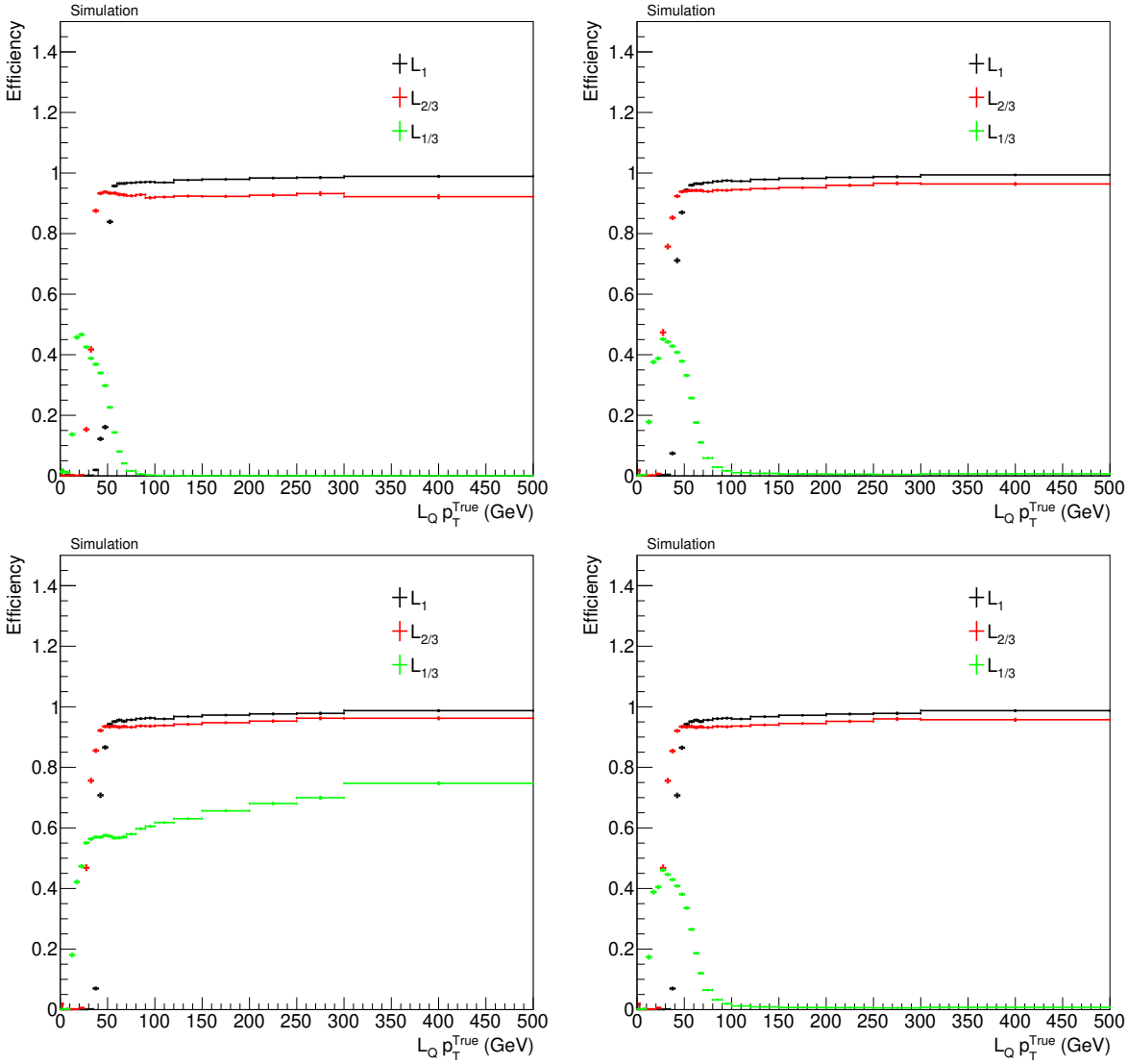


Figure 6.7: Differential efficiency for an FCP of mass 100 GeV to be reconstructed as a track (top left), a tracker muon (top right), a stand-alone muon (bottom left) or a global muon (bottom right). The p_T threshold at 50 GeV arises from the trigger requirement. Interestingly, the impact of the 1/Q enhancement of the reconstructed p_T can be observed here: lower charge particles trigger at lower true p_T. This is why the red curve for Q = 2/3 starts to turn on at $p_T \sim 30$ GeV.

The second consequence of the low ionization power of the FCP is that their dE/dx spectrum is very different from charge 1 particles. Figure 6.8 illustrates that by comparing the reconstructed dE/dx spectrum of tracker hits on track arising from L_1, $L_{2/3}$ and $L_{1/3}$. This feature will be used to discriminate signal from background.

6.5 Datasets

6.5.1 Data

We analyze pp collisions data recorded by CMS in 2016, 2017 and 2018. It corresponds to a total integrated luminosity of 136 fb^{-1} as shown in table 6.2. In CMS, data is split into chronological *eras* listed alphabetically. For instance, in 2016 data was split into seven eras, B (early data taking) to H (late data taking). A change in era occurs whenever the experimental conditions change significantly.

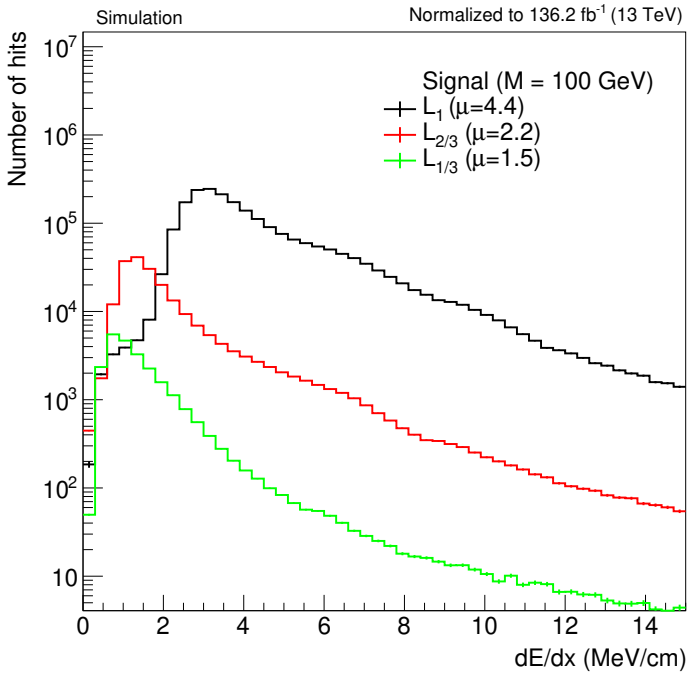


Figure 6.8: Stopping power spectrum for FCP of mass 100 GeV. The lower the charge, the lower the peak of the spectrum. The mean μ of the distribution is also given.

Year	Integrated luminosity (fb^{-1})
2016	35.9
2017	41.5
2018	58.8
Total	136.2

Table 6.2: Data-taking years and corresponding integrated luminosities.

6.5.2 Simulations

We estimate the expected number of events from SM backgrounds from data. However, we use MC simulations to derive signal corrections (as explained later) and the corresponding systematic uncertainties. The background MC processes considered in this search are listed on table 6.3 with the corresponding order in perturbation used at generation and the cross-section. The order at which the cross-section is computed is shown in the parenthesis and is in general different than the order used for generation.

Process	Order at generation	Cross-section (pb)
W($l\nu$)+jets	NLO	61527 (NNLO)
DY(ll)+jets	NLO	5765 (NNLO)
$t\bar{t}$+jets	NLO	832 (NNLO)
WW	LO	76 (LO)
WZ	LO	28 (LO)
ZZ	LO	12 (LO)

Table 6.3: Monte Carlo datasets, order of the computation and order of normalization of the cross-section.

6.6 Triggers

Two distinct effects showed up when studying the signal phenomenology: low charge particles see their reconstructed p_T artificially enhanced by 1/Q and the efficiency to reconstruct particles with $Q \leq 1/3$ is very low. Both these effects yield E_T^{miss} in the global event description. The first one as it brings imbalance from the overestimation of the particle's transverse momentum, and the second one as the track may simply be lost. These two effects are actually working against each other and care should be taken regarding their interpretation. Let us start the discussion with the simpler case in which both FCP tracks are reconstructed. For $Q = 1$ particles, the E_T^{miss} should be exactly zero (up to the detector resolution). For $Q < 1$ particles, the value of E_T^{miss} depends on the transverse momentum of the Z/γ boson, i.e. on the hadronic recoil in the transverse plane. For high enough $p_T^{Z/\gamma}$, the two FCPs are boosted along the direction of the boson momentum and each of them inherits from about half of the boson p_T. Consequently, the p_T overestimation (and therefore the expected E_T^{miss}) is given by:

$$\begin{aligned} E_T^{\text{miss}} &= \left| p_T^{Z/\gamma} - 2 \times \frac{1}{Q} \times \frac{p_T^{Z/\gamma}}{2} \right| \\ &= \left| \frac{Q-1}{Q} \times p_T^{Z/\gamma} \right| \end{aligned} \tag{6.8}$$

As $Q < 1$, the signed value of the E_T^{miss} (or the value of the missing transverse momentum $\vec{p}_T^{\,miss}$) is always the same. We take as a convention the direction of $\vec{p}_T^{\,miss}$ to be positive when pointing towards the hadronic recoil. This is illustrated in figure 6.9 for a boson p_T of 100 and 300 GeV. In the scenario where one track is lost, the picture is a bit more subttle. The lost and the reconstructed tracks both add E_T^{miss} but in opposite directions. Indeed, the lost track adds E_T^{miss} along the direction of the missed particle while the reconstructed track adds E_T^{miss} in the opposite direction. In this case, the missing transverse momentum can be oriented in either directions. Similarly to the "2-reconstructed tracks" case, a simple calculation yields:

$$\begin{aligned} E_T^{\text{miss}} &= \left| p_T^{Z/\gamma} - \frac{1}{2} \times \frac{1}{Q} \times p_T^{Z/\gamma} \right| \\ &= \left| \frac{2Q-1}{2Q} \times p_T^{Z/\gamma} \right| \end{aligned} \tag{6.9}$$

This is shown in the same figure for the two same boson p_T.

With this study in mind, one might be tempted to select events using an E_T^{miss} trigger. Unfortunately, the following observations are to be made. In the "high charge" FCP scenario, i.e. $Q \geq 1/2$, most events have both tracks reconstructed as the tracking efficiency is still high. This means that in that case, the relevant region to look at is the right half of figure 6.9 for the blue curve. Very large boson p_T are needed to reach high E_T^{miss} and the corresponding cross-section, $\sigma(pp \to L_Q \bar{L}_Q + \text{high } p_T \text{ jets})$, will be small. On the other hand, low charge FCP have typically only one track reconstructed. In this case, the region of interest is $Q \in [1/3, 1/2]$ for the red curve. Similarly, large E_T^{miss} would require large hadronic recoil and the cross-section amplitude would suffer from it. At even lower charges ($Q < 1/3$), reconstruction completely fails and no tracks are reconstructed. The E_T^{miss} then simply equals the boson p_T.

This reasoning is verified when looking at the actual reconstructed E_T^{miss} distribution in simulated FCP events for a FCP mass of 100 GeV. This is shown in figure 6.10. Unfortunately, the CMS pure E_T^{miss} trigger has a threshold at 170 GeV which is much larger than the mean E_T^{miss} observed for any charge. Even using the E_T^{miss} triggers computed without the muons, the threshold of 120 GeV would be too high. With these conclusions, we decided to focus entirely on

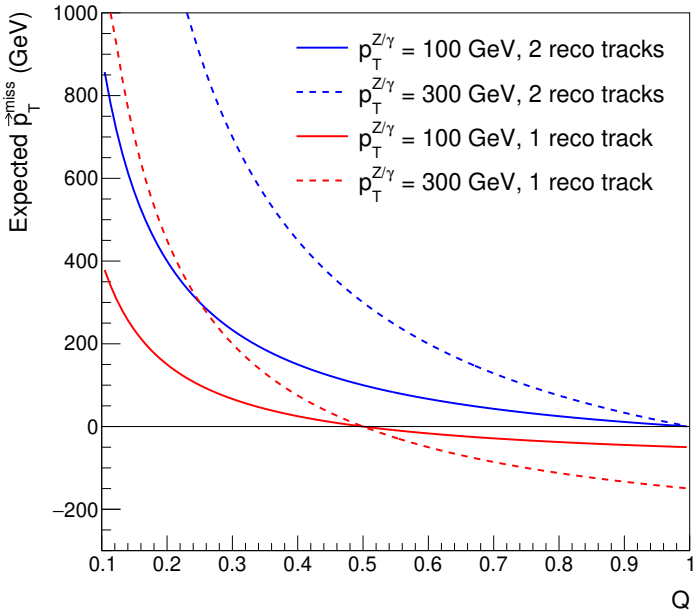


Figure 6.9: Expected $\vec{p}_T^{\,miss}$ in events with one or both FCP tracks reconstructed. The sign of the missing transverse momentum vector is conventionally taken as negative in the direction of the Z/γ boson, i.e. the direction of the boost of the FCP particles. When both tracks are reconstructed, the missing transverse momentum always points in the direction opposite to the FCP particles system. The missing energy increases with increasing recoils and decreasing charges. When only one track is reconstructed, its p_T enhancement competes against the loss of the other track. It "wins" below Q = 1/2.

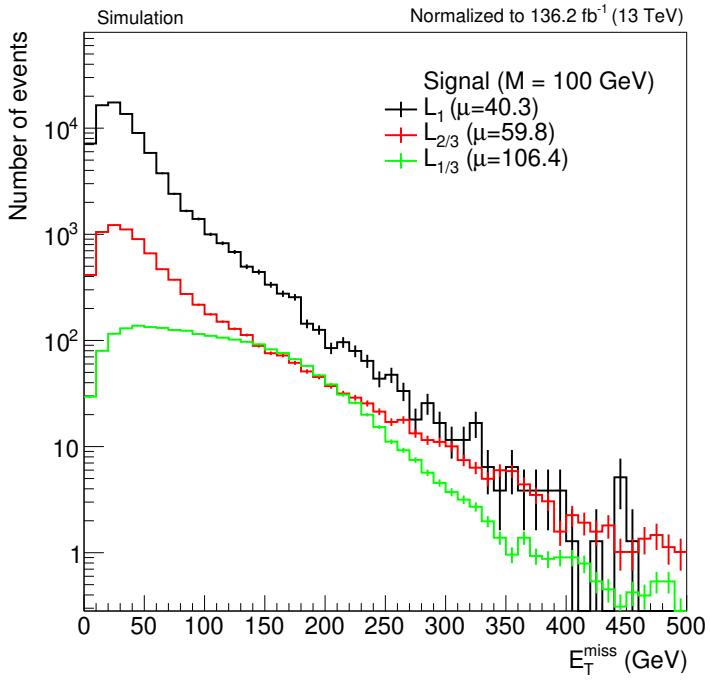


Figure 6.10: Reconstructed E_T^{miss} distribution in simulated FCP events for a FCP mass of 100 GeV. μ stands for the mean of the distribution. As discussed in the text, the expected E_T^{miss} is not large enough for the bulk of the events to pass the CMS E_T^{miss} triggers thresholds.

the muon triggers to collect events of interest.

More explicitely, candidate events are selected with the help of single-muon triggers. These

triggers select the event if a muon is found that has a p_T higher than the trigger threshold. The muon p_T spectrum being dominated by muons from Z and W decays with a transverse momentum around half the boson mass (the so-called Jacobian peak), we must cut higher in the muon p_T in order for our signal not to be drowned in those peaks. This motivates our choice not to use the unprescaled path with the lowest threshold (`HLT_IsoMu24`). In 2016, we the `HLT_TkMu50` path is combined in a logical OR statement with the main `HLT_Mu50` path to obtain the best possible efficiency. In 2017 and 2018, the main path (`HLT_Mu50`) is combined in a global OR statement with the `HLT_OldMu100` and the `HLT_TkMu100` paths to cope with an observed selection inefficiency. This analysis uses tracker tracks as the main objects to select on, which means that to trigger on events with no track is of no use for us. Consequently, we define the trigger efficiency as given by:

$$\varepsilon = \frac{\text{Denominator AND trigger AND } L_{\text{Q}} \text{ track matched to a muon track}}{\text{Events with at least one } L_{\text{Q}} \text{ in CMS}} \tag{6.10}$$

In equation (6.10), the L_{Q} track in the numerator is a reconstructed track, while in the denominator we ask for a generated-level L_{Q} within the CMS tracker acceptance, i.e. $\eta \in$ [-2.5,2.5]. The inclusive trigger efficiency is shown on table 6.4 for different charge and mass scenarios. No corrections are applied at this point.

Table 6.4: Inclusive efficiency of the combined muon triggers for different charge and mass scenarios comparing the three data-taking years.

	$M = 100$ GeV			$M = 500$ GeV		
	Q = 1	Q = 2/3	Q = 1/3	Q = 1	Q = 2/3	Q = 1/3
2016	70.4 %	75.8 %	0.7 %	68.3 %	67.9 %	1.9 %
2017	68.9 %	76.1 %	1.2 %	69.5 %	69.4 %	3.2 %
2018	69.0 %	75.4 %	1.1 %	68.6 %	68.2 %	3.0 %

Figure 6.11 shows the differential efficiency of the L1 seed to the `HLT_Mu50` path (`L1_SingleMu22 OR L1_SingleMu25`) on a sample of simulated signal events for $M_{L_{\text{Q}}} = 100$ GeV reconstructed under 2016 conditions. On Figure 6.12 is shown the differential efficiency of the `HLT_Mu50` path itself on a sample of simulated signal events for $M_{L_{\text{Q}}} = 100$ GeV under the same conditions. The corresponding plots for 2017 and 2018 are given in appendix B. The efficiency is given for a single path as it brings no additional information to draw it for the combined path. The inefficiency visible between 60 and 100 GeV in figure 6.11 (left) is due to the large mass of the L_{Q} particles. In this momentum range, their velocity is significantly lower than the speed of light, causing them to reach the muon chambers later. This effect was checked not to be present (see appendix B for a brief study on timing) in signal events with a L_{Q} mass of 1 GeV nor in Drell-Yan events, which confirms that the large FCP mass is the source of it. In figure 6.11 (right), the same effect is deformed due to the fact that the reconstructed L_{Q} p_T is overestimated by the inverse of the charge.

By comparing the efficiency curves at L1 and HLT, one can conclude that the inefficiency only comes from the Level-1 trigger. At HLT the efficiency curves are essentially the same as to L1, except that a muon is then also defined with its inner track and not only in the muon chambers. This has a direct impact on the reconstruction efficiency for low charge L_{Q} tracks as discussed in section 6.4. If the FCP loses its inner track, the muon is not a valid object to trigger on at HLT anymore (a global muon is required) and the efficiency drops. This explains the difference in the efficiency curves at L1 and HLT for $L_{1/3}$.

One can further notice a difference between the black (L_1) and red ($L_{2/3}$) efficiency curves. $L_{2/3}$ particles start to trigger at lower p_T. This is due to the p_T enhancement effect: a FCP of a given true p_T is actually reconstructed with a p_T larger by a factor $\frac{1}{Q}$.

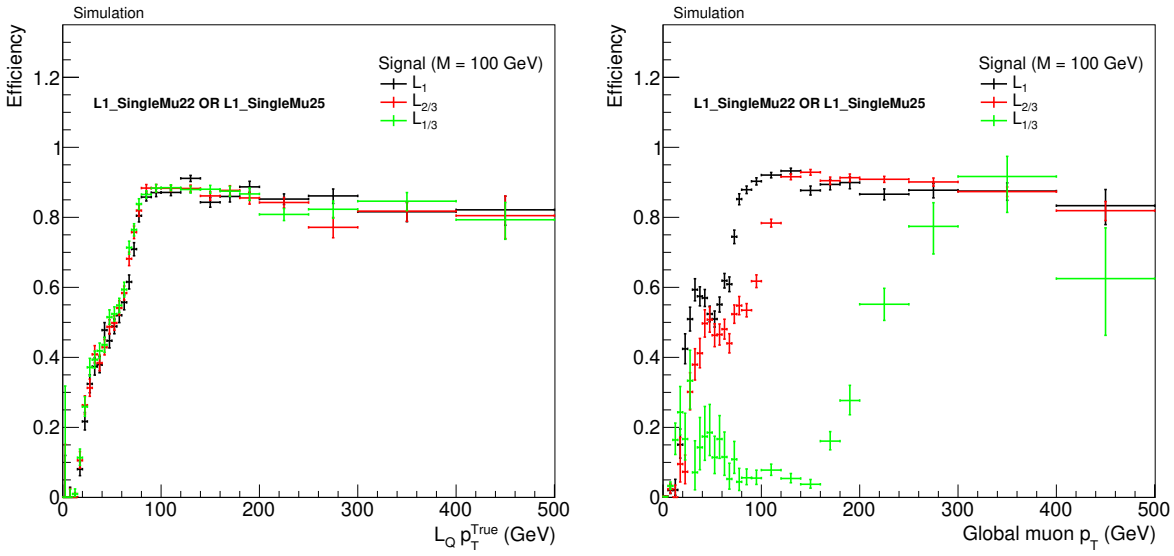


Figure 6.11: Efficiency of the `L1_SingleMu22 OR L1_SingleMu25` seed on signal as a function of the L_{Q} true p_T (left) and as a function of the matching reconstructed muon p_T (right) under 2016 conditions.

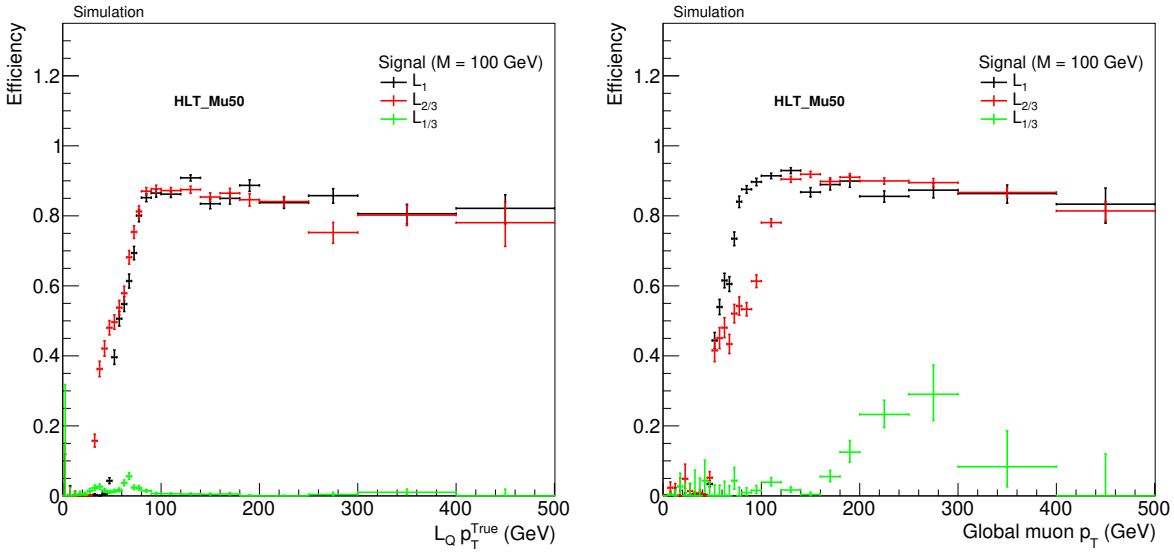


Figure 6.12: Efficiency of the `HLT_Mu50` path on signal as a function of the L_{Q} true p_T (left) and as a function of the matching reconstructed muon p_T (right) under 2016 conditions.

6.7 Physics objects

In this section, we go through the objects used to select our signal events and we define the associated quantities relevant for the analysis. In section 6.9, we show the cut values applied on those same quantities.

6.7.1 Tracks

The main objects in this analysis are the tracks measured in the CMS inner tracker. Track objects contain the following information of interest for the analysis:

- p_T: transverse momentum
- η: pseudorapidity

- d_{xy}: transverse impact parameter of the track with respect to the primary vertex if one is found, otherwise with respect to the beam spot.
- d_z: longitudinal impact parameter of the track with respect to the primary vertex if one is found, otherwise with respect to the beam spot.

We also compute α_{max}, the maximum 3-dimensional angle between the candidate track and any other high p_T (> 35 GeV) object. This variable is set to zero when no high p_T object other than the candidate track itself is found in the event. To each track is associated a collection of hits with charge and pathlength measurement. From this we compute the stopping power dE/dx for each hit by dividing one by the other. The number of hits in the strip and the pixel detectors will also be used as selection criteria. In the process of selecting a candidate track, a matching of the track to a closeby muon is applied. Therefore, we introduce the muon object and its main characteristics in the next subsection.

6.7.2 Muons

The following information is available for a muon object (see section 3.3.3 for a detailed description of the different reconstructions):

- isLoose: whether the muon fulfills loose identification requirements.
- isGlobal: whether the muon is *Global*, i.e. is reconstructed simultaneously in both the muon system and the tracker.
- PF isolation: for a muon μ, this variable is defined by equation (6.11) where all considered particles must be in a ΔR cone of 0.4 around the muon.

$$\begin{aligned} \text{PF isolation} \;=\; & [\Sigma p_T(\text{charged hadrons from the primary vertex}) \\ & +\max(0, \Sigma E_T(\text{neutral hadrons}) + \Sigma E_T(\text{photons}) \\ & -\frac{1}{2}\Sigma p_T(\text{charged hadrons from pileup}))]/p_T(\mu) \end{aligned} \tag{6.11}$$

- time$_{IP}$: time at which the muon emerged from the interaction point (IP) given it propagates at the speed of light. It is centered around zero for light, prompt particles produced in the hard interaction. It takes larger positive values for heavier, i.e. slower particles. Cosmics arrive at random times and therefore show a flat distribution.

6.8 Instrumental sources of background

Before presenting the event selection, we must introduce the sources of background from detector effects. In [136], two instrumental sources of background are described that we address here as well. We explain why they are problematic and how we deal with them.

6.8.1 Radiation damage in the tracker

The pixel detector being the closest subdetector to the beam, it is the most affected by radiation damage. Hadrons created in proton-proton collisions modify the silicon crystal structure by adding what is called "impurities". This generates additional energy states in the silicon junctions that can trap free electrons, in particular those created by ionization of the medium after the passage of a particle in the junction. As a consequence, part of the ionization charge stays inside the crystal and the signal read out is weaker. In other words, radiation degrade the energy of some hits and therefore lower their dE/dx. The hits dE/dx spectrum in data then gets shifted to lower values, making the background more signal-like. This directly translates in more signal-like

tracks with a high number of low dE/dx hits. As we select our signal on the number of low charge hits, this effect enhances the background of the analysis. Figure 6.13 gives an example of this effect by comparing the stopping power in the three layers of the pixel detector in early 2016 data (eras B to F, 19.7 fb^{-1}). The analysis selection is already applied here (see section 6.9). By comparing the MC prediction (where radiation damage is not simulated) to the data, one can see that the first layer is by far the most affected by radiation. We suppress the impact of this effect by ignoring hits from the first pixel layer in the hit collections of the tracks. This does not change the number of tracks in the signal region but has a significant impact on background rejection. This impact will be quantified in section 6.10 when the signal extraction variable will be defined.

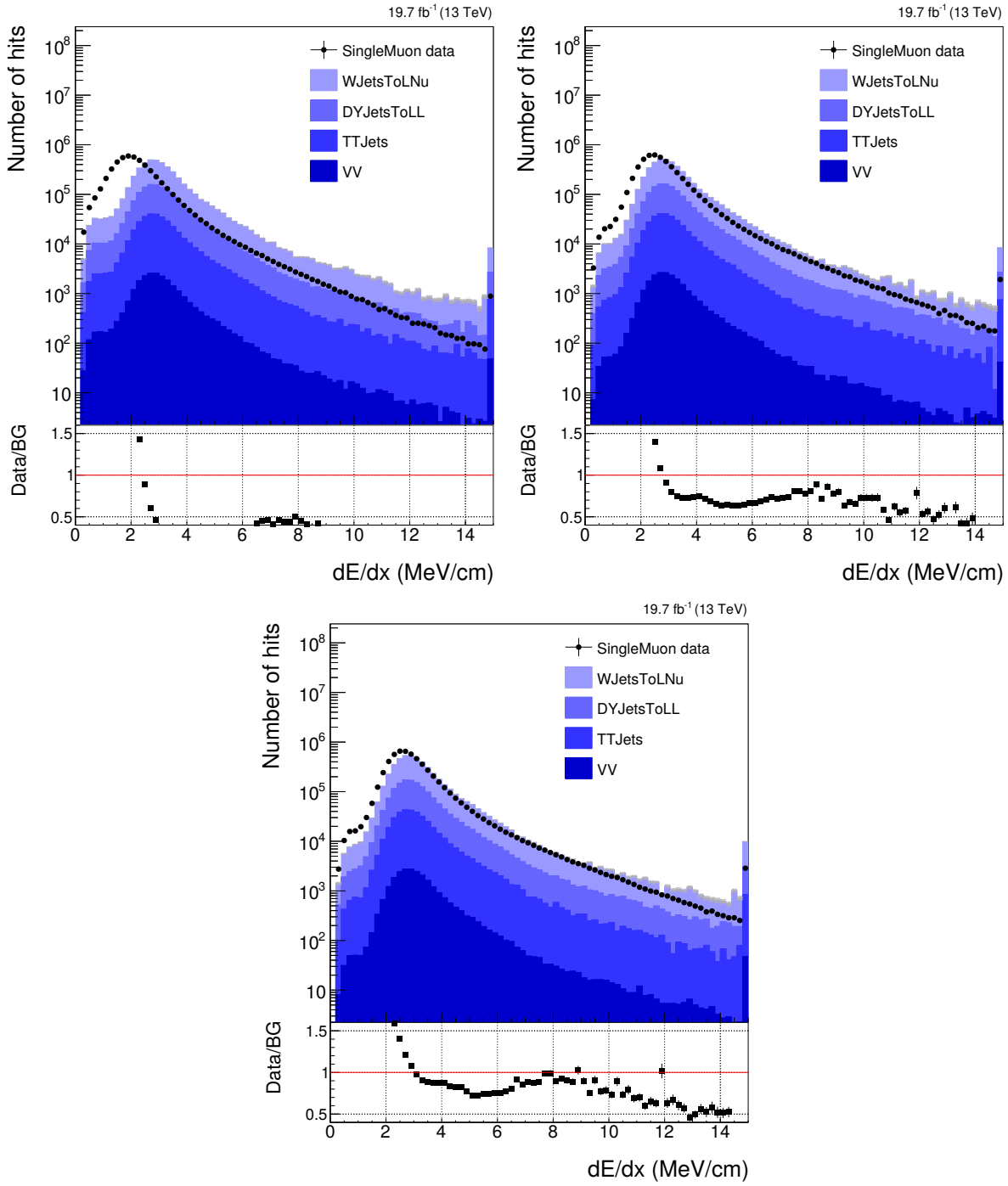


Figure 6.13: Stopping power spectrum for the hits in the pixel barrel for early 2016 data (eras B to F, 19.7 fb^{-1}). Layer 1 (top-left), layer 2 (top right) and layer 3 (bottom).

6.8.2 Hits near the edge of a tracker module

Another experimental effect that affects the expected number of candidate tracks is related to the edges of the tracker modules. When a particle leaves its energy close to the edge of a module, part of the charge is not collected therefore leading to a fake "low charge" hit. Unfortunately, tracks hitting such areas in the barrel have a higher probability of hitting more of them due to the tracker barrel geometry. Those tracks then look exactly like signal tracks, i.e. tracks with a high multiplicity of low dE/dx hits. This effect is dealt with by simply ignoring any track containing at least one hit close to the edge of a tracker module. Due to the $|\eta| < 1.5$ restriction, most tracks are in the barrel as is shown in table 6.5. However, for forward tracks, a significant proportion of their hits is in the TEC. We must therefore reject edge hits in this region as well.

η range		PXB	PXF	TIB	TID	TOB	TEC
<1.5	N_{hits}	2.96	0	5.09	0.36	5.56	1.56
	Prop. (%)	19.3	0	32.6	2.3	35.5	10.2
[0.9,1.5]	N_{hits}	2.97	0	4.39	0.87	2.90	3.81
	Prop. (%)	20.2	0	29.5	5.6	19.7	25.0

Table 6.5: Mean number and proportion of hits (%) per track in the given tracker subsystem.

As opposed to the rectangular geometry of the tracker modules in the barrel, the geometry in the endcaps is trapezoidal and depends on the ring. Therefore, one must transform the module's local X coordinate in order to be able to place a cut on it independent of the local Y coordinate. These transformations are simple $\mathbb{R}^2 \to \mathbb{R}^2$ linear transformations:

$$X' = aX + bY, \tag{6.12}$$

where a and b are determined by solving a system of two equations that depends on the shape of the trapezoid (large side on top or at the bottom) and the sign of X. For TEC ring 5 and negative values of X, it reads:

$$\begin{cases} X_{\min} = aX_{\min} + bY_{\min} \\ X_{\min} = aX_{\max} + bY_{\max} \end{cases} \tag{6.13}$$

where ($X_{\min}$,$Y_{\min}$) and ($X_{\max}$,$Y_{\max}$) are the coordinates of the bottom-left and top-left points of the trapezoid. They define the slope of the left edge of the figure. Their values for the left edge of ring 5 are: ($X_{\min}$,$Y_{\min}$) = (-6.0, -8.0) and ($X_{\max}$,$Y_{\max}$) = (-4.7, 8.0). Let us note that they need not be exactly on the edge: only the slope they define matters for the transformation. Solving the system yields:

$$\begin{cases} a = \dfrac{X_{\min}(Y_{\min} - Y_{\max})}{X_{\max}Y_{\min} - X_{\min}Y_{\max}} \\ b = \dfrac{X_{\min}(X_{\max} - X_{\min})}{X_{\max}Y_{\min} - X_{\min}Y_{\max}} \end{cases} \tag{6.14}$$

Plugging in the values of $X_{\min}$, $Y_{\min}$, $X_{\max}$ and $Y_{\max}$ yields $a = 1.12$ and $b = -0.09$. A similar treatment is done for the positive values of X and for ring 6 and 7. Rings 1 to 4 are outside the $|\eta| < 1.5$ restriction of the analysis. We show the tracker module geometry before and after transformation in rings 5, 6 and 7 of the TEC in figure 6.14.

