Parton Distribution Function Studies and a Measurement of Drell-Yan Produced Muon Pairs at LHCb

Francesco De Lorenzi

The thesis is submitted to University College Dublin in fulfilment of the requirements for the degree of Doctor of Philosophy in Physics

School of Physics

Head of School: Professor **Lorraine Hanlon**

Principal Supervisor: Doctor **Ronan Mc Nulty**

University College Dublin *May 2011*

CERN-THESIS-2011-237
21/03/2011

To my family.

DECLARATION

I hereby certify that the submitted work is my own work, was completed while registered as a candidate for the degree stated on the Title Page, and I have not obtained a degree elsewhere on the basis of the research presented in this submitted work.

CONTENTS

1. INTRODUCTION

The Large Hadron Collider (LHC) at CERN is going to probe our understanding of the theory which describes the subnuclear interaction. For the past few decades, physicists have been able to describe with increasing details the fundamental particles that constitute the Universe and the interactions between them. This understanding is encapsulated in the Standard Model of particle physics, but there are still important gaps in our knowledge. The upcoming experimental data from the LHC might produce unexpected results and unveil new scenarios in our understanding of the model of elementary particles. However, the correct identification of any signal of new physics requires a careful assessment of the Standard Model backgrounds. Given that the vast majority of events are due to strong interactions, a deep understanding of the phenomenology of strong interactions is fundamental in order to fully exploit the physics potential of modern colliders. This thesis describes the contribution of my research activity in the understanding of the strong interaction phenomenology at LHC through the constraint of the Parton Distribution Functions.

The work can be divided into two main parts: the description of a technique to constrain the Parton Distribution Function at LHCb and a study of the Drell-Yan events at LHCb using real data collected between March and October 2010.

In chapter 2 we describe the experimental apparatus which allowed our experimental activity. We describe the LHCb detector and we give details of the subsystems and the software algorithms which allow us to measure the momentum, the position and the energy and to identify the nature of the particles produced in the collisions. A short description of the accelerator system is also given.

In chapter 3 we give an overview of the Standard Model of Particle Physics which is the model that describes the interactions between particles and the forces behind them. A brief description of the structure of field theory is presented. A description of the Higgs mechanism is also provided.

In chapter 4 we introduce the underlying theory of the hadron collisions. We give a description of the theory and how it is possible to perform calculations in order to provide predictions of physical observables with particular attention on the Drell-Yan mechanism.

In chapter 5 we describe how parton distribution functions (PDF) are obtained with global fits to several datapoints and the open questions that these fitting procedures have. The importance of PDF measurements at LHCb is presented. We also give a survey of the current knowledge of the PDFs.

A novel method to constrain PDFs using LHCb data is described in chapter 6. Different implementations of fitting algorithm are described. The statistical consistency of the algorithms has been tested and systematic effects due to model dependence have been assessed. The method also provides a way to estimate the luminosity at LHCb.

Finally, in chapter 7 we describe a measurement of Drell-Yan produced muon Pairs at LHCb. We describe the selection criteria to select the signal and reject the backgrounds. Data driven method to evaluate efficiency and background contamination are presented.

2. THE LARGE HADRON COLLIDER AND THE LHCb EXPERIMENT

The Large Hadron Collider (LHC)[1] is a particle accelerator located in a tunnel 100 m underground at CERN. The original LHC idea dates back to 1991 and it was proposed in order to extend the Standard Model research and, in particular to search for the Higgs Boson.

2.1 The LHC Accelerator Chain

The LHC consists of two beam pipes that contain counter circulating proton beams and is approximately 27 km in circumference. 1232 Super-conducting magnets with a magnetic field of 8.5 Tesla keep the protons in a circular path. The protons, before being injected in the principal LHC ring, are accelerated by a chain of smaller accelerators which progressively increase their energy. The first system is a linear accelerator called LINAC, which brings the proton to the energy of few MeV. Then the protons are driven into the Proton Synchrotron Booster (PSB) and accelerated to 1 GeV by the Proton Synchrotron. Finally they are injected into the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV before being driven into the main LHC ring. At this stage the proton bunches are accelerated by the LHC radiofrequency cavities up to the energy of 3.5 TeV providing a centre of mass energy of 7 TeV, which makes the LHC the highest energy accelerator and collider ever built. The beams are brought into head-on collision every 25 ns in four different points where the particle detectors ATLAS[2], CMS[3], LHCb[4] and ALICE[5] are located and ready to record the particles produced in the collisions. The schematic of the CERN accelerator system is shown in figure 2.1

The LHC program also foresees collisions between Lead Ions which are accelerated to 5 TeV per nucleon which would allow the experiments, in particular the dedicated heavy

ion experiment ALICE, to put light on new states of matter, such as the so called quark gluon plasma. There are two main reasons for using beams of protons in the collider instead of antiprotons and protons, or electrons and positrons. First, since for rare events discovery an high luminosity machine is required, a high number of initial state particles has to be provided; protons are easier to produce compared to antiprotons whilst still having approximately the same total cross-section for interacting. Protons for the LHC beam are extracted from hydrogen atoms using an electromagnetic trap with high efficiency. The second reason is that there is much less synchrotron radiation loss for protons compared to electrons of the same energy. Charged particles traveling along a circular trajectory lose part of their energy per unit time (dt) through the emission of synchrotron radiation according to the relation

$$\frac{dE}{dt} \propto \frac{E^4}{m^4 R} \tag{2.1}$$

where m is the mass of the particle, E its energy and R the radius of the circular path. This means that at a given energy and radius of the accelerator, an electron beam loses a fraction of energy $(m_p/m_e)^4 \sim 10^{12}$ times larger than a proton beam. To use effectively electrons and positrons collisions one needs to build a larger circular accelerator or linear accelerator.

2.2 Luminosity

The interaction rate R of a given process is proportional to the cross-section of the specific process and the proportionality factor is called instantaneous luminosity $\mathcal{L}$. The luminosity depends on the intensity of the beams and can be expressed as:

$$\mathcal{L} = fk\frac{n_1 n_2}{4\pi\sigma_x\sigma_y}R_\phi \tag{2.2}$$

where n_1 and n_2 are the numbers of particles in each of the k bunches in each beam, σ_x and σ_y give the extensions of the bunches perpendicular to the beam direction, f is the revolution frequency and R_ϕ is a luminosity reduction factor due to the non-vanishing LHC beam crossing angle. The total number of events of a given process is given by:

$$N_X = L\sigma_X \tag{2.3}$$

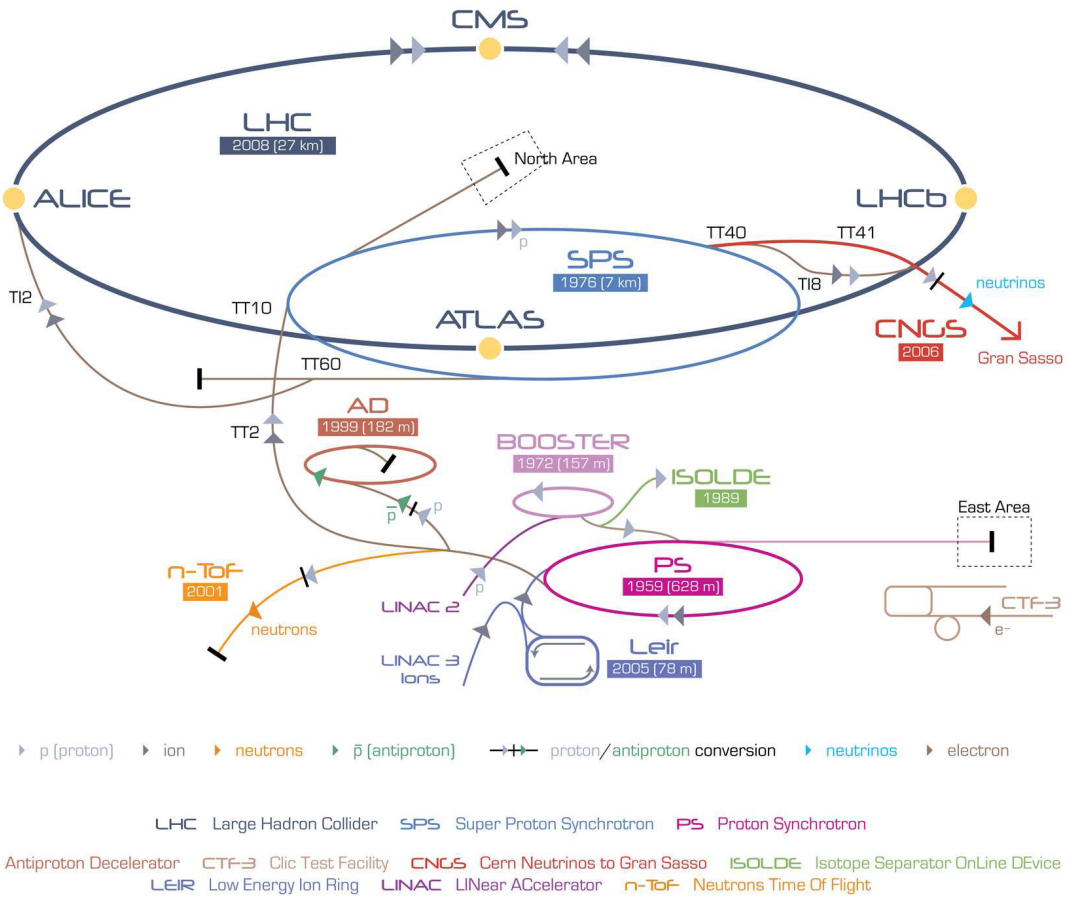


Figure 2.1: Schematic representation of the CERN accelerators system. Taken from [6].

where σ_X is the cross-section of the process and L is called the integrated luminosity. L is defined as the integral of the instantaneous luminosity over time

$$L = \int dt\mathcal{L} \tag{2.4}$$

It is clear that, for any absolute cross-section measurement, the knowledge of the luminosity has a very important role since it allows a comparison with theoretical predictions. Luminosity measurements can be made by either directly measuring the beam parameters or, by measuring the event rate of some accurately predicted physics process. At LHC, the measurement of the beam parameters can be performed through the Van der Mer Scan or at LHCb using the vertex locator (VELO) to measure the characteristics of beam-gas events near the interaction point [7].

For the running of the LHC several different luminosity phases are planned. The first phase was be at around $\mathcal{L} = 10^{31}\ cm^{-2}s^{-1}$ and it was be used for commissioning of

the collider and for first physics results. Then the luminosity will be increased to $\mathcal{L} = 10^{33}\ cm^{-2}s^{-1}$ and, hopefully, later to $\mathcal{L} = 10^{34}\ cm^{-2}s^{-1}$ which is the design luminosity.

2.3 LHCb experiment

LHCb is a dedicated B physics experiment, and aims to exploit the unprecedented quantity of b hadrons produced at the LHC. It is a single-arm spectrometer with a forward angular coverage from approximately 10 *mrad* to 300 (250) *mrad* in the vertical (horizontal) direction. The choice of the detector geometry is justified by the fact that at high energies both the b and $\bar{b}$-hadrons are predominantly produced in the same forward or backward cone as can be seen in figure 2.2 where the azimuthal angle with respect to the beam direction for b and $\bar{b}$ hadron are plotted. The design of the LHCb detector exploits this production topology by only instrumenting a small portion of the forward hemisphere corresponding to the pseudorapidity range of $1.9 < \eta < 4.9$.[1] A schematic drawing of the experiment is shown in figure 2.3, The right-handed coordinate system adopted has the z axis along the beam, and the y axis along the vertical; a first sight, it looks very similar to a fixed target experiment, but it actually looks at the result of the proton proton collision provided by the LHC.

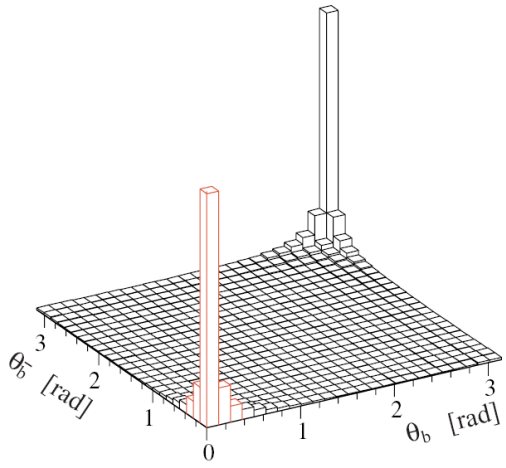


Figure 2.2: Angular distribution of the produced b-hadrons. Taken from [4]

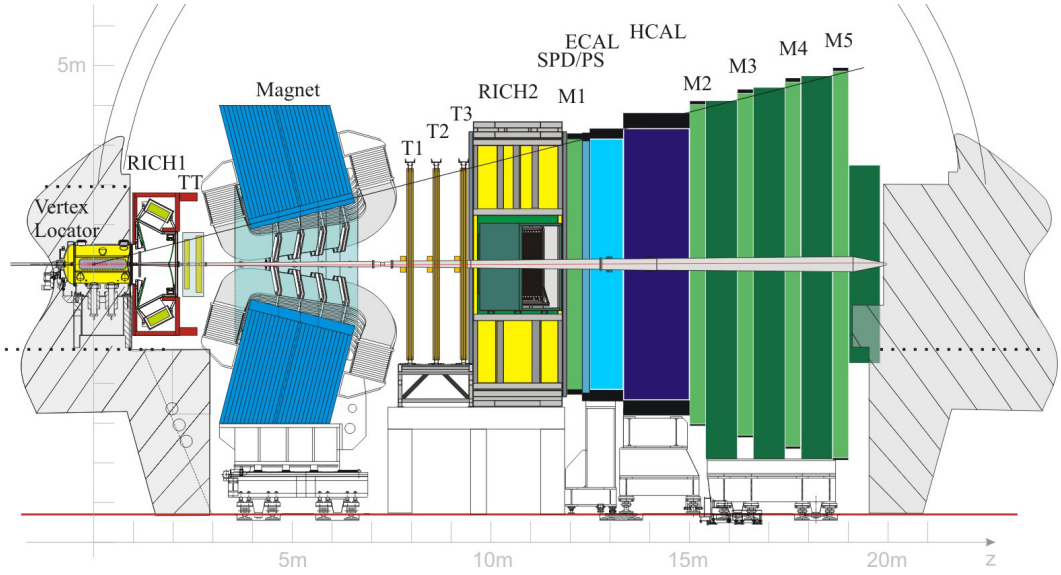


Figure 2.3: Schematic of the LHCb experiment. Taken from [4]

2.3.1 VErtex LOcator

The Vertex Locator (VELO) performs precision measurements of track coordinates close to the interaction region. Its first measurement point is at a radius of 8 mm from the colliding beams. This information allows for accurate reconstruction of both primary and secondary vertexes [8]; the latter being a distinctive feature of b- and c-hadrons. To achieve resolutions for particle lifetime and impact parameter of 40 fs and 20 μm, respectively, the VELO reconstructs vertexes with a resolution of 10 μm [9]. The detector consists of two sets of 21 modules positioned perpendicular to the beam line. It is located about the interaction region within a vacuum maintained vessel. Each module of the detector is formed of two, approximately semi-circular, 300 μm thick, silicon detectors; the R- and ϕ-sensors. A hole of 7 mm in radius is allocated for the passage of the beam when the detector is operated in its "closed" position. The silicon in the R-sensor is implanted within concentric semi-circles centered about a point coincident with the LHC beam position. This allows for a determination of a track r co-ordinate from the z-axis. The orthogonal co-ordinate, the azimuthal angle, is then determined by the ϕ-sensor, whose silicon strips run radially from the detector inner to outer radius. Determination of the third co-ordinate, z, of each hit is provided by knowledge of each modules position along the z-axis. The 21 pairs of R- and ϕ-sensors in each half of the VELO are arranged as shown in Fig. 2.4. The modules are positioned between $z = -18$ cm and $z = 88$ cm. Although the nominal beam interaction point is defined at $z = 0$, the longitudinal uncertainty of the LHC beam is estimated at 5 cm. Consequently, modules are positioned

[1] Particle pseudorapidity, η, is related to the angle between the particle direction of flight and the direction of the incident protons, θ, at which the particle is produced in the following way $\eta = log[tan(\theta)/2]$.

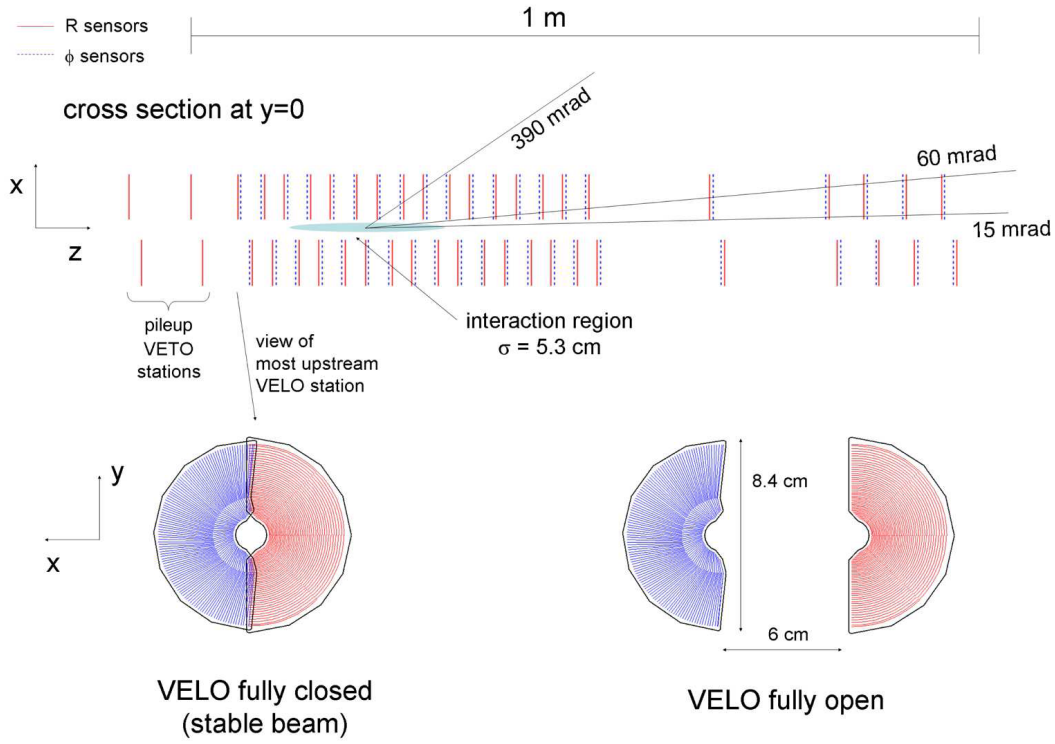


Figure 2.4: Representation of the modules position in the VELO Detector. Interaction region is shown and tracks showing the acceptance are drawn. Taken from [8]

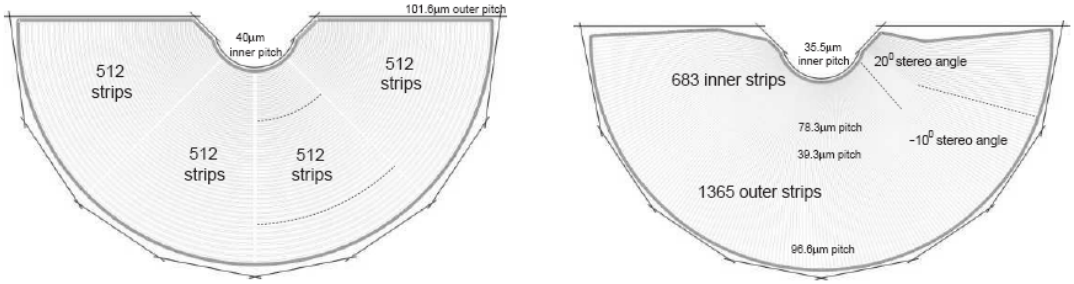


Figure 2.5: Representation of the R (left) and ϕ (right) VELO Modules. Taken from [8]

comprehensively throughout this region. During normal operation, the LHC beam is well confined within the other two dimensions and, hence, poses no danger to the silicon modules just 7 mm away. This aperture increases however during the machine's injection phase when, on account of increased radiation levels, it is necessary to retract the two detector halves by 3 cm. As a result, the VELO detectors are mounted on a controllable

positioning system which inserts or retracts the silicon sensors to either closed or open positions depending on the state of the LHC beam operation.

Vertex Reconstruction

The primary vertex in LHCb is reconstructed with the following iterative procedure [10]:

- A histogram with bin width s of 1 mm is filled with the z coordinates of the points on the VELO seed trajectories. The highest bin is chosen together with its four neighbors on each side. The barycenter is then calculated as the first estimation of the z coordinate of the primary vertex. The tracks in these bins form the original primary vertex.
- Tracks with a large χ^2 contribution to the vertex are eliminated and the remaining tracks are fitted again to a new vertex. This process is repeated until there is no track with a χ^2 contribution larger than 9 (225 for the first iteration to avoid losing proper tracks).
- The above vertex with more than 6 tracks is kept as a primary vertex and the tracks which form the vertex are discarded. The process then iterates to search for the next primary vertex. The process stops if no more primary vertexes can be found. If there is no vertex at all with more than 6 tracks, the original vertex is kept as the only primary vertex.

2.3.2 The Magnet System

Charged particles are bent in the magnetic field of the main dipole magnet [11]. Their momentum is measured from the detection of the particle trajectories. The bending strength of the magnet is determined by the integrated magnetic field, which is

$$\int Bdl = 4.2\ Tm \tag{2.5}$$

The magnetic field is designed such that it is large in the y direction and small in the x and z directions. This bends particles moving in z direction in the (x-z) plane (bending plane). The strength of the main component of the field (By) along the z axis is shown in Figure 2.6, the positions of the tracking detectors are indicated by dashed lines. It is warm magnet design with saddle-shaped coils in a window-frame yoke; since no superconductive technology is involved it is possible to control the systematic effects of the detector, by changing periodically the direction of the magnetic field without any major problems. With superconductive magnets a long time is needed to invert the current flow.

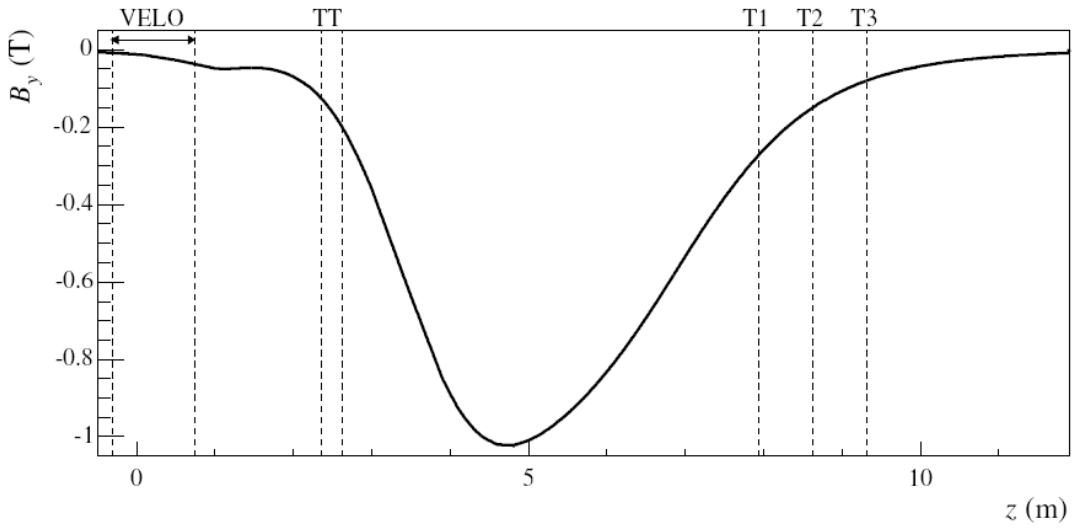


Figure 2.6: Intensity of the component B_y of the magnet field along the z axis. Taken from [11].

2.3.3 Tracking

In addition to the VELO, the LHCb tracking system consists of five tracking stations; two before the magnet and three after. These stations are labeled as TT and T1, T2, T3 in Fig. 2.3. The latter three stations cover the entire geometrical acceptance of the spectrometer. To achieve the excellent tracking performance needed across this area, these three stations are composed of two detector types; Inner and Outer Trackers.

The Trigger Tracker (TT)

The Trigger Tracker[12] is located in front of the magnet. It consists of two stations separated by a distance of 27 cm. Each station of the TT has two layers of silicon strip detectors covering the full acceptance. The strips in the four layers are arranged in (x, u) and (v, x) layers, corresponding to angles with respect to the vertical y axis of $(0°, -5°)$ and $(5°, 0°)$. The stereo angle allows the reconstruction of tracks in three dimensions. The vertical orientation of the strips is chosen to obtain a better resolution in the bending plane of the magnet and therefore a better precision on the measured momentum. The layout for the third TT layer (v) can be seen in Figure 2.7 on the right. Its dimension is 145 cm in the x direction and 132 cm in the y direction. The layout is composed of half modules which consist of a column of seven silicon sensors. In the region above and below the beam pipe, there is one half module on each side. From the middle to the edge, there are seven half modules (in the first two layers) or eight (in the last two layers). According to different occupancies, the seven silicon sensors are divided

into three groups labeled by different colours in Figure 2.8

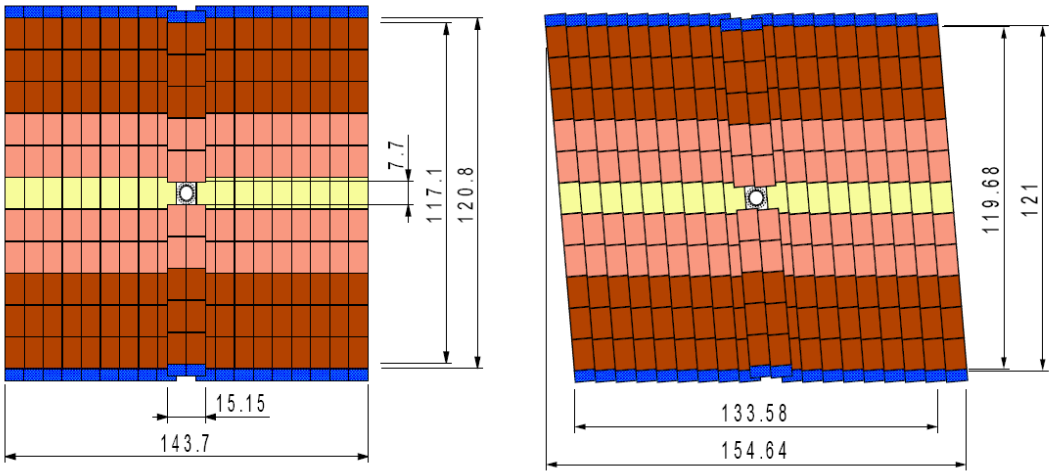


Figure 2.7: The sensor arrangement in the first two TT layers adjacent to the magnet. The colours signify the readout configuration. Taken from[12].

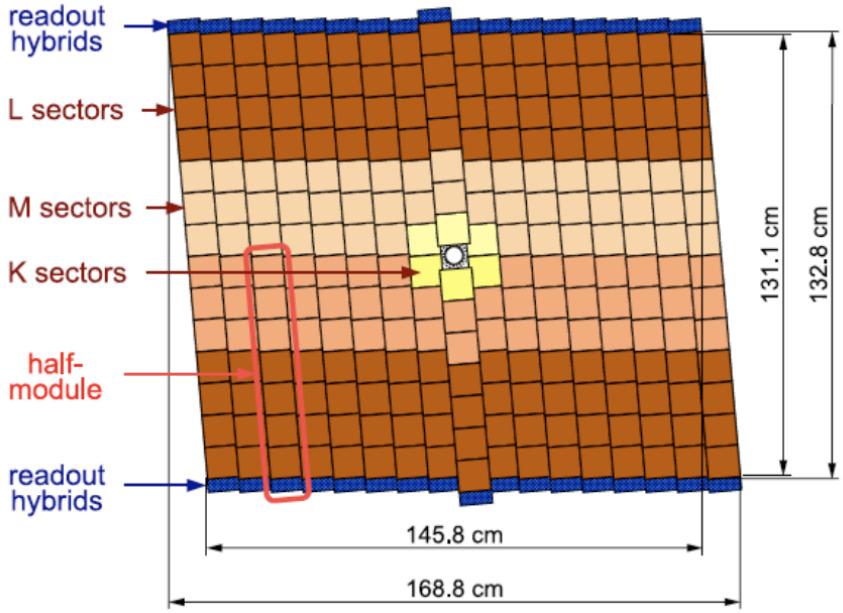


Figure 2.8: The sensor arrangement in the (third) TT layer adjacent to the magnet.Taken from[12].

The T-Stations

The T1-T3 stations are the main tracking system for LHCb. Because of the different occupancies in different regions, each of them is divided into two parts: Inner Tracker (IT) and Outer Tracker (OT).

The **Inner Tracker** (IT) [13] covers the innermost region of the T-Stations which receives the highest flux of charged particles. The IT is placed around the beam pipe in a cross shape, see Figure 2.9. The IT covers approximately 2% of the $6m \times 5m$ acceptance of the T-Stations, but it measures about 20% of the particles. Each station has layers arranged in a (x, u, v, x) configuration, similar to the TT. Two types of silicon $p+$ on n sensors of different thickness are used in the IT. The single sensors, which are $320\mu m$ thick, are placed above and below the beam pipe. The double sensors, which are $410\mu m$ thick, are placed at the sides of the beam pipe. The two types were chosen to ensure sufficiently high signal to noise ratios for each module type while minimizing the material budged of the detector. The strip pitch is $198\mu m$ resulting in a resolution of approximately $50\mu m$ [13]. The maximal occupancy in the IT is below 2%.

The **Outer Tracker** (OT) [14] covers the large region outside the acceptance of

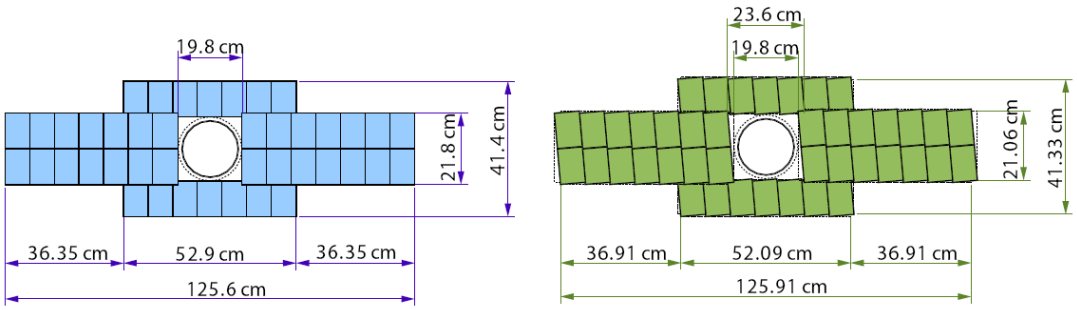


Figure 2.9: Representation of one layer of the IT. Taken from [13].

the Inner Tracker, see Figure 2.10. Charged particles are detected in the OT with gas filled straw tubes serving as drift cells. The gas mixture to operate the OT is $Ar(70\%)/CO_2(30\%)$. The inner diameter of the straws is 5mm and the pitch between two straws is 5.25 mm. At the centre of the straw is a 24 μm thick gold coated tungsten wire which operates as anode. Wire locators are placed every 80 cm to keep the wires at their nominal position. The cathode cell wall consists of two foils: the inner windings are made of a 40 μm thick carbon doped polymer foil (Kapton-XC) and the outer windings are made of a 25 μm Kapton-XC foil with a 12.5 μm aluminium coating. The straws are fixed in a module between two panels which form a stiff gas-tight box.

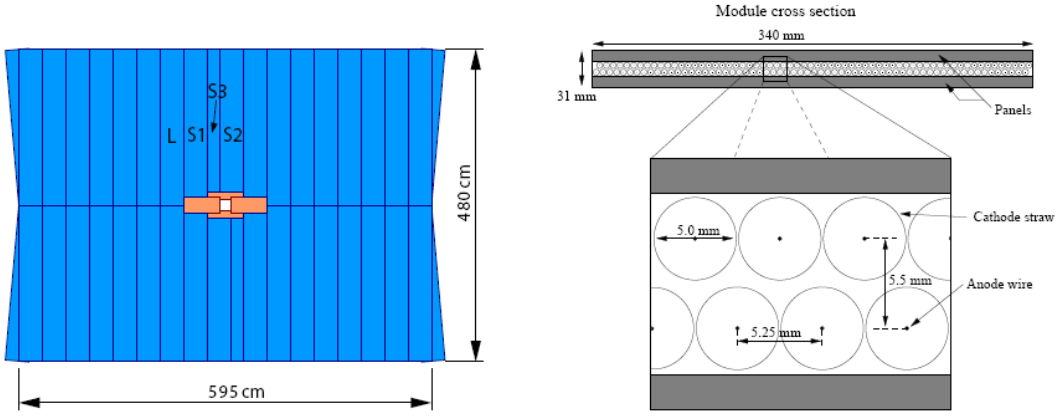


Figure 2.10: Left: Front view of a tracker station. Right: Module cross section of the straw tubes. Taken from [14]

Track Reconstruction

The first step in the track reconstruction is the pattern recognition or track finding, where the correct hits belonging to one track are searched for. The number of reconstructed ghost tracks, i.e., tracks with many wrong hits, should be kept to a minimum. A track is called a ghost if less than 70% of the hits on the track originate from the same generated particle. For example, a real track in the VELO which is matched to a real track in the T-Stations but from adjacent particles is called a ghost track. The tracks are classified in different types, depending on their trajectory in the LHCb spectrometer. They are sketched in Figure 2.11 and can be described as follows:

- **Long tracks** are the best quality physics tracks of LHCb. They traverse the complete tracking system from the VELO to the T-Stations.
- **Upstream tracks** are only reconstructed in the VELO and TT-Stations. They are bent out of the acceptance in the dipole magnet. Although their momentum resolution is reduced, they can be used in some B decay analyses.
- **Downstream tracks** are only reconstructed in the TT and T-Stations. They allow the reconstruction of K_S^0 decays outside of the VELO acceptance.
- **VELO** tracks traverse only the VELO. They allow a precise determination of the primary vertex as they have typically a large polar angle.
- **T-tracks** are only reconstructed in the T-Stations.

The aim of the pattern recognition is to find for each particle the best possible track, i. e., the track which uses measurements in all subdetectors that the particle crossed. To get the maximum number of long tracks, two redundant reconstruction algorithms are used,

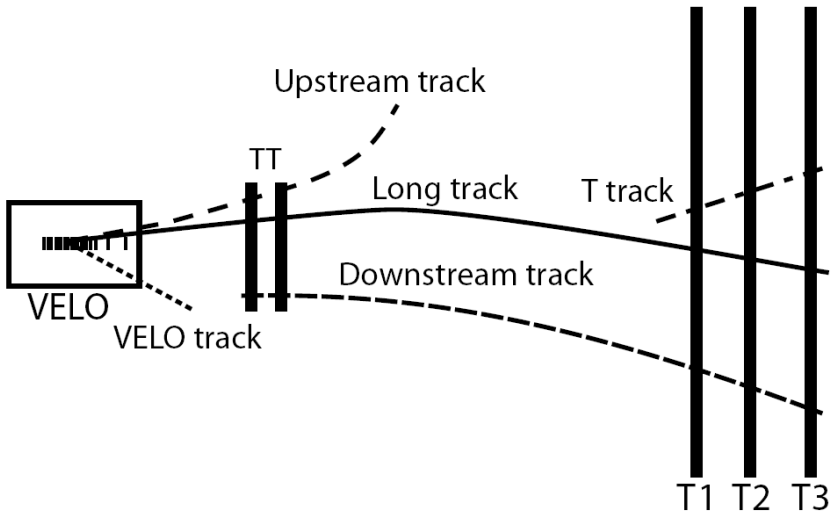


Figure 2.11: Sketch of the five different track types in the LHCb tracking. Taken from [15].

called "forward tracking" and "track matching". As many tracks will be reconstructed by both algorithms, a clone killing algorithm will clean up duplications (tracks are called "clones" if they share at least 70% of their hits). Both algorithms start with standalone VELO tracks, therefore this algorithm will be described first:

In the following, the individual algorithms in the track finding procedure are described:

- **Standalone VELO tracking:** The magnetic field in the VELO is sufficiently low, the tracks can be reconstructed as straight lines. The algorithm [16] starts with trajectories in the $(r - z)$ plane. These tracks are called VELO 2D tracks. The priority is to have a high efficiency and a low ghost rate. The VELO 2D tracks are then extended to 3D tracks by adding hits from the VELO ϕ sensors. Both VELO 2D and 3D track finding algorithms assume that the tracks originate from the same vertex. The resulting track segments serve as seed for the other track finding algorithms.

- **Forward tracking:** This algorithm [15] starts from the VELO tracks and uses them as seeds to find continuations in the T-Stations. The algorithm is based on the idea that a single hit in the T-Stations together with the VELO track segment define the complete trajectory of the track. The trajectory is parameterized by a second order polynomial in y and a third order polynomial in x. Further T-Station hits in a window around the expected position are picked up. The candidates with the most hits are selected and their hits are assigned to the track. Finally, a likelihood is calculated to confirm the correct tracks and discard wrongly reconstructed combinations (ghosts). Hits in TT are picked up if they are close enough to a track through VELO and T station hits.

- **Track matching:** In the track matching, standalone VELO tracks are matched to standalone T-Station tracks. The T-track and the VELO track are matched by extrapolating both track segments to the bending plane of the magnet and evaluating quality criteria as the position in the bending plane or the change of the track slope. TT hits close to the resulting tracks are added afterwards.

2.3.4 RICH Detectors

LHCb uses two ring-imaging Cherenkov (RICH) detectors for the purpose of particle identification. The Cherenkov effect happens when charged particles pass through a dielectric medium at a speed greater than the speed of light in that medium, the particles interact with the medium in such a way as to cause the emission of photons.

Cherenkov radiation is emitted in a cone. The polar angle θ_C is angle at which the photons are emitted is known as the Cherenkov angle, and is related to the speed of light in the medium, v_m, and the speed of the charged particle, v_p.

$$\cos\theta_C = \frac{1}{n\beta} \tag{2.6}$$

where $n = c/v_m$ and $\beta = v_p/c$. An essential part of the LHCb experiment is an efficient system of particle identification, effective over a $1 - 100\ GeV$ momentum range [17]. Two RICH detectors with three Cherenkov radiators are included in LHCb for this purpose. Mirrors are used to focus and reflect photons onto photodetection planes which are outside the LHCb angular acceptance. Pixel Hybrid Photo-Diodes are used to detect the photons. The three radiators used are: Aerogel, C_4F_{10} and CF_4 and, as it is possible to see in figure 2.12 each radiator provides effective $K - \pi$ separation over a different range of momenta. The first RICH detector (see figure 2.3) is positioned immediately downstream of the VELO. Charged particles that are within the LHCb geometric acceptance first traverse an aerogel radiator, before passing through a volume of C_4F_{10} gas. These two radiators will distinguish between charged pions and kaons over a momentum range of $1 - 10\ GeV$ and $10 - 60\ GeV$ respectively. Similar to RICH 1 in design, but further downstream, is RICH 2 which is equipped with a CF_4 radiator to cover momenta up to 100 GeV, over a smaller angular acceptance of 120 $mrad$.

2.3.5 Calorimeters

The Calorimeter system [18] provides measurements of the energies and positions of hadrons, electrons and photons. This information plays an important role in the Level-0 trigger (see Section. 2.3.8) and, ultimately, in particle identification (see 2.3.7). Situated downstream of RICH-2, between the first and second Muon stations, the Calorimeter

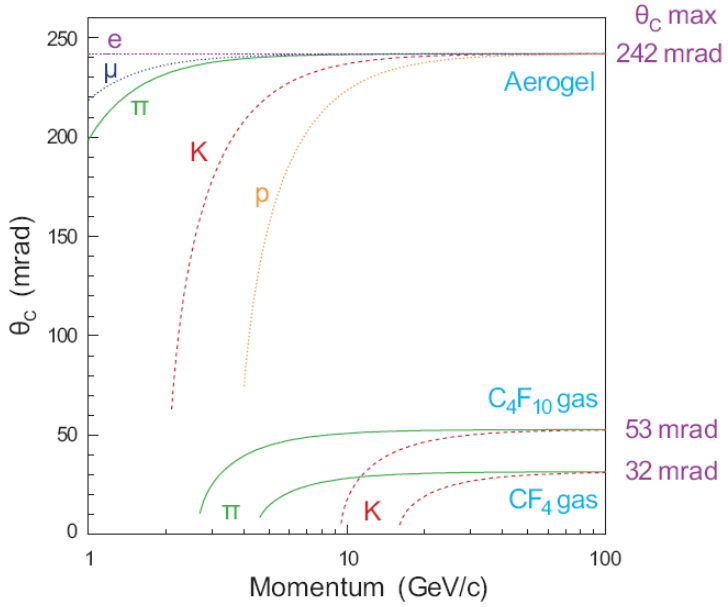


Figure 2.12: Variation of Cherenkov angle with particle momentum, for the three radiators used in the LHCb RICH detectors

system takes the classical structure of an electromagnetic calorimeter (ECAL) followed by a hadron calorimeter (HCAL). One of the most critical sources of information for the Level-0 trigger is accurate electron identification. To achieve this in the extremely demanding environment that is a hadron-collider, longitudinal separation of the electromagnetic (EM) shower is needed in order to reject the overwhelming π^0 background.

