ULB UNIVERSITÉ LIBRE DE BRUXELLES

iihe BRUXELLES BRUSSEL

Faculté des Sciences
Département de Physique

Étude du processus de Drell-Yan à haute masse invariante au LHC à $\sqrt{s} = 8$ TeV

(Study of the Drell-Yan process at high invariant mass at LHC with $\sqrt{s} = 8$ TeV)

Patrick L.S. Connor

Promoteur : Prof. Laurent Favart
Co-promoteur : Dr. Tomislav Ševa

Année académique 2013-2014

Mémoire présenté en vue de l'obtention
du diplôme de Master en Sciences Physiques

Mots-clefs physique des hautes énergies, Modèle Standard, processus de Drell-Yan, haute virtualité, facteurs de forme, quarks, partons, muons, photons, bosons neutres faibles, section efficace, LHC, CMS, limite sur la taille des quarks

Résumé

En 2012, l'expérience CMS récolta des données à une énergie de 8 TeV dans le référentiel de centre de masse, échelle qui n'avait jamais été atteinte auparavant avec des protons en laboratoire. De telles énergies ouvrent de nouvelles régions dans l'espace des phases : notamment, on espère pouvoir étudier les quarks à plus petite échelle que cela n'a encore été fait. Ce mémoire propose une analyse du processus de Drell-Yan dans les données de CMS, ainsi qu'une étude de l'insertion de facteurs de formes des quarks dans l'expression de la section efficace pour essayer de déterminer si les quarks ont une taille en optimisant l'ajustement entre données et simulations.

Sommario

Nel 2012 lo sperimento CMS recordò dati a un'energia così alta come 8 TeV nel sistema di riferimento di centro di masse, scala che no era mai stato arrivata prima in laboratorio. Questa energia apri nuove regioni dello spazio delle fasi: speriamo tra l'altro studiare i quark a più piccola scala. La presente tesi proporre un'analisi del processo di Drell-Yan nei dati del CMS, così come une studo dell'inserimento di fattori di forma dei quark nell'espressione delle sezione d'urto per provare di determinare se i quark abbiano una taglia ottimizzando l'adattamento tra i dati e le simulazioni.

Samenvatting

In 2012 nam de CMS experiment data van een zo hoog energie als 8 TeV in de systeem van centrum of massa, schaal dit nooit vroeger met protonen in laboratorium was bereikt. Zo'n energie opent nieuwe perspectieven voor onderzoek : in het bijzonder wordt de bestudering van quarks aan smallere schalen hopen voorlopen. Dit proefschrift stelt een analysis van Drell-Yans proces in de data van CMS voor, evenals een studie van de plaatsing van vormfactoren van quarks in de formule van de cross section om de gestalte van de quarks te proberen maten door het verschil tussen de data en de simulaties te verminderen.

Abstract

In 2012, CMS experiment took data at an energy so high as 8 TeV in the centre-of-mass system, scale that had never been reached before with protons in laboratory. Such an energy opens new regions of the phase space: consequently, we hope to be able to look at the quarks at smaller scales than ever done before. This master thesis presents an analysis of the Drell-Yan process in CMS data, as well as a study of the introduction of form factors for the quarks in the expression of the cross section to optimize the fitting between data and simulations.

Key-Words high energy physics, Standard Model, Drell-Yan process, high virtuality, form-factors, quarks, partons, muons, photons, neutral weak bosons, cross section, LHC, CMS, limit on quarks' size

Contact: pconnor@ulb.ac.be

Aeneadum genetrix, hominum divomque voluptas,
alma Venus, caeli subter labentia signa
quae mare navigerum, quae terras frugiferentis
concelebras, per te quoniam genus omne animantum
concipitur visitque exortum lumina solis:
te, dea, te fugiunt venti, te nubila caeli
adventumque tuum, tibi suavis daedala tellus
summittit flores, tibi rident aequora ponti
placatumque nitet diffuso lumine caelum.
nam simul ac species patefactast verna diei
et reserata viget genitabilis aura favoni,
aeriae primum volucris te, diva, tuumque
significant initum perculsae corda tua vi.
inde ferae pecudes persultant pabula laeta
et rapidos tranant amnis: ita capta lepore
te sequitur cupide quo quamque inducere pergis.
denique per maria ac montis fluviosque rapacis
frondiferasque domos avium camposque virentis
omnibus incutiens blandum per pectora amorem
efficis ut cupide generatim saecla propagent.
quae quoniam rerum naturam sola gubernas
nec sine te quicquam dias in luminis oras
exoritur neque fit laetum neque amabile quicquam,
te sociam studeo scribendis versibus esse,
quos ego de rerum natura pangere conor
Memmiadae nostro, quem tu, dea, tempore in omni
omnibus ornatum voluisti excellere rebus.
quo magis aeternum da dictis, diva, leporem.

—Lucretius, *De rerum natura*

Acknowledgements

I first of all thank my *promoteur* Pr. Laurent Favart, my *co-promoteur* Dr. Tomislav Ševa and (coming soon Dr.) Alexandre Lénoard (who clearly deserves the title of co-co-promoteur) for the time they have spent with me and for the trust they have granted me. I wish the best to Tomislav, who could not pursue our activities. I also thank my two lecteurs, Pr. Frank Ferrari and Pr. Barbara Clerbaux.

I thank the whole IIHE, among them all Pr. Pascal Van Laer and (coming soon Dr.) Laurent Thomas for the their advice. Thanks, David Vannerom for the beautiful coversheet I have stolen you; Hugo Delannoy, Federicco Lucchetti and Gwenhaël de Wasseige for having suffered me for years while studying particle physics. And finally, thanks to my parents, my brother Thomas, and my friends, especially the musicians from the Conservatoire Royal de Bruxelles, Thierry Stulemeijer, and the Depretz.

Contents

0

Introduction

The motivation of the present master thesis is to address the following question:

« Do the quarks have a size? »

It is a very natural question. Let us recall a few historical facts.

In 1909, Rutherford experiment showed that the atom, if it existed, was possibly made of a positively charged nucleus surrounded by an electronic cloud. In 1913, with *Les Atomes*, Jean Perrin definitely convinced the scientific community of the existence of the atom. Suspecting that the nucleus was made of elementary charges, Hideki Yukawa presented in 1935 a theory to explain how such a positively charged structure could hold in such a small volume as the nucleus with a short-ranged, nuclear interaction. Various experiments were conducted to study those nuclear interactions and led to the discovery of many new particles. Some of them, called the *hadrons*, including the proton, showed symmetrical properties, suggesting similar substructures in turn. During the 1950s, at the SLAC, the scattering electron-proton allowed to measure the size of the proton and during the 1960s, getting higher in energy, the beams used in scattering experiments highlighted structures of the proton and of the neutron at smaller distance scales. They appeared as sets of pointlike particles, each carrying a fraction of the four-momentum of the nucleon, called *partons*. Parallelly, in 1964, Murray Gell-Mann and George Zweig used those symmetries to predict the existence of components of the hadrons: the *quarks*. Subsequent studies showed that the partons were either quarks, as predicted, or gluons, whose rôle is similar to the photon's for nuclear interaction.

Our present motivation stands precisely in the continuity of the study of the size of the tiniest component of matter. Thanks to the CMS experiment at LHC, we now have access to the tera-electron-Volt scale. This motivates the current analysis of the Drell-Yan process, which describes the leptoproduction due to quark annihilation. As a parton process, it is relatively abundant at hadron colliders such as LHC. So, fitting Monte Carlo simulations to the data, we hope to be able to measure a size to quarks, or at least to put a limit to it.

Here is the plan:

1. a few recalls on the hadron-hadron collisions,
2. a study of the Drell-Yan cross section,
3. a description of the experimental setup (LHC and CMS),
4. the analysis of the Drell-Yan process,
5. and lastly the discussion about the quarks' size.

1

Context

THIS FIRST CHAPTER'S AIM is to recall some generalities about the Standard Model of High Energy Physics, and to present the Drell-Yan process as a hadron-hadron collision. The idea of this work will be further explained toward the end of this chapter.

1.1 Standard Model of High Energy Physics [1]

1.1.1 High Energy Units

High Energy Physics (HEP) is the field of physics that studies the constituents of matter at both quantum and relativistic scales. This is reflected in terms of units:

1. electric charges are counted in units of $e = 1.602 \times 10^{-19}$ C,
2. terms of the dynamical equations are of order $\hbar = \frac{h}{2\pi} = 1.055 \times 10^{-34}\,\mathrm{J \cdot s}$,
3. and velocities are measured in units of $c = 2.99 \times 10^{8}$ m/s.

In practice, units of energy are *electron-Volt*, or multiples:

$$1\,\mathrm{GeV} = 1.602,175,565(35) \times 10^{-10}\,\mathrm{J} \tag{1.1}$$

Momentum and mass are respectively measured in GeV/c and GeV/c^2. In practice however, the c is often omitted. $\hbar$ is always forgotten, but if need be, it is straightforward to get it back into the formula using dimensional analysis.

As for « High Energy », using c and $\hbar$ (or $h = 2\pi\hbar$), one can relate an energy E to a wavelength λ:

$$E = \frac{hc}{\lambda} \tag{1.2}$$

This means that getting higher in energy is synonymous with probing small distance scales.

1.1.2 Generalities

At such scales, matter and interactions are both regarded as particles. HEP is also known as Particle Physics.

1. For a general introduction to High Energy Physics, see Donald H. PERKINS [24]; for a general introduction to Quantum Field Theory, see Michael E. PESKINS & Daniel V. SCHROEDER [25]; for a general introduction to the Standard Model, see Francis HALZEN & Alan D. MARTIN [17]; for a good introduction to Gauge theories, see Bjørn FELSAGER [14].

The Standard Model (SM) is a *relativistic quantum field theory* (QFT) that describes elementary particles as degrees of excitation of fields; it is based on the principle of Gauge invariance to formulate the fundamental interactions. It has been developed for decades and though already self-consistent, is still under development.

Mathematically, the behaviour of the fields (typically noted ϕ, Ψ or A_μ) is contained in a Lagrangian $\mathcal{L}$ and the equations of motion (EOM) are derived from it using the Principle of Least Action:

$$S = \int \mathcal{L}\left(\phi, \partial_\mu \phi\right) \, \mathrm{d}^4 x \quad \longrightarrow \quad \text{EOM} \equiv \delta S = 0 \tag{1.3}$$

The Principle of Gauge [2] invariance states that the physics described by the EOM is unchanged under a Gauge transformation of the fields (*i.e.* under a global change of phase, possibly depending on the position in spacetime and possibly being a matrix). This drastically restricts the kinds of terms that may appear in the Lagrangian.

On the four fundamental interactions,

— electromagnetic and weak nuclear interactions are described within the same body of theory, called *Electro Weak* theory (EW), whose Gauge group is $U(1) \times SU(2)$;

— strong nuclear (or *colour*) interaction is described by *Quantum ChromoDynamics* (QCD), the Gauge group being $SU(3)$, and is responsible for the quarks holding together in *hadrons*;

— gravitational interaction is not only too weak to be studied at a smaller scale than 1 mm, but it is also difficult to formalize as a QFT, and is therefore not described by the SM.

The Lagrangian including the interactions terms is constructed from the Lagrangian of free particles by requiring it to be Gauge invariant. It contains terms like:

mass terms	order two in the fields (ex: $-\frac{1}{2} A_\mu A^\mu$)
interaction terms	order three or four in the fields (ex: $e^2 \bar{\Psi} \gamma^\mu \Psi A_\mu$)
kinematic terms	terms involving derivatives (ex: $i \bar{\Psi} \gamma^\mu \partial_\mu \Psi$)

where for instance A represents a vector field and Ψ a spinor field (see Table 1.1a for a summary of the different types of fields). The construction of the different terms of the Lagrangian is detailled in many references, like [17] (more physical) or [25] (more mathematical).

Particles of matter are all $\frac{1}{2}$-spin particles (summarised in Table 1.1b), whereas Gauge particles are all 1-spin particles (summarised in Table 1.1c); there is also the *elementary scalar boson*, which is spinless: it is probably the most intriguing particle, as it is involved in a complex procedure to make the terms of mass for the particles appear without breaking the Gauge invariance, but we are presently not concerned with it.

1.1.3 Cross sections

Figure 1.1 illustrates quite well the principle of scattering experiments in HEP: a target is bombed by a beam of known particles and the dynamics of the collisions is studied using the kinematic distribution of the results.

The main measurable quantity to characterise a scattering is the *cross section*; such a name suggests an area but this is no more valid in quantum physics, as the boundaries

2. Old word for phase.

spin	mass	type of field	symbol	d.o.f.	particles
0	yes	scalar	ϕ	1	BEH scalar boson
$\frac{1}{2}$	no	Weyl spinors	Ψ	2	neutrinos
	yes	Dirac spinors	Ψ	4	charged leptons and quarks
1	no	massless vector	A_μ	2	photon and gluons
	yes	massive vector	W^i_μ	3	weak bosons

(a) Fields in the Standard Model.

type	particles	electromagnetism	weak forces	strong forces
charged leptons	e, μ, τ	yes ($Q = \pm 1$)	yes ($T^3 = 0, \frac{1}{2}$)	no (white)
neutral leptons	ν_e, ν_μ, ν_τ	no ($Q = 0$)		
up quarks	u,c,t	yes ($Q = \frac{2}{3}$)	yes ($T^3 = 0, \frac{1}{2}$)	yes (RGB)
down quarks	d,s,b	yes ($Q = -\frac{1}{3}$)		

(b) Particles of matter are all $\frac{1}{2}$-spin particles. Particles from left to right have been ordered by mass.i The six different types of quarks are called *flavour*. Except for the neutrinos that are neutral, every elementary particle has a matching antiparticle that has the same properties but an opposed electric charge; they are mathematically contained in the same field.

Gauge boson	symbol	carried interaction	range
photon	γ	electromagnetism	infinite
neutral weak boson	Z^0	weak interaction	nuclear scale
charged weak bosons	$W^\pm$		
gluons	g	colour interaction	nuclear scale

(c) Particles of interaction are all 1-spin particles. Gravitation is not described by the SM. There should be eight gluons instead of one, but they are perfectly symmetric and cannot be distinguished from one another; furthermore, as massless, the gluon might be of infinite range, but it is effectively restricted to the nuclear scale.

TABLE 1.1 – Summary of the particles as described by the SM.

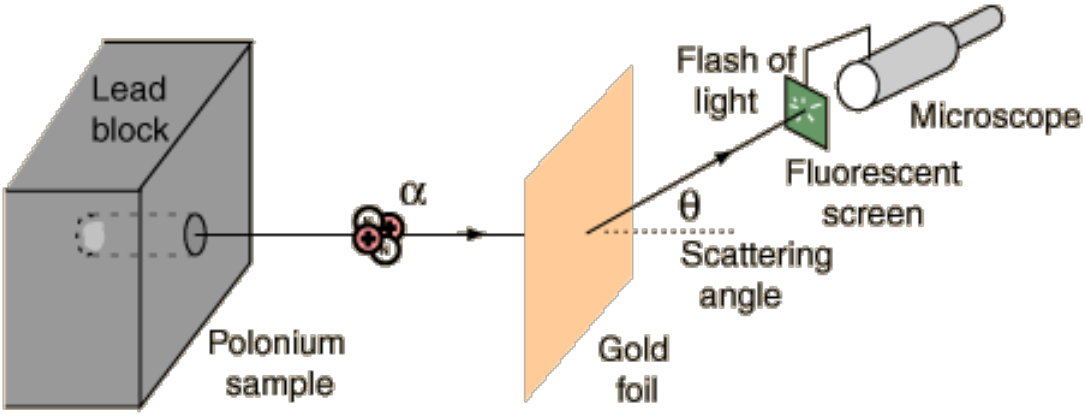


FIGURE 1.1 – Rutherford experiment, achieved by his two collaborators Hans GEIGER and Ernest MARSDEN, is one of the key experiments in modern physics. Golden atoms are thrown α particles on. The angular distribution suggests the existence of a pointlike, positively charged nucleus. Modern experiment are still based on the same principle of bombing target particles to study their nature using the angular distribution of the outgoing particles.

of objects are not well-defined, and it is rather regarded as the quantity that contains the dynamics of the physical process of the scattering.

Techniques have been developed to compute cross sections with the SM. As the exact computation of a cross section cannot usually be done analytically, Richard FEYNMAN and some others developed a pertubative technique to make computations and to represent the interaction at a given order in the *coupling constant*, which is a quantity describing the strength of an interaction. Those representations are called *Feynman diagrams*. This procedure is powerful: to compute the cross section of a process at a given order, just draw the different possible diagrams, and apply the *Feynman rules* (see for instance the appendix of M. PESKIN & D. SCHROEDER) to compute the *scattering matrix element*

$$\mathcal{M} = \langle \text{final state} | \text{ interaction Hamiltonian } | \text{initial state} \rangle \tag{1.4}$$

like $\mathcal{M}(q\bar{q} \to \mu^+\mu^-) = \langle \mu^+\mu^- | H_{\text{int}} | q\bar{q} \rangle$ as we shall do for the Drell-Yan process. Then the *Fermi Golden Rule* (FGR) may be used to compute the cross section. In the simplest case of two incident, non-polarised particles and two resulting particles, FGR is reduced to:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{\overline{|\mathcal{M}|^2}}{64\pi^2 E_{\text{CMS}}^2} \tag{1.5}$$

where the line over the squared matrix element indicates that the polarisations of the incident particles have been averaged and the polarisations of the resulting particles have been summed.

1.1.4 Couplings

The strength of an interaction is given by the prefactor in the interaction term in the Lagrangian. In each case, it can be related to a dimensionless *running coupling constant*,

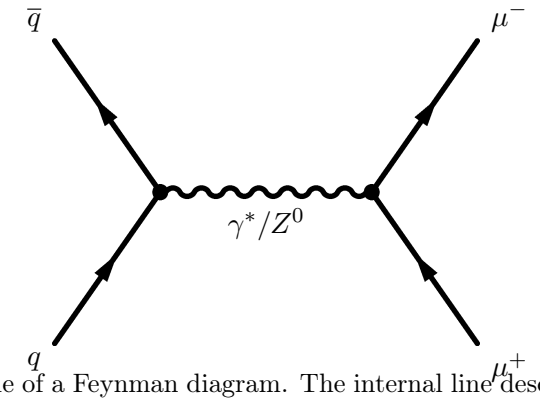


FIGURE 1.2 – Example of a Feynman diagram. The internal line describe a *virtual particle*, whose squared mass Q^2 is called *virtuality* and gives the energy scale of the process. The plain lines with arrows represent currents.

running because it depends [3] on the energy scale Q but constant in all other respects. It is related to the charge, depending on the interaction.

For instance, electromagnetic coupling is the *fine structure constant*. At a low energy, it is given by $\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c}$ (recall $\alpha \propto e^2$) and measured with a very good precision to $7.297,352,5698(24) \times 10^{-3}$ (from PDG booklet [16]).

interaction	coupling	value
electromagnetism	α	$\frac{1}{128.957 \pm 0.020}$
weak force	$\sin^2\theta_W$	0.23116 ± 0.00012
strong force	α_{S}	0.1184 ± 0.0007

TABLE 1.2 – Couplings' values are given at the energy scale $Q^2 = M_Z^2 = (91.2\,\mathrm{GeV})^2$ by convention. α_S is the direct analog to α for the strong force, whereas θ_W, Weinberg angle, is of different nature, as it is involved in the mixing of electromagnetism and weak force within the electroweak theory. The measured couplings correspond to *renormalised* couplings, which means that the internal lines of a Feynman diagram include the loop corrections; this just ends in actualising their effective values.

On the one hand, the values of the couplings of electromagnetism and weak force always allow to use the perturbative treatment required for the computation of the cross sections with Feynman rules. On the other hand, the strong coupling constant α_S is greater than one at low Q^2 and lower only from $Q^2 \geq \Lambda^2_{\mathrm{QCD}}$ where $\Lambda_{\mathrm{QCD}} \simeq 217^{+25}_{-23}\,\mathrm{MeV}$. At higher and higher energy, α_S becomes smaller and smaller, and the quarks and gluons (the only particles sensitive to strong force) behave increasingly like free particles: such a phenomenon is called *asymptotic freedom*.

1.1.5 Further

The SM is a remarkable edifice and its achievements as a predicting theory are numerous. However several mathematical as well as physical questions remain unanswered:

— its Gauge-structure $U(1) \times SU(2) \times SU(3)$ is unexplained,

— the existence of three generations of leptons and quarks is questionable,

— gravitation is not taken into account,

3. See Glenn BARNICH lectures [1].

— there are seventeen free input parameters,

— some chapters of HEP as neutrino oscillations are about to be included, although they need to be checked more thoroughly.

— evidence for dark matter abund in the universe but it can hardly be described as particles of the SM,

— non-perturbative QCD is not easy to study,

— the asymmetry observed between matter and antimatter in the universe is not explained.

The list is not exhaustive, which motivates further investigations if physicists want a fundamental theory of everything. Hundreds (thousands?) of theories try to extend the SM: they are called *Theories Beyond the Standard Model* (BSM).

This document aims at studying the annihilation of quarks of same flavour into a neutral electroweak boson, then decaying into a pair of muons. This annihilation should depend on the size of the quarks; this would be a lead for new physics at smaller distance scales.

1.2 Hadron-hadron collisions [4]

1.2.1 Generalities

1.2.1.1 Hadrons as sets of quarks

Quarks have never been observed free; they are always confined in hadrons [5].

Hadron means *strong* in Old Greek, in relation to the strong force. There are two kinds of hadrons, according to their *valence quarks*: just as the valence electrons that determine the properties of chemical species, the valence quarks determine the properties of the hadrons:

baryons (heavy) which are triplets of quarks or antiquarks,

mesons (middle) which are pairs of quark-antiquark.

The proton is the only stable baryon; the neutron may be stable only if it is bound into a nucleus. Mesons are all unstable. (Leptons (*light*) are no hadrons.)

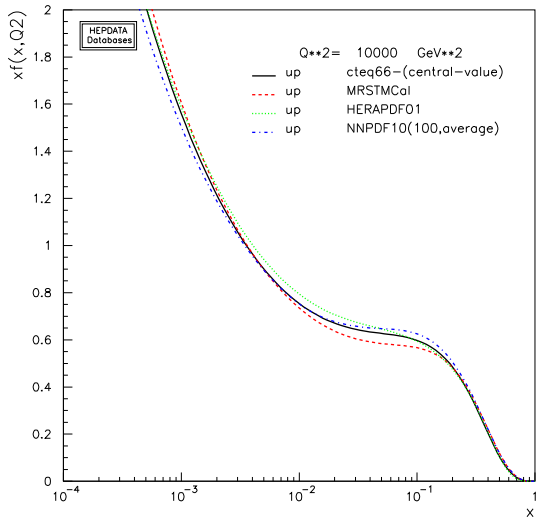
Other quarks may exist in a hadron, namely the *sea* quarks., whose existence is due to QCD fluctuations.

1.2.1.2 Feynman Parton Model

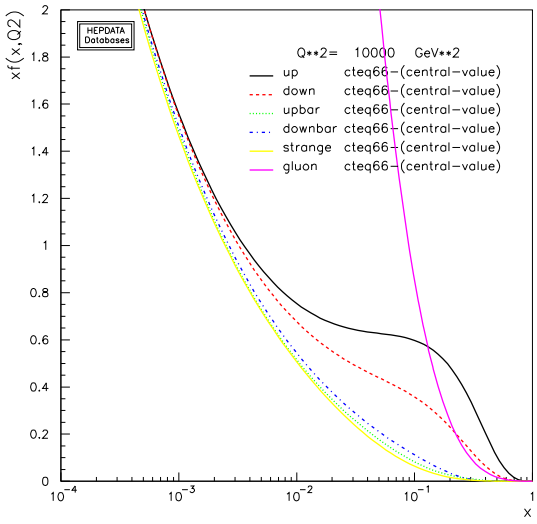
Parton is a generic word for the particles constitutive of the hadrons; it comes from the model imagined by Richard Feynman describing partons as free, pointlike particles moving collinearly inside the hadrons indenpently of the model of the quarks. Partons were identified to quarks and gluons only later. At energy scales $Q^2 \gg \Lambda^2_{\text{QCD}}$, the colour interaction weakens, and partons can be seen as free, though they are not isolated even so: they still live inside the hadrons, each one of them carrying a fraction x of the momentum of the hadron.

4. See Ellis, Stirling and Weber [12].

5. This sentence is actually not perfectly true: a particular state called *quark-gluon plasma* is made of deconfined quarks and gluons, but it happens unders conditions with which we are presently not concerned.



(a) Different PDFs exist to describe the up quark density in the proton, according to the achieved technique of measurement and of fitting.



(b) Quark, antiquark and gluon densities in the proton as predicted by CTEQ66 set. Gluons are the most important part of the content of the proton. The bumps of u and d densities correspond to the three *valence* quarks.

FIGURE 1.3 – Parton Distribution Functions at $Q^2 = (100\,\text{GeV})^2$ (Durham PDF plotter [11]).

The content of hadrons is measured in terms of quark, antiquark and gluon densities, generically called *Parton Distribution Functions* (PDF). The PDFs f depend on the scale of energy Q^2 (the higher the scale Q^2, the more partons inside the hadron) and on the momentum fraction x carried by the parton (see Figure 1.3). The precise analytic form of the PDFs is unknown and several fits and measurements exist, including CTEQ, MRST and HERAPDF among others.

1.2.2 Factorisation [6]

Let us fix the framework of hadron-hadron collisions.

As quarks are always grouped together, hadrons are accelerated to scatter at scales $Q^2 \gg \Lambda^2_{\text{QCD}}$ so that quarks seem asymptoticly free. The measured cross section is that of the hadron process, whereas we are interested in that of the parton process. In other words, the interaction involves the whole hadron and always contain some non-perturbative part. The *QCD factorisation theorem* states that the cross sections of the proton-proton scattering may be related to the parton-parton scattering owing to the PDFs.

Usually, the factorisation can only be achieved at asymptotical freedom scales $Q^2 \gg \Lambda^2_{\text{QCD}}$: at low scales Q^2, the hadron must be seen as a whole. It has been theoretically and experimentally showed for electron-proton scattering from $Q^2 \geq (1\,\text{GeV})^2$ and for most of the processes of proton-proton scattering from $Q^2 \geq (10\,\text{GeV})^2$. Technically, a *factorisation scale* μ_F must be chosen, fixing the resolution at which the proton content is studied. Given the factorisation scale μ_F, the PDFs include all the radiation effects

6. See Laurent FAVART's lectures [13].

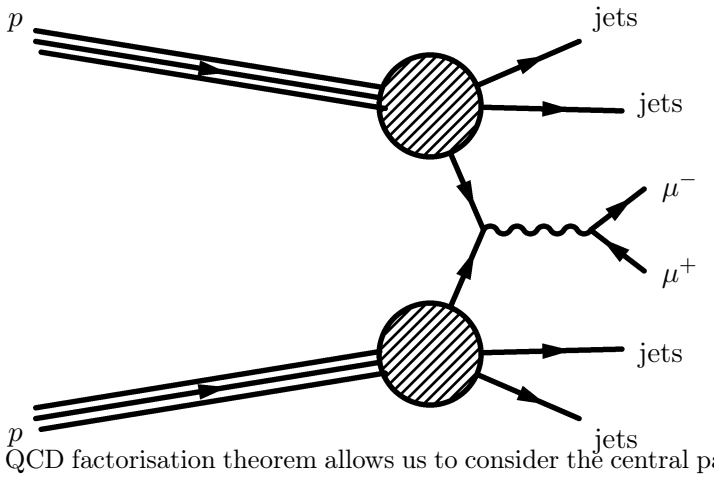


FIGURE 1.4 – The QCD factorisation theorem allows us to consider the central parton event on a very short time scale separately from the rest of the hadron event.

corresponding to the natural evolution of the proton at an energy between its mass and the factorisation scale. The most natural choice for the factorisation scale is the virtuality of the parton process (as on Figure 1.3), *i.e.* $\mu_F^2 = Q^2$.