The X and Y local coordinates of the hits in the PXB and TIB tracker modules are shown in figure 6.15 with the chosen cut value for a subset of 2016 data (10 % of era B). The corresponding figures for the TOB and the TEC are given in appendix C. Only hits with a stopping power lower than 2 MeV/cm are displayed, showing the accumulation of such hits on the edges of the modules. The analysis selection is already applied here (see section 6.9). This mitigation procedure effectively removes tracks from the signal region and should be thought of as a selection

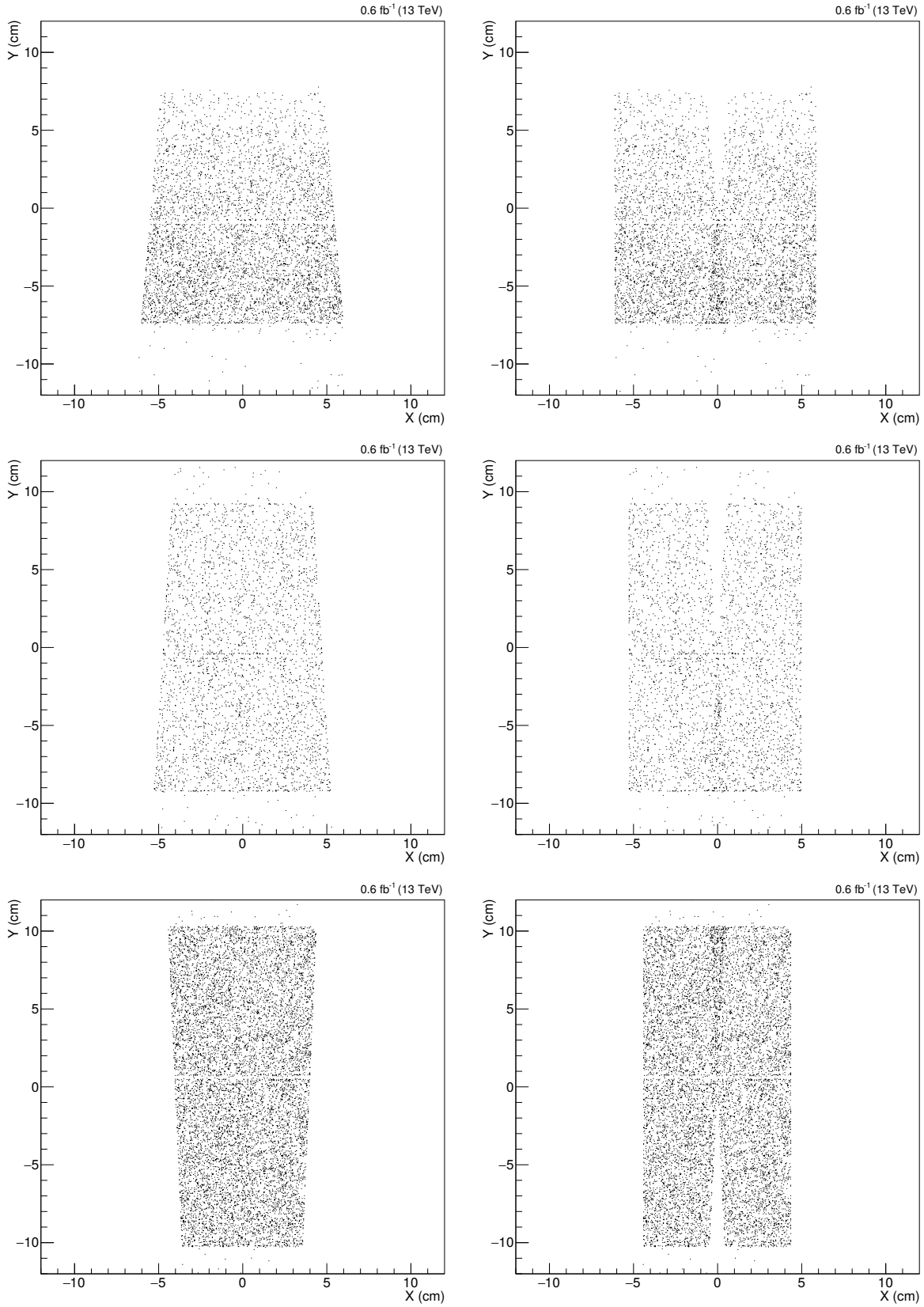


Figure 6.14: Two-dimensional map of the hits local coordinates in the Tracker Endcaps before (left) and after (right) transformation of the X coordinate for rings 5 (top), 6 (middle) and 7 (bottom). Only 10 % of 2016 data (eras B to F) is used.

cut. Its efficiency is given in the event selection cut flow in section 6.9. The impact on the expected number of background is quantified in section 6.10 when the signal extraction variable will be

defined. As these cuts only depend on the geometry of the tracker modules, they do not depend on time (i.e. on the era). Although in 2017 and 2018 a new pixel detector was installed, its module have the same geometry (this was checked explicitely) and therefore the cuts stay the same for the three years.

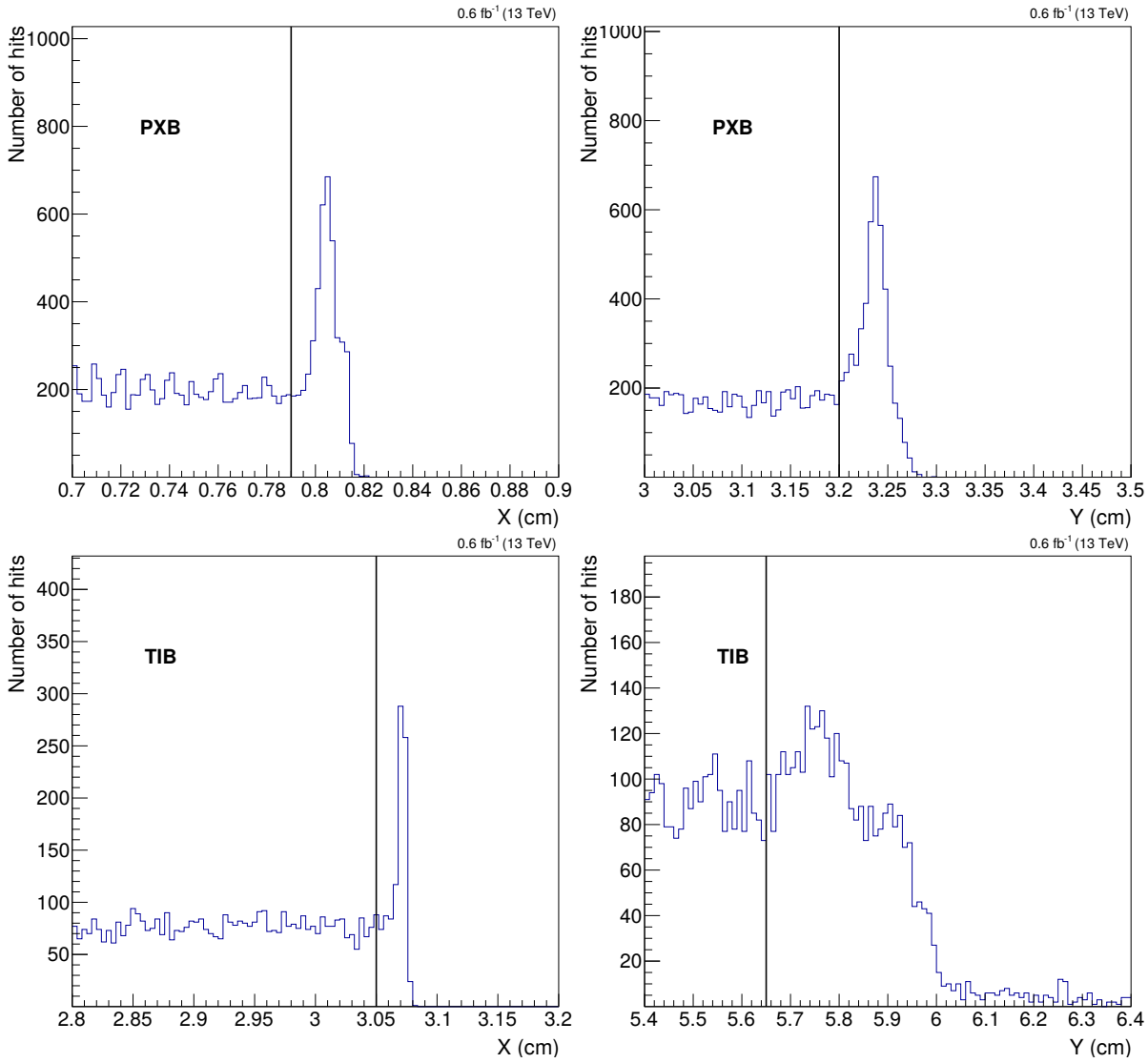


Figure 6.15: Absolute value of the local X (left) and Y (right) coordinate of hits with dE/dx $<$ 2 MeV/cm in the Pixel Barrel (top) and the Tracker Inner Barrel (bottom). Only 10 % of 2016 data (eras B to F) is used. An accumulation of hits is observed at the edges which is removed by cutting where the line is placed.

6.8.3 Signal correction at low dE/dx

Ignoring hits from the first pixel layer is one way to handle background from instrumental sources. However, the remaining tracker layers (both in the pixel and the strips detectors) are also affected by radiation damage and other experimental effects (discussed in section 6.10). From figure 6.13 one notices that simulations do not reproduce well the degradation of the ionization energy, mainly in the low dE/dx region (below 2 MeV/cm). On this figure, the analysis selection (see section 6.9) is applied. Since our signal is simulated using the exact same conditions as for background processes, it is necessary to correct for these data-to-MC discrepancies. The idea of the correction procedure is to degrade the energy of a given small proportion of hits, chosen randomly. Each chosen hit sees its dE/dx reduced by a second random factor, generated according to the Weibull distribution [137] in the $[0, 1]$ range. The choice of the Weibull distribution to generate the random

degradation factors is not motivated by strict physical arguments. It is chosen for its convenient, modulable, shape and the fact that it drops steeply near zero. This allows to avoid the excessive accumulation of hits in the low dE/dx region after correction. The proportion of degraded hits as well as the Weibull parameters are layer-dependent. This ensures a good data-to-MC agreement layer by layer. Given the physical causes of the accumulation of hits in the low region of the dE/dx spectrum in data, i.e. tracker modules not collecting all the deposited energy, this method is well motivated. Figure 6.16 shows the dE/dx spectrum in data and in the simulations (before and after corrections) for the Pixel Barrel (PXB) in early 2016 data (eras B to F, 19.7 fb^{-1}). Again, the analysis selection is applied to the corresponding events. The equivalent figures for the other tracker subdetectors and for eras G and H of 2016 and for 2017 and 2018 are given in appendix D. Once derived on background MC processes, these corrections are applied to the signal MC.

The validity of the correction method is cross-checked with the help of the control region (defined in section 6.9). Corrections derived using BG MC in the search region (also defined in section 6.9) are applied to BG MC in the control region and compared to data in the control region. Figure 6.17 shows the good agreement between data and MC in PXBL2, TIBL1 and TOBL1 for 2016 data (eras B to F, 19.7 fb^{-1}). Later (in section 6.10), when we will have defined the variable we will use to look for a signal, we will have another way of estimating the impact of those corrections.

Although we already split Run 2 data in four subsets according to the radiation damage (see section 6.10), one might wonder whether the corrections are consistent along a given data period. Figure 6.18 shows the corrections applied in the PXB Layer 2, TIB Layer 1 and TOB Layer 1, comparing era B and F of 2016 data. In the PXB and in the TIB, the corrections are quite stable between the two eras although one can observe that era B is slightly overcorrected while era F is slightly undercorrected. This effect is stronger in the TOB L1. Since the correction itself is much stronger than its evolution with time, we do not expect this effect to play any role. It is important to assess that the corrections derived in the full data period do not average a strong change in the behaviour of data with time. This shows that it is not the case. A systematic uncertainty is assigned to these corrections as detailed in section 6.12.

6.8.4 Very low dE/dx hits

Figure 6.19 shows a zoom of the dE/dx spectrum in the $[0, 0.4]$ region for the PXBL2, PXBL3, TIB and TOB in early 2016 data. We observe that in the TIB, some hits populate the very low dE/dx region. We decide to cut them off as we do not expect those hits to come from physical sources. The lower bound is set at 0.25 MeV/cm for all layers. Tracks containing such hits are kept in the selection with simply these hits ingored.

When correcting signal MC (see previous section), we degrade the dE/dx of some hits. This can lead to the concerned hits ending up with an energy below the zero-suppression threshold necessary for detection. If the energy of a corrected hit falls below the lower bound of the corresponding layer, we ignore the hit so that the very low dE/dx region does not get repopulated by the effect of the corrections.

6.9 Event preselection

Prior to applying our signal selection, events are preselected if they contain at least one candidate track. A candidate track is a high p_T, clean track matched to a muon. A track is considered to be matched to a muon is they are found within a ΔR cone smaller than 0.05. Some of the following requirements are applied to the track, others to the muon depending on the nature of the cut. This is made explicit for each cut in the bullet list below. The detailed preselection cut flow with efficiency on signal, background MC and data is given on table 6.6 after trigger selection. The efficiency is defined as the number of events with at least one track passing the following criteria over the total number of events.

Let us state for each criteria the reason it is selected on:

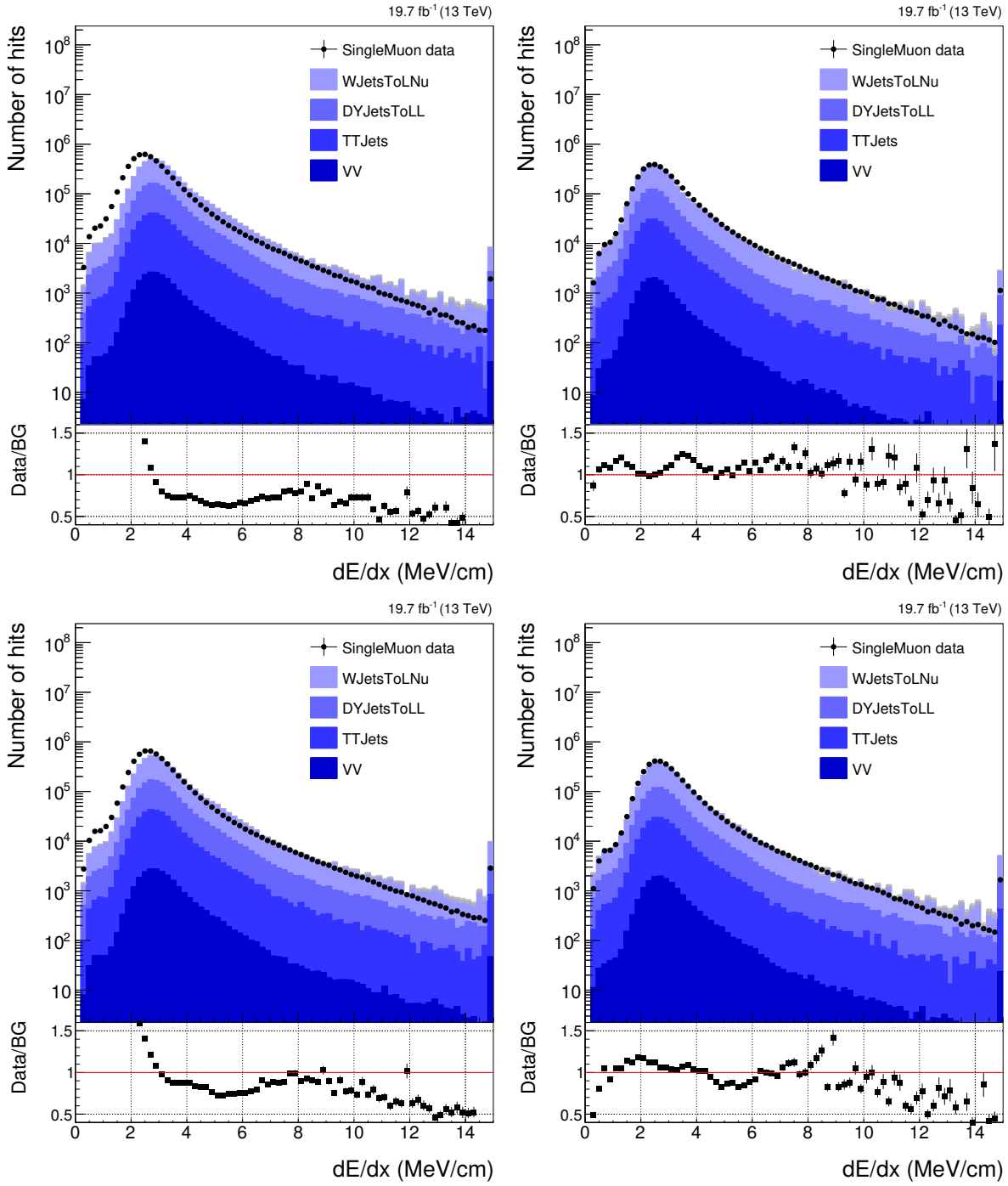


Figure 6.16: dE/dx spectrum before (left) and after (right) corrections in the PXB layer 2 (top) and layer 3 (bottom) in 2016 data (eras B to F, 19.7 fb^{-1}).

- $p_T > 55$ GeV on the tracker track: as explained in the trigger section, we choose to select high p_T tracks to avoid being drowned in the muon spectrum from W and Z decays. At this p_T value, the plateau of the trigger for a Q = 1 low-mass FCP is already reached.
- $|\eta| < 1.5$ on the tracker track. Signal is more central than SM BG.
- $|d_{xy}| < 0.1$ cm and $|d_z| < 0.5$ cm on the tracker track: it imposes the candidate track to point to the primary vertex. We expect the FCP to be produced promptly at the primary vertex. Moreover, this requirement cleans the sample from cosmic muons that can go through the detector at any impact parameter.
- $N_{\text{tracker hits}} > 5$ on the tracker track: the signal-to-background discrimination will involve

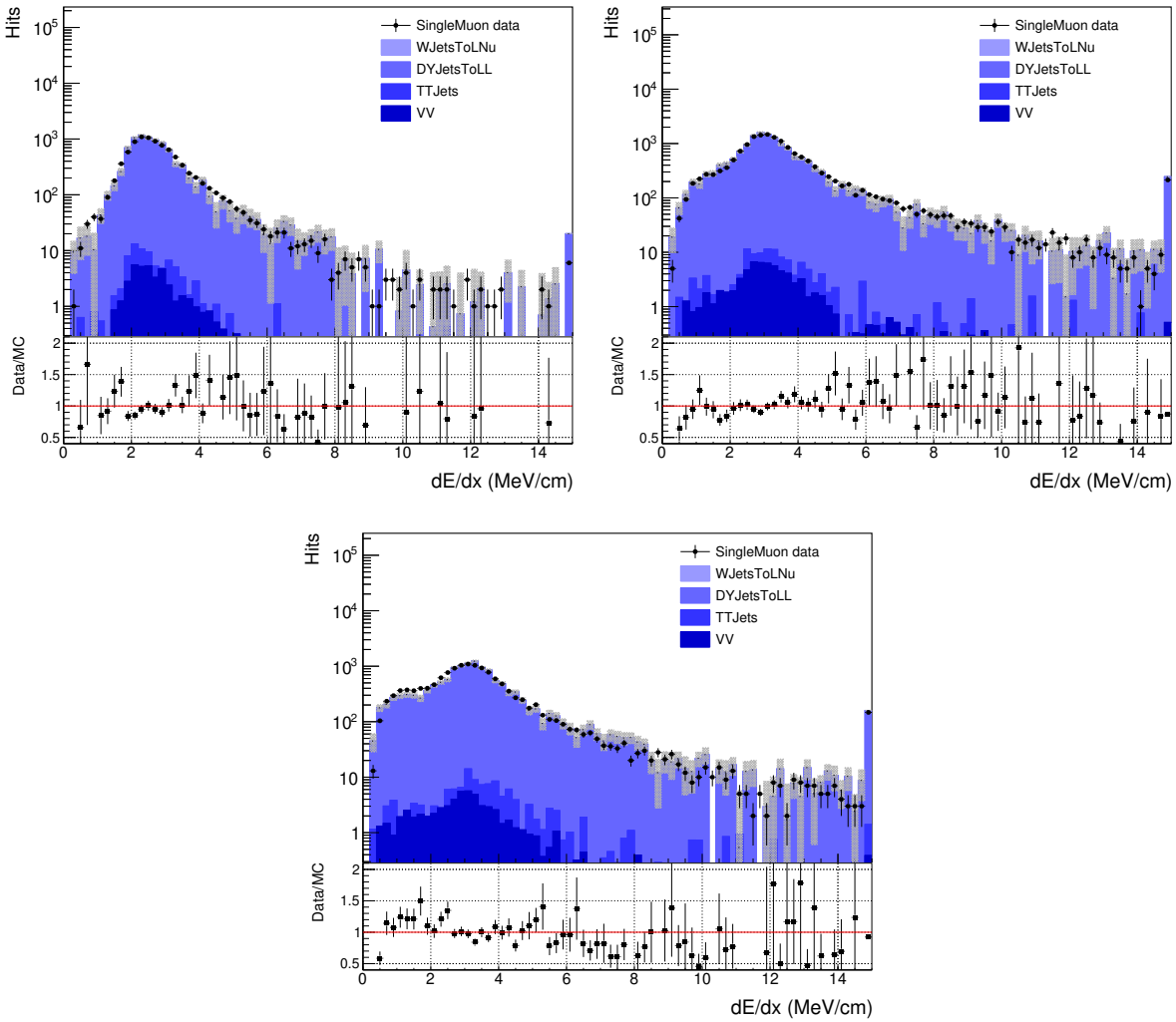


Figure 6.17: dE/dx spectrum after corrections in the PXBL2 (top-left), TIBL1 (top-right) and TOBL1 (bottom). This is 2016 data (eras B to F, 19.7 fb^{-1}) in the control region.

cutting on the number of hits. We need a minimal number of hits per track for the procedure to be relevant.

- $N_{\text{pixel hits}} > 1$ on the tracker track: cosmic muons tend to have a low number of pixel hits due to their out-of-time arrival in the detector. This helps reducing the rate.
- isLoose on the matched muon: ensures that the muon is a PF candidate which allows to apply PF isolation on it.
- PF isolation < 0.15 on the matched muon: signal tracks are expected to be isolated from other activity.
- isGlobalMuon on the matched muon: allows for the use of the timing variable.
- time$_{IP} > 0$ ns on the matched muon: we select positive values, i.e. we favour massive particles. To make sure the time variable is valid, we also impose the corresponding number of degrees of freedom larger than 7.
- $\alpha_{max} < 2.8$ on the tracker track: muons from cosmic rays go through the detector in straight lines, i.e. with a α_{max} close to π.
- Hits near edge on the tracker track: ignoring hits near the edge of a module (see subsection 6.8.2).

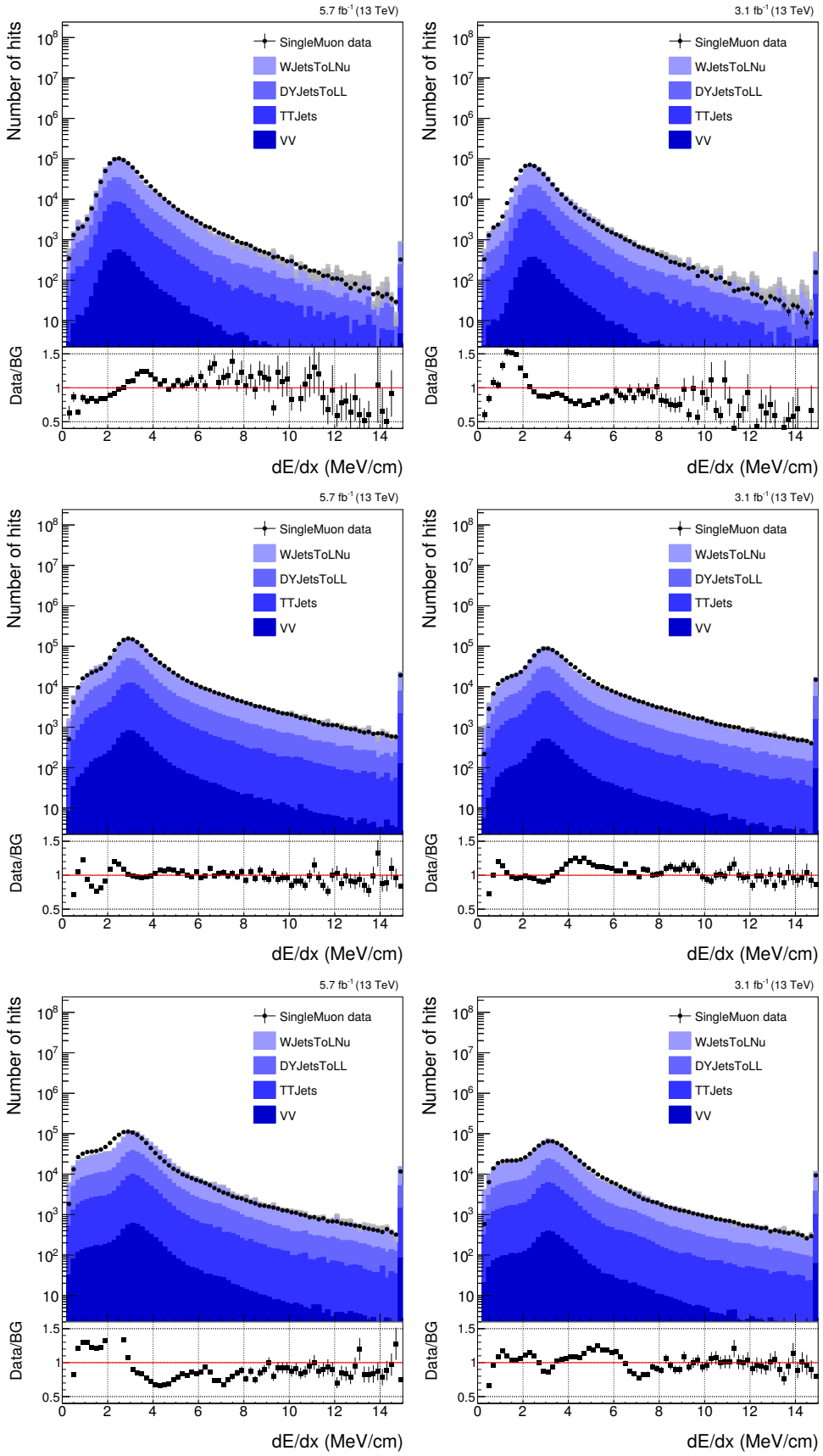


Figure 6.18: dE/dx spectrum after corrections in the PXBL2 (top), TIBL1 (middle) and TOBL1 (bottom) in eras B (left) and F (right) of 2016 data.

- $M_{\mu\mu} < 80$ GeV || $M_{\mu\mu} > 100$ GeV on matched muon pairs: split of the dataset in a search and a control region, the latter being used for sanity checks and background estimation.

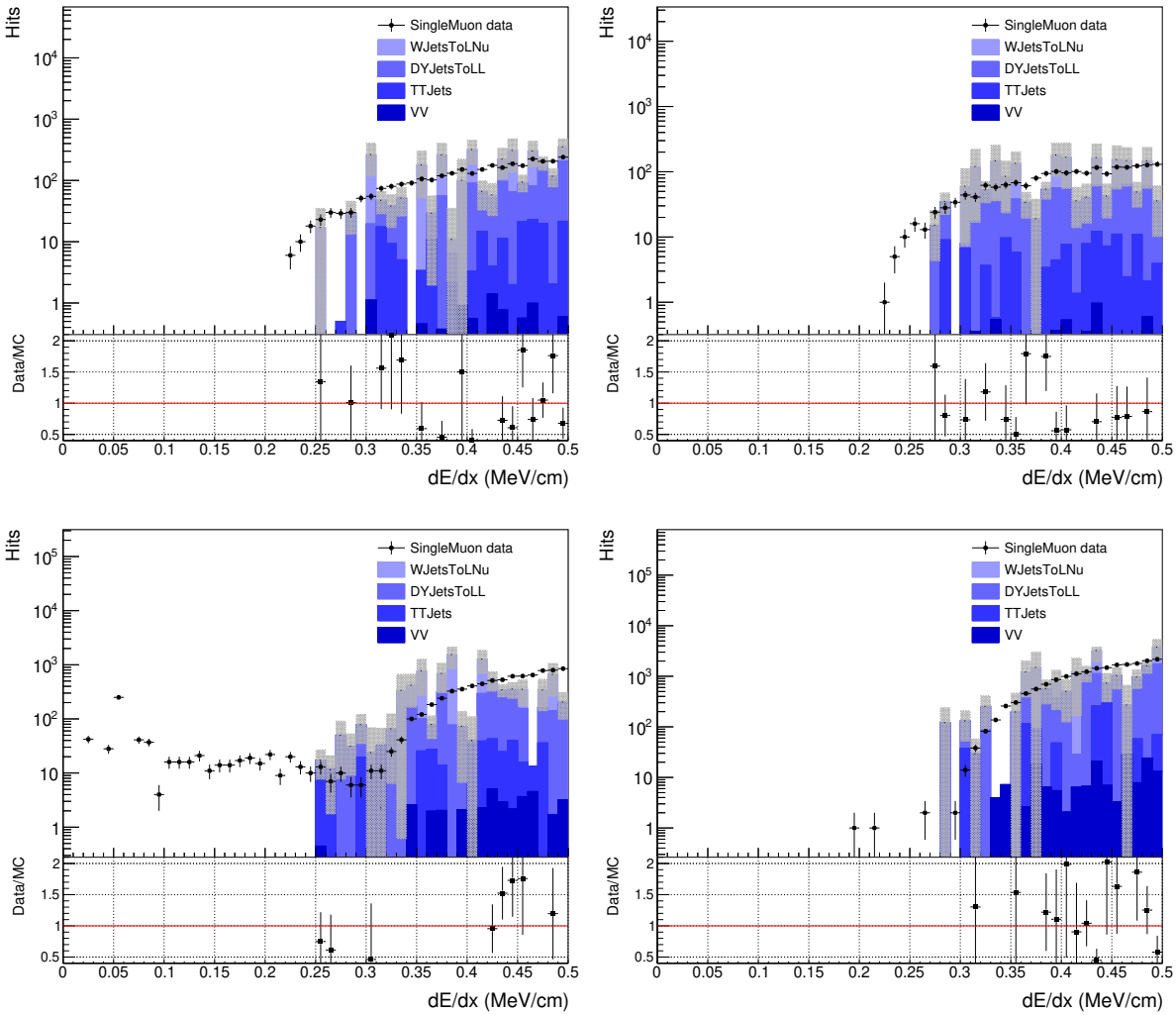


Figure 6.19: Zoom of the dE/dx spectrum in at low values in the PXBL2 (top-left), PXBL3 (top-right), TIB (bottom-left) and TOB (bottom-right) in 2016 data (eras B to F, 19.7 fb^{-1}). Some TIB hits accumulate in the very low dE/dx region.

The control region is made of 2-tracks events around the Z boson mass peak and the search region is the orthogonal dataset, i.e. all 1-track events (for which $M_{\mu\mu}$ is not even defined) and 2-track events off the Z mass peak. Events containing 3 candidate tracks are rare and are rejected (see figure 6.20).

The last requirement in table 6.6 is made to split our dataset into a search and a control region. By construction, it is not met if the event only contains one candidate track. The search region is where we will look for our signal while the control region will be used to check the validity of the method and estimate the number of background events. The control region is made of dimuon events around the Z peak, i.e. a sample of "good" SM muons, the search region being composed of all other events. The two regions are orthogonal by construction. The very high efficiency to select events in the search region seems to indicate that no events are selected in the control region but it is simply that the control region is much smaller than the search region. The ratio of the two is of the order of 0.1 %. This is mainly due to the requirement of having two tracks with a p_T larger than 55 GeV. With this cut, we preferentially select boosted Z bosons hence ending up with a lower cross-section. On top of that, the W process has a larger cross-section with respect to that of Drell-Yan and it is more unlikely to have two good tracks in an event than a single one.
The fact that the control region is much less populated than the search region may lead to some issues when estimating the background. We will discuss this in section 6.11.

Table 6.6: Event preselection cut flow for 2016 data (35.9 fb^{-1}). Only events with at least one track fulfilling the following requirements are kept.

Cut	Signal (M_{L_Q}= 100 GeV) Q=1	Q=2/3	Q=1/3	W($l\nu$)	DY(ll)	Data
HLT_Mu50	18,304	9,062	66	30,317,146	10,430,523	86,811,235
$p_T > 55$ GeV	17,609	8,739	21	20,625,354	6,910,926	60,770,305
$\|\eta\| < 1.5$	11,727	5,678	20	13,851,789	4,704,856	45,677,445
$\|d_{xy}\| < 0.1$ cm	11,723	5,677	20	13,844,154	4,702,002	45,558,210
$\|d_z\| < 0.5$ cm	11,723	5,677	20	13,844,154	4,702,002	45,558,210
$N_{\text{tracker hits}} > 5$	11,710	5,670	19	13,794,153	4,687,442	44,744,559
$N_{\text{pixel hits}} > 1$	11,664	5,627	13	13,663,611	4,645,611	43,891,460
is matched to loose ID muon	11,621	5,599	12	13,549,365	4,618,178	40,808,672
PF iso. < 0.15	11,582	5,578	12	13,287,514	4,546,287	22,180,069
matched muon is global	11,551	5,562	12	13,260,079	4,530,939	22,066,573
time$_{IP} > 0$ ns	11,094	5,347	12	6,102,867	2,144,996	10,847,648
$\alpha_{max} < 2.8$	10,273	4,974	11	6,037,451	2,098,381	10,561,898
Hits near edge	7,299	3,624	8	3,689,246	1,298,085	6,612,732
$M_{\mu\mu} < 80$ GeV || $M_{\mu\mu} > 100$ GeV	7,292	3,619	8	3,689,246	1,289,286	6,603,283
Total	7,292	3,619	8	4,978,532		6,603,283
Total efficiency (%)	39.8	39.9	12.1	12.2	12.4	7.6

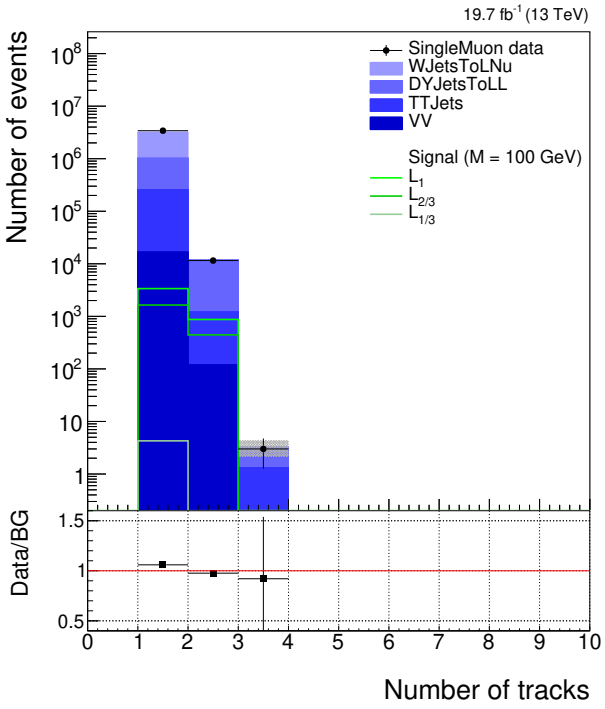


Figure 6.20: Number of candidate tracks before splitting of the dataset into search and control region (2016 data eras B to F, 19.7 fb^{-1}). Very few events contain 3 tracks and they are rejected from the selection.

Figure 6.21 shows the histogram for each variable used to define a candidate track in early 2016 data (eras B to F, 19.7 fb^{-1}). All but the cut on the considered variable are applied on its histogram ("N-1" plots). The cut value is displayed with a black line. Monte Carlo backgrounds are normalized to their cross-section and to the luminosity, whith pile-up reweighting and centrally produced trigger and muon identification scale factors applied [138, 139, 140]. Most cuts reject a similar fraction of data and MC events with the notable exception of the cut on the isolation. Reconstructed objects are less easy to isolate in data due to the noise and the pile-up interactions which explain the observed discrepancy. The estimated time at which the particle was produced time$_{IP} > 0$ shows a very different behaviour for data and background MC

on one hand, and signal MC on the other hand. This is explained by the larger mass of the FCP (here shown for $M_{L_Q} = 100$ GeV), leading to slower particles. The hypothesis on their velocity to be the speed of light therefore does not hold and those particles seem to have been created later.