To obtain the separation required, two additional detectors are placed in front of the ECAL: the scintillator pad (SPD) and preshower (PS) detectors. A total of four separate components, then, compose the Calorimeter system. Each has its own unique specification:

- **SPD:** A plane of 15 mm thick scintillator tiles cover an area 7.6 m wide and 6.2 m high. Each scintillator tile is a square of side either 39.2, 59.0 or 118.4 mm depending on its position in the detector. The light generated in each tile is transmitted via a single wavelength-shifting (WLS) fibre to the input of a multianode photomultipler tube (MaPMT). When used in coincidence with data from the ECAL, SPD information assists in the rejection of $\pi^0 \rightarrow \gamma\gamma$ and photon backgrounds from the electron candidate selection.

- **PS:** A lead converter, 15 mm thick (2.5 interaction lenght equivalent), is sandwiched between the SPD and a second identical scintillator plane. Since electrons will shower within the lead converter as a result of their shorter interaction length, data from the PS scintillators can be used in conjunction with the ECAL to reject

$\pi^{\pm}$ backgrounds from the electron candidate selection.

- **ECAL:** This is a sampling calorimeter formed of 66 alternating layers of lead absorber and scintillator, of width 2 and 4 mm, respectively. The total thickness of the detector corresponds to 25 X_0 which ensures full containment of showers from high energy photons. Standard photomultiplier tubes are used to read out the scintillator light signals.

- **HCAL:** This is a sampling device and consists of alternating layers of iron and scintillator. Unlike the SPD, PS and ECAL planes that are positioned perpendicular to the z-axis, the HCAL scintillator tiles are orientated parallel to it. Due to space constraints the detector thickness corresponds to only 5.6 λ_I. As in the ECAL, standard photomultiplier tubes are used to gather the scintillator light.

The energy resolution of the ECAL modules has been determined in test beam with electrons and muons [18]. The parameterisation

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus b \oplus \frac{c}{E} \tag{2.7}$$

is used, where a, b and c stand for the stochastic, constant and noise terms respectively (E in GeV). Depending on the type of module and beam conditions, the stochastic and constant terms were measured to be $8.5\% < a < 9.5\%$ and $b \sim 0.8\%$. Similarly the energy resolution of the HCAL has been determined. Fits to test beam data return a resolution of

$$\frac{\sigma(E)}{E} = \frac{69\% \pm 5\%}{\sqrt{E}} \oplus (9 \pm 2)\% \tag{2.8}$$

2.3.6 Muon System

The muon system [19] consists of five stations. The first station (M1) is located in front of the calorimeter system to improve the momentum measurement, whereas the other stations (M2 - M5) are located directly behind the hadronic calorimeter, as shown in Figure 2.3. The inner and outer angular acceptances of the muon system are 20 *mrad* (16 *mrad*) and 306 *mrad* (258 *mrad*) in the bending (non-bending) plane respectively. This results in an acceptance of about 20% for muons from inclusive b semileptonic decays. The stations M2-M5 are separated by 80 *cm* thick iron filters, corresponding to a total of 20 interaction lengths. The minimum momentum of a muon to cross the five stations is approximately 6 GeV. A pad readout structure provides binary space point measurements of the muon tracks, allowing fast track finding in the hardware trigger. Each station is divided in four regions with different pad granularity. The granularity of each region is adjusted to the expected particle flux. The regions and their pads increase in size from M1 to M5 so that they are mutually projective towards

the interaction point. The dimensions of the pads decrease towards the inner regions such that their hit occupancy stays roughly constant. In addition, their size in x is smaller than that in y, giving a more accurate momentum measurement. The muon stations are instrumented with multi-wire proportional chambers (MWPC) [20], which fulfill both the requirement from the trigger to collect the signal within 20 ns and the requirement for radiation hardness. The required granularity is obtained by grouping anode wires and the cathode pads of the MWPCs. In the innermost region of the first muon station, where the particle density will be higher, triple-GEM detectors [21] are used.

2.3.7 Particle Identification

The LHCb particle identification (PID) is achieved with the following subdetectors: RICH1, RICH2, calorimeters and muon detectors. In general, the two RICH detectors are used to distinguish hadrons like π, K, electrons are identified with the calorimeters and muons are identified by the muon detectors. The RICH detectors can also provide information for lepton identification; muons only deposit minimum ionisation energy in the calorimeters which can also be used as identification information. As an example, we describe the procedure for the muon identification. The PID likelihood is constructed by combining all the information from the PID detectors; the likelihood function for the muon hypothesis is:

$$L(\mu) = L_{RICH}(\mu)L_{CALO}(\mu)L_{MUON}(\mu); \tag{2.9}$$

where the subscripts of the different likelihood functions indicate different detectors. The main contribution of the muon identification comes from the muon detectors. In our study, the MuonID algorithm consists of three main steps [22]

- Define the field of interest (FOI): The sizes of the FOI are defined separately for each of the 4 regions of the muon system in all the four stations, M2-M5. The centers of the FOI are the points in the muon stations obtained by extrapolating the track direction in M1 to the rest of the stations. The size of the elliptic FOI around the center point is given by the momentum dependent function

  $$p_0 + p_1 \cdot \exp(-p_2 \cdot p) \tag{2.10}$$

  The parameters p_0, p_1 p_2 are currently determined from Monte Carlo by maximizing the efficiency over the whole range while reducing the misidentification rate to the percent level.

- Set a boolean value: Each track is then marked with boolean value called *IsMuon* according to their hits in the FOI for different muon stations. It is required to

have at least one hit in the FOI in a number of stations which depends on the momentum of the tracks. The requirement is listed in Table 2.1

- Building the DLL (Delta Log-Likelihood function) for the muon identification: The misidentification rate is further reduced with the use of the average squared distance of the hits in the FOI to the center point.

Due to the difficult understanding of the DLL in real data, for our analysis we used only the IsMuon flag which is better understood and has a robust approach.

Momentum Range (GeV)	Muon Stations
$3 < p < 6$	$M2 + M3$
$6 < p < 10$	$M2 + M3 + (M4 or M5)$
$p > 10$	$M2 + M3 + M4 + M5$

Table 2.1: Definition of $IsMuon = 1$. For each given momentum range, at least one hit is required in each of the listed muon stations

2.3.8 The Trigger System

LHCb expects $b\bar{b}$ cross-section of 500 μb at 14 TeV, but the inelastic cross-section is expected to be 100 mb. Most of the inelastic cross-section is considered uninteresting in terms of new physics, and it is desirable in terms of data storage and processing to retain only the data with signatures of specific physic channels. The trigger is used to select events that are considered interesting and discard those which are not.

The LHCb trigger is composed of two levels: Level-0 (L0) and High Level Trigger (HLT). L0 is implemented in custom electronics, and reduces the non-empty rate from 30 to 1 MHz at a fixed latency of $4\mu s$. The HLT is a software trigger running in computing nodes forming the Event Filter Farm (EFF). L0 uses information from the calorimeters and muon chambers to provide high E_t and P_t candidates, and the VELO pile-up system provides a fast estimation of the number of proton-proton interactions that occurred in the bunch crossing. L0 positive decisions are sent back to the front-end electronics of all the sub-detectors, which pick-up the pieces of the relevant events from buffers and send them through a read-out network to the EFF.

L0 Trigger

The L0 trigger decision is taken by the L0 Decision Unit, based on the following pieces of information provided by the trigger boards:

- Calorimeter clusters classified as electrons, photons, neutral pions or charged hadrons according to the energy deposition in the different calorimeter layers.
- Single muon and dimuon candidates.
- The result of a fast proton-proton interaction vertex search based on two dedicated layers of the vertex detector (the Pileup System).
- The multiplicities in the Pileup System and in a scintillator layer in front of the calorimeter (SPD).

If the running conditions require it, i.e. the instantaneous luminosity is large and more than one interaction per bunch crossing is present, vetoes on the number of interaction vertices and on multiplicities can be applied to any L0 line. This would allow removing events which are harder to process at the HLT. The Pileup system of the VELO will be used to reconstruct the longitudinal position of the interaction vertices and reject events with two or more such vertices. The multiplicity measured in the Pileup detector, the Scintillator Pad Detector (SPD), and the total E_T measured in the calorimeter are used toset an upper limit to the activity in the detector which are called Global Event Cuts (GEC) which reject events which high activity.

For building muon candidates, straight segments are searched for in the four muon stations downstream of the calorimeter, where the occupancy is relatively low. The search is performed within projective towers, under the assumption that the muon tracks roughly point to the interaction point. The segments found are confirmed by looking for a compatible hit in the muon chamber upstream of the calorimeter. The momentum is then estimated by using a look-up table built under the assumption that the muon tracks originate at the interaction point. A positive decision is taken when a calorimeter or muon candidate above the pt or Et threshold is found. The threshold values for different types of candidates are shown in table 2.2. These thresholds depend on the running conditions and on the relative bandwidth division between the different L0 triggers.

Electron	Photon	Hadron	Muon	Dimuon
2.8 GeV	2.8 GeV	3.84 GeV	1.36	$P_T^1 + P_T^2 > 1.48\ GeV$, $min(P_T^1, P_T^2) > 80\ MeV$

Table 2.2: L0 Trigger thresholds

High Level Trigger

The High Level Trigger (HLT) has access to the full detector informations in one event, thus in principle the HLT could perform the offline selections. Nevertheless, given the 1 MHz output rate of the Level-0 and the computing power limitations, the HLT aims to

reject most of uninteresting events using only part of the full detector data. The HLT has been designed to be as flexible as possible. Its purely software character makes it dynamic and adjustable according to the real needs of the experiment. For that reason the HLT will evolve in time with the knowledge of the experiment and the physics priorities. Figure 2.13 shows a typical trigger flow diagram from the L0 up to the output of the HLT. All calorimeter L0 clusters and muon tracks above threshold are passed to the HLT

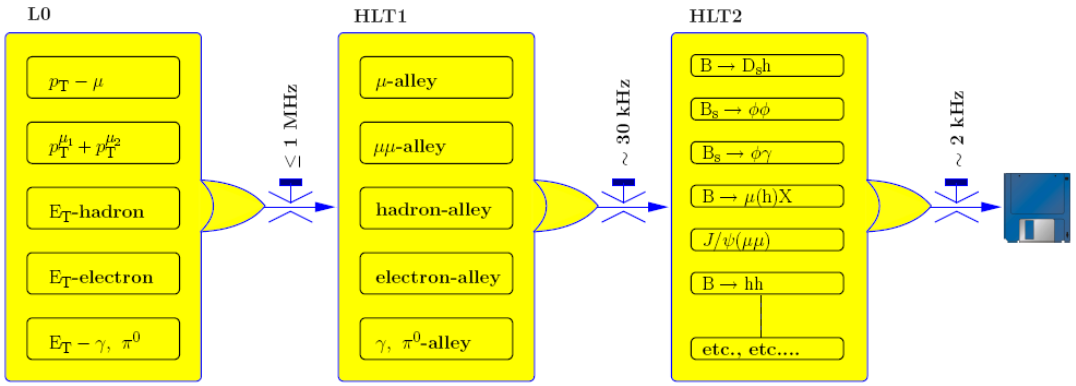


Figure 2.13: Diagram of the three trigger levels. Taken from [23].

and will be referred to as L0 objects henceforward. The HLT is divided in two stages, HLT1 and HLT2. The HLT1 aims to confirm the L0 objects by searching for tracks in the VELO and the T-stations corresponding to the L0 electrons, hadrons or muons. In the case of L0 γ and π^0, the HLT1 aims to confirm the absence of charged particle tracks corresponding to these L0 objects. This stage is called L0 confirmation. The HLT1 output, of about 30 kHz, allows full pattern recognition by the HLT2, which aims to execute a series of inclusive and exclusive trigger algorithms in order to reconstruct partially or totally several kinds of interesting decays.
We now describe some trigger lines which has been used in the analysis that we will describe in chapter 7.

- **Minimum Bias Trigger:** A very loose minimum bias trigger was set up in order to trigger inelastic *pp* collisions. A requirement was made of at least one charged track in the VELO detector or in the downstream tracking system.
- **Hlt1SingleMuon:** The following algorithm flow summarizes this line [24]:
  1. Selection of L0 muons that fired L0 ($P_T > L0PT_{Cut}$ Value)
  2. T Confirmation and VELO Matching: Confirmation with T stations, Confirmation of momentum, Muon ID, T-VELO2d Matching, T-VELO3d Matching.
  3. Single Muon Decision: Apply a P_T cut at 0.8 GeV

- **Hlt2 Drell-Yan lines:** These lines create dimuon (or dielectron) candidates according to some kinematic requirements. The muons transverse momentum has to be larger than 1 GeV and the dilepton invariant mass has to lay in a given range. Four different lines, corresponding to different invariant mass ranges, are identified: $2.5-5\ GeV,\ 5-10\ GeV, 10-20,\ GeV, 20-40\ GeV$. A dedicated line for Z selection with only a lower limit on invariant mass at 40 GV is also present.

2.3.9 LHCb software

The LHCb software is used to reconstruct the particle physics events and analyse these events for physics signatures. This software is important not only for designing analyses before the detector is switched on, but will also be used to help in detector calibration and the analysis of certain channels when the LHC is running. The simulation of events at the LHC consists of two phases: event generation, detector response simulation and digitization which are outlined in turn. Then, as well as with real data, the chain is completed by the event reconstruction and event selection and analysis.

Event generation

Physics events at LHCb are simulated using a software package called Gauss: the LHCb Monte Carlo events are generated using PYTHIA and EvtGen [25]. PYTHIA simulates the production of $b\bar{b}$ pairs in pp collisions through the processes of gluon fusion, gluon splitting and flavour excitation. The EvtGen package then decays the resulting B mesons from a table of predetermined decay channels and rates. The LHCb Monte Carlo samples used in the analysis described in chapter 7 were centrally created using PYTHIA.

Detector Simulation

Once an event has been generated, it is necessary to simulate the passing of particles through the detector and their interactions with it. The package GEANT 4 [26] is used for this purpose, always controlled through the LHCb simulation program GAUSS. The LHCb detector is simulated in detail, including:

- Passive materials such as supports, frames, shielding elements and the LHC beam pipe are simulated, and secondary interactions of particles are tracked down to 10 MeV for hadrons and 1 MeV for electrons and photons.
- The LHCb magnet is simulated using a field map measured on the magnet which will be used in the running detector.

- The spill over of particles from one event into another, is simulated. The probability of neighbouring bunch crossings producing particles is calculated from the instantaneous luminosity.
- Individual subdetectors are fully simulated. For example, effects such as charge sharing, noise, and cross talk are included in the simulation of silicon sensors.

Digitization

Once the detector response to the event has been simulated, it is necessary to simulate the response of the detector and the readout electronics to the hits in the detector. This includes simulating the propagation of electrical signals through the detector, the response of the readout hardware to these signals, and the analogue to digital conversions performed in the readout modules and L0 trigger boards. The digitization is performed using the package BOOLE. BOOLE produces digitized data in the same format as that which will be produced by the LHCb hardware once the machine is running.

Reconstruction

The LHCb reconstruction software BRUNEL creates tracks and RICH Cherenkov rings out of the digitized output provided by BOOLE. BRUNEL contains the full pattern recognition and tracking software which will be used to reconstruct data taken during LHCb running.

Analysis

The tracks the calorimeters information and RICH information output by BRUNEL are the inputs to the LHCb analysis package, DAVINCI, which uses them to reconstruct the particle decays. The final state of interest can then be selected by making the appropriate particle combinations. Finally, Davinci provides the functionality for developing and applying offline selection algorithms that can be used to select physics processes of interest and reject background processes.

3. THE STANDARD MODEL

The Standard Model (SM) of the fundamental interactions is a quantum field theory which has been developed by physicists as an attempt to describe the matter and its interactions as they are known from several experimental results.

The SM provides a systematic description of three fundamental interactions: the strong interaction, the weak interaction and electromagnetic interaction. The description of the interactions is built according to local gauge theories which foresee the presence of massive and massless boson gauge fields as mediators of the forces. Local gauge invariance is a central issue for the construction of this theory: in a gauge theory [27], the Lagrangian must be invariant under local transformations. In a global symmetry, such as a spatial translation, all the laws of physics remain the same, no matter what point in space or time we are dealing with. However, in a local symmetry, the transformation is a function of space and time, yet the laws of physics must remain the same at all points in space-time. These local symmetries can be defined by a group. A group is a set of elements, whereby a mathematical operation between two elements of the group (say multiplication) produces an element of the group. Each group has a generator or number of generators which can produce all elements of the group. In terms of a field theory, a group defines the set of local transformations, and for each generator there exists a gauge field.

The gauge group of the SM is the direct product of three symmetry groups:

$$G_{SM} = SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \tag{3.1}$$

The $SU(3)_C$ symmetry is the colour symmetry and it predicts eight massless vector particles, called *gluons*, which are responsible for the strong interaction. The second symmetry, $SU(2)_L$ concerns the weak interaction and acts on fermion doublets of defined helicity particles (left handed) and it defines three vector bosons W^1 W^2 and W^0. The third group symmetry $U(1)_Y$ is related to the weak hypercharge[1] of the interacting particle and it is mediated by a vector boson B. However, in nature we don't observe an interaction based on the hypercharge but there is mix between the bosons B and W^0

[1] The weak hypercharge Y is defined from $Q = T_3 + \frac{Y}{2}$ where Q is the electric charge and T_3 is the third component of weak isospin which is the conserved quantity of the $SU(1)_Y$ group.

which give rise at the mediator of the electromagnetic interaction, the photon γ and to the mediator of the weak interaction, the Z^0 boson. At the same time two classes of fermionic elementary particles which constitutes the matter are defined: quarks and leptons. The characteristics of the particles predicted by the standard model are summarized in figure 3.1

3.1 Elementary Particles

Leptons: These are divided into three left handed doublets of weak isospin called generations. The doublets consist of a negative electrically charged lepton (electron, muon or tau) and by a neutral lepton called neutrino. In addition three right handed singlets made only of the charged lepton are present; no right handed neutrinos are present in the theory or, more precisely, no interaction that involves right handed neutrinos exists. Charged lepton can experience both electromagnetic and weak interaction while neutrinos only feel the weak interaction. Leptons do not carry colour charge and so they are not affected by the strong interaction.

Quarks: These are also divided into three left handed weak isospin doublets and six right handed singlets. They have a fractional electric charge ($2/3$ and $-1/3$ within the doublet) and they are subject to both electro-weak interaction and to the strong interaction. Apart from the flavour quantum number, which allows them to be separated in six families (u, d, c, s, t and b) quarks carry a colour charge. Each quark can exist in three colours which are mass degenerate, r (red), g (green) and b (blue). Due to the nature of strong interaction which allows the physical existence of only un-coloured states, quarks can only be observed in bound states and free quarks are impossible to observe. *Mesons* are expected to be quark-antiquark bound states, while *baryons* are interpreted as bound states of three quarks.

We now describe the Electromagnetic Interaction and its unification with the weak interaction; the strong interaction and the parton model will be detailed in chapter 4.

3.2 Electromagnetic Interaction

The symmetry group which describes the electromagnetic interaction is $U(1)$ for the electric charge[2], and it is described by an abelian gauge theory called Quantum Electrodynamics or QED. It has only one gauge boson, the photon γ and the coupling constant

[2]This is not the same group as $U(1)_Y$. Here the conserved quantity is the electric charge and not the weak hypercharge.

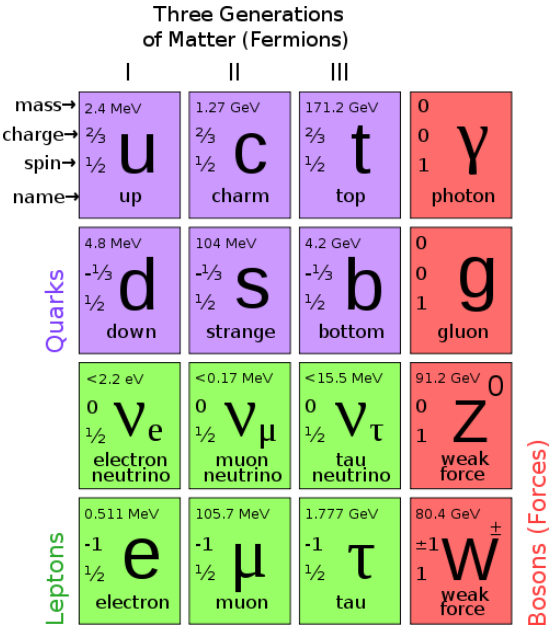


Figure 3.1: Standard model particles.

is the fine structure constant α_{em}, which is related to the electric charge e, by:

$$\alpha_{em} = \frac{e^2}{4\pi\epsilon_0\hbar c} \simeq \frac{1}{137} \tag{3.2}$$

where $\hbar = \frac{h}{2\pi}$ is the reduced Planck constant, c is the speed of light in vacuum; and ϵ_0 is the permittivity of free space.

The lagrangian of the theory for a spin-1/2 field interacting with the electromagnetic field is:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{3.3}$$

where γ^μ are the Dirac matrices, ψ is the spinor field for a spin-1/2 particle and $\bar{\psi}$ its adjoint. The covariant derivative is defined as

$$D_\mu = \partial_\mu + ieA_\mu \tag{3.4}$$

where A_μ is the covariant four-potential of the electromagnetic field of the particle. The electromagnetic field tensor is

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{3.5}$$

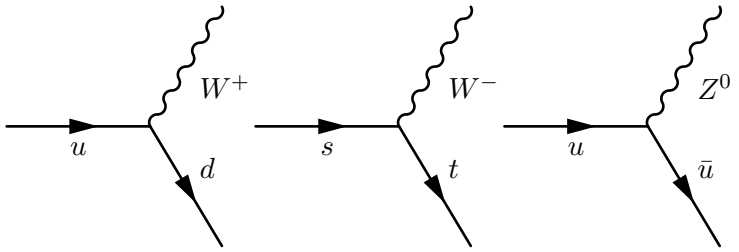


Figure 3.2: Charged and Neutral Weak currents in the quark sector.

The covariant four-potential A_μ represent the massless photon field.

3.3 Weak Interaction and Electroweak unification

To the doublet description of the two spin states of the fermion it is possible to associate a "weak" isospin: $T_3 = 1/2$ for up-type fermions and $T_3 = -1/2$ for down-type fermions. For this reason, the intensity of the interaction, i.e. the coupling constant, is different for the particles in the doublet. It has also been observed from parity violation experiments that charged weak interactions are only possible between left-handed components (right-handed components are isopsin singlet $T = 0$) and that the mediator of the neutral weak interaction couples to left and right-handed particles with different strengths. Since electromagnetic theory is a pure vector interaction an extension to a vector-axial coupling is required [28]. The group describing the spin states of a fermion is SU(2) with the generators W^1, W^2 and W^0. This gauge theory needs bosons to transform between all types of "weak charge" in the theory, i.e. allowing variations $\Delta T_3 = 1$ and also $\Delta T_3 = 0$, as shown in figure 3.2 so, charged current of the kind:

$$j_\mu^- = \bar{\nu}_L \gamma_\mu e_L \qquad j_\mu^+ = \bar{e}_L \gamma_\mu \nu_L \tag{3.6}$$

and a neutral current:

$$j_\mu^3 = \frac{1}{2}\bar{\nu}_L \gamma_\mu \nu_L - \frac{1}{2}\bar{e}_L \gamma_\mu e_L \tag{3.7}$$

are required, where the left component of the spinors is defined from the projection:

$$u_L = \frac{1 - \gamma^5}{2} u. \tag{3.8}$$

The charged currents can be expressed in terms of the vector fields

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp i W_\mu^2) \tag{3.9}$$

while the neutral current cannot be identified with the electromagnetic current and so the W^0 is not the electromagnetic photon. If we admit the existence of a weak neutral current as in equation3.7, it is necessary to extend the gauge group to include a second neutral generator from an abelian factor $U(1)$ giving rise to the group [29]:

$$SU(2) \otimes U(1) \tag{3.10}$$

The pure $SU(2)_L$ Bosons (Weak Isospin $T_3 = 1$) are W^1, W^2 and W^0 while the pure $U(1)_Y$ boson (Weak Isospin $T_3 = 0$) is B^0. The interaction lagrangian for the neutral sector can be written as:

$$\mathcal{L}_n = g\bar{\psi}\gamma_\mu T_3 \psi W_0^\mu + g'\bar{\psi}\gamma_\mu \frac{Y}{2}\psi B^\mu \tag{3.11}$$

where g and g' are the coupling constants and ψ the spinor field. Y is the weak hypercharge defined as $Y = 2Q - T_3$. The actual physical fields are obtained applying a rotation of an angle θ_W in the space of the neutral gauge fields W_μ^0 and B_μ

$$\begin{pmatrix} A \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B^0 \\ W^0 \end{pmatrix} \tag{3.12}$$

and in terms of the new vector fields, A_μ and Z_μ, the lagrangian becomes

$$\mathcal{L}_n = \bar{\psi}\gamma_\mu \left(g \sin\theta_W T_3 + \frac{Y}{2} g' \cos\theta_W \right) \psi A_0^\mu + \bar{\psi}\gamma_\mu \left(g \cos\theta_W T_3 + \frac{Y}{2} g' \sin\theta_W \frac{Y}{2} \right) \psi Z^\mu. \tag{3.13}$$

There are two coupling constants: g for the symmetry group $U(2)$ and g' for the $U(1)_Y$ which are related by

$$g \sin\theta_W = g' \cos\theta_W = e. \tag{3.14}$$

The bosons $W^\pm$ allow the transition between the two component of the weak doublet with a coupling constant $g/\sqrt{2}$. The Z boson couples in a different way to the two fermions of the doublet. The coupling constant, which depends on the third component of the weak isospin T_3, is given by:

$$g_Z = \frac{g}{\cos\theta_W}(T_3 - Q\sin^2\theta_W) \tag{3.15}$$

where Q is the electric charge in unit of e. The neutral current for the coupling of the Z bosons with fermion fields can be re-arranged in order to show the axial-vector nature of this current:

$$\bar{\psi}\gamma^\mu(V - A\gamma^5)\psi Z^\mu \tag{3.16}$$

where the vector coupling is

$$V = T_3 - 2Q\sin^2\theta_W$$

and

$$A = T_3$$

which represents the axial coupling.

3.4 The Origin of the Masses and Higgs Mechanism

One of the unsolved problem of the Standard Model is the mechanism through which the particles get their masses. From the theoretical point of view, it is not possible to preserve the chiral symmetry of the electroweak interaction adding "by hand" classical mass terms for fermions of the kind:

$$m(\bar{\psi}\psi) = m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L + \bar{\psi}_R\psi_R + \bar{\psi}_L\psi_L) \tag{3.17}$$

because the right-left mixed terms are clearly not invariant under $SU(2)$ gauge transformations [30]. To explain the existence of the masses it is necessary to introduce a new mechanism which preserves the gauge invariance of the theory: *The Higgs mechanism* [31], proposed by Peter Higgs in 1964 and by F. Englert, R. Brout [32], foresees the existence of a doublet of scalar complex fields

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \frac{\phi_1 + i\phi_2}{\sqrt{2}} \\ \frac{\phi_3 - i\phi_4}{\sqrt{2}} \end{pmatrix}$$

with ϕ_i real fields. The first component has electric charge equal to 1 while the second is neutral. The weak boson masses have their origin in the Spontaneous Symmetry Breaking Mechanism and the fermion masses derive from the Yukawa coupling with the Higgs doublet[3]. The lagrangian term, which includes the kinematic characteristics of the Higgs Scalar field and its coupling with the W_μ and B_μ, is

$$\mathcal{L} = (D_\mu\phi)^\dagger(D_\mu\phi) - V(\phi^\dagger\phi) \tag{3.18}$$

where the covariant derivative includes the weak fields:

$$\mathcal{L} = (D_\mu\phi) = \left(\partial_\mu - \frac{ig\sigma_j W_\mu^j}{2} - ig'YB_\mu\right)\phi \tag{3.19}$$

[3]Yukawa's interaction, is an interaction between a scalar field ϕ and a Dirac field Ψ of the type $V \approx g\bar{\Psi}\phi\Psi$

where σ_j are the Pauli Matrices. The Higgs potential is not known but in the simplest form is built as:

$$V(\phi^\dagger\phi) = \mu^2\phi^\dagger\phi + \frac{\lambda(\phi^\dagger\phi)^2}{2} \tag{3.20}$$

with a self-interacting term $\lambda\phi^4$. This potential is shown in figure 3.3.

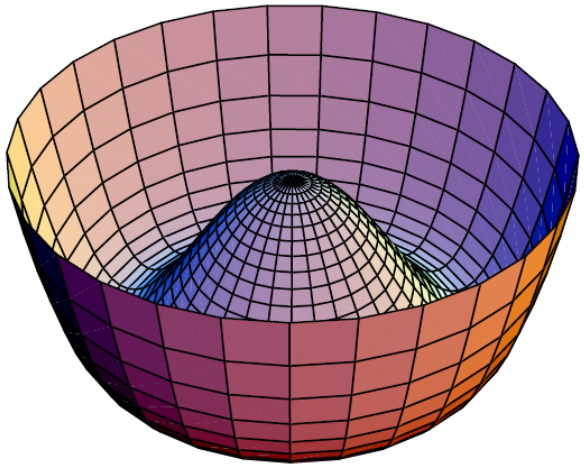


Figure 3.3: Simplest form of the Higgs potential

The parameters μ and λ are arbitrary. If the parameter μ^2 assumes a negative value, the potential has not only one minimum for $\phi = 0$ but a "ring" of minima in the complex plane $|\phi| = \pm\sqrt{-\mu^2/\lambda}$.

The spontaneous symmetry breaking consists in the fact that "Nature" choose one minimum among all those possible (and equiprobable) states keeping a fixed phase in the complex plane. Defined the expectation value of the vacuum of the Higgs field as:

$$v = \sqrt{-\mu^2/\lambda} \tag{3.21}$$

to obtain the masses it is necessary to make some assumption on the Higgs doublet, i.e. $\langle\phi_1\rangle = \langle\phi_2\rangle = \langle\phi_4\rangle = 0$ and $\langle\phi_1\rangle = v$. Expanding the Higgs field around v with the help of four other scalar field θ_1 θ_2 θ_3 and H[33] [4]

$$\phi(x) = e^{i\tau\theta(x)/v}\begin{pmatrix} 0 \\ \frac{v+H(x)}{\sqrt{2}} \end{pmatrix}.$$

The lagrangian is locally invariant in $SU(2)_L$ so, choosing an appropriate gauge we

[4] This is equivalent to rotate the field ϕ in order to obtain $\phi^+ = 0$ and ϕ^0 real. This is possible because of the gauge symmetry.

remove the scalar $\theta_i(x)$ which represent non the physical degrees of freedom. The Higgs doublet becomes

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \tag{3.22}$$

Using this definition of the scalar field in the lagrangian 3.18 and the mixing 3.12 which gives the physical states of the gauge bosons, the interaction terms between the Higgs field and the gauge fields arise. The couplings are hence proportional to the gauge boson masses:

$$\begin{aligned} m_W &= \frac{gv}{2} \\ m_Z &= \frac{m_W}{\cos\theta_W} = \frac{1}{2} v \sqrt{g^2 + g'^2} \\ m_\gamma &= 0 \end{aligned} \tag{3.23}$$

The mechanism which gives mass to the fermions is similar: the lagrangian term that describe the coupling between the Higgs field and the fermion is:

$$\mathcal{L} = G_f \left[(\nu_e, \bar{e})_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} e_R + \bar{e}_R (\phi^-, \bar{\phi^0}) \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \right] \tag{3.24}$$

where G_f depends on the fermion considered. Using the definition of the Higgs doublets in equation 3.22 in the lagrangian defined in equation 3.24, with the Higgs vacuum expectation value definition in equation 3.21 it is possible to obtain the terms

$$-\frac{G_e}{\sqrt{2}} v (\bar{e}_L e_R + \bar{e}_R e_L) - \frac{G_e}{\sqrt{2}} (\bar{e}_L e_R + \bar{e}_R e_L) H = \frac{G}{\sqrt{2}} \bar{e}e - \frac{G}{\sqrt{2}} \bar{e}eH \tag{3.25}$$

It is enough to chose G_e such that $m_e = \frac{G_e v}{\sqrt{2}}$. The term $\bar{e}eH$ represent the coupling of the Higgs boson with the electron field where the coupling constant is:

$$\frac{G_e}{\sqrt{2}} = \frac{m_e}{v} \tag{3.26}$$

The Higgs boson mass terms can be derived from the lagrangian in equation 3.18, using the potential 3.20 and the doublet definition 3.22: Ignoring the gauge fields, the Lagrangian for H has the form

$$\frac{1}{2} (\partial_\mu H)^2 - \lambda v^2 H^2 - \lambda v H^3 - \frac{\lambda}{4}. \tag{3.27}$$

The mass terms correspond to the Higgs self-interacting terms H^2 and the mass of the

Higgs boson is:

$$m_H = \sqrt{2\lambda}v \tag{3.28}$$

Since the parameter λ is arbitrary, the mass of the particle associated to the Higgs field, the Higgs boson, cannot be predicted by the theory.

3.5 Quark Mixing and the CKM matrix

In the Standard Model the quarks have masses through the Higgs Mechanism. However the eigenstates of the weak interaction (u, c, t) and (d, s, b) do not coincide with the mass eigenstate of these quarks (U, C, T) and (D, S, B). Infact charged weak interactions where a quark from one generation changes into a quark from another generation exist [34]. The observed process can be explained using a lagrangian for charged current of the kind:

$$\mathcal{L} = \bar{u}_L \gamma_\mu W^{\pm}_{\gamma_\mu} V_{ud} d_L \tag{3.29}$$

where V_{ud} is the element of a unitary 3×3 matrix which relates the quark weak eigenstates to the quark mass eigenstates.

$$\begin{bmatrix} D \\ S \\ B \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix}.$$

This matrix known as the Cabibbo-Kobayashi-Maskawa (CKM) matrix. In a generic $n \times n$ unitary matrix there are n^2 independent real parameters. Some of them (n_a) can be thought of as rotation angles in the n-dimensional space of generators. The number of these angles is:

$$n_a = \frac{1}{2} n(n-1).$$

The other parameters are just complex phases; their number is

$$n_a = \frac{1}{2} n(n+1).$$

$2n - 1$ oh these phases can be removed just re-defining the left-handed quark fields. In the SM it results that the CKM matrix has 4 independent parameters: 3 angles and one phase. The observed phenomenon of CP violation - a variance of physics under the interchange of particles and anti-particles and a reversal of spatial coordinates - can be accommodated within the SM if the CKM matrix is complex. The complex phase of the CKM matrix results in the violation of the CP symmetry within the SM.
A similar mixing mechanism is present in the leptonic sector. An equivalent matrix is defined for flavor and mass eigenstates of the neutrinos.

4. QCD AND THE PARTON MODEL

Quantum Chromodynamics (QCD) is the underlying theory for the strong interaction thus, all scattering processes at high energy at hadron colliders. The success of the theory has been confirmed by the comparison between the experimental data collected in the last forty years and its predictions.

The formulation of QCD was built by Gell-Mann and Frizsc in 1973 as a renormalizable non-abelian theory based on the symmetry group $SU(3)$ [35, 30]. It includes quarks and gluons as elementary fields. The lagrangian of the theory is

$$\mathcal{L} = -\frac{1}{4} F\mu\nu_a F^a_{\mu\nu} + \bar{\psi}_i(i\gamma^\mu\partial_\mu - m)\psi_i - gG^a_\mu\bar{\psi}_i\gamma^\mu t^a_{ij}\psi_j \tag{4.1}$$

where the gluon field tensor is defined as

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g_s \sum_{b,c} f_{abc} A^b_\mu A^c_\nu \tag{4.2}$$

$\bar{\psi}_i$ are quark fields, m_f is the quark mass, γ^μ are the Dirac matrices and g_s is the coupling constant of the theory. The symbols t^a_{ij} are the $SU(3)$ generators and the f_{abc} are the structure constant of the $SU(3)$ algebra [36]. The matrices t^a form a complete basis of traceless 3×3 matrices. There are 8 such matrices[1], and therefore there are 8 gluons. The basis is chosen in such a way that

$$Tr\left(t^a t^b\right)\frac{1}{2}\delta^{ab} \tag{4.3}$$

The symbols f are defined by the commutator

$$[t^a, t^b] = if^{abc}t^c \tag{4.4}$$

[1] In the adjoint representation of $SU(n)$ group the generators are represented by $(n^2 - 1)$ matrices. In this case $n = 3$ and hence there are 8 matrices.

The colour structure of the Lagrangian may seem complicated at first sight. One simple way to look at it, is to think of quarks as objects having 3 states (called colours). The gluon can be thought as carrying the combination of a colour and an anticolour, except that out of the nine possible combinations the neutral one, formed by the sum of all equal colour-anticolour pairs is subtracted away.

4.1 Hadron Spectroscopy

Quarks are not free particles, but cluster into groups of two or three to form bound states called hadrons. When these hadronic resonances were first being observed, starting in the late 1940's, they were not recognized as a bound states of confined quarks. An accounting system for classification of the hadrons based on their mass and quantum numbers (charge and strangeness) was proposed independently by both Gell-Mann and New'eman in 1961, in which the hadrons are members of an SU(3) symmetry group. The scheme, called the "Eightfold-Way" [37], involved making unique assignments of the quantum numbers to each hadron according to certain patterns, and was substantiated when it correctly predicted the mass and quantum numbers of the hadrons not observed up until that point. The hadrons are separated into two groups according to their spin quantum number, which can be thought of as the particle's inherent angular momentum. Mesons, or two-quark systems have integer spin, while baryons, or three-quark systems, have half-integer spin. To each particle a quantic number called *Baryon Number* was associated; this number is calculated as:

$$B = \frac{1}{3}(n_q - n_{\bar{q}})$$

where n_q is the number of quarks and $n_{\bar{q}}$ is the number of antiquark. To the particles is also associated a flavour number for each of the quark of a given kind in the hadron. The diagrams that show the classification of the particles can be drawn; in figure 4.1 is shown, in figure 4.2 the Baryon Decuplet is shown and in figure 4.3 the Baryon Octet is shown.