For proton-proton scatterings in presence of events occurring at a *hard* scale Q^2 in the final state (*i.e.* with large momentum transfers), the factorisation reads:

$$\begin{aligned}\sigma_{pp'\to X(Q^2)} &= \sum_{a,b=q,\bar{q},g} \int_0^1 \mathrm{d}x_1\, \mathrm{d}x_2\, f_{a/p}(x_1,\mu_F)\, f_{b/p'}(x_2,\mu_F) \quad \times \\ &\times \quad \hat{\sigma}_{ab\to X(Q^2)}(x_1 p_p, x_2 p_{p'}, \mu_F, \alpha_S(\mu_R))\, \Theta\left((x_1 p_p + x_2 p_{p'})^2 - Q^2\right)\end{aligned} \tag{1.6}$$

where the PDFs $f_{a/p}$ describe the density of the parton a in the proton p; $\hat{\sigma}$ denotes the parton cross section (it can be computed thanks to Feynman rules for instance if the process occurs at a perturbative scale); Heaviside function Θ delimits the hard scales. The proton-proton cross section may be seen as the sum on the gluons and the flavours of quarks of the corresponding parton scatterings weighted by the PDFs, at a given factorisation scale μ_F. It is represented in Figure 1.4.

1.2.3 Underlying event

Even so, the scattering of hadrons cannot be reduced to the scattering of partons convoluted with PDFs. Additional hadron activity is contained in what is called the *underlying event*. There are two kinds of them:

Beam remnants	The spectators partons will be responsible of a part of the hadronic production.
Multiple Parton Interaction	Two or more interactions may come from the same pair of scattering hadrons.

The underlying event contributes to the hadron cross section and pollutes the signal of the parton process.

1.3 Drell-Yan process [7]

1.3.1 Description

The Drell-Yan process consists in the electroweak annihilation of a pair of quark-antiquark into a virtual photon or neutral weak boson decaying into a pair of fermion-antifermion:

$$q + \bar{q} \longrightarrow \gamma^*/Z \longrightarrow f + \bar{f} \tag{1.7}$$

On Figure 1.5, it is represented at tree-level.

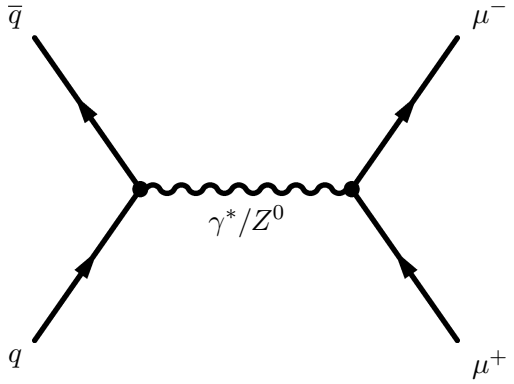


FIGURE 1.5 – Tree-level of the Drell-Yan process. There are diagrams: one for the electromagnetic interaction with the photon and the other for the weak interaction with the neutral weak boson.

The difermion production is often accompanied by strong (and sometimes electromagnetic) X radiations:

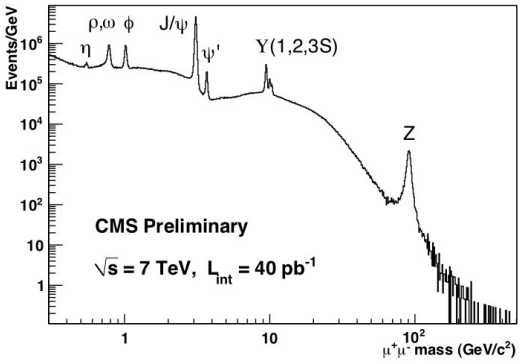
$$q + \bar{q} \longrightarrow f + \bar{f} + X \tag{1.8}$$

The probably two most famous discoveries that it helped to find are the hadron resonance J/Ψ (proving the existence of a new quark flavour: the *charm*) and the weak resonance Z^0 (see Figure 1.6a). With respect to other parton processes, it is rather abundant, and is therefore interesting to study for constraining or measuring physical quantities. For instance, it has been a very good test of the universality of the PDFs, as shown on Figure 1.6b.

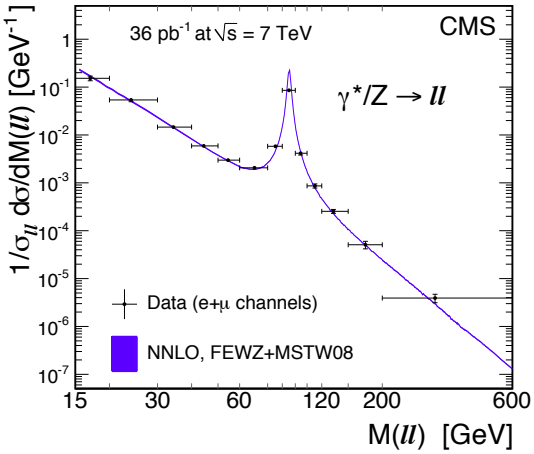
1.3.2 This study

Our study of the Drell-Yan process is guided by the research of the quarks' size in proton-proton collisions. Thanks to the new region of the phase space that the Large Hadron Collider (LHC) has opened, we might find an effect (see Figure 1.7). Such an effect is likely to be very sensitive, so we need a channel as clear as possible. There are two possibilities: either the electron channel, or the muon channel. The analysis is somewhat different for both cases and each one has its advantages and its inconveniences. The problem of electron channel is the important background, whereas the muon channel is much cleaner, and even called the *Golden Channel* at the CMS experiment, as we shall see.

7. For a detailled description of the Drell-Yan process, see the article of I.R. KENYON [19].



(a) Invariant mass spectrum of the two resulting muons of the Drell-Yan process obtained as a preliminary plot of CMS. One can see the peaks of the resonances, as the J/Ψ's or the Z^0's, among others. In our study, we shall be interested by such a plot, but in a range starting at 60 GeV up to 2 TeV, far more on the right.



(b) Test of the universality of the PDFs: the PDFs measured in the electron-proton scattering have been successfully used in the factorisation of the hadron-hadron scattering.

FIGURE 1.6 – Two of the successes of the Drell-Yan process: the resonances and the PDFs.

Therefore we shall focus on the Drell-Yan process in the muon channel:

$$p + p \longrightarrow \mu^+ + \mu^- + X \tag{1.9}$$

where X represents the radiations. Our study is *inclusive*, that is to say that we integrate over the radiations.

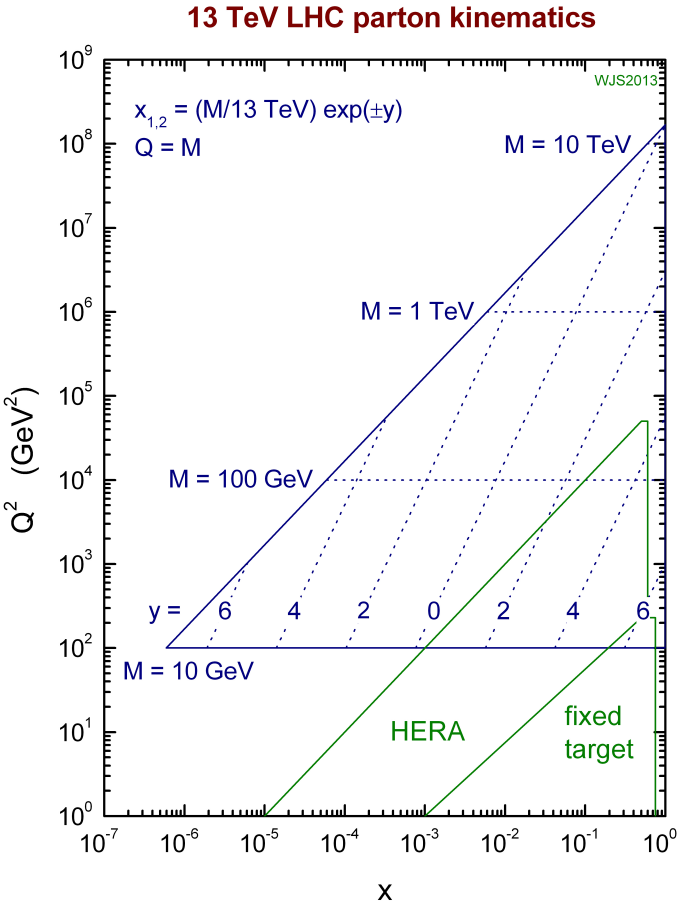


FIGURE 1.7 – Parton phase space covered by LHC experiments. x is the momentum fraction carried by the parton, Q^2 the scale of energy (taken as $Q^2 = M^2$) and y is the rapidity.

Main ideas and summary

So far we have made a few recalls of the main principles of HEP and proton-proton scatterings. The most important idea to keep in mind is the factorisation of the hadron and parton processes.

The context of this study having been drawn up, we can therefore move to the computation of the tree-level cross section of the Drell-Yan process and try to put a parameter that would represent the size of the quarks. The analysis of the Drell-Yan will be performed later, after a description of the CMS experiment at LHC.

2

Form factors in the Drell-Yan process

Now that the context of the present work has been presented, we can move to the study of the cross section of the Drell-Yan process: first, we are computing its tree-level cross section thanks to the Feynman rules; then, we are using the analogy with the study of the electron-proton scattering to insert form factors in the Drell-Yan cross section, to take into account the quarks' size.

In Chapter 1, we explained that the cross section of the hadron-hadron collision could be factoured as the integral of a PDF and of the cross section of the Drell-Yan process for a good choice of the energy scale. In this framework, assuming that the PDFs are known, the discussion can be restricted to the parton cross section.

2.1 Cross section at tree-level

We give here the main steps to compute the cross section of the Drell-Yan (DY) process at tree-level.

— We use the Feynman rules given in the Appendix of Peskin & Schroeder [25]. At

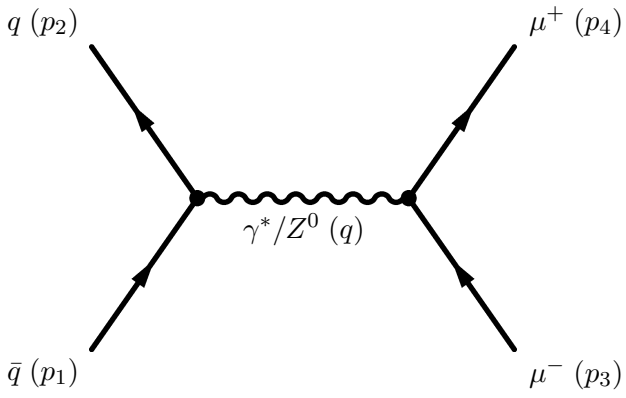


Figure 2.1 – There are two such diagrams at tree-level for the Drell-Yan process: the exchange of a photon γ and the exchange of a neutral weak boson Z^0. The solid, external lines represent the fermion fields, while the wavy line represent the exchanged boson. Note that $q^2 > 0$, as the photon is *virtual*; this is why it is noted γ^*.

tree-level, at energies of the order of the mass of the Z^0, there are two contributions to the DY process (Figure 2.1), and thus two scattering-matrix elements to compute: $\mathcal{M}_\gamma$ and $\mathcal{M}_Z$. The difference lies in the *propagator* $\frac{\eta_{\mu\nu}}{q^2-...}$ and in the couplings:

$\mathcal{M}_\gamma$ The propagator of the photon is $\frac{\eta_{\mu\nu}}{q^2}$ and the couplings are purely vectorial: $e\,Q_{q,l}\,\gamma^\mu$, Q being the charge of the considered fermion.

$\mathcal{M}_Z$ The propagator of the neutral weak boson is $\frac{\eta_{\mu\nu}}{q^2-M_Z^2+i\frac{\Gamma_Z}{M_Z}}$ (where $\Gamma_Z \frac{\hbar}{\tau_{\text{decay}}}$) and the couplings are either vectorial or axial, *i.e.* either $\frac{g}{\cos\theta_W}\,v_{q,l}\,\gamma^\mu$ or $\frac{g}{\cos\theta_W}\,a_{q,l}\,\gamma^\mu\gamma^5$, where the a and the v depend on the considered fermion.

The Dirac matrices γ mixes the components of the *spinors* by u and v to represent the fermion fields in the formulae (see Table 1.1). Applying Feynman rules, we have

$$i\mathcal{M}_\gamma = \bar{v}_1 i Q_q e\gamma^\mu u_2 (-i)\frac{\eta_{\mu\nu}}{q^2}\bar{u}_3 i Q_l e\gamma^\nu v_4 \tag{2.1}$$

$$\begin{aligned} i\mathcal{M}_Z = {} & \bar{v}_1 i\frac{g}{2\cos\theta_W} e\gamma^\mu \left(v_q + a_q\gamma^5\right) u_2 \\ & \times \quad (-i)\frac{\eta_{\mu\nu}}{q^2 - M_Z^2 + iq^2\frac{\Gamma_Z}{M_Z}}\bar{u}_3 i\frac{g}{2\cos\theta_W} e\gamma^\mu \left(v_l + a_l\gamma^5\right) v_4 \end{aligned} \tag{2.2}$$

— Remembering that $e = g\sin\theta_W$ due to the mixing of the electromagnetic and weak interactions in the SM, it is common to define the ratio of the propagators and the couplings

$$\mathcal{R} = \frac{1}{Q_q Q_l \sin^2 2\theta_W}\frac{q^2}{q^2 - m_Z^2 + iq^2\frac{\Gamma_Z}{M_Z}} \tag{2.3}$$

so that

$$\mathcal{M}_\gamma = Q_q Q_l e^2 \frac{1}{q^2}\left(\bar{v}_1\gamma^\mu u_2\right)\left(\bar{u}_3\gamma_\mu v_4\right) \tag{2.4}$$

$$\mathcal{M}_Z = Q_q Q_l e^2 \mathcal{R}\frac{1}{q^2}\left(\bar{v}_1\gamma^\nu(v_q + a_q\gamma^5)u_2\right)\left(\bar{u}_3\gamma_\nu(v_l + a_l\gamma^5)v_4\right) \tag{2.5}$$

— The total scattering matrix element is to be squared and summed on the final polarisation states, and as the beams used to get the data we are going to treat were not polarised, we have to average on the initial polarisation states too:

$$\overline{|\mathcal{M}|^2} = \left(\frac{1}{2}\right)^2 \sum_{\text{spins}} |\mathcal{M}_\gamma|^2 + 2\left(\frac{1}{2}\right)^2 \sum_{\text{spins}} \text{Re}\left(\mathcal{M}_\gamma \mathcal{M}_Z^*\right) + \left(\frac{1}{2}\right)^2 \sum_{\text{spins}} |\mathcal{M}_Z|^2 \tag{2.6}$$

After having used the properties of the traces (cyclic permutation of the γ matrices)

and of the spinors in the ultrarelativistic limit, one gets:

$$\begin{aligned}\overline{|\mathcal{M}_\gamma|^2} &= 8(Q_q Q_l e^2)^2 \frac{1}{q^4}\left((p_1\cdot p_4)(p_2\cdot p_3)\right. \\ &\quad\left. + \;(p_1\cdot p_3)(p_2\cdot p_4)\right)\end{aligned} \tag{2.7}$$

$$\begin{aligned}\overline{2\mathrm{Re}\,\mathcal{M}_\gamma\mathcal{M}_Z^*} &= 8(Q_q Q_l e^2)^2 \frac{1}{q^4}\left(c_+^{\mathrm{int}}(p_1\cdot p_4)(p_2\cdot p_3)\right. \\ &\quad\left. + \;c_-^{\mathrm{int}}(p_1\cdot p_3)(p_2\cdot p_4)\right)\end{aligned} \tag{2.8}$$

$$\begin{aligned}\overline{|\mathcal{M}_Z|^2} &= 8(Q_q Q_l e^2)^2 \frac{1}{q^4}\left(c_+^{Z}(p_1\cdot p_4)(p_2\cdot p_3)\right. \\ &\quad\left. + \;c_-^{Z}(p_1\cdot p_3)(p_2\cdot p_4)\right)\end{aligned} \tag{2.9}$$

where the cs are functions of the couplings.

— We sum and express the average scattering matrix element in terms of Mandelstam variables s, t and u: (Appendix A is devoted to Mandelstam variables)

$$\overline{|\mathcal{M}|^2} = \left(Q_q Q_l e^2\right)^2 \frac{1}{s^2}\left(c_+ u^2 + c_- t^2\right) \tag{2.10}$$

where

$$\begin{aligned}c_\pm &= 1 + 2\mathrm{Re}\,(\mathcal{R})(v_l v_q \pm a_l a_q) \\ &\quad +|\mathcal{R}|^2\left(\left(v_l^2+a_l^2\right)\left(v_q^2+a_q^2\right) \pm 4 v_l a_l v_q a_q\right)\end{aligned} \tag{2.11}$$

— Using FGR A.5 of Appendix A, the differential cross section in terms of Mandelstam variables then reads:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}t} = \frac{\left(Q_q Q_l e^2\right)^2}{16\pi s^4}\left(c_+ u^2 + c_- t^2\right) \tag{2.12}$$

with s considered as fixed and the three variables related to one another by $s+t+u \simeq 0$ in the ultrarelativistic limit.

For $l = \mu$, this cross section is the prediction of the SM to describe the parton scattering $q\bar{q} \longrightarrow \mu^+\mu^-$.

2.2 Cross section including the quarks' size

Now let us imagine that in the new region of the phase space opend by the LHC, the quarks are not seen anymore as pointlike particles. Then, in the CMS data of 2012 (that we are going to analyse) or in a few years with more statistics and at full power of LHC, there might be a deviation from the cross section. The DY process produced at high dimuon invariant mass offers the opportunity to probe the qqZ and $qq\gamma$ vertices at a higher resolution scale.

In order to do this, we shall argue by analogy to the probing of the proton, as the two processes are very similar: first in terms of diagrams at tree-level (Figure 2.2) and secondly in terms of deviation to the predicted cross section with pointlike particles. This motivates a short history of the electron-proton scattering before getting right down into business.

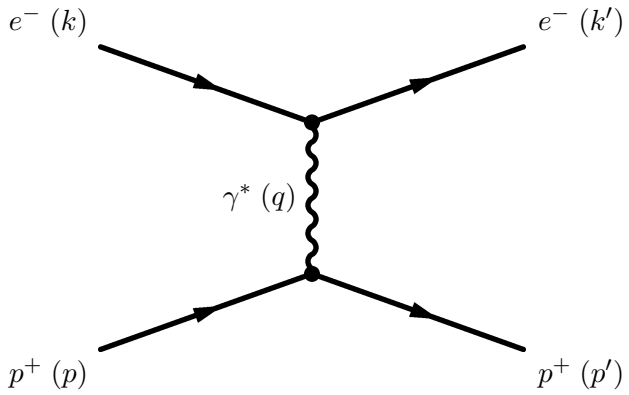


FIGURE 2.2 – Tree-level diagram of the electron-proton scattering with the proton considered as pointlike (valid until a few GeV). It is very similar to the Drell-Yan's, as the dynamics is the same. Note that $q^2 < 0$, as the photon is *virtual*; this is why it is noted γ^*.

2.2.1 History of the probing of the internal structure of the proton [1]

Studying the electron-proton scattering, a group of physicists working at the SLAC [2] during the fifties found that the Mott [3] cross section of the process was to be corrected by a global (form) factor G, that ought to be a function of the four-vector of the virtual photon:

$$\frac{\mathrm{d}\sigma_{\mathrm{exp}}}{\mathrm{d}Q^2} = \frac{\mathrm{d}\sigma_{\mathrm{Mott}}}{\mathrm{d}Q^2} \times |G(q)|^2 \tag{2.13}$$

If the energy scale Q is defined by $Q^2 \equiv -q^2$, such a form factor entails the decrease of the predicted cross section for large values of Q, meaning that *only a fraction of the electric charge is seen by the incoming electron.*

The form factor G can be interpreted as the Fourier transform of the charge distribution of the target $\rho(r)$, and is therefore called *electric form factor.*

$$G_E(q) = \int \rho(\mathbf{r}) \exp(i\mathbf{r} \cdot \mathbf{q}) \, \mathrm{d}^3 r \tag{2.14}$$

Assuming for simplicity that the distribution is spherically symmetric, it is often re-written

1. See Laurent FAVART's lectures [13], C.F. PERDRISAT, V. PUNJABI & M. VANDERHAEGHEN [22] for an article on the form factors and their measurement, and E. PEREZ & E. RIZVI [23] for an article on the structure functions.

2. The group of Robert HOFSTADTER, Jerome I. FRIEDMAN, Henry W. KENDALL and Richard E. TAYLOR from the Stanford Linear Accelerator Center, California.

3. Scattering of a pointlike, $\frac{1}{2}$-electron on a pointlike, spinless hadron. The computation of this cross section is done in the context of non-relativistic nuclear context, with *scattering amplitudes* in the Born approximation. The formalism is quite different from ours (see Charles JOACHAIN [18]).

in a more simple form:

$$G_E(q) = \int \rho(r) \underbrace{\exp(i\mathbf{r}\cdot\mathbf{q})}_{\simeq 1+i\mathbf{r}\cdot\mathbf{q}-\frac{1}{2}(\mathbf{r}\cdot\mathbf{q})^2+\ldots} \mathrm{d}^3r \tag{2.15}$$

$$= 1 + 0 + \frac{1}{2}q^2 \int \rho(r) r^4 \,\mathrm{d}r \int \mathrm{d}\phi \int \cos^2\theta \,\mathrm{d}\cos\theta + \ldots \tag{2.16}$$

$$\simeq 1 + \frac{1}{6}q^2 \int \rho(r) r^2 \,\mathrm{d}^3r \tag{2.17}$$

where we have first Taylor-developed, the terms of odd order vanishing due to the symmetry, and then used a trick of arithmetic to obtain the mean square radius $\langle r^2\rangle = \int \rho(r) r^2 \,\mathrm{d}^3r$ appeared. In such condition, G is just a function of the scale $Q^2 = -q^2$ and reads:

$$G_E(Q^2) = 1 - \frac{1}{6}Q^2\langle r^2\rangle + \text{corrections} \tag{2.18}$$

Note that $\langle r^2\rangle = -6\frac{\mathrm{d}G_E}{\mathrm{d}Q^2}$.

The dipole form factor was found to be in good agreement with the deviation:

$$G_E(Q^2) = \frac{1}{\left(1+\frac{Q^2}{\Lambda^2}\right)^2} \tag{2.19}$$

which corresponds to the Fourier transform of an exponential distribution. In the case of the electron-proton scattering, such a global form factor is valid up to energies of a few hundreds of MeV, with the parameter measured to $\Lambda \simeq 0.84$ GeV, which gives a radius of $\langle r^2\rangle = 0.74 \pm 0.24$ fm to the proton. With data obtained at an energy of 125 MeV, a similar *magnetic form factor* G_M was needed in order to take into account the spin effects. In the laboratory frame, the cross section reads:

$$\frac{\mathrm{d}\sigma_{\text{phys}}}{\mathrm{d}\Omega} = \frac{\mathrm{d}\sigma_{\text{Mott}}}{\mathrm{d}\Omega} \times \left[\frac{\epsilon|G_E(Q^2)|^2 + \tau|G_M(Q^2)|^2}{\epsilon(1+\tau)}\right] \tag{2.20}$$

with the factor $\tau \equiv \frac{Q^2}{4m_p^2}$ and the rate of the longitudinal polarisation of the virtual photon $\epsilon = \left[1 + 2(1+\tau)\tan^2\frac{\theta}{2}\right]^{-1}$. With different polarisations of the virtual photon, both form factors may be measured separately, and it was found that $G_M(Q^2) \simeq \mu_p G_E(Q^2)$ for the proton with the magnetic moment μ_p.

Later, in the end of the sixties still at SLAC, with electron beams up to 21 GeV, the deviation had to include inelastic effects, revealing a substructure to the proton: only a fraction of the whole energy $\sqrt{(p+p')^2}$ in the centre-of-mass frame is consumed. To write the cross section including the inelastic effects, one introduces *structure functions* $F_{1,2}(x, Q^2)$ (instead of form factors $G_{E,M}$) containing a kinematic dependence more in the Lorentz-invariant $x = \frac{Q^2}{2p\cdot q}$, called the Björken variable:

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}x\,\mathrm{d}Q^2} = \frac{4\pi\alpha^2}{xQ^4}\left[xy^2F_1(x,Q^2) + (1-y)F_2(x,Q^2)\right] \tag{2.21}$$

where $y \equiv \frac{p\cdot q}{p\cdot k} = \frac{Q^2}{sx}$ is another Lorentz-invariant quantity used in the framework of inelastic electron-proton scattering.

In both cases, form factors and structure functions mean that only a « fraction of the proton » is seen, but the physical meaning is somewhat different:

form factors Only a part of the electric charge of the proton contributes to the process, but the proton is still considered as a whole, and the interaction is elastic.

structure functions Only a part of the proton, regarded as a structure, is seen by the electron.

Naturally, they are related: imposing $x = 1$, the structure functions can be re-written as combinations of the form factors.

The study of the structure functions at energy scales $Q^2 \sim (10\,\mathrm{GeV})^2$ shows two remarkable properties:

— the structure functions are independent of Q^2 from a few GeV^2 (scale invariance), suggesting the proton was made of pointlike constituents;

— they are related by the *Callan-Gross* relation $F_2(x) = 2xF_1(x)$, suggesting that only one function is enough to describe the structure of the proton with an electromagnetic probe on a very short time scale.

Both could be explained in terms of the parton model introduced by Feynman in 1969, which was described in Subsection 1.2.1. Knowing that the partons are free only asymptotically and that they correspond to the quarks and the gluons, the existence of the QCD fluctuations will break the two properties:

— The partons are still considered as pointlike, but the higher the energy scale the more partons.

— Instead of moving collinearly, the partons may have some transverse momentum and the scattering may make the helicity flip, which requires one more structure function $F_L(x) \equiv F_2(x) - 2xF_1(x)$ (L as only the *longitudinal* helicity of the virtual photon may flip the spin of the parton).

Going at higher energy, while approaching $Q^2 \sim M_Z^2$, the Z boson must be taken into account. First it modifies the $F_{L,2}$; secondly, it is responsible for the parity to be violated, and a third form factor F_3 is defined:

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}x\,\mathrm{d}Q^2} = \frac{2\pi\alpha^2}{xQ^4}\left[Y_+\tilde{F}_2 + Y_-\tilde{F}_3 - y^2\tilde{F}_L\right] \tag{2.22}$$

where $Y_\pm = 1 \pm (1-y)^2$. The $\tilde{F}$ are the structure functions generalised for the weak couplings:

$$\tilde{F}_{L,2}(x,Q^2) = F_{L,2}^{\gamma\gamma}(x,Q^2) - v_e\kappa\chi(Q^2)F_{L,2}^{\gamma Z}(x,Q^2) + (v_e^2+a_e^2)\kappa^2\chi(Q^2)^2F_2^{ZZ}(x,Q^2) \tag{2.23}$$

$$\tilde{F}_3(x,Q^2) = \quad - a_e\kappa\chi(Q^2)F_3^{\gamma Z}(x,Q^2) + 2a_ev_e\kappa^2\chi(Q^2)^2F_3^{ZZ}(x,Q^2) \tag{2.24}$$

where $\kappa = \sin^2 2\theta_W$, and $\chi(Q^2) = \frac{Q^2}{Q^2+M_Z^2}$. The electroweak neutral current structure functions (NCSF) are directly related to the PDFs:

$$\left[F_2^{\gamma\gamma}, F_2^{\gamma Z}, F_2^{ZZ}\right] = x\sum_q \left[Q_q^2, 2Q_qv_q, v_q^2+a_q^2\right]\{q+\bar{q}\} \tag{2.25}$$

$$\left[F_3^{\gamma Z}, F_3^{ZZ}\right] = x\sum_q \left[2Q_qa_q, 2v_qa_q\right]\{q-\bar{q}\} \tag{2.26}$$

where the q and $\bar{q}$ are the quark densities. F_L contains the contribution of the gluons, giving a significant contribution only at high y (or equivalently small x).

2.2.2 Introduction of form factors in the Drell-Yan process

As in electron-proton scattering, we would like to insert form factors in the Drell-Yan cross section, to describe the size of the quarks just as it was done for proton at SLAC.