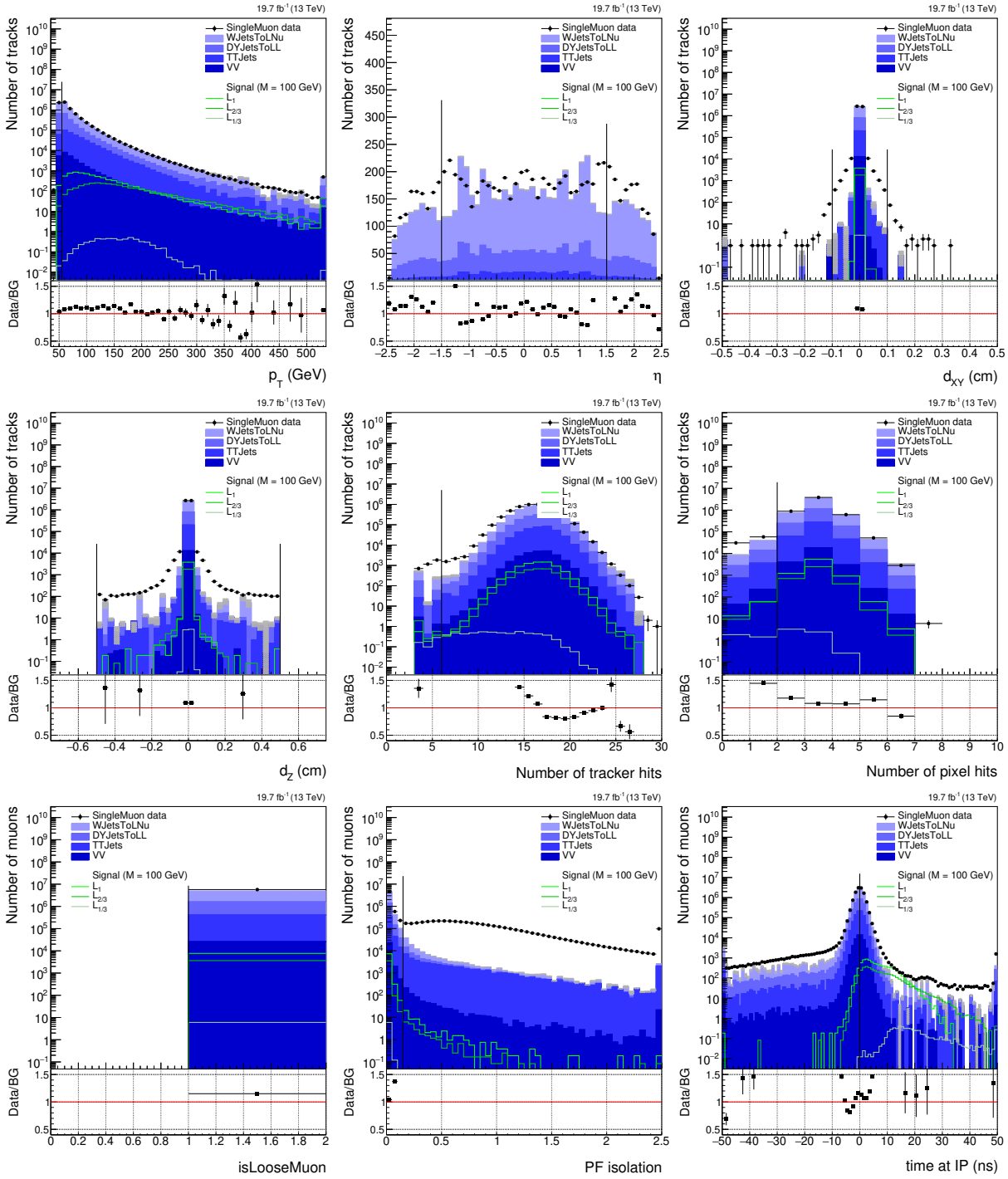


Figure 6.21: Candidate track selection variables comparing data, background MC and signal MC in 2016 data (eras B to F, 19.7 fb^{-1}). All but the cut on the considered variable are applied on its histogram ("N-1" plots).

On top of this series of selection cuts, we apply an extra cut on signal events only to account for an inefficiency at HLT that is not taken care of in simulations. This HLT inefficiency comes from the fact that the dE/dx spectrum is not well simulated: it is softer in data. As a consequence, the number of tracker hits above the zero-suppression threshold is overestimated in signal events leading to a similarly overestimated number of reconstructed tracks. Although we already correct

our signal for this effect "offline", it must be taken into account at trigger level. To do that, we count the number of hits with a dE/dx lower than the threshold defined in section 6.8.4 *after* correction as defined in section 6.8.3. If the total number of hits above threshold has dropped below 6, we remove this track from the collection of candidate tracks. On top of that, we also reject the track if it has two consecutive hits with a dE/dx lower than the threshold. These two extra requirements simulate the scenario in which a track would have too many low dE/dx hits for the reconstruction to work and would therefore not trigger the single muon path. The impact of this cut is given at the end of the next section as it needs the definition of the variable used to look for a signal.

6.10 Signal selection

After selection of candidate tracks, one must come up with an efficient way of distinguishing muon-like tracks (background) from signal-like tracks. We use a single variable to discriminate signal from background: the number of low dE/dx tracker hits per track $N_{\text{hits}}^{\text{dE/dx<2}}$, where "low" means below the 2 MeV/cm threshold. While not indicated on the subscript, a lower bound is also set at dE/dx= 0.25 MeV/cm to remove noisy hits with very low ionization energy as described in section 6.8.4. The upper bound is chosen by comparing the dE/dx spectrum in data and for our signal as was shown in figure 6.8. The track of a fractionally charged particle L_{Q} is expected to contain a large number of such hits as its ionization power is lower.

Figure 6.22 shows $N_{\text{hits}}^{\text{dE/dx<2}}$ for 2016 data (eras B to F, 19.7 fb^{-1}) both in the search and the control region (as defined in section 6.9) next to the dE/dx spectrum for the hits of the corresponding tracks. As we have not yet defined the background estimation method, we blind the $N_{\text{hits}}^{\text{dE/dx<2}}$ distributions at 99 % for the moment. Once well defined, the background estimation procedure will be applied to the full dataset. The control region is scaled by the ratio of the number of tracks in the signal region to the number of tracks in the control region. We show the distributions before and after mitigation of instrumental background. The effect of the mitigation is to ignore inner single hits (for radiation damage) or the track itself (if it contains edge hits) which directly translates into a softer $N_{\text{hits}}^{\text{dE/dx<2}}$ distribution. This is crucial for our background estimation method as it brings the $N_{\text{hits}}^{\text{dE/dx<2}}$ distribution closer to its binomial genuine behaviour. This aspect is developed in section 6.11.

Each data-taking period is different with respect to radiation damage and other effects that lower the collected charge in the tracker hits (e.g. the APV saturation issue in early 2016 data and the pixel DC/DC issue in 2017 data). In particular, those effects increase the proportion of hits with a stopping power below threshold which directly translates into an enhanced background from instrumental sources in the considered subdataset. To properly take these different behaviours into account, the choice of the cut on $N_{\text{hits}}^{\text{dE/dx<2}}$ must be time-dependent. Tables 6.7 to 6.9 give the proportion of low dE/dx hits for each era of the three data-taking years from which we analyze the data. We split the full Run 2 dataset into four periods according to the proportion of low dE/dx hits, i.e. the expected background: early 2016 (B-F), late 2016 (G-H), 2017 and 2018. The large value of p in early 2016 data is explained by what is referred to as "the APV saturation effect". This can be seen on table 6.7 by comparing the numbers from eras B-F with the two last eras G-H where this problem was mitigated. In 2017, the proportion is smaller at first due to the installation of the new damage-free pixel detector. However, an issue with the DC/DC converters in the pixels kept increasing the proportion of defective modules along time. From era E onwards, a mitigation procedure stopped this increase. Data collected in 2018 show a stable proportion though larger than in 2017.

Adding signal to the bottom right plot of figure 6.22 and rescaling, one can observe in figure 6.23 the difference in behaviour between data and low charge signal in the dE/dx spectrum for four different data periods. The corresponding $N_{\text{hits}}^{\text{dE/dx<2}}$ distributions are given in figure 6.23. As expected, signal tracks tend to have a large $N_{\text{hits}}^{\text{dE/dx<2}}$ while SM muons have very few of them. If FCP were to exist, they would appear as an excess in data in the high $N_{\text{hits}}^{\text{dE/dx<2}}$ region:

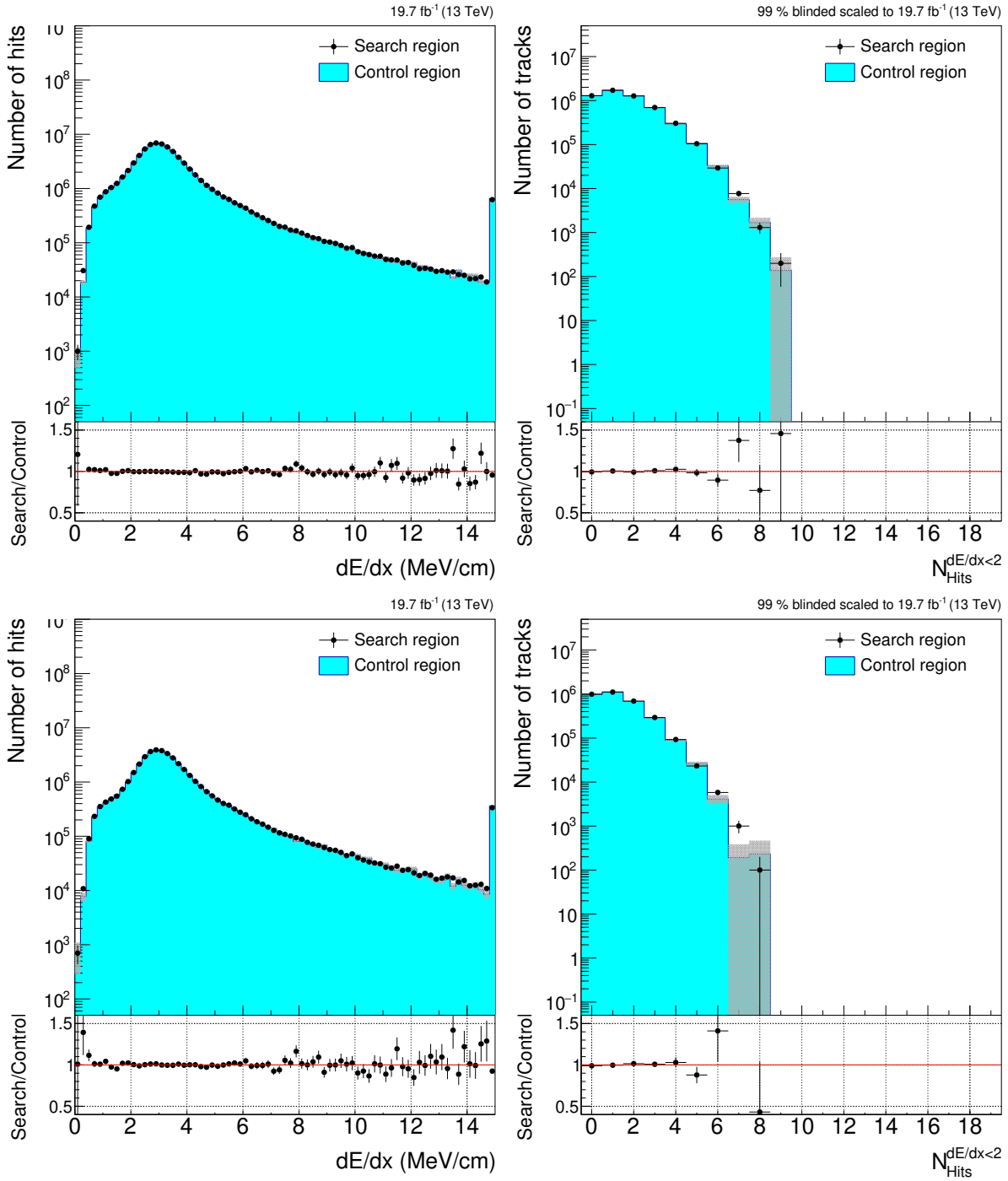


Figure 6.22: Hits dE/dx spectrum (left) and low dE/dx hits multiplicity per track (right) in the search and the control regions of 2016 data (eras B to F, 19.7 fb^{-1}). Top row shows the distributions before cleaning from instrumental background, bottom row gives the same distributions after cleaning.

this is our signal extraction variable. We will not simply do a cut-and-count analysis but use the information from several consecutive bins in the $N_{\text{hits}}^{\text{dE/dx<2}}$ spectrum in a so-called "shape" analysis. Let us point out that we do not make use of the full $N_{\text{hits}}^{\text{dE/dx<2}}$ distribution but only consider the last few bins at high values of $N_{\text{hits}}^{\text{dE/dx<2}}$. The choice of the bins to be included in the limit setting procedure is determined by the ratio of background to signal events: we only consider bins in which signal is not overwhelmed by background and so in which we are sensitive. In early 2016, we will consider bins $N_{\text{hits}}^{\text{dE/dx<2}} = 8$, 9, 10, 11 and ≥ 12. In late 2016 and 2017 data behave similarly and the binning is chosen to be 4, 5, 6 and ≥ 7. Finally in 2018, we choose

Table 6.7: Proportion p of low dE/dx hits per track and mean number μ of hits per track for each era of 2016 data.

	B	C	D	E	F	G	H
Lumi (fb^{-1})	5.7	2.6	4.2	4.0	3.1	7.6	8.7
p (%)	9.1	9.2	9.2	10.2	10.3	3.2	2.8
μ	14.3	14.1	14.1	13.9	14.0	14.7	14.6

Table 6.8: Proportion p of low dE/dx hits per track and mean number μ of hits per track for each era of 2017 data.

	B	C	D	E	F
Lumi (fb^{-1})	4.8	9.6	4.2	9.3	13.5
p (%)	1.4	2.3	3.2	1.5	2.1
μ	15.2	15.3	15.6	15.4	15.2

to consider bins 5, 6, 7 and ≥ 8. The last bin always contains less than 1 expected background event and is inclusive, i.e. it is integrated up to infinity.

In figure 6.25, we compare the hits dE/dx spectrum and the corresponding $N_{\text{hits}}^{\text{dE/dx<2}}$ for signals with masses 100 and 1000 GeV superimposed on 2016 data eras B to F. For a given charge, a M = 1 TeV particle is more ionizing than a 100 GeV particle (see Bethe) which is why the dE/dx spectrum for M = 1 TeV is broader and extends to larger values. Consequently, the $N_{\text{hits}}^{\text{dE/dx<2}}$ distributions are "softer", i.e. they show a lower mean value. In the scenario of a "high charge" (Q $\geq$ 1/2) FCP, this leads to a lower sensitivity since it brings the signal extraction variable closer to a background-like behaviour. However, for low charge particles the picture is different (as we will detail when presenting the results). For low charge FCP, the tracking efficiency is low due to the low ionization and most tracks are lost. With higher mass, this efficiency improves while keeping the ionization lower than that of Q = 1 particles. Therefore in this case the sensitivity improves with the mass.

Now that the $N_{\text{hits}}^{\text{dE/dx<2}}$ variable has been properly defined, we can evaluate the impact of the corrections we derived in section 6.8.3 on the simulations. In figure 6.26, the $N_{\text{hits}}^{\text{dE/dx<2}}$ spectrum in a subset of 2016 data is compared to that in DY simulations before and after the corrections. Although the agreement is not perfect after corrections, it is clear that they have a huge impact on the distributions, bringing the simulations much closer to the spectrum in data.

We can also come back to study the impact of the HLT inefficiency cut. As expected, the impact of this cut is very small on "high charge" signal, with an efficiency of 0.006 % for Q = 1 and 0.003 % on Q = 2/3 FCPs. On the other hand, the rejection efficiency reaches 5 % for Q = 1/3 FCPs. However, the rejected tracks are mostly found in the low $N_{\text{hits}}^{\text{dE/dx<2}}$ region where signal is anyway drown in background. The reason for this is that for a track to be rejected by this cut, it must have a low number of hits to begin with (around 6 or 7), hence an even lower $N_{\text{hits}}^{\text{dE/dx<2}}$.

6.11 Background estimation

The number of background events expected from pp collisions and cosmic muons must be properly evaluated in each considered bin of the $N_{\text{hits}}^{\text{dE/dx<2}}$ spectrum.

Table 6.9: Proportion p of low dE/dx hits per track and mean number μ of hits per track for each era of 2018 data.

	A	B	C	D
Lumi (fb^{-1})	13.5	6.8	6.6	31.1
p (%)	2.8	3.7	4.0	3.9
μ	15.4	15.3	15.4	15.4

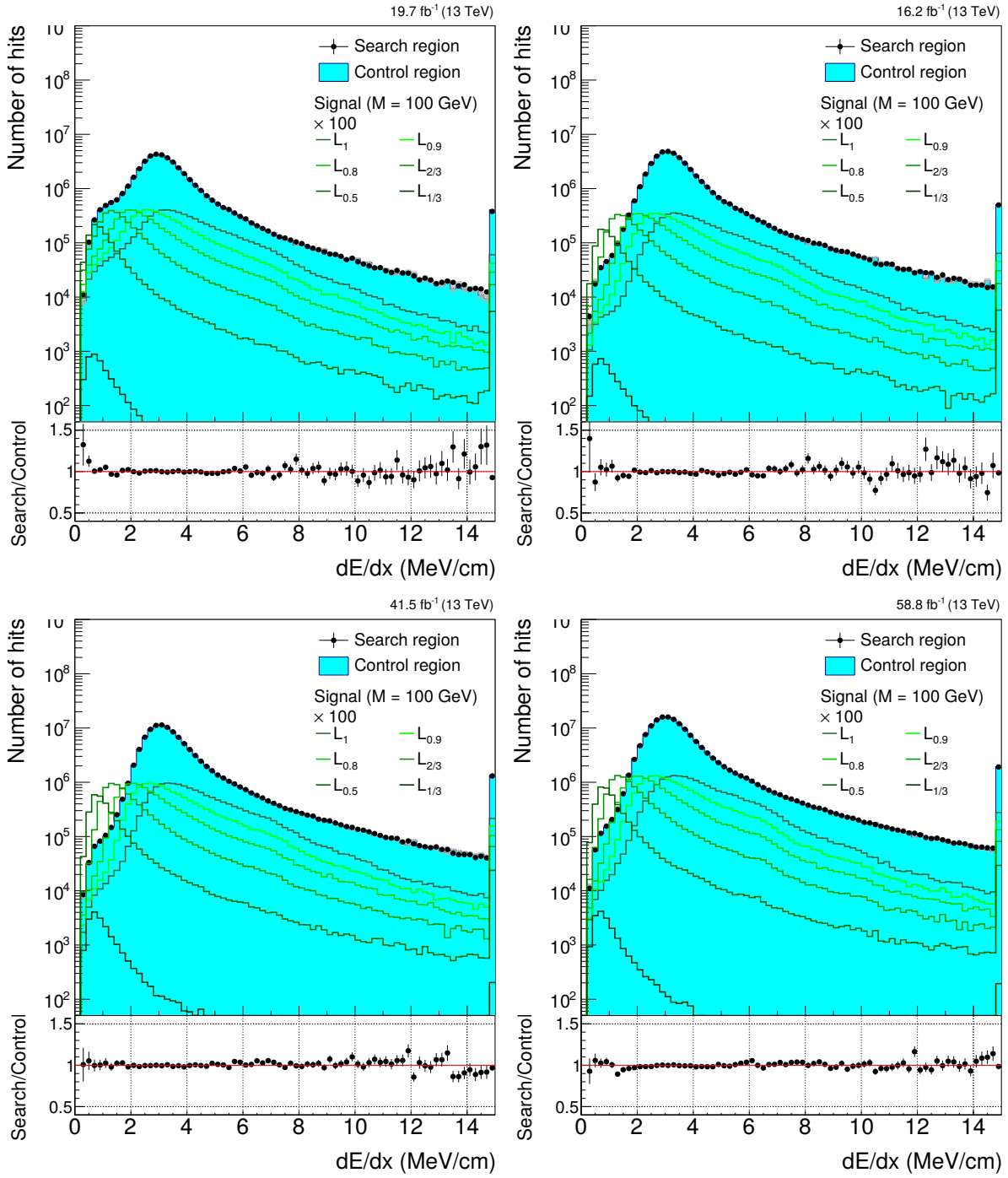


Figure 6.23: Hits dE/dx spectrum in the search and the control regions. On the top-left, 2016 data eras B to F (19.7 fb^{-1}), on the top-right 2016 data eras G and H (16.2 fb^{-1}), on the bottom-left 2017 data (41.5 fb^{-1}) and on the bottom-right 2018 data (58.8 fb^{-1}). Signal MC ($M_{L_Q} = 100$ GeV multiplied by a factor 100) is added for comparison.

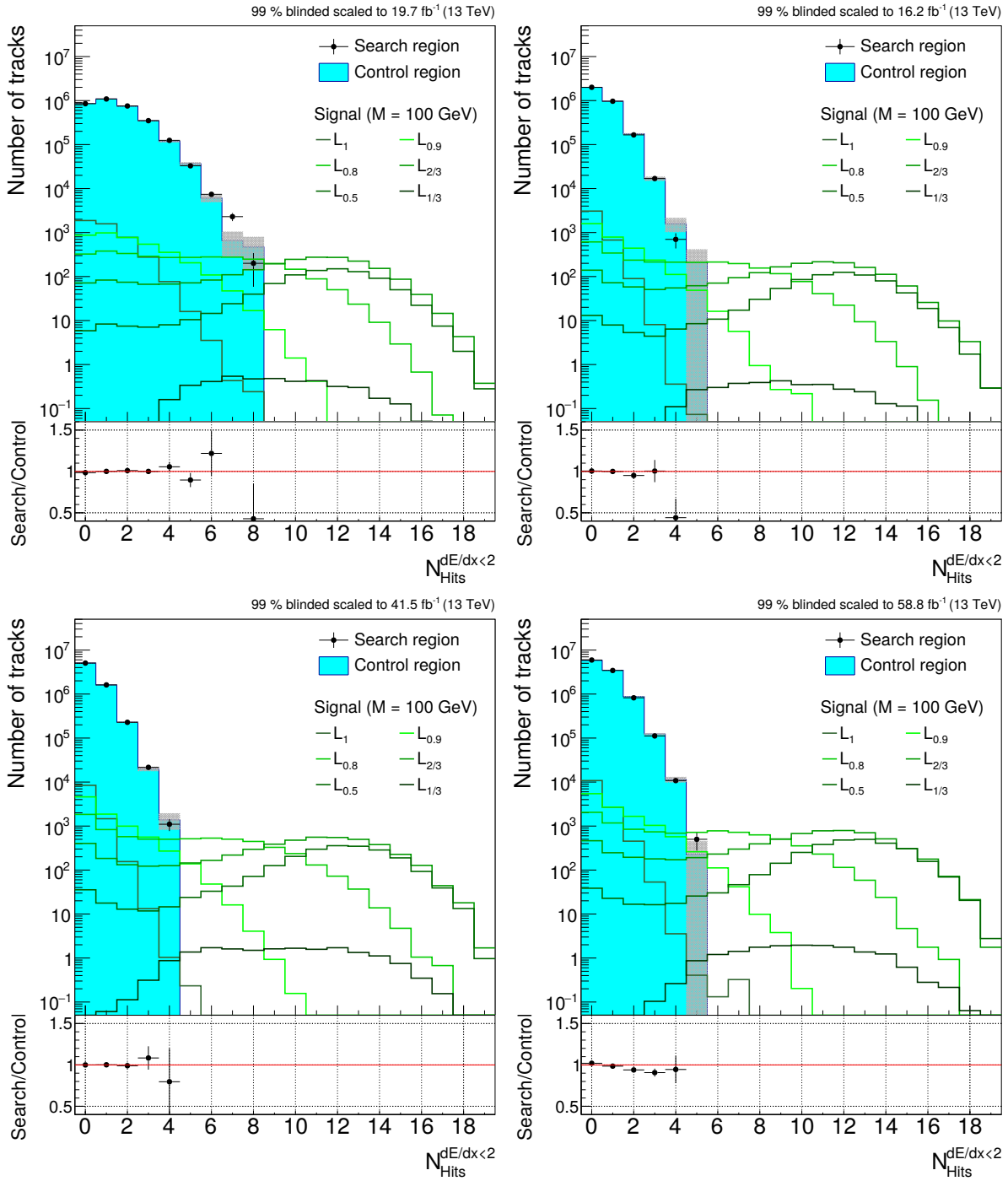


Figure 6.24: Low dE/dx hits multiplicity per track in the search and the control regions. On the top-left, 2016 data eras B to F (19.7 fb^{-1}), on the top-right 2016 data eras G and H (16.2 fb^{-1}), on the bottom-left 2017 data (41.5 fb^{-1}) and on the bottom-right 2018 data (58.8 fb^{-1}). Signal MC ($M_{L_Q} = 100$ GeV multiplied by a factor 100) is added for comparison.

6.11.1 Background from pp collisions

After removal of instrumental background, genuine physics events such as SM W($l\nu$) and Drell-Yan(ll) production still contribute to the overall background. As was shown in figure 6.13, the dE/dx spectrum is not well described by MC simulations. Therefore, we decide to rely mainly on data to estimate and predict the background. The background estimation method consists in using the control region at low $N_{\text{hits}}^{\text{dE/dx<2}}$ to extrapolate to the expected number of tracks at high $N_{\text{hits}}^{\text{dE/dx<2}}$ where the control region lacks statistics. This is justified by the good agreement between the search and the control regions both in the dE/dx spectrum and the low dE/dx hits

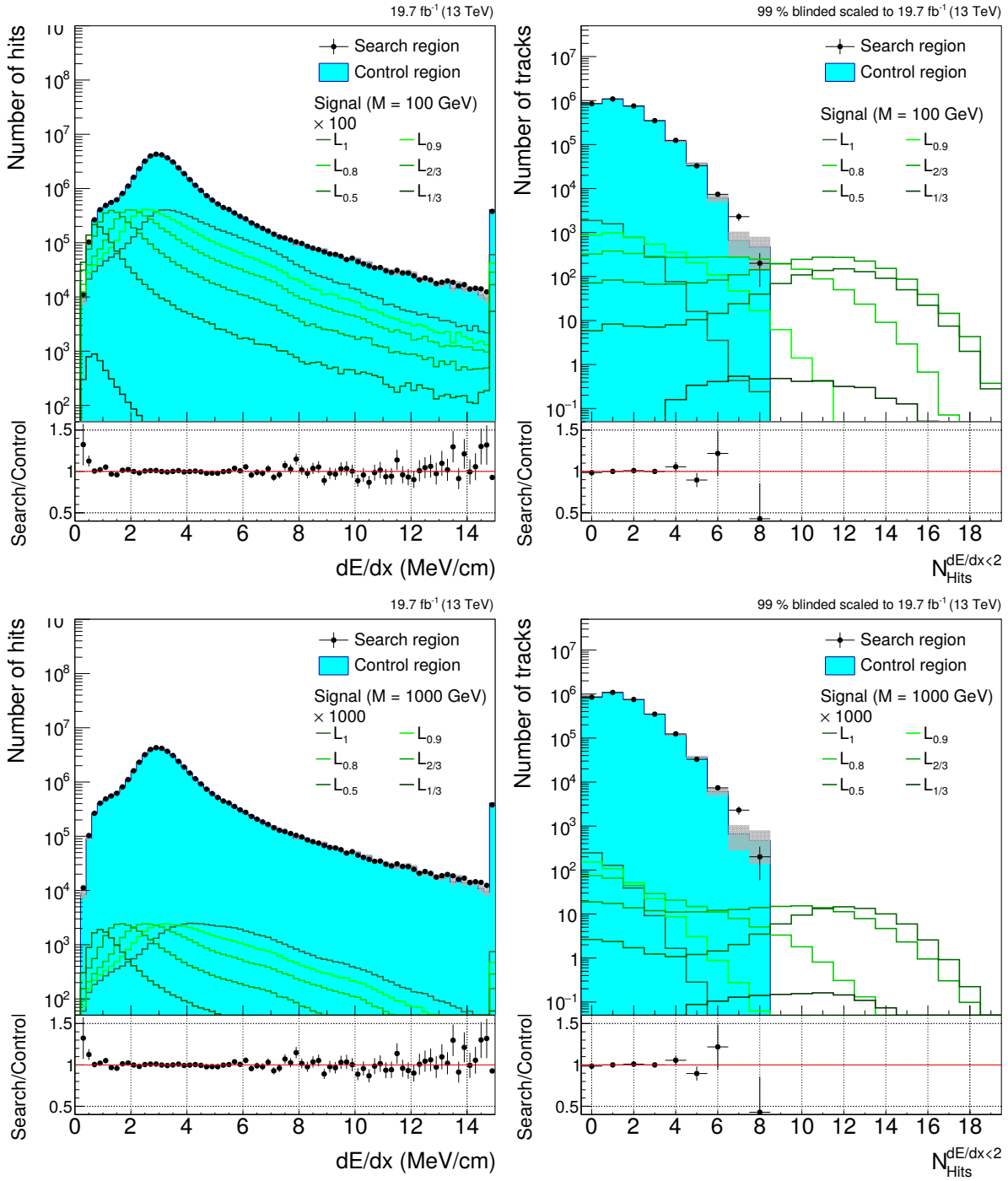


Figure 6.25: Hits dE/dx spectrum (left) and low dE/dx hits multiplicity per track (right) in the search and the control regions for $M_{L_Q} = 100$ (top) and 1000 (bottom) superimposed on 2016 data eras B to F (19.7 fb^{-1}).

multiplicity per track. The extrapolation is done by fitting the control region. Uncorrelated measurements of a given binary variable (hit dE/dx higher or lower than 2 MeV/cm) on identical detector modules should follow a binomial distribution. We would therefore naively use such a function. However, data move away from a strict binomial behaviour due to the fact that each layer is affected by radiation differently, biasing the hypothesis of identical modules. Following a suggestion from [132], we observe that another function that we call the "main" function, fits our data better. It reads:

$$N_{\text{events}}(n) = p_0 n^{p_1 + p_2 n} \tag{6.15}$$

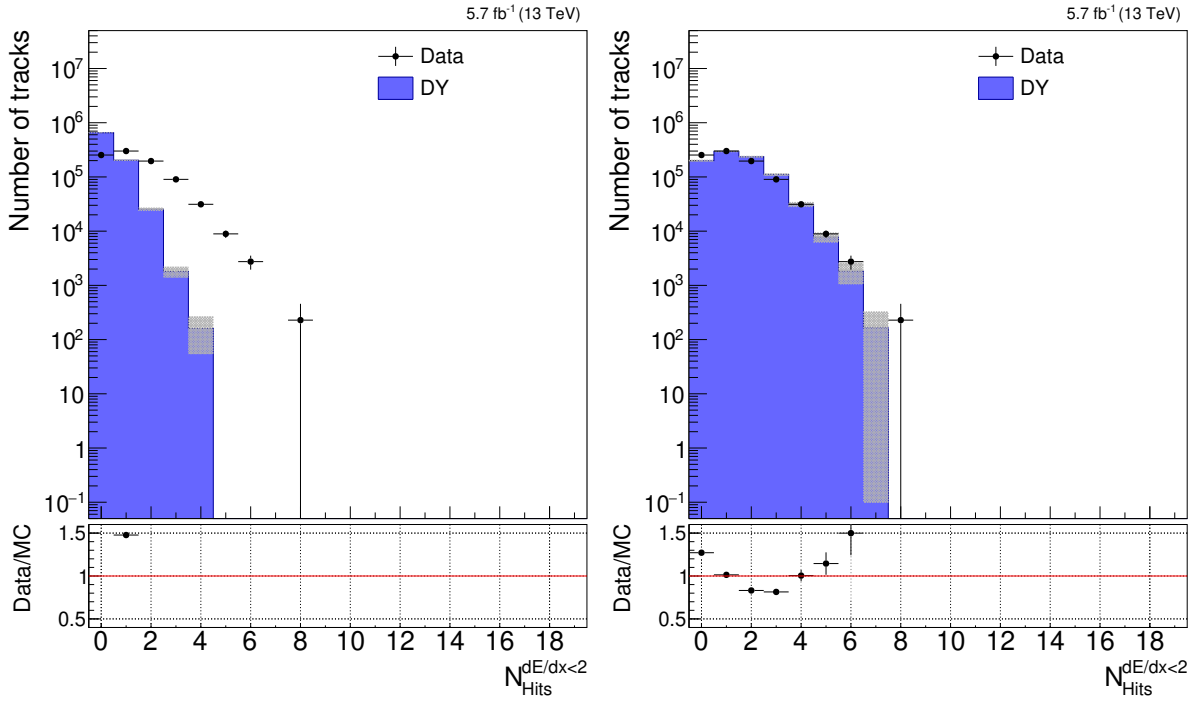


Figure 6.26: $N_{\text{hits}}^{\text{dE/dx<2}}$ in a subset of 2016 data compared with DY simulations before (left) and after (right) application of the corrections derived in section 6.8.3. We insist on the fact that the spectrum is not shifted by a global factor but that the changes are a propagation of the corrections applied on the dE/dx spectra for each layer.

We decide to use this function to fit our control region but to keep the binomial function as an estimate of the fit uncertainty. This means that from physical considerations, we still use as hypothesis that the binomial describes our data reasonably well. Let us then give the expression of the binomial distribution, describing the probability to observe n occurences of a binary experiment of probability p repeated μ times. It reads:

$$N_{\text{events}}(n) = N_0 \binom{\mu}{n} p^n (1-p)^{\mu-n}, \tag{6.16}$$

where

$$\binom{\mu}{n} = \frac{\Gamma(\mu+1)}{\Gamma(n+1)\Gamma(\mu-n+1)}. \tag{6.17}$$

As explained in section 6.10, the two parameters are the proportion p of low dE/dx hits per track and the mean number μ of hits per track. The binomial function then gives the expected number of tracks with a given number of low dE/dx measurements. As shown on tables 6.7 to 6.9, the μ parameter is stable through the years, being larger by one in 2017 and 2018 due to the extra pixel layer. This parameter then seems like a geometrical constant we should not be fitting. We choose to fix this parameter to its mean value and to fit the overall normalization and the p parameter. With no loss of generality, this removes a degeneracy we observed leading to unstable fits.

Figure 6.27 shows the result of the fit for the four subdatasets. The uncertainty on the fit-to-control ratio is split into two parts, one being statistically driven and the other being a systematic uncertainty. The statistical uncertainty is a quadratic combination of the statistical uncertainty on the fit and the statistical uncertainty from the limited number of events in the control region. It is shown by the black vertical lines on the dots. The systematic uncertainty comes from the deviation of the fit with respect to the fit to the binomial function and is displayed in grey bands. These two uncertainties are described in details in section 6.12.

For early 2016 data, the first bin is taken out of the fit as it brings the $N_{\text{hits}}^{\text{dE/dx<2}}$ spectrum too far away from a binomial behaviour. In other words, even after mitigation of instrumental background, the tracker layers in early 2016 data are too much affected by the APV saturation problem which correlates between individual measurements on a track. The difference in dE/dx spectrum between the inner (more damaged) and the outer (less damaged) layers is too large for the $N_{\text{hits}}^{\text{dE/dx<2}}$ to be described by a binomial function over its whole range.

The numerical outputs of the fits are given on table 6.10. The main conclusion one can come to when looking at this table is that the main fit has a much lower χ^2 than the binomial fit. This is another way to motivate the choice of this function as a central value for background estimation.

Table 6.10: Results of the binomial and main fits to the low dE/dx hit multiplicity per track.

		Early 2016	Late 2016	2017	2018
Main fit	p_1	2.84	0.80	-0.04	1.0
	p_2	-0.74	-1.22	-1.15	-1.12
	$\chi^2/\text{N}_{\text{d.o.f.}}$	2.1	0.39	0.04	2.68
Binomial fit	μ (fixed)	14.1	14.7	15.3	15.4
	p (%)	11.4	2.64	2.01	3.25
	$\chi^2/\text{N}_{\text{d.o.f.}}$	1.38	0.64	0.59	6.0

Once the $N_{\text{hits}}^{\text{dE/dx<2}}$ spectrum is fitted for the considered data period, the number of expected candidate tracks from SM processes in pp collisions is directly given by the fit at the considered $N_{\text{hits}}^{\text{dE/dx<2}}$ value.

In figure 6.28, an event display of the highest $N_{\text{hits}}^{\text{dE/dx<2}}$ (= 7) track is shown. It was recorded in 2018.

Now we would like to discuss the question we introduced when we showed that the control region is much smaller than the search region. With a control region much less populated than the signal region, a small source of background may go unnoticed that the fit would therefore not catch. This effect could then show up in the search region as an excess in data. Tracks hitting several module edges is a typical example of such an effect. It yields a signal-like contribution to the $N_{\text{hits}}^{\text{dE/dx<2}}$ spectrum that is too weak to enter in the control region but that would show up as a tail at high $N_{\text{hits}}^{\text{dE/dx<2}}$ in the search region. As we estimate our background from a fit to the control region, this would then look as an excess in data. However, most sources of low dE/dx hits (such as dead strips or the APV saturation issue) should not yield such a tail in the $N_{\text{hits}}^{\text{dE/dx<2}}$ distribution as they do not involve a correlation between hits of the same track. They impact the whole distribution in its bulk which we then fit. On top of that, we believe there is a certain "safety" to this analysis as signal and BG show up really differently in the $N_{\text{hits}}^{\text{dE/dx<2}}$ spectrum. BG shows up as a tail, while signal would look like a whole new bulk at higher $N_{\text{hits}}^{\text{dE/dx<2}}$ values.