One major problem with the Eightfold-Way classification was the existence of the Δ^{++} which, once identified as bound states of three identical quarks, violated the Pauli exclusion principle. The solution proposed, was to introduce a new quantum number: colour. There are three basic colours: red, green and blue, corresponding to the three colour states the quark can be found in. Similarly to anti-charge, the three colours have corresponding anti-colours, $\bar{r}$; $\bar{g}$ and $\bar{b}$. The hadrons observed in nature do not carry a colour charge, rather they are always produced as colour-singlets. This implies that either the total amount of colour is zero, i.e. there is as much colour as anti-colour contributing from the quarks or all three colours are present in equal amounts

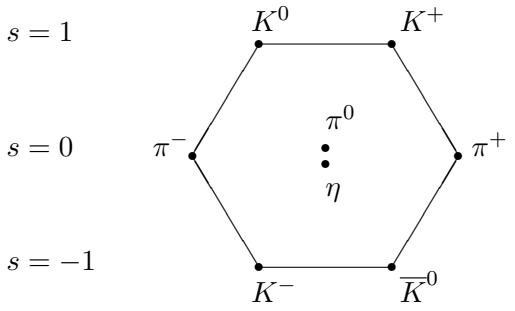


Figure 4.1: The meson octet. Particles along the same horizontal line share the same strangeness, s

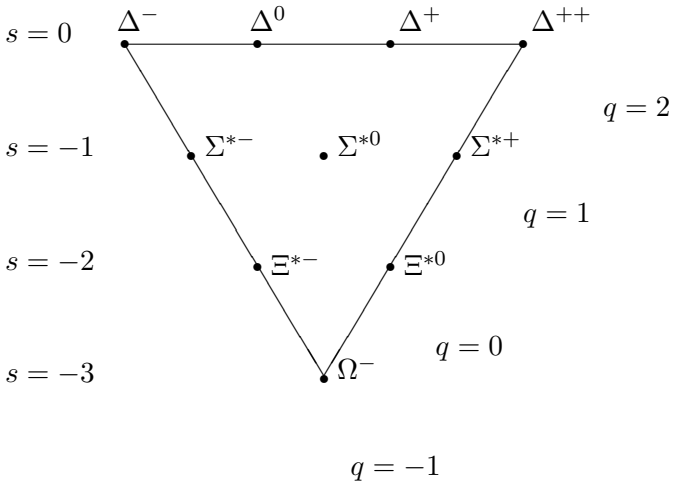


Figure 4.2: The baryon decuplet

4.2 Asymptotic Freedom and Quark Confinement

This theory possesses a number of important properties. In the low-energy regime it is strongly-interacting and produces an attractive force in a quark-antiquark and a three-quarks system. On the other hand, it predicts the asymptotic freedom, according to which the coupling decreases as the energy increases [38, 39] as it is shown in figure 4.4. The scale dependence of the coupling constant is encoded in the β function which has the following perturbative expansion [40],

$$\beta(\alpha_s) = \frac{\partial \alpha_s}{\partial \log \mu} = -\frac{\beta_0}{4\pi}\left(1 + \frac{\beta_1}{4\pi\beta_0}\alpha_s + \mathcal{O}(\alpha_s^2)\right) \tag{4.5}$$

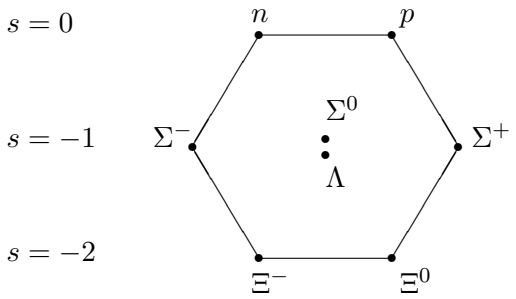


Figure 4.3: The baryon octet

where $\alpha_s = \frac{g_s^2}{4\pi}$. The coefficients β_0 and β_1 are defined as

$$\begin{aligned} \beta_0 &= \frac{33 - 2n_f}{3} \\ \beta_1 &= \frac{2(153 - 19n_f)}{3} \end{aligned} \tag{4.6}$$

where n_f is the number of flavours. If we neglect β_1 and the other higher order terms, at leading order the solution of the equation is

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + \alpha_s(\mu^2)\beta_0 \log(Q^2/\mu^2)} \tag{4.7}$$

QCD is very similar to QED but, since gluons can carry a colour charge while photons don't carry an electric charge, there is a substantial difference i.e. that gluons can interact with other gluons. This direct coupling between gluons has the dramatic effect of changing the strength of the force at different energy scales. The variation of the coupling constant is caused by the vacuum polarisation due to the production and annihilation of virtual quarks and gluons. In an effort to illuminate its effects within QCD we first consider how the corresponding effect arises in QED. Within QED so called charge screening can occur whereby virtual charged pair production occurring near a charged particle will cause the vacuum to become polarised in the region surrounding the particle. This polarisation causes the effective charge of the particle to vary with distance and results in an electromagnetic coupling constant, α_{em}, that increases with increasing momentum transfer, Q^2. In a similar way virtual quark-antiquark pairs also cause a screening of the colour charges of QCD. However, here the effect is more complicated since the gluons also carry colour charge and can alter the strong coupling constant in the same way. The net effect of the polarization caused by virtual gluons in the vacuum is not to screen the field, but to increase it. Although the strong coupling constant is large at low energies, asymptotic freedom allows cross-sections for processes involving the strong interaction to be calculated using perturbation theory so long as those processes occur at high energies. Moreover, since the effective strength of the strong force increases

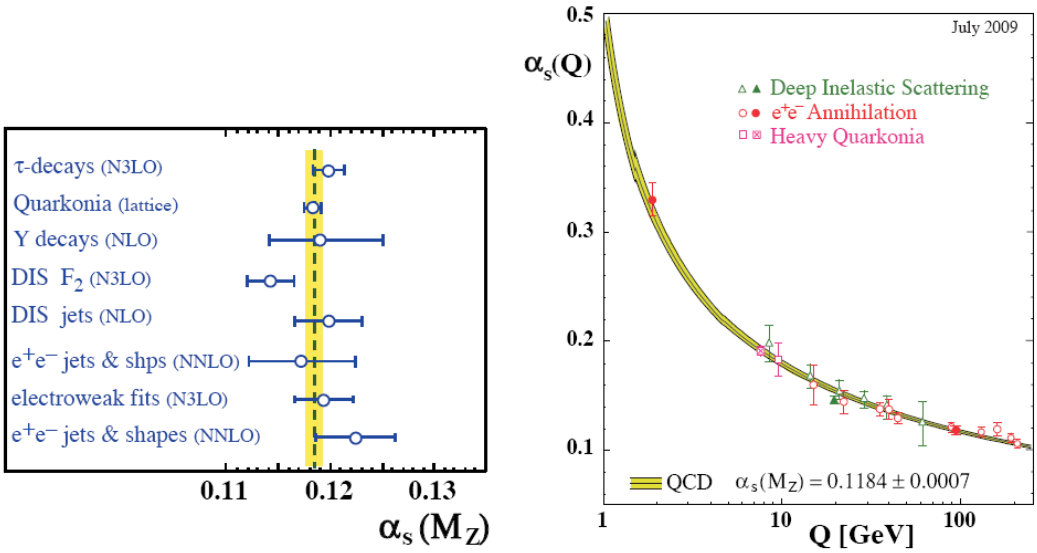


Figure 4.4: Left: Summary of measurements of $\alpha_s(M_Z)$, used as input for the world average value; Right: Summary of measurements of α_s as a function of the respective energy scale Q. Taken from [41]

with distance, if two quarks are separated, the strenght of the force between them will continue increasing until the field gains a sufficient amount of energy to create a new $q\bar{q}$ pair. This effect whereby no finite amount of energy can liberate a quark is called quark confinement and explains why no free quarks are seen in nature.

Even though QCD is asymptotically free, the computation of cross sections for any strong process always involves non-perturbative contributions, because the initial states and final states are not the fundamental degrees of freedom of the theory but compound states of quarks and gluons which cannot be described in perturbation theory. An important property of QCD is the factorisation theorem, which basically enables one to separate in every process a hard part, computable in perturbation theory, from a low energy one, which is process-independent and can be taken as a phenomenological input. The latter, given by the Parton Distribution Functions (PDFs), parametrise our ignorance on the inner structure of the nucleons. More details on the way in which PDFs are extracted from the existing data and on the related phenomenological issue are given in the next chapter.

4.3 The Parton Model

In the 60's by deep inelastic scattering (DIS) experiments were investigating the nucleon structure by colliding high energy electrons (in general leptons, also neutrinos) with

nucleons:

$$l(k) + h(P) \rightarrow l'(k') + X(P_X) \tag{4.8}$$

where in brackets we show the four-momentum of the particles. l is the initial lepton and l' is the final lepton. h is a nucleon and for X we mean any kind of final state product. A surplus of electrons at a large scattering angle was reported. Feynman gave a simple phenomenological explanation for this result: at the small distance probed by the electrons, the nucleon has to be considered as a gas of non interacting point-like particles, called partons. The electron simply scatters elastically with the components of the proton q, each of them carrying a fraction z of the total momentum of the nucleon according to the process

$$l(k) + q(zP) \rightarrow l'(k') + q'(p') \tag{4.9}$$

where q and q' are respectively the initial and final state partons. In the experiments, the deep-inelastic region is identified by the so-called Bjorken limit [42], which corresponds to considering events in which the mass M_X of the hadronic system in the final state is much larger than the nucleon mass. If we label the four-momentum of the exchanged

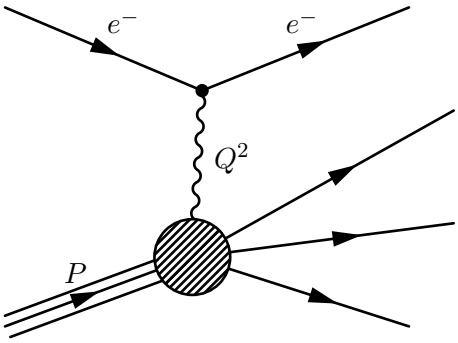


Figure 4.5: Feynman diagram of the deep inelastic scattering process

boson by $q = k - k'$ (where k and k' are the momentum of the lepton in the initial and final state), neglecting the relatively small lepton masses, the standard kinematic variables are defined by:

$$Q^2 = -q^2 = -(k - k') = 2E_l E_{l'}(1 - \cos\theta) \tag{4.10}$$

$$\nu = P \cdot q / M_N = (E_l - E_{l'}) \tag{4.11}$$

where the energy refers to the target rest frame, M_N is the mass of the target, E_l and $E_{l'}$ are the energies of the incoming and scattered lepton and θ is the lepton scattering

angle. The other important kinematic variables are defined as

$$x = x_B = \frac{Q^2}{2pq} = \frac{Q^2}{2M_N\nu} \tag{4.12}$$

$$y = \frac{q \cdot P}{k \cdot P} = \frac{E_l - E_{l'}}{E_l} \tag{4.13}$$

Once the energy of the lepton-nucleon centre of mass S is set, only two variable out of x_B, y and Q^2 are independent of each others, being

$$S = (P + k)^2 = \frac{Q^2}{yx_B} + M_N^2 \tag{4.14}$$

The cross-section for this process can be written as the product of a leptonic and a hadronic factor:

$$\frac{d^2\sigma}{dxdQ^2}Q^2 = \frac{\alpha_{em}}{Q^4}L^{\mu\nu}W_{\mu\nu}. \tag{4.15}$$

The leptonic tensor $L^{\mu\nu}$ it is easy to compute since it is a point-like object, while the hadronic tensor $W_{\mu\nu}$ contain the information on the nucleon structure. With general considerations $W_{\mu\nu}$ can be parametrized with the combination of the four-vectors of the involved particles and a set of arbitrary scalar functions. In an energy regime where the effect of the weak interaction can be neglected (i.e. energy significantly smaller than the Z mass) the tensor can be written [43]:

$$\begin{aligned} W_{\mu\nu} = &\left(-g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2}\right) F_1(x, Q^2) + \frac{1}{P \cdot q}\left(P_\mu - \frac{P \cdot q}{q^2}q_\mu\right)\left(P_\nu - \frac{P \cdot q}{q^2}q_\nu\right) F_2(x, Q^2) \\ &+ \frac{i}{P \cdot q}\varepsilon_\alpha^{\mu\nu\alpha\beta}q_\beta F_3(x, Q^2) \end{aligned} \tag{4.16}$$

where $\varepsilon^{\mu\nu\alpha\beta}$ is the Levi-Civita Tensor. The information on the a-priori unknown structure of the target as seen by the virtual boson is carried by the structure functions F_i. In the parton model the structure functions have a very simple expression, since the hadron is described in terms of probability density distributions for the momentum fractions of its parton constituents, which only depend on the hadron itself and are independent of the process. At large Q^2, the nucleon appears as a beam of pointlike constituents, very weakly interacting among themselves. This is the observation that leads to the formulation of the parton model. One can assume that the cross-section for any process

that involves a hadron with momentum p in the initial state can be written as:

$$\sigma(p) = \sum_i \int_0^1 dz f_i(z) \hat{\sigma}_i(zp) \tag{4.17}$$

where i is the index of the partons. The parton i carries a fraction x of the hadron momentum with a probability density $f_i(x)$ with the condition that

$$\sum_i \int dx x f_i(x) = 1 \tag{4.18}$$

This condition means that the sum of the momenta of the partons is equal to the total momentum of the hadron.

It has also be experimentally observed that the functions F_i also do not depend on the scale Q^2 [44]. Therefore in this simple picture, which was then understood as the $Q^2 \to \infty$ limit, the structure functions are observed to obey an approximate scaling law

$$F(x, Q^2) \to F(x); \tag{4.19}$$

this represent the so called *Bjorken Scaling*.

Let us assume that the parton i is a spin-1/2 massless particle $\mathcal{Q}$ with charge e and momentum zP. We can calculate the tensor $W^i_{\mu\nu}$ for the process:

$$\mathcal{Q}(zP) + \gamma^*(q) \to \mathcal{Q}(p') \tag{4.20}$$

Taking the average over the nucleon and the initial state lepton, only the first line of equation 4.16 survives and it becomes[2]:

$$W^i_{\mu\nu} = \frac{e^2}{2zPq} \left[z^2 P_\mu P_\nu + z P_\mu q_\nu + z P_\nu - g_{\mu\nu} zPq \right] \delta(1-x) \tag{4.21}$$

If we compare this result with the equation 4.16, we obtain that

$$F_1^i = \frac{e^2}{2} \delta(1-x)$$

and

$$F_2^i = e^2 \delta(1-x).$$

[2]The Levi-Civita tensor is a complete antisymmetric tensor i.e. changes sign under the exchange of any pair of indices. So it vanishes when multiplied by a symmetric tensor.

We can now compute the physical structure functions F_1 and F_2. Applying the parton model formula 4.17 we have:

$$W_{\mu\nu} = \sum_i \int_0^1 dz f_i(z) W^i_{\mu\nu}(zp, q) \tag{4.22}$$

Straightforward algebra leads to

$$F_1(x) = \frac{1}{2} \sum_i e_i^2 f_i(x) \tag{4.23}$$

$$F_2(x) = \sum_i e^2 x f_i(x) \tag{4.24}$$

and it also holds that

$$F_1(x) = \frac{1}{2x} F_2(x)$$

The functions f are called Parton Distribution Functions. By summing over all contributing partons, we must recover the quantum numbers of the proton: charge 1, baryon number 1, strangeness 0. So it follows that [46]:

$$\begin{aligned} \int_0^1 \left[u(x) - \bar{u}(x)\right] dx &= 2 \\ \int_0^1 \left[d(x) - \bar{d}(x)\right] dx &= 1 \\ \int_0^1 \left[s(x) - \bar{s}(x)\right] dx &= 0. \end{aligned} \tag{4.25}$$

It is also experimentally observed [47] that at values of Q^2 of a few GeV about half of the proton momentum is carried by the gluons $g(x)$. The momentum sum rule is also given:

$$\sum_0^1 \left[(q_i(x) + \bar{q}_i) + g(x)\right] x dx = 1. \tag{4.26}$$

The gluon contribution cannot be directly detected because the gluon has no electric charge and doesn't interact with the probe photon of the DIS of equation 4.20
As an example of the impressive agreement between theory and experiment, we show in Figure 4.6 the measurement of the total neutral current cross-section. These are the combined data of the two experiments, H1 and ZEUS, carried out at the electron-proton collider HERA, along with results from some fixed-target DIS experiments, spanning five orders of magnitude in both variables x and Q^2.

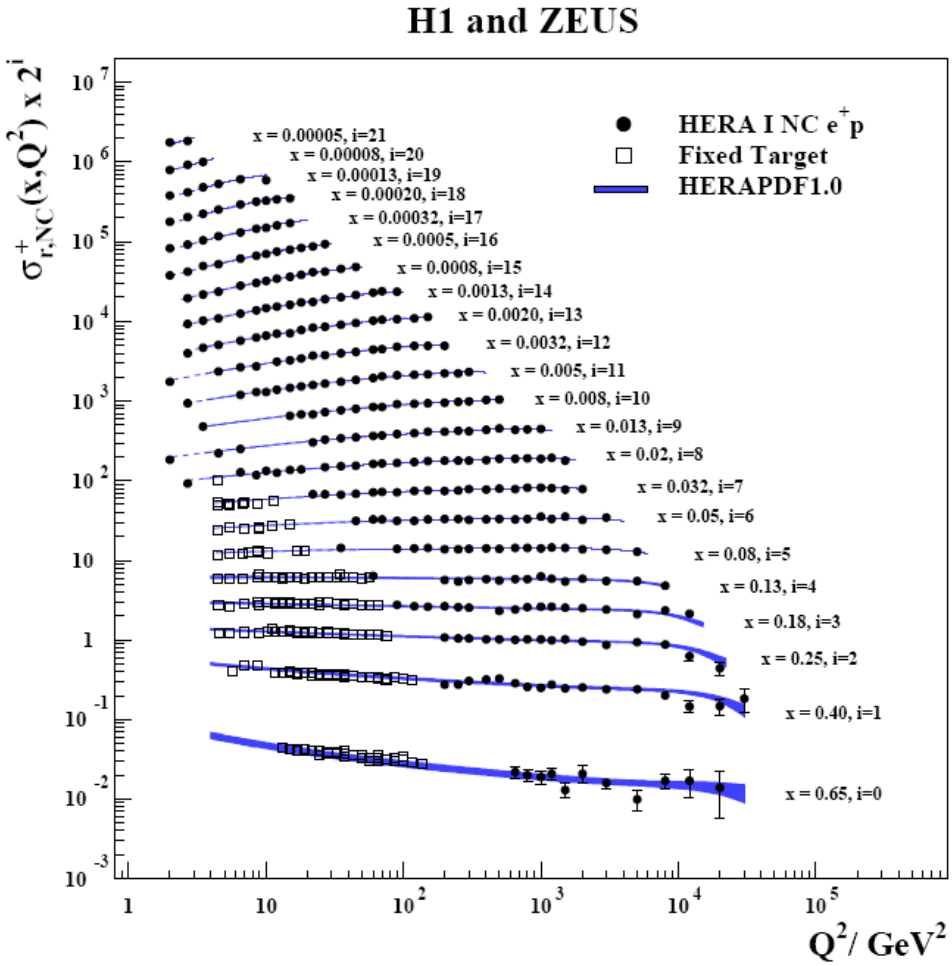


Figure 4.6: Modern DIS data compared to QCD predictions. HERA combined Neutral Current e^+p cross section and fixed-target data as a function of Q^2. The error bars indicate the total experimental uncertainty. The HERAPDF1.0 fit is superimposed. The bands represent the total uncertainty of the fit. Dashed lines are shown for Q^2 values not included in the QCD analysis. Taken from [45]

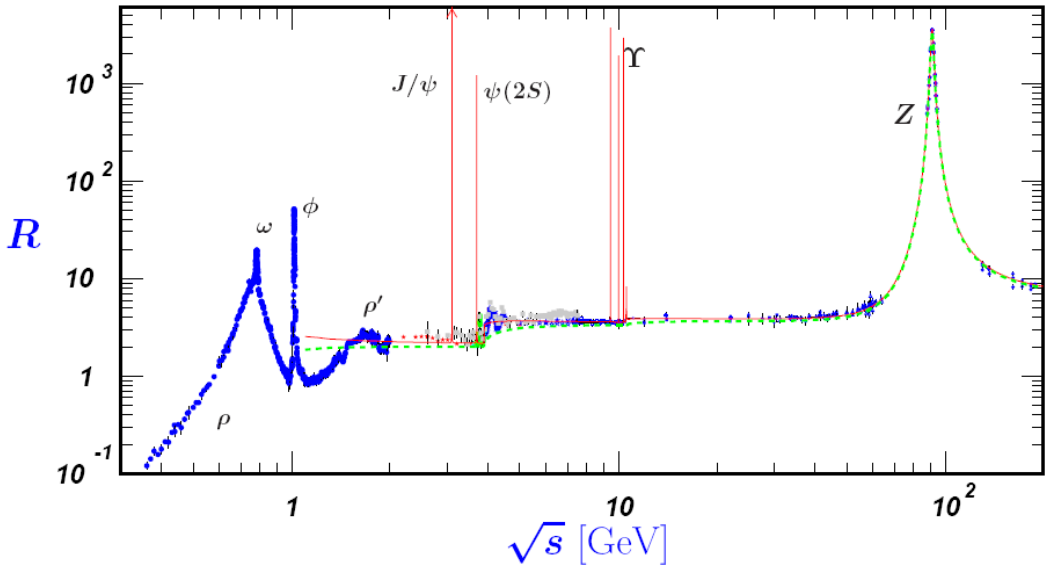


Figure 4.7: World data on the Ratio $R(\sqrt{s}) = \frac{\sigma(e^+e^- \to \text{ hadrons})}{\sigma(e^+e^- \to \mu^+\mu^-)}$. The curves represent different predictions: the broken one (green) is the naive quark-parton model prediction, and the solid one (red) is a 3-loop pQCD prediction. Taken from [49]

4.4 Other Experimental Confirmations

Hadron production in e^+e^- collisions

The idea of the existence of a set of subatomic elementary particles with different masses was confirmed by the measurement of the ratio of the cross-section of hadron production and the dimuon cross-section in electron positron collisions [48].

$$R = \frac{\sigma(e^+e^- \to \text{ hadrons})}{\sigma(e^+e^- \to \mu^+\mu^-)} = 3\sum_f Q_f^2 \tag{4.27}$$

The experimental results of the measurement of this ratio are shown in figure 4.7. This ratio depends on the number of active flavours, by which we mean the number of flavour that, depending on the center of mass energy, are allowed to be produced. The ratio R behaves roughly as step functions in the center-of-mass energy staying constant over certain ranges that now we know correspond to regions between heavy quark flavor thresholds. It is also important to note the consistency of the experimental results with the theoretical predicted multiplicative colour factor (3) which indirectly confirm the presence of three colours.

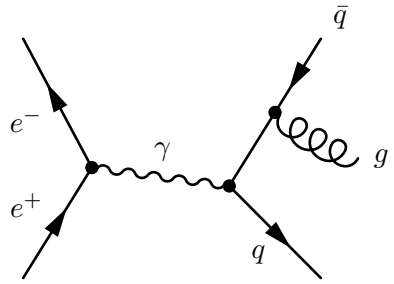


Figure 4.8: Feynman diagram of the three jet production with electron positron collision; an equivalent diagram in which the gluon is radiated from the quark leg has to be considered.

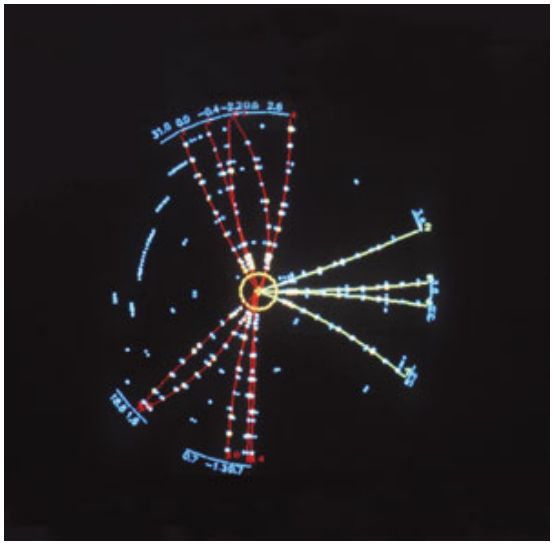


Figure 4.9: A three jet event in an event display of the TASSO experiment. Gluon radiation might cause the hadronic events to appear as a three-jet topology. The energy of the radiated gluon is normally smaller than the energy of the radiating quark. Taken from [6]

Three Jet Events in e^+e^- collisions

The first direct experimental evidence for gluons which confirmed the validity of QCD was obtained when the TASSO experiment, in June 1979, recorded three jet events from a electron positron collision provided by the PETRA accelerator at DESY [50, 51]. The existence of a third jet was related to the emission of one gluon by one of the final state quarks i.e. final states of the kind $gq\bar{q}$ was produced where the two quarks and the gluon hadronize and are detected as jets in the detector. Diagram for three jets production in e^+ e^- collisions are shown in figure 4.8 and an event display of this process at TASSO experiment is shown in figure 4.9. Hence the measurement of the ratio R together with the discovery of the gluon, brought strength and credibility to QCD as a theory of the strong interaction.

4.5 DGLAP evolution equations

In section 4.3 we considered the structure functions to be independent on the scale Q^2. However, when calculating QCD corrections to the process, collinear divergences arises [52]: these are also known as *Scaling Violations*. The scaling violations imply that the parton distributions inside the proton become Q^2-dependent and using resummation techniques it is possible to obtain an evolution equation to evolve the structure functions to different scales. The evolution equation is obtained in [52] and [53]. From this description it is possible to infer a mild dependence upon the energy scale which is determined by

$$\frac{d}{dt}q(x,t) = \frac{\alpha_s(Q)}{2\pi}\int_x^1 \frac{dy}{y}q(y,t)P\left(\frac{x}{y}\right) + \mathcal{O}(\alpha_s^2(Q)). \tag{4.28}$$

This equation is the QCD evolution equations for parton densities, also called the DGLAP or Dokshitzer-Gribov-Lipatov-Altarelli-Parisi equation [52]. P is the splitting function which represent the probability that the parton splits into another two partons. In a more compact way, using the convolution notation, they are expressed

$$\frac{d}{dt}q_i(x,t) = \frac{\alpha_s(Q)}{2\pi}[q_i \otimes P_{qq}] + \frac{\alpha_s(Q)}{2\pi}[g \otimes P_{qq}] \tag{4.29}$$

and for the gluon density:

$$\frac{d}{dt}g_i(x,t) = \frac{\alpha_s(Q)}{2\pi}[\sum_i (q_i + \bar{q}_i) \otimes P_{gq}] + \frac{\alpha_s(Q)}{2\pi}[g \otimes P_{gg}] \tag{4.30}$$

P_{qg} is the Altarelli-Parisi splitting function which universally describes the splitting of a gluon into a quark (or antiquark) and a gluon and P_{gg} is the gluon gluon splitting function. The DGLAP equation is based on the redefinition of the quark distribution; it is necessary to include the contributions from the probability that a gluon is emitted $q \to qg$ and that a quark antiquark pair arise from a gluon ($g \to q\bar{q}$). Likewise the redefined gluon involves contributions from the splitting of $q \to gq$, $\bar{q} \to g\bar{q}$ and $g \to gg$. The net effect is the dependence of the redefined distributions upon the energy scale. The higher the energy the higher the probability that the splittings occur and so, given the PDF at a certain energy scale, the DGLAP equation allows the evolution of the PDF to any other scale in the perturbative region.

4.6 The Factorization Theorem

The factorization theorem is a generalization of the parton model and provides a practical description of hadron interactions. This theorem states that, in an hadronic process (for example a proton proton collision at LHC), the total cross-section production of some particle X, can be written (apart from power suppressed corrections) as:

$$\sigma(s)_{pp\to X} = \sum_{ab} \int dx_a dx_b f_{a/A}(x_a, Q^2) f_{b/B}(x_b, Q^2) \hat{\sigma}_{ab\to X}(\alpha_s(Q), x_1 x_2 s) \tag{4.31}$$

where x_a and x_b are the fraction of momentum of the proton carried by the partons undergoing the hard process. $\hat{\sigma_X}$ is the elementary partonic cross-section and $f_{a/A}(x_a, Q^2)$ are the Parton Distribution functions of the flavour a. The PDF can be interpreted as the probability that a parton of the kind a belonging to the proton A, which carries a fraction of the proton momentum x_a, participates in the partonic interaction. The partonic cross-section has to be computed in perturbation theory while the PDFs, which absorbs all the long distance effects, cannot be obtained by perturbative QCD and have to be measured from experimental observables. However the nature of PDFs is universal, i.e. PDF do not depends on the physical process, and so, once known, it is possible to use them to make predictions of the rate for any given hard process. The QCD evolution equations 4.28 are essential in order to evolve the measured parton densities from one scale Q to a different one.

The scale Q^2 (usually referred to as μ when describing hadron interaction) is called the factorization scale. Qualitatively, is corresponds to the resolution with which the hadron is being probed. The partonic (hard-scattering) cross section is independent of the factorization scale μ at leading order in perturbative QCD, but, as shown in [54] depends logarithmically on μ at next-to-leading order and higher. When calculated to all orders in perturbative QCD, the hadronic cross section is independent of μ. However, at any finite order in perturbation theory, the calculated hadronic cross section depends on μ. This dependence is usually significant at low orders in perturbation theory.

4.7 Drell-Yan Process

We now present one of the easily calculable phenomenon of hadron collisions. The Drell-Yan process is the production of a lepton pair through a quarks antiquark annihilation. This process was proposed for the first time in 1971 [55] to explain the di-lepton production in hadron collisions within the parton model framework. At low energy this occurs through the production of a virtual photon:

$$q\bar{q} \to \gamma^* \to l^+ l^- \tag{4.32}$$

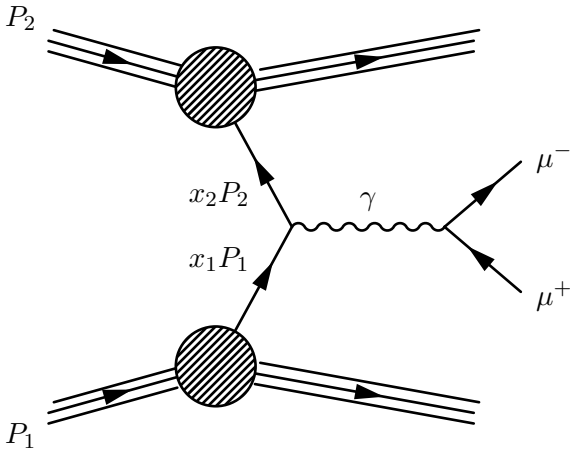


Figure 4.10: Drell-Yan mechanism for muon pairs production.

but, if enough energy is provided in the collision, production of a real Z boson or a real W boson, can occur. Clearly following the charge conservation law, the leptonic decay of the W will produce only one charged lepton and one neutrino while, if Z or γ^* is produced the classical opposite charge signature is generated. Despite this difference, the calculation of the partonic cross-section changes only in the multiplicative terms of the coupling constants and the colour factors and the presence of a massive propagator. The cross section for quark-antiquark annihilation to a lepton pair via an intermediate massive photon is easily obtained from the fundamental QED process [30]

$$e^+ + e^- \rightarrow \mu^+ + \mu^- \tag{4.33}$$

obtaining the partonic cross-section

$$\hat{\sigma}(q\bar{q} \rightarrow l^+l^-) = \frac{1}{N}\frac{4\pi\alpha}{3\hat{s}}Q_q^2 \tag{4.34}$$

where Q_q is charge of the quark. The colour factor $1/3 = 1/N$ arise from the fact that only when the colour of the quark matches with the colour of the antiquark the annihilation into a colour singlet final state can take place. We now consider the differential invariant mass distribution:

$$\frac{d\sigma}{dM^2} = \frac{\hat{\sigma}_0}{N}Q_q^2\delta(\hat{s} - M^2) \qquad \hat{\sigma_0} = \frac{4\pi\alpha^2}{3M^2} \tag{4.35}$$

where M is the dilepton invariant mass. In the center of mass framework of the two colliding hadrons the four-momenta of the parton can be written as:

$$p_1^\mu = \frac{\sqrt{s}}{2}(x_1, 0, 0, x_1) \qquad p_2^\mu = \frac{\sqrt{s}}{2}(x_2, 0, 0, x_2) \tag{4.36}$$

The square of the parton centre of mass energy $\hat{s}$ is related to the corresponding hadronic quantity by $\hat{s} = x_1 x_2 s$. Folding in the PDFs for the initial state quarks and antiquarks in the colliding beams gives the hadronic cross section:

$$\frac{d\sigma}{dM^2} = \frac{\hat{s}_0}{N} \int_0^1 dx_1 dx_2 \delta(x_1 x_2 s - M^2) \left[\sum_k Q_k^2 (q_k(x_1, M^2)\bar{q}_k(x_2, M^2) + [1 \leftrightarrow 2]) \right]. \quad (4.37)$$

From the description in equation 4.36, the lepton pair rapidity is

$$y = \frac{1}{2} \log \left(\frac{x_1}{x_2} \right) \quad (4.38)$$

and hence the relations

$$x_1 = \frac{M}{\sqrt{s}} e^y \qquad x_2 = \frac{M}{\sqrt{s}} e^{-y} \quad (4.39)$$

can also be used to obtain the double differential cross-section in mass and rapidity:

$$\frac{d\sigma}{dM^2 dy} = \frac{\hat{\sigma_0}}{Ms} \left[\sum_k Q_k^2 (q_k(x_1, M^2)\bar{q}_k(x_2, M^2) + [1 \leftrightarrow 2]) \right]. \quad (4.40)$$

Thus different values of M and y probe different values of the parton x of the colliding beams. The formulae relating x_1 and x_2 to M and y of course also apply to the production of any final state with this mass and rapidity.

4.8 Parton Showering

The previous section outlined the procedure for calculating the leading order cross-section for the Drell-Yan process at the LHC. This calculation can be extended to include higher order contributions - in fact in recent years exact calculations for this process have been made that include all terms up to next-to-next-to leading order (NNLO)[56]. Unfortunately, even these state-of-the-art calculations are still missing higher order terms that will contribute to the overall cross-section. The standard procedure for dealing with the higher orders that are not included in a given perturbative calculation is to add initial and final state parton showers to the process in question. This works in the following way. Since the partons carry electromagnetic and/or colour charge, the higher order contributions that are missing will take the form of emissions of gluons and photons. These missing terms can be approximated by adding additional radiative processes whereby a parton is allowed to branch into a parton with lower energy plus, for example, an emitted gluon, i.e. $q \rightarrow qg$. This branching is usually modeled using

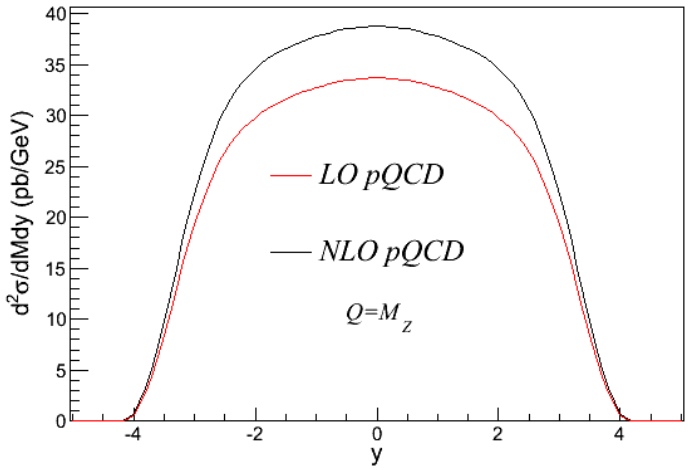


Figure 4.11: Double differential cross-section $d^2\sigma/dydM$ with $Q = M_Z$ for proton proton collision at $\sqrt{s} = 7\ TeV$; in red the LO calculation obtained with the equation 4.40 while in black the calculation with NLO QCD corrections as reported in [35]

the Altarelli-Parisi splitting function [52]. The showering continues until the branching quarks and gluons reach some predefined energy which is above the confinement regime. The Monte-Carlo generators PYTHIA [57] and Herwig [58] both apply a value of 1 GeV.

4.9 Hadronization

The hard scattering process can be calculated using a combination of the Factorisation theorem and perturbation theory and the missing higher order terms in the calculation can be approximated using the parton shower method. At this stage we have an ensemble of unphysical states that consist of coloured quarks and gluons. Hadronization is the mechanism by which quarks and gluons produced in hard processes form the hadrons that are observed in the final state. This is an intrinsically nonperturbative process, for which we only have models at present. Being models, none of them can lay claims to being 'correct', although some may be better founded than others. The best that can be aimed for is internal consistency, a good representation of existing data, and a predictive power for properties not yet studied or results at higher energies. The main current models are the *cluster model* described in [59] and *string fragmentation model* described in [57].

4.10 Underlying Event

In addition to the quarks and gluons that originate from the hard scatter there will also be a large number of coloured partons that come from the dissociation of the colliding hadrons which are converted in hadrons by the hadronization mechanism. These particles represent the so called Underlying Event (UE). The term Underlying Event is often used to describe the the component of the produced particles which do not originate from the hard interaction. Such a definition is likely to contain some ambiguity: experimentally it is not possible to identify the true origin of an hadron since it may have been the result of the hadronization of a number of partons that all have arisen from different sources. It has become common practice to identify the underlying event depending on the process type under study: The underlying event in Drell-Yan muon pair production would simply be all event activity apart from the two final-state muons. By contrast, the underlying event definition in jet events is more complicated and a strategy to associate particles in jet events either with the hard process or with the underlying event has to be developed.

The underlying event in hadron-hadron collisions is only weakly constrained by existing data and comprises many aspects not describable by perturbative QCD. Traditionally, measurements of minimum-bias events were used to parametrize the underlying event [60] and measurements by CDF collaboration at the Tevatron collider characterized the underlying event with Z and Drell-Yan events [61]. Only models including a description of several distinct parton-parton interactions seem to be able to describe the data [62]. The multiple parton interactions (MPI) are implemented in modern event generator programs. The two most widely used underlying event models, are implemented in the event generators Pythia [57] and for Herwig through JIMMY Generator [63].

4.11 Beyond Leading Order Calculations

In order to improve the accuracy of theoretical predictions in QCD phenomenology, higher order terms in the perturbative expansion have been studied and cross sections have been computed at next-to-leading order (NLO) and, in some cases, also at nextto-next-to-leading order (NNLO) [56]. However, in certain regions of the phase space, the expansion in powers of the coupling constant α_S is no longer good. Cross sections contain terms proportional to the logarithm of some kinematical variable ω:

$$\sigma_{pp\to X} = \alpha_S a_0 \ln\omega + \alpha_S^2(a_1 + \ln^2\omega + b_0 \ln\omega) + ... \tag{4.41}$$

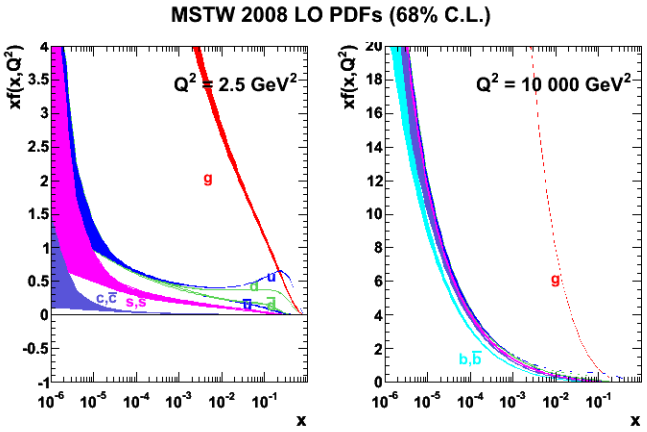
If ω is very large or close to zero the logarithm is large and hence:

$$\alpha_S \ln \omega \sim 1 \tag{4.42}$$

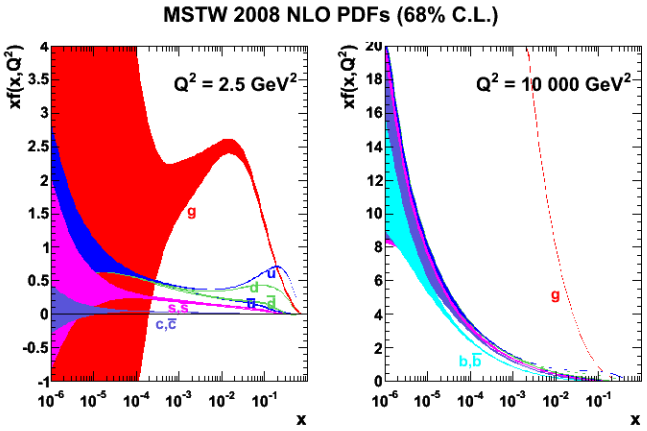
even if the coupling is small; this clearly invalidates the perturbative expansion. In order to get reliable predictions, these logarithms have to be resummed to all orders and the series rearranged as:

$$\sigma_{pp\to X} = \sum_k (\alpha_S \ln \omega)^{k+1} a_k + \alpha_S \sum_k (\alpha_S \ln \omega)^{k+1} b_k + ... \tag{4.43}$$

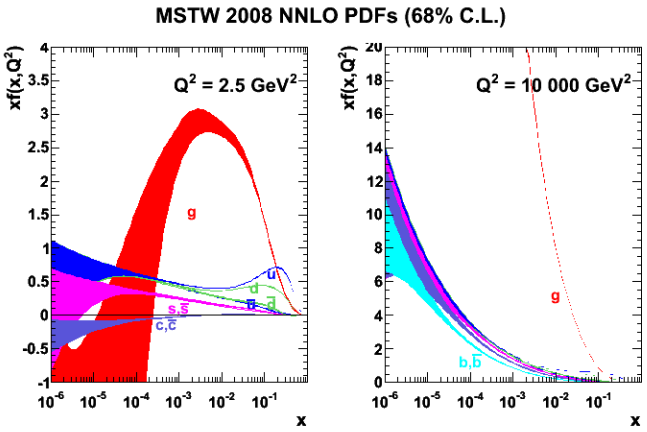
The first term corresponds to a leading order resummation, while the second one to a next-to-leading order one. One may encounter different kinds of logarithms and different techniques have to be used to perform the resummation. Logarithmic enhancement of higher order perturbative contribution may take place when more than one large scale ratio is present. In Deep Inelastic Scattering and Drell-Yan this happen in the two opposite limits when the center-of-mass energy of the partonic collision is much higher than the characteristic scale of the process, or close to the threshold for the production of the final state. These correspond respectively to the small x and large x kinematic regions, where $0 < x < 1$ is defined in terms of the invariant mass M^2 of the non-leptonic final state as $M^2 = \frac{1-x}{x} Q^2$ The corresponding perturbative contributions are respectively enhanced by powers of $\ln(1/x)$ which are due to the emission of collinear partons, and $\ln(1-x)$ which are originated by soft radiation. The theoretical status of small x and large x resummation is somewhat different. Large x logs are well understood and the corresponding perturbative corrections have been determined to all orders with very high accuracy. Indeed, the coefficients that determine their resummation can be extracted from fixed-order perturbative computations. Their resummation for Drell-Yan and Deep Inelastic Scattering was originally derived in [64, 65] and extended on very general grounds in [66]. With the low invariant mass Drell-Yan production in the forward region it is possible to reach value of $x \sim 10^{-6}$; this is one situation where the effect of the resummation can become significant [67] The resummation of logarithms of the hard scale of the process ($\omega = Q^2/\mu^2$), is performed thanks to the DGLAP evolution equation. The class of logarithms of the kind $\ln(1/x)$ can be resummed in the high energy, or small x limit, where the behaviour of the QCD is described by the Balitsky-Fadin-Kuraev-Lipatov (BFKL) evolution equation [68]. This equation, unlike the DGLAP equation which perform an evolution in Q^2, operates on the fractional momentum of the partons x. However, neither small x nor large x resummation and BFKL is systematically incorporated in current parton fits, so data points for which such effects may be important appear to be inconsistent. As an example of the low-x problem in parton fit, we show the results of the MSTW08 fit when performed with different order calculation in pQCD. In particular the gluon distribution (red) at low-x has very different behaviour when fit is performed at LO Fig. 4.12(a), NLO Fig. 4.12(b) and NNLO Fig. 4.12(c).