However, the situation is not directly transposable: the weak boson is probably non-negligible with respect to the photon in the case of the Drell-Yan process at the TeV scale whereas it was clearly negligible in the case of elastic electron-proton scattering at much lower energy. Yet, the general introduction of weak form factor in the matrix element leads to very long formulae for the cross section, which are rather hard to read (256 terms in the cross section, each prefactored with a complicate combination of the form factors...). The first question to ask is rather: « How many measurable quantities are there? »

Our first idea was to use 2.22 with $x = 1$ (as applied to subcomponent of quarks in turn) to restore elasticity and with $\tilde{F}'_L = 0$ to ignore internal fluctuations (a sort of Callan-Gross relation for possible sub-activity inside the quark):

$$\frac{\mathrm{d}\sigma'}{\mathrm{d}t} = \frac{(Q_q Q_\mu e^2)^2}{16\pi s^4}\left[\frac{\tilde{F'}_2 - \tilde{F'}_3}{2}t^2 + \frac{\tilde{F'}_2 + \tilde{F'}_3}{2}u^2\right] \tag{2.27}$$

where we would have « elastic » NCSF F' depending only on $s = Q^2$:

$$\tilde{F'}_2 \simeq F_2'^{\gamma\gamma} - v_e\kappa F_2'^{\gamma Z} + (v_e^2 + a_e^2)\kappa^2 F_2'^{ZZ} \tag{2.28}$$

$$\tilde{F'}_3 \simeq - a_e\kappa F_3'^{\gamma Z} + 2a_e v_e \kappa^2 F_3'^{ZZ} \tag{2.29}$$

as $\chi(Q^2) \longrightarrow 1$. A priori, the different terms in the $\tilde{F}'_{2,3}$ cannot be measured separately, as the prefactors are only couplings. There are only two measurable quantities.

Finally, we return to the initial interpretation of the form factors: the quarks only see partially the electric and weak charges of one another. Let us take the dipole form factors:

$$Q_q^2 \longmapsto Q_q^2 \times \frac{1}{\left(1 + \frac{Q^2}{\Lambda_E^2}\right)^2} \tag{2.30}$$

and similarly for the weak charge with a parameter Λ_W. We thus have an electromagnetic form factor and a weak form factor. Let us rewrite 2.12:

$$\frac{\mathrm{d}\sigma'}{\mathrm{d}t} = \frac{\left(Q_q Q_l e^2\right)^2}{16\pi s^4}\frac{1}{\left(1 + \frac{Q^2}{\Lambda_E^2}\right)^2}\left(\tilde{c}_+ u^2 + \tilde{c}_- t^2\right) \tag{2.31}$$

where the electric form factor has been written explicitly, while the weak form factor has been set inside the $\tilde{c}$s, as involved the parity violation. It is interesting to note that the weak form factor does contribute globally as well as the electric form factor, but this contribution may not be factoured out from the $\tilde{c}$s. Therefore, it will also be more difficult to measure, as we shall explain after the analysis.

As a first measurement, we shall thus consider no weak form factor in the coefficients $\tilde{c}s$ and only introduce a global electroweak dipole form factor in the cross section to deduce or limit the electric root-mean-square radius $\langle r^2\rangle = -6\frac{\mathrm{d}G_E}{\mathrm{d}Q^2}$ of the quarks without distinction between the flavours. The effect of such a global form factor is to lower the curve of the cross section mostly at high energy scales, as we shall illustrate in Chapter **??**. Later, if an effect is seen, it would be possible to measure the weak contribution thanks the forward-backward asymmetry.

Summary and perspectives

We first computed the parton cross section of the Drell-Yan process, knowing that it could be related to the hadron cross section with the factorisation theorem. Then we wanted to introduce a new parameter for the size of the quarks in the cross section.

The insertion of quarks form factors in the Drell-Yan cross section has been discussed by analogy with the electron-proton scattering. The situation is similar on both points of view on experimental situation and on Feynman diagrams. Having discussed how many form factors were needed in the cross section and we came to the conclusion that two of them were measurable. One, the electric, can only be global, while the other has to introduce the parity violation of weak interaction.

However, for simplicity, we shall only focus on one global form factor and extract or limit the root-mean-square radius of the quarks. For this, we are going to compare a simulation with real data: adapting the simulation with the global correction could help to increase the agreement between the simulation and the data, or to compute a new limit.

3

Experimental setup

The present pages are devoted to the description of the experimental setup that has been used to acquire the data that we are going to analyse. We will give an overview of the CERN, the LHC and the CMS experiment. Then the description will be directed towards the muon detection at CMS, as they play an important rôle in our analysis. The discussion of the detection of the particles will have consequences on the analysis, in next chapter.

3.1 CERN

3.1.1 Generalities

The *European Organization for Nuclear Research* is an international centre for experimental physics. HEP needs big and expensive experiments, involving hundreds or even thousands of physicists and engineers. Twelve European countries decided to associate their programmes in 1952. Nowadays, countries from all over the world have joined the team: twenty-one countries are full members and tens of others are just associated at different degrees.

The CERN is located near Geneva: some parts are in Switzerland, others in France. It is the biggest centre for HEP in the world. The internal budget of the CERN is of the order of a billion euros a year, around the price of one coffee per European citizen.

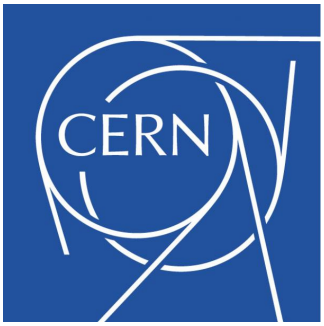


Figure 3.1 – Logo of the European Organization for Nuclear Research.

Most of the employees are technicians and engineers to who control and repair the machines, whereas most physicists rather stay in their universities and only come from time to time to visit, meet, or participate to an acquisition.

3.1.2 Experiments

There are many different experiments at the CERN, but most of them are *accelerator experiments*. An accelerator experiment consists of a scattering experiment such as Rutherfor's (Figure 1.1), where the beams of particles are first accelerated before scattering. In

practice, accelerating and scattering particles are two different jobs: there is an accelerator complex leading the beams into the different *detectors*, where the scattering occurs. One of these accelerators is called the Large Hadron Collider (LHC) and one of those detectors is called the Compact Muon Solenoid (CMS).

The data we are going to analyse in the second part comes from beams that were accelerated at LHC and that collided at CMS. So let us study them in more details.

3.2 Large Hadron Collider

3.2.1 Generalities [1]

One of the major experiments, and surely the most mediatic, is the LHC. The acronym means *Large Hadron Collider*. The project was born in the eighties and the story of its building is quite epic. The tinier the scale of distance, the higher the scale of energy: one of the goals was to build a machine that would open new regions of the phase space (Figure 1.7 shows the regions of phase space that recent experiments are covering in terms of parton kinematics). In 2012, thirty years later, the LHC started to work but not yet at full power.

A *collider* is a type of HEP experiment: it is an accelerator that speeds up beams of particles in opposite directions to make them collide in flight, by opposition to a fixed-target setup. The fixed-target experiments are easier to set up, and many are still used today (actually more than colliders). They present many technical advantages: the luminosity can be much increased thanks to the density of the target, the beams may be made of unstable or neutral particles, etc. However, in such an experiment, the energy in the centre-of-mass system grows as $\sqrt{E}$ where E is the kinetic energy of the incident beam, whereas in colliders, it grows as $\sqrt{E_1 E_2}$ where $E_{1,2}$ are the energies of the two beams. Colliders are more difficult and restricted but allow to go up much higher in energy.

Furthermore, there are two main kinds of colliders: *linear* and *ring* colliders. Both are limited in energy, the former by the length and the latter by the synchrotron radiation [2]. The LHC is a *large* ring, and the accelerated particles are heavy, as *hadrons*, so as to limit the synchrotron radiation as much as possible.

The circumference of LHC is of $26,659\,\mathrm{m}$ (see Figure 3.2). In practice, the particles are protons or lead nuclei [3]; here we are going to treat proton collisions.

3.2.2 Luminosity

The *luminosity* is a quantity used to measure the production of particles of an accelerator. For a circular collider with identical beams:

$$L_{\text{inst}} = fn\frac{N^2}{A} \tag{3.1}$$

where f is the revolution frequency (around $11\,\mathrm{kHz}$), n the number of *bunches* in the ring (around 2800), N the average number of particles in one bunch (around 10^{11}) and A the

1. For a few key figures of LHC, see the brochure [3] available online.

2. A particle turning in an electromagnetic field radiates light. The energy given by the magnet to accelerate the particle is partly lost as light.

3. Lead is the biggest stable nucleus that can be used in collider experiments, around 208 the mass of a proton, and lead-lead scattering should occur at centre-of-mass energies of $1148\,\mathrm{TeV}$ when the LHC is fully at work.

cross-sectional area of the beams (around $10^{-5}\,\mathrm{cm}^2$). This quantity is characteristic of the production of the machine per unit of time and is called the *instantaneous luminosity*.

Engineers try to optimise this quantity, while physicists who analyse the data rather use the *integrated luminosity*, as only the total quantity of data matters (Figure 3.3).

The total luminosity of the data with which we are working is:

$$L_{\mathrm{int}} = \int L_{\mathrm{inst}}\,\mathrm{d}t = 19.6 \pm 0.05\,\mathrm{fb}^{-1} \tag{3.2}$$

with an uncertainty of 2.6% ([7]).

The rate of collisions is

$$R = \sigma \times L_{\mathrm{inst}} \tag{3.3}$$

and thus the number of collisions is

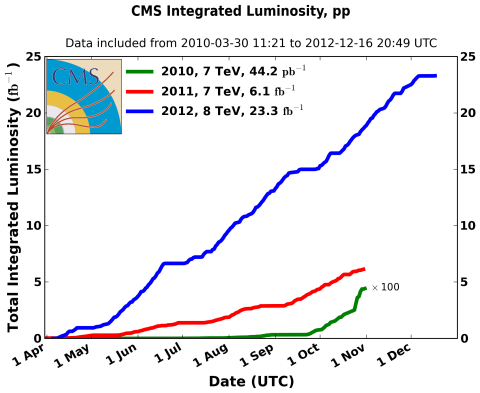
$$N = \sigma \times L_{\mathrm{int}} \tag{3.4}$$

The cross section σ may be total as well as differential. But if the LHC delivers a certain luminosity, some detectors may be defective, and the luminosity must be corrected by a factor of efficiency, that can evolve with time (Figure 3.3). We shall come back to this relation in the chapter of analysis.

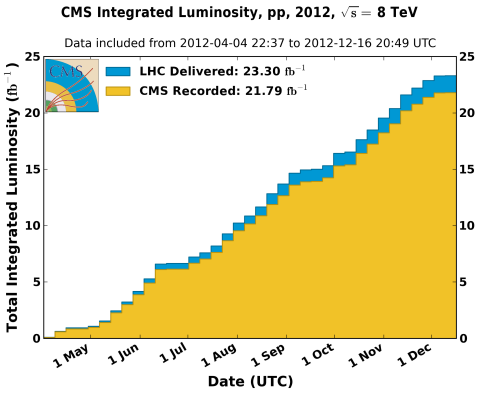
3.2.3 Working

First the particles are accelerated to their cruising speed in the ring; secondly they collide in the detectors. The two steps are very different.

FIGURE 3.2 – Comparison in size of the LHC complex and the airport of Geneva.



(a) Superposition of the evolution of the integrated luminosities during the different *runs* of LHC. One can see the stops of the machine and the remarkable increase of luminosity that was achieved. We can notice that the total integrated luminosity is higher than 3.2: some runs of the LHC are not taken into account in our data.



(b) This plot illustrates the difference between the delivered luminosity and the recorded luminosity. The correction is due to the efficiency of the detectors.

FIGURE 3.3 – Integrated luminosity at CMS and LHC.

3.2.3.1 Acceleration [4]

The techniques used to accelerate particles are relativistic but entirely classical; there are almost no quantum effect. Particles are only manipulated through electromagnetic interactions. There are several steps of acceleration before the beams of particles access to the LHC; An overview of the of the beam complex of acceleration of the CERN is given on Figure 3.4.

First they must be extracted from a hydrogen gas bottle and accelerated in smaller rings than the LHC. Then the particles must be grouped by *bunches*. Those steps are achieved thanks to cavity resonators.

The mean size of a beam that is travelling in the pipe is around the size of a knitting needle. Magnets are not only used to direct the beams but also to focus them. Indeed the envelope of this oscillation must be controlled to avoid losses due to collisions with the wall of the pipe, and when a beam approaches the detection zone, its diameter must be strongly reduced to increase the number of collisions per bunch crossing.

3.2.3.2 Detection

There are four big detectors at LHC: CMS, ATLAS, LHCb and ALICE. Each one is a big collaboration involving tens of institutes.

For proton-proton scatterings, the number of collisions per bunch crossing is around twenty; for data in 2012 (ours), bunches cross every 50 ns in the CMS and ATLAS detectors at an energy in the centre-of-mass system of 8 TeV (in 2015, it should reach 13 TeV for a bunch crossing every 25 ns). Consequently, the production of particles is extremely important: there may be hundreds of particles produced at each bunch crossing. To treat

4. See Django MANGLUNKI's lectures at CERN [20].

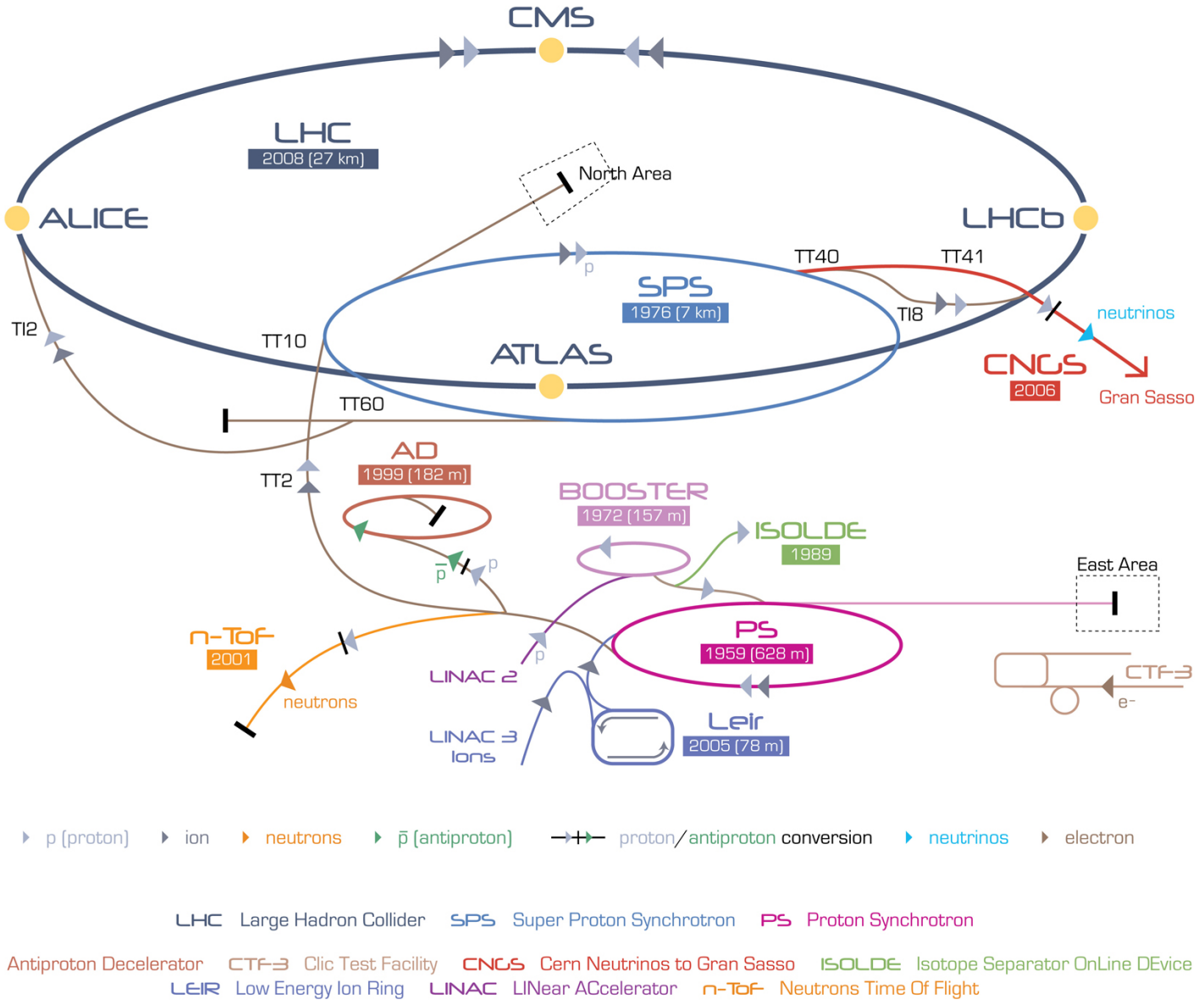


FIGURE 3.4 – CERN accelerator complex. The protons are extracted from a simple hydrogen bottle based at the Linac 2. They are collected in the Proton Synchrotron Booster that accelerate them from 50 MeV to 1.4 GeV, then to the Proton Synchrotron that speeds them up to 26 GeV and separates the beams in smaller bunches, then to the Super Proton Synchrotron that makes them go to 450 GeV. And finally, they can go to the Large Hadron Collider, whose nominal energy per beam should be 7 TeV. Lead nuclei can be as well accelerated, but start from the Linac 3 and go through the Leir.

such an amount of particles and to be precise, the detectors need to be gigantic: CMS is 18 m high (Figure 3.6), and ATLAS more than 20 m!

Let us focus on CMS. The main principles of detection hold for the other experiments.

3.3 Compact Muon Solenoid

3.3.1 Generalities

CMS is the one experience at CERN in which the ULB and the VUB are involved, as well as all the Belgian groups. The technical proposal was approved in 1996 and its construction officially began in 1997. The design goals were defined as follows:

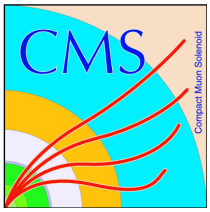


FIGURE 3.5 – CMS logo.

— very good and redundant muon system of detection,

— very good electromagnetic calorimeter,

— very good tracker,

— financially okay

by order of priority. Table 3.1 gives a few key figures as a guide to CMS.

CMS	Comparison with an A380 airplane
100 m underground	$1,25\times$ of the length
21 m long	$\frac{1}{3}\times$ the size of the fuselage
15 m diameter	height
14,000 T	280 airplanes
4300 people	9 airplanes

TABLE 3.1 – A few key figures (cf. [6])

The acronym stands for *Compact Muon Solenoid*:

Compact means that most of the devices are placed inside the magnet, the trackers and the calorimeters except for the muon chambers (this is a technical exploit to build such a titanic magnet);

Muon means that the detector has been designed to be very sensitive to the muons, as they are involved in some important channels;

Solenoid stands by opposition to *toroidal*, which is one of the two possible geometries for the magnetic field to be parallel to the beams in order to act only on the product particles.

As its name might suggest, CMS is designed for particles scattering in the Centre-of-Mass System, so as to significantly increase the energy of scattering in the laboratory frame (see above).

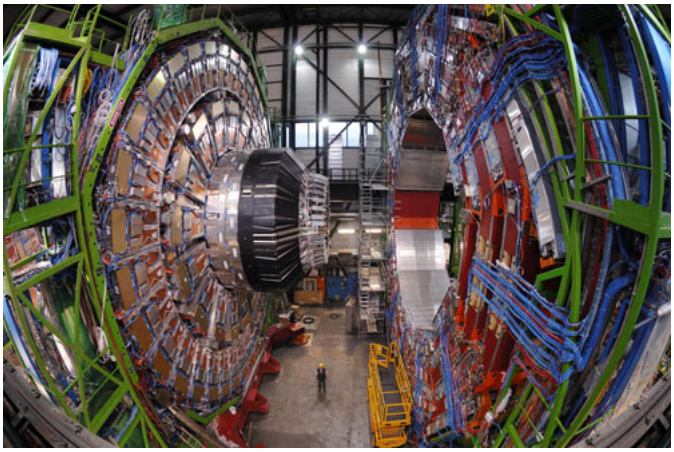


FIGURE 3.6 – A view of the detector under construction.

3.3.2 Detection [5]

3.3.2.1 General principles

The particles seen in the detectors are considered as classical particles. An event is described by the list of the particles, their momenta and the angles of emission (see 3.7 for the conventional coordinate system) of the different processes that have happened at the interaction point. The detector is designed to measure all this.

Most of the particles created at the interaction point do not live long enough to be directly detected. The particles to be discovered or studied are seen through their decay

5. See S. TAVERNIER [26] for more details on the interaction of particles with media and the Volume II of the CERN manual on LHC [5] for more details on the detectors.

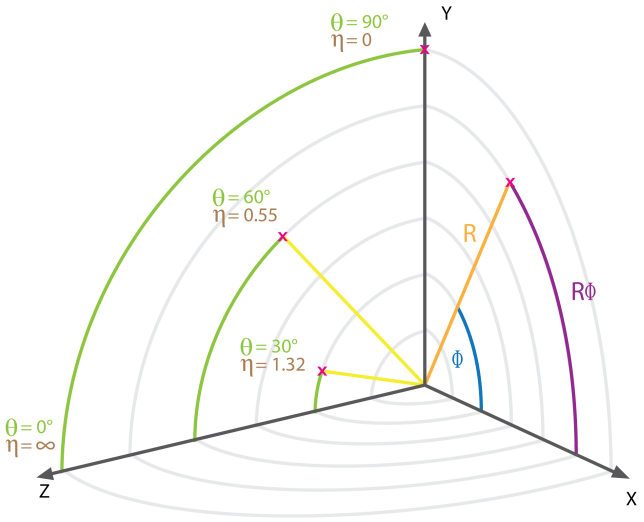


FIGURE 3.7 – Conventional coordinates at a collider. z is the axis of the beam, $\eta = -\ln\tan\frac{\theta}{2}$, ϕ the azimuthal angle. The pseudorapidity is an equivalent quantity to the polar angle θ but as asymptotically equal to the rapidity, it is almost Lorentz invariant.

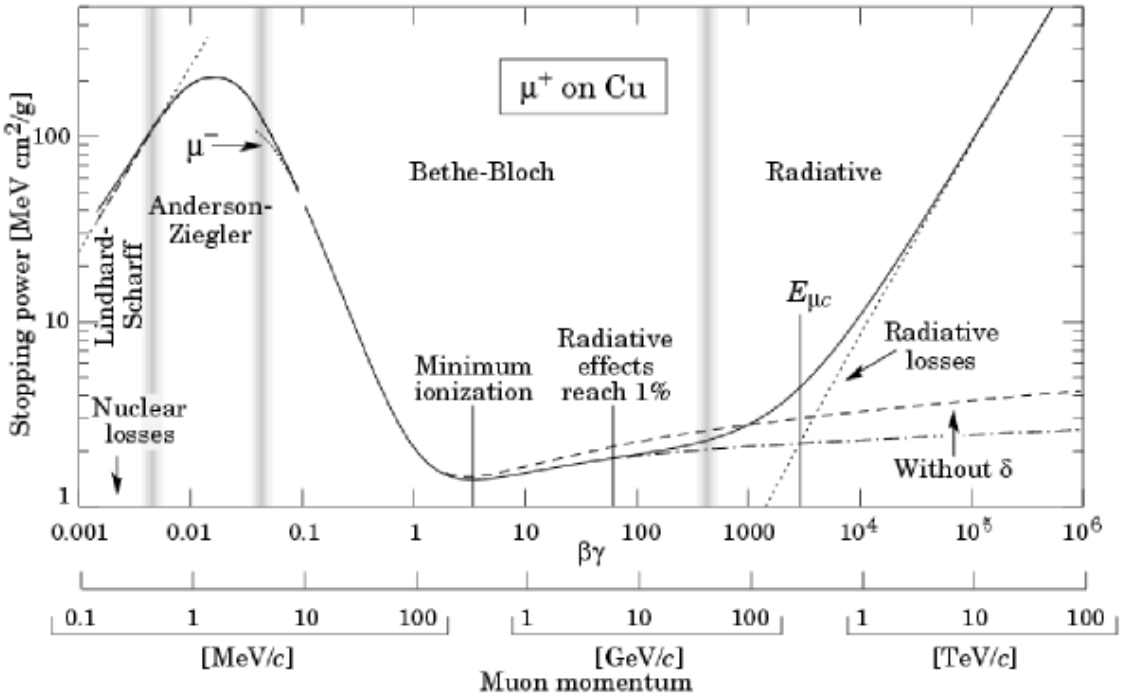


FIGURE 3.8 – Effect of muon passing through a medium as a function of its energy. At CMS, the interaction that matters are between 10 GeV and 1 TeV: in such a range it can been considered at the *minimum potential ionisation* (MIP).

products, which are: photons (γ), electrons (e), muons (μ), charged pions (π), charged and neutral kaons (K), protons (p), neutrons (n), and their antiparticles. The neutrinos do not decay but interact too weakly to be detected in accelerator experiments.

Those particles' interactions with media are very well-known:

Photons
- — *photoelectric effect*,
- — *Compton effect*
- — and *pair production*.

Only the pair production is significant.

Electrons and muons
- — *Bremsstrahlung*
- — *ionisation*,
- — and *multiple scattering*.

For electrons, only the Brehmsstrahlung is significant and for the muons that are heavier, Bremsstrahlung is negligible for energies up to 1 TeV (See Figure 3.8).

Protons, charged pions and charged kaons As muons, they should interact because of their charge, but this effect is negligible. They rather interact thanks to nuclear effects.

Neutrons, neutral kaons Only have nuclear interactions.

To complicate the situation, mesons may also decay in flight: for instance, the charged pions may weakly decay into muons and fake muon events.

In practice, the particles produced at the interaction point go through the *tracker*, and should stop in one of the *calorimeters*, except for muons that continue through the magnet and through the *muon chambers* as well as neutrinos that are not seen:

tracker The aim is to reconstruct the trajectory of a produced particle according to the curvature of its trajectory; only charged particles are seen by the tracker. Their trajectories are curved thanks to the magnetic field, and the curvature of a trajectory can be related to its transverse momentum ($p_T = qB\rho$ where ρ is the curvature radius of the charged particle).

calorimeter The calorimeter stops the incident particles and measures their energy deposits. If possible, a deposit will be associated to one of the charged particles seen by the tracker at reconstruction. Otherwise, it will be considered as the deposit of a neutral particle.

muon chambers Calorimetry cannot be used to detect muons. Muon chambers are a kind of external tracker designed especially for muons.

Now that we know the main principles of detection, we can apply them to understand the subdetectors of CMS.

3.3.2.2 Application at CMS

Figure 3.9 shows the ideal path of a particle that comes from the interaction point and goes through the detector.

Tracker The aim of the tracker is to reconstruct the trajectories of the charged particles. The trajectory may then be used to determine the charge, the momentum and the interaction in which is it involved, as several interactions happen in the central part of the detector at the same bunch crossing (*pile-up*).

In a semi-conductor, the valence band of the electrons is just below the conduction band: the excitation of a valence electron makes it go up in the conduction band. A detector is made of a *npn*-doped junctions made in silicon, in which an electric potential of a few hundred Volts is applied to optimise the ratio signal over noise. It is nowadays the fastest technology for tracking.

At CMS, there are two trackers:

silicon strip tracker From $R = 55\,\text{cm}$ to $R = 110\,\text{cm}$, the tracker is constituted of layers of doped semi-conductor detectors. The strips measure $25\,\text{cm} \times 180\,\mu\text{m}^2$ and are arranged in stereo to get the two components of the coordinate.

pixel tracker Around the pipe, from $R = 20\,\text{cm}$ to $R = 55\,\text{cm}$, as the resolution must be very accurate and the pixel cells measure $100 \times 150\,\mu\text{m}^2$ to be as precise as the silicon strip tracker, but the principles of working and detection are the same. The pixels are made into materials that are resistant to the important radiations.

Tracking goes up to $|\eta| = 2.5$. The interaction points will be associated by reconstruction algorithms to tracks and to energy deposits in the calorimeters or signals in the muon chambers.