6.11.2 Sanity check on background simulations

In order to build up confidence in our background estimation method, we apply the same procedure to a Drell-Yan Monte Carlo sample. We compare the two fit functions on the control and the search region in two scenarios: uncorrected ans corrected simulations. "Corrected" means that the dE/dx spectrum is corrected as prescribed in section 6.8.3 to match that in data. The corrections are exactly the same as those applied to the signal simulations. The results are shown in figure 6.29 and the corresponding numbers are given in table 6.11.

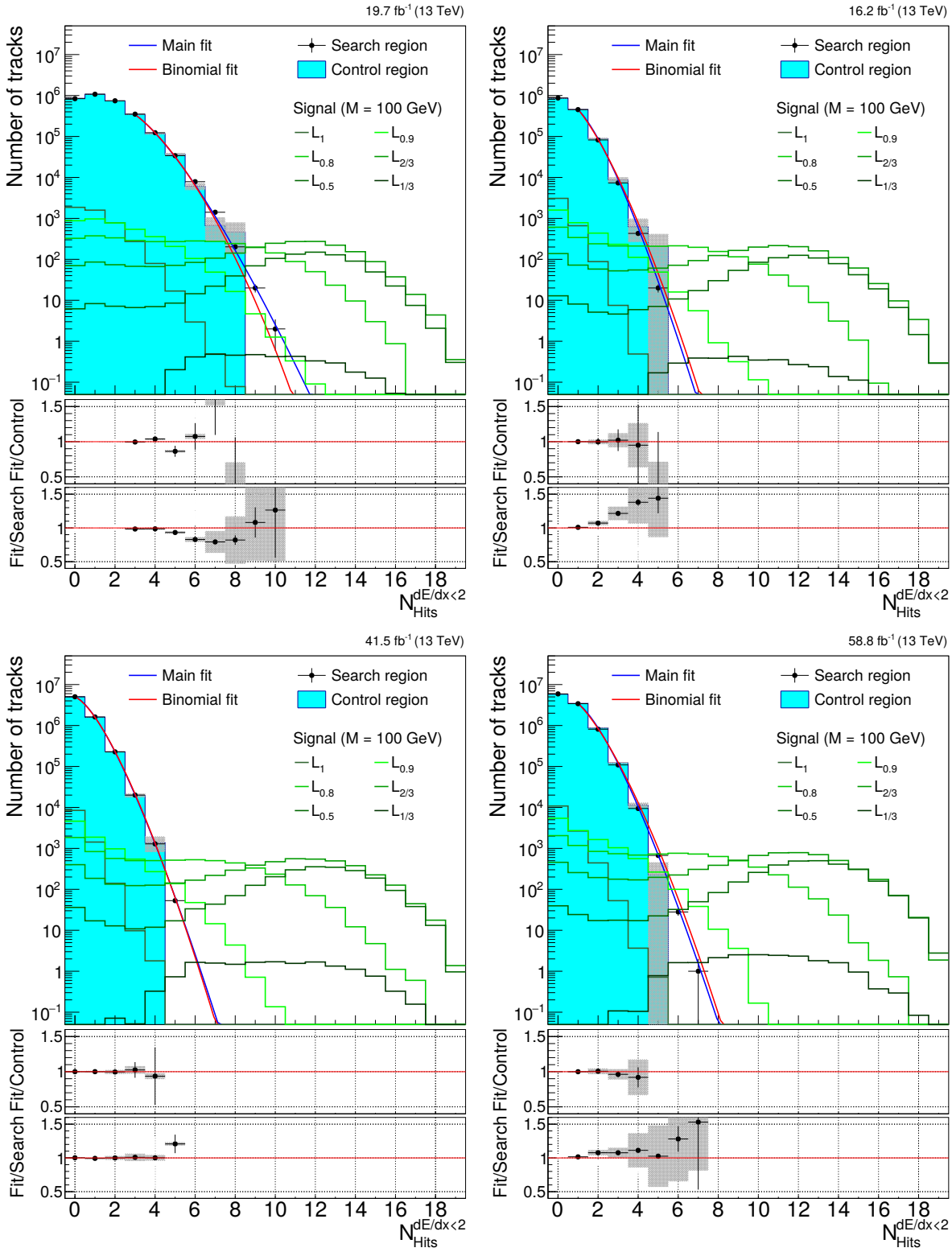


Figure 6.27: $N_{\text{hits}}^{\text{dE/dx<2}}$ spectrum and fit (red curve) of the control region in early 2016 (top-left), late 2016 (top-right), 2017 (bottom-left) and 2018 (bottom-right). On the ratio plot, the uncertainty is split into statistical (black vertical lines on the dots) and systematic (grey bands) sources. The statistical sources are made of the number of events in the control region and the uncertainty on the parameters of the main fit. The systematic uncertainty comes from the deviation from the binomial distribution.

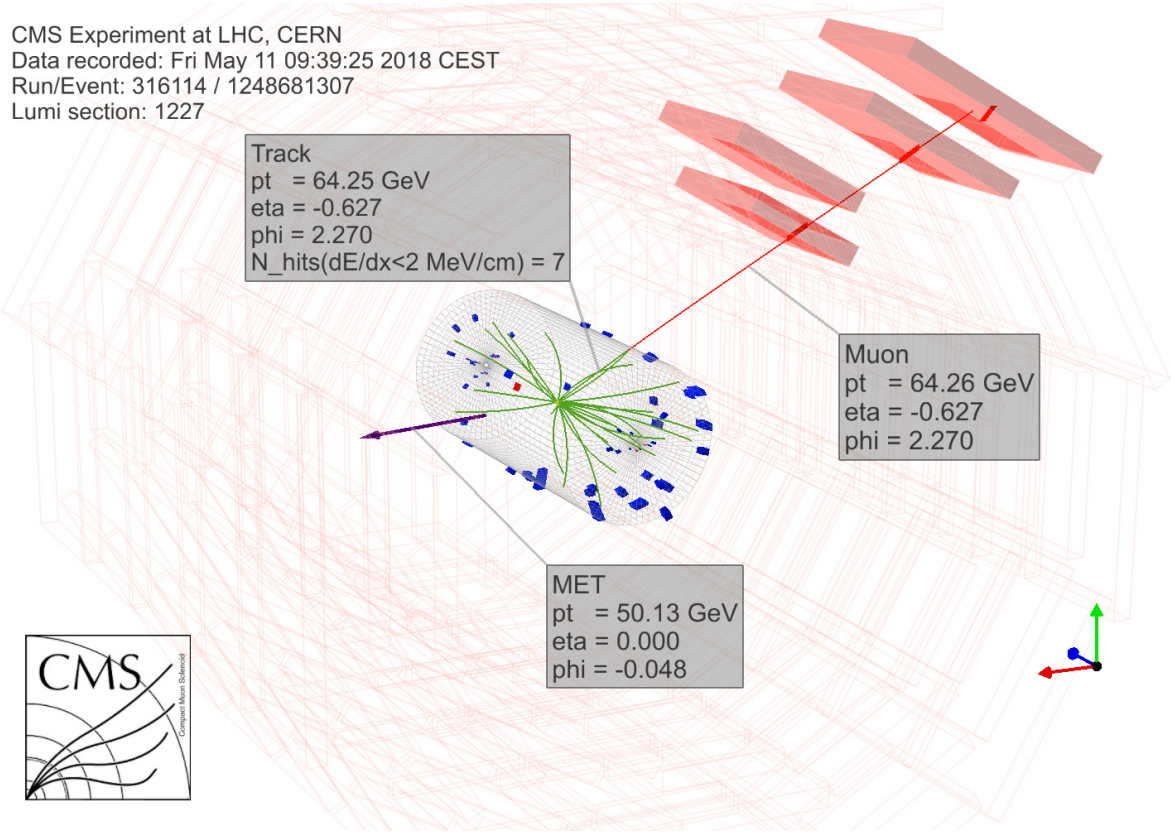


Figure 6.28: Event display of the highest $N_{\text{hits}}^{\text{dE/dx<2}}$ ($= 7$) track event, recorded in 2018. The kinematics and topology of the event makes it look like a W($\mu\nu$) event or a Z($\mu\mu$) event where one muon would have been missed. What really qualifies it as a good candidate in the search for low charge particles is the large number of $N_{\text{hits}}^{\text{dE/dx<2}}$ on the inner track.

6.11.3 Background from cosmic rays

Since muons from cosmic rays are not time-synchronized on the collisions, their tracks may satisfy the candidate track selection criteria much more easily and stand out in the signal region if it is slightly out of time and thus inducing correlated low dE/dx hits. This expectation of low dE/dx hits from cosmic rays is why several specific cosmic-rejection cuts have been added to the selection.

In the previous incarnation of this analysis [132], the background from cosmic rays was estimated from a $0.1 < |d_{\text{xy}}| < 1.1$ cm and $\theta_{\text{max}} > 2.8$ sideband (see section 5.1 in [132]). An estimate for the signal region of 0.0048 ± 0.0024 events was obtained. Another sideband estimate using $0.5 < |d_{\text{z}}| < 50$ cm instead of the θ_{max} cut yields a compatible 0.0065 ± 0.0046 events. This background estimate was on par with the estimated background from pp collisions.

A difference with the previous analysis though, is that now we integrate a significantly larger dataset over the same period of time, while the cosmic background is expected to scale with the length of the data taking, not with the luminosity. Assuming equal running time for both the year 2011 and for the years 2016, 2017, and 2018, we thus expect the cosmic background to be suppressed with respect to that from pp collisions by a factor equaling the luminosity ratio $L_{2016+2017+2018}/(3 \times L_{2015}) = (36 + 42 + 59)/(3 \times 5) = 9.1$. In addition, the collision backgrounds are now actually larger due to radiation damage (in particular in the first half of 2016) while the cosmic background most probably is not growing with the same proportion as it arises from real low dEdx hits. So the suppression factor for cosmic rays will in reality be even larger.

Given the at least order of magnitude difference between the cosmic and collision background and the fact that the collision background is typically very small in the signal region with a significant systematic uncertainty, the cosmic background is expected to play no role, and we can safely neglect it.

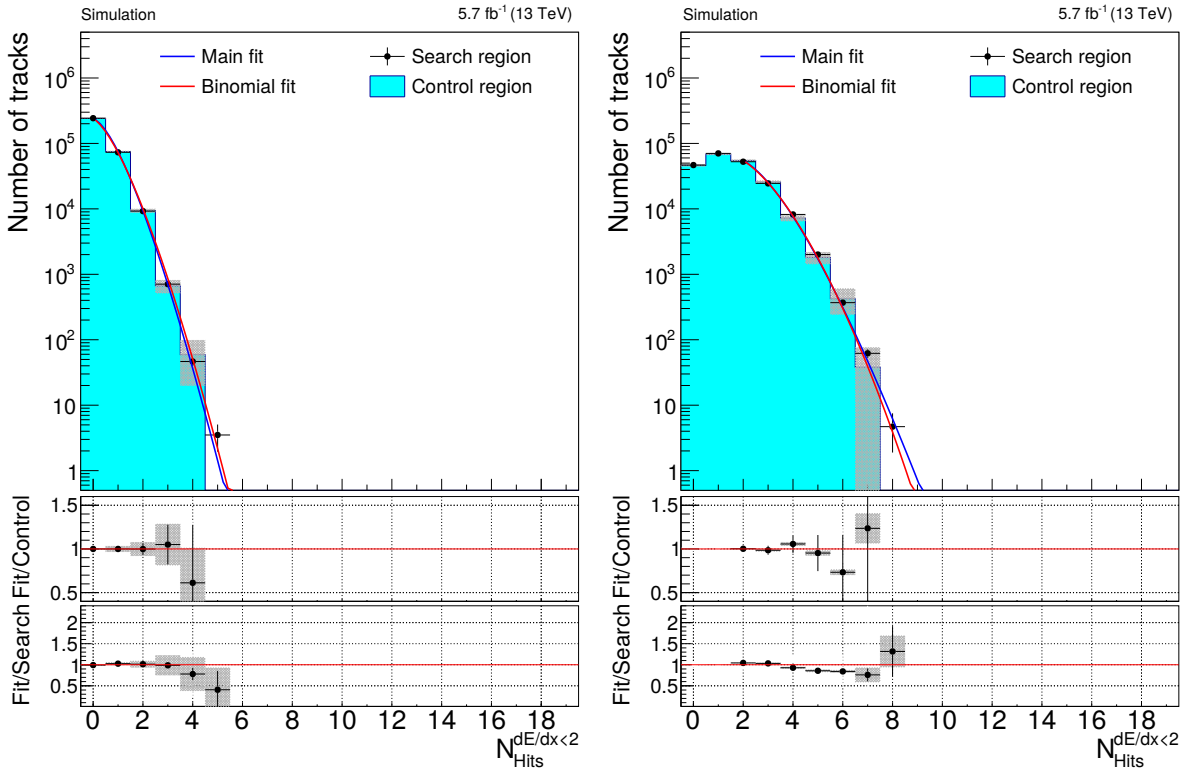


Figure 6.29: Application of the background estimation procedure on simulated Drell-Yan events for uncorrected (left) and corrected (right) dE/dx spectrum.

6.12 Systematic uncertainties

In this section, we give the detail on the methods used to compute the main systematic uncertainties. Systematic uncertainties are considered fully correlated between bins from a given data period. When combining data periods, systematic undertainties are considered fully uncorrelated.

6.12.1 Uncertainties on the background

Fit of the control region

The systematic uncertainty in the fit to the control region is built of two components: the uncertainty statistically driven by the number of events in the control region and the uncertainty coming from the deviation of the main fit from the binomial behaviour.

To derive the statistical component of this uncertainty, we generate 10,000 $N_{\text{hits}}^{\text{dE/dx<2}}$ toy distributions. The number of events in each bin of the toy distribution is randomly generated according to a Poisson distribution with a mean given by the main fit function. Next, each toy distribution is fitted with the main fit function and, for each $N_{\text{hits}}^{\text{dE/dx<2}}$ bin, the obtained fit value is stored. This way, we obtain a distribution of the main fit central value of which the width quantifies the statistical uncertainty on it. This is shown in figure 6.30 for early 2016 data in the bins considered for the search and is done separately for each data period. To each bin is associated an asymmetric uncertainty from the up and down standard deviation of the distribution.

The uncertainty coming from the deviation from the binomial function is given by the relative deviation of the binomial fit function with respect to the main function:

$$\text{Systematic uncertainty} = \frac{|\text{fit}^{\text{main}} - \text{fit}^{\text{binomial}}|}{\text{fit}^{\text{main}}}. \tag{6.18}$$

As opposed to the uncertainty on the signal corrections, the uncertainties on the fit only depend on the cut value. Tables 6.12 to 6.14 list the expected number of background tracks in

Table 6.11: Results of the binomial and main fits to the low dE/dx hit multiplicity per track distribution in DY simulations. The fits start at bin $N_{\text{hits}}^{\text{dE/dx<2}} = 0$ (uncorrected DY) and bin $N_{\text{hits}}^{\text{dE/dx<2}} = 2$ (corrected DY).

		Uncorrected DY	Corrected DY
Main fit	p_1	0.02	2.42
	p_2	-1.24	-0.76
	$\chi^2/\text{N}_{\text{d.o.f.}}$	0.20	0.30
Binomial fit	μ (fixed)	14.0	14.0
	p (%)	2.1	10.2
	$\chi^2/\text{N}_{\text{d.o.f.}}$	1.78	0.35

the considered $N_{\text{hits}}^{\text{dE/dx<2}}$ bins and the corresponding numerical value of the fit statistical and systematic uncertainties for the four data periods.

		$N_{\text{hits}}^{\text{dE/dx<2}}$ = 8	= 9	= 10	= 11	≥ 12
Early 2016	$\text{N}_{\text{BG}}^{\text{expected}}$	166	21.6	2.53	0.27	0.03
	Stat. (%)	$^{+45.1}_{-35.6}$	$^{+69.1}_{-44.1}$	$^{+100.9}_{-51.4}$	$^{+138.9}_{-56.8}$	$^{+161.2}_{-54.2}$
	Syst. (%)	35.0	56.4	75.5	88.9	96.5

Table 6.12: Expected number of background tracks and systematic uncertainties (%) on the fit to the control region in the considered $N_{\text{hits}}^{\text{dE/dx<2}}$ bins for early 2016 data.

		$N_{\text{hits}}^{\text{dE/dx<2}}$ = 4	= 5	= 6	≥ 7
Late 2016	$\text{N}_{\text{BG}}^{\text{expected}}$	593	28.8	1.09	0.03
	Stat. (%)	$^{+38.0}_{-28.7}$	$^{+76.8}_{-43.0}$	$^{+131.9}_{-53.0}$	$^{+191}_{-55.3}$
	Syst. (%)	31.5	57.5	83.2	102
2017	$\text{N}_{\text{BG}}^{\text{expected}}$	1292	64.0	2.57	0.09
	Stat. (%)	$^{+12.0}_{-11.3}$	$^{+18.3}_{-16.2}$	$^{+25.4}_{-21.3}$	$^{+32.6}_{-25.4}$
	Syst. (%)	4.1	2.4	14.4	31.2

Table 6.13: Expected number of background tracks and systematic uncertainties (%) on the fit to the control region in the considered $N_{\text{hits}}^{\text{dE/dx<2}}$ bins for late 2016 and 2017 data.

		$N_{\text{hits}}^{\text{dE/dx<2}}$ = 5	= 6	= 7	≥ 8
2018	$\text{N}_{\text{BG}}^{\text{expected}}$	690	35.9	1.53	0.06
	Stat. (%)	$^{+17.5}_{-14.9}$	$^{+27.6}_{-21.6}$	$^{+40.0}_{-28.0}$	$^{+52.7}_{-33.2}$
	Syst. (%)	45.4	62.7	71.8	67.0

Table 6.14: Expected number of background tracks and systematic uncertainties (%) on the fit to the control region in the considered $N_{\text{hits}}^{\text{dE/dx<2}}$ bins for 2018 data.

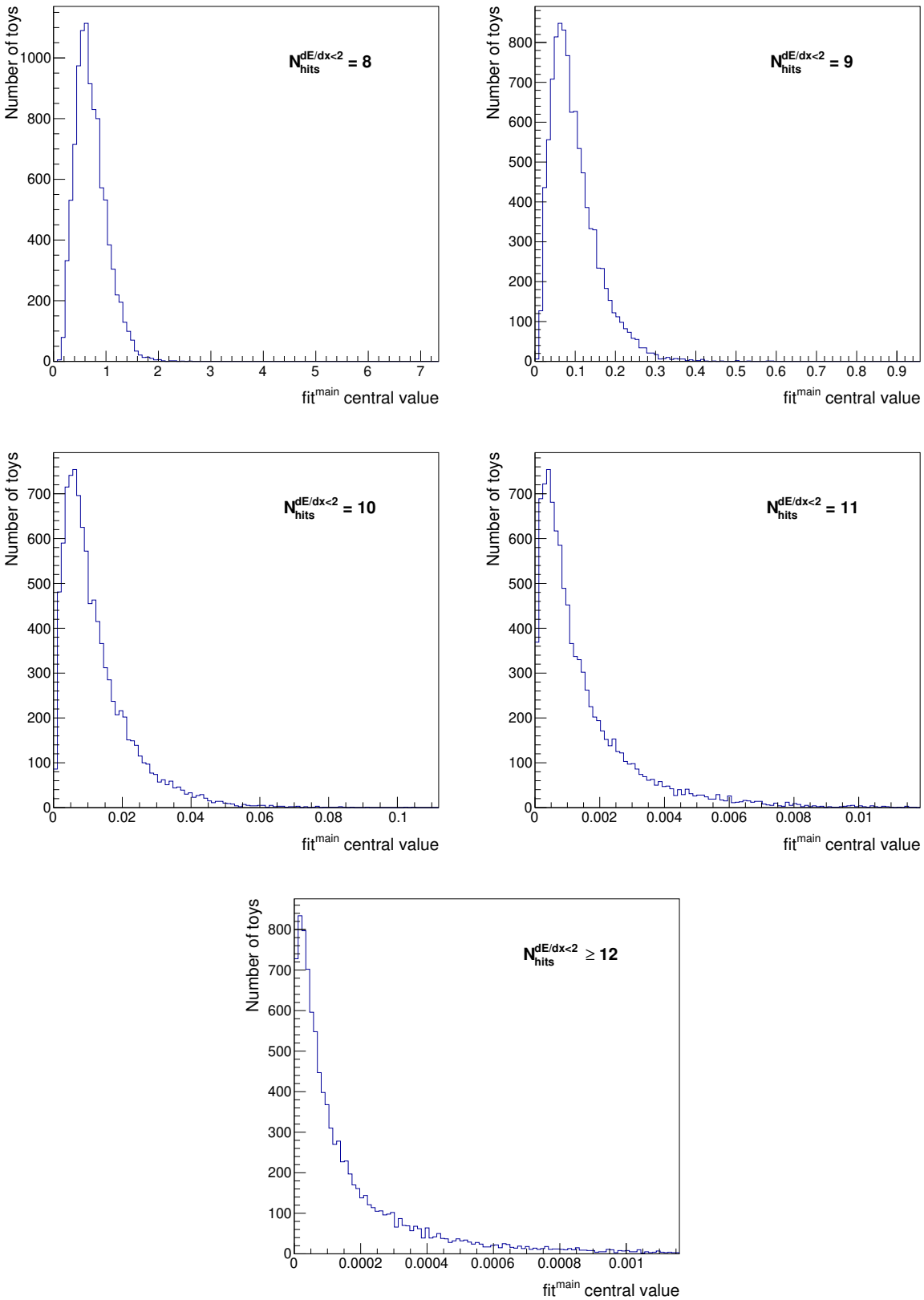


Figure 6.30: Distribution of the main fit value in $N_{\text{hits}}^{\text{dE/dx<2}}$ bins 8 to 11 (exclusive) and 12 (inclusive) for early 2016 data in 10,000 toys. The asymmetry grows with the bins up to a point where we must take the effect into account. The up and down standard deviations are used as estimations of the statistical uncertainties of the fit in the considered bin.

6.12.2 Uncertainties on the signal

Luminosity

The signal is scaled to the luminosity, hence a systematic uncertainty should be assigned to it. An uncertainty of 2.5 % in 2016 and 2018 and of 2.3 % in 2017 is considered [141, 142, 143].

Monte Carlo statistics

The signal cross-section is computed with PYTHIA8 for different mass points and scaled with the square of the charge. 50,000 events are generated for each mass scenario leading to a statistical uncertainty of 0.3 %. We consider it small enough to be ignored.

Signal selection

Uncertainties coming from the discrepancies between the selection of signal and data events are to be taken into account. We go through each of them here.

We apply centrally derived trigger scale factors for the paths defined in section 6.6. However, these scale factors are computed over a denominator of high-p_T ID muons while we use loose ID muons. We evaluate the efficiency of the two identification criteria on tracks passing the full preselection but the identification requirement. We then define the systematic uncertainty as the relative difference between the two efficiencies. It shows to be lower than 1 % in all mass scenarios and for all three data-taking years. Consequently, we apply a 1 % uncertainty on all scenarios to take this into account.

We do not apply identification or isolation scale factors as they are equal to one up to the level of a percent in the phase-space of the analysis (high p_T, central muons). An uncertainty of 1 % is applied to cope for the fact that the loose isolation scale factors are derived for a medium and not a loose ID.

Comparing the effect of the cut on $\alpha_{\max}$ in signal and in BG MC shows that they behave the same up to the percent level. We add a systematic uncertainty of 1 % to be conservative.

The timeAtIP variable is expected to show a strongly different behaviour in signal and in data as it depends on the mass of the particle. We use the inefficiency of this cut on signal as the uncertainty itself. It is observed to be constant across the years and to decrease with higher masses. It is of 6 % for a signal mass of 100 GeV, 3 % for 300 GeV, 2 % for 500 GeV and 1 % for 1000 GeV.

Signal correction at low dE/dx

As described in section 6.8.3, the signal simulation is corrected for the fact that the CMS simulation underestimates the proportion of low dE/dx hits. The effect of these corrections is to move hits from the bulk of the dE/dx distribution to the < 2 MeV/cm region. The lower the charge, the less the signal $N_{\text{hits}}^{\text{dE/dx}<2}$ is affected because it already lives in this region. Thus those corrections are charge-dependent. Although derived using background MC in the search region, we will use the background MC in the control region to derive the uncertainty on this correction method. We have explicitely verified that the corrections bring the simulations close to data in the control region (see figure 6.17) which acts as a closure test. The procedure to derive systematic uncerainties on these corrections is the following. We redefine the dE/dx interval to be Q-dependent, I $= [\frac{0.25}{Q^2}, \frac{2}{Q^2}]$, and count the number of hits in it $N_{\text{hits}}^{\text{dE/dx}\in\text{I}}$. Then, we evaluate this new variable at each $N_{\text{hits}}^{\text{dE/dx}<2}$ bin in data $N_{\text{hits}}^{\text{dE/dx}\in\text{I,data}}$ and in DY $N_{\text{hits}}^{\text{dE/dx}\in\text{I,DY}}$ both in the control region. With these numbers, we estimate the uncertainty as:

$$\text{Systematic uncertainty} = \frac{|N_{\text{hits}}^{\text{dE/dx}\in I_Q,\text{data}} - N_{\text{hits}}^{\text{dE/dx}\in I_Q,\text{MC}}|}{N_{\text{hits}}^{\text{dE/dx}\in I_Q,\text{data}}}. \tag{6.19}$$

This uncertainty depends on the signal charge as well as on the considered $N_{\text{hits}}^{\text{dE/dx<2}}$ bin. As for the expected number of background events, the last bin is taken inclusively up. One can directly visualize these systematic uncertainties by comparing data to Drell-Yan simulated events in figure 6.31. It gives the distribution of the number of hits in the interval I (as defined above) for the six charge points considered. With the exception of Q = 1/3, the data to MC agreement is quite good.

It happens that some bins do not contain simulated events, or reversely that no data is present but there are simulated events. In these cases, the uncertainty is taken as the statistical uncertainty on the non-empty bin. When both data and DY yield a non-zero bin content, the uncertainty is taken as the largest of the following three items: the statistical uncertainty on data, the statistical uncertainty on DY and the uncertainty derived from the aforementioned method. This ensures a conservative approach.

Tables 6.15 to 6.17 give the uncertainties for the four data periods and for different signal charges and cuts on $N_{\text{hits}}^{\text{dE/dx<2}}$.

	Signal charge	$N_{\text{hits}}^{\text{dE/dx<2}}$ 8	9	10	11	≥12
Early 2016	2/3	34.7	24.3	20.0	10.1	14.4
	1/3	39.2	23.3	9.9	2.2	18.2

Table 6.15: Systematic uncertainties (%) associated with the correction of the signal simulation dE/dx spectrum for early 2016 data.

	Signal charge	$N_{\text{hits}}^{\text{dE/dx<2}}$ 4	5	6	≥7
Late 2016	2/3	91.7	25.9	72.4	11.4
	1/3	60.9	22.8	55.2	11.3
2017	2/3	38.0	40.9	28.2	86.2
	1/3	46.4	66.3	46.6	27.9

Table 6.16: Systematic uncertainties (%) associated with the correction of the signal simulation dE/dx spectrum for late 2016 data and 2017 data.

	Signal charge	$N_{\text{hits}}^{\text{dE/dx<2}}$ 5	6	7	≥8
2018	2/3	51.3	34.9	23.3	14.8
	1/3	44.1	28.4	14.0	15.7

Table 6.17: Systematic uncertainties (%) associated with the correction of the signal simulation dE/dx spectrum for 2018 data.

6.13 Results

We compute expected and observed upper limits on the FCP production cross-section $\sigma_{pp\to L_{\text{Q}}\bar{L}_{\text{Q}}}$ for a given charge and mass scenario. Limits are set at 95 % confidence level using the CL_s method [144, 145]. Let us briefly explain how this method works. First of all, a test statistic q_r must be defined in order to find the signal strength r that is the most likely to have produced

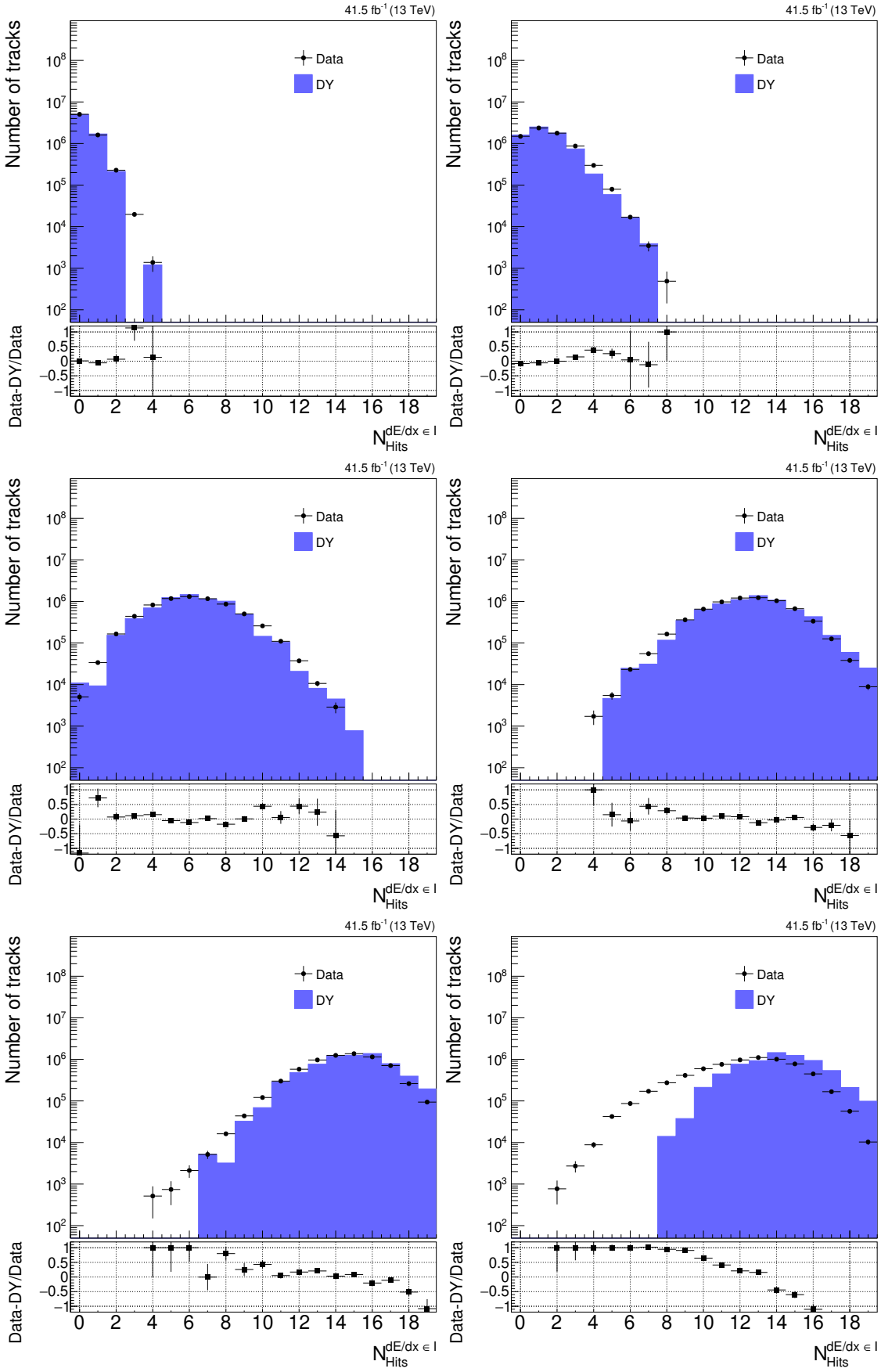


Figure 6.31: Data (2017) to MC comparison of the $N_{\text{hits}}^{\text{dE/dx}\in I}$ distribution for the six charge points. From top-left to bottom-right, Q = 1 down to Q = 1/3.

the observed data under the signal + background (s+b) hypothesis. At the LHC, the chosen test statistic is the following profile likelihood ratio:

$$q_r = -2\ln\left[\frac{\mathcal{L}(\text{data}|rs+b,\hat{\theta}_r)}{\mathcal{L}(\text{data}|\hat{r}s+b,\hat{\theta})}\right] \tag{6.20}$$

The two $\mathcal{L}$ functions should be read as "the likelihood for a number of signal events rs (or $\hat{r}s$) and a number of background events b to have produced the observed data, given the nuisance parameters $\hat{\theta}_r$ (or $\hat{\theta}$). The idea of a profile likelihood is to express all parameters as a function of the one that is being estimated (here the signal strength). In the numerator, the nuisance parameters are left free floating (this is the meaning of the hat) for the considered value of the signal strength. In the denominator, both the signal strength and the nuisance parameters may optimize their value. Maximizing each likelihood yields a value of the test statistic for a given value of r. To compute the CL_s for a given signal strength, we need to derive the probability distribution functions of the expected value of the test statistic in the background-only and signal + background hypotheses, f_r^b and f_r^{s+b}. To do so, toy datasets are generated based on the prior knowledge of the observed data. For each toy, a different value of the test statistic will be computed and its probability distribution function can be drawn. This is shown in figure 6.32 in early 2016 data for a signal of charge 2/3 and of mass 100 GeV and a signal strength $r = 3.8$.

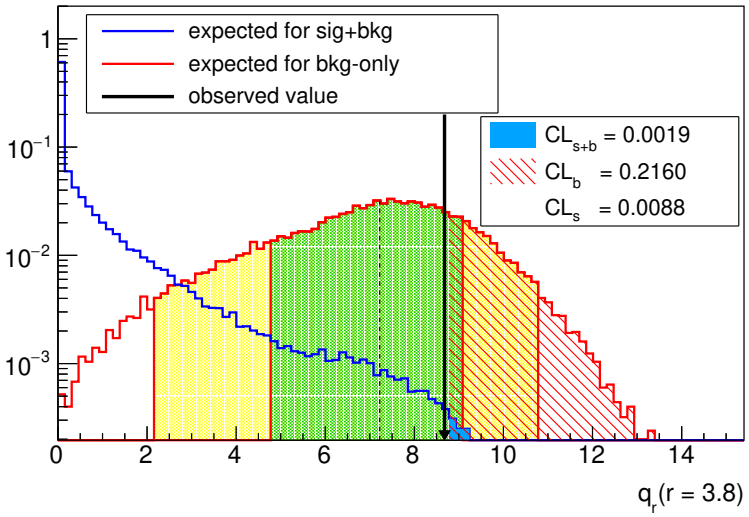


Figure 6.32: Distribution of the test statistic in early 2016 data for a signal of charge 2/3 and of mass 100 GeV and a signal strength $r = 3.8$. The black down arrow gives the value of the test statistic observed in data while the blue and the red distributions are the probability distribution functions in the $s + b$ and b scenarios generated from 100,000 toy datasets. The blue filled area defines the probability CL_{s+b} for the test statistic in a $s+b$ scenario to be larger than the observed one. The hashed region is similarly defined for CL_b in the b scenario. The ratio $\text{CL}_{s+b}/\text{CL}_b$ yields the CL_s corresponding to the signal strength of 3.8.

The $\text{CL}_s(r)$ is defined as the ratio of the two probabilities:

$$\text{CL}_s(r) = \frac{\text{CL}_{s+b}(r)}{\text{CL}_b(r)}, \tag{6.21}$$

where $\text{CL}_{s+b}(r) = p_r^{s+b}$ is the probability that a signal strength larger than the observed one optimizes the test statistic in the $s + b$ scenario and $\text{CL}_b(r) = p_r^b$ is similarly defined in the background-only scenario. More explicitely, it gives:

$$p_r^{s+b} = \int_{q_r^{\text{obs}}}^{\infty} f_r^{s+b} \tag{6.22}$$

$$p_r^{b} = \int_{q_r^{\text{obs}}}^{\infty} f_r^{b} \tag{6.23}$$

These integrals are represented in figure 6.32 as a blue and a hashed area for p_r^{s+b} and p_r^b respectively. We see that a signal strength value of 3.8 yields a confidence level of 1-0.0088 = 99.12 %. To find the signal strength corresponding to a CL_s of 0.05 (confidence level of 95 % which is the standard level), we produce a grid of signal strength values and we compute the CL_s for each signal strength value in 2000 toy datasets. The CL_s value is then plotted as a function of the signal strength and fitted in order to extract the signal strength value corresponding to a confidence level of 95 % (CL_s 0.05). This is done automatically by the CMS-recommended tool to extract the upper limits. This is illustrated in figure 6.33 for early 2016 data and a FCP mass of 100 GeV and a charge of 2/3. The amplitude of the error bars on the CL_s values (in black) are determined by the number of toys generated at each signal strength point. The central red vertical line gives the value of the signal strentgh corresponding to a CL_s of 0.05. The two other red lines give the left and right statistical uncertainties on the computed signal strength. They get closer to the central value (i.e. the uncertainties get smaller) for increasing numbers of toys and increasing densities of the signal strength grid. By density we mean the number of signal strength points per signal strength interval.