(a)



(b)



(c)

Figure 4.12: MSTW08 parton fit performed at different order in perturbation theory.

5. PARTON DISTRIBUTION FUNCTIONS

In this chapter several aspects of the determination of parton distribution functions (PDF) are explored. In section 5.1 we are going to describe how global fits are perform in order to extract PDF from experimental data. We describe both experimental and theoretical inputs of the so-called global analyses and the statistical issues associated to the determination of a set of functions from a finite number of data points. Statistical tools and formalism used in the global fits are described in sections 5.3 and 5.4 and a the current knowledge of the PDF is given in section 5.5

5.1 Global fits

Parton Distribution Functions are universal, process-independent, non-perturbative quantities which must be extracted from the comparison to the available experimental data with the functions (that depends on several parameters) which are supposed describe the structure of the proton. Global analysis of parton distributions involves making use of experimental data from many physical processes and using the parton evolution equations to extract a set of universal parton distributions which best fit the existing data. These distributions can then be used in predicting all other physical observables at energy scales far beyond those presently achievable [69]. To perform a global analysis some steps are necessary:

1. Develop a program to numerically solve the evolution equations 4.28;
2. Make a choice of experimental data sets, such that the data can give the best constraints on the parton distributions;
3. Choose the parametric form for the input parton distributions at a given scale μ_0, and then evolve the distributions to any other values of μ_f;

4. Use the evolved distributions to calculate the χ^2 between theory and data, and choose an algorithm to minimize the χ^2 by adjusting the parameterizations of the input distributions;

5. Parameterize the final parton distributions at discrete values of x and μ_f by some analytical functions.

In all high-energy data, deeply inelastic scattering of leptons on nucleon and nuclear targets are the primary source of information on parton distributions, because of its high-statistics. Such data are known to be mostly sensitive to certain combinations of quark distributions. Drell-Yan lepton pair production, and direct photons at large transverse momenta provide important complementary information on antiquark and gluon distributions. Most data used in obtaining recent parton distributions are at fixed target energies, HERA data and Tevatron data.

PDFs are derived from a global analysis of experimental data from a wide range of hard processes in the framework of the perturbative quantum chromodynamics (pQCD). In a global analysis it is desirable to include as many experimental data sets as possible in order to constrain PDFs and reduce as much as possible their uncertainty. However the large amount of experimental data-points ($\sim$ 1200) do not come from a uniform set of measurement, but consist of a collection of measurement from many experiments, on several physical observables. The complexity of the set of data makes difficult the application of the standard statistical tools. For example when different data-sets of the same or similar processes are included, they may not be consistent according to the standard statistical tests, even if each of individual data-set is self-consistent. Several approaches have been proposed and adopted by the fitting groups; later we will describe some of this solutions.

5.1.1 Theoretical uncertainties

Besides the experimental uncertainties, a global fit procedure has to deal with some theoretical assumptions that can lead to different results of the fit and so can be considered as sources of uncertainty. The main source of uncertainties comes from the higher order contributions: only recently have parton fits been developed at NNLO, but not all processes included in the fits, for example the inclusive jet calculations, are not available at NNLO. In addition some hadronic observables require a too long calculation to be evaluated at high order in the framework of global fitting hence a *K-factor*[1] [70] approximation is often enforced. The others important assumptions are the choice of the parametrization and the flavour decomposition. In principle a set of 11 different partons has to be considered

$$u, \bar{u}, \quad d, \bar{d}, \quad s, \bar{s}, \quad c, \bar{c}, \quad b, \bar{b}, \quad g$$

[1] *K-factor* approximation consists in the assumption that physical observable calculated at high order, (i.e. NLO or NNLO) can be obtained through calculation at lower order multiplying the result by an appropriate constant factor K

since $m_b, m_b \gg \Lambda_{QCD}$ heavy parton distributions can be determined perturbatively. Only a set of 7 independent distributions are left:

$$u_V = u - \bar{u}, \quad d_V = d - \bar{d}, \quad sea = 2 * (\bar{u} + \bar{d} + \bar{s}),$$

$$s^+ = s + \bar{s}, \quad s^- = s - \bar{s}, \quad \bar{d} - \bar{u}, \quad g$$

Global fits have to rely on some assumption since the data cannot constrain all of these 7 distribution. This may cause a bias in the fit results. For example in the past, two typical assumption were to consider a null strange valence distribution ($s^- = 0$) and the $\bar{u}$ content equal to $\bar{d}$. Recent Global Fits including Drell-Yan and charged-current data allow to free these constraint and to determine independently $\bar{d},\ \bar{u}\ s,\ \bar{s}$.

The crucial assumption of the PDF fitting is the choice of the parametrization. Usually PDFs are parametrized at an initial scale with a polynomial function such as:

$$xf(x, Q_0^2) = (1 - x)^\eta (1 + \epsilon x^{0.5} + \gamma x) x^\delta \tag{5.1}$$

The parameter η is motivated by the consideration that in the limit $x \to 1$ PDFs are supposed to be 0 while the parameter δ drives the behaviour in the small-x limit. Since PDFs describe our lack of knowledge of the non perturbative proton structure, there is no reason to prefer one or another parametric form and so it is clear that the fit results are deeply affected by the theoretical prejudice which is imposed by the choice of the parametrization. For this reason a completely different approach based on the use of neural networks has been developed by the NNPDF collaboration that provides a very generic and flexible parametric form for the PDFs. This method is going to be described in section 5.4

5.2 The fitting procedure

In general (irrespective of the choice of parametrization form and the choice of the datasets included in the fit), global fits use a χ^2 minimization to extract the PDFs from the data i.e.s one wants to determine the set of parameters $\{a\}$ which minimize the χ^2:

$$\chi^2(\{a\}) = \sum_{i,j}^{N_{data}} \left(F_i^{exp} - F_i^{theo}(\{a\})\right) C_{ij}^{-1} \left(F_j^{exp} - F_j^{theo}(\{a\})\right) \tag{5.2}$$

where i and j run on the N_{data} experimental data-points, F_i^{exp} are the experimentally measured values, $F_i^{theo}(\{a\})$ are the theoretical predictions for that particular observ-

ables and they are a function of the set of parameters $\{a\}$ which describe the set of PDFs. C is the experimental covariance matrix which provide all the information about statistical, correlated and uncorrelated systematic uncertainties of the data-points ensemble. Once the parameters that give the best description of the data are extracted, an estimate of their uncertainties is required. We are going to describe two different methods of evaluating uncertainties on the PDFs.

5.3 Hessian Method

One of the most efficient approaches of studying the uncertainties in a global analysis is through a quadratic expansion of the χ^2 around its global minimum. This method is also known as the Error Matrix or Hessian method [71]. The error determination through the hessian method is widely used in the PDFs fitting community and has been deployed in the PDF sets MSTW, CTEQ, Alekhin and HERAPDF [72, 73, 74, 75] using a polynomial parametrization of the kind described in equation 5.1.

Assuming that the deviation in χ^2 for the global fit from the minimum value χ^2_0 is quadratic in the deviation of the parameters a_i from their values at the minimum a_i^0 it is possible to write

$$\Delta\chi^2 = \chi^2 - \chi^2_0 = \sum_{i,j}^{n} H_{ij}(a_i - a_i^0)(a_j - a_j^0) \tag{5.3}$$

where n is the number of free input parameters and H_{ij} is the element of the Hessian Matrix and it is defined as

$$H_{ij} = \frac{\partial^2\chi^2(\{a\})}{\partial a_i \partial a_j} \tag{5.4}$$

In this case the standard linear propagation of the errors allows to calculate the errors on any variable F depending on the PDF parameters $\{a\}$ using the formula:

$$(\delta F)^2 = \Delta\chi^2 \sum_{i,j}^{n} \frac{\partial F}{\partial a_i} V_{ij}(a) \frac{\partial F}{\partial a_j} \tag{5.5}$$

where $V = H^{-1}$ is the covariance matrix of the parameters produced by the global fit. $\Delta\chi^2$ is the allowed variation of the χ^2: uncertainties on the parameters are usually obtained finding the values of each parameter that generates a deterioration of the χ^2 of a given amount $\Delta\chi^2$. The choice of $\Delta\chi^2$, that will be described in section 5.3.1, is a critical feature of the PDF Global Analysis.
However, as demonstrated in [76], it is convenient to diagonalize the covariance matrix

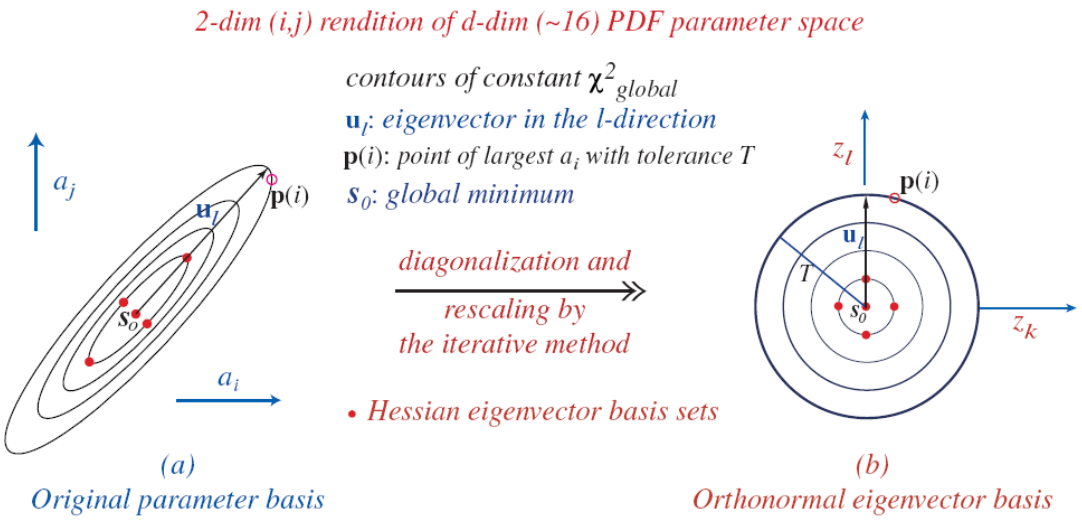


Figure 5.1: Schematic representation of the hessian method, taken from [77]

and use a set of complete orthonormal eigenvectors v_{ik} with eigenvalues λ_k

$$\sum_{j}^{n} V_{ij}(a) v_{jk} = \lambda_k v_{ik}. \tag{5.6}$$

Since the variations in some direction in parameter space lead to a deterioration of the quality of the fit far more quickly than others and the eigenvalues λ_k span several orders of magnitude it is helpful to use rescaled eigenvectors

$$e_{ik} = \sqrt{\lambda_k} v_{ik}$$

therefore the parameter displacement from the global minimum of the fit can be written

$$\Delta a_i = (a_i - a_i^0) = \sum_{k=1}^{n} e_{ik} z_k \tag{5.7}$$

or, using the orthogonality of the eigenvectors,

$$z_i = (\lambda_i)^{-1} \sum_{k=1}^{n} e_{ki} \Delta a_k \tag{5.8}$$

in which z_i are normalized combinations of the Δa_k which define the orthogonal directions in the space of the deviation of the parton parameters. This means that the z_i are the contribution to the χ^2 displacement from the global minimum. The increase of the χ^2 is:

$$\Delta\chi^2 = \sum_{i=1}^{n} z_i^2 \tag{5.9}$$

To determine errors on any variable F function of the parton parameters is now only necessary to evaluate

$$(\delta F)^2 = \Delta\chi^2 \sum_{i=1}^{n} \left(\frac{\partial F}{\partial z_i}\right)^2 \tag{5.10}$$

which corresponds to the usual practice of adding in quadrature uncorrelated uncertainties. This description provides an easier way (compared to the one in eq 5.5) to determine uncertainties since the uncertainties on the parameters can be treated as independent. When the parameter displacement 5.7 assumes positive or negative values it can produce a different χ^2 deterioration and so it is convenient to define parton sets $S_k^{\pm}$for each eigenvector direction from eq. 5.7

$$\Delta a_i(S_k^{\pm}) = \pm t e_{ik} \tag{5.11}$$

where t is the tolerance defined as $t = \sqrt{\Delta\chi^2}$ and $\Delta\chi^2$ is the allowed deterioration in the fit quality for the error determination. The set $S_k^{\pm}$ are defined by positive and negative displacement of magnitude t along each of the n eigenvector directions. Assuming the quadratic behaviour of the χ^2 the uncertainty can be re-written as

$$(\delta F)^2 = \frac{1}{2}\sum_{k=1}^{n} \big(F(S_k^+) - F(S_k^-)\big)^2 \tag{5.12}$$

The differences between S_k^+ and S_k^- are due to non quadratic shapes of the χ^2; so if one wants to obtain the uncertainty on any function these asymmetries have to be taken into account. The simple equation 5.10 has to replaced by the more sophisticated asymmetric uncertainties:

$$(\delta F)_+^2 = \sum_{k=1}^{n} \big(max(F(S_k^+) - F(S_k^0), F(S_k^-) - F(S_k^0), 0)\big)^2 \tag{5.13}$$

$$(\delta F)_-^2 = \sum_{k=1}^{n} \big(max(F(S_k^0) - F(S_k^+), F(S_k^0) - F(S_k^-), 0)\big)^2 \tag{5.14}$$

where S_k^0 is the best fit set of partons.

5.3.1 Choice of the tolerance

Standard statistics textbooks state that if a set of measurements is self-consistent, $\Delta\chi^2 = 1$ (i.e. the tolerance t is set to one) should allow a description of the data within one standard deviation error. This means that 68% of the datapoints lie within the band generated when using a set of paramaters that produces an increment in the χ^2 equal to one respect to the global minimum. However, due to the complicated nature of the global fitting procedure which includes a large amount of independent datapoints, this scholastic approach produces unrealistic small uncertainties [78]. The inconsistencies are partly due to some failures of the theoretical approximations to work properly in the full data range and they are strictly related to the theoretical assumption outlined in section 5.1.1. On the other hand some source of experimental error could also be underestimated, but both problems are in practice extremely difficult to surmount. The problem of the tolerance has been an hot topic in the PDF fitting community and each working group developed its own strategy. We briefly describe here some of them:

- **Consistent data:** One solution to this problem can be the choice of a subset of consistent data-points; this allows one to keep the usual definition of error with $\Delta\chi^2 = 1$ but it downgrades the analysis to a non full global fit. This is case of the PDF sets Alekhin02 [74] and HERAPDF [75].
- **Average tolerance:** The CTEQ collaboration tried to formulate a reasonable setting for the tolerance t by estimating the range of overall χ^2 along each of the eigenvector directions within which a good fit to all data-sets (within 90% C.L.) can be obtained. Then an average of the ranges over eigenvector directions is calculated. This method leads to an evaluation of the tolerance of $t \sim \sqrt{100}$ [79].
- **Dynamic determination:** Recently the MSTW group developed a new way of determining the tolerance. This method uses the so called "hypothesis-testing" criterion considering the eigenvectors PDF sets as the alternative hypothesis that have to be tested. It demands that, for each of the eigenvectors directions each data set must be described within its one-sigma (or 90% CL) limit. At the end of the procedure a value of tolerance of each eigenvector is provided and it is used to compute the $S_k^{\pm}$ sets [72].

These two last methods give a reasonable estimate of the errors but it does not rely on any real mathematical or statistical foundation and, in general, it has the net effect of inflating the experimental errors. The tolerance issue has been overcome by the NNPDF group using a Monte Carlo approach in the error determination. We will describe this method in section 5.4

5.4 The NNPDF Method

The NNPDF collaboration developed an innovative method of fitting the Parton Distribution Functions. The novelty is twofold: the PDFs are parametrized with a neural network and uncertainties are estimated with a Monte Carlo method. The use of neural networks provides a robust and flexible parametrization of the parton distributions at the initial scale, while the use of the Monte Carlo sampling allows to evaluate all quantities such as the uncertainty or the correlation of the PDFs with a robust statistical meaning [80]. We now briefly describe these features:

- **The Monte Carlo sampling:** A faithful estimation of errors on PDFs have to face several issues: first of all when you want to estimate an error of a single quantity it is easy to define the 1-σ error. It is also easy to define for multiple quantities a 1-σ contours but in the case of the PDFs it is necessary to define an "error band" in the space of functions, i.e. the probability density $\mathcal{P}[f]$ in the space of functions $f(x)$. Hence given a quantity which depends on PDFs $\mathcal{O}(f)$ (or a PDF itself) its average is given by integrating over the space of functions V

$$\langle \mathcal{O} \rangle = \int_V df \mathcal{P}[\{f_i\}] \mathcal{O}(\{f_i\}) \tag{5.15}$$

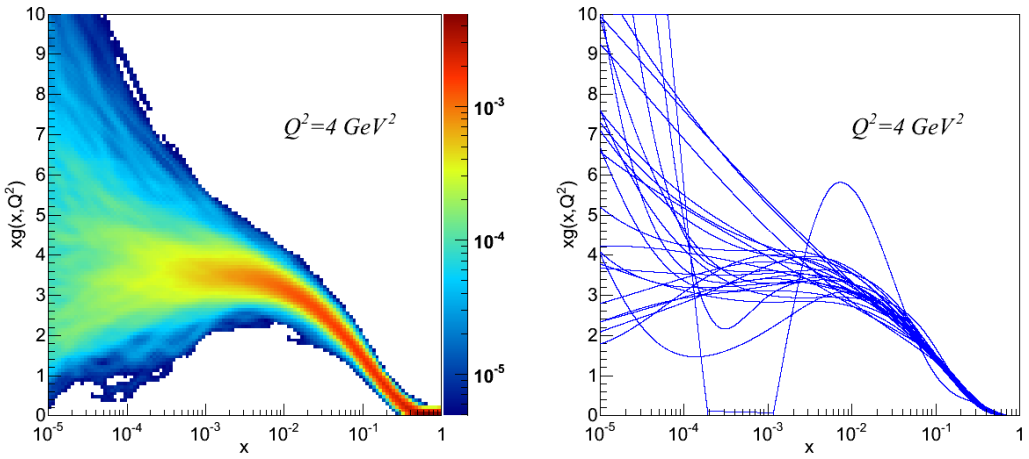
  In the NNPDF approach the probability measure is represented by a Monte Carlo sample in the space of PDFs defined in two steps: first the generation of an ensemble of replicas of the original data set, such that it reproduces the statistical distribution of the experimental data, followed by its projection into the space of PDFs through the fitting procedure.

- **The Neural Net Parametrization:** To avoid the introduction of theoretical bias in the PDFs determination it is necessary to ensure the maximum flexibility for the functional form. Lack of redundancy can lead to an artificially reduction of the parton uncertainty in regions where data do not constrain enough the PDFs. This flexibility can be achieved in different ways, for example using more complex polynomial function or with a neural network. In the NNPDF method each of the independent PDFs is parametrized with a multi-layer neural network with 37 parameters for each PDF (i.e. 185 free parameters for the full set) which can easily accommodate any functional form.

Therefore the method can be summarized in two steps: First one generates a Monte Carlo ensemble of replicas of the original data points that one wants to include into the fit. The ensemble is generated with the probability distribution of the data and contains all the available experimental information. Each element in the Monte Carlo set is a replica of the experimental data and contains as many data points as are originally

available.

The second step consists in the construction of a set of parton distributions from each replica of the data. Each independent PDF at a given scale is parametrized by an individual neural network. Physical observables are computed from parton distributions by evolving the initial scale parton distributions to the scales of the experimental measurements by using the DGLAP evolution equations. Physical observables are computed by convoluting the evolved parton distributions with hard partonic cross sections. The best fit set of parton distribution is determined by comparing the theoretical computation of the observable for a given PDF set with their replica experimental values by a χ^2 minimization.[2]



(a) (b)

Figure 5.2: NNPDF Gluon distribution at $Q^2 = 4\ GeV^2$: (a) is a 2D representation of the probability distribution sampled by 1000 replicas while (b) are 25 members of the set of Monte Carlo replicas of the gluon distribution.

The result of these two stages is an ensemble of equally likely Monte Carlo replicas of the PDF set that can be used to make predictions on physical observables; In figure 5.2 the result of the NNPDF fitting method for the gluon distribution is shown: in 5.2(b) 25 Monte Carlo replicas of the gluon distribution are plotted while in 5.2(a) the probability density function of the gluon distribution described by the ensemble of 1000 NNPDF replicas is plotted.

[2]Here the best fit is not given by the absolute minimum of the χ^2 because, due to the extreme flexibility of the neural net, it is necessary that the best fit does not attempt to reproduce random fluctuations of the data but only the underlying physics. For this reason a strategy to choose the best fit was developed in [81]

It is possible to use each replica of the PDFs set inside a Monte Carlo generator[3] and generate physical observables. The best estimation of the observable $\mathcal{O}$ is obtained averaging on all different outputs of the Monte Carlo generator:

$$\langle \mathcal{O}\left[f\right]\rangle = \frac{1}{N}\sum_{k=1}^{N} \mathcal{O}\left[f^{(k)}\right] \tag{5.16}$$

where N is the number of replicas, $\mathcal{O}$ is a generic function of the PDF set and $f^{(k)}$ are PDF replicas. This form is the discrete version of the integral evaluation in 5.15.
The uncertainty on the observable can be expressed through the usual definition of the standard deviation.

$$\sigma^2_{\mathcal{O}} = \left(\frac{1}{N-1}\sum_{k=1}^{N}\left(\mathcal{O}\left[f^{(k)}\right] - \langle \mathcal{O}\left[f\right]\rangle\right)^2\right) \tag{5.17}$$

Similarly it is possible to define correlations between different observables $\mathcal{A}$ and $\mathcal{B}$, for example:

$$\sigma_{\mathcal{AB}} = \frac{1}{N-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\mathcal{A}[f^{(i)}] - \langle \mathcal{A}\left[f\right]\rangle\right)\left(\mathcal{B}[f^{(j)}] - \langle \mathcal{B}\left[f\right]\rangle\right) \tag{5.18}$$

A covariance matrix can also be built. For example, each bin of a rapidity distribution is considered an observable, and so, to describe the correlations between M bins, one needs to build an $M \times M$ matrix where the elements are defined as

$$M_{ij} = \sigma_{ij}$$

where the elements σ_{ij} are calculated as equation 5.18.

5.5 Current Knowledge of the PDFs

In this section we give a brief review of the curent knowledge of the PDF; Experimental and kinematics's range of datapoints included in the global analysis are described in section 5.5.1 and the results of global fits and a comparison of different PDF set is given in section 5.5.2.

[3]For our study we used MCFM 5.7

5.5.1 Experimental data and Kinematical Range

In this section we present the datasets included in a typical global analysis, namely the NNPDF2.0 fit [80] and the consequent knowledge of the PDFs that these data provide. The kinematical region covered by these data is shown in figure 5.3

Fixed target data

The core of a global analysis is based on a comprehensive set of experimental data from deep-inelastic scattering with various lepton beams and nucleon targets. These data include the proton and deuteron structure functions determined in fixed-target experiments by the BCDMS [82, 83] and NMC [84, 85] collaborations. They provide the most accurate and up-to-date information on the valence region of parton distributions. They are supplemented with data on the structure functions from SLAC [86] which, though rather older and less precise, improve the kinematic coverage in the large x region. Altogether these data cover the middle- to large-x and smaller Q^2 region of the kinematical range, corresponding to the lower-right corner in figure 5.3.

DIS data

Collider experiments have explored a larger kinematical range in great detail. Neutral and charged current cross-sections, from H1 [87, 88, 89, 90] and ZEUS [91, 92, 93, 94, 95] collaborations were used. Both neutral and charged current scattering data from charged lepton beams and neutrino scattering data enable one to disentangle the quark and antiquark distributions. In the most recent global analysis [80] these data are replaced by the HERA-I run with the combined set of Ref. [96]. The combined HERA-I dataset has a better accuracy than the one expected on purely statistical grounds from the combination of previous H1 and ZEUS data because of the reduction of systematics from the cross-calibration of the two experiments. These HERA data sets yield information in a much wider region of the (x, Q^2) plane, in both the small-x and the large-Q^2 directions. In the fits are also included neutrino DIS data: specifically, the large, up-to-date, and consistent set of neutrino and antineutrino scattering data by the CHORUS collaboration [97] are used. These data have a similar kinematic coverage to the fixed target charged lepton DIS data. In the global fits like [98] the analysis supplemented by data on deep-inelastic neutrino production of charm from NuTeV [99] (dimuon data, henceforth) which give an handle on the strange distribution. NuTeV dimuon data overlap with the rest of fixed target experiments, providing information of the proton strangeness for $x > 10^{-2}$.

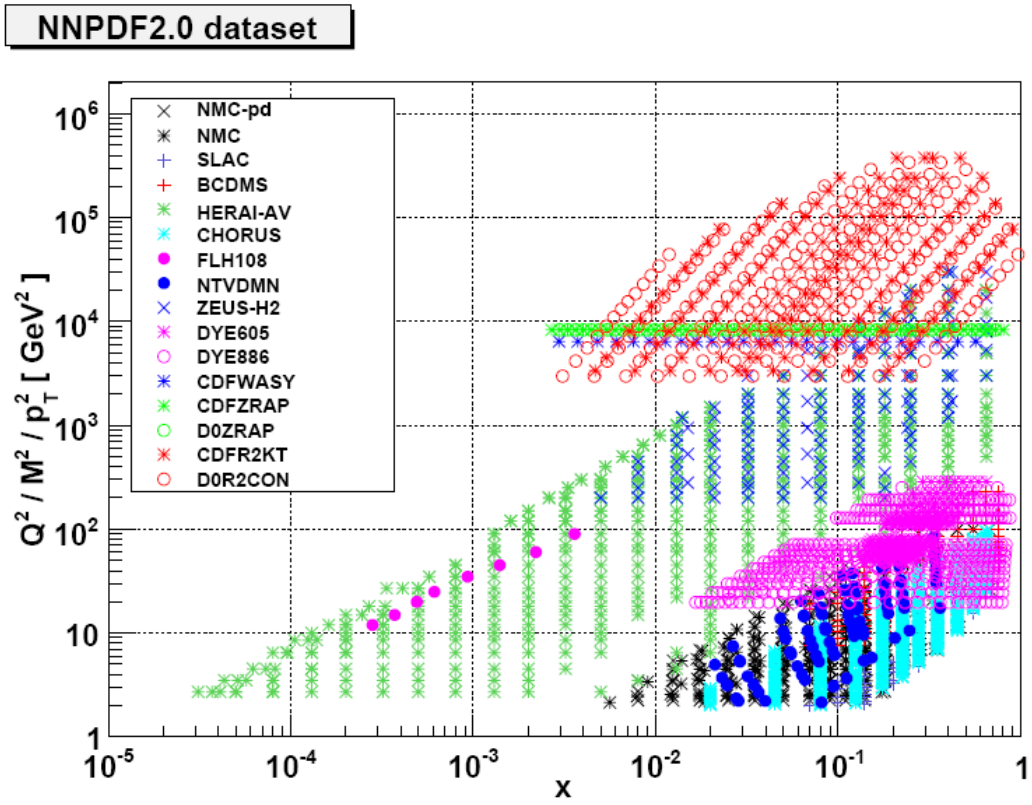


Figure 5.3: Experimental data which enter the NNPDF2.0 analysis.

Hadronic Data

Usually, in the parton fitting three classes of hadronic processes were included into the fit: Drell-Yan production in fixed target experiments, collider weak vector boson production, and collider inclusive jet production.

The fixed target Drell-Yan data included are usually E605 and E866 datasets. The former provides the absolute cross-section for Drell-Yan production from a proton beam on a copper target [100]. E886, also known as NuSea, is based on the experimental set-up of E605. The absolute cross section measurements on a proton target is described in [101, 102], while the cross section ratio between deuteron and proton targets can be found in [103]. These datasets provide information in a high-x and mid-Q^2 region.

The weak boson production data included in the fits are the D0 and CDF Z rapidity distribution and the CDF W boson asymmetry. The D0 Z rapidity distribution measurement was performed at Tevatron Run II and is described in [104]. The CDF Z rapidity distribution is analogous to its D0 counterpart, and it is described in [105]. The CDF W boson asymmetry measurement, also performed at Tevatron Run II, is described in

[106] Vector boson production explores values of $Q^2 = M_Z^2$ and $Q^2 = M_W^2$ with x down to values of $9 \cdot 10^{-2}$.

Finally also the inclusive jet production cross section as a function of the transverse momentum p_T of the jet for fixed rapidity bins are included as described in [107, 108]. As it is possible to see from figure 5.3 (red empty circle and red stars) these measurements provide a coverage at very high-Q^2.

5.5.2 Parton Distributions

The result of the fits of these datasets leads to the production of a set of parton distribution functions equipped with the uncertainties. We show the parton sets in figure 5.4 for MSTW08, and in figure 5.5 for the set HERAPDF where the experimental, model and parameterization uncertainties are shown separately.

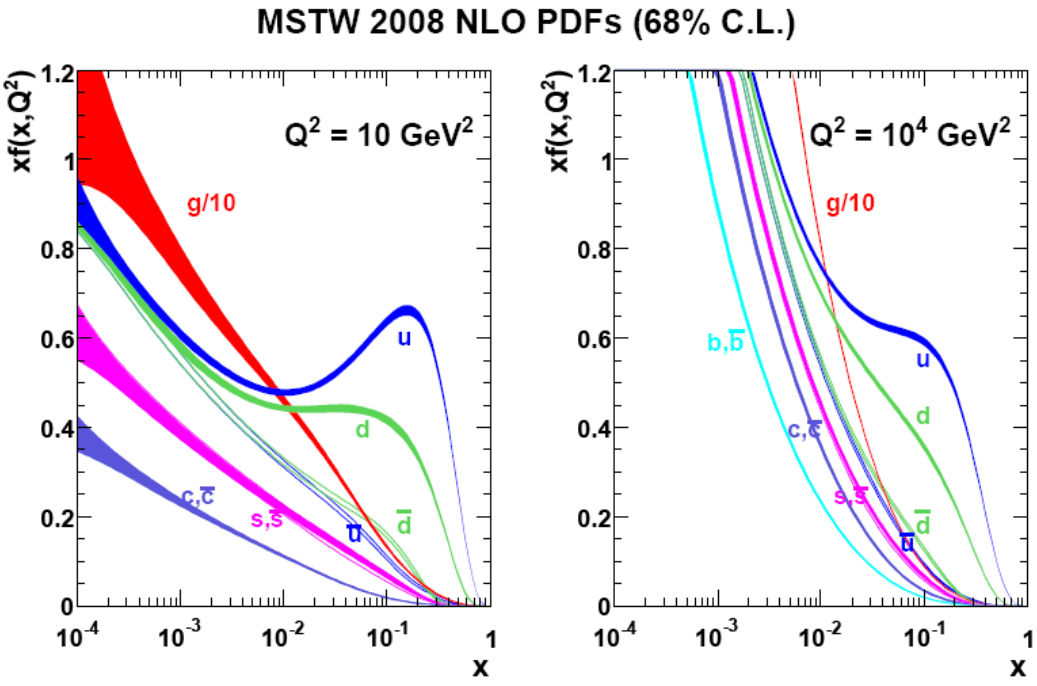


Figure 5.4: MSTW08 NLO PDFs at $Q^2 = 10\ GeV^2$ and $Q^2 = 10^2\ GeV^2$. 68% CL error bands are shown. Taken from [72].

In figure 5.6 we show the NNPDF2.0 sets compared to MSTW08 and CTEQ6.6. These next-leading order parton fits contain the same amount of experimental information. The only difference is in the different treatment of the heavy quark masses. This comparison shows that the most updated PDF sets provide a description of the proton

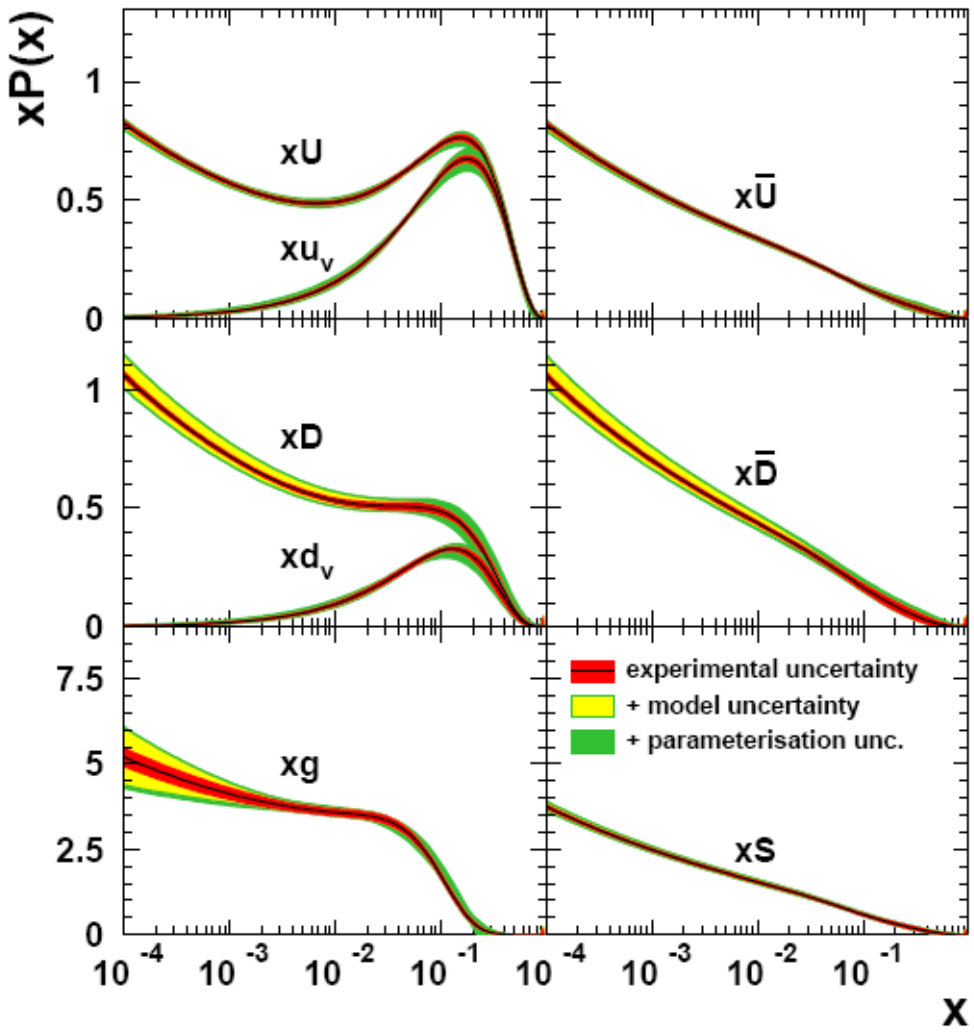


Figure 5.5: HERAPDF parton fit at $Q^2 = 4\ GeV^2$ The experimental, model and parametrisation uncertainties are shown separately. Taken from [75]

structure which is not identical but, often consistent within the uncertainties. At small-x the gluon remains basically unconstrained by data and error bands are larger. Most NNPDF2.0 uncertainties are comparable to the CTEQ6.6 and MSTW08 ones; there are however some interesting exceptions. The uncertainty on strangeness, which NNPDF2.0 parametrises with as many parameters as any other PDF [80] is rather larger than those of MSTW08 and CTEQ6.6, in which these PDFs are parametrised with a very small number of parameters. The NNPDF2.0 uncertainty on total quark singlet (which contains a sizable strange contribution) is also larger. The uncertainty on the small-x gluon is significantly larger than that found by CTEQ6.6, but comparable to that of MSTW08, which has an extra parameter to describe the small-x gluon in comparison to CTEQ6.6

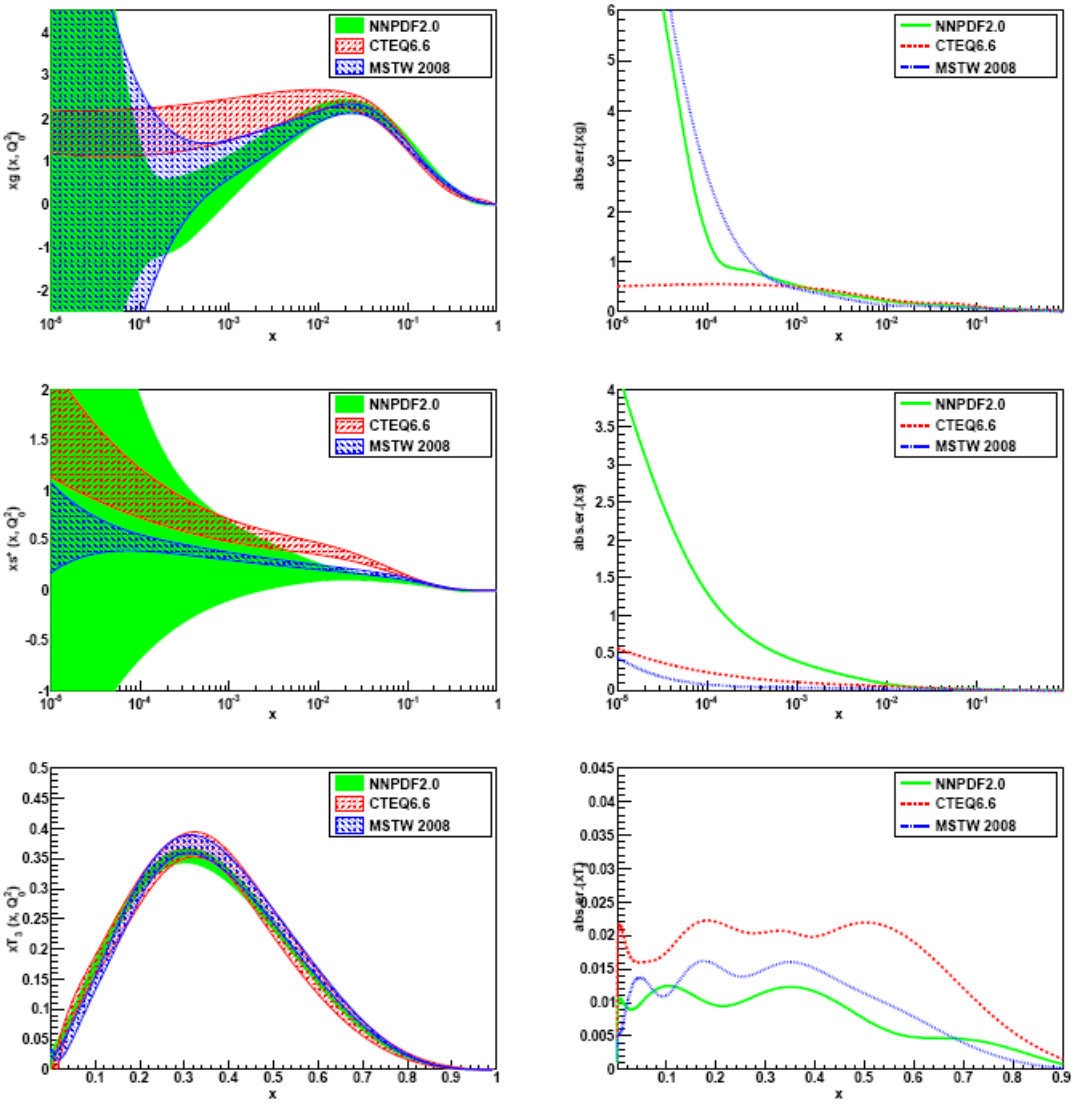
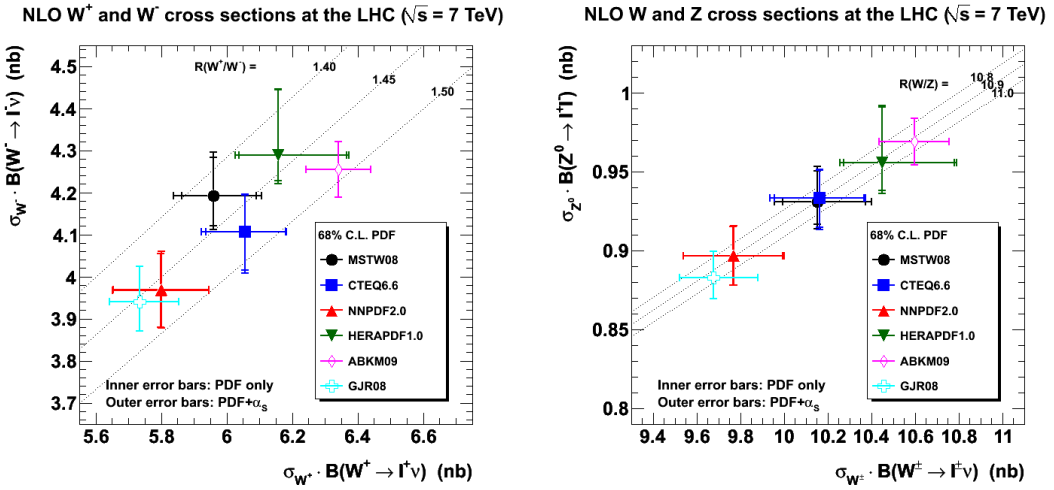


Figure 5.6: The gluon $g(x)$ (top), the total strangeness (middle) $s^+(x) =$ and the triplet (bottom) $T_3(x) = u^+(x) - d^+(x)$ (where $q^+ = q + \bar{q}$) at the initial scale $Q_0^2 = 2GeV^2$ from the NNPDF2.0 analysis. Both the PDFs (left) and their absolute uncertainty (right) are shown, compared to MSTW08 and CTEQ6.6 PDFs

[72]. The uncertainty on the triplet combination is rather smaller in NNPDF2.0 than either MSTW08 or CTEQ6.6.