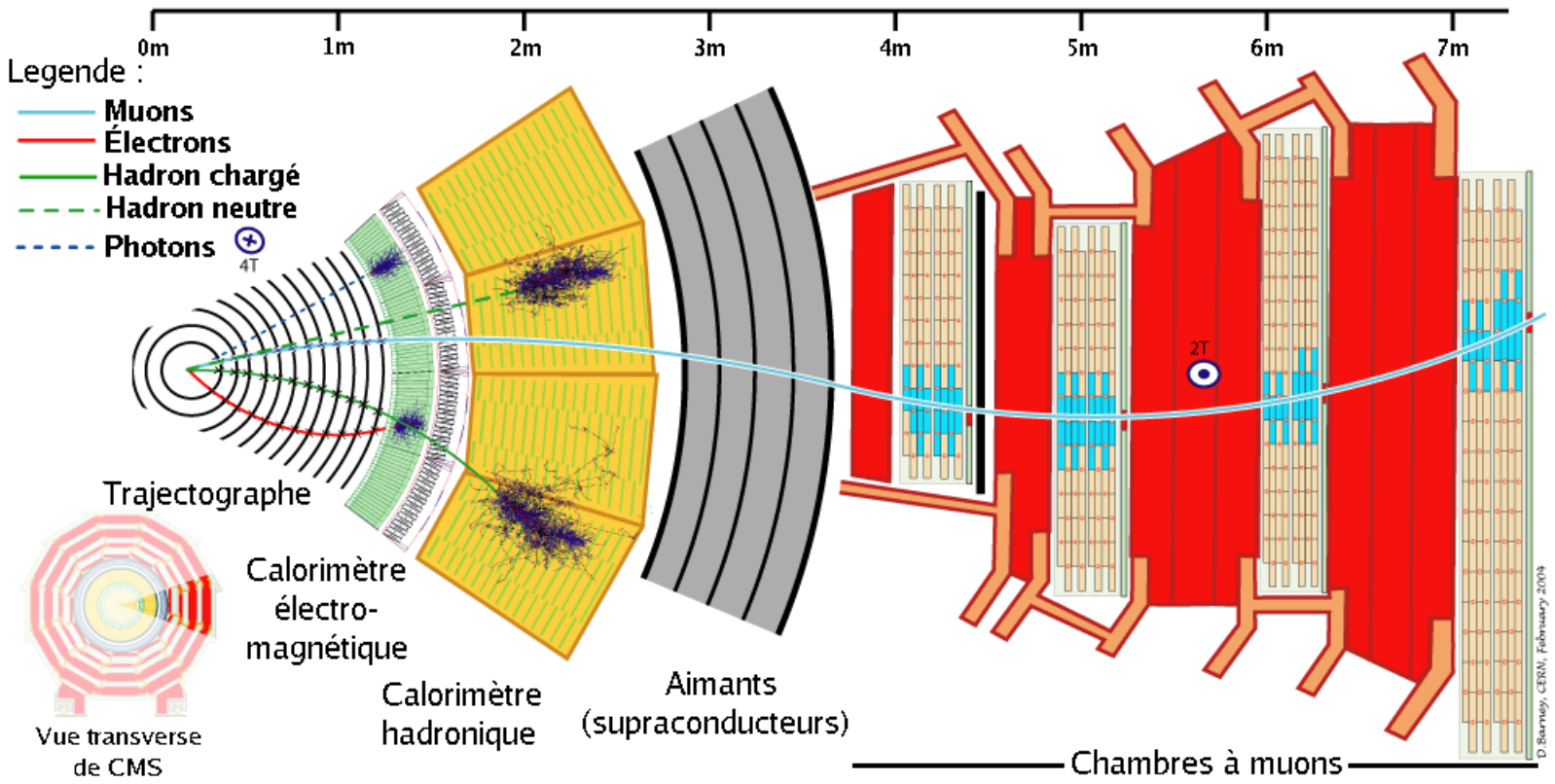


FIGURE 3.9 – Diagrammatic view of the barrel of the detector (endcaps have a similar structure). The ideal trajectories of the different types of particles have been drawn. A particle coming from the interaction point first goes through the tracker; it is seen only if it is charged. Then it goes through the electromagnetic calorimeters, that should only stop photons and electrons. If it has not been absorbed yet, the particle continues through the hadronic calorimeter, that should stop hadrons. Only a muon or a neutrino should still survive after those steps. It goes through the magnet and the steel structure, and goes through the muon chambers. It may hit the detector if it is a muon. A neutrino always goes out.

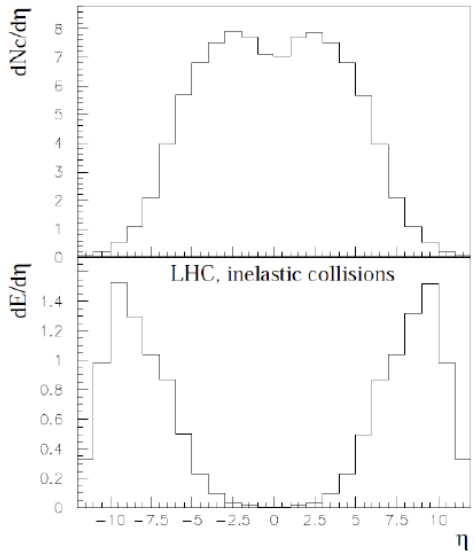


FIGURE 3.10 – Simulation of the number of particles and energy as functions of the pseudorapidity. Most of the activity remains in the tube. Calorimeters only sees up to $|\eta| < 5.6$ and trackers and muon chambers only up to $|\eta| < 2.5$.

Calorimeters There are two kinds of calorimeters: *electromagnetic* calorimeter (ECAL) and *hadronic* calorimeters (HCAL). Indeed, leptons and hadrons do not have the same interactions with matter, and it is better to separate their detection.

ECAL The ECAL is made of a lead tungstate crystals that makes the electrons and positrons radiate, and that makes the photons produce electron-positron pairs; this scintillator medium is transparent for photons of a certain wavelength which is subsequently collected by photodetectors.

HCAL Similarly, the HCAL exploits the nuclear interactions of the hadrons to produce lighter hadrons and photons; it is made of alternate layers of absorbers and scintillators, the former used to slow down the hadrons and the latter to collect the produced photons.

Muons are not concerned by the calorimeters and are supposed to go through without any loss of energy as they are heavy enough to neglect Bremsstrahlung and have no nuclear interactions.

Muon Chambers The activity is somewhat different according to the angle of emission, as Figure 3.10. Thus different technologies should be used for the muon chambers. They work up to $|\eta| = 2.6$ as further, the activity it too important to track the particles (as the simulations show, on the same figure).

There are three types of muon chambers. They all are gaseous detectors: the principle is to fill a volume with a gas and an electric field, so that an incoming particle may ionise it and create an electric signal. But there are different types of gaseous detectors, according to the handling of the electric signal. In the case of the muon chambers:

Cathode Strip Chambers CSCs are based on the same principle as multi-wire proportional chambers. They consist of arrays of anode wires and cathode strips arranged perpendicularly within a gas volume. When a muon goes through, it

ionises the gas: its position is determined by the intensity of currents induced on the strips by the motion of the electrons and the ions. There are six such layers in the endcaps of CMS for as much as 540 CSCs.

Drift Tubes DTs are placed in the barrel ($|\eta| < 1.6$). Its dimensions are $4\,\text{cm} \times 2\,\text{m} \times 2.5\,\text{m}$ and an anode wire is stretched in it. DTs are put in layers perpendicularly to the trajectory of a muon. When a muon goes through a drift tube, it ionises the gas and the freed charges move in the electric field and induce a signal. The delay of the signal is used to determine the position of the muon inside the tube. As a DT gives only one coordinate, three groups of up to sixty DTs are arranged successively: the two extreme measure its perpendicular coordinate while the middle one measures its coordinate parallel to the beam. All in all, there are 1400 DTs at CMS.

Resistive Plate Chambers RPCs consist of an anode plate and a cathode plate separated by a gas volume. The material that are used are highly resistive, so that the chamber works with very high electric fields ($E \simeq 50\,\text{kV/cm}$). There are both 480 RPCs arranged in four concentric cylinders in the barrel ($|\eta| < 1.6$) and 288 in four disks in each endcap ($|\eta| > 1.6$). They are used to get a first approximation of the momentum of the muons in real time but their spatial resolution is moderate.

The muon chambers are combined with the tracker to reconstruct the trajectories of the muons and improve the resolutions on their transverse momenta.

The geometry of the detector with its coordinates is shown on Figure 3.11.

3.3.3 Reconstruction

To reconstruct an event the first thing to do is to use the tracker information to reconstruction the tracks. Then, the tracks must be associated to reconstructed energy deposits in the calorimeters or to impacts in the muon chambers. If a deposit is not associated with any track, it is associated to a neutral particle.

The general algorithms of reconstruction of the whole event are very complicate subjects: it involves magnetic effects, the pile-up (the superposition of the many events that happen at the same bunch crossing), the superposition of the trajectories coming from the same events, decays in flight, the detectors efficiencies, etc. Even determining the nature of the particles that has left a track is not always obvious.

3.3.3.1 Tracking algorithm

Kalman filter A simplified version of the tracking algorithm, called *Kalman filter*[6], is worth describing. Consider a particle, having just passed through the successive layers of the detector and left impacts. The procedure is iterative. In words:

1. One starts with the measured position on the first layer.
2. The equations of motion are used to predict the position on the second layer. It should be close to the measured position, but possibly different. A compromise, tempered by the resolutions of the measurement and of the prediction, is used to determine the most likely position, called *filtered position*.

6. See the article [15] where the method is fully described.

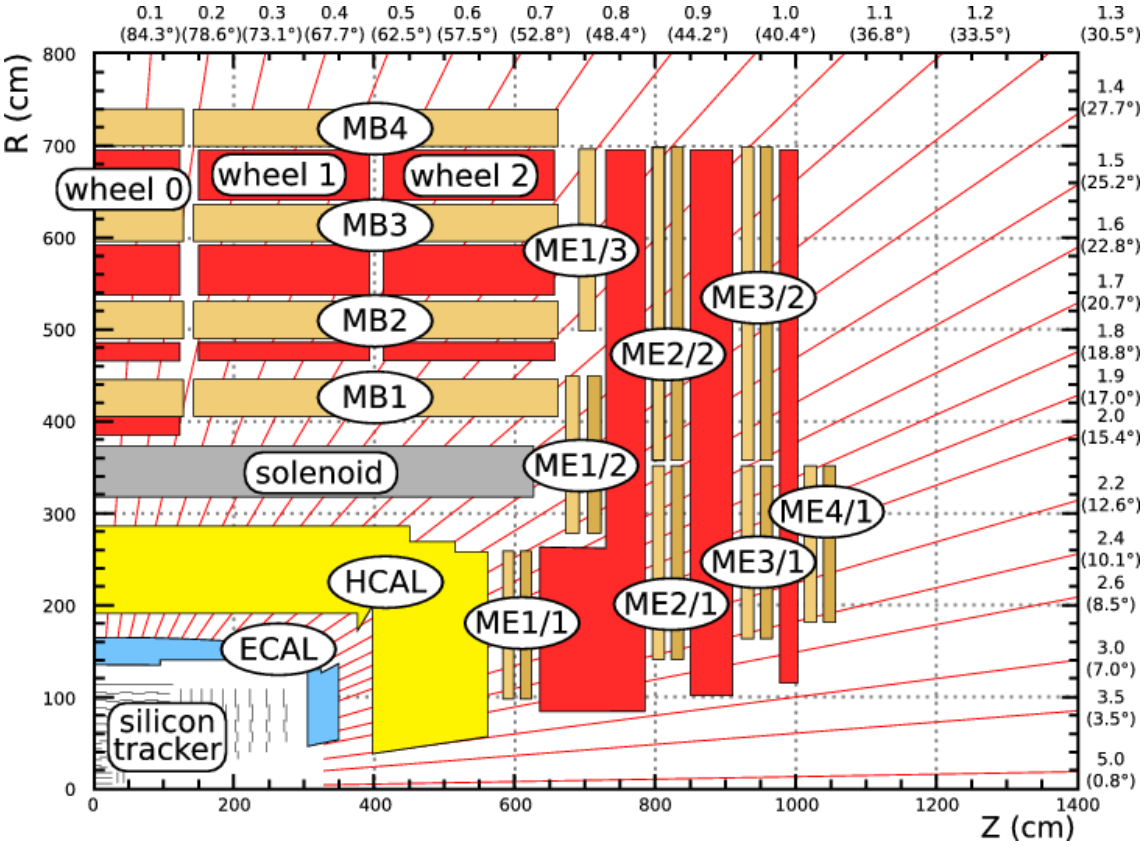


FIGURE 3.11 – Geometry of a quarter of the CMS detector and notably the muon chambers: ME stands for « muon chamber endcaps » (CSCs + RPCs) and MB for « muon chamber barrel » (DTs and RPCs).

3. The filtered position on the ith layer is determined by a weighted average of the measured position on the layer and the predicted one from the previous layers.

Many refinements exist.

In the simple case of the Kalman filter for particles going through infinitely thin, parallel layers with no dead angle and perfect efficiency [7], the resolution of the detectors, due to multiple scattering, limits to about ten the number of layers. In more realistic cases, there are more sources of uncertainties: the efficiency of the detectors, their thickness, etc. This still reduces the resolution.

Resolution on the momentum Usually, the resolution on the momentum is determined by the following formula:

$$\left|\frac{\sigma_p}{p}\right|^2 = \left(\frac{a_n p_T}{0.3BL_T^2}\right)^2 \sigma_X^2 + \left(\frac{0.06}{B\beta}\right)^2 \frac{1}{X_0 L_T \sin\theta} \tag{3.5}$$

where $a_n = \sqrt{\frac{720}{n+4}}$ is a typical parameter of the current configuration (n is the index of the layer), B is the strength of the magnetic field in Tesla, L_T the length of the tracker in

7. Realised at a practical training with Pascal VANLAER.

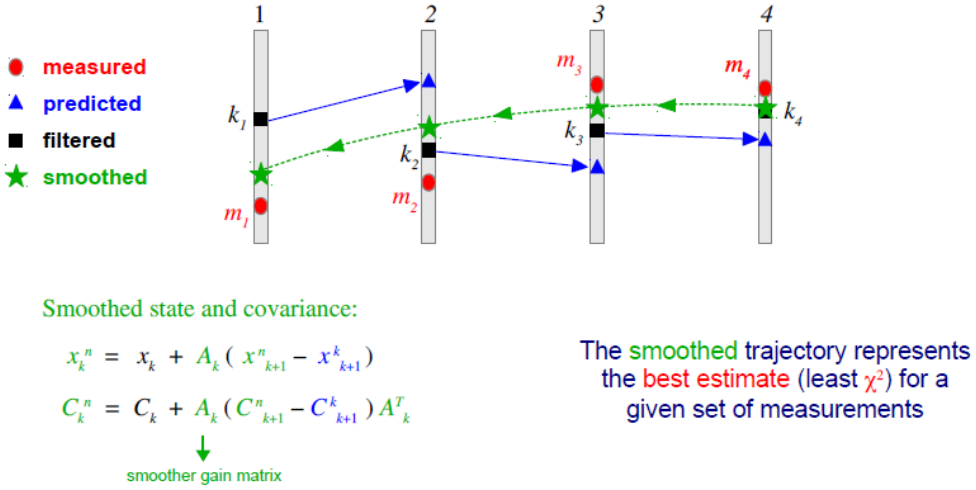


FIGURE 3.12 – Illustration of Kalman filter. x is the position, k the iteration; C is the covariant matrix; A is the *gain matrix* and operates Kalman filter.

metres, p_T the transverse momentum in GeV/c, θ the polar angle, σ_X the spatial resolution of the devices in metres, and X_0 the characteristic length of the material. It means that at low energy, the momentum resolution is limited by the multiple scattering while at high energy, it is limited by the determination of the curvature (the uncertainty from the polar angle is negligeable).

Finally, the number of layers in the tracker was fixed to ten. Figure 3.13 shows the evolution of the relative momentum resolution that can be reached in practice thanks to Kalman filter.

Muon tracks In the case of the muons, the muon chamber contribute to the tracking. This is achieved in several steps: *local* reconstruction inside a muon chamber, *standalone* reconstruction by combination of the whole muon system, *global* reconstruction by combination of the muon system and the tracker. Sometimes, a muon seen by the tracker does not leave enough impacts in the muon system to reconstruct a track with this algorithm, but the extrapolation of the track to the muon chamber can be used to identify the track as that of a muon.

3.3.3.2 Identification

Electrons and positrons are reconstructed thanks to the association of a curved track and a deposit of energy [8].

Photons are identified only thanks to the energy deposits in the ECAL.

Hadrons that are produced from a gluon radiation are most of the time emitted in the same direction: thus the ones that are suspected to come from the same radiation are grouped as *jets*.

In the case of the muons, the identification is quite easy for most of the case as they must have been seen in the muon chambers. Several criteria have been defined:

8. Unfortunately, there are important backgrounds due to the pion decays that superpose to the signal. Nothing similar exists for the muons, which explains why we have chosen the channel for our analysis.

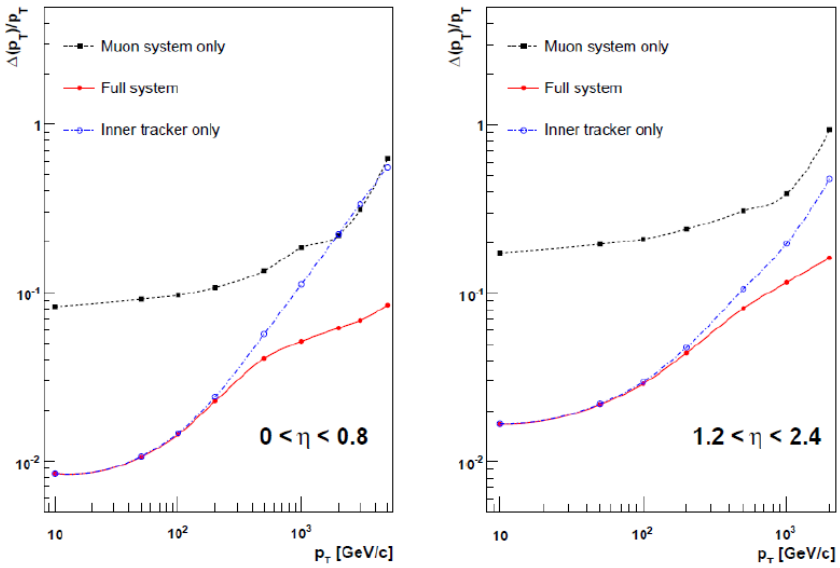


FIGURE 3.13 – Resolution of the transverse momentum of the muons. The combinaton of the trackers and of the muons chambers clearly increases the resolution on the transverse momentum.

tight	optimised for weak boson studies, it means that the muon must have been reconstructed from both the tracker and the muon chamber, and must have been seen in two different places of the chambers;
soft	optimised for b quark decays, the muon must have been tracked in seen in one CSC or in one DT (lower value of p_T are then accepted);
loose	optimised for multiple muons events, the muon is identified after the full combination of the tracker, the muon chamber and even the calorimeter;
high-p_T	optimised for studies where $p_T \geq 200\,\text{GeV}$.

The choice and the purpose of these algorithms are dictated by the *Muon POG*, *i.e.* Muon Physics Object Group, which is the CMS group that work on the reconstruction of muons. In our analysis, we shall use the tight muons (we shall repeat it in due time).

3.3.4 Trigger system

The LHC is designed to run at a very high collision rate, so high as 20 MHz; in other words the bunches should follow one another every 50 ns. Each bunch crossing provides ten to forty collisions (pile-up).; let this be multiplied by the hundreds of particles that have been produced for each collision, knowing that one particle must be detected in several parts of the detectors, which must be combined to reconstruct its trajectory and determine its nature: no computer is able to treat so many data at such a rate.

That is why an important *trigger system* has been set up. Only a very few events are accepted, at a rate that hard disks are able to follow: only one event out of a few hundred thousands! At CMS, there are two levels:

Level1 This is a hardware system: only thanks to the calorimeters and the muon chambers, the Level1 accepts or rejects an event (a bunch crossing). This must be done online. The stream delivered by the Level1 trigger is of order of 100 kHz.

High Level The High Level Trigger is a software system: It is divided is several steps:

- Level 2 pursues the analysis with the information from the calorimeters and from the muon chamber.
- Level 2.5 combines the information at level 2 with tracks in the pixel tracker to delimit region of extrapolation in the strip tracker.
- Level 3 combines level 2.5 with the strip tracker.

The stream delivered by the High Level Trigger is of order of 100 Hz.

The system *triggers* on the basis of certain criteria; one example of criterion is the « mu17mu8 », which means that the system triggers when it sees a first muon with transverse momentum of 17 GeV and a second muon with transverse momentum of 8 GeV. This criterion is the one that is used to select the data that we are going to analyse. The rate of such events needs no rescaling as its rate is low enough (sometimes, a criterion should activate the system more often than it can register them; in such cases, the measured data must be rescaled to do as if the detector had seen them all).

The whole event is registered, including every single electronic signal. It is sent to the computing centre of the CERN, called the Tier 0, that in turn will share the data with other computing centres called Tier 1 and 2, all over the world in member countries. The VUB and the ULB provides one of the Tier 2. There, the data will be available for users.

3.3.5 Storage

When the trigger accepts an event, it is stored entirely at the Tier0, the computing centre of the CERN in a format called FEVT (FullEVenT). Each raw event weights a few MB. But they are not ready for use: the raw data must be converted into physical variables. The mere reconstruction algorithm of the trigger does not take into account the inefficiencies of the detectors and does not reconstruct necessarily the full event.

The offline reconstrution is achieved by CMSSW, the official framework of CMS. It proceeds to a first reconstruction and produces RECO files (reconstructed level), weighing around 500 MB, and AOD (Analysis Object Data), weighing around 50 MB per event; the former contain many variables describing the working of the devices, while the latter only contains the useful variables for data analysis, as we are concerned. From those files, a second reconstruction is worked to classify the events; several algorithms exist and the one that we use in this master thesis is the *Particle Flow*.

The events are grouped by period of acquisition in *JSON* files. Indeed, as the instantaneous luminosity varies because of the dispersion of the beams for instance; so each JSON file is given a luminosity and some technical characteristics of the acquisition. The downloading and the merging of the JSON files from the computing centres is applied with a platform called *Crab*.

Summary

Detection is a vast subject.

The particles have followed a long procedure of extraction and acceleration until they are grouped by bunches and ready to collide; those steps do not matter in our analysis and we only need to know that the centre-of-mass energy of the proton-proton scattering and the total luminosity of the sample of data to analyse.

Bunches cross and many particles are produced in the pipe; most of them decay. The remaining ones, whose behaviour in media is very well-known, travel through the devices of the detector: first the particles go through the tracker; the aim is to reconstruct the trajectories of the charged particles with the Kalman filter (and thus deduce the momenta); it must be precise and the energy lost must be negligible. Secondly, the particles pass through the electromagnetic calorimeter, which stops the photon thanks to pair creation effect and stops the positrons and electrons thanks to Bremsstrahlung; the energy deposits should be their initial energy. Thirdly, the remaining particles go to the hadronic calorimeter; the same procedure applies to the hadrons, but adapted to their nuclear interactions. Besides, the muons are the only particles left that can still be detected; this is achieved by the muon chambers.

The L1 trigger decides wether an event is interesting to keep or not only on the basis of the calorimeter and the muon chambers; if okay, the HLT refines the selection of calorimeters and reconstructs the tracks. After the selection (a few tens per second at most) the data is sent to the computing centres of the CERN and its partners (Tier 0, 1, 2), and parsed in files that are handleable by the average physicist.

4

Analysis

From the trigger that we described a few pages ago to the final plots, there is a long way to go. In this chapter we are detailling the different steps of the analysis, including the treatment of the systematic uncertainties, of CMS data of 2012 corresponding to the dimuon production in proton-proton scatterings at 8 TeV of energy in the centre-of-mass frame.

Before describing the analysis, it is worth mentionning Monte Carlo simulations. Indeed, they play different rôles in HEP analysis, as analytic computations are either impossible or too time-consuming. As far as we are concerned, we shall use simulation to describe the detectors effect and to compare the data with our knowledge.

Then we are moving to the analysis, describing each step:

— getting the files of data and simulations from the computing centres of the CERN,

— choosing the phase space,

— subtracting the background,

— inverting the effect of the apparatus (*unfolding* the data),

— and finally computing the systematic uncertainties.

We are following the evolution of the differential cross section in the dimuon invariant mass spectrum step by step. Indeed, in the next chapter, we shall use it to discuss the quarks' size. But many other plots will be drawn as well, such as the transverse momentum of the muon pairs, the angle with the beam axis, etc.

4.1 Raw data and simulation

The raw data as well as the simulation are to be treated in parallel. We have described the acquisition in Chapter 3; now we are having a few words on simulations.

4.1.1 Simulation

There are many different uses for Monte Carlo simulations. Each specific use will be detailed in due time. For now, let's just explain what a simulation is made of.

There are two different levels of simulation:

generated level Would-be-physical distribution of four-vectors. Such a simulation is doable by any personal computer. The most used [1] programs of simulation of physical events are *MadGraph*, *Sherpa*, *Pythia*, etc.

reconstructed level Would-be-measured distribution of the physical variables computed from the electronic measurements. It is obtained from the generated level simulation after it has gone through a simulation of the detector. It is a very time-consuming treatment that cannot be achieved by common computers. The simulation of one event might take ten minutes at the CERN computing centres! The most used program of simulation of the detectors is *Geant4*.

Both will be useful: the former at the end of the analysis to compare the matching of the data with the simulation, and the latter will be used to extract the Drell-Yan signal from the whole CMS sample and to treat the non-physical effects (*i.e.* effects due to the detectors) on the distributions.

On Figure 4.1, the global steps of the simulation are represented in comparison with the global steps of real data acquisition. We have already described some of those steps in Chapter 3; now let us focus on the simulators.

4.1.1.1 Event generator

An event generator such as MadGraph takes a process as input and some parameters such as the value of the couplings, the number of events, the PDF to use, etc. In our study, we have asked $pp \longrightarrow \mu\mu X$ (where X stands for the hadronic production) to the MadGraph

1. At CMS, actually, but there is no limitation, it is only historical.

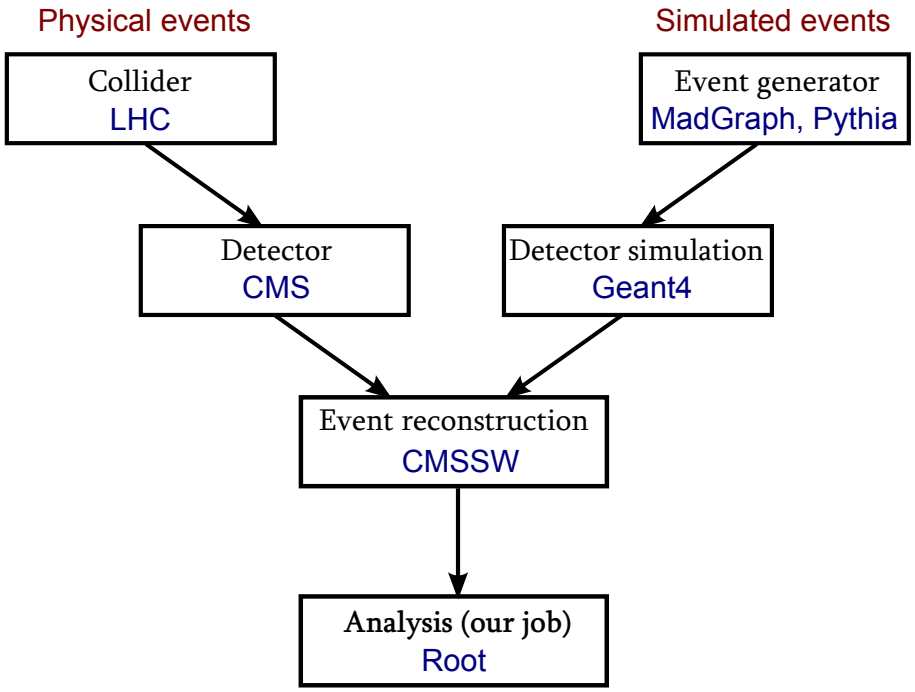


FIGURE 4.1 – Diagram with the different programs used from the acquisition and simulations of the samples to the analysis.

generator. Based on the different Feynman diagrams at tree-level, this generator produces *events* and gives a distribution of four-vectors as output.

On this basis, the generator Pythia 8 first adds the *parton shower* resulting from the radiation from both initial and final states on time scale of order $\tau \sim \frac{1}{\sqrt{s}}$ (with $s \gg \Lambda^2_{\mathrm{QCD}}$); secondly, it adds the effect of hadronisation, as partons cannot be considered as free longer than time scales of $\tau \sim \frac{1}{\Lambda_{\mathrm{QCD}}}$. Thirdly, it adds the *underlying event* (remind from Subsection 1.2.3):

Beam remnants	The debris of the proton produce hadronisation as well.
Multiple Parton Scattering	Additional interaction may be provided by the two interacting protons.