Figure 6.34 presents the limits on $\sigma_{pp\to L_{2/3}\bar{L}_{2/3}}$ in the four data periods. We observe the impact of the experimental conditions on the excluded cross-section. The less-irradiated detector in 2017 and 2018 yields the lower excluded cross-section thanks to a larger dataset and a new pixel detector with an extra layer and a shorter radiation history.

In figure 6.35 we show the limits on the signal production cross-section computed on the full dataset. Six different charge scenarios are presented, ranging from Q = 1 down to Q = 1/3. For each charge scenario, we observe a certain slope of the upper limit as a function of the mass. This slope is different for different charges. To understand this, let us recall that the ionization energy deposited by the FCP in the tracker modules increases with the charge and the mass of the particle. If the ionization energy gets so low that the hit is not readout, FCP tracks may be missed by the tracking alogrithm by lack of a decent number of hits. This being said, two different regimes can be met:

- High charge FCP (down to Q = 2/3 included) have enough associated hits for their track to be reconstructed no matter their mass. In that case, the higher the mass, the higher the hit dE/dx. Since we select our signal on the number of low dE/dx, we are less sensitive to higher masses and the upper limit increases. This explains the rising slope of the upper limit.

- For low charge FCP (Q = 0.5 and 1/3), we are in a regime where the particle feels the effect of tracking failure. In this case, a higher mass means more reconstructed hits, therefore a better tracking efficiency and a higher signal selection efficiency.

There is a compromise between the loss of sensitivity due to high ionization and the loss of sensitivity due to a low signal selection. As a consequence, the best limits are obtained for intermediate charges Q = 2/3 and 0.5. The CMS analysis published in 2013 [132] excluded the existence of such particles up to masses of 310 GeV (Q = 2/3) and 140 GeV (Q = 1/3). As is shown here, we push those limits to 765 GeV (Q = 2/3) and 245 GeV (Q = 1/3). Moreover, we interpolate with more charge points from 1 down to Q = 1/3. The difference in gain between charge 2/3 and 1/3 comes from the fact that the harsher experimental conditions in Run 2 have a stronger impact on lower charges. The dE/dx spectrum in Run 2 shows more hits in the low

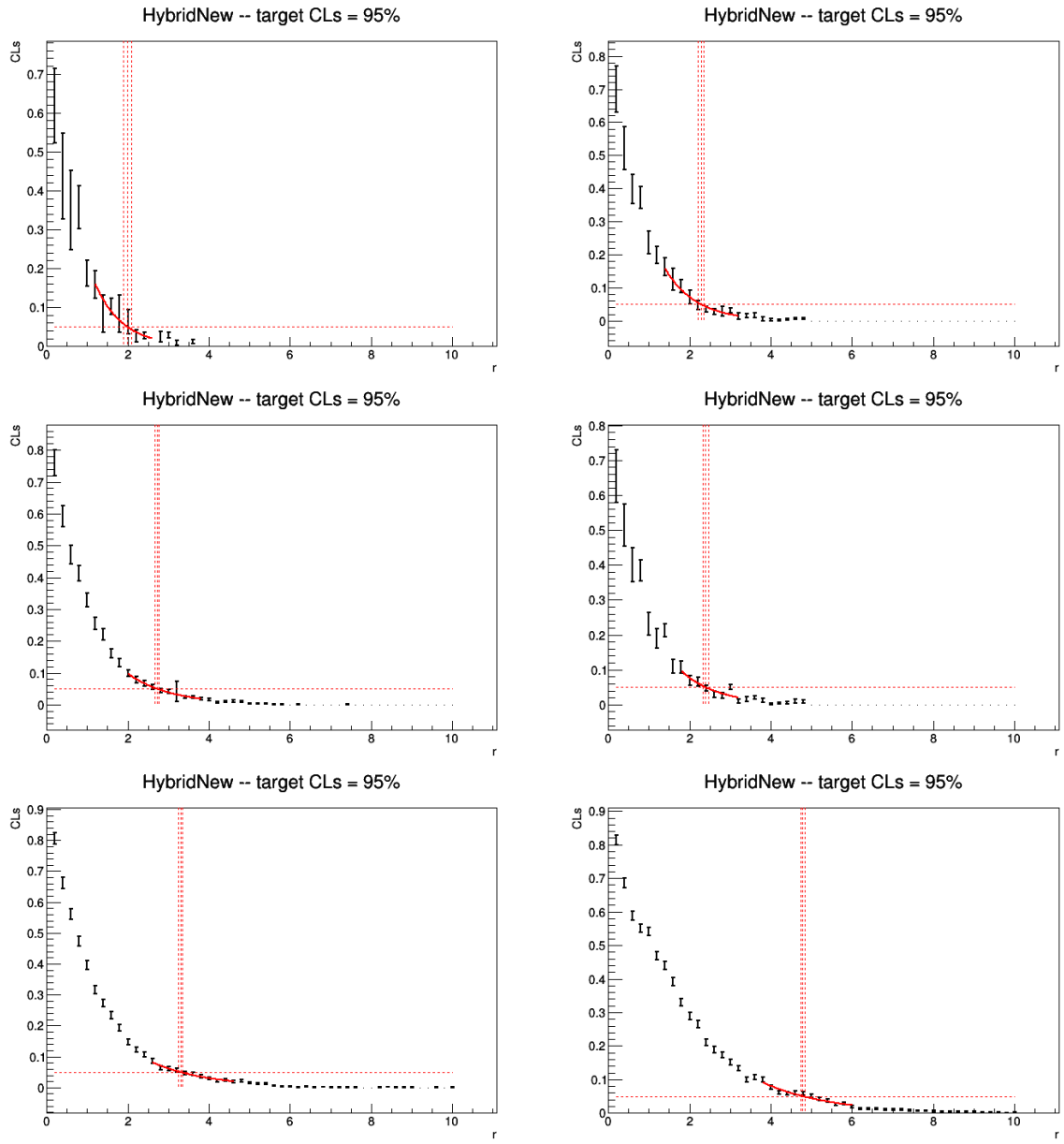


Figure 6.33: CL_s as a function of the signal strength in early 2016 data for a FCP of mass 100 GeV and a charge of 2/3. The number of signal events has been scaled by a factor 1/1000 to make the signal strentgh close to unity. The top row shows this distribution for the -2 sigma (2.75 % quantile, left) and -1 sigma (16 % quantile, right) uncertainty bands on the expected limit, the middle row shows the central value (50 % quantile, left) and the observed limit (right) and the bottom row gives the upper 1 (84 % quantile, left) and 2 (97.5 % quantile, right) sigma uncertainty bands on the expected limit. "HybridNew" is the technical name of the method used to produce the limit. It corresponds to the description in the text.

dE/dx region, even for the cleanest dataset (2017).

Figure 6.36 shows the upper limits computed with and without the systematic uncertainties on the background prediction. Here, the limits have been computed using the asymptotic formula described in reference [146]. Only the deviation of the main fit prediction from the binomial behaviour is removed, the statistical component is kept. One can see that the impact is very small on the limits, as expected from the very small background prediction.

We would like to mention here a recent paper predicting the existence of heavy long-lived fractionally charged leptons [147]. Without entering into the details of their model, we simply

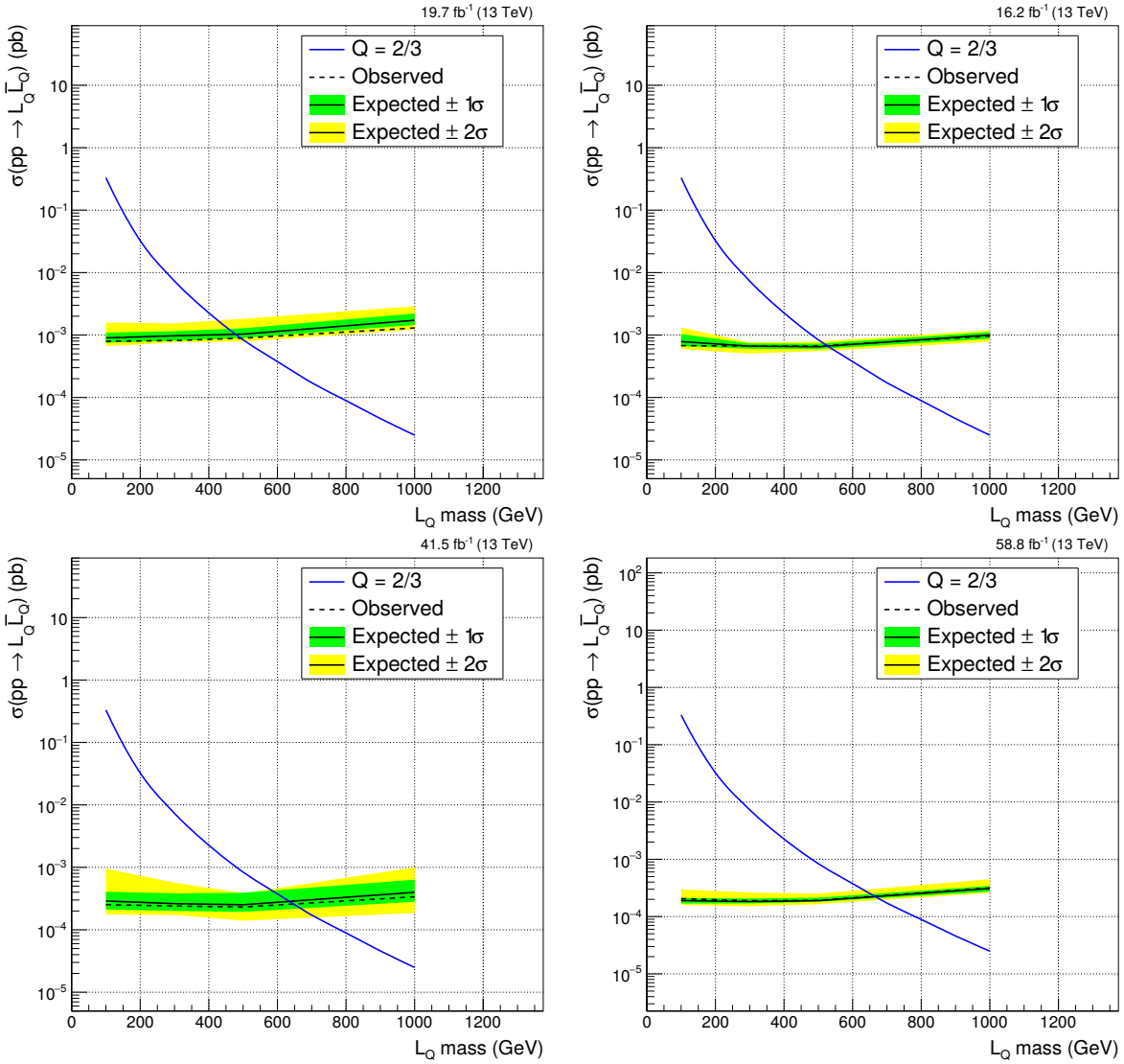


Figure 6.34: Limits on the Q = 2/3 signal cross-section using the early 2016 (top-left), late 2016 (top-right), 2017 (bottom-left) and 2018 (bottom-right) data. The green bands give the +/- 1 sigma envelope while the yellow bands give the +/- 2 sigma envelope.

note that it yields a cross-section of the order of 10^{-3} pb for the production of pairs of Q = 2/3 particles for masses between 500 and 1000 GeV. As our upper limit in this scenario is ten times smaller, and although their computation has been done at 14 TeV (which means the cross-section at 13 TeV would be smaller), the authors would probably be interested in reading through our analysis to determine whether our results allow to exclude part of the phase-space in which their model lives.

Finally, we combine the limits for each charge points in a single charge-mass two-dimensional plot, as was given by figure 6.1. This is shown in figure 6.37 with comparison to previous CMS analyses. Not only do we exclude higher masses (or lower cross-sections) for the charge 2/3 and 1/3 scenarios, we also extend the exclusion region to other charge points.

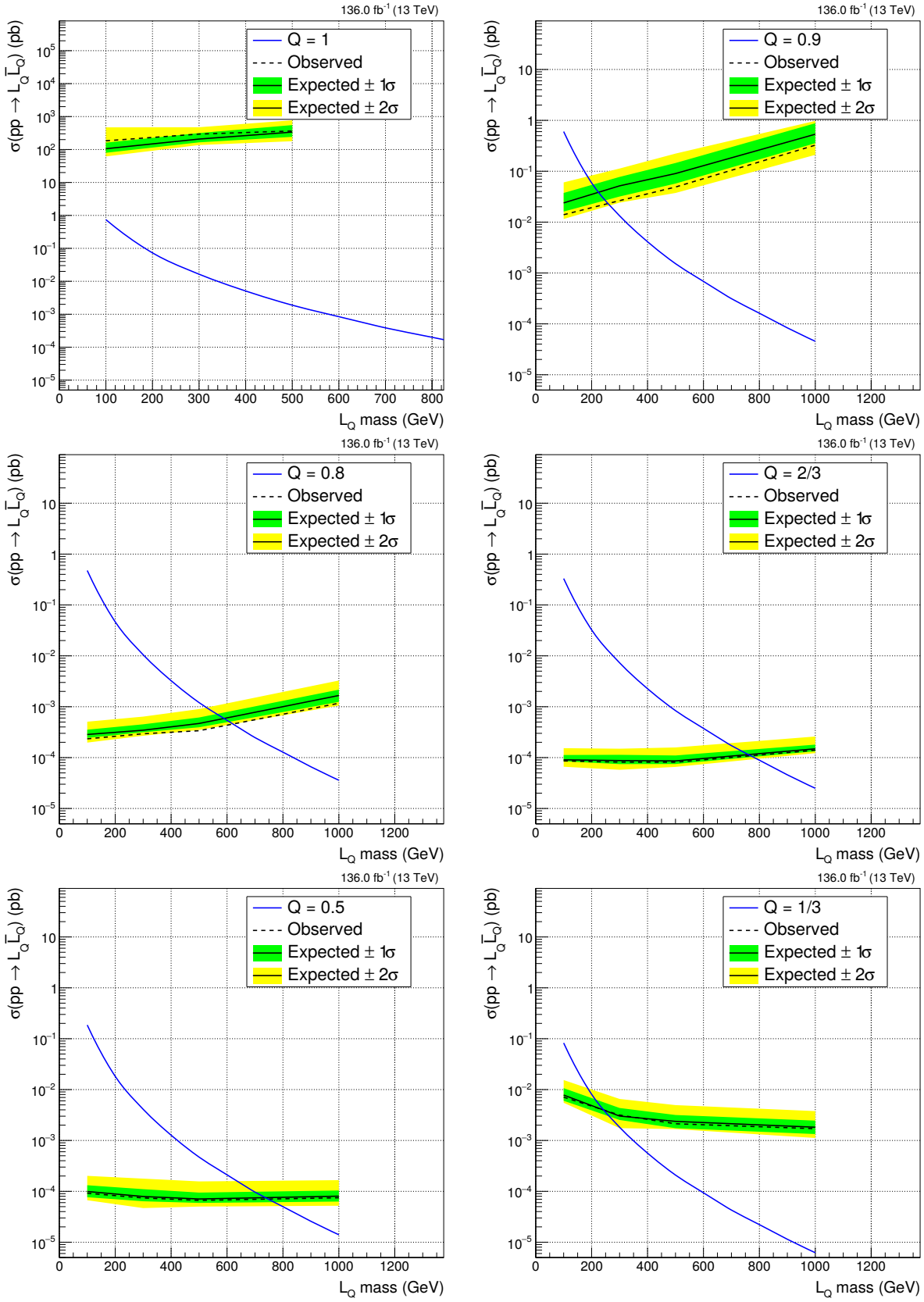


Figure 6.35: Limit on the Q = 1, 0.9, 0.8, 2/3, 0.5 and 1/3 signal cross-section using the full dataset. For Q = 1, no limit is available for a signal of mass M = 1 TeV as zero signal events are expected due to a lack of simulated events.

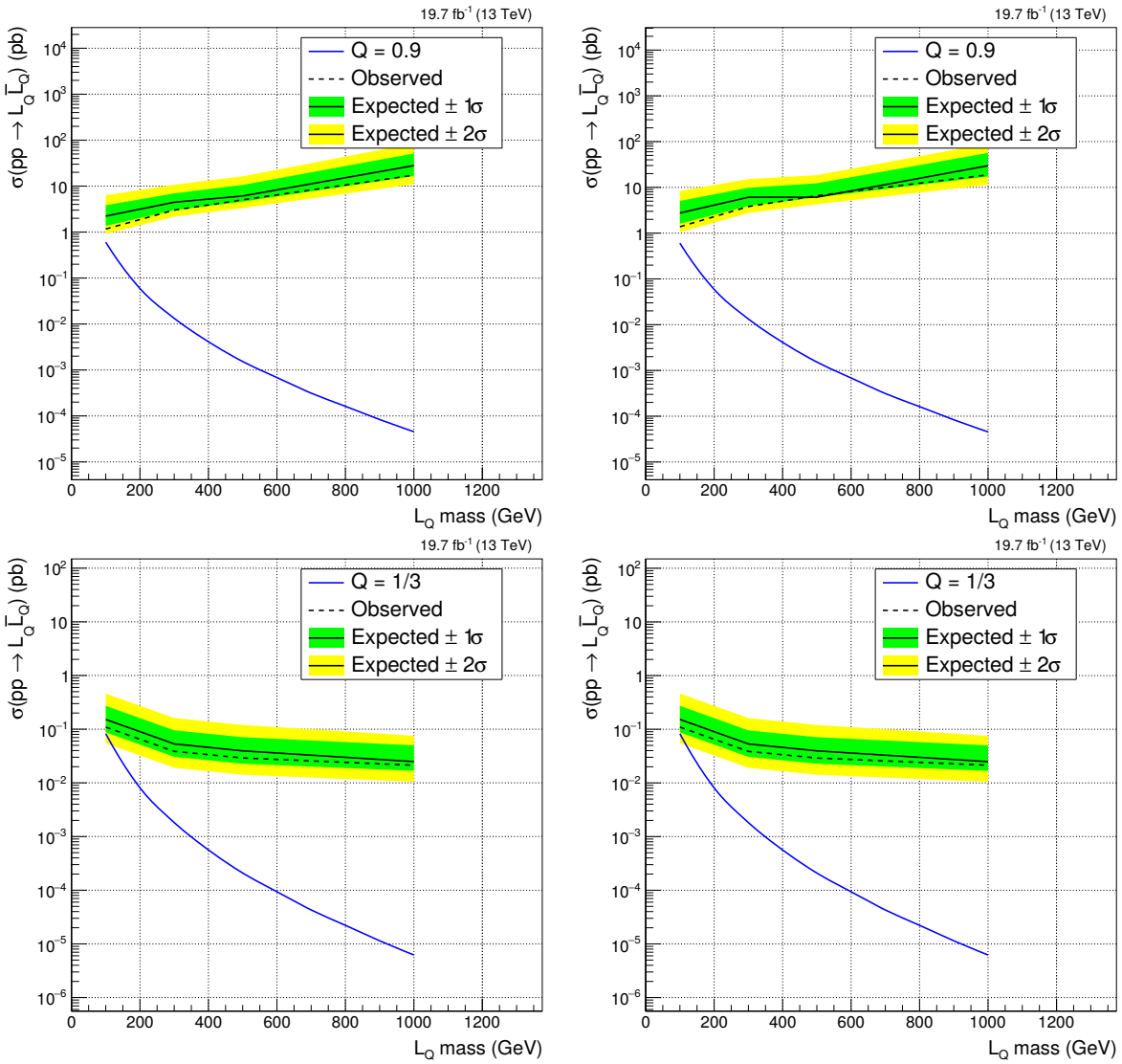


Figure 6.36: Asymptotic upper limits on the signal cross-section without (left) and with (right) systematic uncertainties on the background prediction for Q = 0.9 (top) and Q = 1/3 (bottom) particles for early 2016 data. Only the uncertainty coming from the deviation of the main fit from the binomial is removed, the statistical component is kept.

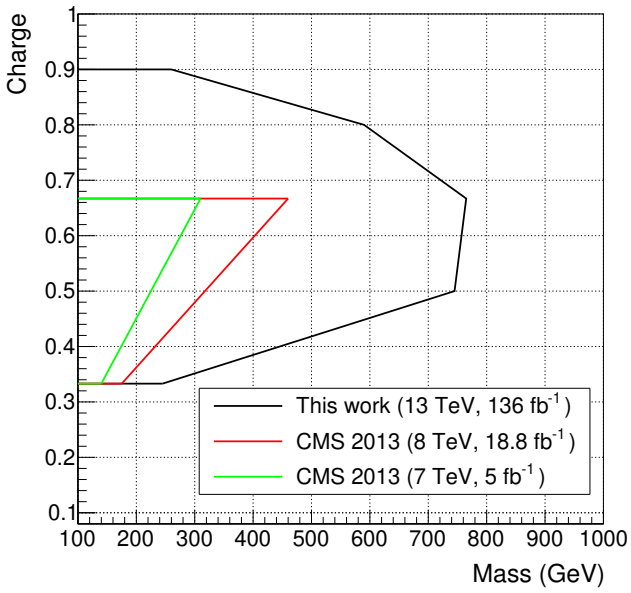


Figure 6.37: Exclusion limits on the signal cross-section in the FCP charge-mass plane. The FCP charge is given in units of the electron charge. The regions on the left of the curves are excluded at 95 % confidence level. The space between the discrete charge-mass points is interpolated with straight lines. Our work allows to push further the existing limits, to exclude higher charge particles and to better interpolate between the 1/3 and 2/3 limit points.

Chapter conclusion

We have presented a search for long-lived particles carrying a fractional electromagnetic charge using 136 fb^{-1} of data collected with the CMS detector at $\sqrt{s} = 13$ TeV. The target range is Q $\in$ [1/3,1]. We have shown that the existence of such particles is well motivated in the context of supersymmetric theories, string theories and Grand Unified Theories. The main existing constraints are from LEP at low masses and from the observation of the CMB anisotropy and the DM relic abundance at high masses. As low charge particles deposit less energy by ionization in matter, the analysis strategy is to look for tracks in the detector with a large number of low dE/dx hits. We had to take into account that the detector is not ideal but is subject to radiation damage and other effects that correlate the different tracker layers. Therefore, we had to mitigate these effects by ignoring some hits or even full tracks in the track candidates collection. We also derived and applied corrections to the signal dE/dx spectrum to cope with this. Then, we estimated our background by fitting our control region with a parametric function. This function is close to a binomial but fits data better, especially when it deviates strongly from the ideal binomial behaviour. Comparing the prediction from our background estimation procedure with the observations, we see no excess in data. Consequently, we set upper limits on the signal cross-section using the CL$_s$ method at 95 % confidence level. Cross-sections as low as 10^{-4} pb are excluded for Q = 2/3, corresponding to a lower limit on the FCP mass at 765 GeV.

Conclusion

In this thesis, we have presented the results of two searches for new physics using proton-proton collision data delivered by CERN's LHC. The data were collected during the LHC Run 2 (i.e. between 2016 and 2018) with the CMS detector at a center-of-mass energy of 13 TeV and corresponds to an integrated luminosity of 136 fb^{-1}.

In a first analysis, we have searched for a signal of Dark Matter as Weakly Interacting Massive Particles (WIMPs) in 2016 data. These particles are expected to escape the detector unnoticed leading to missing energy in the event. In order for the detector to record a signal, we target particular events in which one or several jets are produced alongside the DM particles. Due to the large cross-section of processes with Initial State Radiation (ISR) produced in a QCD interaction, this "monojet" analysis is one of the most sentitive channel for DM searches at colliders. In such channel, a DM signal would show up as an excess of data in the missing transverse energy spectrum. Therefore, the analysis strategy is to carry out a "shape" analysis in this spectrum by comparing the expectation from Standard Model backgrounds to the observations. No excess is observed in the dataset considered and upper limits on the signal production cross-section are set at 95 % confidence level. The existence of a SM-DM mediator is excluded up to a mass of 1.8 TeV in the vector scenario, which corresponds to a DM mass up 700 GeV. These results are complementary with non-collider results in direct and indirect DM detection experiments.

We then present the results of a search for long-lived fractionally charged particles (FCP), i.e. particles with an electric charge lower than that of the electron. Those particles are expected to deposit less energy by ionization in the CMS tracker with respect to Standard Model particles. As a consequence, FCP tracks are expected to contain a significantly higher number of hits with a low stopping power dE/dx. The analysis strategy is to count the number of low dE/dx hits per track and to look for an excess in data in the spectrum of this variable. The expected background from Standard Model tracks is predicted using a fit from a control region made of dimuon events around the Z boson mass peak. A shape analysis is performed in the spectrum of the signal extraction variable and no excess in data is found. Consequently, upper limits on the signal cross-section are set at 95 % confidence level. FCP with masses as large as 800 GeV are excluded for a charge of 2/3 in units of the electron charge. These are the world best results for such signals in the [100 GeV - 1 TeV] mass range and in the [1/3 - 1] charge range.

Regarding this last analysis, we considered several new ideas and potentially interesting developments. We believe there is room for improvement mainly in two ways that we mention here as a potential starting point for others to carry them forward. A first obvious upgrade this analysis could benefit from would be to enlarge the statistics in the control region. This would strenghten the background prediction procedure hence enhancing the sensitivity to a low excess in data. It would be helpful in particular in the search for high charge ($Q \in [0.8,1]$) particles, where the sensitivity is driven by bins with non-zero background. A way to do this would be to use events collected with the low p_T threshold triggers. We explicitely did not use those triggers in order not to drown our signal region with SM muons but it would be possible to enrich the control region with these events. Now in this case, the control region and the signal region would not be built exactly from the same dataset. Care should be taken as the p_T spectrum of the muons in the two datasets would be different. Some corrections due to the use of different triggers should

also probably be applied.

We have seen in figure 6.7 that very low charge particles ($Q \leq 1/3$) have a high chance to have no reconstructed track in the silicon tracker due to the too low number of reconstructed hits. However, the track reconstruction efficiency stays reasonable (60 - 70 %) in the muon chambers. A strategy to target very low charge particles would then be to have a "standalone muons" analysis. The idea would be to require two isolated muons that only have a track in the muon system, drastically suppressing the SM background processes. As the stopping power is not recorded in the muon system, no dE/dx-based analysis is possible. In other words, no identification of the new particle charge is possible. Therefore, the strategy would be entirely different: either count the number of SA muons or perform a "bump hunting" search. Comparing the number of "no track" SA muons expected from background to that from our signal model could be done in the p_T spectrum that we know is enhanced for fractionally charged particles. After suppression of as much background as possible, an excess at high p_T would be a way to identify a signal of new physics. Another strategy would be to select SA muon pairs and look for a bump-shaped excess in their invariant mass spectrum. This procedure would be focused on the search for a potential new mediator decaying to low charge particles. This could be an interesting and relevant physics project to carry out in the coming months with the existing data. Given the current efforts to improve the tracking algorithm in the muon system and to integrate it to already at Level-1, this is a nice opportuinity to carry on a SA muon analysis in the LHC Run 3.

Two drawbacks from these strategies are easily identified. First of all, there are no unprescaled SA muon triggers in CMS. One would then have to use muon triggers and select muons with no tracker track, dramatically reducing the number of events. To cope with this, one could complement the search dataset with events collected using jet or photon triggers. Interestingly, for low charge particles where one or both tracks are lost, the events with a high p_T jet would resemble monojet events. Either because the low charge particles have become invisible to CMS, or because of the effects discussed in section 6.6 that would bring $\mathrm{E}_T^{\mathrm{miss}}$ in the event. In this scenario of FCP pair production recoiling against a high p_T hadronic activity, we find ourself in the position of using the missing transverse energy as a relevant variable for the analysis. The second potential issue, especially for a dimuon bump-hunting analysis, is the poor p_T resolution on SA muons. This low resolution could potentially make us miss a narrow resonance in the dimuon invariant mass spectrum.

Invisible particles and dark sectors will be intensively searched for in the near future. The LHC and its four experiments are being upgraded with new hardware technologies and more advanced algorithms. Considering that the center-of-mass energy may be raised to 14 TeV but most of all that an enormous amount of data is foreseen (3000 fb^{-1} are expected to be recorded at the High-Lumi LHC), a significantly increased sensitivity will be reachable in the coming years. But CMS is not the only experiment looking for dark matter or low charge particles. The MilliQan experiment, currently in a test phase, will search for particles with charges as low as $\sim 10^{-3}$ e by looking at the pp collisions occuring inside CMS but from 17 meters away in a rock-shielded cavern. The SHIP experiment is currently under construction with the aim to uncover a signal of very weakly interacting long-lived particles using the proton beam from the CERN SPS. Non-collider experiments are also on track for the quest for new physics. The IceCube Neutrino Observatory has produced the world-best limits on the spin-dependent DM-nucleon cross-section. The collaboration is currently in the process of upgrading the core of the detector to increase its resolution and lower the energy threshold at which it can detect astrophysical neutrinos. CDMS, a direct detection experiment, is about to take its first data in its upgraded version (SuperCDMS). Many other experiments around the globe are pushing further the efforts to unravel the mysteries of the dark and the hidden sectors. There is still much science to do.

Acknowledgements

De toute évidence, la première personne que j'aimerais remercier et pour laquelle j'ai développé une affection particulière depuis huit ans déjà, est Barbara Clerbaux. Barbara qui m'a soutenu, conseillé et encouragé non seulement sur les sujets scientifiques mais bien au-delà lorsque la thèse m'a embarqué sur des chemins plus administratifs ou stratégiques. Elle m'a aidé, et m'aide toujours, à penser ma thèse comme une formation globale, un parcours de vie qui dépasse les frontières de la physique. J'ai rencontré Barbara à l'automne 2011, lorsque j'ai dû trouver un projet pour ma troisième année de bachelier. Elle m'a fait lire et vulgariser un article de CMS sur la recherche du boson scalaire qu'en toute honneteté je n'ai pas vraiment compris. Il faut dire qu'à l'époque, j'étais ingénieur. En deuxième année de master, je la recontacte pour lui demander d'encadrer mon mémoire. Elle accepte, et c'est alors qu'elle me forme réellement à ce qu'est la physique des particules élémentaires. Parti pour Paris m'initier aux mystères de la physique théorique, nous sommes alors en 2014, Barbara me rapatrie à Bruxelles pour commencer une thèse de doctorat. Je ne sais pas s'il sagit de confiance ou de politesse, mais je ne peux que constater que jamais elle n'a refusé de m'encadrer sur un de mes projets. Aujourd'hui, et après quatre années à chercher des choses qui n'existent pas, je suis le plus épanoui des doctorants. Barbara est un être humain extraordinaire avec qui il est heureux de travailler. Je pars l'année prochaine pour les Etats-Unis, mais je compte bien repasser par la Belgique et l'IIHE donc je ne te dis pas au revoir, patron, mais à très bientôt. *Ta force m'impressionne mais si tant est que tu en aies besoin, j'aimerais te souhaiter bon courage, à toi et à tes filles.*

Steven Lowette is very close to the ideal thesis supervisor. Not mentioning the fact that he is an outstanding physicist, he also takes great interest in the work and development of his PhD students to which he dedicates a significant amount of time. But don't take me wrong: he is not over your shoulder all the time. He is at the same time highly available (maybe a bit too available, we had numerous Skype discussions after 10pm) and leaves me working my own way. Steven has stayed very human in his way of working and supervising which in the CMS Collaboration is something that is not always found. I met him at the beginning of my thesis, when we decided to have a joint PhD between the ULB and the VUB. I could not have taken a better decision. Admittedly, duplicating the administrative work was painful but working with Steven was definitely worth it. And I really hate administrative work, so this is a very serious thing for me to say. Steven likes doing physics and I believe he passed that on to me. We have spent hours dicussing physics matters, trying to understand the funny behaviour of low charge particles in the detector or finding ideas on how to deal with upcoming issues. And we had quite a few on our road. Thank you boss for all your energy and interest, I will try to continue to do physics as well as I saw you do it.

I would like to thank the members of my jury, namely Michel Tytgat, Freya Blekman, Shahram Rahatlou and Giacomo Bruno. They helped me go further in my thesis and understand subttle physics effects that I would have missed otherwise. Une pensée particulière pour Michel avec qui j'ai eu la chance d'interagir un peu plus régulièrement. Merci de m'avoir aidé et surtout d'avoir eu la patience et l'intérêt de présider le comité d'accompagnement et le jury d'une thèse en physique expérimentale. J'espère à l'avenir avoir l'occasion de discuter avec toi des livres que tu m'as conseillés.

En plus de mes promoteurs et de mon jury, j'ai eu la chance d'interagir régulièrement avec d'autres physiciens de mon laboratoire. J'aimerais commencer par Laurent Favart, notre co-directeur, avec qui j'ai eu de nombreuses discussions sur la physique, la collaboration CMS et même sur le monde et la société en général. J'ai eu de plus le plaisir d'encadrer les séances d'exercice d'un de ses cours, ce qui m'a permis d'approfondir nos échanges sur la physique des particules et sur l'enseignement. Laurent, merci pour ta gentillesse, ton sens physique et ta disponibilité.

J'aimerais également remercier Pascal Vanlaer qui a toujours été à l'écoute de mes questions, prenant le temps d'y répondre avec un niveau de pédagogie et de détails surprenant. Merci Pascal d'avoir pris le temps, et merci pour l'ambiance conviviale que tu fais régner dans le laboratoire.

Je n'aime pas le reconnaitre, mais une thèse en physique des particules est impossible à faire aboutir sans l'aide de postdoctorants bienveillants. Ils m'ont fourni les trucs et astuces, les liens vers des sites utiles, les morceaux de code ou bien simplement le soutien dont j'avais besoin pour avancer. Merci donc à Laurent Thomas, qui depuis 2013 lorsqu'il encadrait mon mémoire, me sert de référence dans CMS et à l'IIHE. Merci à Cécile Caillol pour sa disponibilité et sa gentillesse, j'espère qu'on aura l'occasion de se revoir à Bruxelles ou ailleurs. Thanks to Giuseppe Fasanella for his help and support in the first years of my thesis. I also want to thank Raffaele Gerosa with whom I collaborated a lot and who was so helpful and available while not working in the same physical location.

A special thank to my office mates during these four years who made the thesis a little more fun. First in the upper office: Giuseppe, Cécile and Diego. And the people from downstairs: Anastasia, Hugues, Qun, Pierre-Alexandre and Mostafa.

N'oublions pas les compères de l'époque mémoire: Gwen, Hugo et Patrick. Chacun à votre façon vous avez rendus ces quatre années plus amusantes. Gwen de ton pôle sud, et Patrick de Hambourg, vous avez tout de même fait partie intrinsèque de cette expérience pour moi. Hugo, avec ton année d'avance, tu m'as souvent bien aidé à m'orienter dans ce fatras que peut être CMS. Je vous souhaite à tous les trois de trouver votre chemin, comme j'essaye de trouver le mien.

Un immense merci à Audrey et Marleen d'avoir fait en sorte que tout semble toujours facile. Déplacements, écoles, inscriptions, elles font tenir la baraque comme on dit. Sans elles, rien n'est possible. Merci aussi aux IT, en particulier à Romain, Shkelzen et Olivier qui m'ont sauvé la vie un nombre incalculable de fois. Ils forment également une partie nécessaire de la vie d'un thésard en physique des particules.

The end of a thesis is the beginning of something else. A warm thanks to Catherine De Clercq and Juan Antonio Aguilar Sánchez for their help with my application to a postdoctoral position at the MIT as a member of the IceCube Collaboration.

Thank to all IIHE people who make this laboratory a wonderful place to work. See you all very soon!