5.6 Phenomenology at LHCb

The most important impact of the determination of the PDFs is the prediction of physical observables at future and present experiments involving hadronic collisions. It is clear that different determinations of PDFs would lead to different predictions; for this reason the PDF fitting community performed some benchmarking exercises. The most recent work has been performed to compare the cross-sections predictions and their uncertainties for the initial LHC run at 7 TeV. Here we only show in fig. 5.7 the results on the W and Z production cross-section. New experimental data will soon be available



(a) (b)

Figure 5.7: Comparison of Z and W production cross-section times lepton branching ratio for different PDF sets. Taken from [109, 110]

from the LHC experiments and their inclusion in the global fits will help to constrain the PDF in new kinematical regions. It is interesting to see which are the physics observables that are most sensitive to the PDFs. Candidates channels that are likely to be sensitivity to PDFs at LHC energy are: Z, W and Drell-Yan rapidity distribution, $t\bar{t}$, di-jet. A simple way of evaluating the sensitivity is to plot the uncertainty on the predicted observable due to the PDF uncertainties. Then, from an experimental point of view, it is just a matter of estimating the achievable precision on the measurement of that observable. Another way, when considering PDF sets built using the hessian method, is to calculate the contribution of each eigenvector direction to the uncertainty on the observable studied; given an observable F it can can be generated using either the best-fit value or any of the member of the set $S_k^{\pm}$. Then one can calculate the deviation for the observable F:

$$E_k = \frac{F(S_k^{\pm}) - F(S_0)}{F(S_0)} \tag{5.19}$$

where S_0 is the central value of the PDF set. E_k represent the contribution (relative to the central value) to the uncertainty on the physics observable provided by each eigenvector that, in a similar form, has been used in equation 5.13 to obtain the uncertainty on the physics observable. This representation shows immediately which eigenvector directions mainly contribute to the uncertainty on physics observable and that can be improved using new data. For example in figure 5.8 we report the quantity 5.19 where $F = d\sigma_Z/dy$ i.e. rapidity distribution for Z boson production LHC. This plot shows possible sensitivity to certain eigenvector (e.g. eigenvectors 1, 8, 15) for Z produced at high rapidities and also a total independence on some other (e.g. eigenvectors 4, 5, 16). It is here evident which is the peculiarity and importance of the LHCb experiment: since LHCb is fully equipped in the forward region ($1.9 < y < 4.9$) interactions at high rapidities can be reconstructed. To be produced at high rapidity these objects must have hard momentum component along the beam direction and this can only occur when the momenta carried by the partons are highly asymmetric [4] i.e. one soft parton and the other one which is very energetic. For this reason the measure of the Z cross-section in the forward region will probe two distinct regions of the (x, Q^2) plane. In fig. 5.9 the kinematic range for particle production at the LHC is shown. The fraction of momentum parton involved in the hard interaction and the rapidity of the particle produced can be calculated as described in equation 4.39. Using this relation it is possible to calculate which region of x can be probed by measuring the process which involve the production of a particle of mass M in a given range of rapidity. In the case of LHCb, Z and W production can probe values of $x \sim 0.2\ 10^{-4}$. For the Drell-Yan production with Q in the region $2.5 - 5\ GeV$ the lowest value of x reachable is $\sim 2\ 10^{-6}$, which is beyond the range tested at HERA. Hence due to its forward rapidity coverage and the ability to reconstruct low invariant mass muon pair systems, LHCb will cover an un-explored region of kinematical space where Parton Distribution Functions are known with less precision. This aspect can be also deduced looking at the theoretical uncertainties on the rapidity distribution for $W^{\pm}$, and Z/γ^*; The largest contribution to these uncertainties arises from PDFs uncertainties. In figure 5.10 the percentage uncertainties on the production cross-sections for $W^{\pm}$, Z and Drell-Yan process at three different mass are reported: the uncertainties due to PDF become large at high rapidity for all the channels studied (corresponding to rapidity LHCb acceptance) and for the Drell-Yan pair production, the smaller the mass, the larger the uncertainty. These considerations lead at the decision to perform a measurement of the di-muon production through the Drell-Yan mechanism that we reported in chapter 7. We now describe in detail how it is possible to see which PDFs can be constrained at LHCb measuring some electroweak physical observables. A detailed description can be found in [112].
As already stated, at LHCb the reconstructibile events are produced by asymmetric events i.e. $x_1 > x_2$; this allows some simplifications that helps to understand what is involved in these measurements. Since at high x the sea distribution vanishes it is possible to assume:

$$q_1(x_1)\bar{q}_2(x_2) + \bar{q}_1(x_1)q_2(x_2) \approx q_1(x_1)\bar{q}_2(x_2)$$

[4] Exactly as it happens at fixed target experiments

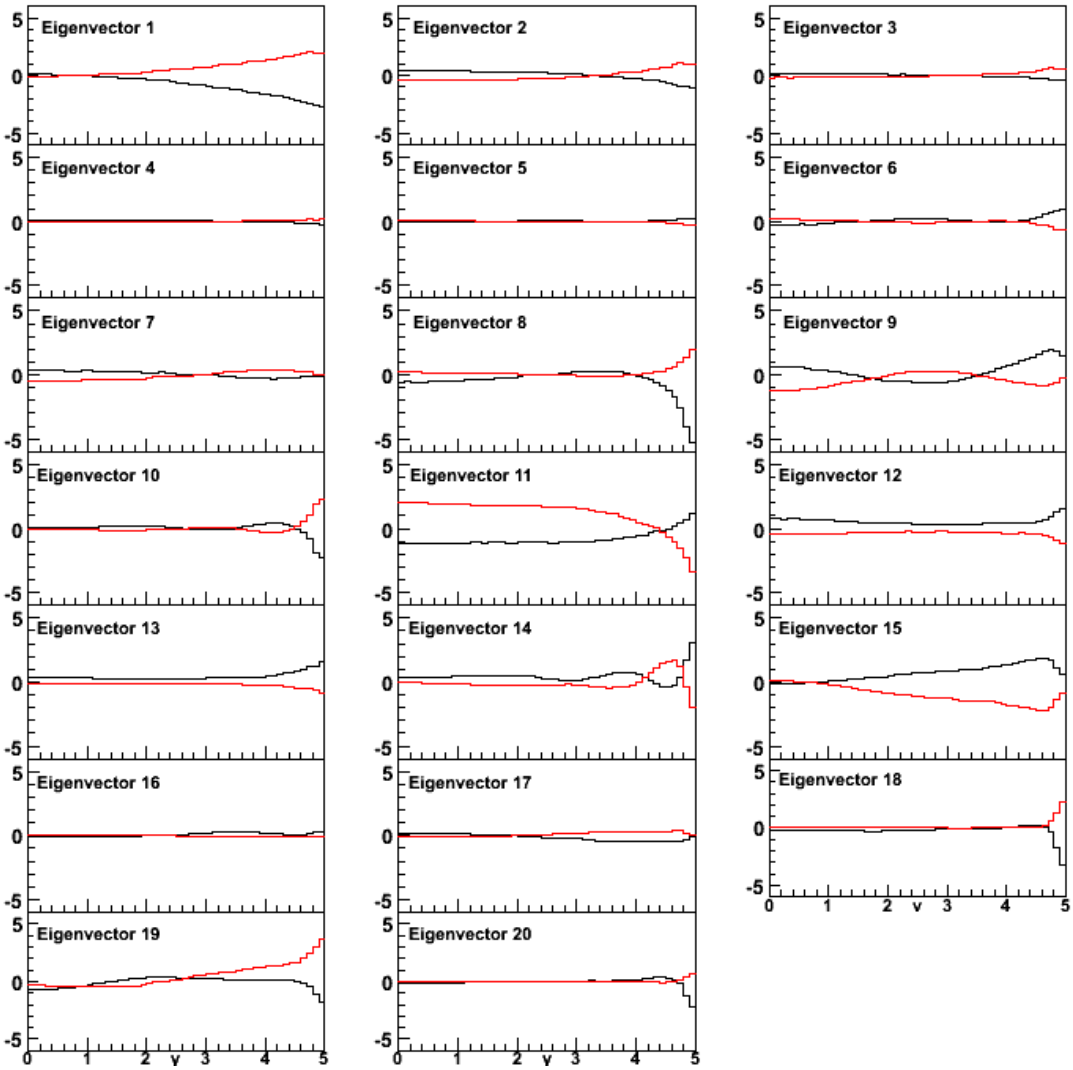


Figure 5.8: Contribution (percentage relative to the central value) as defined in equation 5.19 to the uncertainty on Z rapidity distribution at LHC for the 20 eigenvectors the parton set MSTW08. In red the element S_k^+ has been used, while in black the equivalent for the members S_k^- is plotted.

and also at small x_2 it is possible to assume that:

$$\bar{u}(x_2) = \bar{d}(x_2).$$

Given this assumption and considering that cross-sections are dominated by the up and down distribution, we now can describe some physical observables in terms of parton

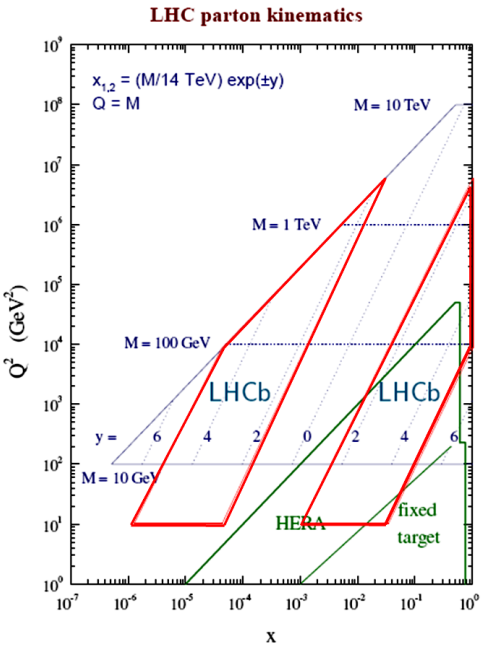


Figure 5.9: Plot of Q^2 versus x for the production of a particle of mass M. In red is highlighted the region that LHCb is able to probe. Taken from [109]

distributions. First of all the Z total production cross-sections:

$$\sigma_Z \simeq A_u u(x_1)\bar{u}(x_2) + A_d d(x_1)\bar{d}(x_2)$$

with $A_u = v_u^2 + a_u^2$ and $A_d = v_d^2 + a_d^2$ where a_i and v_i are the vector and axial couplings. Since $A_d > A_u$ there is a small enhancement for the down contribution. Here the uncertainty is dominated by the up valence distribution and so this distribution can be probed. Similarly for the Drell-Yan production

$$\sigma_{\gamma*} \simeq 4/9 u(x_1)\bar{u}(x_2) + 1/9 d(x_1)\bar{d}(x_2).$$

The uncertainty on $\sigma_{\gamma*}$ is driven by very small-x parton distributions not very well determined by HERA. In particular it is dominated by DGLAP evolution uncertainties, gluon distribution uncertainties and small-x physics described in section 4.11.
For W^+ production we find:

$$\sigma_{W^+} \simeq u(x_1)\bar{d}(x_2)$$

which probes the up valence distribution and the down sea, while for W^-

$$\sigma_{W^-} \simeq d(x_1)\bar{u}(x_2)$$

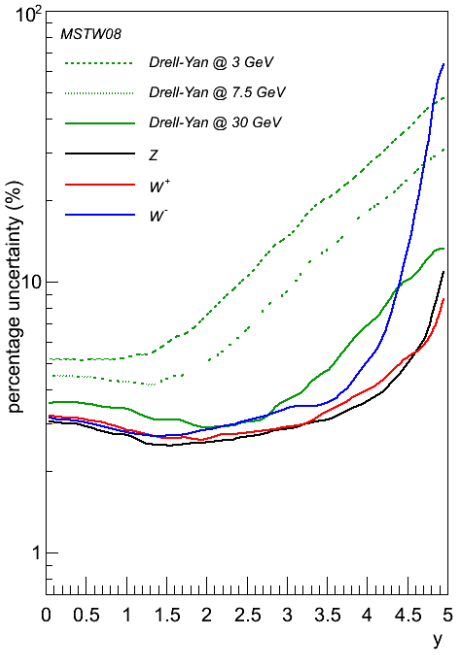


Figure 5.10: Percentage uncertainty due to PDFs on the cross-section production of Z, W and Drell-Yan as function of the boson rapidity for the parton set MSTW08. These uncertainties are calculated with the prescription defined in 5.12 using MCFM[111] with NLO calculations.

which is sensitive to the high-x d valence distribution. It also interesting to look at some of the ratio between those uncertainties; ratio are good physical observables and experimentally very well determined because systematic uncertainties often cancels, for example it is not necessary to have a precise determination of the luminosity. The significant ratios are

$$\begin{aligned} R_{Z/W} = \frac{\sigma(Z)}{\sigma(W^+) + \sigma(W^-)} &\simeq \frac{Au(\tilde{x}_1) + \bar{u}(\tilde{x}_2) + Bd(\tilde{x}_1) + \bar{d}(\tilde{x}_2)}{u(x_1)\bar{d}(x_2) + d(x_1)\bar{u}(x_2)} \\ &\simeq \frac{Au(\tilde{x}_1) + Bd(\tilde{x}_1)}{u(x_1) + d(x_1)} \end{aligned} \tag{5.20}$$

$\tilde{x}$ refers to the fraction of momentum carried by the partons for the production of the Z which are larger than those for W production. Another significant ratio is the W charge

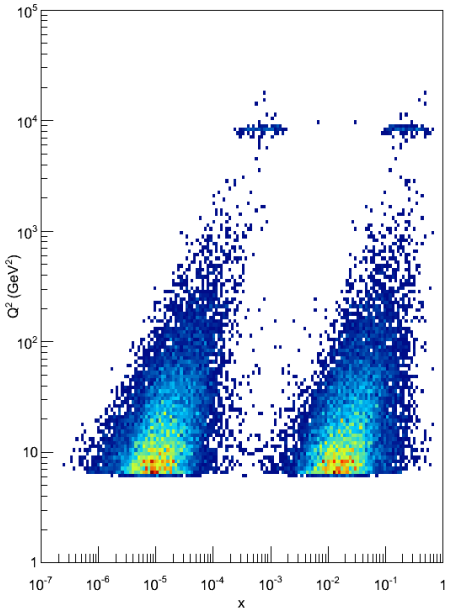


Figure 5.11: (Q^2, x) values for $Z/\gamma^* \to \mu\mu$ events fully reconstructible in LHCb. This plot has been obtained with PYTHIA.

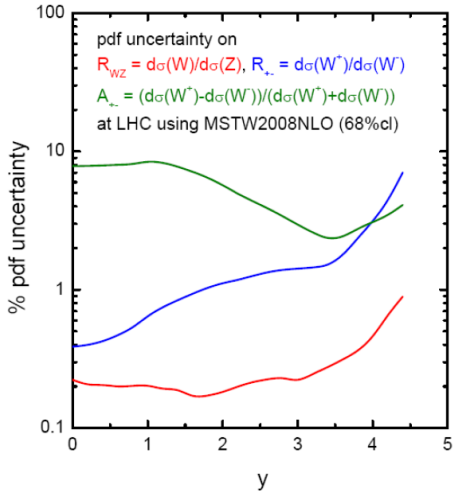


Figure 5.12: Percentage uncertainty due to PDFs on the cross-section ratios function of the boson rapidity for the parton set MSTW08. These uncertainties are calculated with the prescription defined in 5.12 using MCFM[111] with NLO calculations.

asymmetry:

$$A_{\pm} = \frac{\sigma(W^+) - \sigma(W^-)}{\sigma(W^+) + \sigma(W^-)} \simeq \frac{u(x_1)\bar{u}(x_2) - d(x_1)\bar{d}(x_2)}{u(x_1)\bar{u}(x_2) + d(x_1)\bar{d}(x_2)} \simeq \frac{u_V(x_1) - d_V(x_1)}{u(x_1) + d(x_1)} \tag{5.21}$$

that is a good test for the valence quarks.
The last significant ratio is:

$$R_{\pm} = \frac{\sigma(W^-)}{\sigma(W^+)} \simeq \frac{d(x_1)\bar{d}(x_2)}{u(x_1)\bar{d}(x_2)} \simeq \frac{d(x_1)}{u(x_1)} \tag{5.22}$$

In fig. 5.12 are reported the uncertainties on those ratios; The ratio $R_{Z/W}$ is very well predicted and uncertainty is rather small in the whole rapidity range and thus constitutes a precise electroweak prediction for the LHC. The uncertainty on $R_{\pm}$ increases strongly at high rapidity thus LHCb is in a unique position to make a precise measurement of this observable and to constrain the d/u ratio at large x. Finally, the charge asymmetry $A_{\pm}$ is the observable which is predicted with least precision and it is not strongly rapidity dependent. A measurement at LHC constrains the difference in the u and d valence quarks.

5.7 Pseudo-experiment generation

In the previous sections we described how PDFs are extracted from experimental data, how uncertainties are estimated and which is the impact on some significant physical observables. We expect that the measurements of physics observable at LHC agree with the theoretical predictions at least within the theoretical and experimental uncertainties; however, in the perspective of predicting the impact of new data to our knowledge of the PDFs, it is important to describe the physics observable according to the probability distribution predicted by the PDF set. In this way, generating several pseudo-experiments we are able to describe the outcome of repeated measurements of an observable. Here, for "observable" we mean for example the total cross-section of a given process or the cross-section at a given rapidity or an asymmetry as described in section 5.6. The generated pseudo-experiments can be used to test any kind of procedure which has the purpose to asses the improvement on PDF knowledge given the new data. Now we describe how it is possible to generate pseudo-experiments and, once again, we need give a different description when using PDF sets built with the hessian method and the NNPDF set.

Hessian PDF sets
In section 5.3 we described how uncertainties are evaluated and projected in an orthogonal basis. This description allows the uncertainties arising from each eigenvector direction to be dealt with as uncorrelated uncertainties (see eq. 5.12). If one wants to generate a physics observable $\mathcal{O}$ according to the probability distribution predicted by a PDF set, one has to describe this observable as a linear combination of observables generated with the central value of the PDF and the observable generated when moving

each eigenvector of 1σ (as defined in the set $S_k^{\pm}$):

$$\mathcal{O} = \mathcal{O}(S_0) + \sum_{k=1}^{N_e} R_k \cdot \left[\mathcal{O}(S_k^{\pm}) - \mathcal{O}(S_0)\right] \tag{5.23}$$

where N_e is the number of eigenvector directions, $\mathcal{O}(S_0)$ is the observable generated with the PDF central values and $\mathcal{O}(S_k^{\pm})$ is the observable generated when moving the k^{th} eigenvector of 1σ. R_k are a set of N_e random numbers sampled from a Gaussian distribution centered on 0 and with width equal to 1. To produce a different pseudo-experiment it is sufficient to obtain a new set of random numbers R_k. The procedure that we described, practically, is a sampling of the probability density of the PDFs in the space of the eigenvectors. In figure 5.13(a) we show 4 different pseudo-experiments in which we generated the Z rapidity distribution with the PDF set MSTW08.

NNPDF set

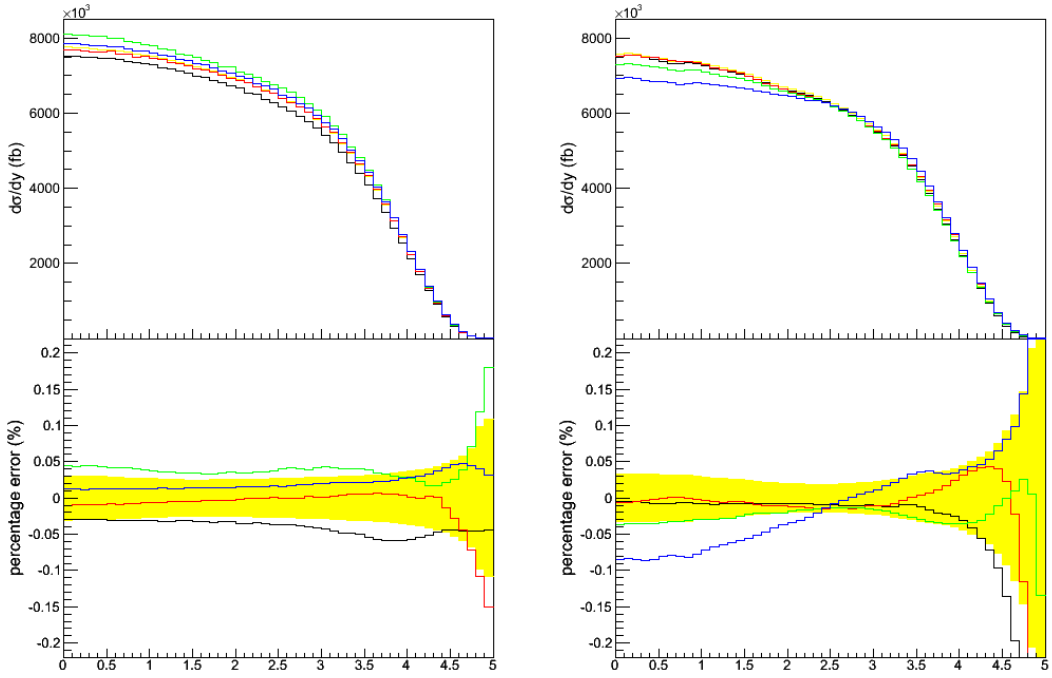
When using the PDF set NNPDF to generate a set of repeated pseudo-experiment of a single observable the procedure is simple. It is only required to sample from a gaussian distribution around the mean value predicted with a spread equal to the uncertainty. Given an observable with a central value $\langle\mathcal{O}(f)\rangle$ that is calculated with the equation 5.16, and the uncertainty $\sigma_{\mathcal{O}}$ as in 5.17 the pseudo-experiment can be obtained as:

$$\mathcal{O} = \langle\mathcal{O}(f)\rangle + \tilde{R} \cdot \sigma_{\mathcal{O}} \tag{5.24}$$

where $\tilde{R}$ is a unit gaussian random number. In the case of multiple observables the procedure is more sophisticated because we need to take into account the correlations; first of all it is necessary to build the covariance (or correlation matrix) of the variables we want to study using the equation 5.18 to calculate each term of the matrix. Given N_o the number of observables, we used this $N_o \times N_o$ matrix to generate a set of N_o correlated random numbers $R_c = \{R_c^0, R_c^1, ..., R_c^{N_o}\}$ starting from an uncorrelated unit gaussian distributed set of random numbers R using the Cholesky-decomposed matrix U of the correlation matrix doing $R_c = RU$. The i^{th} observable of a given pseudo-experiment is now described as:

$$\mathcal{O}_i = \langle\mathcal{O}(f)\rangle + R_c^i \cdot \sigma_{\mathcal{O}} \tag{5.25}$$

To generate a different pseudo-experiment one needs to repeat this procedure changing the set of uncorrelated random numbers R. In figure 5.13(b) we show 4 different pseudo-experiments in which we generated the Z rapidity distribution with the PDF set NNPDF20. In this case, given that we have 50 rapidity bins, we used a 50×50 correlation matrix, obtained using the equation 5.18, that describes the correlations between the rapidity bins.



(a) (b)

Figure 5.13: Example of four different pseudo-experiments. We generated Z rapidity distribution (a) for the NNPDF method, (b) for the hessian method using MSTW08 PDF set. On bottom percentage deviation of each pseudo-experiment from the central value and (in yellow) 1σ uncertainty are plotted.

6. THE FITTING ALGORITHM

In this chapter we report on the analysis of parton distribution functions. In section 6.1 and 6.2 we describe the development of a fitting algorithm to measure the luminosity at LHCb and constrain PDFs. This fitting algorithm has two different implementations according to the PDF set that was considered. In section 6.3 we describe how we validated our algorithm using pseudo-experiments. Finally in section 6.4 we present expected results for luminosity determination and for PDF constraints with future LHC data.

6.1 PDFs built using the Hessian method

As reported in section 5.3, PDF sets obtained using the hessian method are described using an ensemble of independent eigenvectors. The best estimation for a physical observable, for example the rapidity distribution, $f(y)$, can be generated using the central value of the PDF and a 1 σ uncertainty (due to PDF) can be evaluated using equation 5.12. If we want to the describe the probability distribution for this observable by the model, we first calculate the linear combination of the observable obtained using the PDF best-fit, $f_0(y)$, and the observable $f_i(y)$ obtained by moving the i^{th} eigenvector by one-sigma:

$$f(y) = f_0(y) + \sum_{i=1}^{N} \left[f_0(y) - f_i(y) \right] \lambda_i. \tag{6.1}$$

$f_i(y)$ is obtained using the error set $S_i^{\pm}$ as defined in section 5.3, which practically means that one needs to generate the rapidity distribution using as underlying PDF the i-th element of the set $S_i^{\pm}$ and the PDF central value. Then, the one-sigma contour is mapped out by sampling λ_i from a unit Gaussian with mean zero.

Equation 6.1 also provides a form with which data can be fitted. Given binned data with N_j events in the j^{th} bin with an uncertainty δ_j, the normalization and shape of the

distribution can be fitted by minimizing

$$\chi^2(\lambda_0, \lambda_i) = \sum_{j}^{\#bins} \left[\frac{N_j - \lambda_0 \left(f_0^j + \sum_{i=1}^{N} \lambda_i (f_0^j - f_i^j) \right)}{\delta_j} \right]^2 + \sum_{i>0} \lambda_i^2 \tag{6.2}$$

where λ_0 is the overall normalization or luminosity and λ_i are the fitted eigenvalues. The definition 6.2 compares the number of events N_j in a given bin and the number of events predicted by the definition 6.1 for that bin. The fit extracts the PDF eigenvalues and the luminosity λ_0 from the shape of the rapidity distribution. In this fit there are $N+1$ free parameters but there are $\#bins - 1$ degrees of freedom because the N PDF parameters are constrained by the extra term $\sum_{i>0} \lambda_i^2$ in equation 6.2 which constrain the values of λ_i to be distributed according to a Gaussian probability distribution centered on zero and with unit width.

6.1.1 Luminosity Measurement

The parameter λ_0 represents the normalization of the histogram and it can be interpreted as the integrated luminosity of the data sample analyzed. Therefore, in addition to constraining the PDF, this analysis provides a novel method to use a process, for example Z and W bosons production, as standard candles at the LHC in order to monitor the luminosity and give an absolute benchmark for other cross-section measurements.

6.1.2 PDF constraint

The eigenvalues, by definition of the model, have an uncertainty equal to 1 but it is possible that after this fitting operation the uncertainty on the eigenvalues can be smaller than 1. This implies that we have a better constraint of the PDF. The Hessian method provides a diagonalized covariance matrix of the parameters with unitarian uncertainty (i.e. the covariance matrix is equal to the Identity Matrix). The output of our algorithm is a set of N eigenvalues and a $N \times N$ covariance matrix V of the parameters in which values on the main diagonal are smaller than 1 and terms off-diagonal can be different from zero. To evaluate the PDF constraint it is possible to calculate the uncertainties of the parton set given this new covariance matrix. Using the usual equation 5.5 for propagation of uncertainties the new uncertainties on PDFs are calculated as

$$\delta F = \sqrt{\sum_{ij} \frac{\partial F}{\partial \lambda_i} V_{ij} \frac{\partial F}{\partial \lambda_j}} \tag{6.3}$$

where λ_i are the fitted parameters and F is a PDF (e.g. u_V, d_V , *sea* , *gluon* etc.). The variations of F with respect to the parameters $\frac{\partial F}{\partial \lambda_i}$ are exactly:

$$\frac{\partial F}{\partial \lambda_i} = S_i^{\pm}. \tag{6.4}$$

6.2 PDFs built using a Monte Carlo Method

A different method can be used with any PDF set which uses a Monte Carlo Method to sample the probability distribution of the PDFs and which is distributed with a set of Monte Carlo replicas of the PDF set. As described in 5.4, NNPDF uses a neural network to extract the PDFs and a Monte Carlo estimate of the uncertainties; This technique produces a set of equally likely Monte Carlo replicas of the PDFs. The implementation of the fitting algorithm for this kind of PDF set is twofold: a description of the extraction of luminosity is given in section 6.2.1 while in section 6.2.2 we present how it is possible to constrain the PDF with a re-weighting of the replicas.

6.2.1 Luminosity Measurement

Given a set of LHCb data binned in M rapidity bins we estimate the luminosity by minimizing the χ^2 with respect to the parameter λ_0:

$$\chi^2 = a^T U a = \sum_{ij}^{M} (N_i - \lambda_0 f_i^0) U_{ij}^{-1} (N_j - \lambda_0 f_j^0) \tag{6.5}$$

where N_k is the content of the k^{th} rapidity bin, f_k^0 is the central value of the ensemble of replicas (i.e. the theory prediction) for the k^{th} bin and U_{ij} is the matrix element of the covariance matrix. The matrix U is obtained as the sum $U = C + V$, where C is the covariance matrix whose elements are obtained with the prescription of equation 5.18 and describes the correlations between the contents of the rapidity bins predicted by the model. In this case, it is a $M \times M$ matrix. V is the covariance matrix arising from experimental uncertainty. The luminosity is the multiplicative factor λ_0 before the predicted cross-section f_k^0 in that given rapidity bin.

6.2.2 PDF constraint

To constrain the PDF we developed a method described for the first time in [113]. Given a set of Monte Carlo replicas of the PDF set obtained from a global fit it is possible to

estimate the impact of a set of N_{obs} new measured data points

$$\{x\} = x_1, x_2, ..., x_{N_{obs}}.$$

using the method of statistical inference and update the initial probability density $\mathcal{P}_{old}(f)$ by taking into account the new data and obtaining an improved probability density $\mathcal{P}_{new}(f)$.

By the probability $\mathcal{P}_{old}(f)$ for the PDF f, what we actually mean is the probability $\mathcal{P}(f|K)$, where K denotes all the data used in the determination, their associated errors and the theoretical assumptions made to perform the global fit. If we then wish to extend the dataset by including new data x, the new probability $\mathcal{P}_{new}(f)$ is then $\mathcal{P}(f|xK)$. The new probability is then determined from the old probability using the so called sampling distribution $\mathcal{P}(x|fK)$ and multiplicative rule for probabilities, also known as Bayes theorem. This theorem states that, given $P(A)$ the probability for the event A, $P(B)$ the probability for B and $P(A|B)$ the conditional probability that A happens given B, it holds that

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}. \tag{6.6}$$

The theorem can be extended to more than two events. For three events we have:

$$P(AB|C) = P(A|BC)P(B|C) = P(B|AC)P(A|C). \tag{6.7}$$

Applying this theorem in our case we obtain

$$\mathcal{P}(f|xK)\mathcal{P}(x|K) = \mathcal{P}(x|fK)\mathcal{P}(f|K), \tag{6.8}$$

and hence

$$\mathcal{P}(f|xK) = \mathcal{P}(x|fK)\mathcal{P}(f|K)/\mathcal{P}(x|K), \tag{6.9}$$

Note that $\mathcal{P}(x|K)$ does not depend on the PDF and can thus be determined simply by insisting that $\mathcal{P}(f|xK)$ is properly normalized. We replace $\mathcal{P}(f|K)$ with $\mathcal{P}(f|K)df$, $\mathcal{P}(f|xK)$ with $\mathcal{P}(f|xK)df$) [1] and we calculate the normalization as the integral over the PDF space.

$$\mathcal{P}(x|K) = \int \mathcal{P}(x|fK)\mathcal{P}(f|K)df \tag{6.10}$$

Now we can rewrite the probability $\mathcal{P}(f|xK)$:

$$\mathcal{P}_{new} = \mathcal{P}(f|xK) = \mathcal{P}(x|fK)\mathcal{P}(f|K) \Big/ \int \mathcal{P}(x|fK)\mathcal{P}(f|K)df \tag{6.11}$$

[1] here the notation df represent the infinitesimal element of the space of the PDF.

Bayes theorem works only in the case of discrete variables but our datapoints are continuous variables. The calculation has to be performed within the multi-dimensional probability density $\mathcal{P}(x|fK)d^n x$, in the space of the data points in a limit in which the volume element $d^n x$ goes to zero. The derivation of the normalization term

$$\mathcal{N}_\chi = 1/\int \mathcal{P}(x|fK)\mathcal{P}(f|K)df$$

can be found in [114].

To apply this method we need to know the relative probabilities of the new data for different choices of PDF. Since the new data are assumed to have Gaussian errors, these probabilities will be directly proportional to the probability density of the χ^2 of the new data, given the PDFs f [115]

$$\mathcal{P}(x|fK) \propto \mathcal{P}(\chi^2|f) \propto \chi^2(x,f)^{n/2-1}e^{-\frac{1}{2}\chi^2(x,f)} \tag{6.12}$$

where in general

$$\chi^2(x,f) = \sum_{i,j}(x_i - X_i[f])C_{ij}^{-1}(x_j - X_j[f]) \tag{6.13}$$

and $X_i[f]$ is the value predicted for the data x_i using the PDF f. C is the experimental covariance matrix. The probability $\mathcal{P}_{new}$ in equation 6.11 becomes now

$$\mathcal{P}_{new} = \mathcal{P}(\chi^2|f)\mathcal{P}_{old}(f)\mathcal{N}_\chi \tag{6.14}$$

where the probability $\mathcal{P}_{old}(f)$ is defined by the ensemble of the Monte Carlo replicas. If we multiply on both sides by an observable $\mathcal{O}(f)$ and we integrate over the PDF space we obtain a new description for the best estimate of the observable:

$$\begin{aligned}
\langle\mathcal{O}\rangle_{new} &= \int \mathcal{O}(f)\mathcal{P}_{new}(f)df \\
&= \int \mathcal{N}_\chi\mathcal{O}(f)\mathcal{P}(\chi^2|f)\mathcal{P}_{old}(f)df \\
&= \frac{1}{N}\sum_{k=1}^{N_{rep}} \mathcal{N}_\chi\mathcal{P}(\chi^2|f_k)\mathcal{O}(f_k) \\
&= \frac{1}{N}\sum_{k=1}^{N_{rep}} w_k\mathcal{O}(f_k)
\end{aligned} \tag{6.15}$$

in which the probability $\mathcal{P}(\chi^2|f_k) = w_k$ can be seen as a weight:

$$w_k = \mathcal{N}_\chi \mathcal{P}(\chi^2|f_k) = \mathcal{N}_\chi \chi^2(x, f_k)^{n/2-1} e^{-\frac{1}{2}\chi^2(x,f_k)}. \tag{6.16}$$

The normalization factor $\mathcal{N}_\chi$ calculated in [114] is:

$$1/\mathcal{N}_\chi = \frac{1}{N} \sum_{i=0}^{N_{rep}} \chi_i^2(x, f)^{n/2-1} e^{-\frac{1}{2}\chi_i^2(x,f)}. \tag{6.17}$$

and hence, it also holds that

$$N = \sum_{k=1}^{N_{rep}} w_k$$

The weights w_k, when divided by N, are then simply the probabilities of the replicas f_k given the χ^2 of the new data.

Practically, to obtain these weights we made the following steps: First we generated the rapidity distributions with each Monte Carlo replica of the PDF set. Second, given a set of data binned in M rapidity bins, we compared the prediction of a single replica with the data calculating the χ^2.

$$\chi_k^2 = \sum_{bin=1}^{M} \left(\frac{N_{bin} - F_k(y)}{\Delta_{bin}} \right)^2 \tag{6.18}$$

where k is the index of the replica, N_{bin} is the experiment result in that rapidity-bin and $F_k(y)$ is the value of the rapidity distribution generated with the k^{th} replica. Finally we associate a χ^2 to each replica of the PDFs. The probability for each replica, given the new data, can be evaluated as the weight defined in equation 6.16. It is now possible to compute the central values and uncertainties of the PDF $\mathcal{F}$ using a weighted definition: for the central value,

$$\langle \mathcal{F} \rangle = \frac{1}{N} \sum_{k=1}^{N_{rep}} w_k \mathcal{F}(f_k), \tag{6.19}$$

and for the uncertainties

$$\sigma_{\mathcal{F}}^2 = \frac{1}{N-1} \sum_{k=1}^{N_{rep}} w_k (\mathcal{F}(f_k) - \langle \mathcal{F}(f_k) \rangle)^2. \tag{6.20}$$

6.3 Test and Validation

To prove the validity of these methods we performed some statistical tests using LHCb-like pseudo-experiments. We generated pseudo-experiments with energy of centre of mass collision of 7 TeV and values of integrated luminosity of 0.1 fb^{-1}, 1 fb^{-1} and 10 fb^{-1}. For this study we considered the rapidity distribution that can be measured by LHCb of the processes:

$$\begin{aligned} Z &\rightarrow \mu^+\mu^- \\ W^+ &\rightarrow \mu^+\nu_\mu \\ W^- &\rightarrow \mu^-\bar{\nu}_\mu \\ \gamma^* &\rightarrow \mu^+\mu^- \end{aligned}$$

For the dimuon Drell-Yan production four different invariant mass bins have been con-

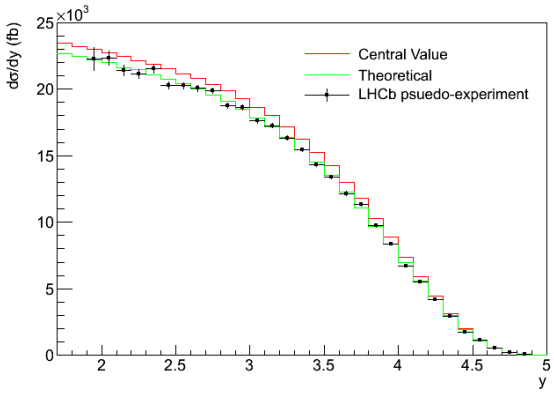


Figure 6.1: Simulated measurement of the Z rapidity distribution at LHCb. In red the distribution predicted using the PDF central value, in green a pseudo-experiment generated as described in section 5.7. The dots are the LHCb reconstructed data which are rescaled by the acceptance in figure 6.2 and by the efficiencies described in section 6.3.1.

sidered: $2.5-5$ GeV, $5-10$ GeV, $10-20$ GeV and $20-40$ GeV. Pseudo-experiment generation had to take into account two main conditions: firstly, the physics observables have to be generated according to the probability distribution predicted by the models. This condition is accomplished by using the procedure described in section 5.7. The second condition is the faithful representation of the LHCb experiment performance in the measurement of the physics observables with a reasonable determination of the experimental uncertainties. The main features we need to consider to achieve this description are: geometrical acceptance, reconstruction efficiency, trigger and selection efficiencies. Given these efficiencies, we are able to describe the statistical uncertainties in each rapidity bin and to simulate the Poissonian fluctuations. For example, assuming for a given integrated luminosity that we expect to reconstruct and record at LHCb a number

N_Z of Z bosons at a given rapidity y, with efficiency ε_Z, the total number of events of that kind produced by the collisions would be

$$\frac{N_Z(y)}{\varepsilon_Z}$$

and the statistical uncertainty on this corrected number of events is

$$\frac{\sqrt{N_Z(y)}}{\varepsilon_Z}.$$

The effects of background contamination has not been included in these distributions. An example of a generated rapidity distribution is shown in Figure 6.1. We now describe briefly how the efficiencies for the different channels have been estimated. Note that this is only a summary of several studies performed within the LHCb collaboration. We have used the estimated efficiencies of these analyses to derive our results.

6.3.1 $Z \rightarrow \mu^+\mu^-$ production

The signature of the production of a Z boson decaying into two muons is very distinctive; we expect to reconstruct two muons of opposite charge with high transverse momentum ($p_T \sim 45\ GeV$) coming from the primary vertex and the invariant mass of the dimuon system near the value of the Z mass ($\sim 91\ GeV$). The accepted mass the dimuon system is in the range $81 < M < 101\ GeV$. A complete description of the analysis can be found in [116].

Geometrical Acceptance: We calculated the geometric acceptance using MCFM[111] with a NLO calculation and NLO PDF set. We generated a sample of $Z \rightarrow \mu\mu$ decay and for each rapidity value of the Z boson, we calculated the fraction of events with two muons inside LHCb acceptance ($1.9 < \eta < 4.9$). The acceptance function is shown in figure 6.2. The differential cross-section measured is corrected bin by bin dividing by the acceptance function in figure 6.2.