Finally, at each step, the program includes the decay of resonances, such as neutral weak bosons, or of short-lifetime hadrons, such as neutral pions.

4.1.1.2 Pile-up

A few tens of interactions superpose in the detector. The detection and the reconstruction must be simulated in the same conditions as the real events.

4.1.1.3 Detector simulator

Geant 4 is a library used to simulate the effect of a particle into a medium after the hadronisation through the different parts of the CMS detector. The result is treated by the trigger system as in real cases.

And finally, it is parsed and formatted by CMSSW, exactly as the sample of real data, and stored in the Tiers as well.

4.1.2 Raw data files

As we said in Subsection 3.3.5, the most information that is contained in raw data files consist of the least electronic signal of every device of the detector: briefly, CMSSW executes a few steps to produce files that are usable by analysts.

Finally, after an operation of preselecting and downloading, the samples are registered in Root files, Root being the analysis framework, and structured in what ITs call *trees*. It is a kind of big table: each event (*i.e.* bunch crossing) is indiced and contains every variable of every detected particles, plus a few variables describing the detection itself (see Figure 4.3). We are going to write programs to handle the trees, to apply some operations of selection or and produce histograms.

Simulations and data are both structured in similar trees, in order to be processed by the same programs, with the difference that simulation may also contain the variables describing the generated events.

Each simulation tree is characterised by two numbers:

— the total cross section of the sample,

— and the number of generated events.

It is important to know them to rescale the simulations up to the data. For instance, in the simulation of the Drell-Yan process, there are $30,459,503$ events for a cross section of $3531.8\,\mathrm{pb}^{-1}$. It is used to compute the luminosity of the apparatus that would have been

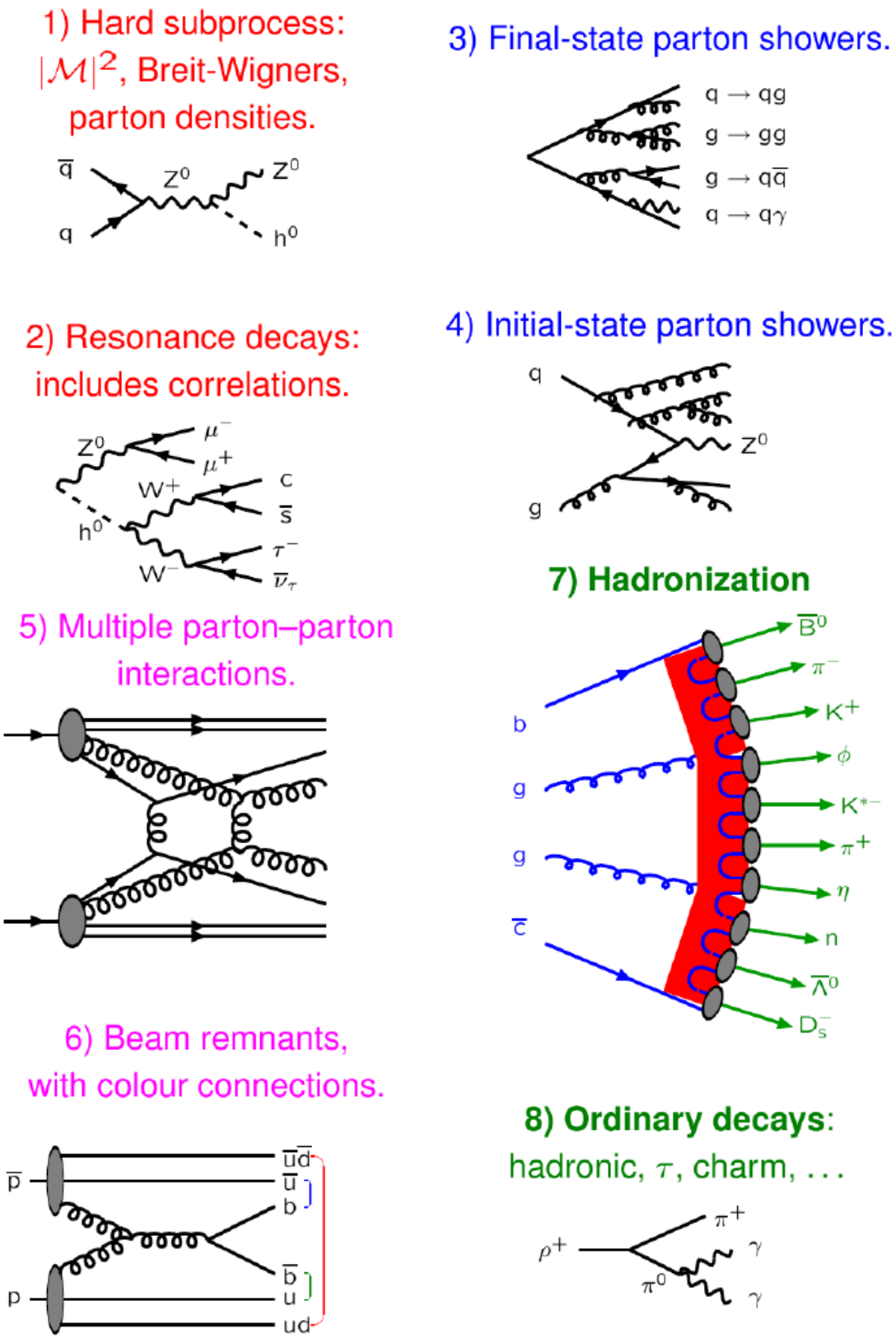


FIGURE 4.2 – The different steps taken into account by any general event generator: the matrix element (in red), the parton showers (in blue), the underlying event (in pink) and the hadronisation and decays (in green). (Taken in [21].)

```
*****************************************************************
*    Row   * Instance * patMuonEn * patMuonPt * patMuonEt * patMuonPh *
*****************************************************************
*        0 *        0 * 75.826713 * 58.472574 * 0.7525565 * 2.4983778 *
*        0 *        1 * 31.639717 * 19.350358 * -1.074569 * 0.6122089 *
*        1 *        0 * 59.041366 * 46.718661 * -0.711223 * 1.7127188 *
*        1 *        1 * 74.980163 * 45.802941 * -1.076057 * -1.535101 *
*        2 *        0 * 101.33430 * 97.870609 * -0.265266 * 1.8077528 *
*        2 *        1 * 77.597969 * 58.148255 * -0.796668 * -1.392483 *
*        3 *        0 * 74.698837 * 57.227445 * -0.762777 * 1.4577105 *
*        3 *        1 * 27.349140 * 22.010372 * 0.6831272 * -1.473299 *
*        4 *        0 * 73.489898 * 55.347356 * -0.789053 * 0.7308084 *
*        4 *        1 * 73.625007 * 37.430202 * -1.297687 * -2.367869 *
*        5 *        0 * 51.722972 * 23.844239 * 1.4095576 * 0.4479529 *
*        5 *        1 * 32.845230 * 22.781791 * 0.9083657 * -2.962432 *
*        6 *        0 * 27.054798 * 23.278871 * -0.562122 * -2.226291 *
*        6 *        1 * 33.332004 * 19.364769 * -1.138550 * 0.5744602 *
*        7 *        0 * 38.414707 * 38.379736 * 0.0425952 * 1.6834195 *
*        7 *        1 * 49.085502 * 29.820984 * -1.082961 * -1.401216 *
*        8 *        0 * 70.534759 * 43.938239 * -1.051211 * 2.7209761 *
*        8 *        1 * 126.71962 * 42.528790 * -1.755513 * -0.468613 *
*        9 *        0 * 73.092834 * 72.466954 * 0.1313263 * -2.840864 *
*        9 *        1 * 41.805084 * 15.103089 * -1.676908 * 1.0523428 *
*       10 *        0 * 174.97541 * 53.674928 * 1.8504450 * -2.982701 *
*       10 *        1 * 96.682334 * 41.089199 * 1.5002702 * -0.410037 *
*       11 *        0 * 90.539115 * 45.646089 * -1.307379 * 0.2093711 *
*       11 *        1 * 48.645042 * 43.365438 * -0.488571 * -2.970951 *
*       12 *        0 * 71.209014 * 35.658180 * -1.315216 * 1.6586117 *
```

FIGURE 4.3 – For instance, to give an idea of what a tree looks like, I have asked the computer to give for each muon (`instance`) of every event (`row`) the energy (`patMuonEn`), the transverse momentum(`patMuonPt`), the pseudorapidity (`patMuonEt`), and the azimuthal angle (`patMuonPh`). This continues up to a few millions of events. Such a view of the tree is useful as a control step but it goes without saying that it is not directly handlable.

used if it were real data, thanks to the relation 3.4; in this case, the MC sample of the Drell-Yan process at tree-level must be rescaled by a factor of 2.26 and the MC is thus said to be *rescaled to the luminosity of the data*.

MC samples like the backgrounds are always provided with a cross section and a number of events, given as parameters of to the event generator, and will be rescaled as well.

Remark

Most of the practical work begins here. In practice, this means coding C++ codes to handle the trees. All in all, nearly four thousands lines of code have been written to proceed to the selection, the processing and the display of the different plots. A few details concerning the coding are described in Appendix B.

4.2 Selection

Our final goal is to study the agreement of a sample of data and a sample of simulation. This agreement can be improved by a smart selection on the events and on the phase space, according to the working of the detector, as the algorithm of reconstruction which has been used, or its topology.

4.2.1 Cutoffs

There are two kinds of cutoffs: over kinematic variables and over the detector variables.

1. The detector variables specify the conditions of reconstruction and identification of the muon. Those variables are studied by the *Muon POG*, which the group at CMS responsible for the optimisation of the reconstruction algorithm.

   Trigger We described the trigger system in Chapter 3 (see Subsection 3.3.4). The trigger variable contains an index specifying on which criterion the event has been accepted by the trigger system. In our case, a muon with $p_T = 17\,\text{GeV}$ and a second muon with $p_T = 8\,\text{GeV}$ must have been seen in the detector.

   Identification Several algorithms exist for muon reconstruction (see Subsection 3.3.3). We choose the *tight* selection, as it is the most suited for bosonic processes.

   Isolation A charged meson can decay weakly in flight and produce a muon.To eliminate those muons from our data, each muon is required to be isolated: there must be no hadronic activity around the reconstructed trajectory of the particle. We select thus events such that $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta phi^2} \geq 0.3$, as recommended by the Muon POG for tight muons.

   Primary vertex The reconstructed track can be extrapolated to the pipe. This may show whether the particle directly comes from the scattering of two protons (primary vertex) or from the decay of one of the resulting particles (secondary vertex). In the analysis of the Drell-Yan, the muons have too come from the primary vertex, which means that the trajectories of the muons should cross at less than 2 mm from the beam axis.

2. The kinematic cutoffs define the region of the phase space in which data and simulations are compared. The choice of the phase space is oriented according to the limitations of the apparatus and by the goal of the analysis.

   Dimuon invariant mass $M_{\mu\mu} \geq 60\,\text{GeV}$
   We are mostly interested in events of very high invariant mass. This fixes the hard scale of the event.

   Transverse momentum $p_T \geq 20\,\text{GeV}$
   We take a margin from the value of selection of the trigger, as it works on basic algorithms of reconstruction.

   Angle with the axis $|\eta| \leq 2.4$
   The angular aperture of the detector cannot be total because of the pipe. Moreover, the magnetic effects near the pipe

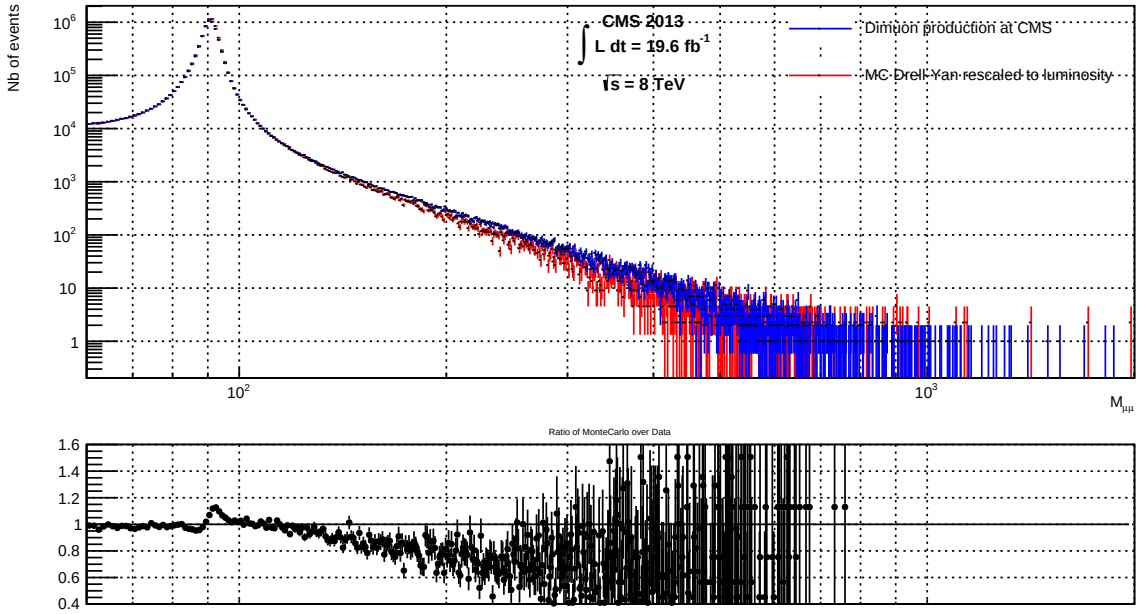


FIGURE 4.4 – Superposition of the invariant mass spectrum from 60 GeV to 2 TeV of both the dimuon signal and of the MC simulation of the Drell-Yan rescaled to the luminosity, directly after selection. The ratio of the MC over the Data is given in the lower part.

make the detection difficult near the pipe. Tracking is limited to $|\eta| \leq 2.5$; again, we take a margin.

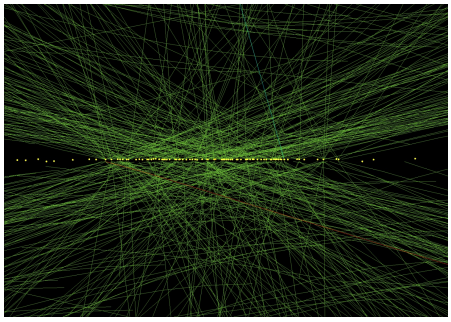
In practice, I wrote a C++ class `HSelect` using ROOT libraries to loop variable by variable and event by event on the tree and build a histogram for each variable and for the muon pair variables. Such a program needs around ten minutes to run over a sample of several millions of events.

The application of this selection on the trees of data and MC give the result of Figure 4.4. The agreement between the data and the MC is not bad at all, at least till a few hundred GeV. But actually, the efficiencies of the cutoffs on the detector variables slightly differ between the data and the MC, as simulations are never perfect. Before debating on the agreement, MC must be applied correction *scale factors*.

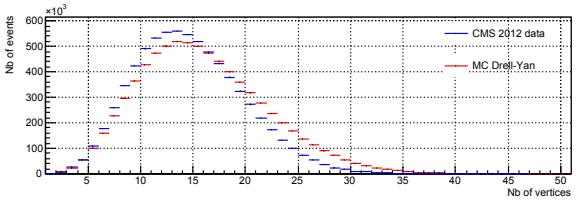
4.2.2 Corrections to the simulation of the detector

4.2.2.1 Scale factors

The difference of efficiency for the selection on a detector variable means that the detector is not perfectly simulated. Indeed, simulating the detector to the percent level is a very difficult task: there are so many devices, which may break or lose in efficiency. It is not straightforward to measure the efficiency of a detector while using it to acquire data. As a consequence, the efficiencies of the detector variable cutoffs differs between the data and simulation from a few percent (see Table 4.1). In order to work with events that have been selected exactly on the same criteria *scale factors* $\frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC}}}$ (SF) have to be applied on the events of the simulation to correct the selection achieved by the detector variables. It is much faster than making the simulation of the detector run again, as it is very time consuming. In theory, a scale factor should be applied in every point of the phase space,



(a) Reconstruction of an event. One can here see the exceptional superposition of 78 events at the same bunch crossing.



(b) The distribution of the number of interaction per bunch crossing is not perfectly simulated: indeed, the distribution are centered around the same value.

FIGURE 4.5 – The pile-up is the superposition of several tens of scatterings.

i.e. for every value of the transverse momentum, of the pseudorapidity and of the polar angle; in practice, it is done region of the phase space by region of the phase space. The values are computed by the Muon POG.

4.2.2.2 Pile-up

As we have said several times in this work, an event does not corresponds to a scattering but to a bunch crossing. The *pile-up* is the superposition of the different scatterings at the same bunch crossing (see Figure 4.5a). The pile-up is one more difficulty to simulate an event and it must be corrected as well, as is illustrated on Figure 4.5b.

4.2.2.3 Application of the correction

The application of the scale factors is done by the same class `HSelect`, event by event just after the selection procedure. The correction that must be applied is computed by the Muon POG too and applied at the same time as the scale factors. Figure 4.6 shows the result of the selection with the application of the scale factors on the MC (notice that now the height of the bins for the MC are not integer anymore).

The agreement is good up to a few hundreds of GeV. Further, the binning (bin width of 1 GeV) is too narrow for the resolution on this variable. Lastly, around the peak of the neutral weak boson (Z-peak), the ratio makes a little wave, due to the uncertainty on the absolute calibration of the detector. This is also due to the binning that is too thin.

Finally, the ratio is more around 0.95 than around 1, as the data includes the whole dimuon production whereas the MC only includes the Drell-Yan process. We are now going to treat the background.

min $\|\eta_1\|$	max $\|\eta_1\|$	min $\|\eta_2\|$	max $\|\eta_2\|$	trigger correction
0	0.9	0	0.9	0.97
0	0.9	0.9	1.2	0.96
0	0.9	1.2	2.1	0.95
0	0.9	2.1	2.4	0.96
0.9	1.2	0	0.9	0.93
0.9	1.2	0.9	1.2	0.94
0.9	1.2	1.2	2.1	0.95
0.9	1.2	2.1	2.4	0.95
1.2	2.1	0	0.9	0.94
1.2	2.1	0.9	1.2	0.91
1.2	2.1	1.2	2.1	0.92
1.2	2.1	2.1	2.4	0.95
2.1	2.4	0	0.9	0.98
2.1	2.4	0.9	1.2	0.95
2.1	2.4	1.2	2.1	0.91
2.1	2.4	2.1	2.4	0.96

(a) Trigger correction to be applied to the pair of muons as a function of the pseudorapidities.

min $\|\eta\|$	max $\|\eta\|$	min p_T	max p_T	Isolation correction	Identification correction
0	0.9	10	20	0.947 05	0.984 868
0	0.9	20	25	0.974 978	0.988 681
0	0.9	25	30	0.997 129	0.993 889
0	0.9	30	35	0.993 863	0.994 164
0	0.9	35	40	0.993 442	0.994 084
0	0.9	40	50	0.994 101	0.992 47
0	0.9	50	60	0.995 544	0.990 978
0	0.9	60	90	0.999 036	0.990 444
0	0.9	90	140	1.001 04	1.003 85
0	0.9	140	500	1.0003	1.027 98
0.9	1.2	10	20	0.951 836	0.986 855
0.9	1.2	20	25	0.988 368	0.987 375
0.9	1.2	25	30	1.000 83	0.994 212
0.9	1.2	30	35	0.998 546	0.990 593
0.9	1.2	35	40	0.999 14	0.990 353
0.9	1.2	40	50	0.998 176	0.989 641
0.9	1.2	50	60	0.998 696	0.991 311
0.9	1.2	60	90	0.999 132	0.986 31
0.9	1.2	90	140	0.999 559	1.011 91
0.9	1.2	140	500	0.996 767	0.955 563
1.2	2.1	10	20	0.980 045	1.012 35
1.2	2.1	20	25	0.997 342	1.001 55
1.2	2.1	25	30	1.007 84	0.999 149
1.2	2.1	30	35	1.006 85	0.997 573
1.2	2.1	35	40	1.0037	0.996 585
1.2	2.1	40	50	1.002 09	0.997 431
1.2	2.1	50	60	1.001 25	0.997 521
1.2	2.1	60	90	1.000 65	0.993 942
1.2	2.1	90	140	0.999 878	1.019 22
1.2	2.1	140	500	0.999 89	1.016 48
2.1	2.4	20	500	1.025	0.994

(b) Identification and isolation corrections to be applied to each muon as a function of its pseudorapidity and its transverse momentum.

TABLE 4.1 – Scale factors for efficiency correction. They are applied to the Monte Carlo distributions.

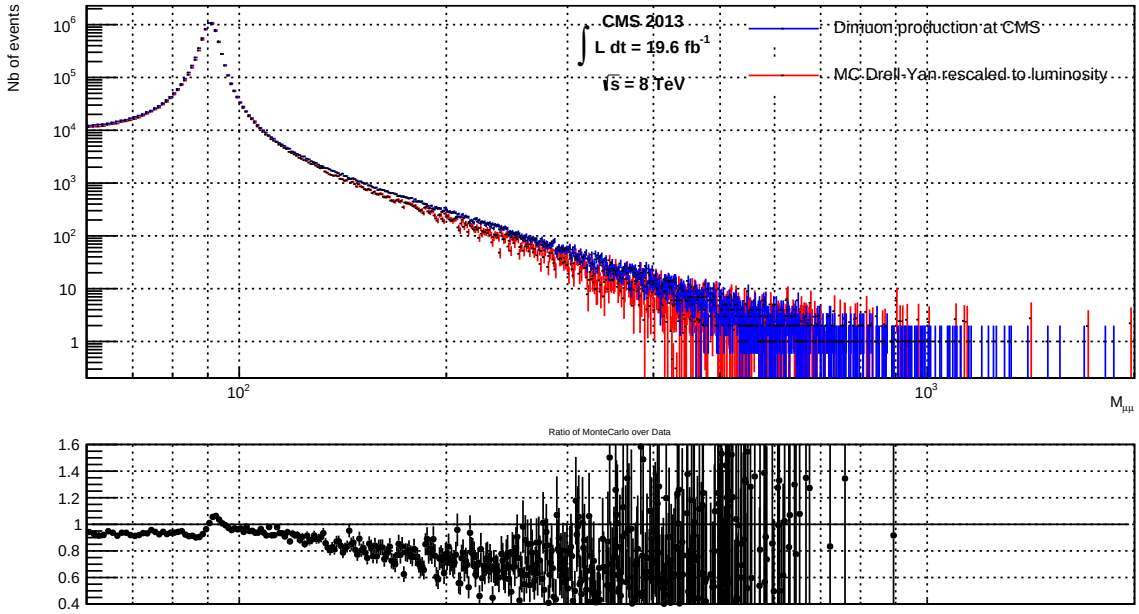


FIGURE 4.6 – Superposition of the dimuon signal and of the MC simulation of the Drell-Yan rescaled to the luminosity and corrected by the scale factors and for the pile-up.

4.3 Extraction of the Drell-Yan signal from the dimuon production

The MC backgrounds are first added to the MC Drell-Yan to be compared to the data. Then we are going to subtract the simulations of the background to the total data to extract the Drell-Yan part of the data. (Let us keep in mind that our goal is to obtain an invariant mass spectrum of the Drell-Yan process.)

4.3.1 Backgrounds

Many other processes may contribute to the dimuon production. The question is to determine which Feynman diagrams (*i.e.* terms in the perturbative development of the evolution of the interaction Hamiltonian) are significant? Provided that the Drell-Yan process is treated at tree-level, the processes that mainly contribute are the following:

$q\bar{q} \to \tau\tau$ — The Drell-Yan process can produce a pair of tauons, which decay weakly into neutrinos and muons.

diboson production — The process involves two virtual bosons, which decay in turn into two pairs of fermions. Two of the four fermions might be muons.

semileptonic decay of top quark — The decay of top quarks may contain muons. Several top backgrounds exist:

- the strong annihilation of two partons into a pair of top-antitop,
- the tW channel, where each may produce a muon,
- and the exchange of a W between two quarks of different flavours may produce a pair top-bottom.

Type	Sample	Nb. of events	B.R. $\times \sigma / \text{pb}^{-1}$
Drell-Yan	DYmumu	30 459 503	3531.8
	DYtautau	30 459 503	3531.8
$gg \to t\bar{t}$	TTJets	6 923 652	245
Diboson processes	WWJets2L2Nu	1 933 235	0.11×54.838
	WZJets2L2Q	3 215 990	0.07×33.21
	WZJets3LNu	1 995 334	0.07×33.21
	ZZJets2L2Nu	954 911	0.014×17.654
	ZZJets2L2Q	1 936 727	0.004×17.654
	ZZJets4L	4 807 893	0.001×17.654
Single-top processes	T (t-channel)	3 758 227	56.5
	T (s-channel)	259 961	3.79
	TW production	497 658	11.1
	Tbar (t-channel)	1 903 681	30.7
	Tbar (s-channel)	139 974	1.76
	TbarW production	493 460	11.1

TABLE 4.2 – Parameters of the samples of MonteCarlo simulations. The Drell-Yan process clearly dominates at the chosen hard scale.

Qualitatively, this would be true for the less massive quarks, but in practice, their contribution is negligible.

When the process does not involve a Z boson, the muons have the wrong kinematics to come from the Drell-Yan and thus deform the signal. Figure 4.7 gives the Feynman diagrams that mostly contribute to the dimuon production with their keyword and Table 4.2 the cross section and the number of event per MC sample (including the Drell-Yan).

As they were parsed with CMSSW, the background trees have the same structure as the MC Drell-Yan and CMS data. They are thus run by the same program `HSelect`, and then rescaled and stacked by another class, called `HProcess`.

Figure 4.8a shows the comparison between the data and the simulation. The deficit seems to be filled, as the ratio is more around one.

4.3.2 Extraction of the Drell-Yan signal

The extraction of the Drell-Yan part of the dimuon production signal is achieved by the class `HProcess`.

Now we see that the agreement is good (Figure 4.8b), the Drell-Yan part of the measured dimuon production may be estimated by subtraction of the MC simulations to the data. A definitive conclusion on the agreement cannot be given prior to computing the systematic uncertainties. But we postpone this to the very ending of the chapter, and move to the treatment of the bias due to the reconstruction on the distributions.