Enfin merci à ma famille, mes parents et ma soeur de m'avoir supporté (je trouve ce mot plus juste que soutenu) durant ces quatre drôles d'années. Vous m'aimez, je crois, et c'est ce qui me rend invincible. Je vous aime au moins autant, soyez en certains. Je pars mais je reviens, vous savez bien que ça finit toujours comme ça.

Je finis en remerciant mon équipière, celle qui joue avec moi. Celle qui m'a rencontré au mauvais moment mais qui a quand même choisi de me suivre. De quel courage fait-elle preuve, celle qui me regarde et me sourit. Merci pour tout, merci pour nous, allons le plus loin possible sur le plus beau des chemins.

Bibliography

[1] L. de Broglie. *La physique nouvelle et les quanta*. Flammarion, 1937.

[2] D. I. Mendeleev. Über die Beziehungen der Eigenshaften zu den Atomgewichten der Elemente. *Zeitschrift für Chemie*, 12:405–406, 1869.

[3] J. J. Thomson. XL. Cathode Rays. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 44(269):293–316, 1897. doi: 10.1080/14786449708621070.

[4] J. J. Thomson. XXIV. On the structure of the atom: an investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle; with application of the results to the theory of atomic structure. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 7(39): 237–265, 1904. doi: 10.1080/14786440409463107.

[5] E. Rutherford. The scattering of alpha and beta particles by matter and the structure of the atom. *Phil. Mag. Ser.6*, 21:669–688, 1911. doi: 10.1080/14786440508637080.

[6] N. Bohr. I. On the constitution of atoms and molecules. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 26(151):1–25, 1913. doi: 10.1080/14786441308634955.

[7] A. Romer. Proton or prouton?: Rutherford and the depths of the atom. *American Journal of Physics*, 65(8):707–716, August 1997. ISSN 0002-9505 (print), 1943-2909 (electronic). doi: https://doi.org/10.1119/1.18640.

[8] J. Chadwick. Possible Existence of a Neutron. *Nature*, 129:312, 1932. doi: 10.1038/129312a0.

[9] M. Planck. Zur Theorie des Gesetzes der Energieverteilung im Normalspektrum. (German) [On the theory of distribution of energy in the normal spectrum]. *Verhandlungen der Deutschen Physikalischen Gesellschaft*, 2(17):237–245, December 1900.

[10] A. Einstein. Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik*, 322(6):132–148, 1905. doi: 10.1002/andp.19053220607.

[11] A. Einstein. Zur Elektrodynamik bewegter Körper. *Annalen der Physik*, 322(10):891–921, 1905. doi: 10.1002/andp.19053221004.

[12] W. Gordon. Der comptoneffekt nach der schrödingerschen theorie. *Zeitschrift für Physik*, 40(1):117–133, Jan 1926. doi: 10.1007/BF01390840.

[13] O. Klein. Quantentheorie und fünfdimensionale relativitätstheorie. *Zeitschrift für Physik*, 37(12):895–906, Dec 1926. doi: 10.1007/BF01397481.

[14] P. A. M. Dirac and R. H. Fowler. The quantum theory of the electron. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 117(778):610–624, 1928. doi: 10.1098/rspa.1928.0023.

[15] L. de Broglie. Recherches sur la théorie des quanta. *Annals Phys.*, 2:22–128, 1925.

[16] J. D. Cockcroft and E. T. S. Walton and E. Rutherford . Experiments with high velocity positive ions. II. -The disintegration of elements by high velocity protons. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 137(831):229–242, 1932. doi: 10.1098/rspa.1932.0133.

[17] C. Allardice, E. R. Trapnell, E. Fermi, and L. Fermi. 40th anniversary: The first reactor. Report DOE/NE-0046, U.S. Department of Energy, dec 1982.

[18] C. L. Cowan and F. Reines and F. B. Harrison and H. W. Kruse and A. D. McGuire. Detection of the Free Neutrino: a Confirmation. *Science*, 124(3212):103–104, 1956. ISSN 0036-8075. doi: 10.1126/science.124.3212.103. URL `https://science.sciencemag.org/content/124/3212/103`.

[19] C. Itzykson and J. B. Zuber. *Quantum Field Theory*. International Series In Pure and Applied Physics. McGraw-Hill, New York, 1980. ISBN 9780486445687, 0486445682.

[20] C. Cohen-Tannoudji and B. Diu and F. Laloë. *Mécanique quantique II*. 1973.

[21] S. L. Glashow. The renormalizability of vector meson interactions. *Nucl. Phys.*, 10:107–117, 1959. doi: 10.1016/0029-5582(59)90196-8.

[22] A. Salam and J. C. Ward. Electromagnetic and weak interactions. *Phys. Lett.*, 13:168–171, 1964. doi: 10.1016/0031-9163(64)90711-5.

[23] S. Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967. doi: 10.1103/PhysRevLett.19.1264.

[24] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, 1964. doi: 10.1103/PhysRevLett.13.321. [,157(1964)].

[25] P. W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13: 508–509, 1964. doi: 10.1103/PhysRevLett.13.508. [,160(1964)].

[26] S. Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30 – 61, 2012. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2012.08.021. URL `http://www.sciencedirect.com/science/article/pii/S0370269312008581`.

[27] G. Aad et al. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1 – 29, 2012. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2012.08.020. URL `http://www.sciencedirect.com/science/article/pii/S037026931200857X`.

[28] F. Wilczek. The Cosmic Asymmetry Between Matter And Antimatter. *Sci. Am.*, 243N6: 60–68, 1980. [Spektrum Wiss.1981,no.2,90(1981)].

[29] M. Tanabashi and others. Review of Particle Physics. *Phys. Rev. D*, 98:030001, Aug 2018. doi: 10.1103/PhysRevD.98.030001. URL `https://link.aps.org/doi/10.1103/PhysRevD.98.030001`.

[30] A. Kellerbauer et al. Proposed antimatter gravity measurement with an antihydrogen beam. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 266(3):351 – 356, 2008. ISSN 0168-583X. doi: https://doi.org/10.1016/j.nimb.2007.12.010. URL `http://www.sciencedirect.com/science/article/pii/S0168583X07017740`. Low Energy Positron and Positronium Physics.

[31] W. Bertsche et al. The ALPHA experiment: A cold antihydrogen trap. *AIP Conf. Proc.*, 796(1):301–308, 2005. doi: 10.1063/1.2130184.

[32] C. Malbrunot and others. The ASACUSA antihydrogen and hydrogen program: results and prospects. *Philosophical Transactions of the Royal Society of London Series A*, 376 (2116):20170273, Mar 2018, 1710.03288. doi: 10.1098/rsta.2017.0273.

[33] J. DiSciacca and others. One-Particle Measurement of the Antiproton Magnetic Moment. *Phys. Rev. Lett.*, 110:130801, Mar 2013. doi: 10.1103/PhysRevLett.110.130801. URL https://link.aps.org/doi/10.1103/PhysRevLett.110.130801.

[34] C. Smorra et al. BASE – The Baryon Antibaryon Symmetry Experiment. *Eur. Phys. J. ST*, 224(16):3055–3108, 2015, 1604.08820. doi: 10.1140/epjst/e2015-02607-4,10.1140/epjst/e2015-02336-2.

[35] A. Gando et al. Search for Majorana Neutrinos near the Inverted Mass Hierarchy Region with KamLAND-Zen. *Phys. Rev. Lett.*, 117(8):082503, 2016, 1605.02889. doi: 10.1103/PhysRevLett.117.109903,10.1103/PhysRevLett.117.082503. [Addendum: Phys. Rev. Lett.117,no.10,109903(2016)].

[36] M. Auger et al. Search for Neutrinoless Double-Beta Decay in ^{136}Xe with EXO-200. *Phys. Rev. Lett.*, 109:032505, 2012, 1205.5608. doi: 10.1103/PhysRevLett.109.032505.

[37] R. Davis. A review of the homestake solar neutrino experiment. *Progress in Particle and Nuclear Physics*, 32:13 – 32, 1994. ISSN 0146-6410. doi: https://doi.org/10.1016/0146-6410(94)90004-3. URL http://www.sciencedirect.com/science/article/pii/0146641094900043.

[38] Y. V. Kozlov and V. P. Martem'yanov and K. N. Mukhin. Neutrino mass problem: the state of the art. *Physics-Uspekhi*, 40(8):807–842, aug 1997. doi: 10.1070/pu1997v040n08abeh000273. URL https://doi.org/10.1070%2Fpu1997v040n08abeh000273.

[39] S. Mertens. Direct Neutrino Mass Experiments. *J. Phys. Conf. Ser.*, 718(2):022013, 2016, 1605.01579. doi: 10.1088/1742-6596/718/2/022013.

[40] K. G. Begeman and A. H. Broeils and R. H. Sanders. Extended rotation curves of spiral galaxies - Dark haloes and modified dynamics. *Monthly Notices of the RAS*, 249:523–537, Apr 1991. doi: 10.1093/mnras/249.3.523. URL http://adsabs.harvard.edu/abs/1991MNRAS.249..523B. Provided by the SAO/NASA Astrophysics Data System.

[41] P. A. R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.*, 594:A13, 2016, 1502.01589. doi: 10.1051/0004-6361/201525830.

[42] T. Marrodán Undagoitia and L. Rauch. Dark matter direct-detection experiments. *J. Phys.*, G43(1):013001, 2016, 1509.08767. doi: 10.1088/0954-3899/43/1/013001.

[43] E. Aprile et al. Dark Matter Search Results from a One Ton-Year Exposure of XENON1T. *Phys. Rev. Lett.*, 121(11):111302, 2018, 1805.12562. doi: 10.1103/PhysRevLett.121.111302.

[44] D. S. Akerib et al. Results from a search for dark matter in the complete LUX exposure. *Phys. Rev. Lett.*, 118(2):021303, 2017, 1608.07648. doi: 10.1103/PhysRevLett.118.021303.

[45] X. Ren and others. Constraining Dark Matter Models with a Light Mediator at the PandaX-II Experiment. *Phys. Rev. Lett.*, 121:021304, Jul 2018. doi: 10.1103/PhysRevLett.121.021304. URL https://link.aps.org/doi/10.1103/PhysRevLett.121.021304.

[46] C. Karwin et al. Dark Matter Interpretation of the Fermi-LAT Observation Toward the Galactic Center. *Phys. Rev.*, D95(10):103005, 2017, 1612.05687. doi: 10.1103/PhysRevD.95.103005.

[47] M. G. Aartsen et al. Search for Neutrinos from Dark Matter Self-Annihilations in the center of the Milky Way with 3 years of IceCube/DeepCore. *Eur. Phys. J.*, C77(9):627, 2017, 1705.08103. doi: 10.1140/epjc/s10052-017-5213-y.

[48] M. Boezio and others. PAMELA and indirect dark matter searches. *New Journal of Physics*, 11(10):105023, oct 2009. doi: 10.1088/1367-2630/11/10/105023. URL `https://doi.org/10.1088%2F1367-2630%2F11%2F10%2F105023`.

[49] J. M. Gaskins. A review of indirect searches for particle dark matter. *Contemp. Phys.*, 57 (4):496–525, 2016, 1604.00014. doi: 10.1080/00107514.2016.1175160.

[50] R. Jora. About electrodynamics, standard model and the quantization of the electrical charge. *Int. J. Mod. Phys.*, A33(33):1850205, 2018, 1805.09526. doi: 10.1142/S0217751X18502056.

[51] Y. Gouverneur. Anomaly constraints and number of families. In *Proceedings, 37th Rencontres de Moriond on Electroweak Interactions and Unified Theories: Les Arcs, France, March 9-16, 2002*, pages 399–404, 2002, hep-ph/0205137.

[52] M. E. Peskin and D. V. Schroeder. *An introduction to quantum field theory*. Westview, 1995.

[53] B. Holdom. Two U(1)'s and Epsilon Charge Shifts. *Phys. Lett.*, 166B:196–198, 1986. doi: 10.1016/0370-2693(86)91377-8.

[54] CERN. HL-LHC Project Schedule. URL `https://project-hl-lhc-industry.web.cern.ch/`.

[55] O. Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004. URL `https://cds.cern.ch/record/782076`.

[56] L. Evans and P. Bryant. LHC Machine. *JINST*, 3:S08001, 2008. doi: 10.1088/1748-0221/3/08/S08001.

[57] S. Jolly. *An Intra-pulse Fast Feedback System for a Future Linear Collider*. PhD thesis, Exeter College, Oxford, 2003.

[58] P. Barnes and J. K. Cockcroft and S. Jacques and M. Vickers. How do Synchrotrons Work? URL `http://pd.chem.ucl.ac.uk/pdnn/inst2/work.htm`. Advanced Certificate in Powder Diffraction on the Web, Birkbeck College, University of London.

[59] N. Ryckx and L. Rivkin and S. Redaelli. *Combined energy ramp and beta star squeeze at the Large Hadron Collider*. PhD thesis, 01 2012.

[60] G. Antchev and others. Luminosity-Independent Measurement of the Proton-Proton Total Cross Section at $\sqrt{s} = 8$ TeV. *Phys. Rev. Lett.*, 111:012001, Jul 2013. doi: {10.1103/PhysRevLett.111.012001}. URL `https://link.aps.org/doi/10.1103/PhysRevLett.111.012001`.

[61] G. I. Veres. Soft QCD at the CMS And ATLAS Experiments. In *Proceedings, 51st Rencontres de Moriond on QCD and High Energy Interactions: La Thuile, Italy, March 19-26, 2016*, pages 143–146. ARISF, ARISF, 2016.

[62] M. H. F. Van De Klundert. Soft QCD at CMS and ATLAS. Technical Report CMS-CR-2017-121, CERN, Geneva, Apr 2017. URL `https://cds.cern.ch/record/2263388`.

[63] A. Holzner. 78 reconstructed vertices in event from high-pileup run 198609. CMS Collection., Sep 2012. URL `http://cds.cern.ch/record/1479324`.

[64] The CMS Collaboration. CMS Luminosity - Public Results, 2019. URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[65] R. Bruce et al. LHC Run 2: Results and challenges. In *Proceedings, 57th ICFA Advanced Beam Dynamics Workshop on High-Intensity and High-Brightness Hadron Beams (HB2016): Malmö, Sweden, July 3-8, 2016*, page MOAM5P50, 2016. doi: 10.18429/JACoW-HB2016-MOAM5P50.

[66] M. Lamont. Performance 2016, July 2016. URL https://lhc-commissioning.web.cern.ch/lhc-commissioning/performance/2016-performance.htm.

[67] R. Barate et al. Search for the standard model Higgs boson at LEP. *Phys. Lett.*, B565: 61–75, 2003, hep-ex/0306033. doi: 10.1016/S0370-2693(03)00614-2.

[68] J. Kuti, L. Lin, and Y. Shen. Upper bound on the higgs-boson mass in the standard model. *Phys. Rev. Lett.*, 61:678–681, Aug 1988. doi: 10.1103/PhysRevLett.61.678. URL https://link.aps.org/doi/10.1103/PhysRevLett.61.678.

[69] M. Spira. Higgs boson production at the LHC. In *The Higgs puzzle - what can we learn from LEP-2, LHC, NLC and FMC? Proceedings, Ringberg Workshop, Tegernsee, Germany, December 8-13, 1996*, pages 122–130, 1997, hep-ph/9703355.

[70] The CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004–S08004, aug 2008. doi: 10.1088/1748-0221/3/08/s08004. URL https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08004.

[71] L. Taylor. CMS detector design, November 2011. URL http://cms.web.cern.ch/news/cms-detector-design.

[72] N. Pacifico and M. De Palma. *Radiation damage study on innovative silicon sensors for the CMS tracker upgrade.* PhD thesis, Dec 2011. URL http://cds.cern.ch/record/2130663. Presented 07 May 2012.

[73] S. Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014, 1405.6569. doi: 10.1088/1748-0221/9/10/P10009.

[74] The CMS Collaboration. CMS Tracker Detector Performance Results, 2019. URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK.

[75] A. Dominguez et al. CMS Technical Design Report for the Pixel Detector Upgrade. Technical Report CERN-LHCC-2012-016. CMS-TDR-11, Sep 2012.

[76] M. Axer. *Development of a Test System for the Quality Assurance of Silicon Microstrip Detectors for the Inner Tracking System of the CMS Experiment.* PhD thesis, Aachen, Tech. Hochsch., 2003.

[77] M. Jansová. *Recherche du partenaire supersymétrique du quark top et mesure des propriétés des dpôts dans le trajectographe à pistes de silicium de l'expérience CMS au Run 2.* PhD thesis, Université de Strasbourg, September 2018.

[78] F. Hartmann. Evolution of Silicon Sensor Technology in Particle Physics. *Springer Tracts Mod. Phys.*, 275:pp.1–372, 2017. doi: 10.1007/978-3-319-64436-3.

[79] R. Wunstorf, W.M. Bugg, J. Walter, F.W. Garber, and D. Larson. Investigations of donor and acceptor removal and long term annealing in silicon with different boron/phosphorus ratios. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 377(2):228 – 233, 1996. ISSN 0168-9002. doi: https://doi.org/10.1016/0168-9002(96)00217-3. URL http://www.sciencedirect.

com/science/article/pii/0168900296002173. Proceedings of the Seventh European Symposium on Semiconductor.

[80] G. L. Bayatian et al. CMS Physics. 2006.

[81] A. Annenkov et al. Radiation damage Kinetics in PWO crystals. Technical Report CMS-NOTE-1997-008, CERN, Geneva, Feb 1997. URL http://cds.cern.ch/record/687197.

[82] S. Lu. Comparison of hadron shower data with simulations. In *International Linear Collider Workshop (LCWS10 and ILC10) Beijing, China, March 26-30, 2010*, 2010, 1007.2947.

[83] T. A. Gabriel et al. Energy dependence of hadronic activity. *Nucl. Instrum. Methods Phys. Res., A*, 338(ANL-HEP-PR-93-69. LBL-34589):336–347. 25 p, Sep 1993. URL http://cds.cern.ch/record/254124.

[84] A. M. Sirunyan et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(CMS-MUO-16-001. CMS-MUO-16-001-003. 06):P06015. 53 p, Apr 2018. doi: 10.1088/1748-0221/13/06/P06015. URL https://cds.cern.ch/record/2313130.

[85] S. R. Davis. Interactive Slice of the CMS detector. Aug 2016. URL http://cds.cern.ch/record/2205172.

[86] A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017, 1706.04965. doi: 10.1088/1748-0221/12/10/P10003.

[87] R. Frühwirth. Application of kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2):444 – 450, 1987. ISSN 0168-9002. doi: https://doi.org/10.1016/0168-9002(87)90887-4.

[88] S. Chatrchyan et al. The Performance of the CMS Muon Detector in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV at the LHC. *JINST*, 8:P11002, 2013, 1306.6905. doi: 10.1088/1748-0221/8/11/P11002.

[89] A. M. Sirunyan et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(06):P06015, 2018, 1804.04528. doi: 10.1088/1748-0221/13/06/P06015.

[90] R. Fruhwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth.*, A262:444–450, 1987. doi: 10.1016/0168-9002(87)90887-4.

[91] V. Khachatryan et al. Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV. *JINST*, 10(06):P06005, 2015, 1502.02701. doi: 10.1088/1748-0221/10/06/P06005.

[92] W. Adam et al. Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC. 2005.

[93] V. Khachatryan et al. Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV. *JINST*, 10(08):P08010, 2015, 1502.02702. doi: 10.1088/1748-0221/10/08/P08010.

[94] A. M. Sirunyan et al. Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(10):P10005, 2018, 1809.02816. doi: 10.1088/1748-0221/13/10/P10005.

[95] P. Loch. Introduction To Hadronic Final State Reconstruction in Collider Experiment (Part V), March 2010. ATLAS Collaboration, The University of Arizona.

[96] M. Cacciari, G. P. Salam, and G. Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04: 063, 2008, 0802.1189. doi: 10.1088/1126-6708/2008/04/063.

[97] S. D. Ellis and D. E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev.*, D48:3160–3166, 1993, hep-ph/9305266. doi: 10.1103/PhysRevD.48.3160.

[98] Jet algorithms performance in 13 TeV data. Technical Report CMS-PAS-JME-16-003, CERN, Geneva, 2017. URL `http://cds.cern.ch/record/2256875`.

[99] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, Geneva, Apr 2009. URL `http://cds.cern.ch/record/1194487`.

[100] T. Plehn. Lectures on LHC Physics. *Lect. Notes Phys.*, 844:1–193, 2012, 0910.4182. doi: 10.1007/978-3-642-24040-9.

[101] S. Cittolin, A. Rácz, and P. Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project.* Technical Design Report CMS. CERN, Geneva, 2002. URL `http://cds.cern.ch/record/578006`.

[102] A. M. Sirunyan et al. Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector. *Submitted to: JINST*, 2019, 1903.06078.

[103] S. Redaelli. Beam Cleaning and Collimation Systems. *CERN Yellow Reports*, 2(0):403, 2016. URL `https://e-publishing.cern.ch/index.php/CYR/article/view/243`.

[104] W. Bialas and D. A. Petyt. Mitigation of anomalous APD signals in the CMS ECAL. *Journal of Instrumentation*, 8(03):C03020, 2013. URL `http://stacks.iop.org/1748-0221/8/i=03/a=C03020`.

[105] A. Apresyan on behalf of the CMS Collaboration). Identification and Mitigation of Anomalous Signals in CMS Hadronic Calorimeter. *Journal of Physics: Conference Series*, 404(1): 012044, 2012. URL `http://stacks.iop.org/1742-6596/404/i=1/a=012044`.

[106] C.-E. Wulz and CMS Collaboration. Measurement technology for the CMS experiment. *Measurement Science and Technology*, 18(8):2424–2431, jul 2007. doi: 10.1088/0957-0233/18/8/018. URL `https://doi.org/10.1088%2F0957-0233%2F18%2F8%2F018`.

[107] J. Abdallah et al. Simplified Models for Dark Matter Searches at the LHC. *Phys. Dark Univ.*, 9-10:8–23, 2015, 1506.03116. doi: 10.1016/j.dark.2015.08.001.

[108] D. Abercrombie et al. Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum. 2015, 1507.00966.

[109] B. Grinstein. Minimal flavor violation. In *4th International Workshop on the CKM Unitarity Triangle (CKM 2006) Nagoya, Japan, December 12-16, 2006*, 2007, 0706.4185.

[110] A. M. Sirunyan and others. Search for new physics in final states with an energetic jet or a hadronically decaying W or Z boson and transverse momentum imbalance at $\sqrt{s} = 13$ TeV. *Phys. Rev. D*, 97:092005, May 2018. doi: 10.1103/PhysRevD.97.092005. URL `https://link.aps.org/doi/10.1103/PhysRevD.97.092005`.

[111] The CMS collaboration. Identification of b-quark jets with the CMS experiment. *Journal of Instrumentation*, 8(04):P04013–P04013, apr 2013. doi: 10.1088/1748-0221/8/04/p04013. URL `https://doi.org/10.1088%2F1748-0221%2F8%2F04%2Fp04013`.

[112] J. Thaler and K. Van Tilburg. Identifying boosted objects with n-subjettiness. *Journal of High Energy Physics*, 2011(3):15, Mar 2011. ISSN 1029-8479. doi: 10.1007/JHEP03(2011)015. URL https://doi.org/10.1007/JHEP03(2011)015.

[113] E. W. Kolb. Cosmology and the Unexpected. *Subnucl. Ser.*, 45:337–363, 2009, 0709.3102. doi: 10.1142/9789814293242_0020.

[114] J. Liu, X. Chen, and X. Ji. Current status of direct dark matter detection experiments. *Nature Phys.*, 13(3):212–216, 2017, 1709.00688. doi: 10.1038/nphys4039.

[115] G. Busoni et al. Recommendations on presenting LHC searches for missing transverse energy signals using simplified s-channel models of dark matter. 2016, 1603.04156.

[116] A. M. Sirunyan et al. Search for invisible decays of a Higgs boson produced through vector boson fusion in proton-proton collisions at $\sqrt{s} = 13$ TeV. 2018, 1809.05937. doi: 10.1016/j.physletb.2019.04.025.

[117] K. Arnold and others. VBFNLO: A parton level Monte Carlo for processes with electroweak bosons. 2008, 0811.4559.

[118] J. Baglio and others. VBFNLO: A parton level Monte Carlo for processes with electroweak bosons – Manual for version 2.7.0. 2011, 1107.4038.

[119] J. Baglio and others. Release note - VBFNLO 2.7.0. 2014, 1404.3940.

[120] S. A. Abel et al. Kinetic Mixing of the Photon with Hidden U(1)s in String Phenomenology. *JHEP*, 07:124, 2008, 0803.1449. doi: 10.1088/1126-6708/2008/07/124.

[121] P. Langacker. The Physics of Heavy Z' Gauge Bosons. *Rev. Mod. Phys.*, 81:1199–1228, 2009. doi: 10.1103/RevModPhys.81.1199.

[122] P. Langacker. The Physics of New U(1)-prime Gauge Bosons. *AIP Conf. Proc.*, 1200(1): 55–63, 2010. doi: 10.1063/1.3327671.

[123] D. E. Morrissey, D. Poland, and K. M. Zurek. Abelian Hidden Sectors at a GeV. *JHEP*, 07:050, 2009, 0904.2567. doi: 10.1088/1126-6708/2009/07/050.

[124] R. Essig et al. Working Group Report: New Light Weakly Coupled Particles. In *Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013*, 2013, 1311.0029. URL http://www.slac.stanford.edu/econf/C1307292/docs/IntensityFrontier/NewLight-17.pdf.

[125] N. Vinyoles and H. Vogel. Minicharged Particles from the Sun: A Cutting-Edge Bound. *JCAP*, 1603(03):002, 2016, 1511.01122. doi: 10.1088/1475-7516/2016/03/002.

[126] Sacha Davidson, Steen Hannestad, and Georg Raffelt. Updated bounds on millicharged particles. *JHEP*, 05:003, 2000, hep-ph/0001179. doi: 10.1088/1126-6708/2000/05/003.

[127] D. N. Spergel et al. First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters. *Astrophys. J. Suppl.*, 148:175–194, 2003, astro-ph/0302209. doi: 10.1086/377226.

[128] S. Davidson, B. Campbell, and D. C. Bailey. Limits on particles of small electric charge. *Phys. Rev.*, D43:2314–2321, 1991. doi: 10.1103/PhysRevD.43.2314.

[129] P.-K. Hui, A. Kusenko, and V. Takhistov. Dark Cosmic Rays. *Phys. Lett.*, B768:18–22, 2017, 1611.04599. doi: 10.1016/j.physletb.2017.02.035.

[130] A. Haas, C. S. Hill, E. Izaguirre, and I. Yavin. Looking for milli-charged particles with a new experiment at the LHC. *Phys. Lett.*, B746:117–120, 2015, 1410.6816. doi: 10.1016/j.physletb.2015.04.062.

[131] S. Alekhin et al. A facility to Search for Hidden Particles at the CERN SPS: the SHiP physics case. *Rept. Prog. Phys.*, 79(12):124201, 2016, 1504.04855. doi: 10.1088/0034-4885/79/12/124201.

[132] S. Chatrchyan et al. Search for fractionally charged particles in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev.*, D87(9):092008, 2013, 1210.2311. doi: 10.1103/PhysRevD.87.092008.

[133] S. Chatrchyan et al. Searches for Long-Lived Charged Particles in pp Collisions at $\sqrt{s}$=7 and 8 TeV. *JHEP*, 07:122, 2013, 1305.0491. doi: 10.1007/JHEP07(2013)122.

[134] T. Sjostrand, S. Mrenna, and P. Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006, hep-ph/0603175. doi: 10.1088/1126-6708/2006/05/026.

[135] T. Sjostrand, S. Mrenna, and P. Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008, 0710.3820. doi: 10.1016/j.cpc.2008.01.036.

[136] M. L. Perl, E. R. Lee, and D. Loomba. Searches for fractionally charged particles. *Ann. Rev. Nucl. Part. Sci.*, 59:47–65, 2009. doi: 10.1146/annurev-nucl-121908-122035.

[137] W. Weibull et al. A statistical distribution function of wide applicability. *Journal of applied mechanics*, 18(3):293–297, 1951.

[138] The CMS Collaboration. Reference muon id, isolation and trigger efficiencies for 2016 legacy re-reco data, 2019. URL `https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffs2016LegacyRereco`.

[139] The CMS Collaboration. Reference muon id, isolation and trigger efficiencies for 2017 data, 2019. URL `https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffs2017`.

[140] The CMS Collaboration. Reference muon id, isolation and trigger efficiencies for 2018 legacy re-reco data, 2019. URL `https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffs2018`.

[141] CMS Luminosity Measurements for the 2016 Data Taking Period. Technical Report CMS-PAS-LUM-17-001, CERN, Geneva, 2017. URL `https://cds.cern.ch/record/2257069`.

[142] CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-LUM-17-004, CERN, Geneva, 2018. URL `https://cds.cern.ch/record/2621960`.

[143] CMS Luminosity Measurements for the 2018 Data Taking Period at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-LUM-18-002, CERN, Geneva, 2018.

[144] T. Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth.*, A434:435–443, 1999, hep-ex/9902006. doi: 10.1016/S0168-9002(99)00498-2.

[145] A. L. Read. Presentation of search results: the CLs technique. *Journal of Physics G: Nuclear and Particle Physics*, 28(10):2693–2704, sep 2002. doi: 10.1088/0954-3899/28/10/313. URL `https://doi.org/10.1088%2F0954-3899%2F28%2F10%2F313`.

[146] E. Gross G. Cowan, K. Cranmer and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J.*, C71:1554, 2011, 1007.1727. doi: 10.1140/epjc/s10052-011-1554-0,10.1140/epjc/s10052-013-2501-z. [Erratum: Eur. Phys. J.C73,2501(2013)].

[147] E. Ramirez Barreto and D. Romero Abad. Heavy long-lived fractionally charged leptons in novel $3-3-1$ model. 2019, 1907.02613.

Appendices

Appendix A

Validation of signal simulation

We validate our signal simulation against central production by comparing significant distributions. Figure A.1 shows the p_T spectrum in events with at least one FCP in $|\eta|$ <2.6 at generated level and corresponding distribution for the matching track. The spectra are compared for Q = 1 and Q = 2/3. In figure A.2, the dE/dx spectrum for the same events is compared for the same charge points. From these histograms, one can see that our simulation matches the central CMS production very well.

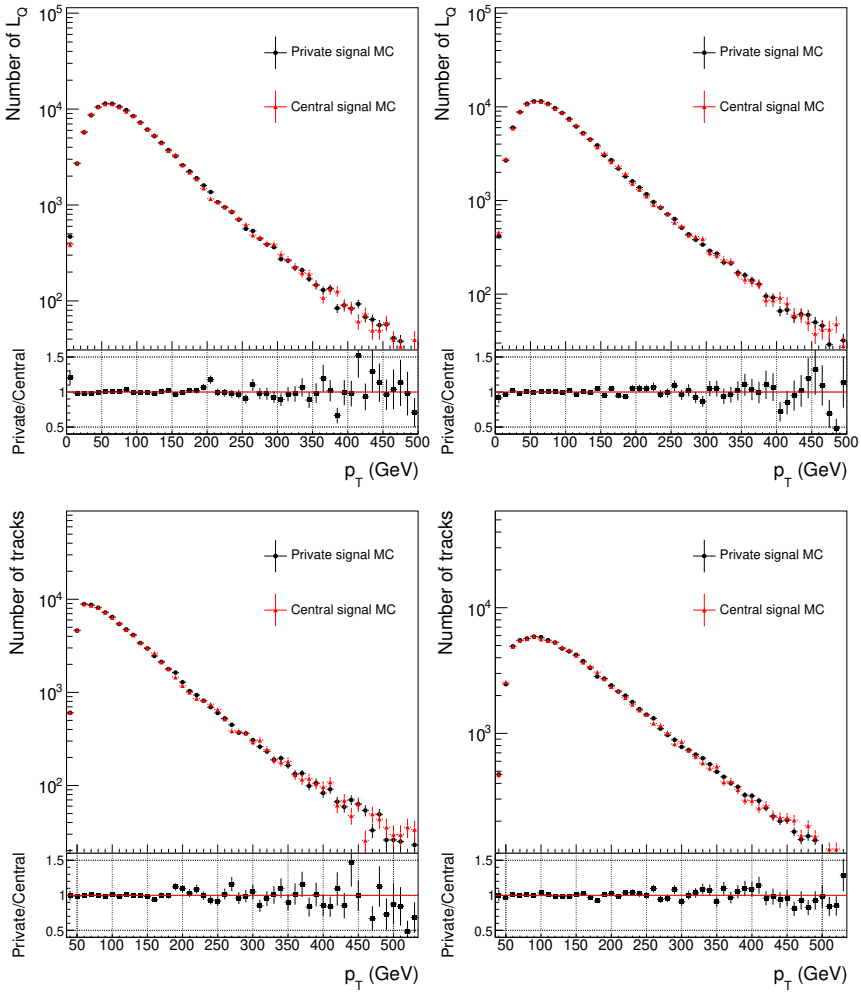


Figure A.1: Comparison of the FCP p_T spectrum (top) and the reconstructed track p_T spectrum (bottom) for our private signal simulation and central CMS production. Left plots are for Q = 1, right plots for Q = 2/3.

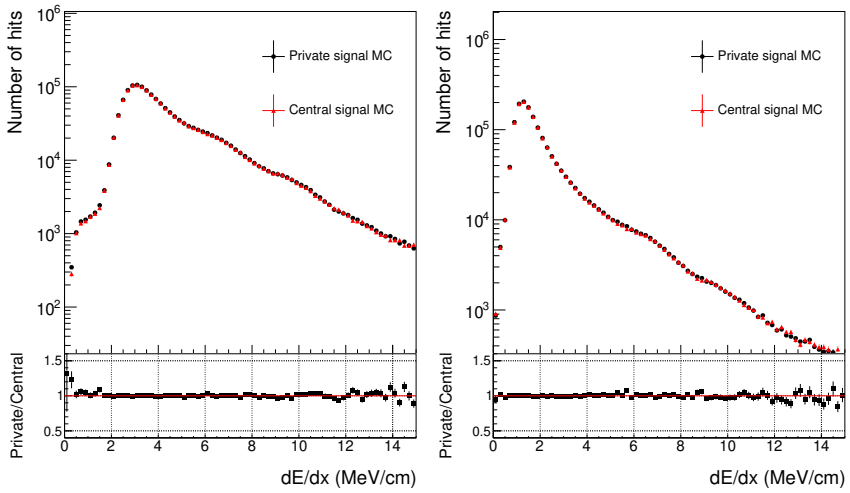


Figure A.2: Comparison of the FCP hits dE/dx spectrum for our private signal simulation and central CMS production. Left plots are for Q = 1, right plots for Q = 2/3.

Appendix B

Trigger inefficency

The trigger efficiency curve show an inefficiency at low p_T for high mass FCP. This is observed to vanish when considering a light (M = 1 GeV) FCP as seen in figure B.1, leading us to the conclusion that this is related to the mass, i.e. the low velocity of the particle.

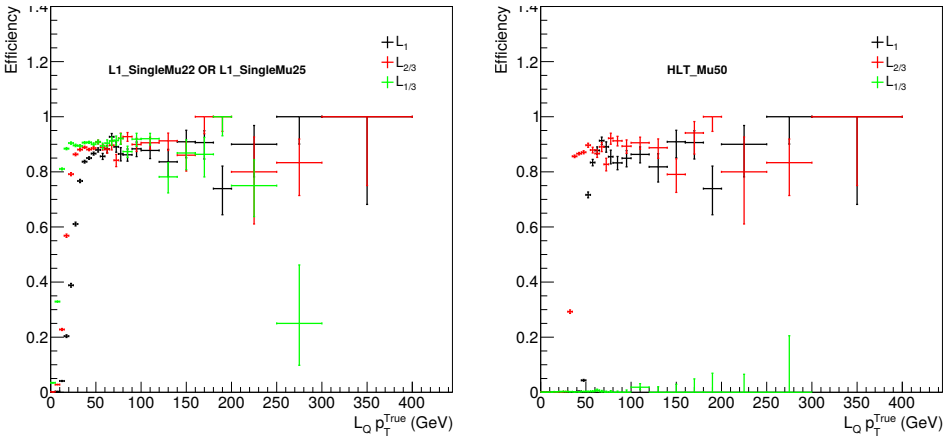


Figure B.1: L1 (left) seed and HLT_Mu50 trigger efficiency for a FCP of M = 1 GeV. The inefficiency observed at large masses is not present.