Reconstruction and Trigger Efficiency: The reconstruction efficiency is defined to be the fraction of events within the LHCb geometric acceptance that can be reconstructed offline. It depends on the efficiency for reconstructing the track, the efficiency for reconstructing the required number of hits in the muon chambers and matching them to this track and on the efficiency of any muon identification criteria. All these efficiencies are calculated separately for the two muons and then are multiplied together to get the efficiency for the muon pair. A detailed description of the procedure can be found in [117]. Using the official LHCb simulation, described in 2.3.9, a sample of Z events inside LHCb has been taken and the fraction that passes the reconstruction requirements has been evaluated. This efficiency has been estimated to be 92 % [116]. Similarly, the total trigger efficiency is defined from the combination of the conditional probabilities that an

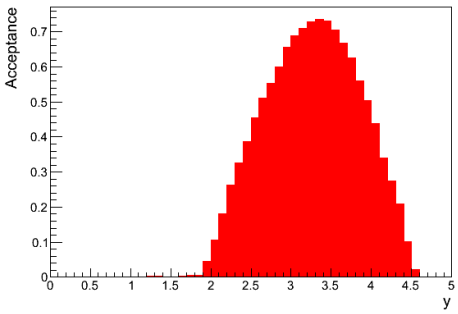


Figure 6.2: Acceptance as function of the Z boson rapidity

event passes each trigger stage. In the same way this efficiency has been estimated to be 90 %.

Selection Efficiency: To select a pure sample of Z and to reduce the background contribution other cuts are necessary.It is required that highest transverse momentum must satisfy $p_T > 20\ GeV$ and that the lowest transverse momentum $p_T > 20\ GeV$. Furthermore, it is required that the impact parameter significance is smaller than 5 to be sure that the muons arise from the primary vertex. The signal acceptance for these offline kinematic cuts is defined to be the fraction of offline reconstructed Z events passing all of the trigger stages that also satisfy these kinematic requirements. This efficiency has been estimated to be ~ 90 %.

6.3.2 $W \rightarrow \mu\nu$ production

The signature for a $W \rightarrow \mu\nu$ decay consists of a single high p_T isolated muon and little other activity in the event. Given that we are able to reconstruct only the muon track, the cross-section will be measured as a function of the muon pseudo-rapidity and only in the LHCb rapidity coverage; thus we don't need to calculate a geometric acceptance. A complete description of the analysis can be found in [118].

Reconstruction and Trigger Efficiency: The definition and the methodology used to calculate the efficiency for $W \rightarrow \mu\nu$ is broadly the same that was used for $Z \rightarrow \mu^+\mu^-$. The only difference arises from the fact that one muon is required in the computation of the efficiencies. The reconstruction efficiency has been estimated to be ~ 96 % while it has been calculated $\sim$ 91% efficiency for the trigger system.

Selection Efficiency: To select these events a hard cut on the muon transverse momentum is applied ($p_T > 20\ GeV$). To suppress backgrounds an additional cut is imposed on the transverse momentum asymmetry:

$$A_{p_T} = \frac{p_T - p_T^{rest}}{p_T + p_T^{rest}}$$

where p_T^{rest} is the transverse momentum of the vector sum of all the tracks excluding the muon track. A cut on this variable is set to be $A_{p_T} > 0.65$ which together with the cut on the transverse momentum leads to an efficiency of ~ 35 %.

6.3.3 Low mass Drell-Yan production

A detailed description of the analysis will be given in chapter 7. Here we show in table 6.1 the overall efficiency we considered to generate the pseudo experiments that were obtained in a previous study [119].

Mass Range (GeV)	Efficiency
$2.5 - 5$	0.39
$5 - 10$	0.75
$10 - 20$	0.80
$20 - 40$	0.80

Table 6.1: Overall efficiency for the Drell-Yan production in four different mass ranges taken from [119].

6.3.4 Statistical Consistency

To be sure that our fitting algorithm is statistically correct, we performed some tests calculating significant statistical quantities that can show weakness or inefficiency of the procedure. We tested the two different implementations described in equation 6.1 and section 6.2.1. Our algorithms extract one or more parameters from the data, so we generated pseudo-data sets with a known set of parameters[2], ran the fitting algorithm and evaluate whether the algorithm had reproduced correctly the parameters used in the generation. We generated an ensemble of 1000 LHCb-like pseudo-experiments following the probability distribution in the space of the PDF, as described in section 5.7, the experimental probability distribution using the efficiency summarized in the sections 6.3.1, 6.3.2 and 6.3.3. At generation stage, for both Hessian PDF sets and NNPDF, the parameter λ_0, that corresponds to the luminosity, has been set to a fixed value.We ran the algorithm on each pseudo-experiment and we recorded the parameters extracted, the uncertainties on the determination of those parameters and the χ^2 of the fit.

[2]Only the luminosity parameter in the case of the implementation given in 6.2.1 is used and luminosity and PDF parameters for the implementation described in section 6.1 are used

χ^2 distribution

If the statistical procedure is correct, we expect that the χ^2 of the fit on the 1000 pseudo-experiments are distributed according to the χ^2 distribution.

$$f(\chi^2, N) = (\chi^2)^{N/2-1} e^{-\frac{1}{2}\chi^2} \tag{6.21}$$

where N is the number of degrees of freedom. If this is not the case, it means that the algorithm is not able to reproduce correctly the shape of the rapidity distribution with the function provided. In figure 6.3 we show an example of the χ^2 distribution when fitting 1000 pseudo-experiments and, overlaid, the expected χ^2 distribution. Since the expected and observed distributions agree, there are no anomalies in the fitting procedure.

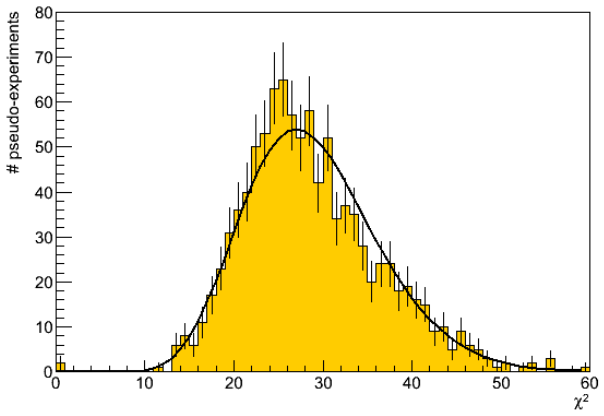


Figure 6.3: χ^2 distribution when running the algorithm on 1000 pseudo-experiment of Z rapidity distribution with the MSTW08 model. The distribution has a mean around 29 which is consistent with the number of the degrees of freedom of the fit, namely 30 rapidity bins minus 1 free parameter. The black solid line superimposed is the χ^2 distribution for 29 degrees of freedom.

Pull Functions

An excellent check of the statistical coherence of the procedure is the evaluation of the pull distributions. For each parameter we calculated:

$$P_k = \frac{\lambda_k^{fit} - \lambda_k^{gen}}{\sigma_\lambda} \tag{6.22}$$

where λ_k^{fit} is the parameter obtained with the fit algorithm, σ_λ the error on the fitted parameter and λ_k^{gen} is the value of the parameter used to generate the pseudo-experiment. For a correct statistical procedure, plotting P_k for each pseudo-experiment, we expect to obtain a gaussian distribution centered on 0 with width equal to 1. Any deviation from this shape points out a weakness or a bias in the algorithm. If the mean value of the pull distributions is significantly larger (smaller) than zero it means that the algorithm is overestimating (underestimating) the value of the parameter. If the width of the pull distributions is larger (smaller) than one, the algorithm is underestimating (overestimating) the uncertainty on the fitted parameter.As can be seen in Figure 6.4, where we show the pull function for the 20 estimated PDF parameters, no bias on the parameter determination is found, nor can a wrong estimation of the uncertainty be spotted.

We performed these tests on the rapidity distributions of the physics channels described in sections 6.3.1,6.3.2 and 6.3.3. The pull distributions and the χ^2 distributions don't show any deviations from the expected behaviour. This confirms that the fitting algorithms we have described are statistically-sound.

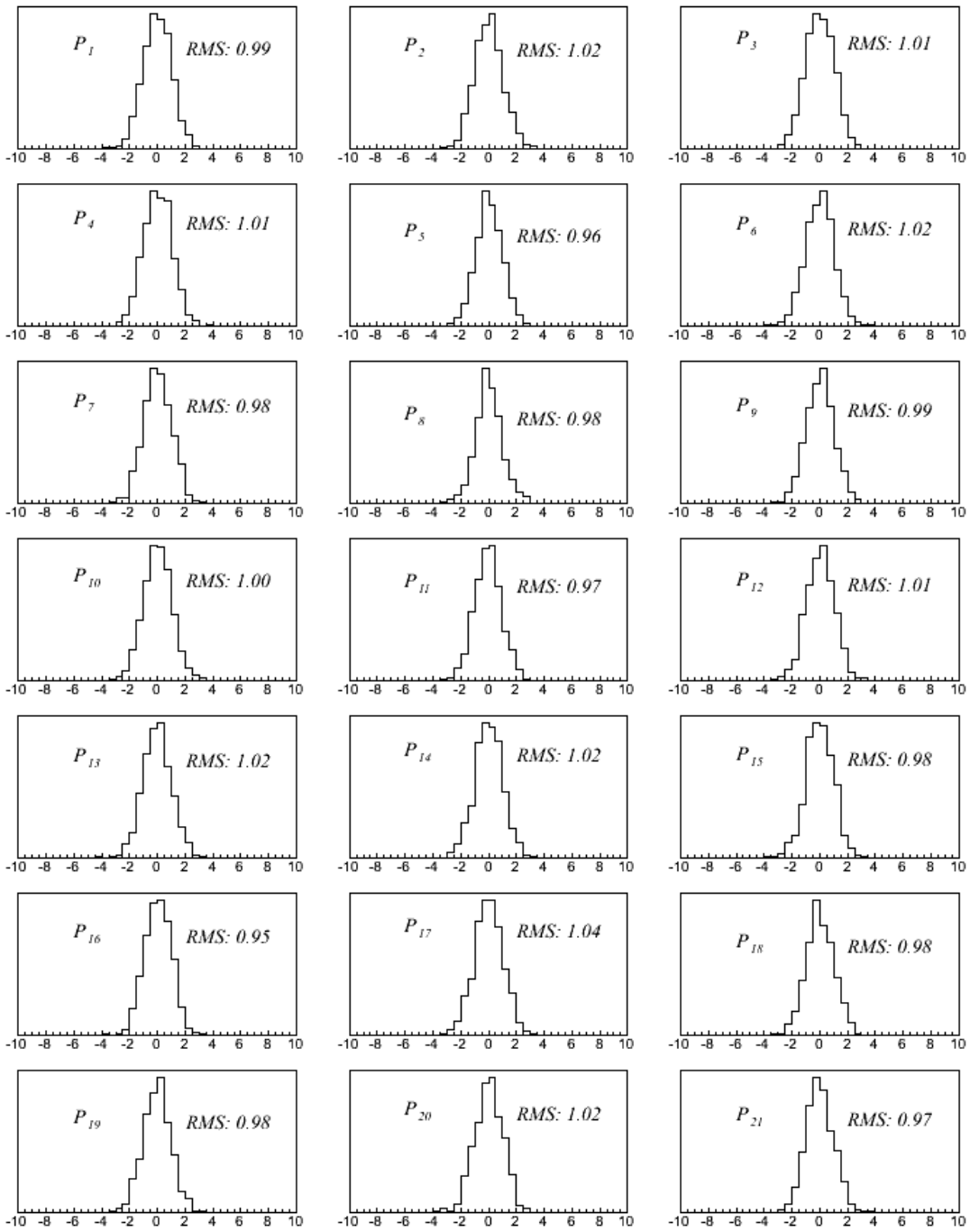


Figure 6.4: Pull distributions when running the algorithm on 1000 pseudo-experiments where Z rapidity distribution are simulated. $P_1 - P_{20}$ are the pull functions of the 20 PDF (eigenvalues) for the MSTW08 model.

6.4 Results

6.4.1 Luminosity determination

The algorithms have been tested with pseudo-experiments that have been generated using the PDF sets MSTW08, CTEQ66, Alekhin02 and NNPDF. We fit the pseudo-datasets using the same model that had been used to generate the distribution. The estimated luminosities returned by the fit on one thousand pseudo-experiments are distributed around the truth luminosity values set at generation level; the RMS of the distribution is the statistical precision achieved. This precision depends on luminosity because with small luminosity (i.e. larger statistical uncertainties) statistical fluctuations prevent an accurate fitting of the rapidity distributions and hence a precise extraction of the parameters. The precision on the luminosity also depends on the model, primarily due to the different structure and flexibility of the PDF set. Percentage precisions for different models and different values of integrated luminosity are shown in table 6.2. For all the models considered, the precisions achieved improves with the luminosity and the best precision is obtained fitting at the same time the $W^{\pm}$ and Z rapidity distribution.

	MSTW08	CTEQ66	Alekhin02	NNPDF1.0
$0.1 fb^{-1}$				
W^+	1.8	2.4	2.0	2.7
W^-	1.9	2.6	2.2	2.3
Z	1.9	2.4	2.2	2.2
WWZ	1.7	2.3	1.8	1.9
$1 fb^{-1}$				
W^+	1.6	2.2	1.8	2.3
W^-	1.6	2.3	2.1	2.2
Z	1.7	2.1	1.9	1.9
WWZ	1.3	2.1	1.4	1.6
$10 fb^{-1}$				
W^+	1.3	2.0	1.5	2.1
W^-	1.2	1.9	1.6	2.1
Z	1.4	1.9	1.9	1.8
WWZ	0.8	1.7	1.0	1.3

Table 6.2: Percentage statistical uncertainty on luminosity obtained from fits to $W^{\pm}$, Z and a combined sample of these (WWZ) corresponding to 0.1, 1 and 10 fb^{-1} with PDFs described by MSTW, CTEQ, Alekhin and NNPDF.

6.4.2 PDF constraints

We present here the performance of the algorithm and its expected impact on the PDF constraints. For the case of the Hessian PDFs, in figure 6.5 we show the correlation between the value of the parameter estimated by the fit (x-axis) and the true value of the parameter set at generation level (y-axis) when running the algorithm on 1000 Z pseudo-experiments for 1 fb^{-1} of integrated luminosity. For some parameters a strong correlation is not observed; this is due to the lack of sensitivity of the Z rapidity distribution to a particular direction of the eigenvector space, i.e. some eigenvectors provide a small contribution to the cross-section (a feature deducible also from figure 5.12). However this is no cause of concern because the fit returns a larger value for the estimated parameters uncertainties and, as already explained, the pull distributions confirm that the method is statistically valid. In figure 6.6 the equivalent plots for parameters extracted from a Drell-Yan rapidity distribution in the mass region $5 - 10\ GeV$ is shown; correlations are more pronounced due to the larger sensitivity to the PDF eigenvectors for this process.

We now present here the expected improvement on *u-valence, d-valence, sea , s^+ and gluon* distributions when introducing 1 fb^{-1} of LHCb electroweak data. The improvement is calculated as the ratio between the uncertainty on the PDF after and before the fitting procedure. This ratio is plotted as function of the bjorken-x variable in figures 6.7 and 6.8. If no improvement is obtained, the ratio will be 1, while if an improvement is present the ratio will assume values smaller than 1. In figure 6.7, results when fitting the rapidity distributions of Z, W^+, W^- and a combined fit of the three are plotted. For the $W^{\pm}$ and Z production the best result is obtained when performing a combined fit of the three rapidity distributions. About a 30% improvement can be expected over the whole x range for each of the PDFs. The combined fit WWZ provides a larger improvement with respect to the fitting procedure applied to the channels singly. The results when fitting the Drell-Yan rapidity distributions in four different mass bins are shown in figure 6.8. The results agree with the qualitative arguments given in section 5.6: the largest improvement is given by the lowest mass bin of the Drell-Yan process. With 1 fb^{-1} of LHCb data the *sea* and the *gluon* distributions at $x \sim 10^{-5}$ and at high-x can be improved by up to 70% while in mid-x regions the improvement is about 50%. Other PDFs can be improved over the full x range by about 60%. Differences in the size and shape of the improvement depend on the model used; these differences arise from features of the PDF set such as the form and flexibility of the parametrization and the datasets included in the global fits. Results for different PDF sets are shown in appendix A.

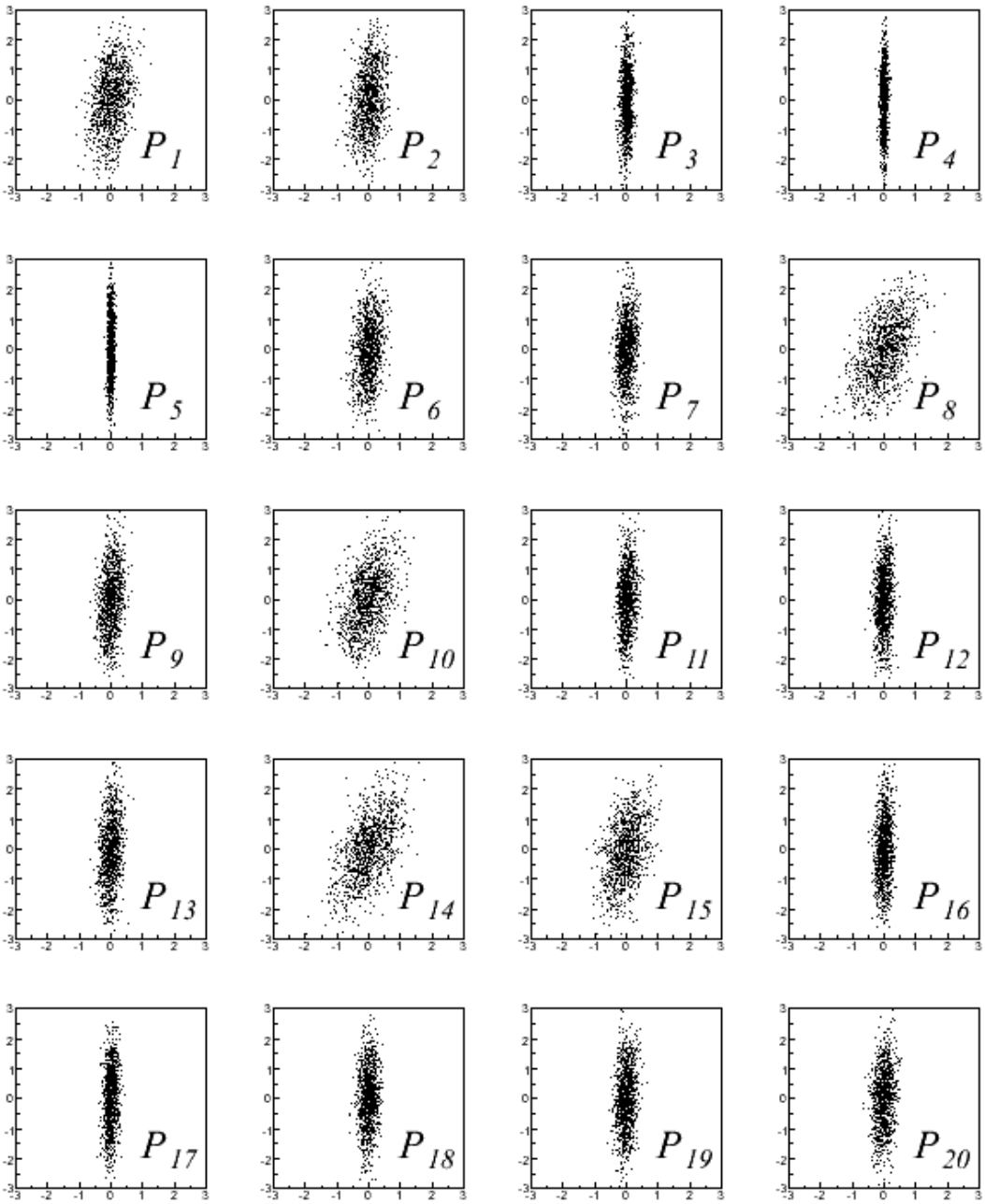


Figure 6.5: Scatter plot of the truth values (y-axis) of the parameter used in the pseudo-experiment generation against the fitted values (x-axis) for each of the 20 MSTW08 eigenvectors when running the algorithm on 1000 pseudo-experiments of Z rapidity distribution.

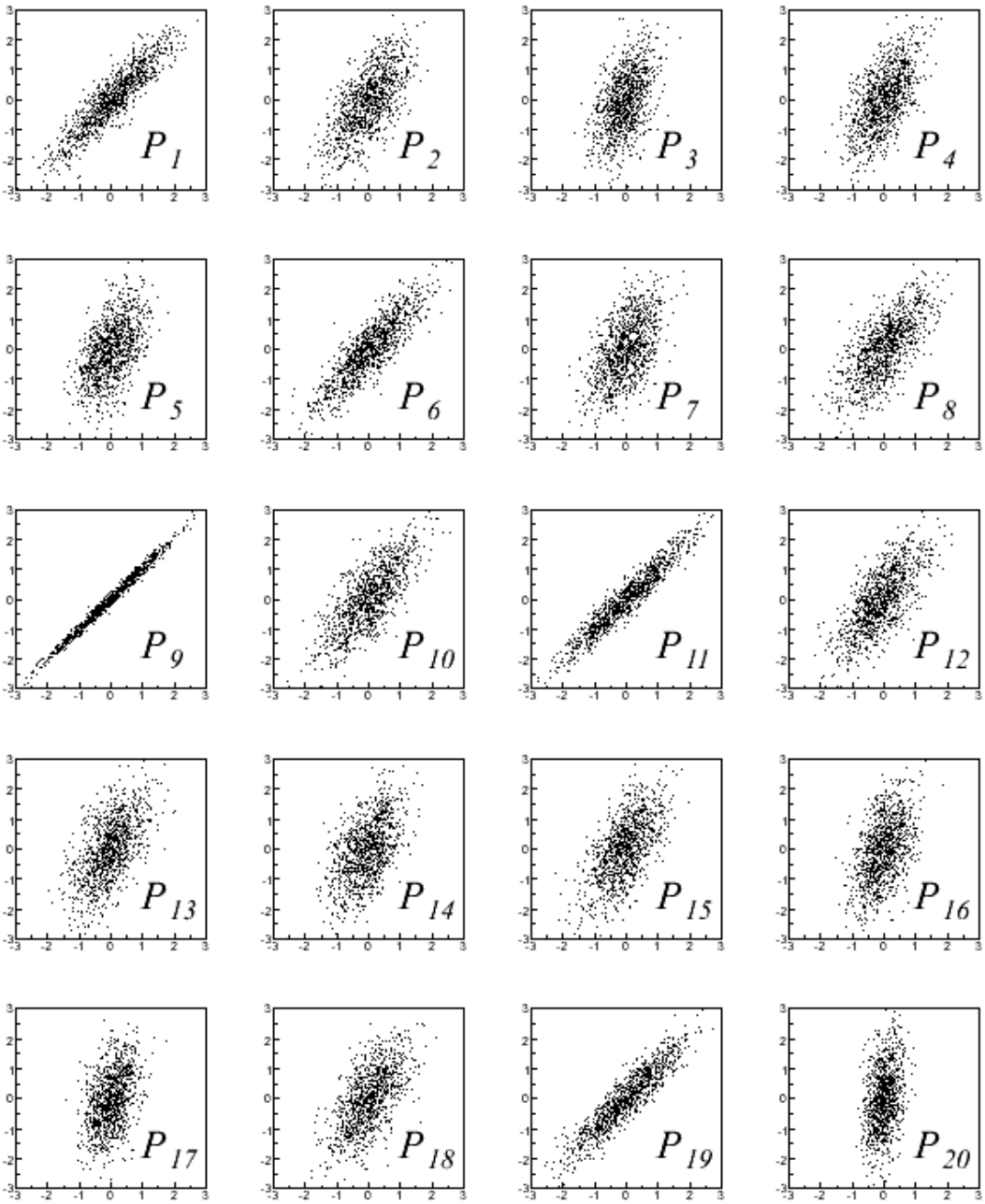


Figure 6.6: Scatter plot of the truth values (y-axis) of the parameter used in the pseudo-experiment generation against the fitted values (x-axis) for each of the 20 MSTW08 eigenvectors when running the algorithm on 1000 pseudo-experiments of Drell-Yan rapidity distribution in the mass region $5 - 10\ GeV$.

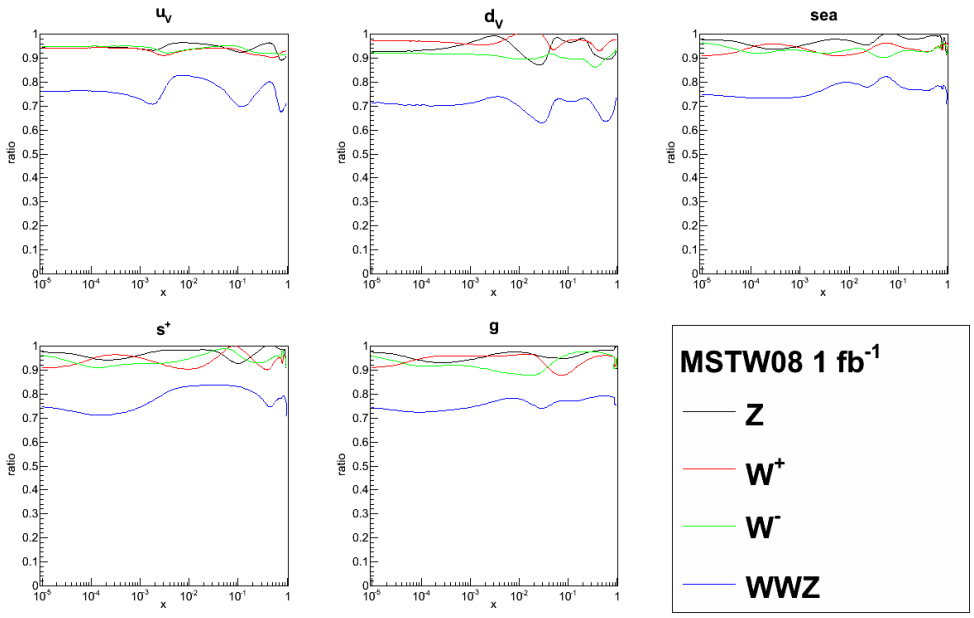


Figure 6.7: Improvement when fitting W^+, W^- Z rapidity distributions and a combined fit of the three using 1 fb^{-1} of LHCb data for the PDF set MSTW08.

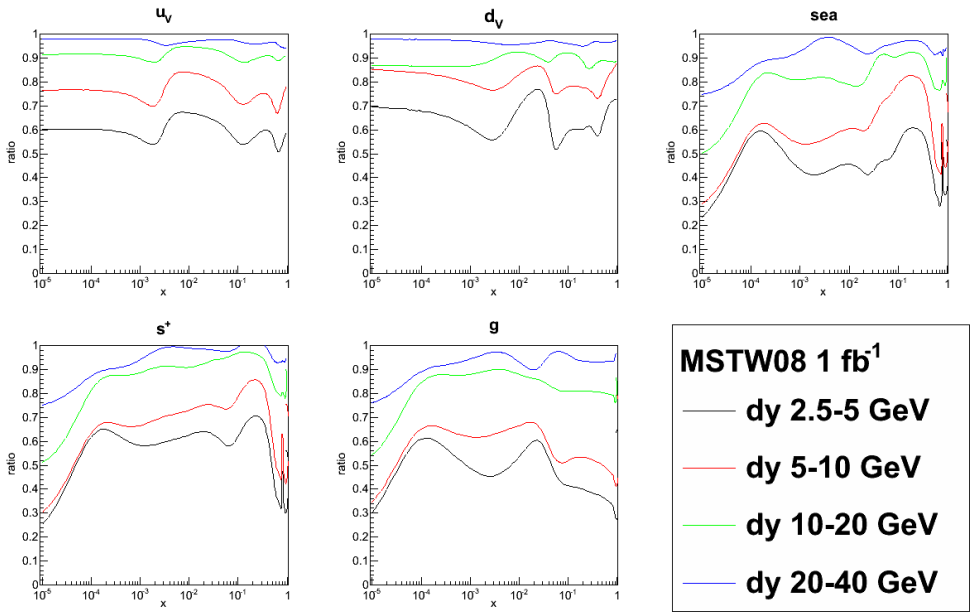


Figure 6.8: Improvement when fitting Drell-Yan rapidity distributions in 4 different mass bins using 1 fb^{-1} of LHCb data for the PDF set MSTW08.

6.5 Systematic effects on luminosity determination

The luminosity determination is affected by a systematic uncertainty if the PDF set used in the fit is different from the one used at generation level. In fact, different PDF sets predict different values of a physics observable, therefore, to use a physics channel with the real data to monitor the luminosity, we need to investigate the effect of the use of a different PDF set and, possibly, quantify the bias that can occur. Consequently, we chose one of the available PDF sets as the true description of the proton structure and generated 1000 pseudo-experiments according to this model before running our fitting algorithm using a different PDF set to extract the luminosity. In this procedure the possibility that observables generated with different PDF sets might be incompatible has to be considered. The statistical estimator of the χ^2 probability helps to discriminate between models and to decide whether a luminosity estimate can be considered trustworthy. We decide to consider only fits with χ^2 probability larger than 1% and so three outcomes are possible. Firstly, the data is compatible with more than one PDF set. In this case, a theory systematic can be determined from the difference in the fit with different model. Secondly, the data is compatible with only one PDF set. In this case it is inappropriate to take as a systematic the difference with the luminosity as fit using a different PDF set. One might reasonably conclude that variations in the theory are already accounted for by the flexibility in the fit that has the freedom to change the contribution of the eigenvectors. In the third case, the data is incompatible with all the PDF sets. In this case, since the models are not describing the data at the 1% level, one would have little confidence in trying to claim that the luminosity is known to this level. The experimental results must be fed back into the global fits where perhaps a different parametrisation could accommodate them.

The PDF set MSTW08 has been chosen as the true model while the sets NNPDF, CTEQ66 and Alekhin02 are used as test models i.e. we have generated pseudo-data using MSTW08 and fit using the other PDF sets. Figure 6.9 shows the results for 1000 pseudo experiments performed using 1 fb^{-1} of data. When fit using MSTW08 (as shown on the x-axis), there is no bias in the results and the spread of values gives the statistical uncertainty in the technique as presented in Table 6.2. However, when we

	CTEQ66	Alekhin02	NNPDF1.0
W^+	-2.9	-1.2	-3.2
W^-	1.2	3.5	3.5
Z	-0.6	-3.9	2.4
WWZ	2.0	-6.4	2.0

Table 6.3: Percentage bias on luminosity obtained when generating pseudo-experiments with MSTW08 model and fitting with different models.

fit using CTEQ66, although a strong correlation in the result is present, a clear bias

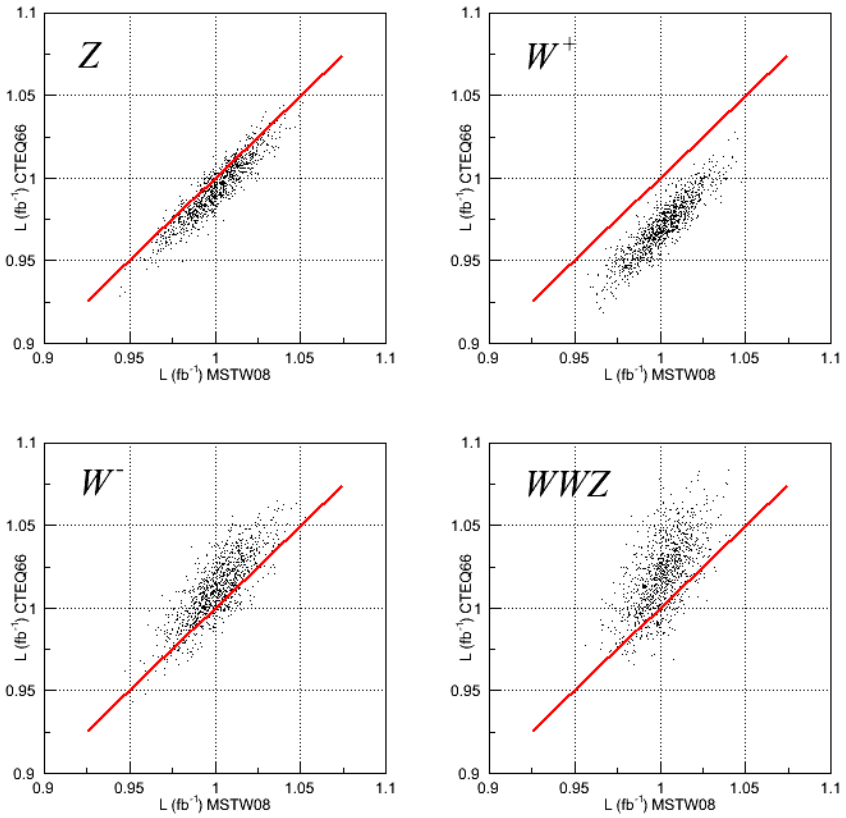


Figure 6.9: Values of the luminosity obtained when generating pseudo experiment with PDF set MSTW08. On the horizontal axis the values when fitting with the MSTW08 and on the vertical when fitting with CTEQ66. The red line represents the ideal situation when both the models give the same estimate of the luminosity.

is also observed. This bias is minimum when fitting a Z rapidity distribution while for W^- there is a general underestimation of the luminosity. For W^+ and a combined fit WWZ the bias is modest while the correlation between the luminosity estimates is less marked. In table 6.3 the percentage biases for the different models are shown. In the model Alekhin02 a large bias due to the substantial differences of this parton fit is present; Alekhin02 considers only a subset of data (namely DIS data) and so it is difficult to compare with other global fits which include hadronic data. NNPDF, on the other hand, provide a flexible structure which allows larger uncertainty and it permits a better agreement with a different PDF set. An added advantage of fitting the differential distributions is that we now have some ability to distinguish between models. The shape for the differential cross-section as predicted by MSTW08 may not be consistent with that predicted by other models, and this will be highlighted in the χ^2 probability of the fit. With the requirement that the χ^2 probability of the fit be above 1%, table 6.4 and 6.5 detail the percentage of times that the χ^2 probability cut is

	CTEQ66	Alekhin02	NNPDF1.0
W^+	2.3	11.7	5.2
W^-	0.7	6.8	4.5
Z	2.1	87.6	4.0
WWZ	10.3	90.2	15.4

Table 6.4: Percentage of times the fits with different PDF sets has a χ^2 probability smaller than 1% where pseudo-data for LHCb has been generated with MSTW08 for an integrated luminosity of 0.1 fb^{-1}

	CTEQ66	Alekhin02	NNPDF1.0
W^+	8.5	80	9.8
W^-	2.9	60	12.1
Z	5.9	100	30.5
WWZ	72.5	100	90

Table 6.5: Percentage of times the fits with different PDF sets has a χ^2 probability smaller than 1% where pseudo-data for LHCb has been generated with MSTW08 for an integrated luminosity of 1 fb^{-1}

not passed when fitting with CTEQ66, Alekhin or NNPDF. Table 6.4 shows result for 0.1 fb^{-1} of data and table 6.5 for 1 fb^{-1}. With 1 fb^{-1} of data, most of the fits with the Alekhin PDF set fail. Thus. the Alekhin PDF set can be shown to be incompatible with MSTW08 with this amount of data. CTEQ66 and NNPDF are indistinguishable from MSTW08 with this amount of data.

6.6 Conclusions

In this chapter we described a fitting algorithm which is able to measure the luminosity of a dataset and constrain the PDFs. This algorithm fits the rapidity distributions for Z, $W^{\pm}$ and Drell-Yan production at LHCb. The algorithm has two different implementations according to the structure of the PDF set considered: those built with the Hessian method and with a Monte Carlo method. The luminosity can be determined with a statistical precision of 1% when fitting a dataset which corresponds to an integrated luminosity of 10 fb^{-1}. Some ability to distinguish between models has been shown and systematics effects due to the choice of the model have been described. These systematics are the dominant effect in the luminosity determination. PDF constraints have also been discussed: expected results for PDF constraint show LHCb's great potentiality in giving informations, in particular in the low-x region, on the PDFs. Drell-Yan production in the lowest invariant mass bin appears to be the most interesting channel that can

be studied at LHCb and can provide the largest constraints on the PDFs. With 1 fb^{-1} of LHCb data the *sea* and the *gluon* distributions at $x \sim 10^{-5}$ and at high-x can be improved up to 70% while in mid-x regions the improvement is about 50%.

7. MEASURING DRELL-YAN PRODUCTION AT LHCB

In this chapter we describe how to reconstruct the dimuon invariant mass spectrum at LHCb and to separate the Drell-Yan signal from the background contributions. We present the strategy to select dimuon pairs produced through the Drell-Yan mechanism, the background estimation and finally we will describe the results achieved for the extraction of the Drell-Yan signal from LHCb data. The theoretical prediction for this process has been discussed in chapter 4, calculations for production cross-section were described in section 4.7 and the importance of the measurement of the Drell-Yan cross-section at LHCb has been pointed out in section 5.6 and in [112].

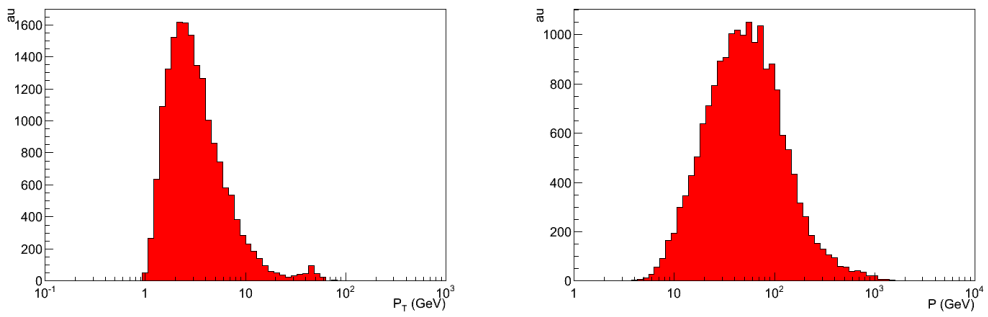
7.1 Signal Events

To perform this analysis and study the characteristics of the Drell-Yan signal in LHCb we used a Monte Carlo sample of Drell-Yan events. The samples have been generated within the official LHCb software framework. The Monte Carlo generator used to generate to events is PYTHIA 6.418 [57] Tune A with the default UE model, using the CTEQ5L [120] set of PDF which are interfaced with the generation program through the LHAPDF library [121]. The PDF set CTEQ5L is a leading order QCD distribution which should be appropriate for simple calculations and for use in Monte Carlo programs such as PYTHIA. For this study we didn't study the effect of different PDF sets because the kinematic, geometric and detector reconstructed variables we want to study have a very mild dependence on PDF. The detector effects were simulated using Gauss. The detector digitization was performed by Boole and event reconstruction by Brunel as described in section 2.3.9. To speed up the generation process the generated events were required to contain a muon arising from γ^*/Z with a direction of flight that lies within 400 $mrad$ of the beam axis and a transverse momentum greater than 1 GeV. The full amplitude has been simulated including the on-shell Z boson production and the γ^*/Z interference diagram. We generated these events in three different samples

with a different lower cutoff on the virtual photon momentum, which corresponds to a requirement on the minimum di-muon invariant mass accepted. The cutoffs are set to $m > 2\ GeV$, $m > 5\ GeV$ and $m > 10\ GeV$. The reason for this choice is twofold. The cross-section in equation 4.37 for the Drell-Yan process depends on the inverse of the invariant mass of the dimuon system and, therefore at very low values of invariant mass the cross-section becomes very large and hence a cutoff is required to stay away from this singularity. Furthermore, after generating a finite number of events, most of them will have an invariant mass very close to the cutoff where the cross-section is largest. This means that the number of events at higher values of invariant mass will be small; generating different samples with different cutoff value allows us to map out the full mass spectrum with an adequate number of events.

7.1.1 Signal Characteristics

Drell-Yan muon pair production has a distinctive topology: we look for events with two muons of opposite charge arising from the primary vertex of the interaction and with transverse momentum larger than 1 GeV. This last requirement is redundant since in LHCb, tracks with $P_T < 1\ GeV$ do not reach the muons stations and so cannot be identified as muons. In figure 7.1(a) we show the transverse momentum spectrum of the Drell-Yan muons and in figure 7.1(b) momentum spectrum.



(a) (b)

Figure 7.1: Dimuon Drell-Yan event characteristics: (a) Transverse momentum (P_T) distribution of the muons in GeV. (b) Momentum (P) distribution of the muons. In the transverse momentum distribution it is possible to see an enhancement at $P_T \sim 45\ GeV$ due to the production of a real Z boson decaying into two muons which have high transverse momentum.

In figure 7.2 we show the Impact Parameter significance distribution for muons generated through the Drell-Yan mechanism. Impact Parameter significance is defined as the muon impact parameter[1] divided by its uncertainty, thus muons arising from the interaction

[1]the impact parameter is defined as the distance of closest approach between the particle's track and

point are expected to have small impact parameter significance. Impact Parameter significance (IPs) is defined as the muon impact parameter divided by its uncertainty.

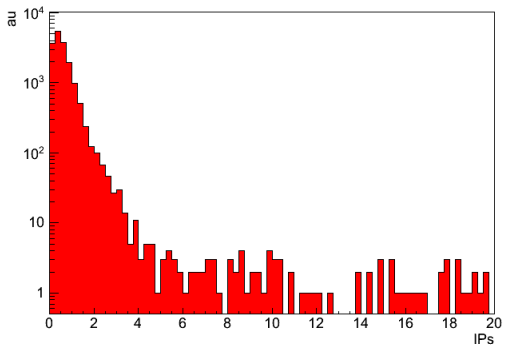


Figure 7.2: Muon impact parameter significance distribution for muons from Drell-Yan produced events.