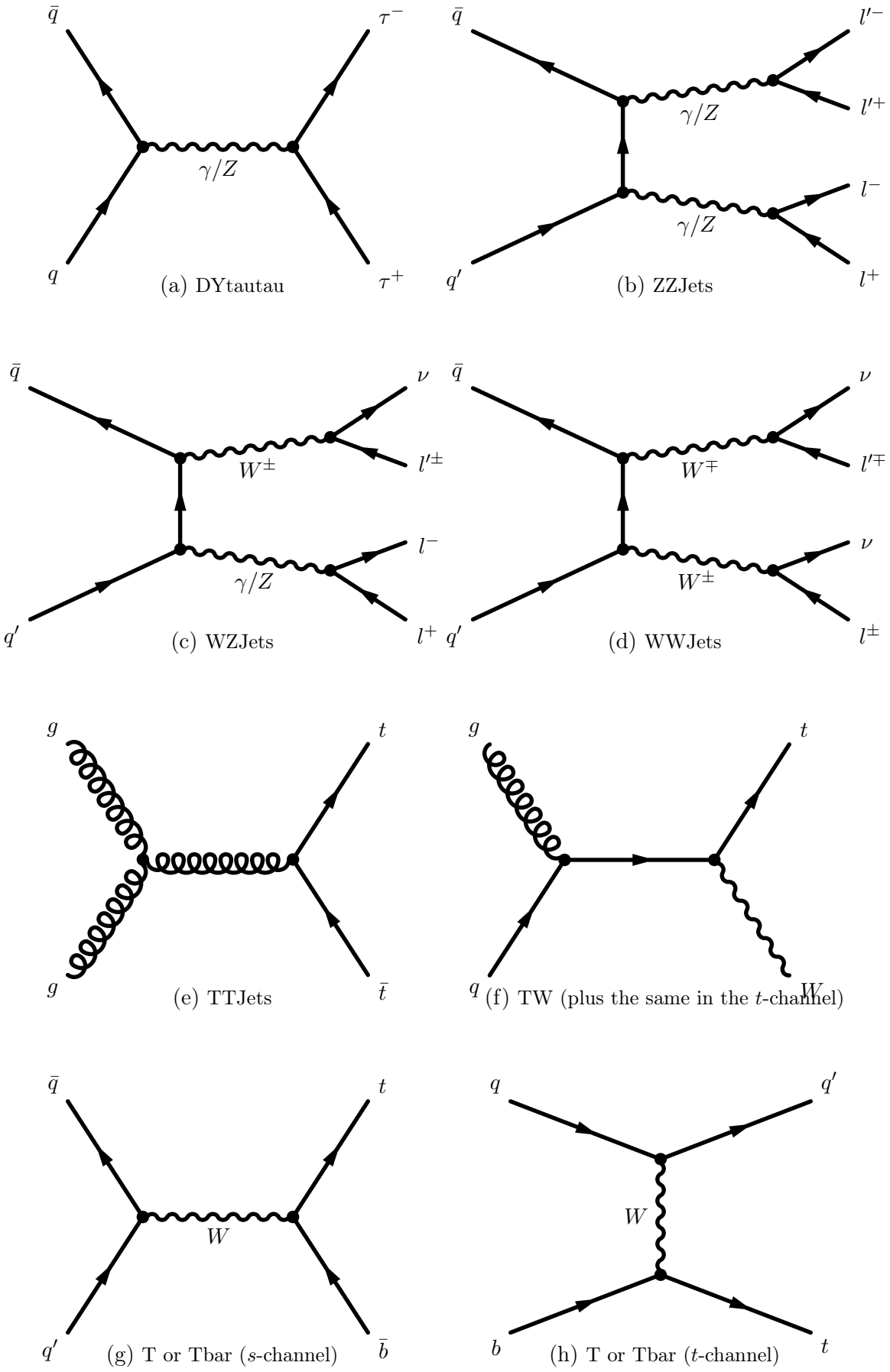
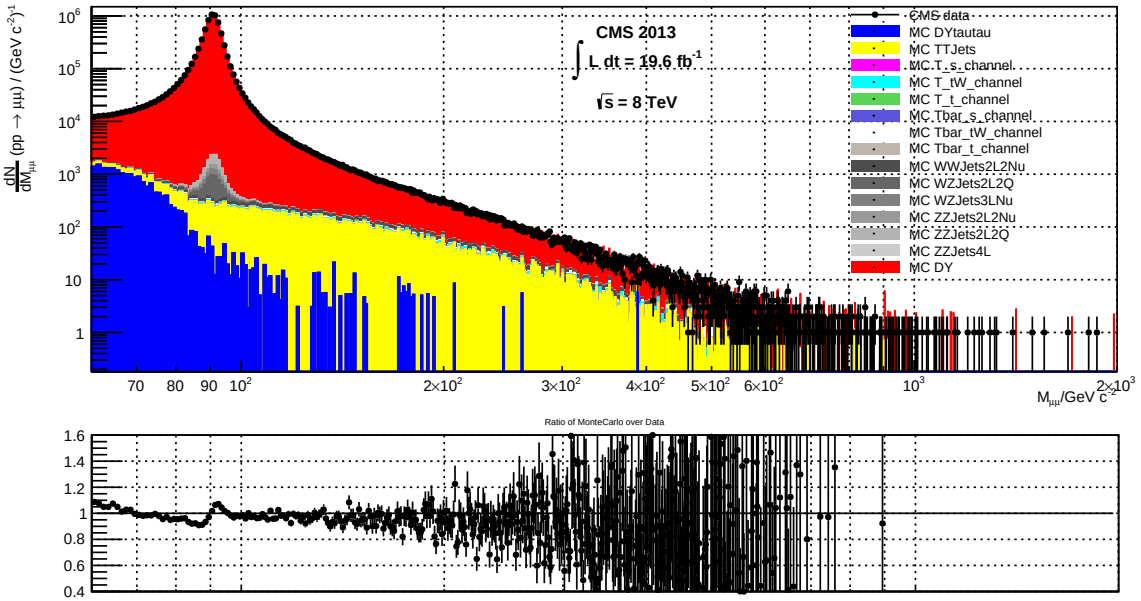
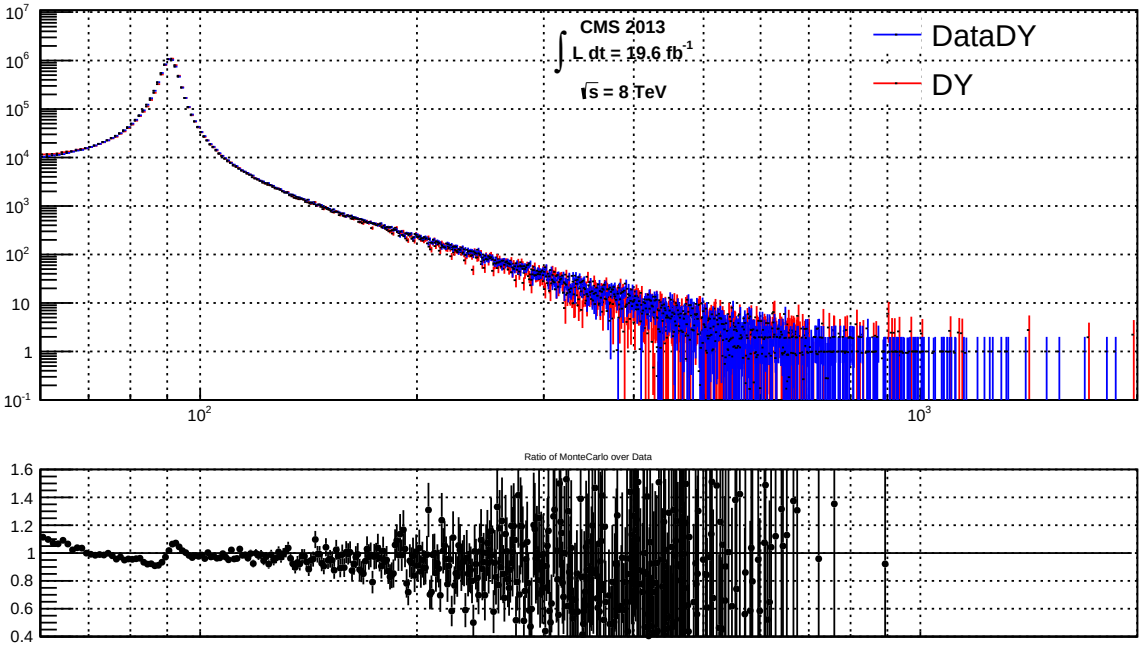


FIGURE 4.7 – Feynman diagrams of the most relevant contributions. The keywords are used in the legend of the plots, but are not conventional. For simplicity, neither the protons nor the decay products of the product particles have not been drawn. In the single-top contribution, the second muon comes from the decay products of B-mesons.



(a) Comparison of the stack of the MC simulations (Drell-Yan plus backgrounds) to the dimuon production at CMS (bold points). The backgrounds that still involve a Z^0 are called *resonant backgrounds* as they contribute to the peak as well.



(b) Extraction of the Drell-Yan signal from the dimuon production at CMS compared with the MC Drell-Yan distribution.

FIGURE 4.8 – Treatment of the backgrounds.

4.4 Unfolding

While acquiring data, there are two kind of bias:

— the bias due to the intrinsic working of the detectors,

— the bias due to the algorithm of reconstruction of the particles.

The aim of unfolding is to reverse those two effects.

We are here first detailling what a bias looks like in our data, and then the procedure of unfolding. By the way, the method provides a way to fix the size of the bins according to the resolution of the detector.

4.4.1 Situation

4.4.1.1 Effects of the bias

It is a universal problem in experimental physics to ensure that measurements are not biased. At CMS, an event to reconstruct is a big bag of bones: each scattering must be reconstructed while tens of them have happened simultaneously. It goes without saying that each particle of every scattering of the event is not perfectly reconstructed.

For instance, let us consider a muon that has been produced with $p_T = 30\,\text{GeV}$ and $\eta = 2.42$. If it radiates a photon, p_T and η vary. The photon may not be seen by the detector, or associated with the wrong particle while reconstructing, and consequently, the muon is attributed wrong values for p_T and η. All in all, three effects may be distinguished:

1. If the transverse momentum is measured to 25 GeV, the muon still passes the cutoff $p_T > 20$ GeV. This is the *migration effect*.
2. Now if the pseudorapidity is measured at 2.39, then the muon passes the cutoff on the pseudrapidity $|\eta| \leq 2.4$ while it should not have! This is a *fake* muon, more problematic for a fine analysis such as ours.
3. Similarly, a muon may also be *missed* if it is generated in the right region of the phase space but not reconstructed.

The reversing of these effect is done thanks to the simulations.

4.4.1.2 Treatment of the bias

Formally, let R and G be the reconstructed and generated distributions of, say, the invariant mass spectrum. On a vector point of view, there should exist a matrix D such that:

$$R = D \times G \tag{4.1}$$

Such a matrix is called the *response matrix*, and it represents the global action from the generation to the reconstruction. It describes the migration of the values of a given variable in terms of probabilities: for instance, the response matrix contains the probabilities for an event produced at, say, a value of 90 GeV of invariant mass to be reconstructed at any other value, say, 91 GeV. Given the choice of the phase space (kinematic as well as every other variable taking place in the selection), it should be unique. This allows us to construct the response matrix thanks to our simulations at generated and reconstructed [2] levels. We leave the exact procedure of construction for later and focus on the method.

2. Corrected by the scale factors.

Unfolding the data means using this matrix D to get the physical distribution from the measured distribution. Let us assume that the response matrix D is square (one can always arrange for this). Then one would naively want to invert and apply it on the measured distribution M to get the physical distribution Φ:

$$\Phi = D^{-1} \times M \tag{4.2}$$

But any square matrix is not necessarily invertible. And even if D can be expected to be invertible with large statistics, it seems that negative matrix elements are unavoidable when large fluctuations occur, leading to unfolded distributions with negative bin heights.

Such problem of stability have encouraged to develop other techniques of *unfolding*. G. D'Agostini suggested a method based on *Bayes theorem* (see [10]). Let us first focus on its theoretical formulation, and then apply it to our extraction of the Drell-Yan signal to get a physical distribution.

4.4.2 D'Agostini's method

G. D'Agostini suggests the problem to be treated on a probability point of view, as the matrix response contains probabilities $\mathbb{P}$[measured value, knowing generated value]. It is an iterative procedure that exploits Bayes theorem on conditional probabilities.

Its advantages are numerous:

— it is theoretically well-grounded;

— the bins may have different widths;

— the domains of definition of the causes and of the effects may differ from one another;

— no matrix invertion;

— it provides the correlation matrix of the results;

— unfolding libraries using D'Agostini method are available online;

— etc. (many other advantages do not concern our study)

4.4.2.1 Bayes theorem

Let us begin by some recalls.

Let A and B be two events of an event space Ω. The conditional probability that A happens if B has happened is given and noted by:

$$\mathbb{P}[A|B] \equiv \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} \tag{4.3}$$

Bayes theorem relates $\mathbb{P}[A|B]$ and $\mathbb{P}[B|A]$:

$$\mathbb{P}[A|B] = \mathbb{P}[B|A] \frac{\mathbb{P}[A]}{\mathbb{P}[B]} \tag{4.4}$$

The proof is direct by using twice the definition 4.3. If both $\{A_i\}$ and $\{B_j\}$ partition Ω, it may be extended:

$$\mathbb{P}[A_i|B_j] = \frac{\mathbb{P}[B_j|A_i]\,\mathbb{P}[A_i]}{\sum_k \mathbb{P}[B_j|A_k]\,\mathbb{P}[A_k]} \tag{4.5}$$

4.4.2.2 Unfolding algorithm

Let us talk in general terms and say *causes* C_i ($i = 1, \ldots, n_C$) and *effects* E_j ($j = 1, \ldots, n_E$) instead of generated and reconstructed level. This will allow us to include the fake and miss event properly.

The goal is to determine the cause distribution $\{C_i\}$ from the effect distribution $\{E_j\}$ in terms of probabilities:

$$n(C_i) = \frac{1}{\epsilon_i} \sum_{j=1}^{n_E} n(E_j) \mathbb{P}\left[C_i | E_j\right] \tag{4.6}$$

where ϵ_i is the efficiencies of the detector to reconstruct the ith cause

$$\epsilon_i \equiv \sum_{j=1}^{n_E} \mathbb{P}\left[E_j | C_i\right] \tag{4.7}$$

In this expression, we only know $n(E_j)$ and we want to know $n(C_i)$; $\mathbb{P}\left[C_i | E_j\right]$ can be treated thanks to Bayes theorem, re-written as:

$$\mathbb{P}\left[C_i | E_j\right] = \frac{\mathbb{P}\left[E_j | C_i\right] n(C_i)}{\sum_{k=1}^{n_C} \mathbb{P}\left[E_j | C_k\right] n(C_k)} \tag{4.8}$$

where we have used $\mathbb{P}\left[C_i\right] = \frac{n(C_i)}{\sum_{l=1}^{n_C} n(C_l)}$. The two relations are cyclic, and the iteration is funded on this observation.

1. At the first iteration, the *true* distribution $\{n(C_i)\}$ (noted without index) is unknown and a uniform distribution $\{n_0(C_i)\}$ can be put in 4.8 in place of it.
2. Then the $\{n_1(C_i)\}$ obtained thanks to 4.6 is a distribution that lies between $\{n_0(C_i)\}$ and the true distribution.
3. The procedure can be iterated: given $\{n_k(C_i)\}$, put it in 4.8 to find $\{n_{k+1}(C_i)\}$.

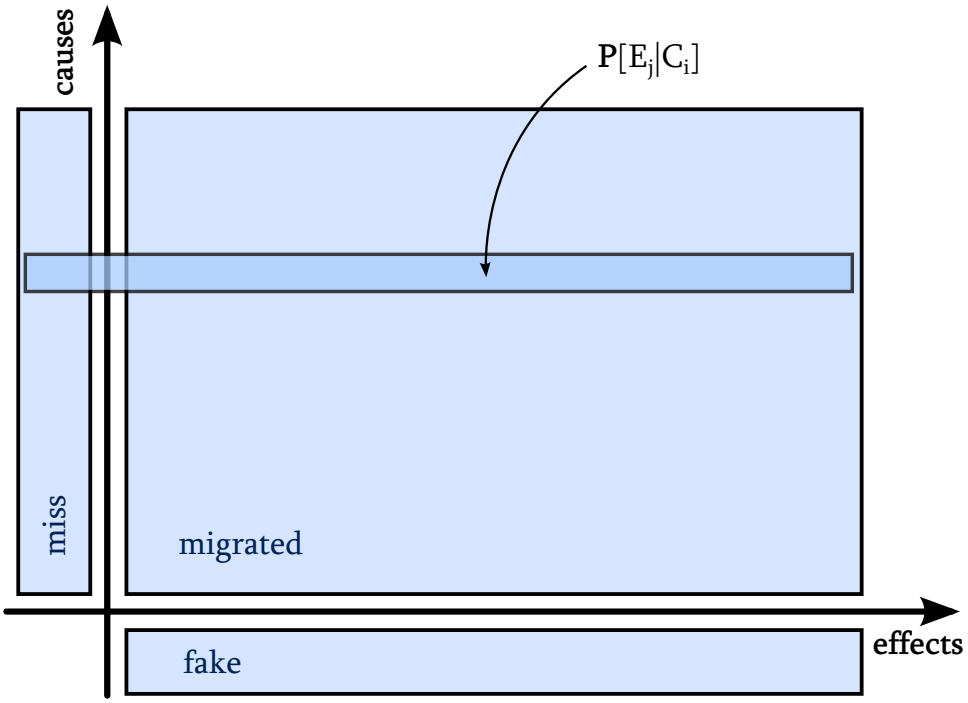
Actually there is no need to iterate 10^{lot} times: after a few iterations, the uncertainties grow more and more while the evoluted distribution moves less and less or begins to oscillate [3]. In practice, around four iterations are enough to get a good estimation of the final distribution.

4.4.2.3 Construction of the response matrix

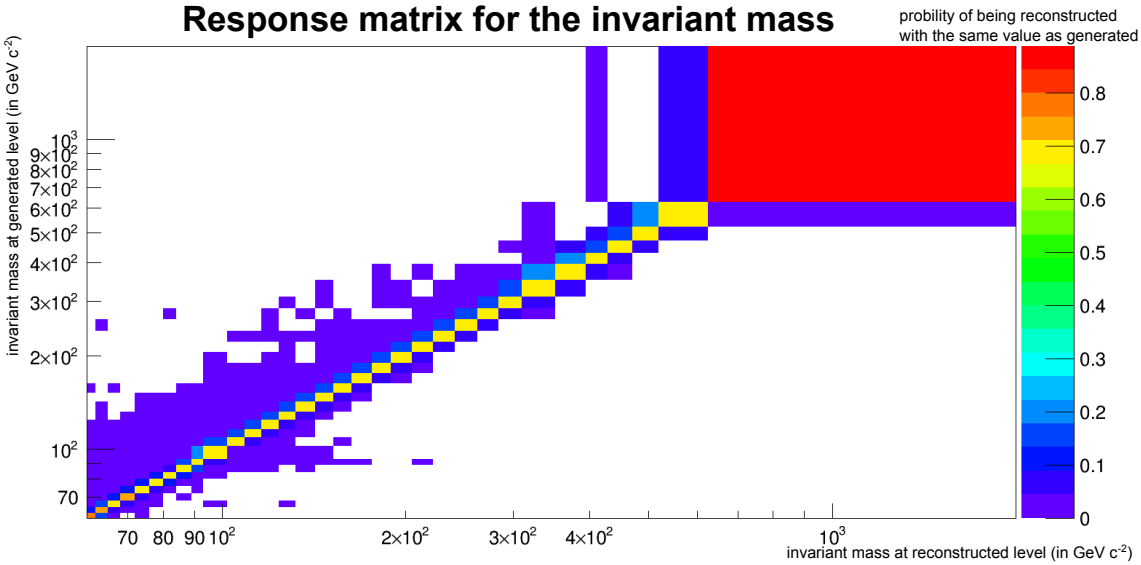
The MC samples for the Drell-Yan process are given at generated and reconstructed levels, which means that each event that is treated is given at both levels. Each level is tested independently and save independently. Now, if both levels are successfully tested, the response matrix is completed as well. At the end of the selection, the response matrix that is partial, as it only describes migrated but neither fake nor miss events.

In a way, fake and miss events are a kind of migration. Indeed, D'Agostini method makes no assumption on the order of the bins or on their widths. Supplementary bins C_0 as a cause for the fake and E_0 as an effect for the miss are appendiced to the partial response matrix that has been put up while selecting. The last step consists of normalising line by line in order to get the probabilities $\mathbb{P}\left[\text{effect}|\text{cause}\right]$. The procedure is illustrated on Figure 4.9a, and the final result is shown on Figure 4.9b.

3. Actually, I have found no proof of convergence.



(a) Schematic response matrix, including the supplementary bins to take into account the miss and the fake events while unfolding.



(b) Response matrix (X-axis corresponds to the effects and Y-axis to the causes). The lines have been normalised so as to give the probabilities of being reconstructed with the same value as generated. This matrix does not include the supplementary bins for the fake and miss muons.

FIGURE 4.9 – Response matrix.

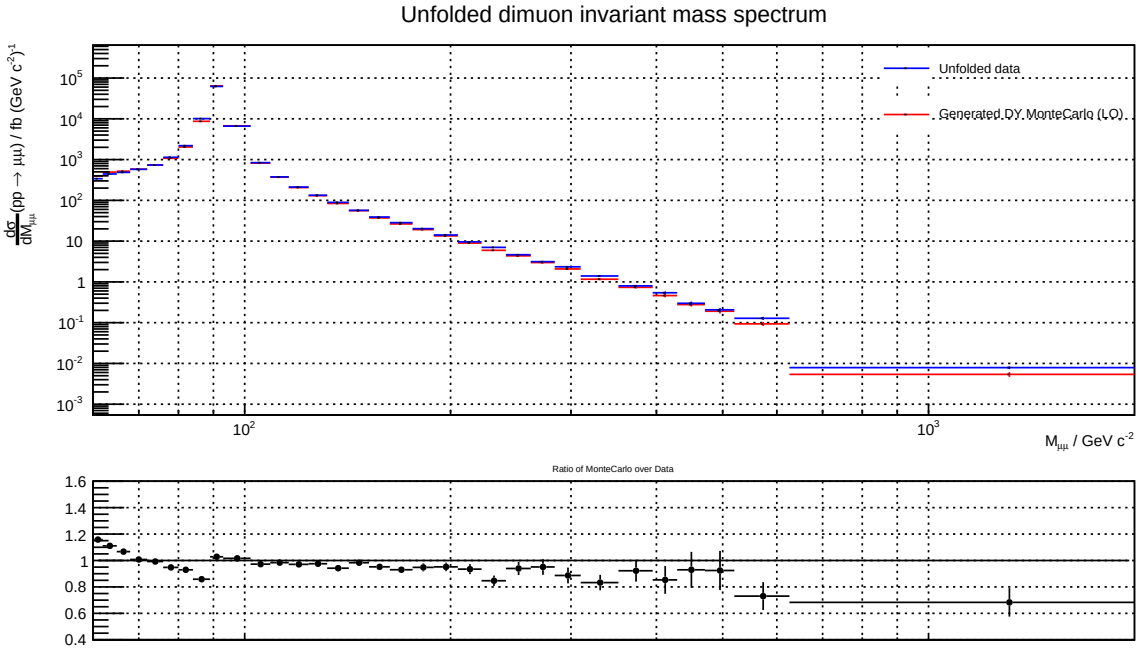


FIGURE 4.10 – Unfolded distribution of the Drell-Yan process compared to the Drell-Yan simulation.

4.4.3 Resolution

The response matrix is also a way to recover the detector resolution by requiring that the bins on the diagonal contain one sigma of the events, *i.e.* 68% of the events have not migrated. We add the second condition of having at least a hundred events (in practice, this only holds for the very last bin).

4.4.4 Unfolded data

A partial response matrix is made during the selection, by `HSelect`, *i.e.* without the two more bins for fake and miss muons. Most of the unfolding procedure is achieved by the standard library `RooUnfold`.

`RooUnfold` library is handled by `HProcess`. Four iterations are applied on the Drell-Yan extracted from the dimuon production. The result is shown on Figure 4.10.

It is time to compute the systematic uncertainties and then conclude on the agreement.

4.5 Uncertainties and final plots

After a few words on the statistical uncertainties, and the computation of the systematic uncertainties. we will be able to discuss the plots, at the end of this section.

4.5.1 Statistic uncertainty

Each bin i of a histogram is considered to follow a Gaussian law $\mathcal{N}(N_i, N_i)$, where N_i is the number ob events in the bin. If the events have been applied scale factors, the error is corrected by the same factor. If the bin content is rescaled, the error must be rescaled

as well. And when two histograms, say two backgrounds, are summed, their errors have to sum quadratically.

At the unfolding procedure, the statistical uncertainty increases rapidly, caused by the little statistics of most off-diagonal bins of the response matrix. Quantitatively, it is automatically computed by the framework that we are using, which is `RooUnfold` (see Appendix B).

4.5.2 Systematic uncertainty

On the other hand, the systematic uncertainties (the bias on the final values due to the numerical treatments of data) depend on the analysis. In our study, they are due to the following steps:

— the cross sections used to generate the MC distributions for the backgrounds,

— the choice of the PDF used to generate the MC Drell-Yan distribution,

— the pile-up correction to the simulation of the detector,

— the SF to correct the simulation of the detector,

— the measurement of the integrated luminosity,

— and the procedure of unfolding.

To estimate the systematic uncertainties of each source on the final result (up and down as it may be asymmetric), we have used the *offset method*.

Then we assume that the distribution of the errors are Gaussian: this is subject to discussion, as the errors may be asymmetric, but it is often difficult to find the right distribution of uncertainty, The Gaussian approximation is enough if one just wants an *estimation* of the uncertainty. Then, as Gaussian, the errors just add quadratically.

4.5.2.1 Pile-up

Let us consider first of all the pile-up correction to simplify the discussion.

The whole procedure of treatment, from the selection of events to the unfolding, is run with the central values shifted by the raised and lowered corrections. At the end of the procedure, the uncertainties are defined as follows:

$$\sigma_i^- = \min\big(f_i(\text{central value}), f_i(\text{raised value}), f_i(\text{lowered value})\big) \tag{4.9}$$

where f stands for the whole procedure; and similarly for σ_i^+. In practice, the lowered and raised values are estimated by shifting by one unit the distribution of number of vertices per bunch crossing.

The same kind of procedure is applied for every source of uncertainty.

4.5.2.2 PDFs

The PDFs of the Drell-Yan cross section at the generated level can be changed just by reweighting the events, thanks to a library called *LHAPDF* (see Appendix B). Here, instead of considering the three results for the raised, central and lowered values, one considers the results for a few choices of PDF and takes the minimum (resp. maximum) value bin by bin as lower (resp. upper) error $\sigma_i^\pm$. In our study, we have considered the following sets:

— HERAPDF,

— NNPDF,

— MRST,

— CTEQ6 (default).

None of them has computed at the LHC.

Note that they dominate the systematic uncertainties.

4.5.2.3 Cross sections of the backgrounds

We have considered an error of 5% on the cross sections of the backgrounds, mostly due to the PDF. As it is just a global factor, the procedure to find the lower and upper results just sums up to rescaling a taking the difference between the rescaled and central values.

This is a minor effect.

4.5.2.4 Scale factors

The Muon POG [4] gives the tables of scale factors up to 500 GeV of transverse momentum and recommends an error of 2%. Higher, it is common to consider the same value and to take 4% of uncertainty, as we have done.

This correction is a small effect.

4.5.2.5 Luminosity

As explained in Subsection 3.2.2, the luminosity is measured while CMS is acquiring data. The error on the total luminosity is of 2.6%. It is just a global factor and the computation, and the computation is easy, as for the cross sections.

4.5.2.6 Unfolding procedure

The building of the unfolding matrix is done with the statistics at our disposal. Most of the off-diagonal bins have large errors.

I did not find the time to compute properly the systematic of the unfolding. I have rather used the the basic unfolding *bin by bin* and taken the difference as an estimation of the error due to the unfolding.

4.5.3 Final plots

We are now able to conclude on the agreement of the data and the simulations, both at detector and physical levels. For more readability, the errors have been drawn around the ratio as well: the statistical error is in dark gray and the systematic error in light gray.

4.5.3.1 At detector level

The final plot of the dimuon invariant mass is given on Figure 4.11: it represents the number of events [5] measured for a given invariant mass.

4. Given online on the internal pages of CMS.

5. The number of events is given by the area and not by the height of a considered bin.

The agreement between the data and the MC prediction is very good till 500 GeV (*i.e.* compatible within one sigma). The misalignment between the peaks of the simulation and of the data does not raise any problem. But there clearly is a tendency to go down at high invariant mass.

4.5.3.2 At unfolded level

The final plot given on Figure 4.12 presents the measured differential cross section obtained after subtraction of the background (thanks to the MC distributions), unfolding and rebinning (thanks to the response matrix built with the simulations of the processes and of the detector), and dividing by the luminosity (according to 3.4).

The agreement between the unfolded level and the generated MC prediction is still good, until 500 GeV. Further, one notices the excess of data with respect to MC prediction

We are now going to discuss what could explain the tendency of going down at high invariant mass. Indeed, recall from Chapter 2 that we would like to put a form factor in the cross section; but the MC distribution can only be lowered.