Appendix C

Local coordinates for the TOB and the TEC

We show here the X and Y local coordinates for the TOB (figure C.1) and ring 5 to 7 of the TEC (figures C.2 to C.4).

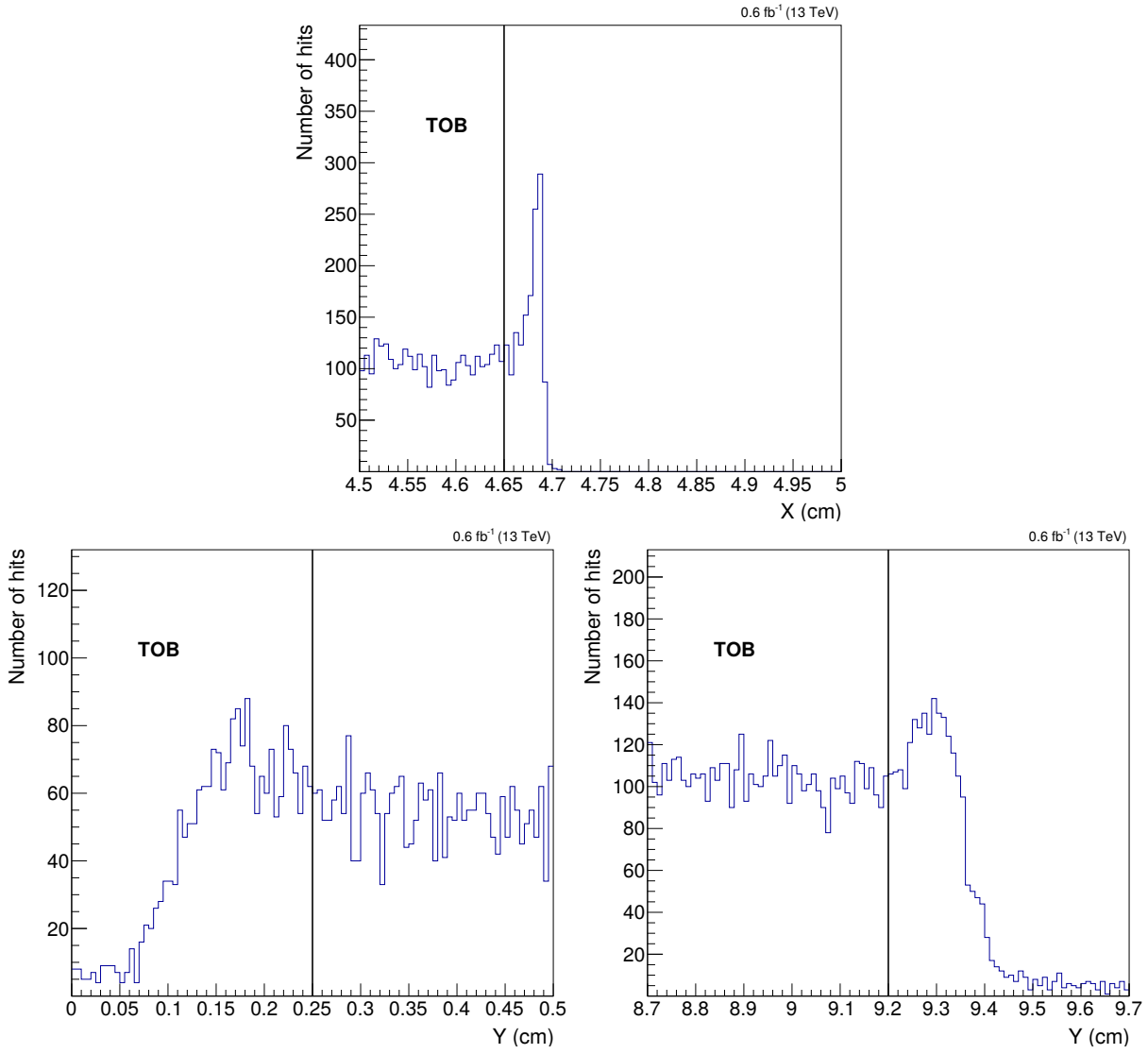


Figure C.1: Absolute value of the local X (top) and Y (bottom) coordinate of hits with dE/dx < 2 MeV/cm in the Tracker Outer Barrel. Only 10 % of 2016 data (eras B to F) is used. An accumulation of hits is observed at the edges which is removed by cutting where the line is placed.

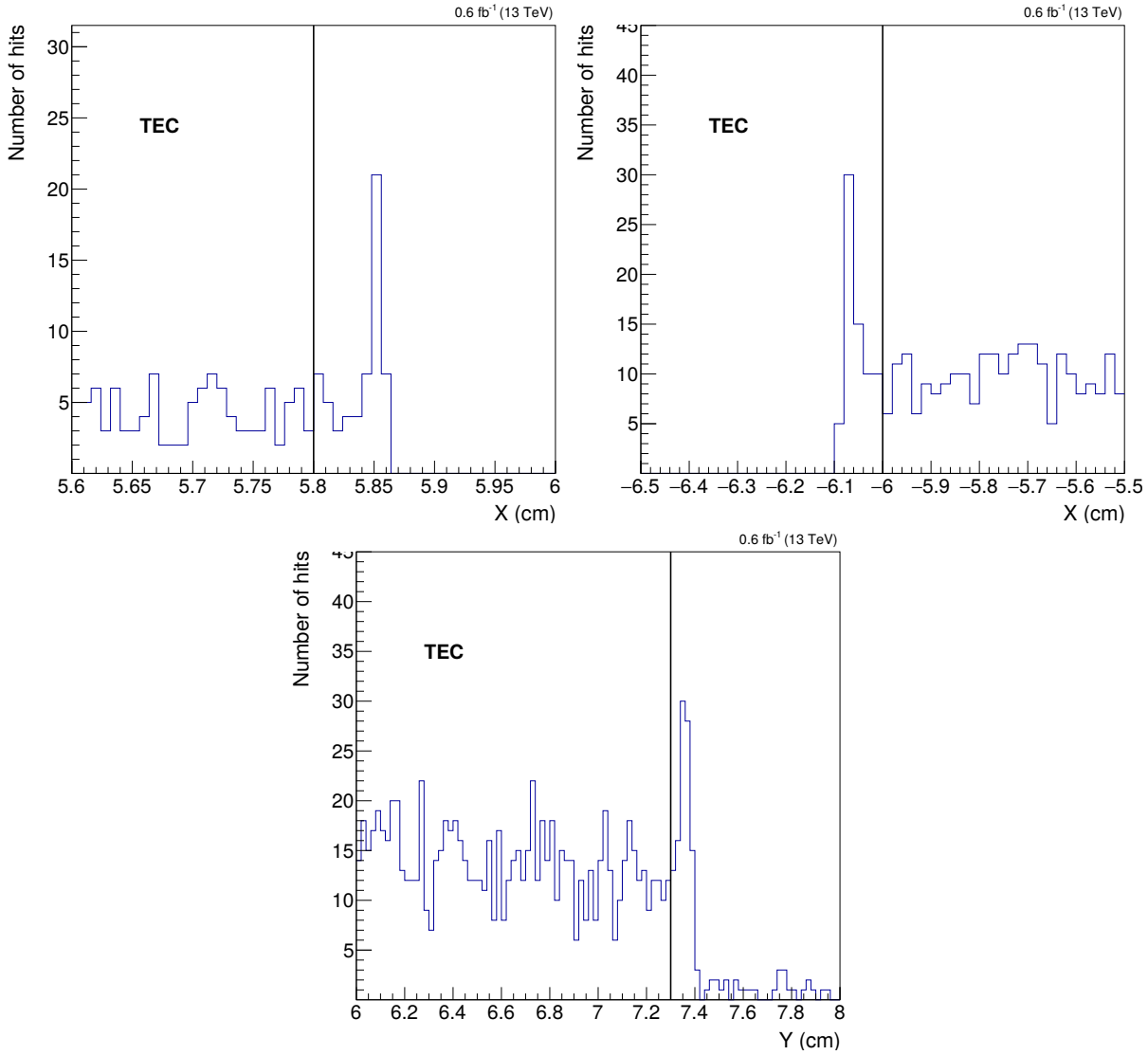


Figure C.2: Local X coordinate (top) and absolute value of the local Y coordinate (bottom) of hits with dE/dx $<$ 2 MeV/cm in the Tracker Endcaps ring 5. The transformations of the X coordinate being not perfect for this ring, they lead to slightly different values in the negative and positive sides of X which is why we do not sum show them separetely. Only 10 % of 2016 data (eras B to F) is used. An accumulation of hits is observed at the edges which is removed by cutting where the line is placed.

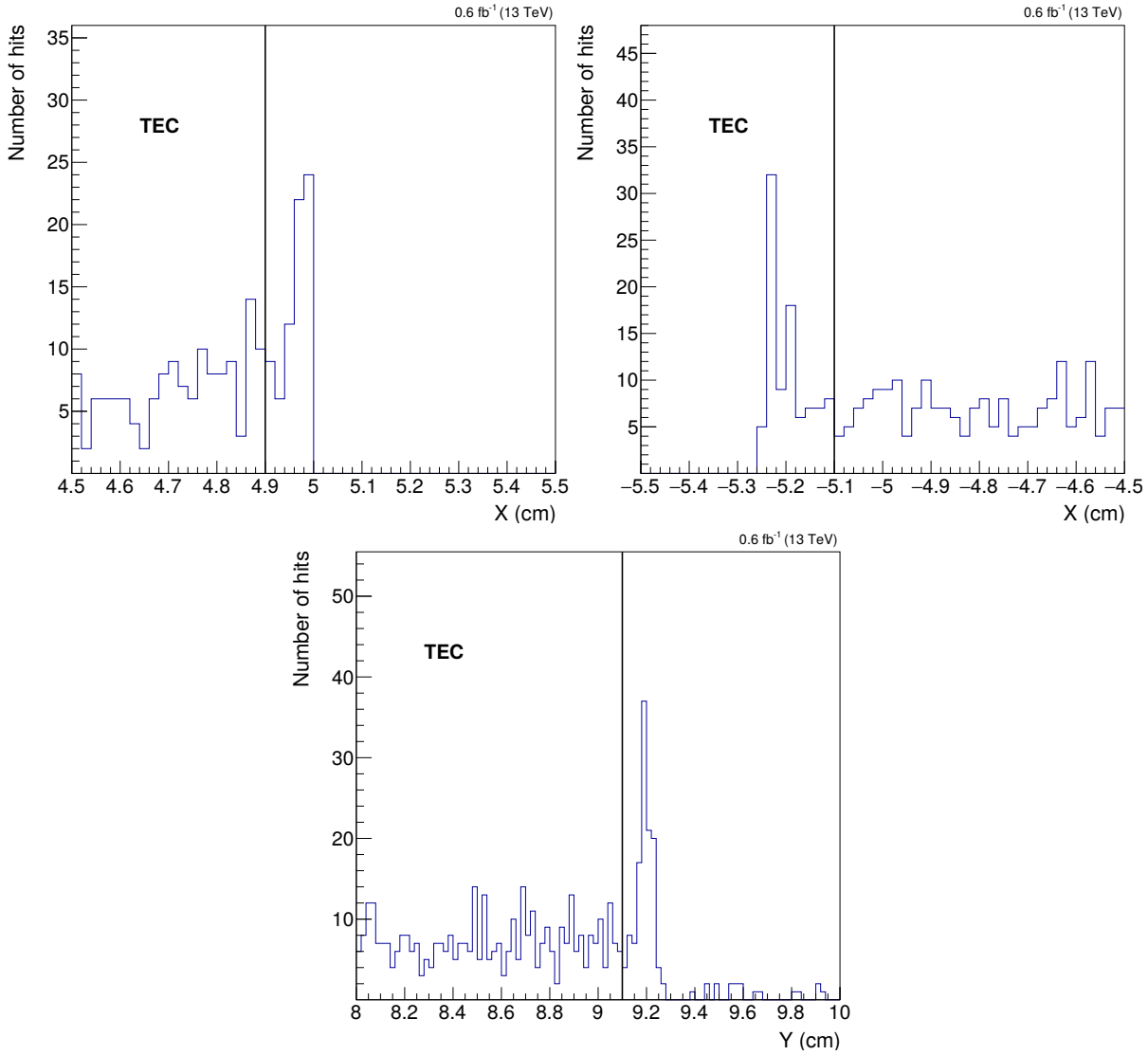


Figure C.3: Local X coordinate (top) and absolute value of the local Y (bottom) coordinate of hits with dE/dx $<$ 2 MeV/cm in the Tracker Endcaps ring 6. The transformations of the X coordinate being not perfect for this ring, they lead to slightly different values in the negative and positive sides of X which is why we do not sum show them separetely. Only 10 % of 2016 data (eras B to F) is used. An accumulation of hits is observed at the edges which is removed by cutting where the line is placed.

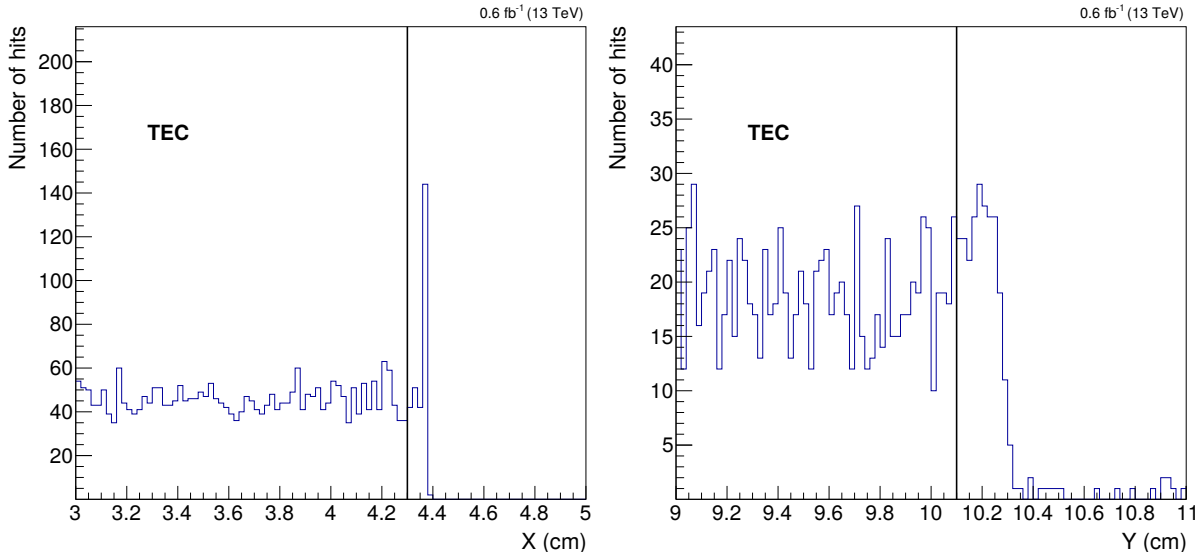


Figure C.4: Absolute value of the local X (left) and Y (right) coordinates of hits with dE/dx $<$ 2 MeV/cm in the Tracker Endcaps ring 7. Only 10 % of 2016 data (eras B to F) is used. An accumulation of hits is observed at the edges which is removed by cutting where the line is placed.

Appendix D

Signal corrections

We show the effect of the corrections derived in BG MC to be applied to the signal hits dE/dx spectra. The dE/dx spectra are given before and after correction in early 2016 (figures D.1 to D.4), late 2016 (figures D.5 to D.8), 2017 (figures D.9 to D.12) and 2018 (figures D.13 to D.16).

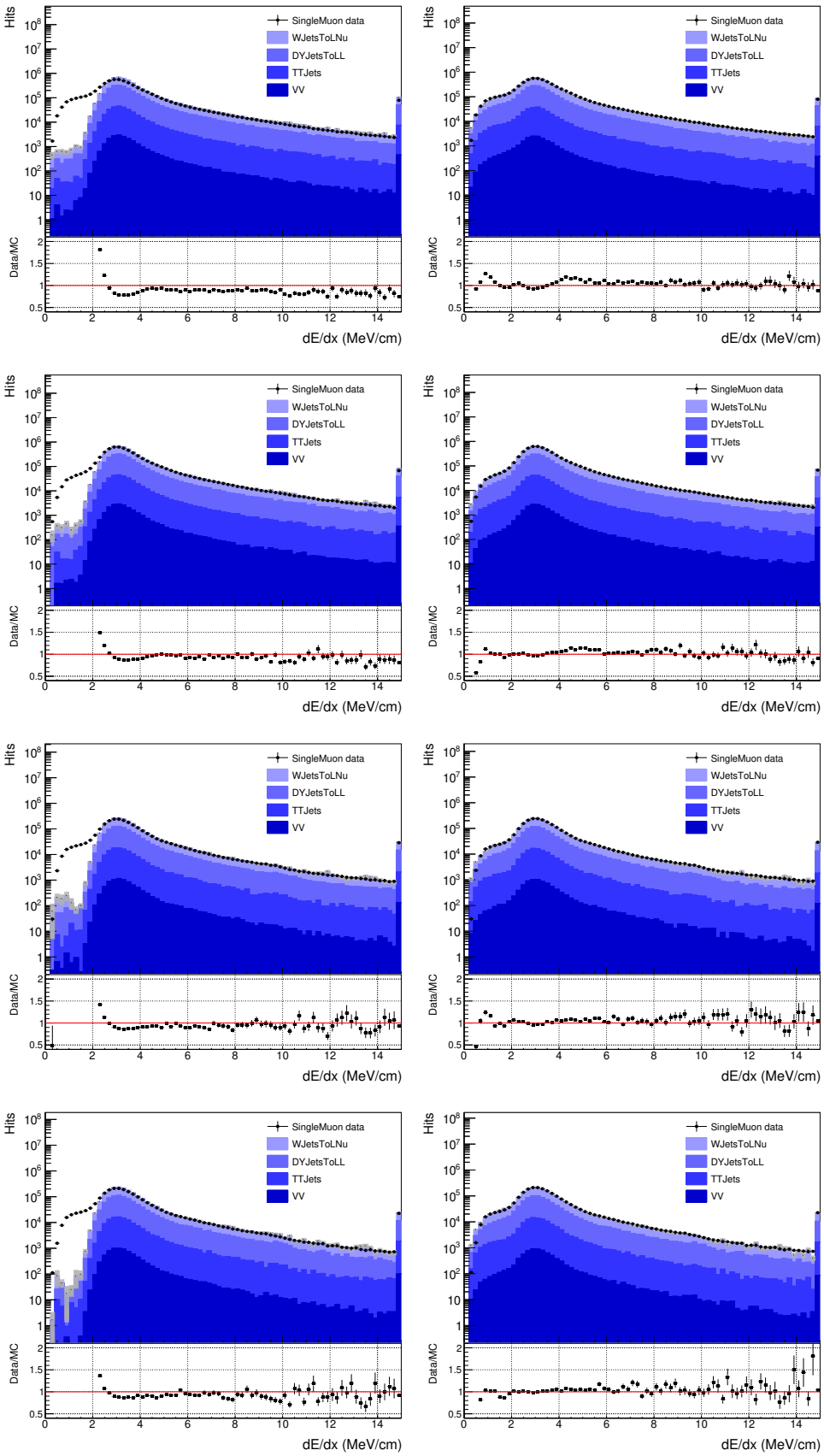


Figure D.1: dE/dx spectrum before (left) and after (right) corrections in the TIB layer 1 (top) to 4 (bottom) in early 2016 data (19.7 fb^{-1}).

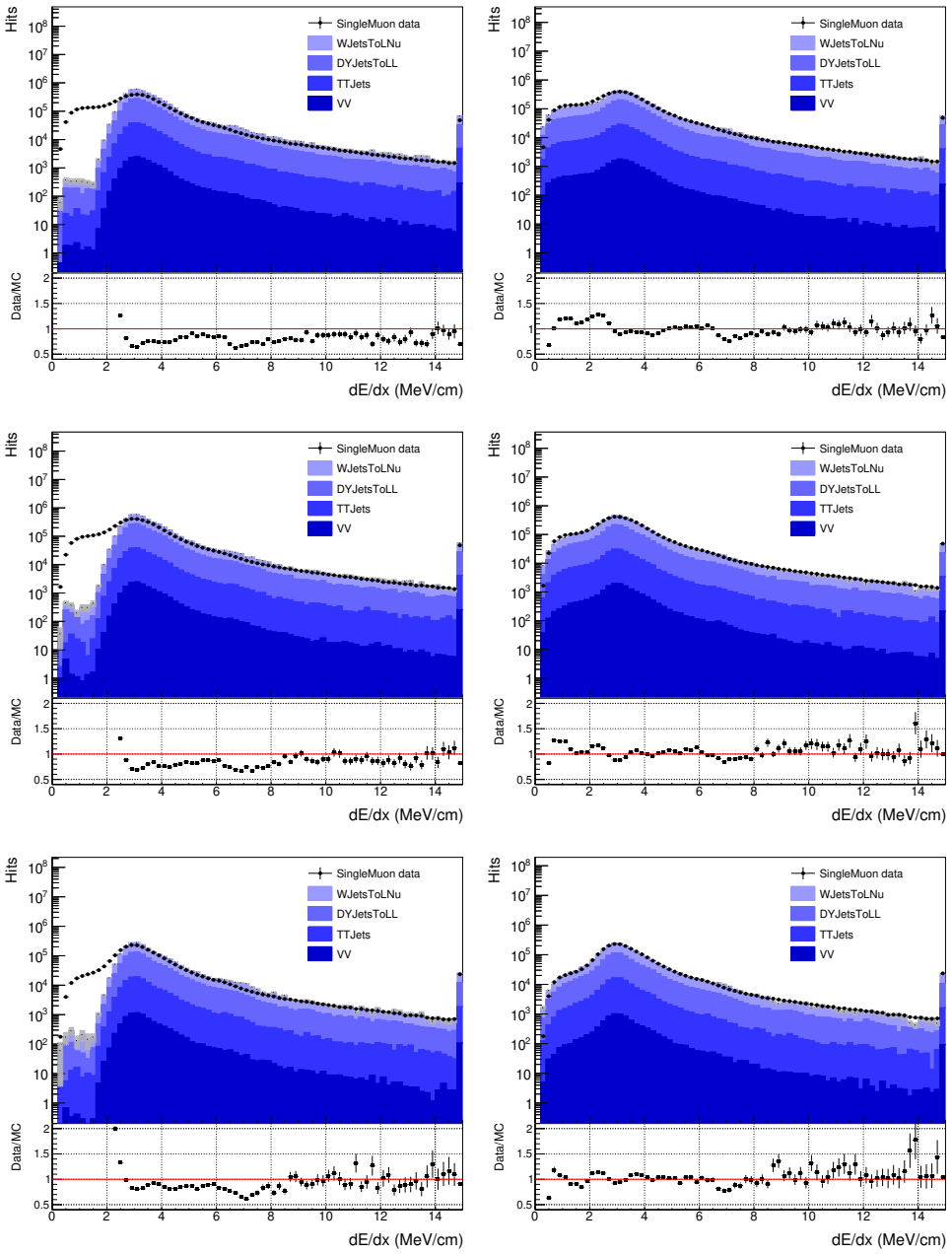


Figure D.2: dE/dx spectrum before (left) and after (right) corrections in the TOB layer 1 (top) to 3 (bottom) in early 2016 data (19.7 fb^{-1}).

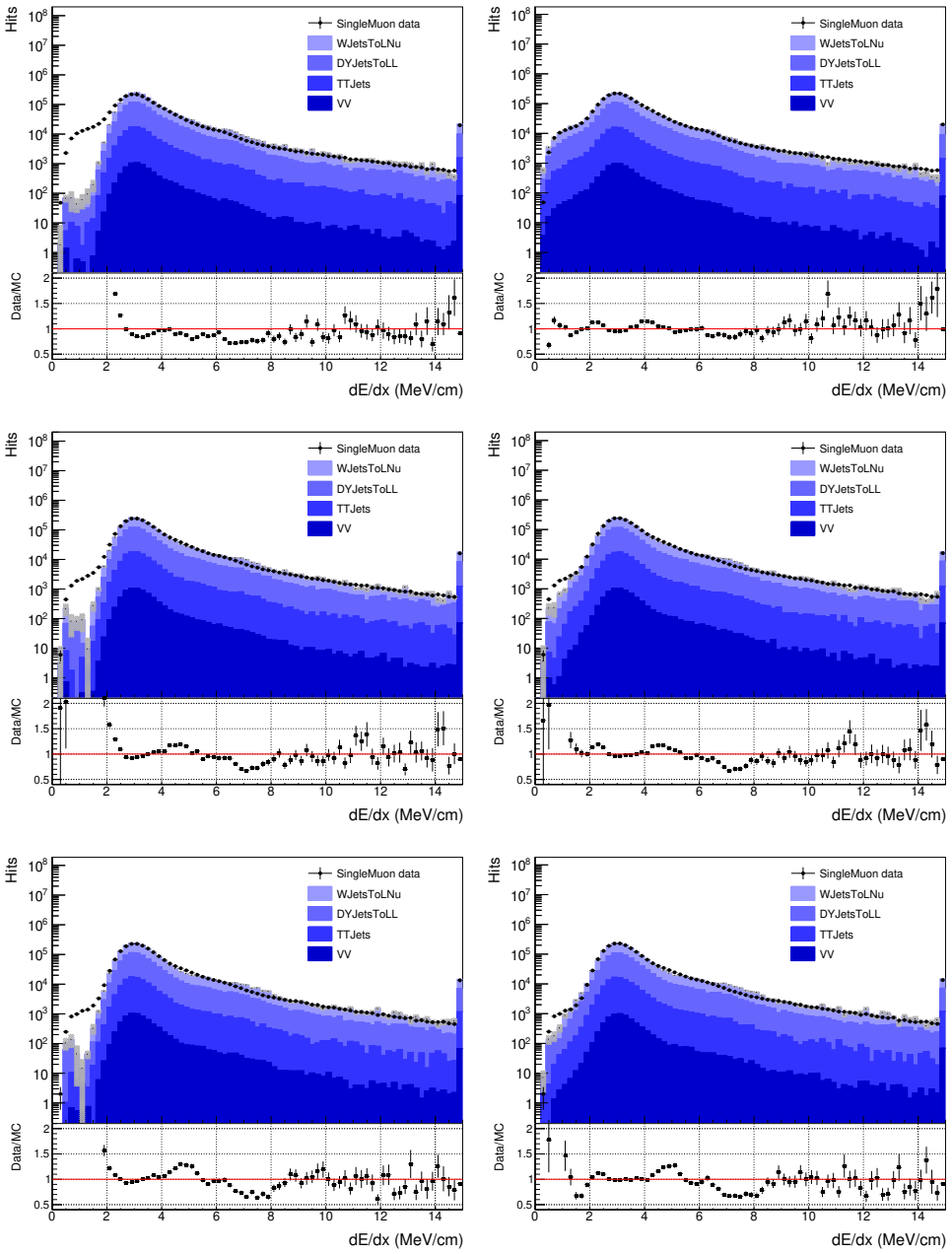


Figure D.3: dE/dx spectrum before (left) and after (right) corrections in the TOB layer 4 (top) to 6 (bottom) in early 2016 data (19.7 fb^{-1}).

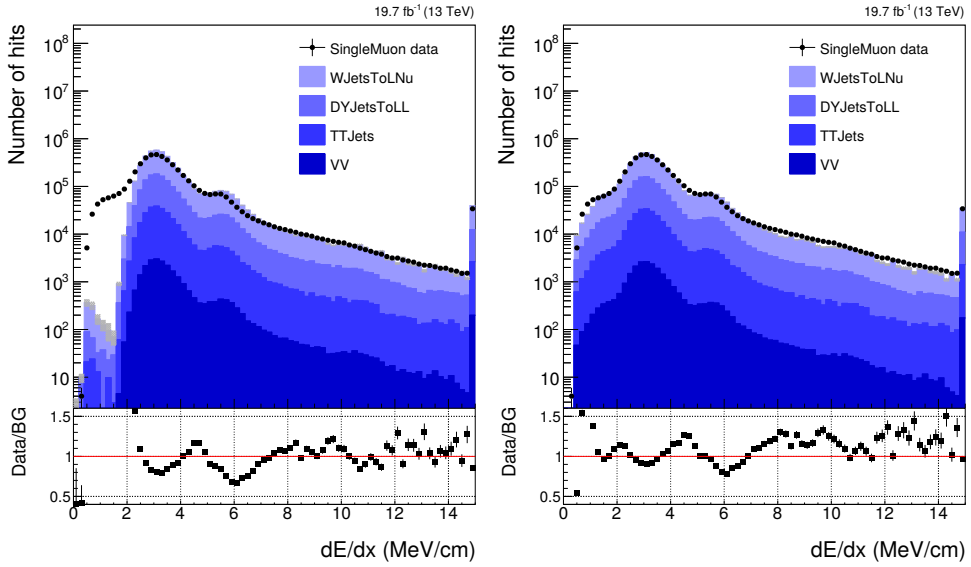


Figure D.4: dE/dx spectrum before (left) and after (right) corrections in the TEC in early 2016 data (19.7 fb^{-1}).

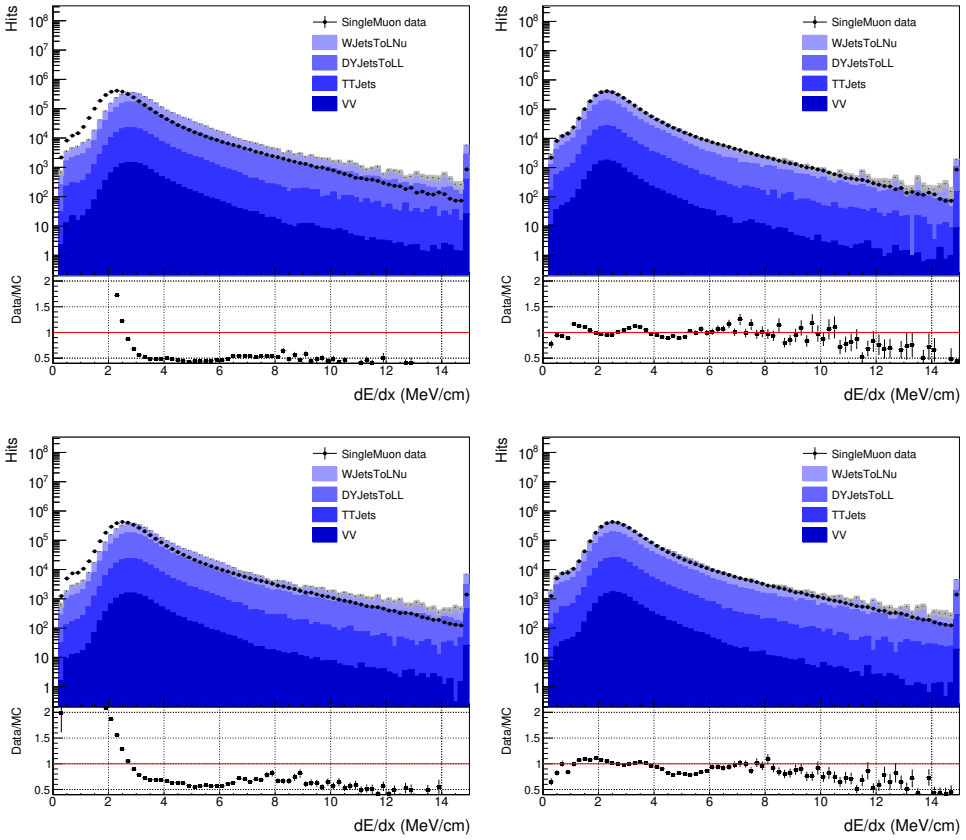


Figure D.5: dE/dx spectrum before (left) and after (right) corrections in the PXB layer 2 (top) and layer 3 (bottom) in late 2016 data (16.2 fb^{-1}).

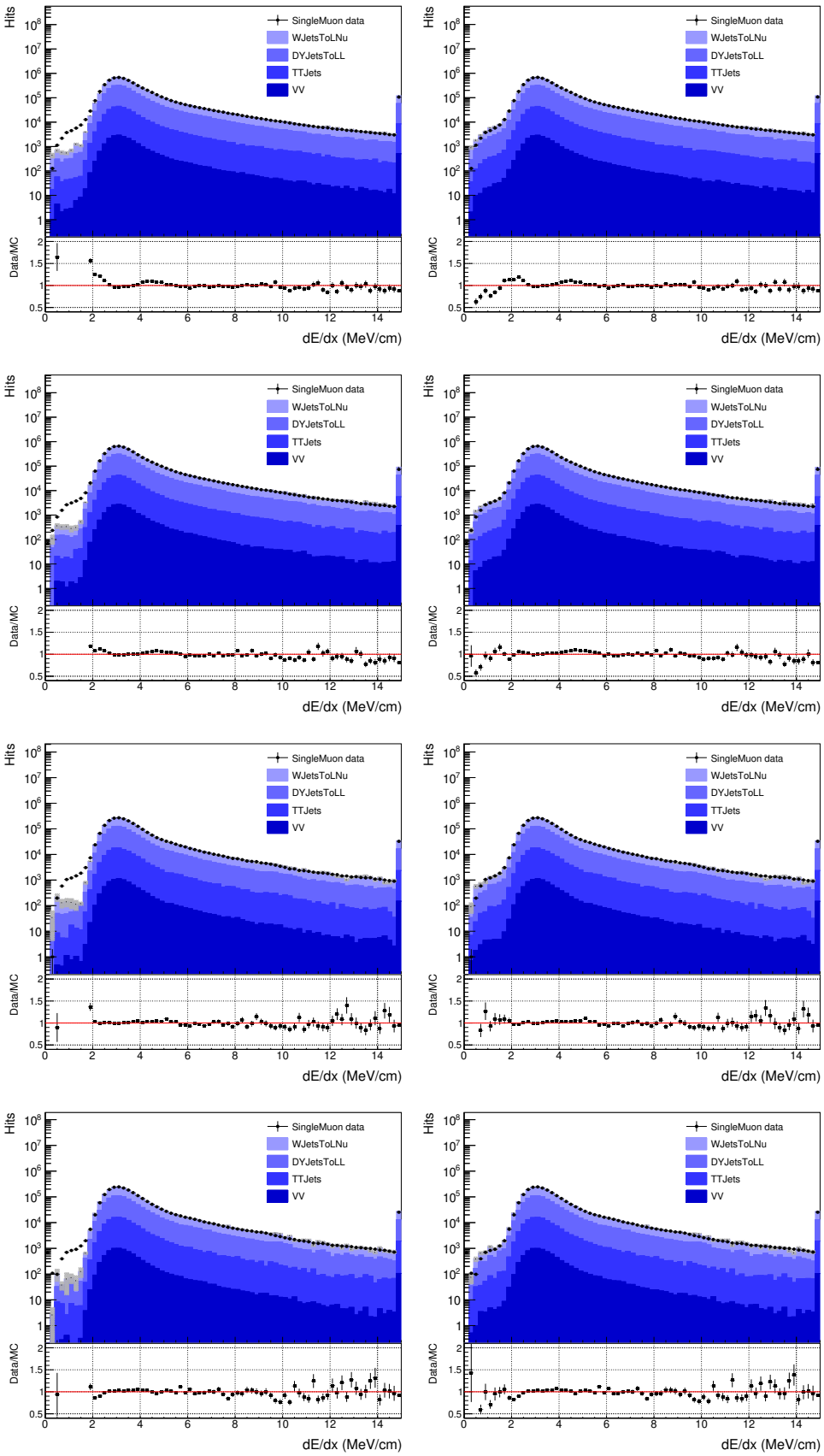


Figure D.6: dE/dx spectrum before (left) and after (right) corrections in the TIB layer 1 (top) to 4 (bottom) in late 2016 data (16.2 fb^{-1}).