7.2 Background Events

The background processes we investigated all produce two identified opposite charge muons in the LHCb acceptance with $P_T > 1\ GeV$. In the low invariant mass and momentum region most of these muon pairs arise from semi-leptonic heavy quark (HQ) decays and hadron mis-identification.

7.2.1 Heavy Quark Decay

To study the heavy quark decay background we used two samples where both the hadrons with a b and c quark produced in the initial interaction are forced to decay semi-leptonically to muons. In particular to speed up the generation process we required that for the *inclusive* B sample at least one B hadron in 400 mrad around the z axis was produced and for the *inclusive* C sample at least one D hadron in 400 mrad around the z axis was produced. Both of the samples consist of about 21 million events but, for the *inclusive* B sample, only 87685 combinations of opposite sign muons are present and, for the *inclusive* C sample, 31969 combinations of this kind can be found.

Table 7.1 presents a comparison of the cross-sections of signal events and background events which have two opposite sign reconstructible muons with invariant mass $M >$

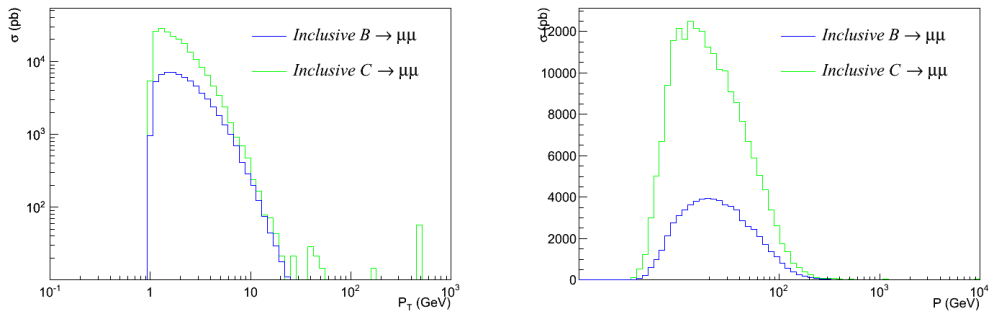
the primary vertex

2.5 GeV^2; these numbers are obtained using PYTHIA and they are meant only to give an idea of the order of magnitude of the cross-sections since the they are affected by large theoretical uncertainties that for heavy quark production [122, 123, 124], can be of the order of a factor 2. The cross-section for the signal that we want to study is

Sample	Cross-section (mb)
$Z/\gamma^* \to \mu\mu$ $M > 2$ GeV	3.26 10^{-6}
$Z/\gamma^* \to \mu\mu$ $M > 5$ GeV	1.53 10^{-6}
$Z/\gamma^* \to \mu\mu$ $M > 10$ GeV	1.91 10^{-7}
Inclusive $B \to \mu\mu + X$	9.64 10^{-4}
Inclusive $C \to \mu\mu + X$	1.96 10^{-4}

Table 7.1: Cross-sections for signal and background events considering events with two opposite sign muons with invariant mass larger that 2.5 GeV in LHCb. These cross-sections are obtained with PYTHIA 6.418

about two order of magnitude smaller than the background events. In figure 7.3 the momentum and the transverse momentum of the background events are plotted. In



(a) (b)

Figure 7.3: Background events characteristics: (a) Transverse momentum (P_T) distribution of the muons. (b) Momentum (P) distribution of the muons.

Figure 7.4 the impact parameter significance of the muons coming from heavy quark decay are reported. This distribution has a long tail which is due to the long-lived nature of B and D hadrons, which after being produced, travel for a few millimeters from the primary vertex generating a non zero impact parameter. This feature is not present in the distribution in figure 7.2 for Drell-Yan process because the muons arise directly from the primary vertex. For this reason this variable can be used to reject heavy quark decays events.

[2]Here we have followed the standard LHCb definition of reconstructible, i.e. a reconstructible long track is one that, at Monte-Carlo truth level, has 3 radial and 3 ϕ VELO clusters and one x cluster and one stereo cluster in each of the three T stations

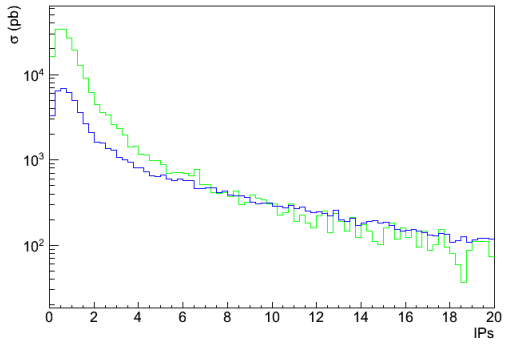


Figure 7.4: Muon impact parameter significance distribution of the muons from heavy quark decay. Note that these distributions have a very long tail compared to the Drell-Yan Impact parameter distribution in Fig.7.2

7.2.2 Hadron Mis-identification

The most important source of background events comes from combinations of oppositely charged pions or kaons that are both mis-identified as muons. There are two mechanisms that can generate a hadron mis-identification: the first one, called *punchthrough*, consists in a high momentum pion (or a kaon) which is not stopped in the hadronic calorimeter and reaches the muon detector. This kind of charged track satisfies all the requirements to be identified as a muon, even though they are not genuine muons. The second mechanism occurs when a pion (or a kaon) which is produced from the primary vertex, decays into a muon and a neutrino. In this case, the muon detected is a genuine muon, but it has not been produced through the Drell-Yan mechanism. A study of the mis-Identification rate has been performed in [116]; we now summarize a description of the hadron mis-ID while in section 7.4.1 we will describe the technique we used to estimate these backgrounds sources with a data driven method.

Punchthrough

The hadron punchthrough occurs when a pion/kaon is not contained in the calorimeter systems and reaches the LHCb muon system. The amount of material that the pion went through corresponds to 21 interaction lengths[3]. The probability of punchthrough is proportional to the momentum of the pion/kaon; a parametrization of this probability as a function of the momentum can be obtained from the data provided by the RD5 collaboration [125] which measured for a wide spectrum of momenta the absorption of pions and kaons. In figure 7.5 we show the results of the RD5 collaboration of

[3]The nuclear interaction length of a material is the mean distance that a hadron will travel in that material before it undergoes an interaction.

punchthrough probability as a function of the momentum. As described in [116], it is possible to fit these data points with a straight line to parametrize the effect.

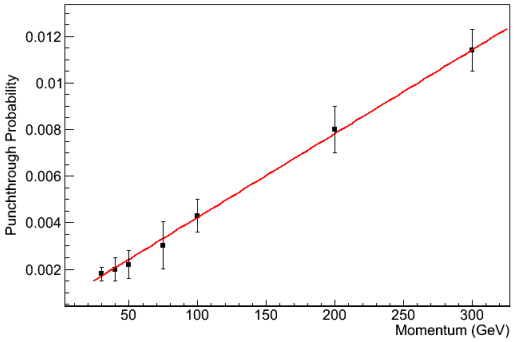


Figure 7.5: Punchthrough as a function of the pion/kaon momentum measured by the RD5 collaboration for the equivalent of 3.5 meters of Iron.

Decay in Flight

When a pion or a kaon is produced directly from the primary interaction, there is a certain probability that it decays into a muon and a neutrino after it flew outside the VELO detector. Then the muon easily reaches the muon stations and hits on those can be associated to the original pion (kaon) track. The probability of the pion/kaon decay is easy to compute theoretically since it is possible to calculate the mean lifetime of pions and kaons in the laboratory frame. We work out the mean lifetime, τ_L, in the Lab frame boosting the mean lifetime at rest:

$$\tau_L = \gamma\tau_R = \frac{E\tau_R}{M} \approx \frac{\tau_R}{M}P \tag{7.1}$$

where τ_R is the mean lifetime calculated in the rest frame and M the mass of the particle. The boost is performed multiplying by the γ relativistic Lorentz factor and the energy E of the decaying particle is approximated with its momentum P. The probability that it travels a distance x_0 or smaller is given by:

$$P(x_0) = 1 - e^{-x_0/c\tau_L} = 1 - e^{-Mx_0/c\tau_R P} = 1 - e^{-\Lambda/P} \tag{7.2}$$

since the distance between the VELO detector and the muon system is about 15 m, considering the appropriate mean lifetime at rest and masses for pions and kaons the

coefficients Λ are:

$$\begin{aligned} \Lambda_\pi &= 0.269 \\ \Lambda_K &= 1.994 \end{aligned}$$

This function provides a conservative description of the mis-ID probability since it assumes that every pion/kaon that decays in this way will result in the association of its VELO track and the corresponding muon chamber hits. Not every muon track that comes from a pion/kaon decay is matched with the original pion/kaon track. In the case that the pion (kaon) decays inside the VELO, the muon track will not point to the primary vertex, it will have a large impact parameter and so this track will be removed by requiring a small impact parameter.

7.3 Signal Selection and Background rejection

To extract the Drell-Yan signal from the huge backgrounds, we developed a selection strategy based on two different kind of cuts: the first uses simple kinematic and geometric variables and the second exploits features of the underlying event.

Track Quality requirements

To ensure that the tracks we use in the analysis are well reconstructed we ask, for each muon, that the χ^2 probability of the track fitting is greater than 1%, and that the fractional momentum uncertainty is less than 10%. As it has been shown in [126], these requirements significantly reduce the ghost rate [4] and badly measured low momenta muons.

Kinematic and Geometric requirements

First of all we look for events with, at least, two opposite sign muons. Then the following kinematic and geometric cuts are required:

- The dimuon invariant mass must be larger than 2.5 GeV
- Both muons must have $P_T > 1$ GeV
- Both muons must have $P > 10$ GeV

[4] Definition of the ghost tracks has been given in section 2.3.3

- Both muons must have $IPs < 5$

Momentum and transverse momentum of the muons for signal events in figure 7.1 and for backgrounds events in figure 7.3 are shown. The cut on the muon momentum has been set to maximize the significance S of the signal,

$$S = \frac{s}{\sqrt{b}} \tag{7.3}$$

i.e the ratio between the number of signal event s which pass the cut divided by the square root of the number of background events b which pass the cut. Here we considered as background both heavy quark decays and hadron mis-Identification. In figure 7.6 we show the significance as a function of the muon momentum cut. The requirement on the

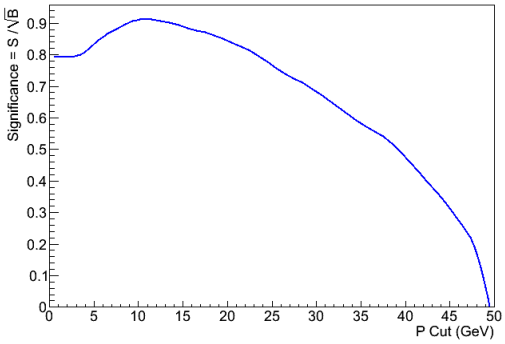


Figure 7.6: Significance of the Drell-Yan Signal as function of the cut on the muon momentum when considering as background processes heavy quark decay into muon pair. The significance is maximized when the cut on the muon momentum is $P_\mu > 10\ GeV$.

impact parameter significance of the muons is chosen to preserve the Drell-Yan signal, which is completely confined for values of $IPs < 5$ as can be seen in figure 7.2. In figure 7.7 we present the efficiency of the IPs cut for a Drell-Yan Monte Carlo sample and the heavy quark decay background. the efficiency is defined as the fraction of events with $IPs < IPs\ Cut$. Requiring that the impact parameter significance is smaller than five for at least one of the muons would remove 30% of events from B decays and 10% from C decays, while rejecting only 1% of Drell-Yan signal.

Removing the J/ψ resonance

A significant part of the remaining background is due to the prompt production of the J/ψ meson: this resonance is detected as a narrow peak in the dimuon invariant mass spectrum at 3.096 GeV, while the Drell-Yan signal produces a continuum spectrum. Removing the J/ψ peak leaves only the continuum part of the background. To estimate

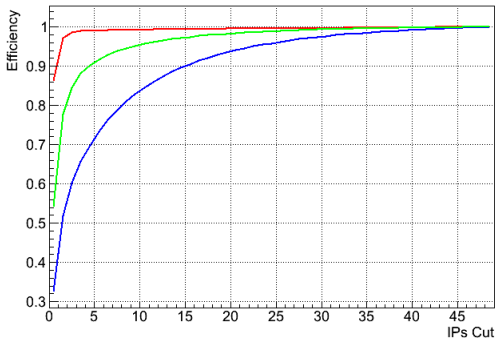
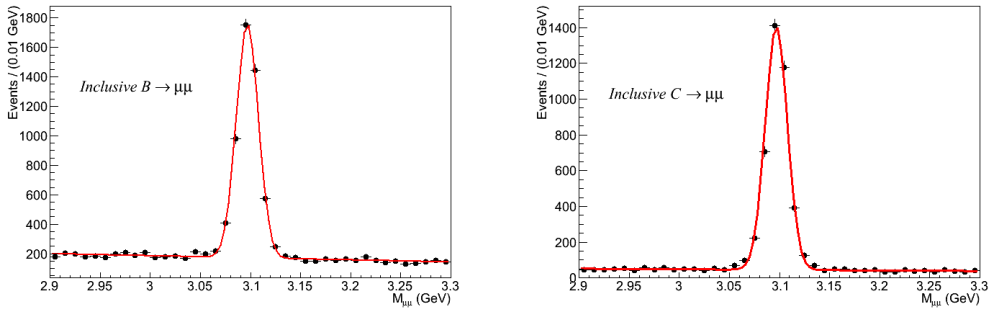


Figure 7.7: Efficiency of the impact parameter significance cut: the efficiency is defined as the fraction of events with $IPs < IPs\ Cut$. In red the efficiency for Drell-Yan in blue for inclusive $B \to \mu\mu$ and green for inclusive $C \to \mu\mu$.

the contribution of the J/ψ peak to the background cross-section, we selected a region of interest around the peak $2.9 < M < 3.3 GeV$ and fitted the peak with a function which has been obtained adding a gaussian and a polynomial. The gaussian models the resonant peak while the polynomial the continuum contribution. The integral of the gaussian gives an estimate of how many events belong to the peak. In Figure 7.8 we show the fits with a gaussian distribution plus a polynomial to the J/ψ peak in the B and C inclusive Monte Carlo samples. The integral of the gaussian distribution is also shown. For the B inclusive sample about 3400 out of 8041 events pass the kinematic cuts (i.e. $\sim 42\%$) while for the C inclusive sample 3509 out of 5895 (i.e. $\sim 60\%$) belongs to the J/ψ. To reject these events we require the dimuon invariant mass is not in the region $3.07 < M < 3.13\ GeV$.



(a) (b)

Figure 7.8: Fit of the J/ψ mass peak for the inclusive B and inclusive C Monte Carlo samples to estimate the size of the peak contribution.

The global effect of these kinematic cuts can be inferred from figures 7.9 and 7.10 in

which the total efficiency of the selection cuts for signal and heavy flavour backgrounds is plotted as a function of the invariant mass. For the Drell-Yan signal for invariant mass above 10 GeV the efficiency is about 95% with fluctuations due to the limited size of the Monte Carlo sample analyzed. At low invariant mass the efficiency decreases up to 65% for 2.5 $GeV < M < 6.5$ GeV. As expected the efficiency for the background samples is smaller and floats between 10% and 30%; no dimuon candidates with invariant mass $M > 20$ GeV pass the selection criteria. This is partly due to finite statistics of the simulated events. In table 7.2 the effective cross-sections for expected signal and backgrounds events which pass the kinematic cuts are reported. The background is still large compared to the expected signal. For this reason we need to introduce additional selection criteria based on the underlying event shape.

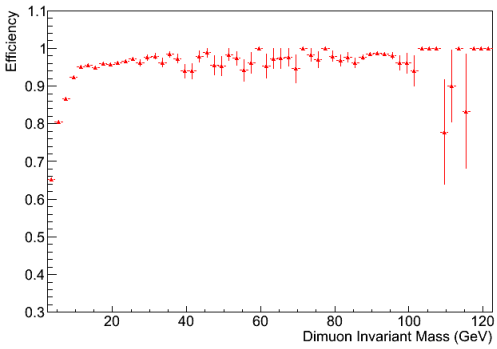


Figure 7.9: Efficiency of the kinematic and geometric cuts for the Monte Carlo Drell-Yan sample as a function of the invariant mass. The efficiency is defined as the number of events which pass the selection cuts divided by the total number of events.

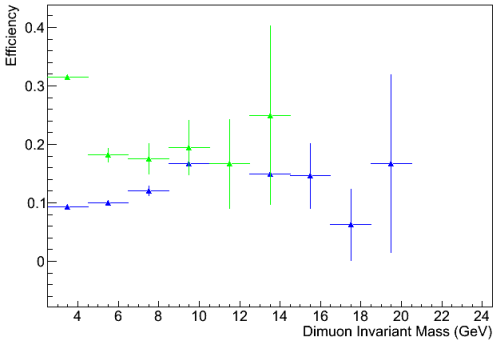


Figure 7.10: Efficiency of the kinematic and geometric cuts for the Monte Carlo B Inclusive (in blue) and C Inclusive (green) samples as a function of the invariant mass.

Sample	Cross-section (mb)	Stat. Uncertainty (%)
$Z/\gamma^* \to \mu\mu$ $M > 2$ GeV	2.29 10^{-6}	0.9
Inclusive $B \to \mu\mu + X$	8.84 10^{-5}	2.5
Inclusive $C \to \mu\mu + X$	4.16 10^{-5}	2.7

Table 7.2: Effective cross-sections for signal and background events considering events with two opposite sign muons in LHCb which pass the kinematic cuts. Statistical uncertainties are calculated with binomial errors on the number of events in the sample used.

Cone Variables Cuts

The distinguishing feature of the Drell-Yan process is the isolated nature of the muon production. Muons generated through the Drell-Yan mechanism appear in the detector as "isolated", because the underlying event particles, generated by soft interaction and proton remnants, tends to be physically distant from the muons and the underlying event is weakly correlated with muons. We have chosen four variables with the purpose of describing the underlying event in the Drell-Yan events and in the background events. These variables are each asymmetries between energy deposits x and y defined as:

$$A(x, y) = \frac{x - y}{x + y} \tag{7.4}$$

For each dimuon candidate which passes the kinematic cuts we define two regions in the (η, ϕ) space around the muons. Given a muon with $\eta = \eta_\mu$ and $\phi = \phi_\mu$ we calculate the distance R between the muon and a charged track with pseudorapidity $\eta = \eta_t$ and azimuthal direction $\phi = \phi_t$

$$R = \sqrt{(\eta_\mu - \eta_t)^2 + (\phi_\mu - \phi_t)^2} \tag{7.5}$$

If R is smaller (larger) than a given quantity ρ we say that the track lies within (outside) the region. For our analysis it was decided to set $\rho = 1$. The region defined for $R < 1$ in the (η, ϕ) space has the typical shape of a cone around the muon. To build the asymmetries we looked at three quantities: the transverse momentum of the muon P_T^μ, the transverse momentum of the vector sum of the momenta of the other charged particles in the cone, P_T^{cone}, and the transverse momentum of the vector sum of all the charged particles outside of both cones, P_T^{rest}. The other variable we used is the transverse momentum of the dimuon system P_T^{sum}.

We identified four asymmetry variables as defined in equation 7.4:

- $A(P_T^\mu, P_T^{cone})$: Asymmetry between the muon transverse momentum and the transverse momentum of the vector sum of the momenta of the other particles in the

cone. This variable is evaluated for both the positive and the negative muon of the dimuon candidate.

- $A(P_T^{sum}, P_T^{rest})$: Asymmetry between the transverse momentum of the dimuon system and the transverse momentum of the vector sum of the other charged particles outside both of the cones around the muons.
- $A(P_T^{cones}, P_T^{rest})$: Asymmetry between the transverse momentum of vector sum of the momenta the tracks which are inside the cones around the muons and the transverse momentum of the tracks outside both of the cones.

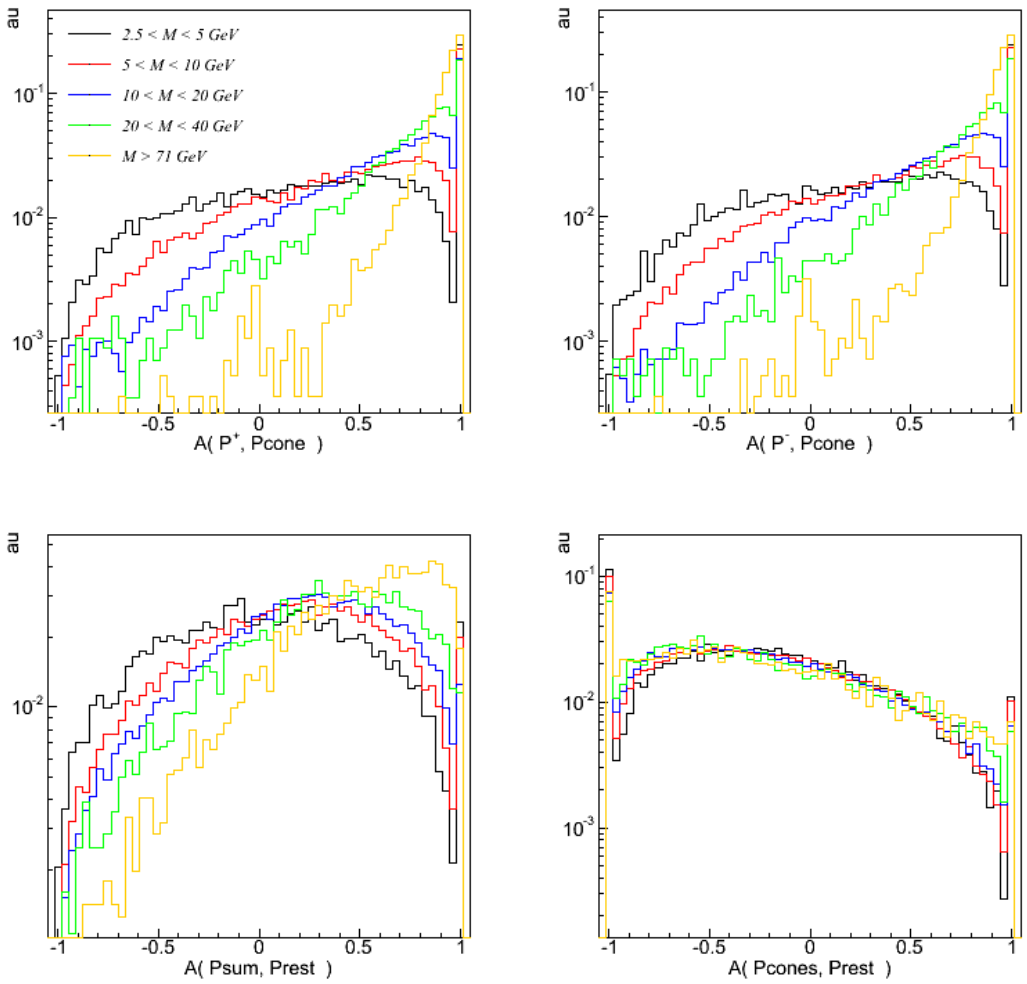


Figure 7.11: The four asymmetry variables for Drell-Yan signal for different invariant mass bins as indicated by the colour key. In this figure the distributions are normalized to unity to compare the shapes.

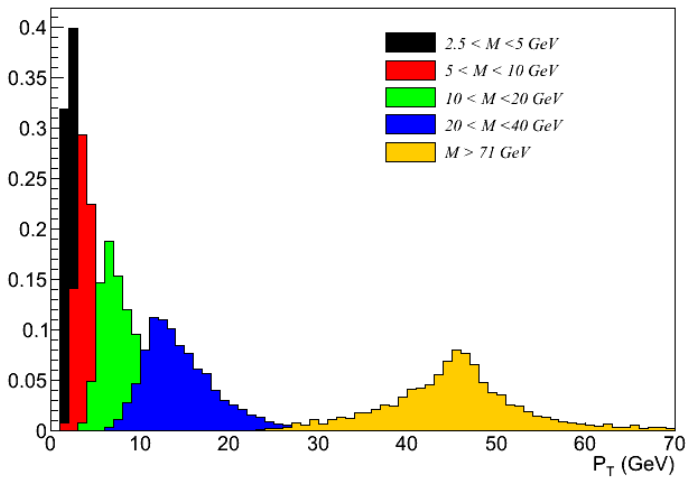
These variables, due to the asymmetry definition, can assume values between -1 and 1. In figure 7.11 the asymmetry variables for the Drell-Yan signal are shown in five different invariant mass regions; these distributions mildly depend on the invariant mass, for example for the variables $A(P_T^{\mu}, P_T^{cone})$ the higher the invariant mass of the dimuon pair produced, the more the asymmetry distributions are peaked towards one. However, these differences are not due to a change of the shape of the underlying event in the different mass ranges. In figure 7.12(b) we show the transverse momentum of the vector sum of the momenta of the charged particles in the cone (P_T^{cone}) for different invariant mass regions. The distributions are similar showing that this variable does not depend upon the invariant mass of the dimuon pair produced. In figure 7.12(a) we show the transverse momentum of the muons in different mass regions from which it is possible to see that for high invariant masses the muon transverse momentum is larger than at low invariant mass. Another proof that the underlying event shape does not depend upon the energy scale of the hard interaction can be inferred by looking at the distribution $A(P_T^{cones}, P_T^{rest})$. This variable do not depend on the energy of the muons but only on the energy deposits inside and outside the cones. This distribution is independent from the invariant mass of the dimuon too.

In figure 7.13 we show the comparison for the asymmetry variables for signal and heavy quark backgrounds. The choice of the cuts on the four variables is a complex task since correlations between these asymmetries can occur. Applying a cut on one of the asymmetry variables can radically change the shape of the other variables and so the selection strategy has to be carefully assessed. In previous studies a cut on likelihood estimator, based on the four asymmetry variables, was proposed [127]; using this technique it is possible to obtain high signal sensitivity [127, 119]. As an initial approach on real data, we decided to use a simple selection based only on the isolation cone variable of the two muons $A(P_T^{\mu}, P_T^{cone})$. The significance is maximized by requiring for both muons that the asymmetry is larger than 0.5. In figure 7.14 we show the signal significance as a function of the cut on the asymmetry variable

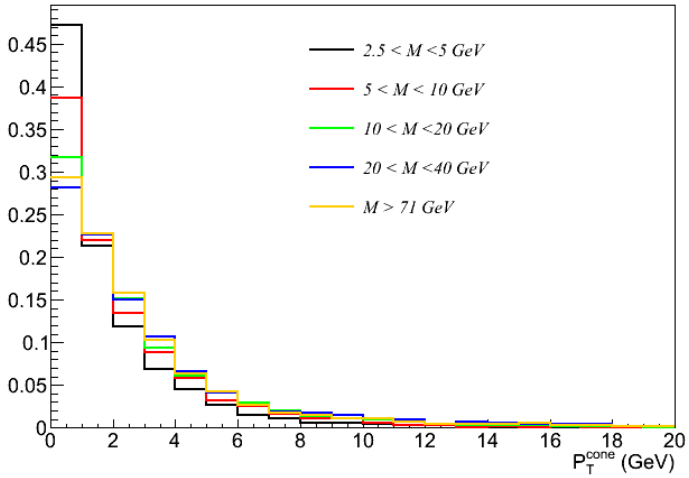
In figure 7.16 the overall efficiency in five different mass regions is plotted and it is also shown in table 7.3. Given these efficiencies, in figure 7.15 we show the invariant mass spectrum for Drell-Yan signal and heavy quarks decay background. The *inclusive C* cross-section is still larger than the Drell-Yan below 10 GeV while the B contribution is everywhere smaller.

7.4 Data Driven Methods

In any measurement there are effects that are not completely described by the Monte Carlo simulations; inaccurate detector descriptions, mis-alignment of the tracking system, track quality descriptions, reconstruction and software performance fall into this category. But not only detector effects can be imprecise. The underlying physics of the fundamental process, in the new LHC energy regime, may not be correctly described in



(a)



(b)

Figure 7.12: In (a) we show the transverse momentum (P_T) of the muons for the invariant mass ranges $2.5-5\ GeV$, $5-10\ GeV$, $10-20\ GeV$, $20-40\ GeV$ and for $M > 71\ GeV$ of the dimuon system. In (b) the the transverse momentum of the vector sum of the momenta of the charged particles in the cone (P_T^{cone}) for the same invariant mass ranges. The distributions are normalized to unity to compare the shapes.

the Monte Carlo generator. For example cross-sections for certain processes could be underestimated or overestimated as has already been discovered by LHCb which measured the cross-section of b quark production to be half that implemented in Pythia [128]. In this section we will describe data driven methods we used to measure the muon mis-ID

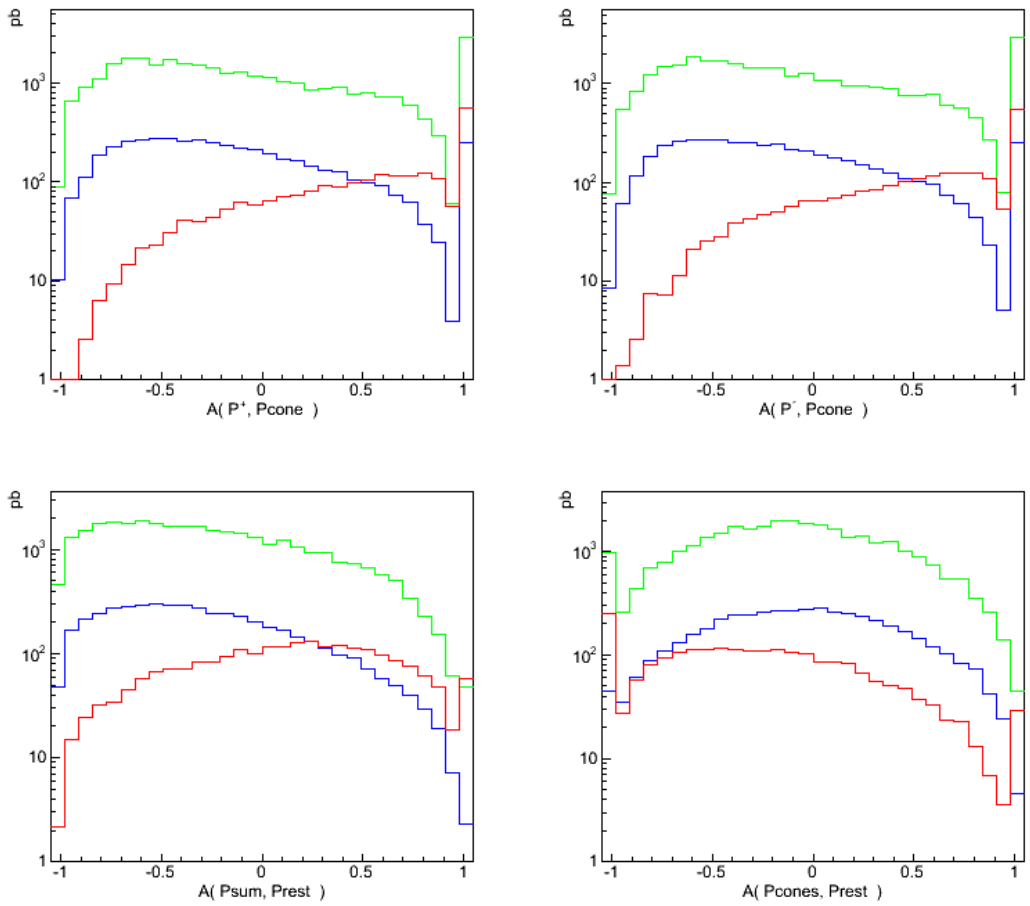


Figure 7.13: The four asymmetry variables for Drell-Yan signal and for heavy quark decay. In red the Drell-Yan signal, in blue the Inclusive B and in green the Inclusive C

Mass Range (GeV)	Efficiency
$2.5 - 5$	0.22
$5 - 10$	0.34
$10 - 20$	0.50
$20 - 40$	0.70

Table 7.3: Overall efficiency for the Drell-Yan production in four different mass ranges with asymmetry cone cut at 0.5

rate and to describe the asymmetry variables. These methods allow us to reduce our dependence on the Monte Carlo descriptions.

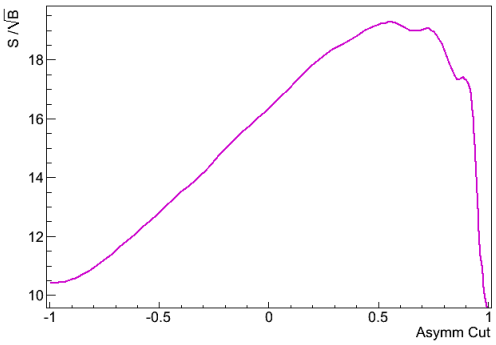


Figure 7.14: Significance of the Drell-Yan Signal as function of the cut on the asymmetry variable $A(P_T^{\mu}, P_T^{cone})$ when considering as background processes heavy quark decay into muon pair.

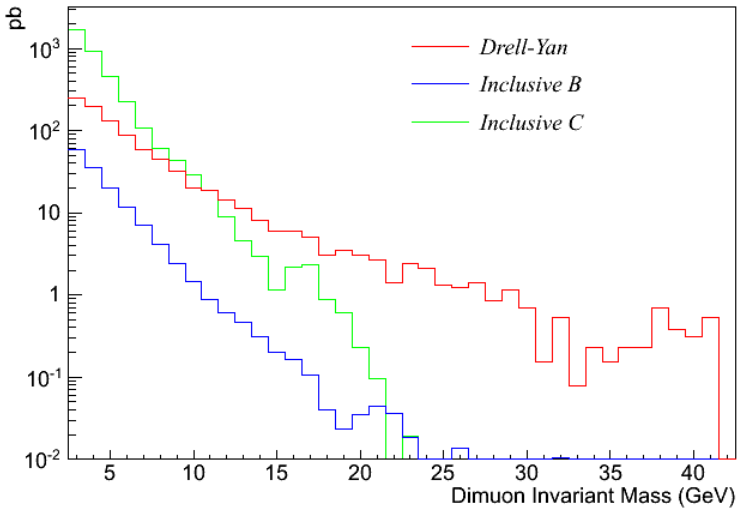


Figure 7.15: Effective cross-section after all the selection cuts are applied as a function of the dimuon invariant mass.

7.4.1 Measure of the Muon Fake Rate

Our aim is to measure the probability that a track, which is not produced by a genuine muon, is identified by the LHCb particle identification algorithm as a muon. We have already described in section 7.2.2 that the pion/kaon punchthrough and decay in flight are the largest sources of hadron mis-identification. To estimate this from the data we take all tracks where we know that the percentage of true muons is small. Then we count how many tracks are identified as muons. The ratio between the number of the tracks identified as a muon and the total number of tracks gives an upper limit on the

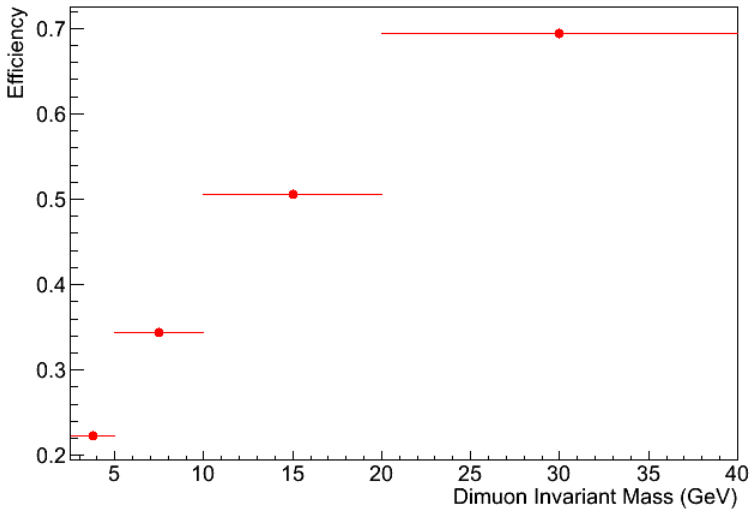


Figure 7.16: Overall efficiency in the four invariant mass regions.

mis-Identification rate. For our study a sample of minimum bias events has been used; we can considered that the contamination of real muons in the events which pass a minimum bias trigger is very small, since the requirements for a minimum bias events are very loose and are described in section 2.3.8.
We have taken all the *long tracks*[5] in minimum bias events, which are the charged tracks that potentially can be muon track. Figure 7.17 shows the fraction of those tracks passing the minimal muon ID requirement, *isMuon*[6], as a function of the track momentum P_{tr}. It is possible to fit the distribution obtained with the function

$$f(P_{tr}) = 1 - e^{-p_0/P_{tr}} + p_1 + p_2 P_{tr} \tag{7.6}$$

that describe the mis-Identification processes described in 7.2.2 as the sum of the punchthrough and decay in flight contribution. The parameters returned by the fit are:

$$\begin{aligned} p_0 &= (0.1453 \pm 1.4\ 10^{-3})\ GeV \\ p_1 &= 8.44\ 10^{-7} \pm 1.087\ 10^{-6} \\ p_2 &= (7.58\ 10^{-4} \pm 9.1\ 10^{-5})\ GeV^{-1} \end{aligned}$$

with a χ^2 per degrees of freedom of 1.11.
With this parametrization it is possible to build a differential cross-section distribution in invariant mass of fake muon pairs. Taking a sample corresponding to an integrated

[5]Definition of long tracks can be found in section 2.3.3.
[6]isMuon requirement has been described in section 2.3.7

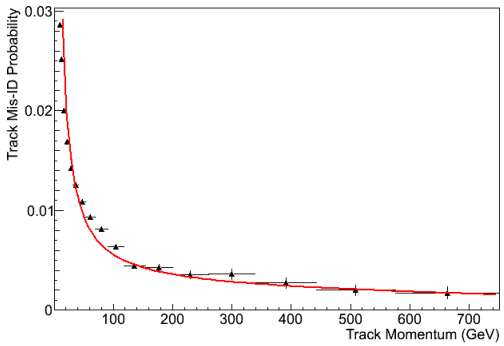


Figure 7.17: Mis-Identification probability as a function of the track momentum. In black the mis-ID rate from minimum bias events and in red the fitted parametrization are shown.

luminosity 16 pb^{-1}, which passed the minimum bias trigger (described in 2.3.8) we built pairs of opposite sign *long tracks* passing the selection cuts described in section 7.3. To each combination we gave a weight W_c, which represents the probability that the combination is made of fake muons, obtained as the product of the probability that each of the muons is a fake

$$W_c = f(P_+) \cdot f(P_-) \tag{7.7}$$

where P_+ and P_- are respectively the momentum of the positive and negative charged tracks. The fake probability for the muons is calculated with the parametrization 7.6. In figure 7.18 we show the estimated differential cross-section as a function of invariant mass for the fake dimuon pairs obtained from real data. Since minimum bias events are produce with a very high rate, this trigger line is prescaled by a factor 10000, which means that only one event in 10000 accepted by the trigger line is recorded on tape. With the available statistics of minimum bias events, we are not able to predict the misID cross-section for invariant masses larger than 30 GeV but only to set an upper limit of 0.1 pb/GeV. However this rate can be considered negligible at higher invariant masses.

7.4.2 Measure of the Asymmetry Variables Distributions

As mentioned in section 7.3, the requirements on the asymmetry variables are crucial for the extraction of the Drell-Yan signal; however the underlying physical processes are not very well understood and described by the Monte Carlo generators. Hence it is very important to use a data driven method to avoid any assumption and prejudice introduced by the Monte Carlo description of these variables. As it possible to see in figure 7.11, the asymmetry distributions for Z and for Drell-Yan in different invariant mass regions look different. We now attempt to predict these distributions using muons produced in the decay of the Z boson. Since Z boson events can be clearly isolated in

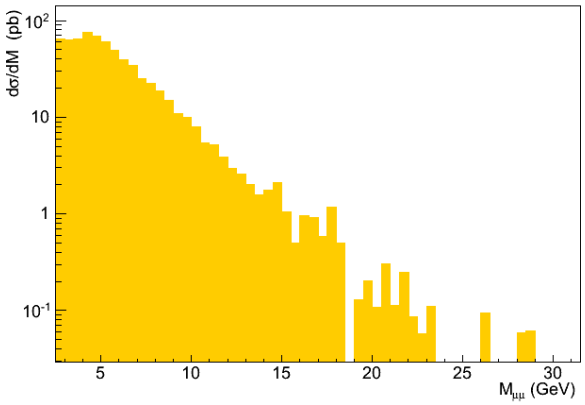


Figure 7.18: Predicted fake dimuon invariant mass obtained from minimum bias real data

data, this is a data-driven method for extracting the asymmetry shapes in each mass region. First of all we selected a sample of clean reconstructed $Z \rightarrow \mu^+\mu^-$ decay from LHCb data as reported in [126]. We looked for events with two opposite sign muons with transverse momenta larger than 15 GeV which combine to give an invariant mass greater than 81 GeV and less than 101 GeV. No impact parameter cuts or isolation criteria are imposed but track quality requirements are applied i.e. χ^2-probability of the muon track is greater than 0.1%, and the fractional momentum uncertainty is less than 10%. Once we have selected the Z decay, to obtain the asymmetry distriributions for a Drell-Yan muon pair of invariant mass M_{ph} ,the algorithm executes the following steps:

- We apply a Lorentz boost to each muon $P^\mu = (P^\mu_x, P^\mu_y, P^\mu_z, P^\mu_t)$ coming from the Z into the center of mass system of Z. We boost the muons in a direction $\vec{b}$

$$\vec{b} = \begin{pmatrix} -P^Z_x/P^Z_t \\ -P^Z_y/P^Z_t \\ -P^Z_z/P^Z_t \end{pmatrix} \tag{7.8}$$

  where P^Z_i are the component of the Z four-momentum. The boosted muon momentum $\tilde{P}^\mu$ is obtained with

$$\begin{aligned} \vec{\tilde{P}}^\mu &= \vec{P}^\mu + \frac{\gamma - 1}{\beta^2}(\vec{b} \cdot \vec{P}^\mu)\vec{b} + \gamma P^\mu_t \vec{b} \\ \tilde{P}^\mu_t &= \gamma(P^\mu_t + \vec{b} \cdot \vec{P}^\mu). \end{aligned}$$

  where

$$\gamma = \frac{1}{\sqrt{1-\beta^2}} \tag{7.9}$$

β being the modulus of vector $\vec{b}$. The boosted muons are combined giving

$$\tilde{\vec{P}}^{Z} = \tilde{P}^{\mu+} + \tilde{P}^{\mu-}.$$

- To rescale the four momenta of the boosted muons in order to obtain the invariant mass M_{ph}, we multiplied each component of the boosted muons momenta by a factor
$$\frac{M_{ph}}{M_Z}$$
where M_Z is not the nominal Z mass but the invariant mass of the initial dimuon pair.