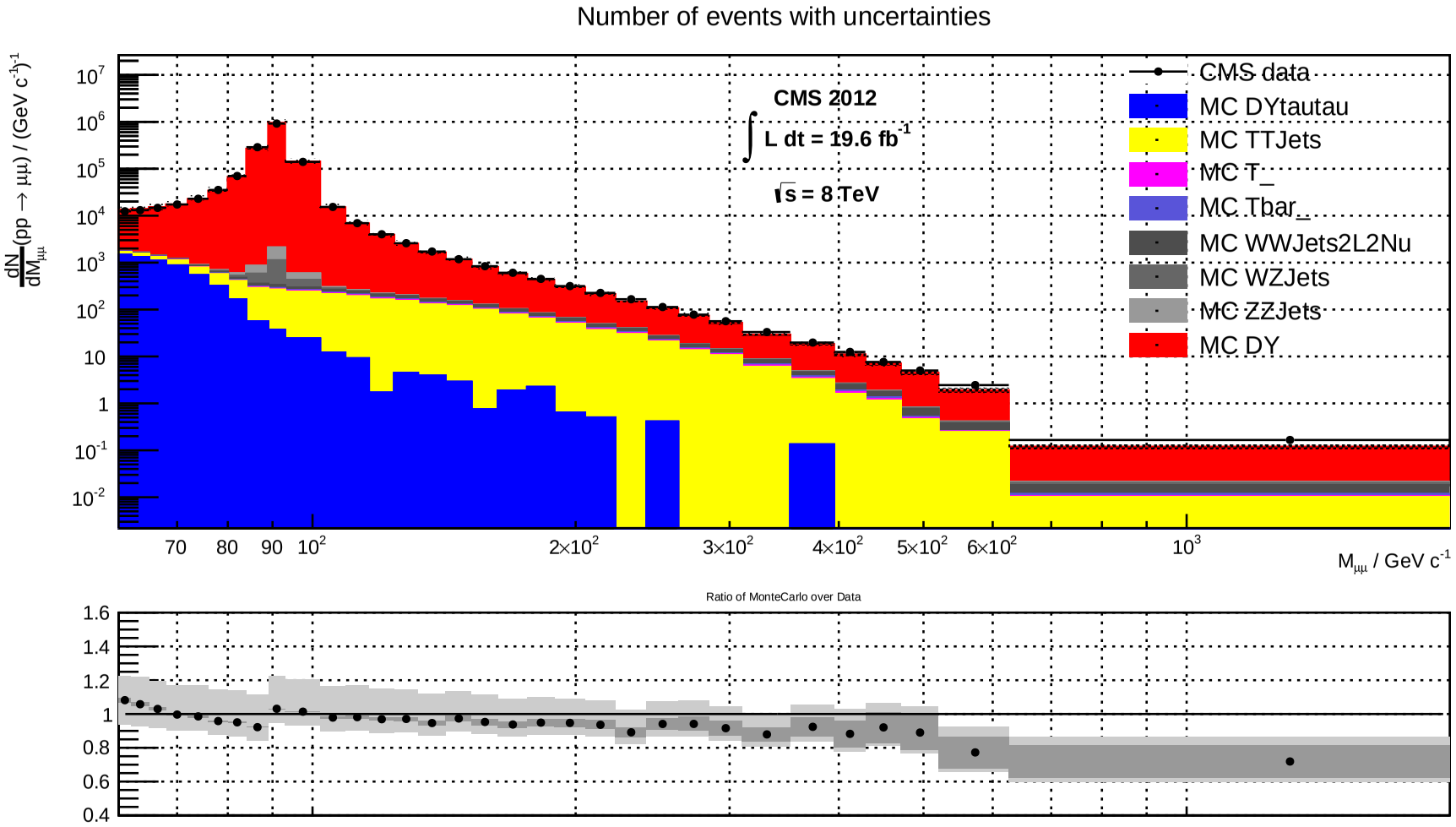


FIGURE 4.11 – Final plot of dimuon invariant mass spectrum of the Drell-Yan process acquired at CMS in 2012 at $\sqrt{s} = 8\,\text{TeV}$ in the centre-of-mass system from 60 GeV to 2 TeV. The MC simulations are stacked and compared to the Data. The statistic and systematics uncertainties are drawn in hatching around the stack but are quite small.

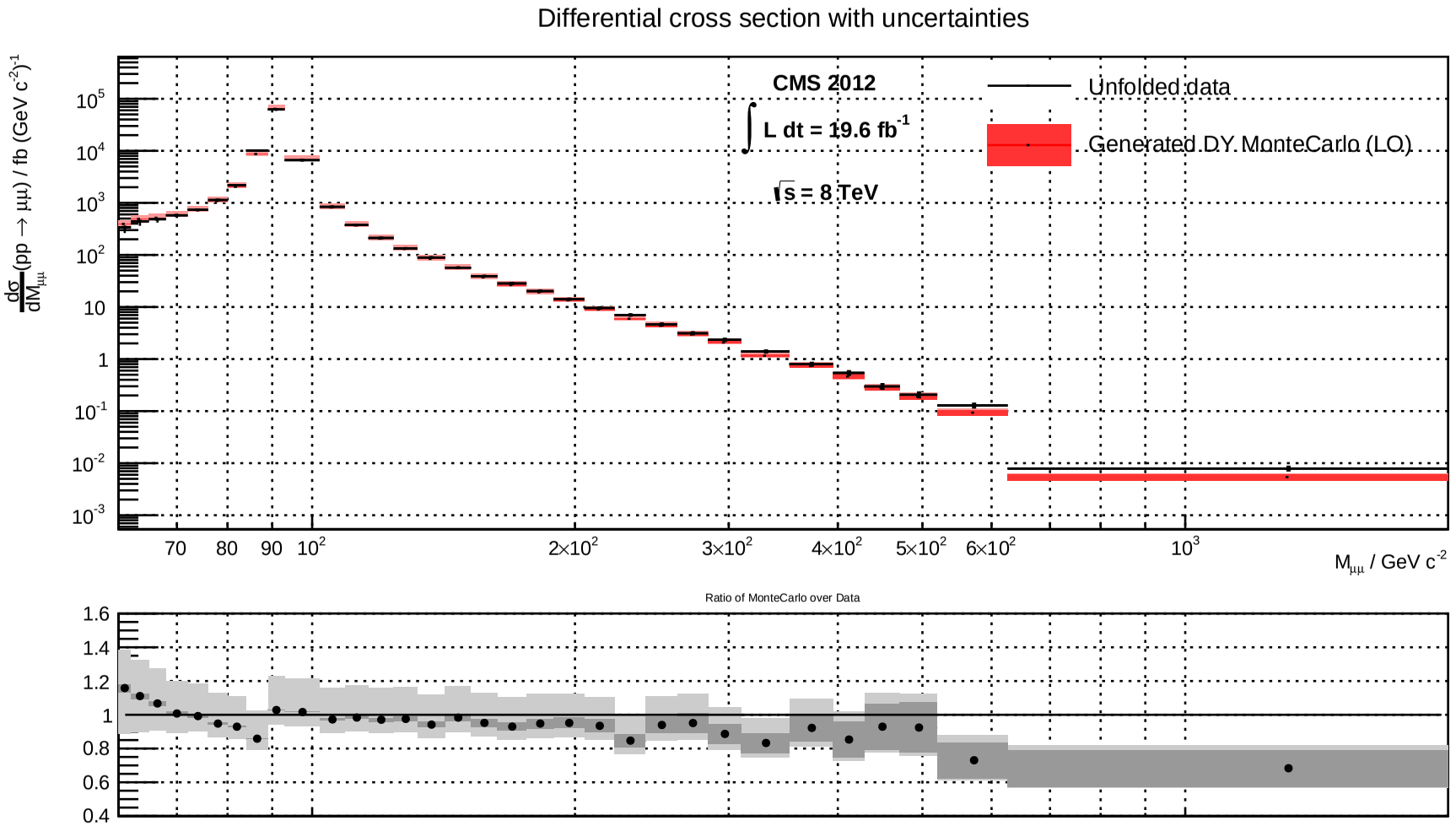


FIGURE 4.12 – Inclusive differential cross section in the dimuon invariant mass of the Drell-Yan process after rebinning and unfolding of CMS data in 2012 at $\sqrt{s} = 8\,\text{TeV}$ in the centre-of-mass system and of MadGraph prediction from 60 GeV to 2 TeV.

4.6 Short discussion of the current analysis [6]

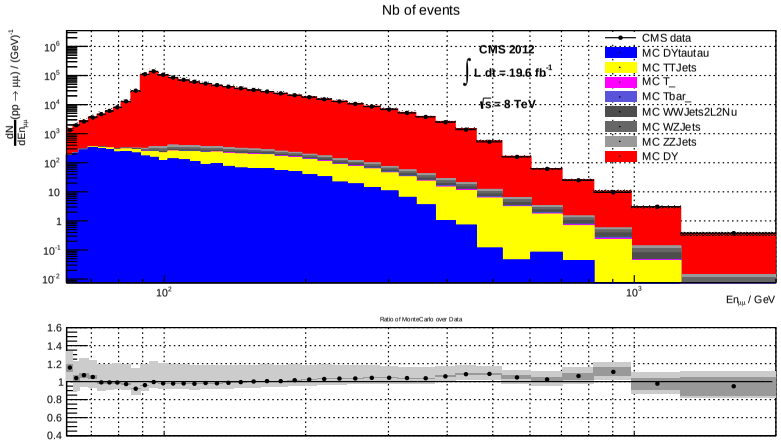
Figures 4.13 and 4.14 show the result of the analysis on the kinematic variables (some more control plots are in Appendix C). The agreement seems rather good, save for high transverse momenta [7].

It has to be said that the resolution of the highest transverse momentum might be overestimated in the MC prediction; this could induce an effect such as the one that we see for the plot of the transverse momentum of the muon pair. Indeed, the muon transverse momentum is not easy to estimate in the tail of the distribution. Such an effect is worth studying, but we are presently not going to do it: this agreement is globally correct, and we can work out a limit value for the quarks' size from it.

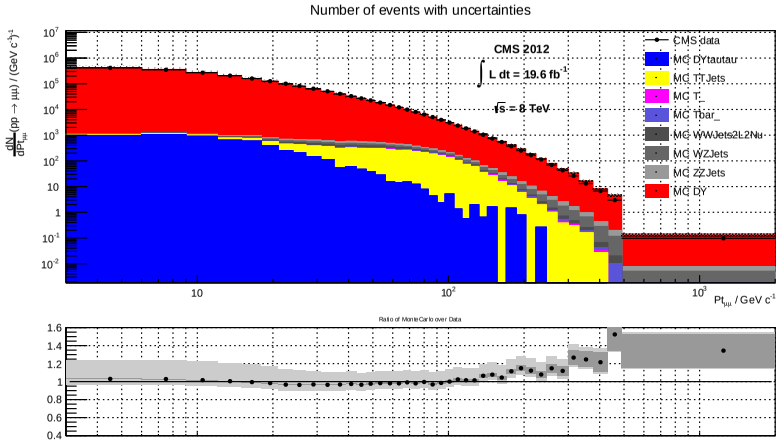
A zoom on the errors shows that the statistical error clearly dominates. It might be interesting to repeat the analysis with data at higher energy in few years.

6. The confrontation with other studies will be done in Chapter **??**.

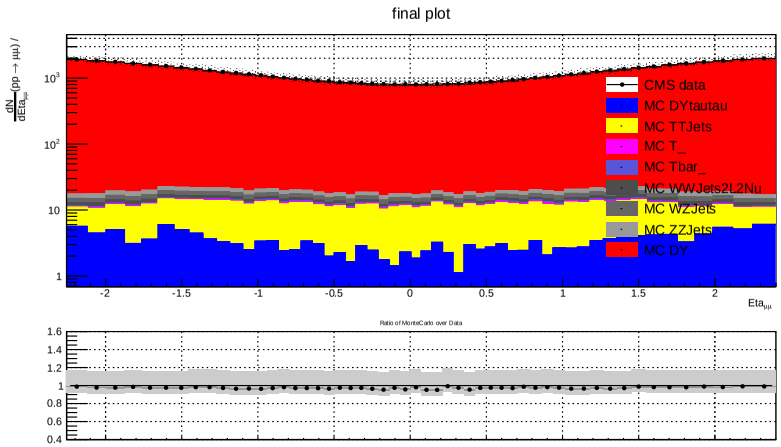
7. Such an overstimation of the decrease of the transverse momentum by MadGraph has also been observed by other internal CMS notes.



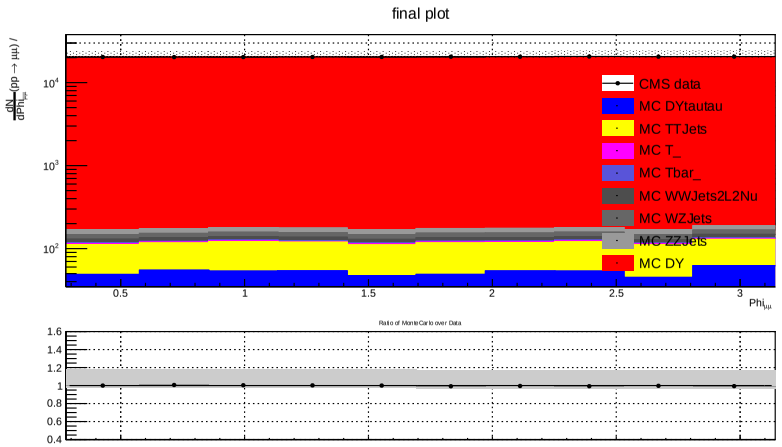
(a) Energy.



(b) Transverse momentum.

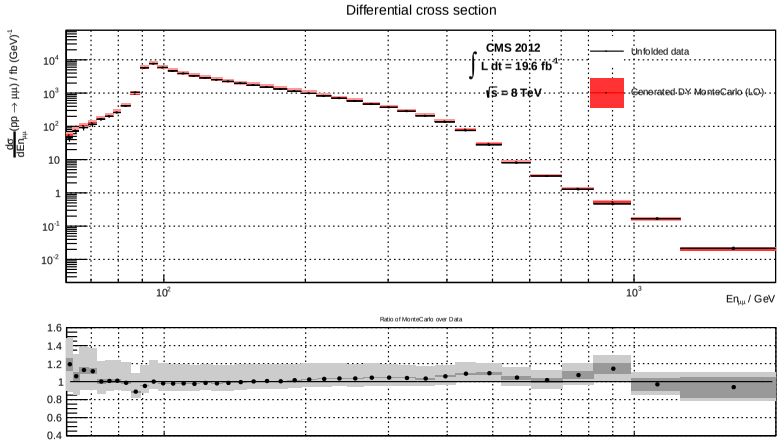


(c) Pseudorapidity.

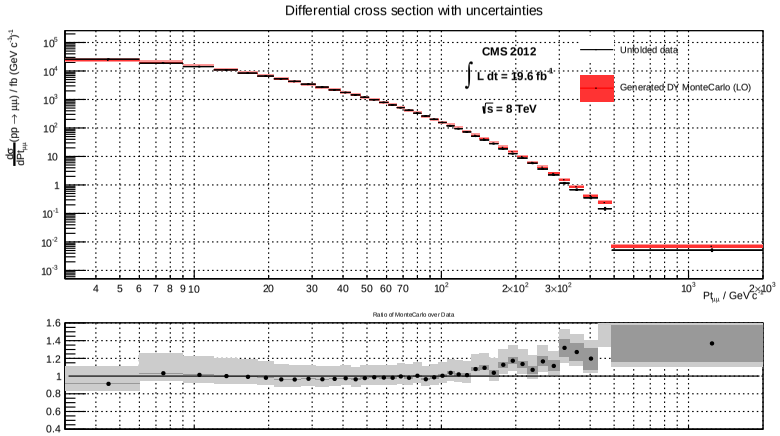


(d) Azimuthal angle.

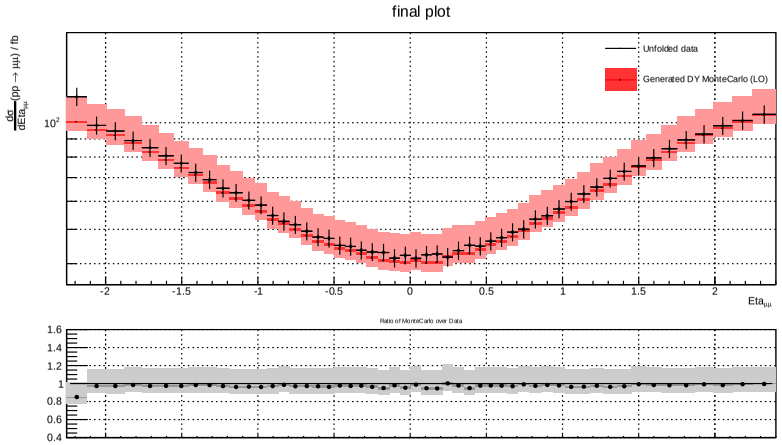
FIGURE 4.13 – Muon pair variables at detector level.



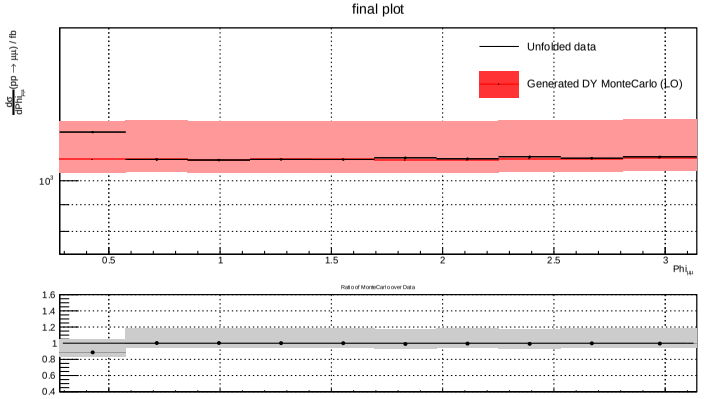
(a) Energy.



(b) Transverse momentum.



(c) Pseudorapidity.



(d) Azimuthal angle.

FIGURE 4.14 – Muon pair variables at detector level.

5

Results

??

NOW THAT THE DIFFERENTIAL CROSS SECTION has been obtained, we are able to measure a size to the quarks. We are first explaining briefly the computation; then we are comparing our measurement to other attempts.

5.1 The size of the quarks

A priori, each event at the parton level should be reweighted by a factor $\frac{\sigma_{\text{FF}}}{\sigma_{\text{th}}}$ one by one during the making of the histogram. But as the form factor that we are including in the cross section is just a global dipole, and as $Q^2 = M^2_{\mu\mu}$, it can be factoured out of the hadron cross section:

$$\frac{\sigma_{\text{FF}}}{\sigma_{\text{th}}} = |G(Q^2)|^2 = \frac{1}{(1 + \frac{Q^2}{\Lambda^2})^4} \tag{5.1}$$

and finally directly applied on the final histogram of the MC differential cross section 4.12. As the MC simulation goes down at high dimuon invariant mass, it is more sensible to look for a new limit on the size of the quarks than to pretend measuring a size. In other words, we are going to test the compatibility at two standard deviations between our data and the reweighted generated MC

5.1.1 Procedure [1]

The procedure is the following: the number of event of each bin is considered as the central value of a Gaussian law whose uncertainty is the total uncertainty (systematic plus statistic uncertainties of both distributions). Thus its square is a χ^2 distribution of degree 1, and their sum follows a χ^2 distribution of degree k, where k is the number of bins considered:

$$\chi^2_k = \sum_{i=1}^{k} \left(\frac{N_i^{\Lambda} - N_i^{\text{obs}}}{\sigma_i^{\text{tot}}} \right)^2 \tag{5.2}$$

We then run on the value of Λ and test the value of χ^2_k to reject values below Λ at two standard deviations of confidence level.

1. The χ^2 is fully detailed in the course [2] (available online).

5.1.2 Upper limit on the root-mean-square radius

In our case, at two standard deviations, we have got

$$\Lambda = 4.9 \times 10^{-19}\,\mathrm{m} \tag{5.3}$$

where Λ is the parameter of the global dipole of 2.31. We deduce the upper RMS radius:

$$\sqrt{\langle r^2 \rangle} \leq 2.7^{-19}\,\mathrm{m} \tag{5.4}$$

with a confidence level of 95%, which is around half the value measured at HERA just below.

Note that our calculus assumes pointlike muons.

5.2 Confrontation

5.2.1 Electron-proton collisions [2]

The H1 Collaboration has studied the size and the compositeness of the quark. A simple global form factor is considered:

$$f(Q^2) = 1 - \frac{1}{6} Q^2 \langle r^2 \rangle \tag{5.5}$$

They measured $\langle r^2 \rangle \leq 6.5 \times 10^{-19}$ m with the full data of H1, which, by the way, is *the world's last strongest limit on the quark size*!

5.2.2 Hadron-hadron collisions [3]

Effective four-fermion *contact interactions* (CI) would be a sign for quark compositeness. Quark and lepton compositeness, as sets of *preons* for example [4], would be a way to explain the three generations of leptons and quarks. Interaction terms of order 4, prefactored with a new coupling λ, are introduced in the Lagrangian of the SM and lead to a correction of the differential cross section. It is another method to introduce a new parameter in the cross section of proton-proton collisions. However, the effect opposite to our method: instead of lowering the cross section, the new coupling λ makes the curve raise.

The CMS and ATLAS Collaborations have both studied CI with 2011 data with $\sqrt{s} = 7$ TeV of energy in the centre-of-mass system and with four times less luminosity. Figure 5.2 shows that there is less sensitivity in both studies than in ours and no upper limit for the quarks' size is given in any of both papers.

2. See [9] for the article.
3. See [8] for CMS' analysis and [4] for ATLAS' analysis.
4. According to the theory of Jogesh Pati and Abdus Salam conceived in the 70'.

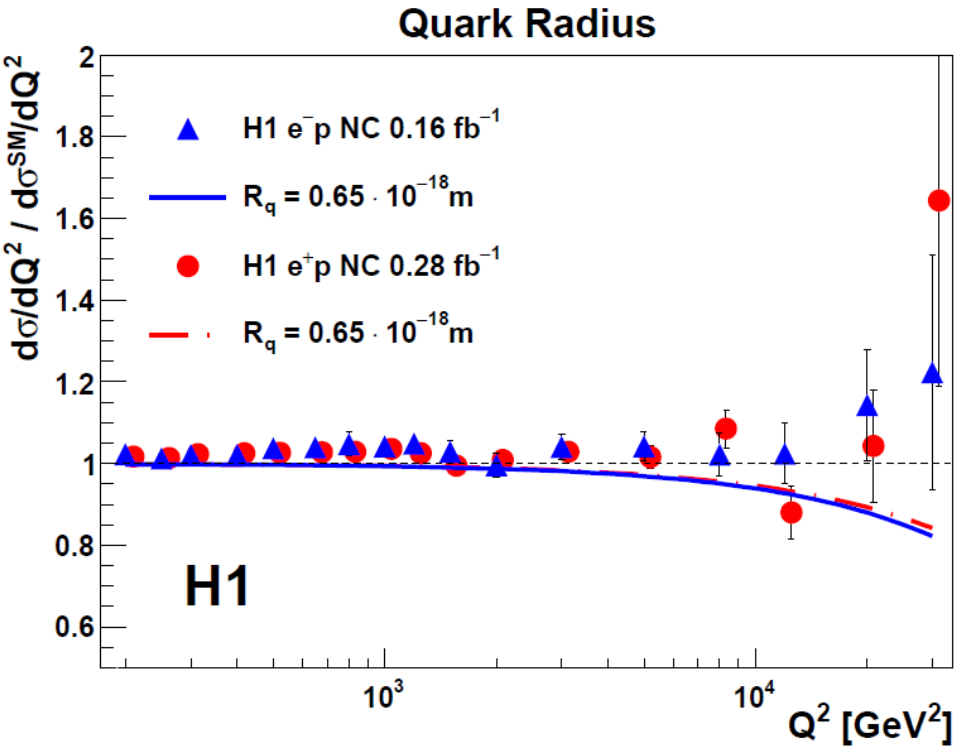
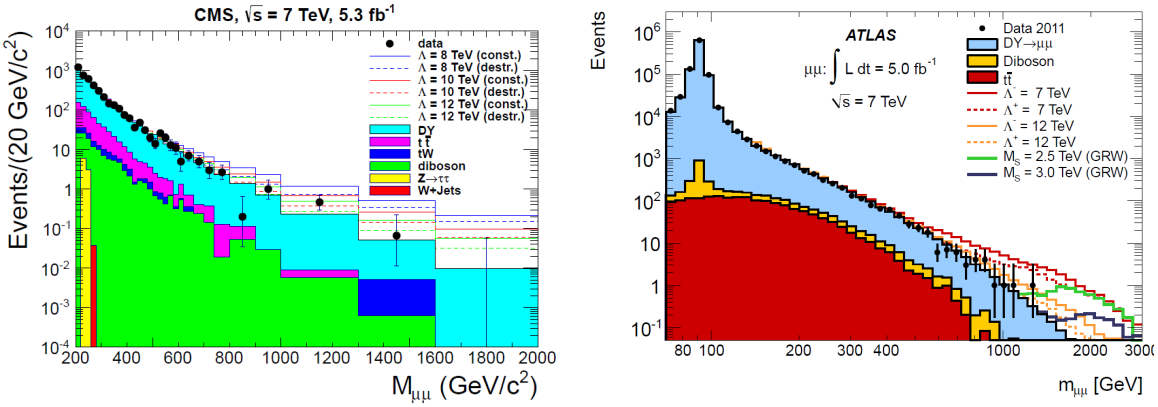


FIGURE 5.1 – Ratio of the measured neutral current cross section $\frac{d\sigma}{dQ^2}$ normalised to the SM prediction. The curves correspond to the exclusion limits at two standard deviations obtained from the full H1 data and the points to the experimental measurements.



(a) CMS analysis.

(b) ATLAS analysis.

FIGURE 5.2 – Dimuon invariance mass plot for the research for quark CI (the λ parameter corresponds to the coupling).

5.3 Final discussion of the measurement

So the conclusion is nice: we might well have derived a new upper limit for the size of the quarks. It could even be improved if the treatment with the treatment of the uncertainty on the procedure of unfolding.

But at least two reservations should be expressed:

1. the scale factors to correct the MC simulations are not computed by the Muon POG higher than $p_T = 500$ GeV;
2. the backgrounds at such an energy scale are not well known.

We may thus conclude that this method is efficient for putting a limit on the size of the quarks but perhaps not to measure it. It will certainly be worth while using it when CMS turns at higher energy and higher luminosity.

A

Mandelstam variables

MANDELSTAM VARIABLES are useful to compute common cross section. We here take some time to describe them more.

A.1 Framework

Definition In the framework of Figure A.1, Mandelstam variables are three conventional, Lorentz-invariant, numerical quantities commonly used in the computation of cross sections. They are defined as follows:

$$s = (p_1 + p_2)^2 \tag{A.1}$$
$$t = (p_1 - p_3)^2 \tag{A.2}$$
$$u = (p_1 - p_4)^2 \tag{A.3}$$

with the conservation of four-momentum:

$$p_1 + p_2 = p_3 + p_4 \tag{A.4}$$

Differential cross section In terms of the Mandelstam variables and in the centre-of-mass frame, the differential cross section reads

$$\frac{\mathrm{d}\sigma}{\mathrm{d}t} = \frac{1}{64\pi s}\frac{1}{|p_{\mathrm{cm}}|^2}\overline{|\mathcal{M}|^2} \tag{A.5}$$

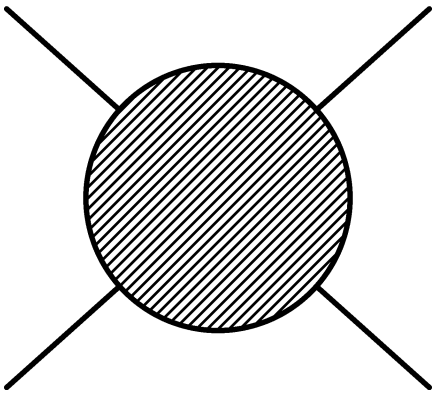


FIGURE A.1 – General diagram of the kinematics.

A.2 Properties

General Let m_i be the masses of the interacting particles:

$$s + t + u = \sum_{i=1}^{4} m_i^2 \tag{A.6}$$

Center-of-mass frame Let θ_{CMS} be the angle between the incident outgoing fluxes:

$$t = \frac{-s}{2}(1 - \cos\theta_{\text{CMS}}) \tag{A.7}$$

$$u = \frac{-s}{2}(1 + \cos\theta_{\text{CMS}}) \tag{A.8}$$

$$\tag{A.9}$$

and the differential cross section reads

$$\frac{\mathrm{d}\sigma}{\mathrm{d}t} = \frac{1}{16\pi s^2}\overline{|M|^2} \tag{A.10}$$

High energy limit The masses being insignificant:

$$s \approx 2p_1 \cdot p_2 \tag{A.11}$$

$$t \approx -2p_1 \cdot p_3 \tag{A.12}$$

$$u \approx -2p_1 \cdot p_4 \tag{A.13}$$

Deeply Inelastic Scattering The inelasticity y may be related to the Mandelstam variables:

$$y = \frac{-t}{s} \tag{A.14}$$

A.3 Channels

Definition For *tree-level* diagrams, that is, diagram without loops, one defines three classes, or *channels*, of reactions:

— s-channel,

— t-channel,

— and u-channel.