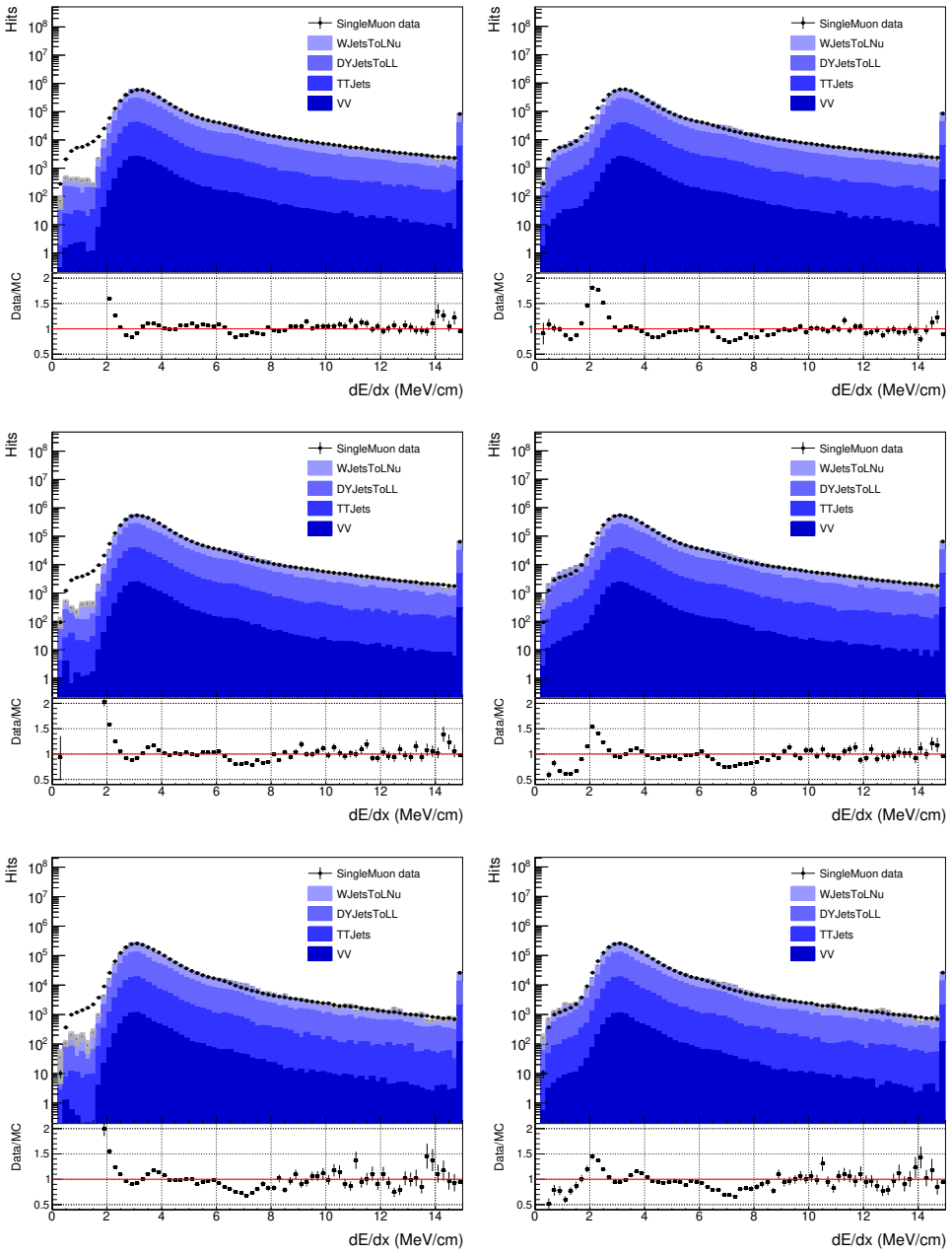


Figure D.7: dE/dx spectrum before (left) and after (right) corrections in the TOB layer 1 (top) to 3 (bottom) in late 2016 data (16.2 fb^{-1}).

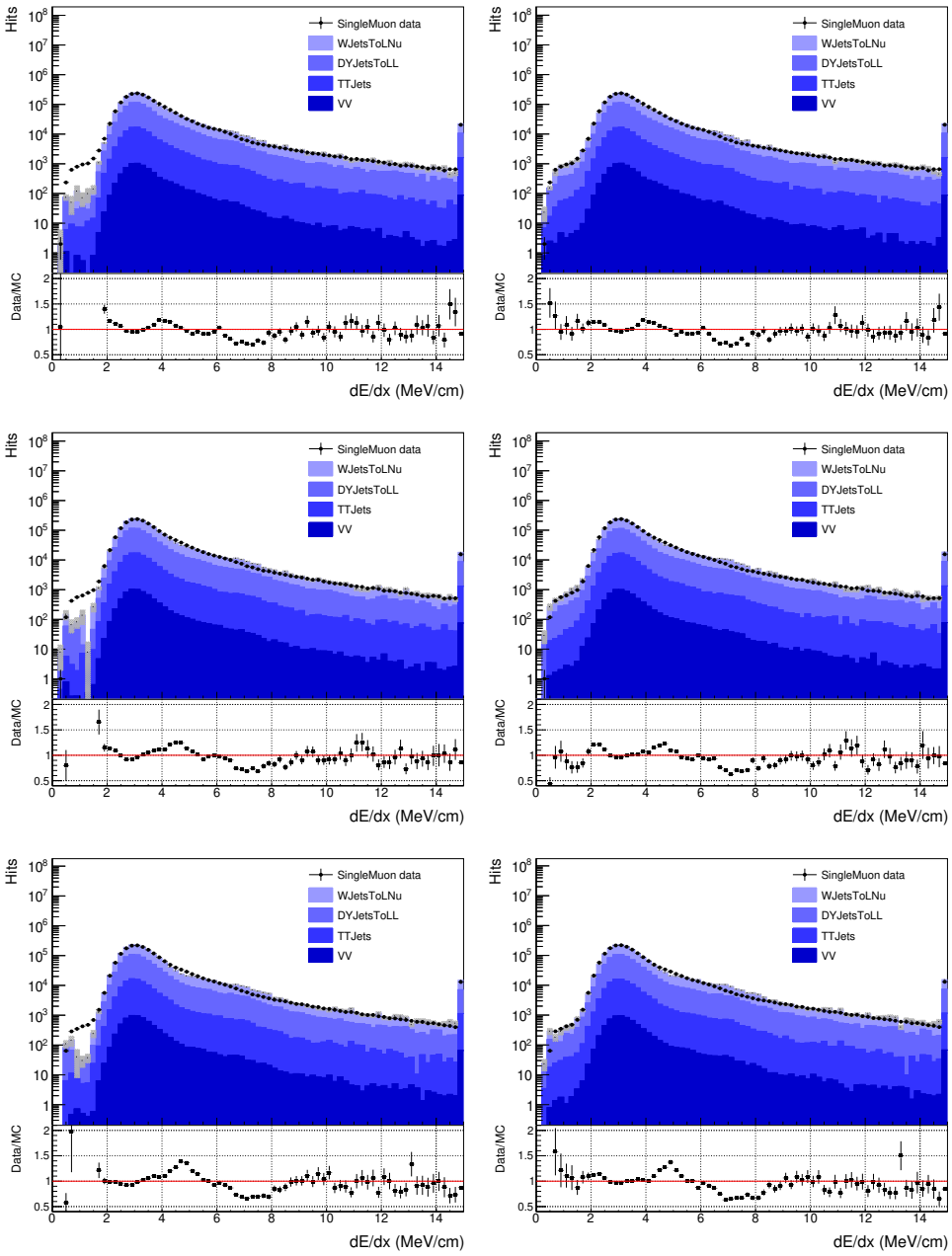


Figure D.8: dE/dx spectrum before (left) and after (right) corrections in the TOB layer 4 (top) to 6 (bottom) in late 2016 data (16.2 fb^{-1}).

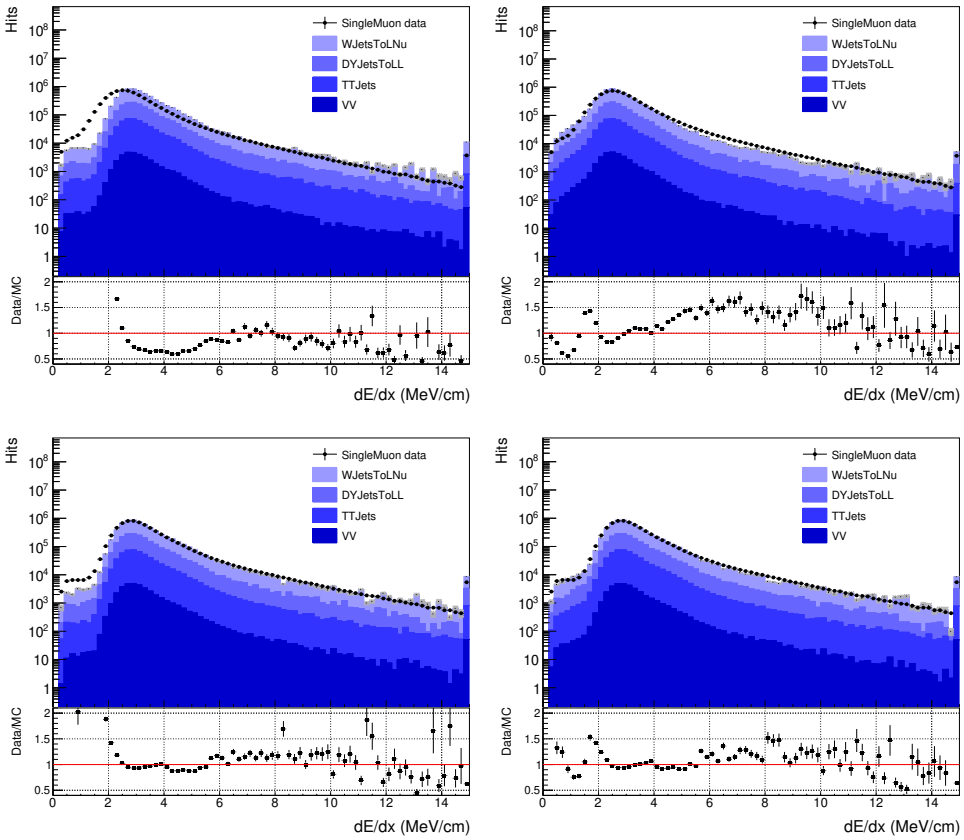


Figure D.9: dE/dx spectrum before (left) and after (right) corrections in the PXB layer 2 (top) and layer 3 (bottom) in 2017 data (41.5 fb^{-1}).

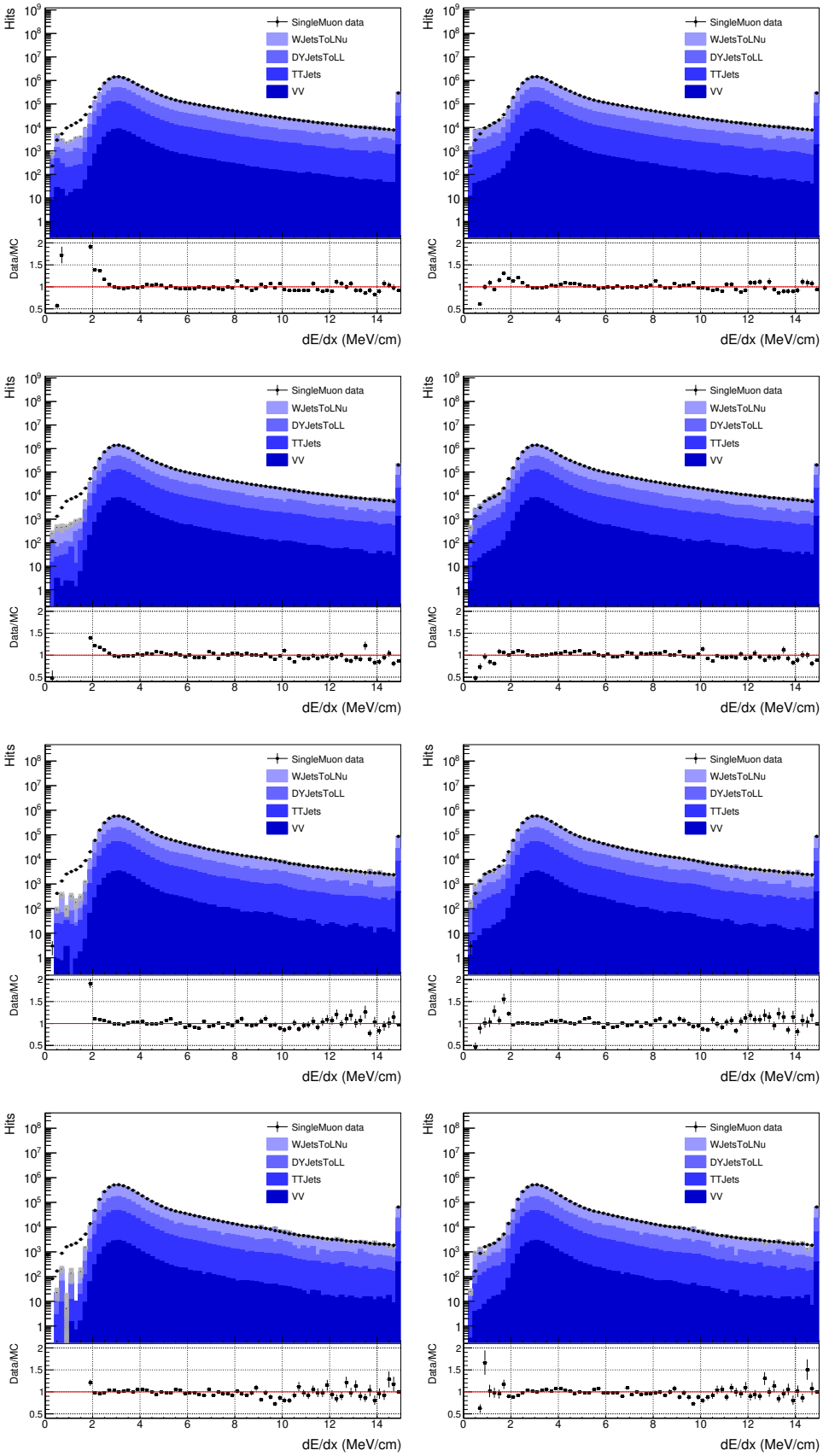


Figure D.10: dE/dx spectrum before (left) and after (right) corrections in the TIB layer 1 (top) to 4 (bottom) in 2017 data (41.5 fb^{-1}).

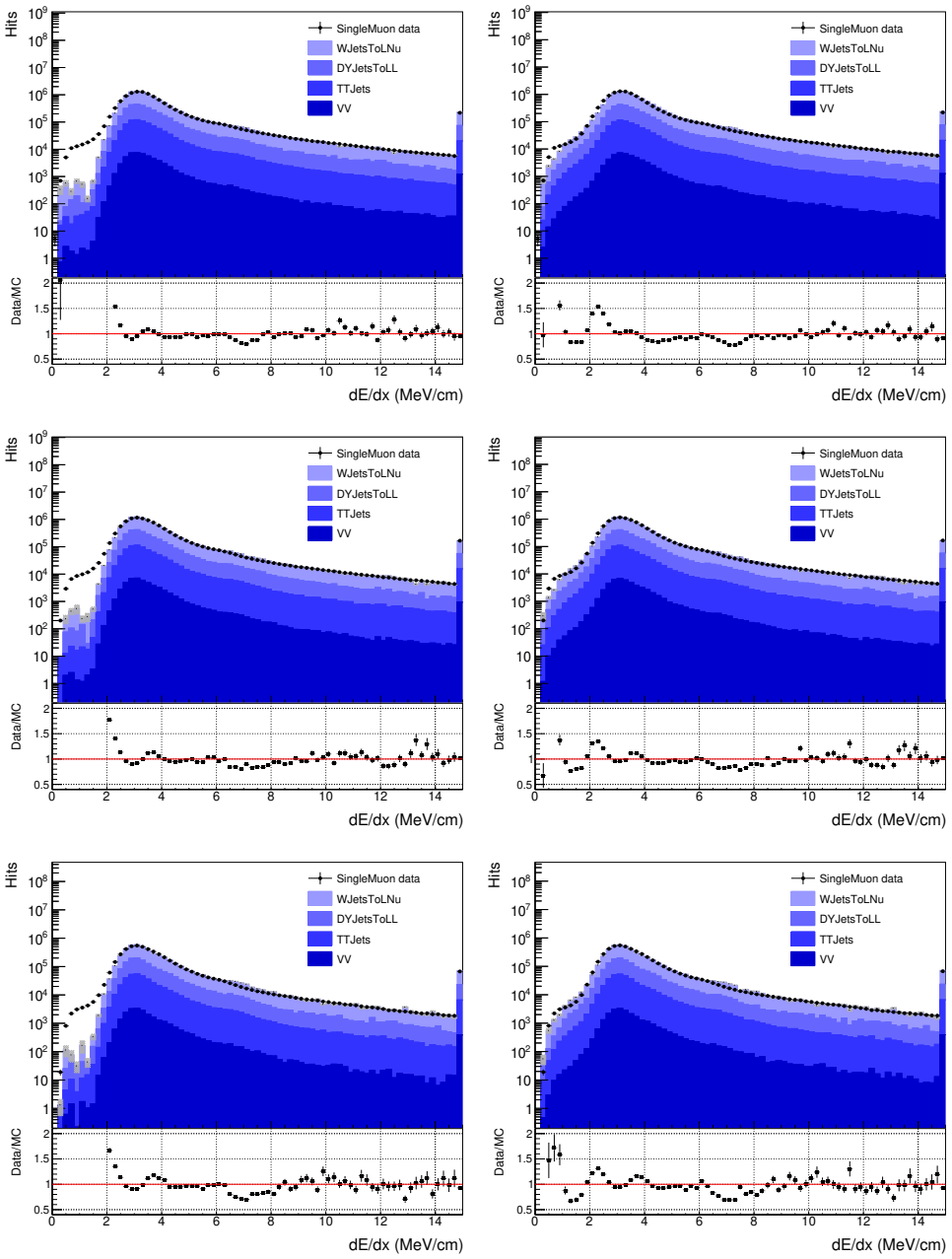


Figure D.11: dE/dx spectrum before (left) and after (right) corrections in the TOB layer 1 (top) to 3 (bottom) in 2017 data (41.5 fb^{-1}).

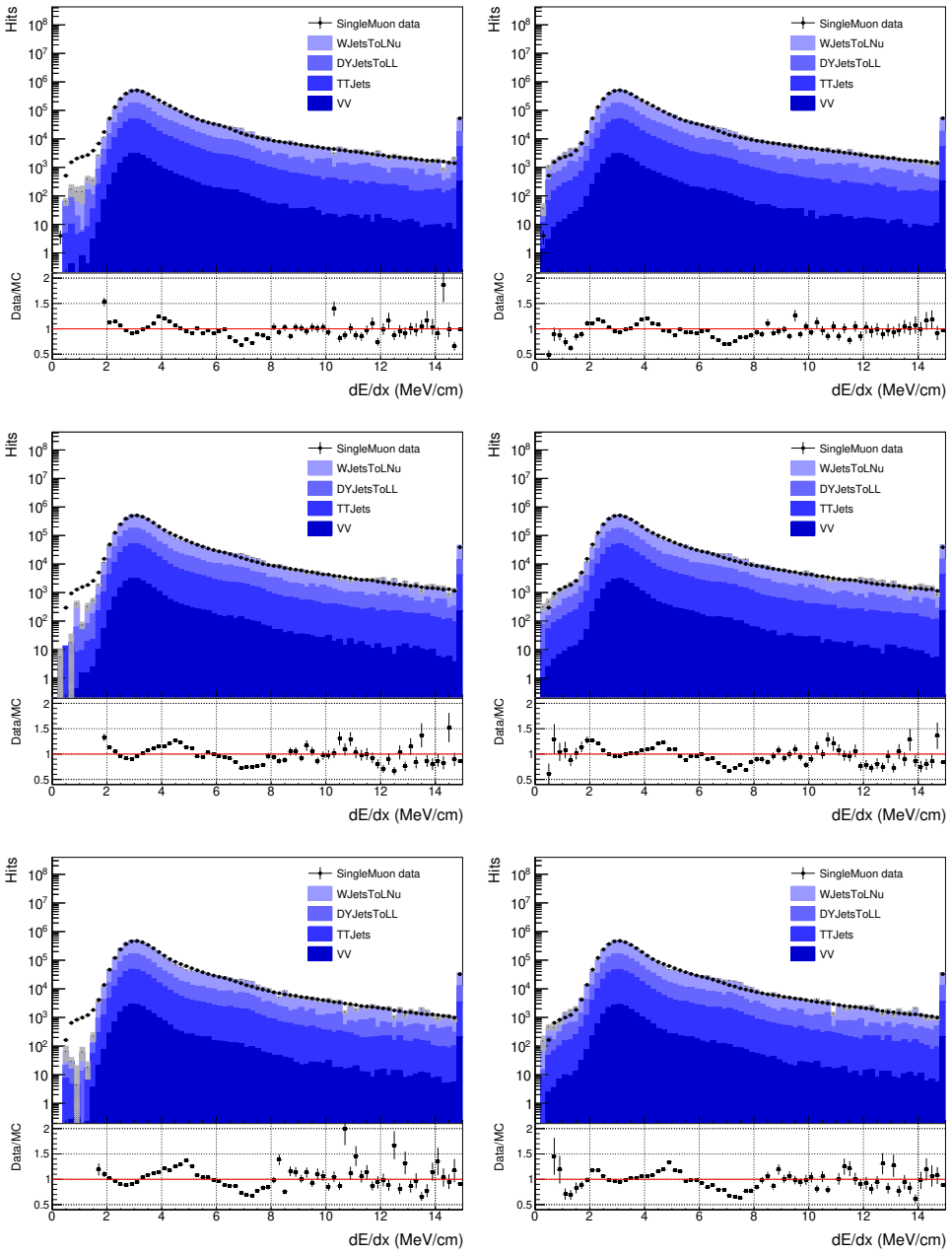


Figure D.12: dE/dx spectrum before (left) and after (right) corrections in the TOB layer 4 (top) to 6 (bottom) in 2017 data (41.5 fb^{-1}).

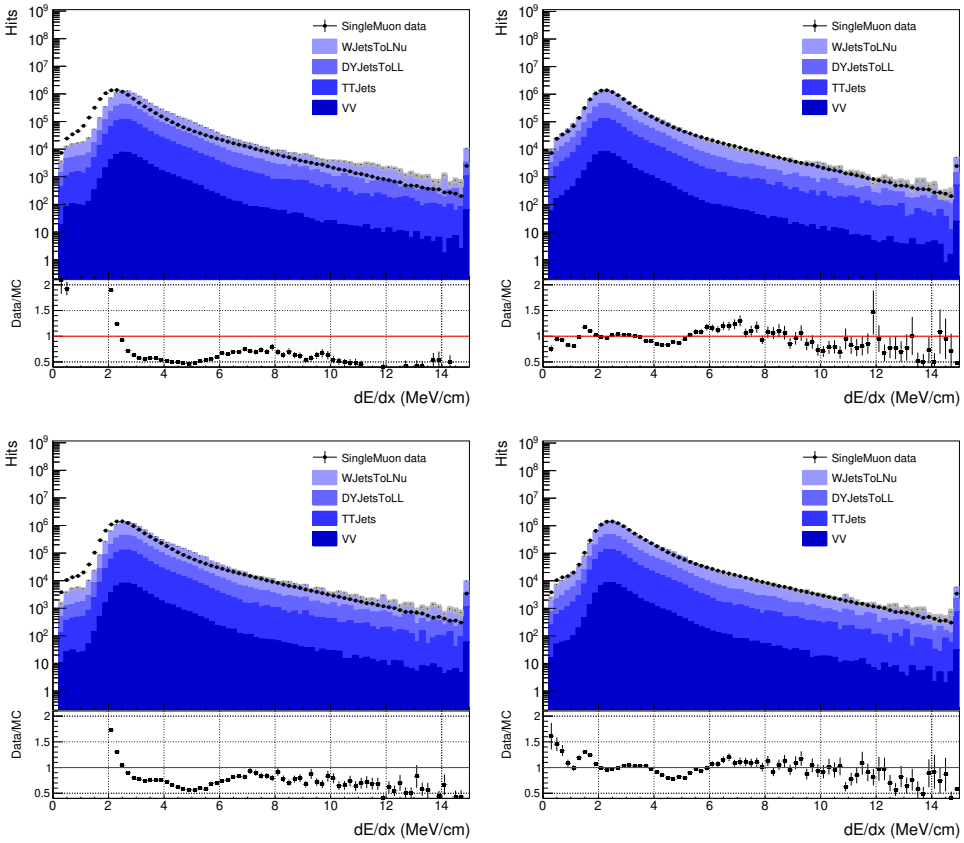


Figure D.13: dE/dx spectrum before (left) and after (right) corrections in the PXB layer 2 (top) and layer 3 (bottom) in 2018 data (58.8 fb^{-1}).

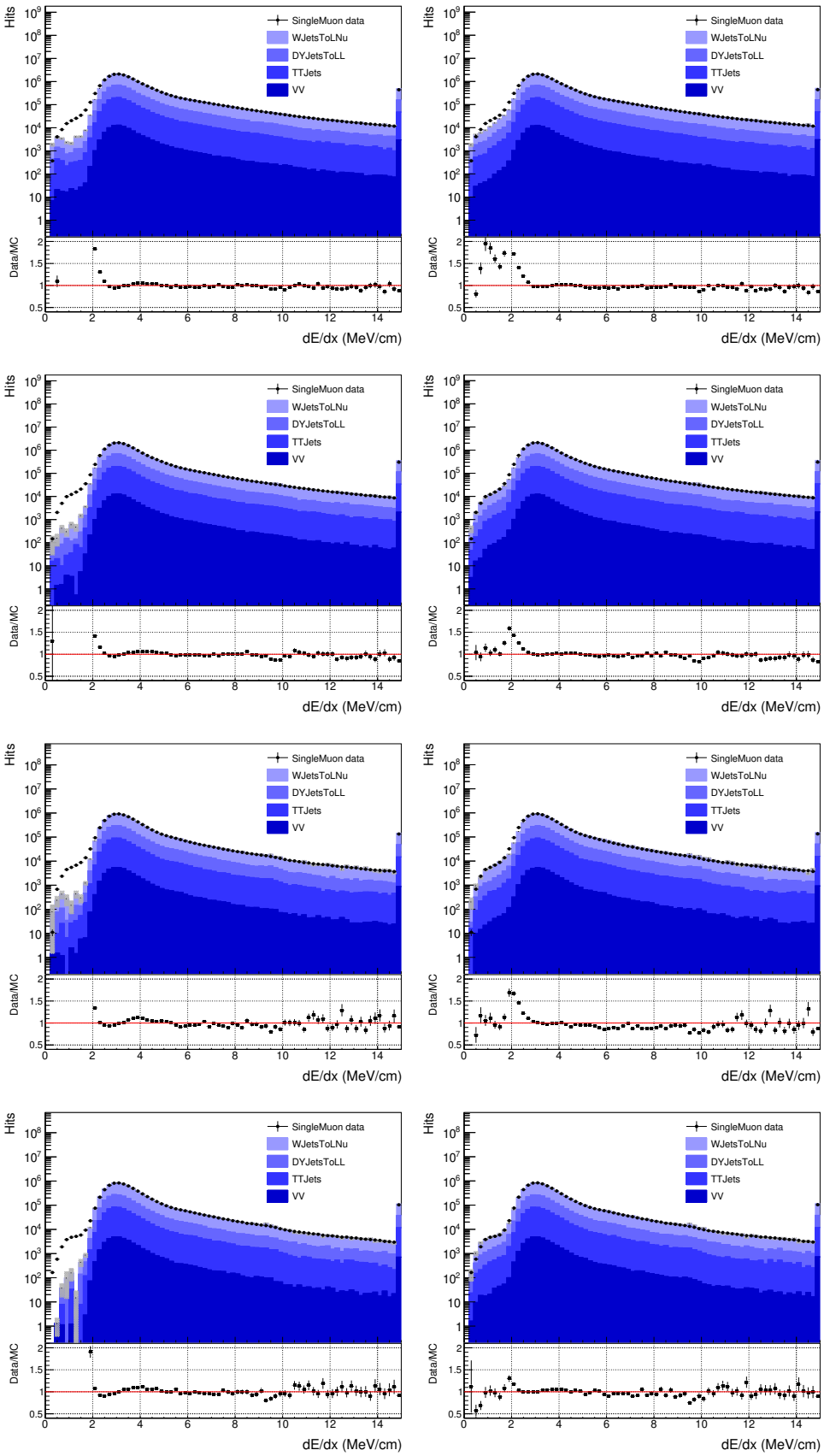


Figure D.14: dE/dx spectrum before (left) and after (right) corrections in the TIB layer 1 (top) to 4 (bottom) in 2018 data (58.8 fb^{-1}).

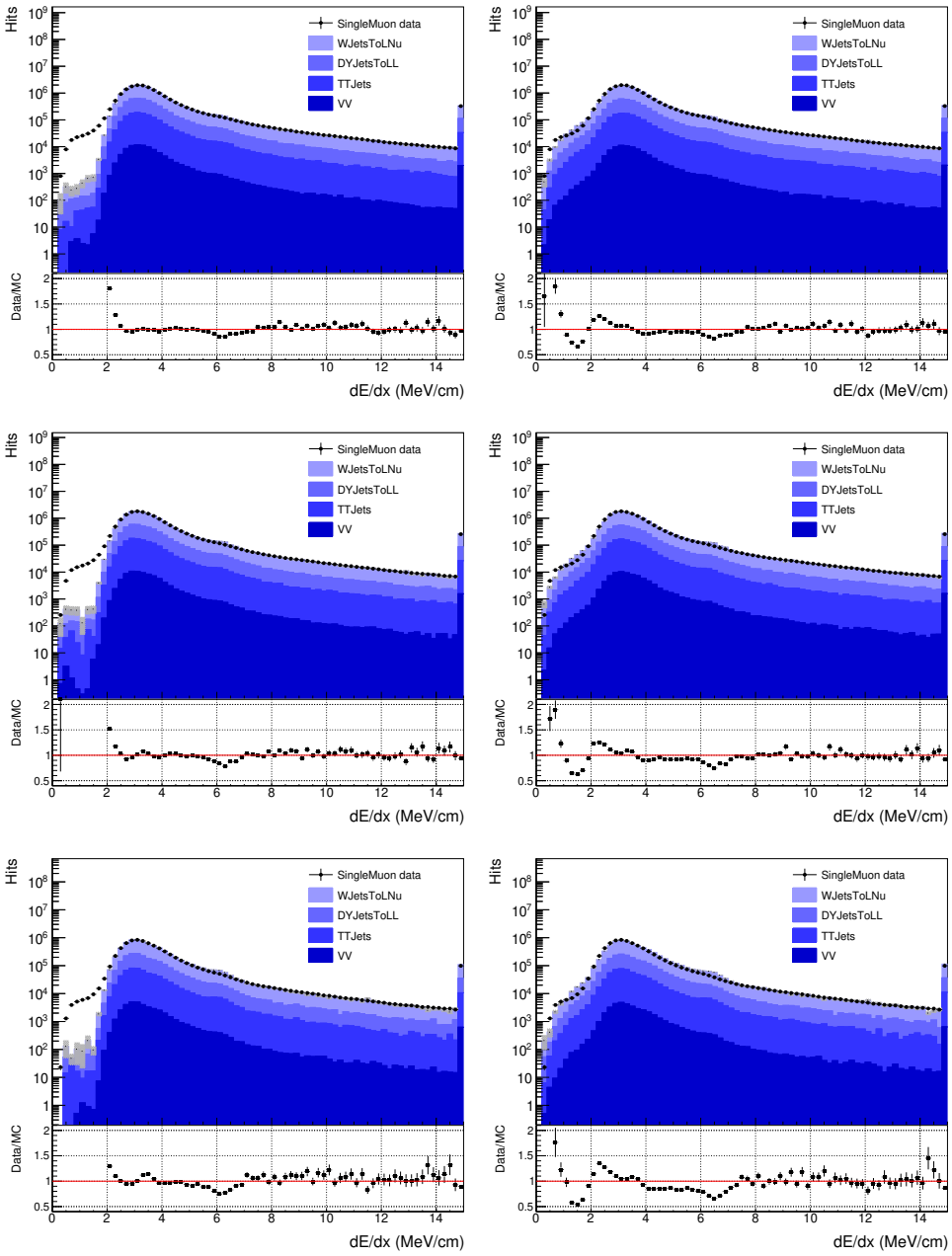


Figure D.15: dE/dx spectrum before (left) and after (right) corrections in the TOB layer 1 (top) to 3 (bottom) in 2018 data (58.8 fb^{-1}).

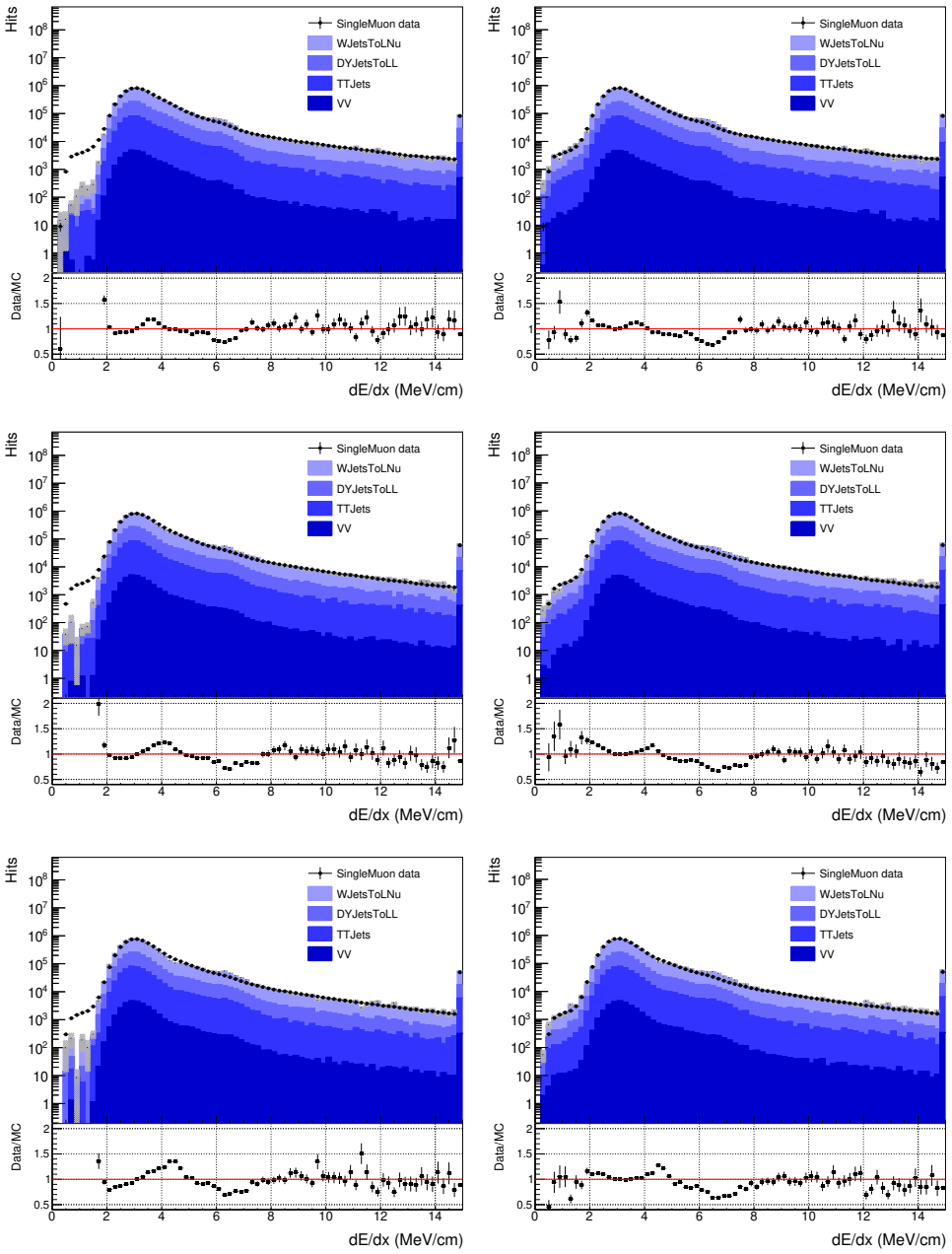


Figure D.16: dE/dx spectrum before (left) and after (right) corrections in the TOB layer 4 (top) to 6 (bottom) in 2018 data (58.8 fb^{-1}).