The terminology is inherited from nuclear physics (see Charles Joachain [18]). The s-channel is the only way to discover new resonances.

Changing of channel Two processes might be physically different but look very similar as Feynman diagrams. Indeed, the dynamics of a process depends on the couplings of the diagram whereas the kinematics depends only on its external lines. In other words, turning a diagram does not change its dynamics, but only permute indices in the relation A.4. In terms of Mandelstam variables, going from the s-channel to the t-channel reduces to the permutation of the variables s and t in the scattering matrix element $\mathcal{M}$. For instance, in

this master thesis, the deeply inelastic scattering is a process in the t-channel involving two fermions resulting into two fermions, whereas the Drell-Yan process is in the s-channel, involving a pair of fermion-antifermion resulting in a pair of fermion-antifermion; the cross section of the Drell-Yan process including form factors can be deduced from the cross section of the electron-proton scattering.

B

Programs

MOST OF THE PRACTICAL WORK achieved for this master thesis remains in the writing of `C++` codes. This appendix details them briefly.

B.1 General organisation

Any analysis is unique, but of course, some steps are common. For example, the unfolding (see Chapter 4) is usual, and has been written in `RooUnfold`. Let us review one by one the different libraries that I have used all along the analysis: first the ones that I have not programmed, secondly my own ones.

B.2 Libraries

B.2.1 `Root`

`Root` is a standard platform for HEP programming, mostly written in C++ by René BRUN and Fons RADEMAKERS. It provides hundreds of classes to help histogramming, fitting, computing, etc.

The daily use is `Root` consists of coding macros and using them interactively in a prompt window. All my codes have been written within this framework.

B.2.2 `RooUnfold`

`RooUnfold` is a standard library for unfolding the real data thanks to the response matrix, programmed by Tom ADYE. It provides several method:

— bin-by-bin,

— D'Agostini method,

— SVD,

— TUnfold,

— ...

We have used it to unfold bin-by-bin and with D'Agostini method. It automatically manages the evolution of the statistical error through the procedure.

B.2.3 Pile-Up correction

The correction for the pile-up is provided by the Muon POG. It directly uses the trees while the selection is operating to compute correction factors to rescale the simulation, as if the pile-up had been correctly simulated.

B.2.4 LHAPDF

Les Houches Accord PDF is « general purpose C++ interpolator, used for evaluating PDFs from discretised data files ». It has been programmed by physicists and ITs from Durham University.

It can be used to compute the Parton Density Functions, describing the constitution of the hadrons. It provides many measurements from different experiments with various techniques of fitting:

— CTEQ6

— HERAPDF

— NNPDF

— ...

We have used it to compute the systematics of the MC simulation of the Drell-Yan process (Section 4.5).

An online PDF plotter is available on `http://hepdata.cedar.ac.uk/pdf/pdf3.html`.

B.3 Personal codes

B.3.1 `HSelect`

`HSelect` is a big class that is used to extract histograms from the trees. The same code is used for every simulation and for the data.

It applies to

— the selection,

— the scale factors to the simulations (possibly biased for the computation of the systematics),

— the pile-up correction to the simulations (same remark),

— the rescaling of the MC for the choice of the PDF (only used to compute the systematics),

— and is already programmed to correct the MC simulations with reweighting factors containing non-global form factors, though it has been used yet to measure the size of the quarks.

It produces many plots: for the kinematic variables and for the detector variables. It also builds the response matrices for the kinematic variables.

The histograms are saved in `.root` files (one per tree).

B.3.2 `HProcess`

`HProcess` combines the histrograms generated from the trees intro *processed* histogrammed, *i.e.* :

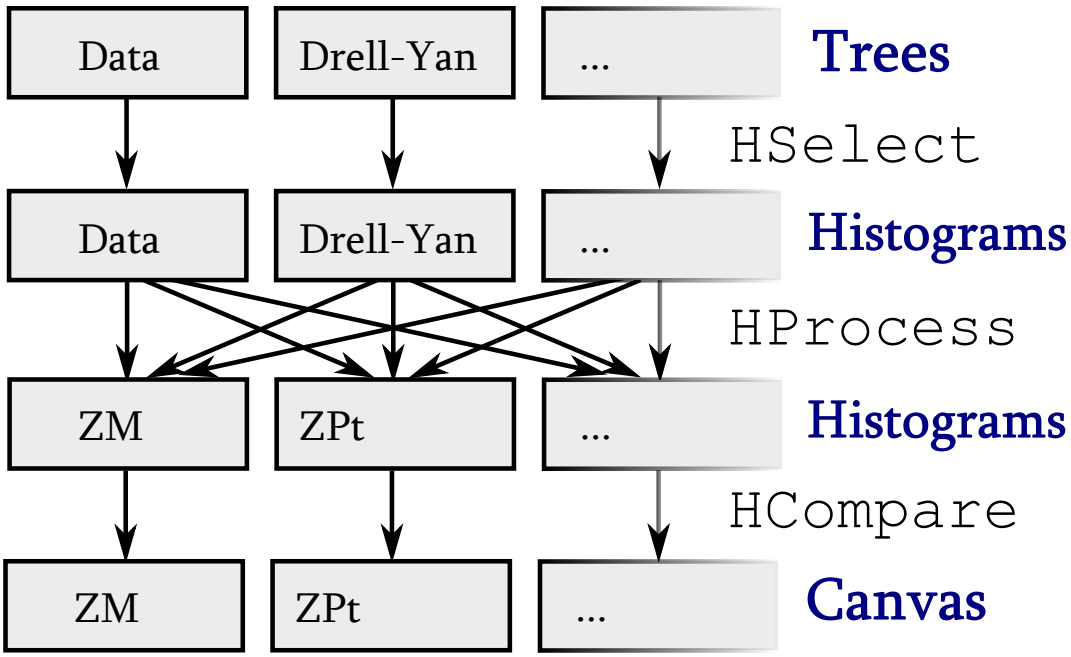


FIGURE B.1 – Diagram of the organisation of the classes

.

— it rescales to the luminosity;

— it builds the stack of the MC;

— it rebins and unfolds thanks to the response matrix;

— and it computes the systematics.

One variable is treated at a time, and the final histograms are saved in a .root file.

B.3.3 `HCompare`

`HCompare` is used to superpose two histograms and compute their ratio. It also draws the uncertainties.

Using `Root` to display delicate histograms is not easy and takes much time.

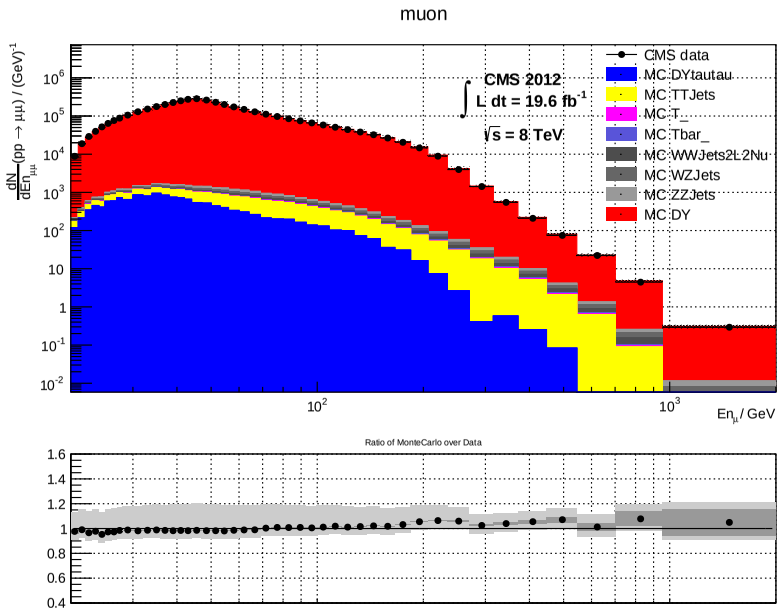
C

Control Plots

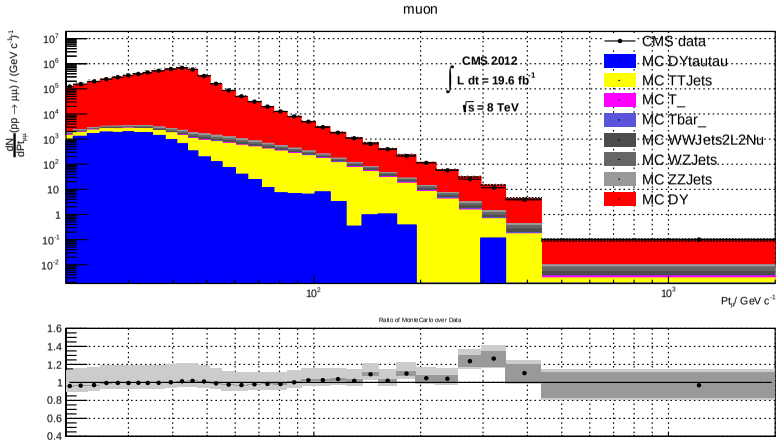
The analysis is made of many steps. It is necessary to control regularly on different variables at each step that the analysis is doing well. To lighten the discussion of the analysis in Chapter 4, most of the control plots have been postponed in this appendix.

C.1 Selection of the muons

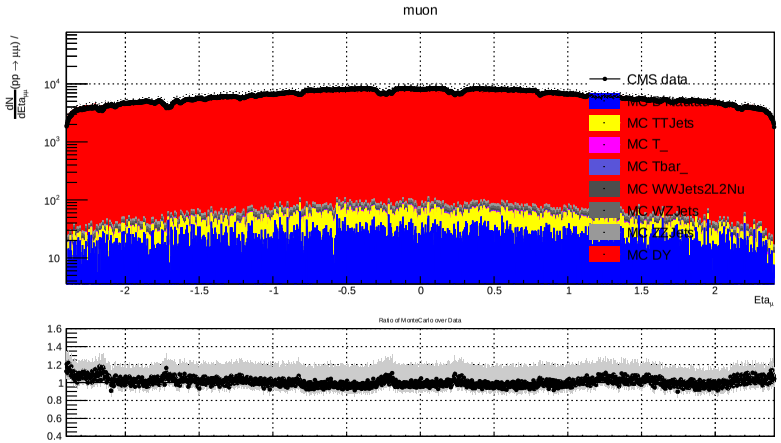
The very first check-up to do is to look at the muon variables step by step. We give on Figure C.1 the result. The agreement is almost perfect, save a little fluctuation for the transverse momentum of the muons.



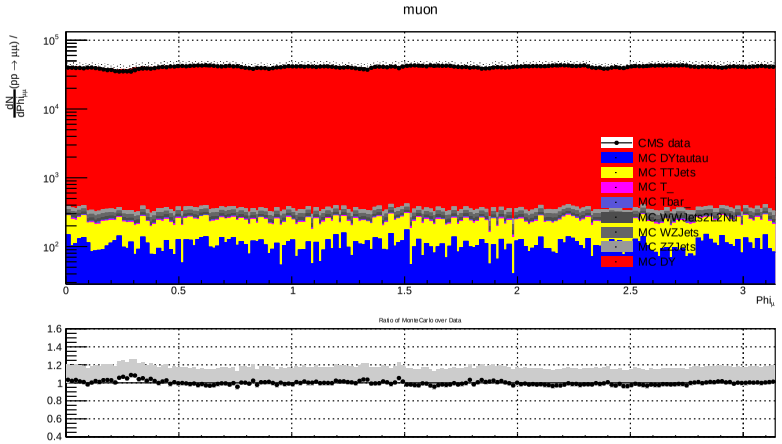
(a) Energy.



(b) Transverse momentum.



(c) Pseudorapidity.



(d) Azimuthal angle.

FIGURE C.1 – Muon variables at detector level.

C.2 Unfolding

C.2.1 Bin by bin

The bin-by-bin unfolding consists of taking the ratio « bin by bin » of the distributions of the DY simulations at both simulated and generated levels to describe the effect of the detector. This method is not as clean as D'Agostini's, as it treats the migration effect as a efficiency effect for each bin, but is much easier to use and confirms the result obtained in Chapter 4.

C.2.2 Miss and Fake

During the selection operated on the DY simulation at generated and reconstructed levels, some events might pass the cutoffs at only one level:

miss the event has passed the generated level but not the reconstructed level;

fake the event has passed the reconstructed level but not the generated level.

On Figure C.3, one can see the distribution of invariant mass of fake and miss events. There are much more miss than fake.

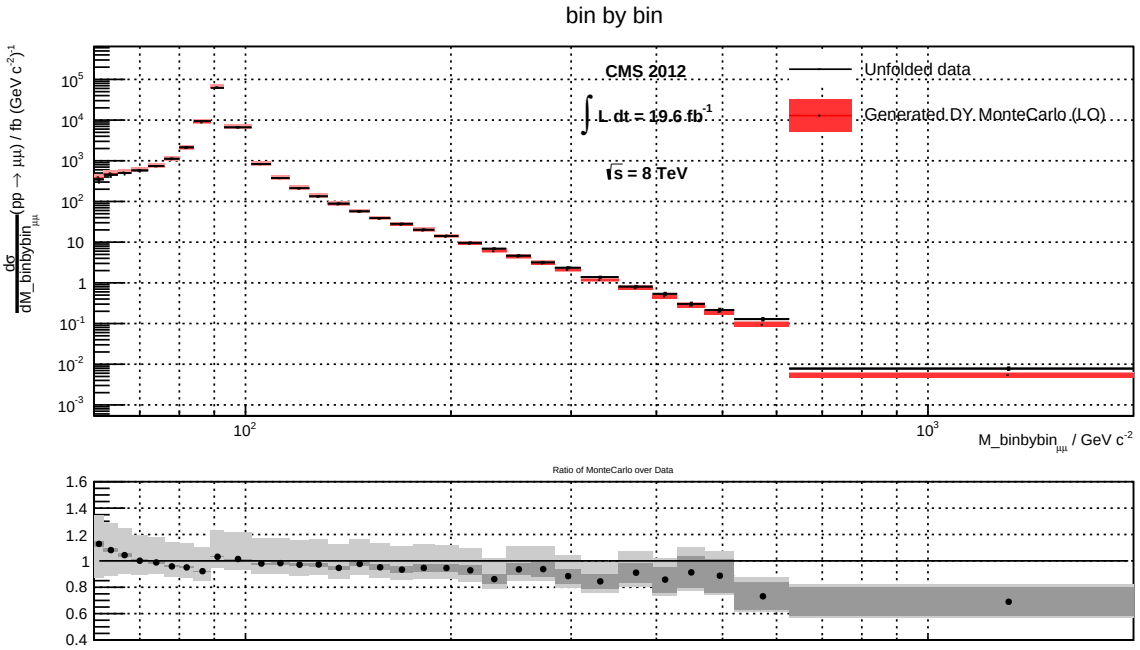


FIGURE C.2 – Unfolding bin by bin.

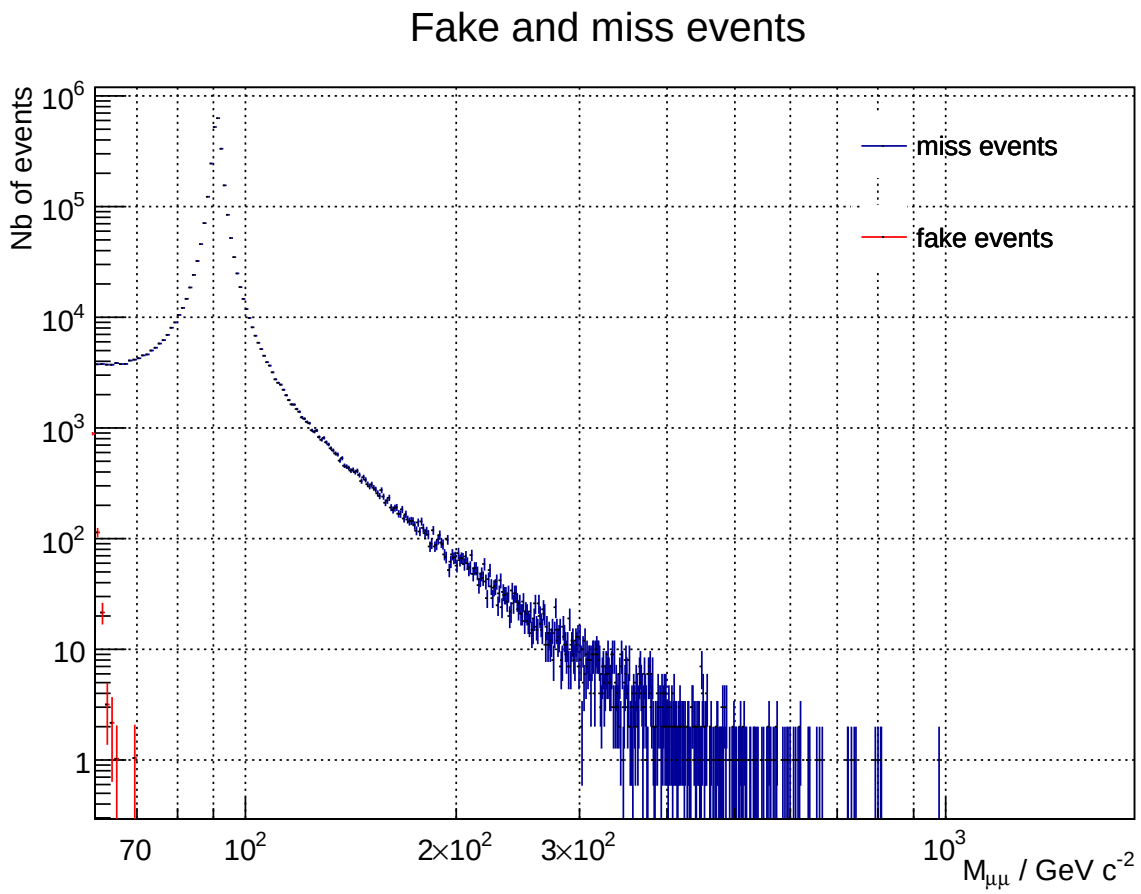


FIGURE C.3 – Invariant mass plot of the fake and miss muon pairs.

C.3 Other

C.3.1 Momentum fraction

The generated level of the MC Drell-Yan contains some information on the incident quarks, as the momentum fraction. It is represented on Figure C.4.

C.3.2 Number of vertices

Evidence for an elastic background might have been found in the number of tracks (Figure C.5). Unfortunately, there is no such evidence at small values. The deviation in the agreement is due to the fact that it is extremely difficult to simulate the radiations.

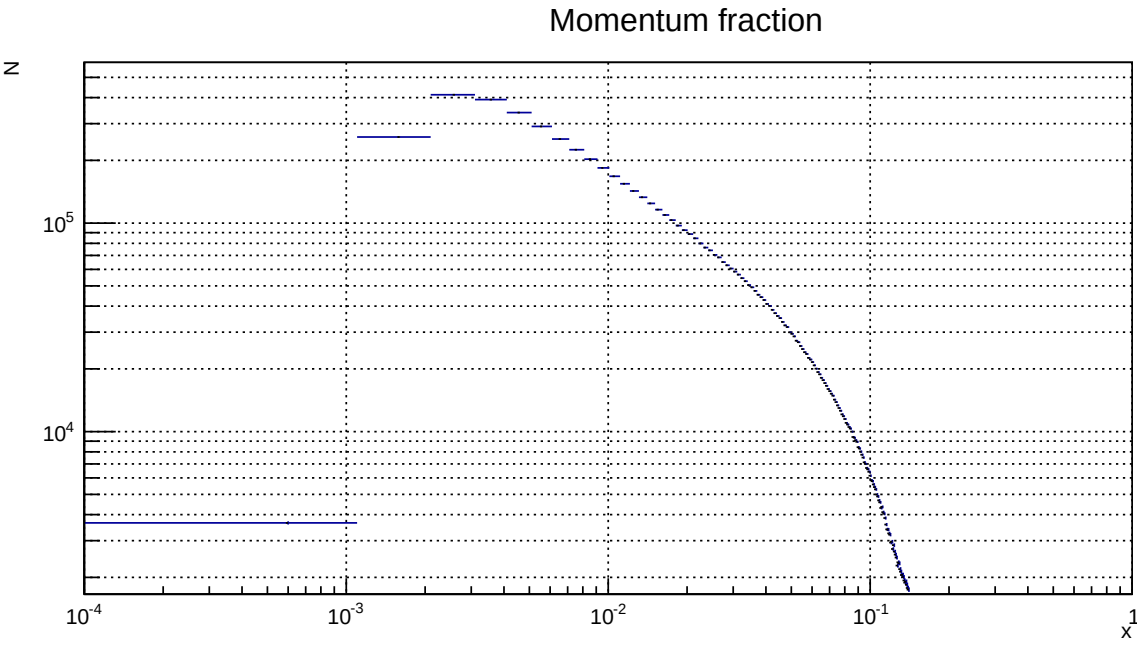


FIGURE C.4 – Momentum fraction of the incident quarks.

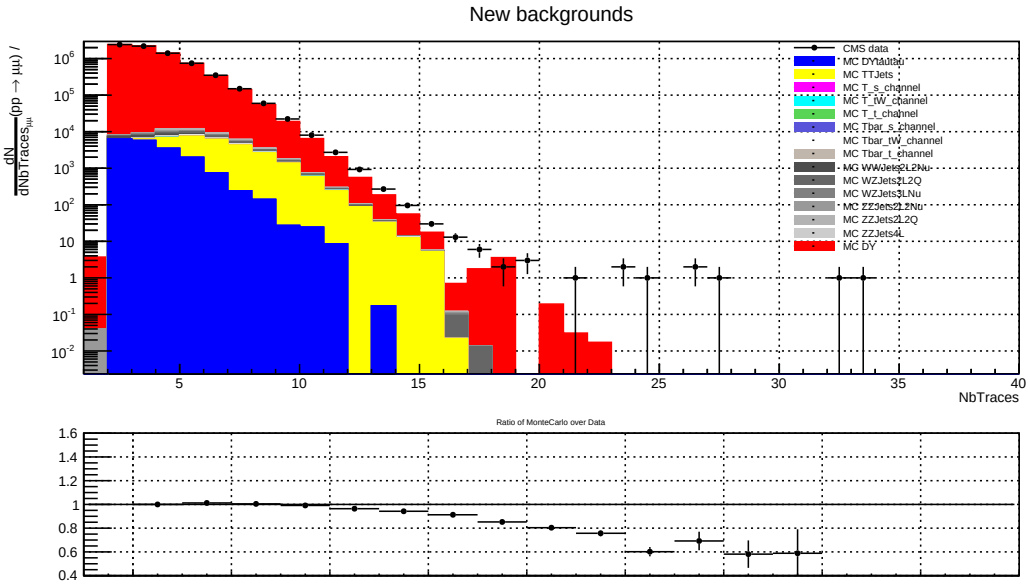


FIGURE C.5 – Number of events as a function of the number of tracks.

Conclusions and Perspectives

In this master thesis, we have studied the Drell-Yan process at leading order in the muon channel and put a limit on the size the quarks.

In order to do this, we have studied some theoretical preliminaries on hadron-hadron collisions and the experimental setup of the CMS detector: we first recalled the structure of the proton in terms of Parton Densities Functions, and we showed how to use them to treat a parton process with the factorisation theorem. We then introduced the Drell-Yan process, and explained why it was a good choice to measure a thin effect as the size of the quarks. We computed the cross section and discussed the insertion of form factors in its analytical form.

We then moved to the description of the experiment setup, the CMS detector at LHC. We explained as much as necessary the different steps of muon detection. Keeping the notions of trigger and reconstruction track in mind, we could begin the analysis: selection of the events, making the histograms, combining the histograms of MC and data, unfolding, rebinning and finally computing the uncertainties. We obtained an invariant mass plot of the dimuon invariant mass spectrum and we observed a slight divergence at high values. We checked on other diagrams that our data was indeed utilisable.

Eventually, we sought to use it to put an upper limit on the size of the quarks. Indeed, no experience had ever been done at so high energy as CMS (and parallelly ATLAS), opening new regions of the phase space for high energy scale and high momentum fraction. We succeeded in this task and found a new upper limit.

However, it could be still improved by considering more uncertainties, like the unfolding or the reconstruction of the particles. It will also be improved, as the LHC is going to run at higher energy and at higher instantaneous luminosity. The statistics will become more important and more interesting.

On the other hand, the simulations should be studied in detail. It is known that MadGraph produces to much radiation at high transverse momentum.

We may thus conclude as so: our goal has been reached, but there is still much work to do to improve the sensitivity of the technique.

Bibliography

[1] Glenn Barnich. Théorie quantique des champs 2. Course of 1st master year, ULB, 2013.

[2] Julien Barnier. Tout ce que vous n'avez jamais voulu savoir sur le χ^2 sans jamais avoir eu envie de le demander. Course of the École Normale Supérieure de Lyon, available online., 2013.

[3] CERN. LHC the guide. http://cds.cern.ch/record/1165534/files/CERN-Brochure-2009-003-Eng.pdf.

[4] ATLAS collaboration. Search for contact interactions and large extra dimensions in delepton events from pp collisions at $\sqrt{s} = 7$ tev with the ATLAS detector. *Physical Review D*, 2013. Submitted.

[5] Cern collaboration. *The CERN Large Hadron Collider: Accelerator and Experiments*, 2009.

[6] CMS collaboration. CMS public web site. `http://cms.web.cern.ch`. Accessed in january 2014.

[7] CMS Collaboration. Measurement of cms luminosity. *Internal analysis note*, 2012.

[8] CMS collaboration. Search for contact interactions in $\mu^+\mu^-$ events in pp collisions at $\sqrt{s} = 7$ tev. *Physical Review D*, 2013. Submitted.

[9] H1 collaboration. Search for contact interactions in $e^{\pm}p$ collisions at HERA. *Physical Letter B*, 2011. Submitted.

[10] G. D'Agostini. A multidimensional unfolding method based on bayes' theorem. *Nuclear Intstruments & Methods in Physics Research A*, 362:487–498, 1995.

[11] Durham HepData Project. Durham PDF plotter. http://hepdata.cedar.ac.uk/pdf/pdf3.html.

[12] Ellis, Stirling, and Weber. *QCD and Collider physics*. Cambridge University Press, 1996.

[13] Laurent Favart. Physique auprès des collisionneurs. Course of 1st master year, ULB., 2013.

[14] Bjorn Felsager. *Geometry, Particles, and Fields*. Springer, 1998.

[15] R. Frühwirth. Application of kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research A*, 262:444–450, 1987.

[16] Particle Data Group. Particle physics booklet, 2012.

[17] Francis Halzen and Alan D. Martin. *Quarks & Leptons.* John Wiley & Sons, Inc., 1984.

[18] Charles Joachain. *Quantum collision theory.* Elsevier Science Ltd, 1984.

[19] I. R. Kenyon. The drell-yan process. *Reports on Progress in Physics*, 45:1261–1315, 1982.

[20] Django Manglunki. Calcul, technique et réalisation des accélérateurs de particules. Course of 2nd master yead of civil engineering, ULB, given at CERN, 2013.

[21] John Morris. Hep analysis. Two-day seminar on Monte Carlo in HEP analysis at Queen Mary University of London, 2012.

[22] C.F. Perdrisat, V. Punjabi, and M. Vanderhaeghen. Nucleon electromagnetic form factors. *Progress in Particle and Nuclear Physics*, 2013.

[23] E. Perez and E. Rizvi. The quark and gluon structure of the proton. *Reports on Progress in Physics*, august 2012. Submitted.

[24] Donald H. Perkins. *Introduction to High Energy Physics.* Addison-Wesley Publishing Company, Inc., 3rd edition, 1987.

[25] Micheal E. Peskin and Dan V. Schroeder. *An Introduction to Quantum Field Theory.* Frontiers of Physics, 1995.

[26] S. Tavernier. *Experimental techniques in Nuclear and Particle Physics.* Springer-Verlag, 2010.

List of Figures

List of Tables

Index