- We boost back the muons momenta $\tilde{\vec{P}}^{\mu+}$ and $\tilde{\vec{P}}^{\mu-}$ in the direction:
$$\vec{b}_{ph} = \begin{pmatrix} P_x^Z/\tilde{P}_t^Z \\ P_y^Z/\tilde{P}_t^Z \\ P_z^Z/\tilde{P}_t^Z \end{pmatrix} \tag{7.10}$$
where
$$\tilde{P}_t^Z = \sqrt{(P_x^Z)^2 + (P_y^Z)^2 + (P_z^Z)^2 + M_{ph}^2}.$$
A new pair of muons with momenta $\hat{\vec{P}}^{\mu+}$ and $\hat{\vec{P}}^{\mu-}$ is obtained from this operation.

- The momenta $\hat{\vec{P}}^{\mu+}$ and $\hat{\vec{P}}^{\mu-}$ are used to calculate the asymmetry variables.

We tested this method on the Drell-Yan Monte Carlo samples. In figure 7.19 we show the result of the test of the scaling algorithm for the mass range $10\ GeV < M < 20\ GeV$ for the variable $A(P_T^{\mu}, P_T^{cone})$. The rescaling algorithm is able to reproduce in a quite satisfactory way the asymmetry distribution at low invariant mass. The variable $A(P_T^{cones}, P_T^{rest})$, does not depend on the momenta of the muons and hence rescaling is not necessary. On the other hand, rescaling does not work for the quantity $A(P_T^{sum}, P_T^{rest})$, as can be seen in figure 7.20 where the rescaled distribution does not describe this asymmetry.

As a last test, we performed a check of the agreement of the rescaled Z data, the rescaled Z simulation and the Monte Carlo shapes in each of the mass ranges. We applied the rescaling algorithm to a clean sample of reconstructed Z decay from the LHCb data taking which corresponds approximately to $\sim 16\ pb^{-1}$ of integrated luminosity. In figure 7.21 the data-Monte Carlo comparison is shown. In red the asymmetry is obtained using Monte Carlo dimuon pairs in the different mass ranges while the black dots represent the asymmetry obtained using the rescaling algorithm on a Z Monte Carlo sample. The blue distribution is obtained taking from real data a sample of reconstructed $Z \to \mu\mu$ decay and applying the rescaling algorithm.

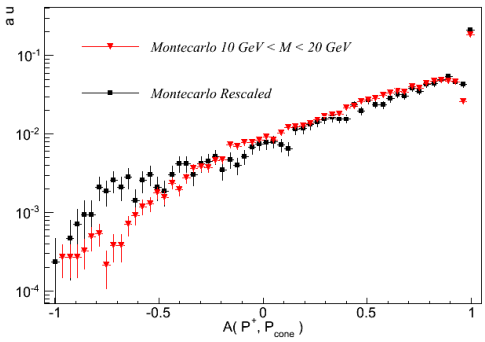


Figure 7.19: Asymmetry distribution $A(P_T^{\mu}, P_T^{cone})$ for the invariant mass range 10 GeV < M < 20 GeV of the Monte Carlo Drell-Yan signal. In red the asymmetry obtained using Monte Carlo dimuon pair in the mass range while the black dots represent the asymmetry obtained using the rescaling algorithm. We rescaled at value of invariant mass M_{ph} = 13 GeV which is the mean value of the mass in that range.

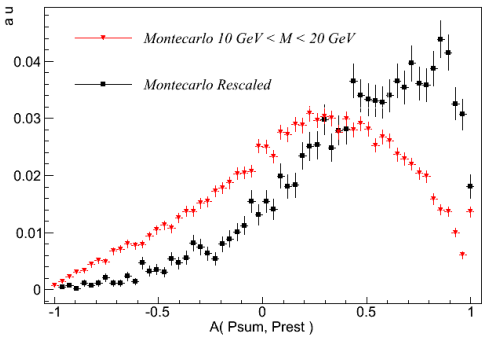


Figure 7.20: Asymmetry distribution $A(P_T^{sum}, P_T^{rest})$ for the invariant mass range 10 GeV < M < 20 GeV of the Monte Carlo Drell-Yan signal. In red the asymmetry obtained using Monte Carlo dimuon pair in the mass range while the black dots represent the asymmetry obtained using the rescaling algorithm. In this case the rescale algorithm is not reliable and not able to reproduce the asymmetry distribution.

7.5 Real Data Analysis

In this section we are going to describe the analysis we performed on real LHCb data. The intention is not to make a cross-section measurement because this is not yet possible with the amount of data that LHC has delivered and the partial understanding of the detector performance. However we want to show the potential of LHCb to make this measurement in the future. The main aim of this analysis is thus to produce the invariant mass distribution for dimuon pairs with LHCb real data and compute the signal and background contributions. We applied the selection criteria described in section 7.3 to events passing the Drell-Yan dedicated trigger lines described in section 2.3.8. For this

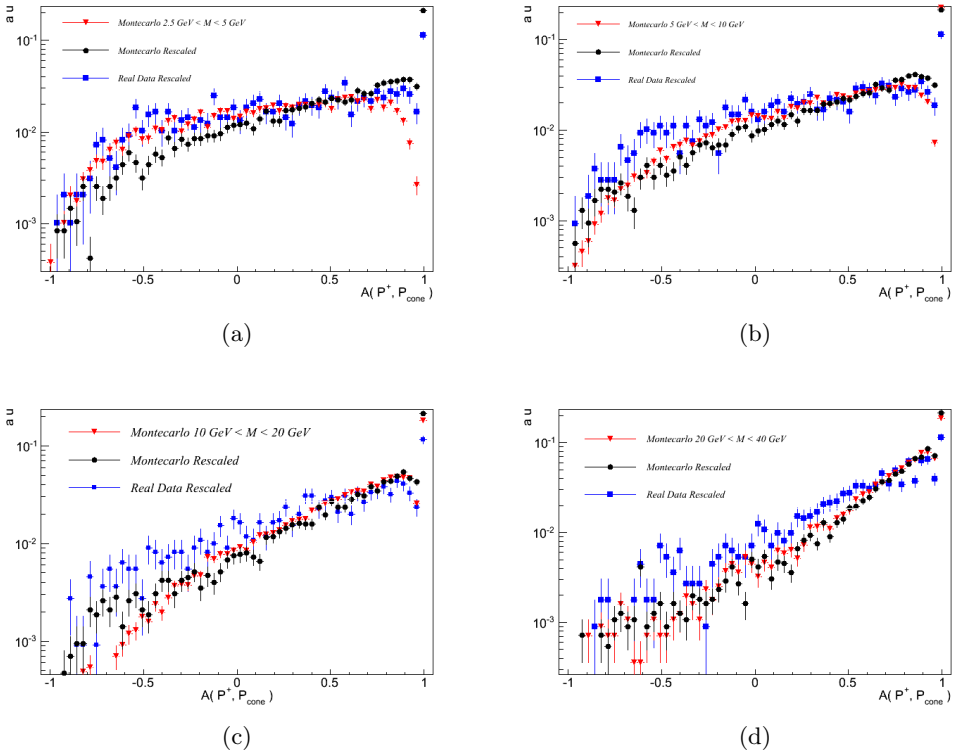


Figure 7.21: Asymmetry distribution $A(P_T^{\mu}, P_T^{cone})$ for the invariant different mass ranges: In blue the distribution rescaled staring from a sample of Z taken from LHCb real data. In red the asymmetry obtained using Monte Carlo dimuon pair in the mass range while the black dots represent the asymmetry obtained using the rescaling algorithm on a Z Monte Carlo sample. The distributions are normalized to unity in order to compare the shapes.

study we required that the events must have been selected by the Hlt1 single muon trigger lines *HLT1SingleMuonNoIPL0*, *HLT1SingleMuonNoIPL0HiPt* and the Hlt2 dimuon lines *Hlt2DiMuonDY1*, *Hlt2DiMuonDY2*, *Hlt2DiMuonDY3*, *Hlt2DiMuonDY4* and *Hlt2DiMuonUnbiasedZmm* which, all together, cover the full invariant mass spectrum from 2.5 GeV without any upper limit.

7.5.1 Dimuon Spectrum

In figure 7.22 we show the raw dimuon invariant mass spectrum selected by these trigger lines and by our offline selection. In black the real data corresponding to an integrated luminosity of about 16 pb^{-1} is plotted. Monte Carlo predictions for the Drell-Yan signal is shown in red, the prediction for B is in blue and the prediction for C is in green. In orange the contribution of K/π mis-identification, which has been obtained from

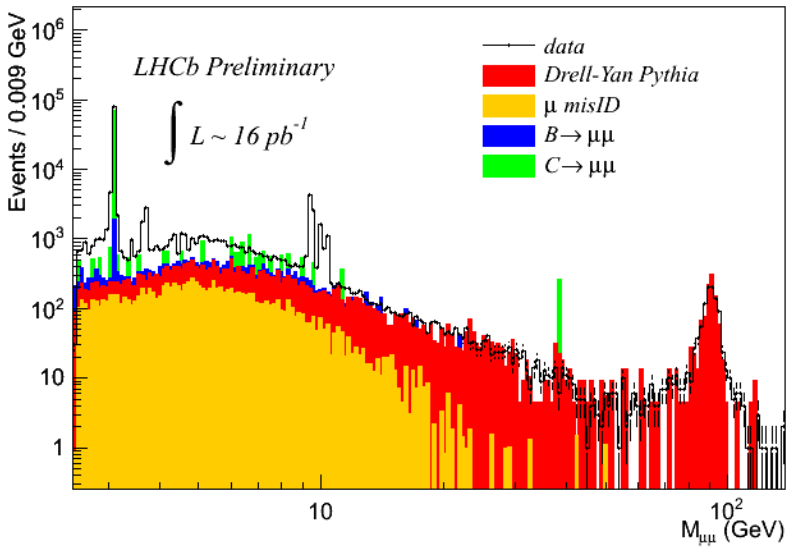


Figure 7.22: Real data: full dimuon spectrum corresponding to an integrated luminosity of about 16 pb^{-1}. Different signal and background contributions to the spectrum are shown with different colors.

data using the procedure outlined in 7.4.1 is plotted. This figure shows the ability of LHCb to reconstruct the full dimuon spectrum but there is no complete understanding of the signal and background contributions. In figure 7.23 restricted mass regions of the spectrum are plotted to show the reconstructed resonances: J/ψ at $M \sim 3.096\ GeV$, $\psi(2S)$ at $M \sim 3.686\ GeV$, $\Upsilon(1S)$, $\Upsilon(2S)$ and $\Upsilon(3S)$ respectively at $M \sim 9.46\ GeV$, $M \sim 10.02\ GeV$ and $M \sim 10.35\ GeV$. Finally at $M \sim 91.1\ GeV$ the peak of the Z boson can be clearly seen.

The agreement of the full dimuon spectrum is not completely satisfactory, in particular the Monte Carlo predictions look smaller that what we see in the data. The first thing we had to check is if the overall contribution of heavy quark decay is correctly described in Monte Carlo and, if not, a correction has to be found. We already know from a previous LHCb measurement [128] that the B cross-sections in the LHC energy regime and in the LHCb forward region it is not consistent with the prediction given by PYTHIA. We now describe how we measured the heavy quark contribution using a data driven method.

7.5.2 Heavy quarks contribution

A simple way to control the heavy quark contribution is to use the lifetime information of the decayed particle. We know that B and D mesons are long lived particles and so, at large values, the impact parameter significance distribution would be dominated by

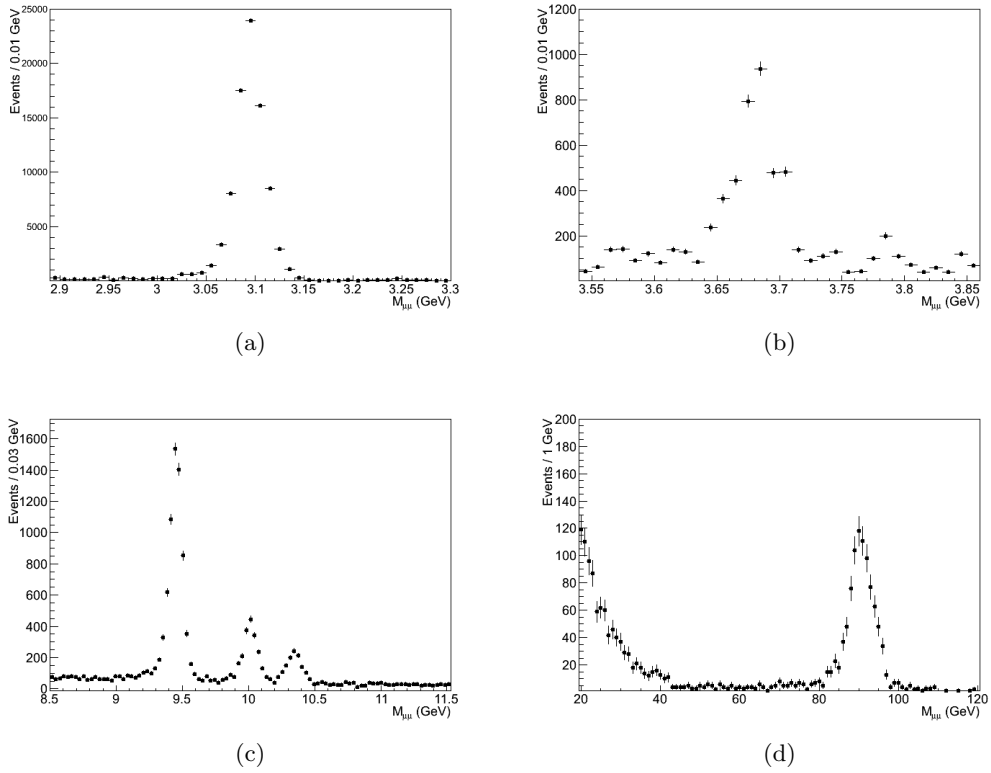


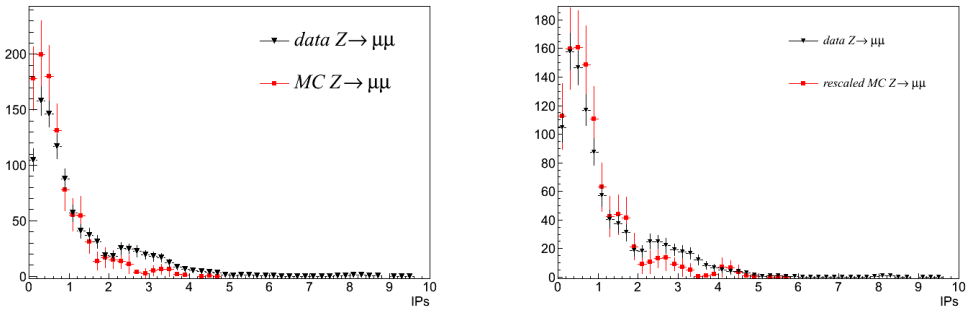
Figure 7.23: Dimuon invariant mass in restricted mass regions selected by the trigger and selections optimized for Drell-Yan events reconstruction. J/ψ (a), $\psi(2S)$ (b), Υ mesons (c) and the Z peak (d).

the heavy quark decays.

Understanding the Impact Parameter Significance

First of all, we need to be sure that the impact parameter significance distribution looks the same in Monte Carlo and data, at least for short living particles. We compared this distribution for Z candidates in data and in Monte Carlo. The distribution is shown in figure 7.24(a) and it is seen that the impact parameter significance (IPs) is wider in data respect to Monte Carlo. A tracking study [129] compared the impact parameter significance in data and Monte Carlo; the result of this study is summarized in figure 7.25 where IPs for minimum bias events in data and Monte Carlo are plotted as a function of $1/P_T$. The ratio of these histograms provide a correction factor (as a function of $1/P_T$) which can be used to rescale the IPs, event by event, to make the IPs distribution in Monte Carlo more like real data. In figure 7.24 we show a comparison of the IPs in data and Monte Carlo before and after we smeared. On average, for muons coming from a

Z, the IPs is 27% larger in data, with respect to the Monte Carlo.



(a) (b)

Figure 7.24: Comparison of the Impact Parameter Significance in data (black) and in Monte Carlo (red) without (a) and with (b) a rescaling factor.

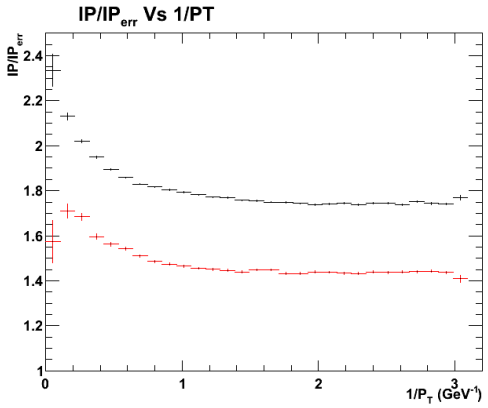


Figure 7.25: Comparison of the impact parameter significance in data (black) and Monte Carlo (red) as function of $1/P_T$ obtained using a minimum bias sample.

B and C cross-section

Due to the long lived nature of heavy quark decays, if we require that at least one of the muon has impact parameter significance larger than 7 we expect to obtain a sample of events which is completely made of heavy quark decay without any Drell-Yan contamination. Given this assumption we check that the IPs distribution above 7 that we measure in the data has the same number of events predicted by the Monte Carlo. Here, instead of using the cross-section predicted by PYTHIA, we used the LHCb

measurement for b production cross-section which is [128]:

$$\sigma(pp \to b\bar{b}X) = (284 \pm 20 \pm 49)\mu b \tag{7.11}$$

With this cross-section, we expect, in 16 pb^{-1}, $N_B^{>7} = 1980$ events[7] passing the selection cuts with IPs larger than 7. The same calculation for the *inclusive C* sample gives $N_C^{>7} = 6005$ events passing the selection cuts with IPs larger than 7. In our data we observe $N_{data}^{>7} = 14686$ events of this kind. Given that the B cross-section is measured by LHCb with 18% accuracy while the theoretical uncertainties on the c production can be very large, we look for a correction factor that is necessary to apply to the *inclusive C* Monte Carlo prediction in order to obtain the correct normalization of the IPs distribution above 7. Given that the number of events in the data has to be the sum of the number of heavy quark events predicted:

$$N_{data}^{>7} = N_B^{>7} + k_c N_C^{>7} \tag{7.12}$$

we obtain for the correction factor $k_c = 2.1$, which means that in data the $pp \to c\bar{c}X$ cross-section is 2.1 times larger than the PYTHIA prediction. In figure 7.26 we show the impact parameter distribution where the C contribution has been scaled by a factor 2.1.

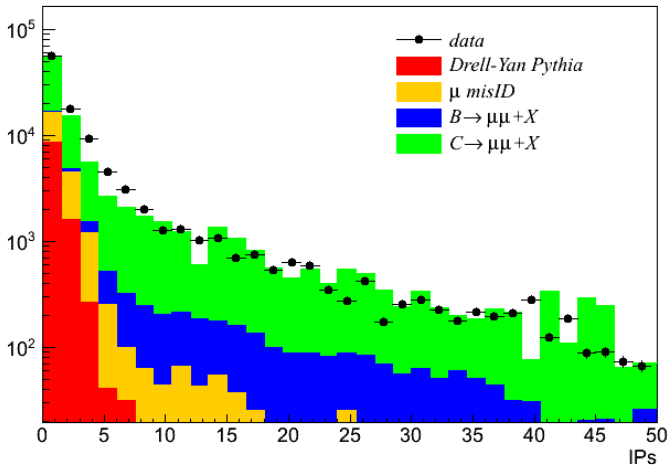


Figure 7.26: Impact parameter significance distribution for data and Monte Carlo predictions. The *inclusive C* Monte Carlo predictions is scaled up by a factor 2.11.

[7]The apex $^{>7}$ stands for $IPs > 7$ and $^{<5}$ for $IPs < 5$

Heavy quark mass spectrum

Since the number of events of the Monte Carlo sample passing the selection cuts is very small (38 events in the inclusive B and 95 in the inclusive C sample), there is not enough statistics to describe accurately the shape of the full mass spectrum. We need to build this from data. We enrich the sample with heavy quark decays by requiring that at least one muon has $IPs > 7$. This gives the shape of the invariant mass distribution for heavy quark decays. It is normalized to the expected number of heavy quarks decays in our signal as follows:

We calculate the number of event expected in 16 pb^{-1} of integrated luminosity, with IPs smaller than five. We obtained for the *inclusive B* sample $N_B^{<5} = 731$ and for the *inclusive C* sample $N_B^{<5} = 8349$. Here we considered the cross-section for $b\bar{b}X$ measured by LHCb and $c\bar{c}X$ multiplied by the factor 2.11 obtained in the previous section. The total number of heavy quark decays expected in 16 pb^{-1} with IPs larger than 7 is:

$$N_{HQ}^{>7} = N_B^{>7} + k_c N_C^{>7} \sim 14686$$

while the number of expected heavy quark decays with IPs smaller than 5 is:

$$N_{HQ}^{<5} = N_B^{<5} + N_C^{<5} \sim 9080$$

With this information we are able to calculate the fraction of heavy quarks events with $IPs > 7$ with respect to the number of events with $IPs < 5$:

$$K_{HQ} = \frac{N_{HQ}^{>7}}{N_{HQ}^{<5}}.$$

7.5.3 Like Sign Combinations

Another way to estimate the hadron mis-identification background, is to search for combinations of muons with same charge (i.e $\mu^+\mu^+$ and $\mu^-\mu^-$). The only processes underlying the like-sign muon production is the doubly semileptonic decays of B mesons, so all the other combinations of this kind are due to hadron mis-Identification. Since the probability for combining two opposite sign mis-identified muons is the same as that for combining two same sign mis-identified muons, we expect that the like-signal invariant mass distribution looks similar to the mis-Id except for the contribution at low invariant mass of the double semileptonic decay.
Furthermore, we need to consider that at low invariant mass the J/ψ resonance contributes with a large amount of real muons that can be combined together with another same sign mis-identified muon to produce a large mis-ID contribution. Using the *inclusive C* Monte Carlo sample we evaluated the expected number of this source of back-

ground making like-sign combinations where the combination is made by a real muon from the $J/\Psi \rightarrow \mu\mu$ decay or from a semileptonic decay $c \rightarrow \mu + X$ and the other one is a mis-identified muon. This allows us to estimate the mis-ID background when only one of the muon is mis-identified. In 16 pb^{-1} we expect about 12506 of this kind of events. We only considered the *inclusive C* sample because it is a larger source of single muons and dimuons from J/ψ than the *inclusive B* sample.

In figure 7.27, we show the expected dimuon invariant mass for the mis-ID backgrounds (i.e. when both muons and only one muon is mis-identified) and overlay the same distribution obtained using like-sign muons. The two distributions have been normalised and their consistency gives us a confirmation that the muon mis-identification background is correctly estimated.

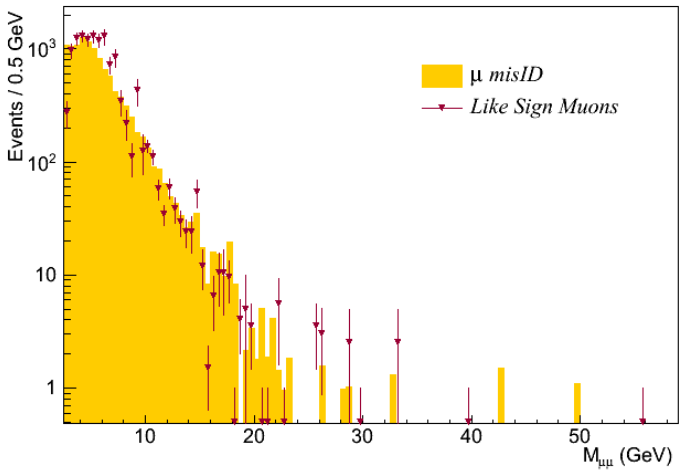


Figure 7.27: Like Sign dimuon spectrum compared with the Mis-ID distribution.

7.5.4 Dimuon Invariant Mass Spectrum

Finally, in figure 7.28 we show the dimuon spectrum which has been built using the procedure outlined in the previous section. A much better agreement is now obtained between data and simulation. With the preliminary studies we have performed, our understanding of the signal and background contribution on the full invariant mass spectrum is quite good. However, the overall agreement at low invariant mass is not perfect since we observe more events than expected. This may be due to the description of the hadron mis-identification which is the largest contribution at low masses. It may also be due to the theoretical description of the Drell-Yan process as described in section 4.11. However more work is needed to suppress the backgrounds before data can be compared to theory in this region.

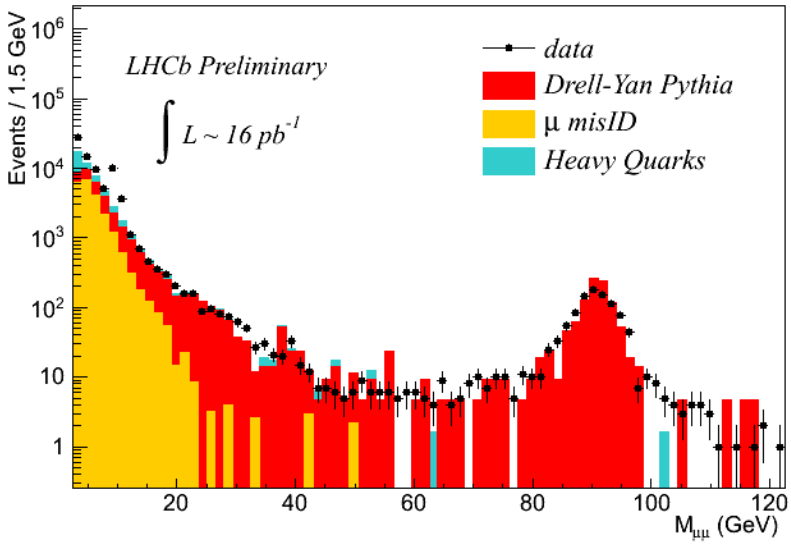


Figure 7.28: Dimuon spectrum where data and backgrounds expectation.

A. PDF CONSTRAINT FOR DIFFERENT PDF SETS

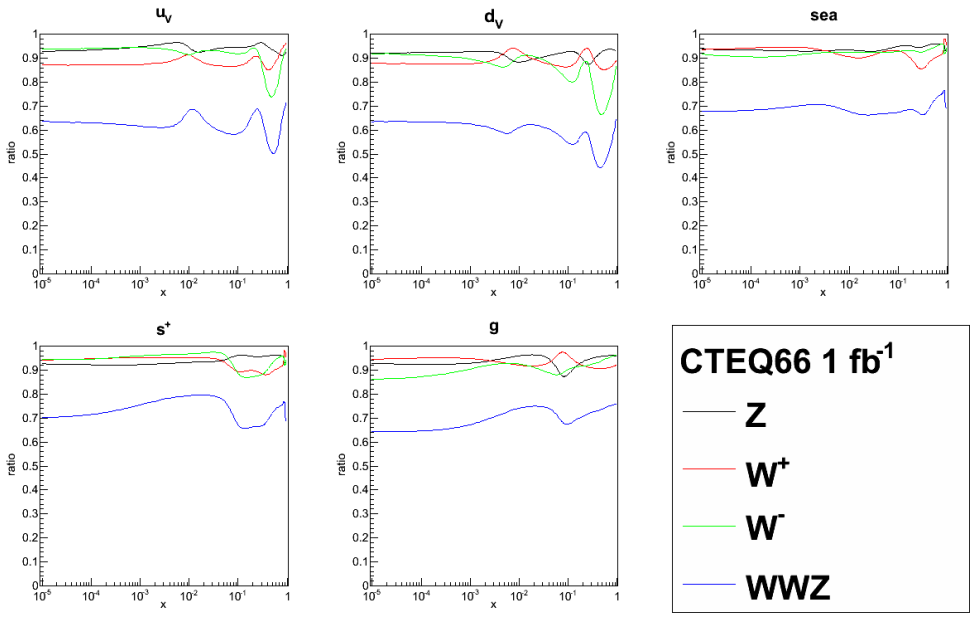


Figure A.1: Improvement when fitting W^+, W^- Z rapidity distributions and a combined fit of the three using 1 fb^{-1} of LHCb data for the PDF set CTEQ66.

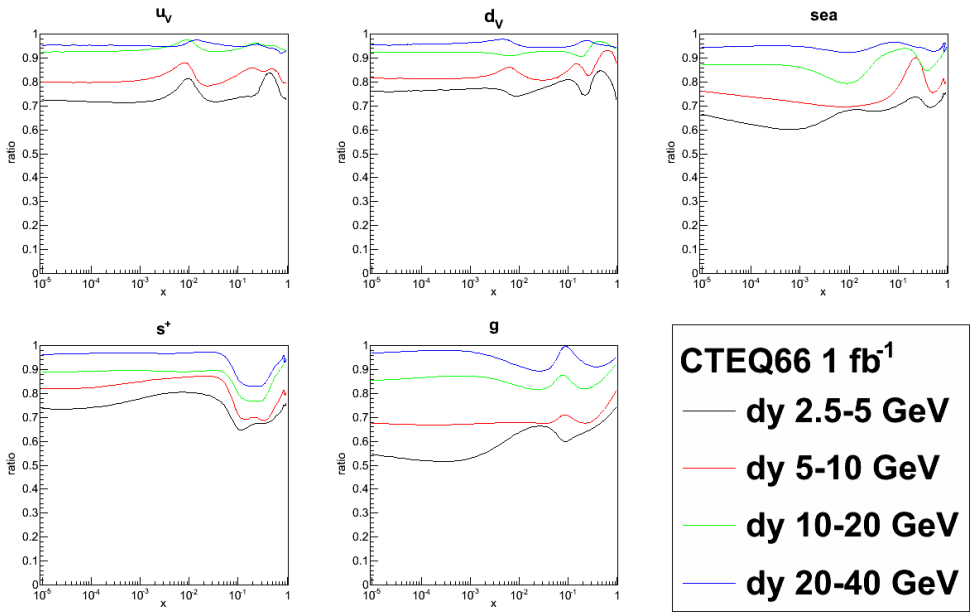


Figure A.2: Improvement when fitting Drell-Yan rapidity distributions in 4 different mass bins using 1 fb^{-1} of LHCb data for the PDF set CTEQ66.

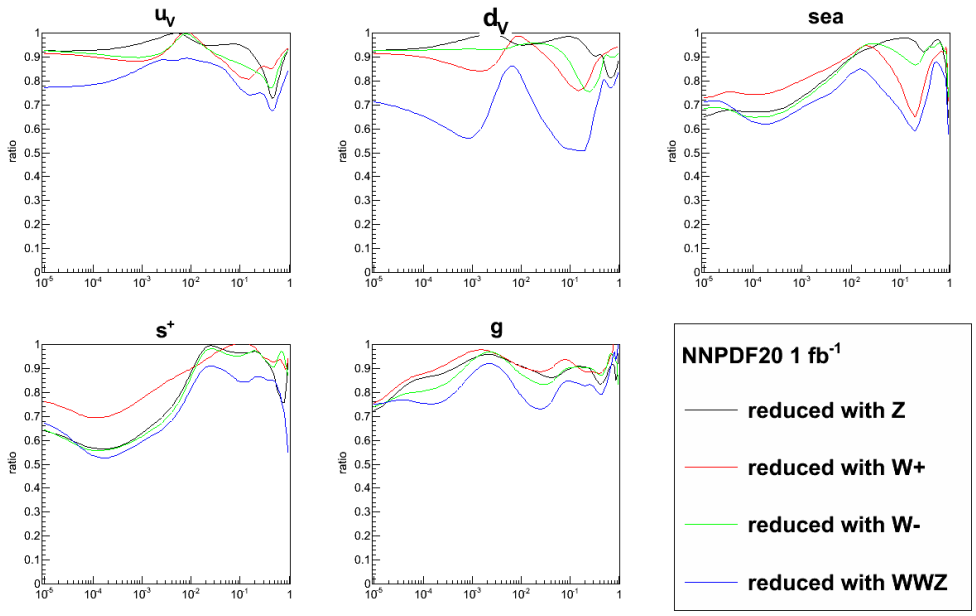


Figure A.3: Improvement when fitting W^+, W^- Z rapidity distributions and a combined fit of the three using 1 fb^{-1} of LHCb data for the PDF set NNPDF2.0

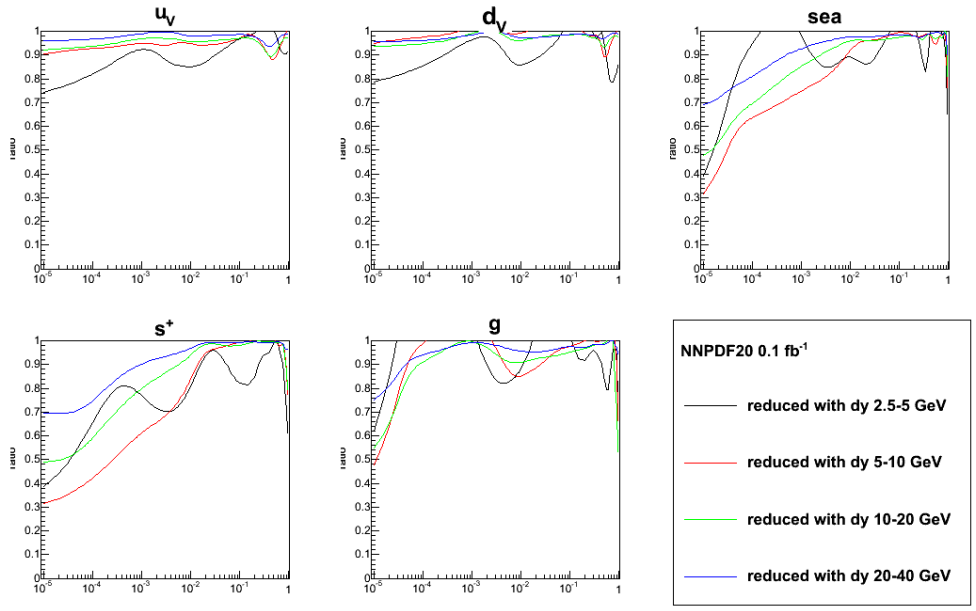


Figure A.4: Improvement when fitting Drell-Yan rapidity distributions in 4 different mass bins using 0.1 fb^{-1} of LHCb data for the PDF set NNPDF2.0

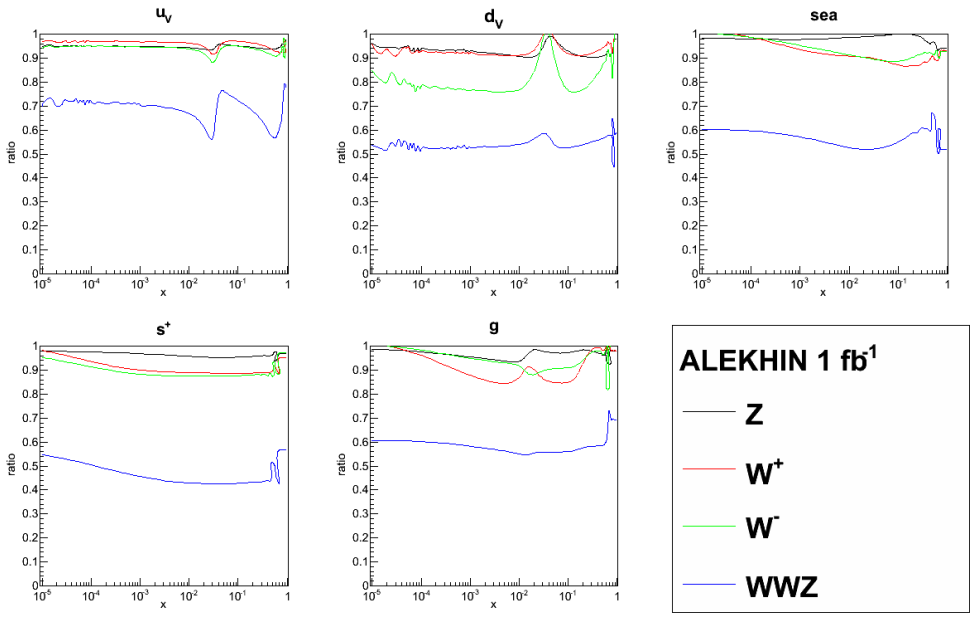


Figure A.5: Improvement when fitting W^+, W^- Z rapidity distributions and a combined fit of the three using 1 fb^{-1} of LHCb data for the PDF set ALEKHIN.

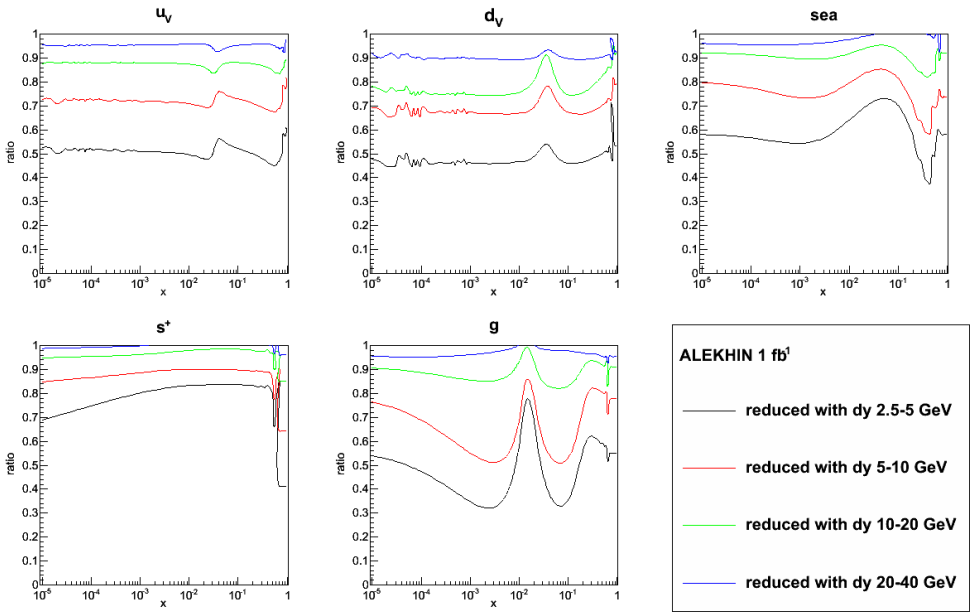


Figure A.6: Improvement when fitting Drell-Yan rapidity distributions in 4 different mass bins using 1 fb^{-1} of LHCb data for the PDF set ALEKHIN.

BIBLIOGRAPHY

[1] Lyndon Evans and Philip Bryant. LHC machine. *Journal of Instrumentation*, 3(08):S08001, 2008.

[2] The ATLAS Collaboration. The ATLAS experiment. *Journal of Instrumentation*, 3(08):S08003, 2008.

[3] The CMS Collaboration. The CMS experiment. *Journal of Instrumentation*, 3(08):S08004, 2008.

[4] The LHCb Collaboration. The LHCb experiment. *Journal of Instrumentation*, 3(08):S08005, 2008.

[5] The Alice Collaboration. The Alice experiment. *Journal of Instrumentation*, 3(08):S08002, 2008.

[6] *CERN - the European Organization for Nuclear Research, http://www.cern.ch* .

[7] M Ferro-Luzzi. Proposal for an absolute luminosity determination in colliding beam experiments using vertex detection of beam-gas interactions. *Nucl. Instrum. Methods Phys. Res., A*, 553(CERN-PH-EP-2005-023. 3):388–399. 17 p, May 2005.

[8] LHCb Collaboration. *LHCb VELO (VErtex LOcator): Technical Design Report.* Technical Design Report LHCb. CERN, Geneva, 2001.

[9] M L McCubbin. Optimising the LHCb VELO detector. *Nucl. Instrum. Methods Phys. Res., A*, 473(1-2):163–166, 2001.

[10] M Krasowski, M Kucharczyk, W Manner, G Polok, and M Witek. Primary vertex reconstruction. Technical Report LHCb-2007-011. CERN-LHCb-2007-011, CERN, Geneva, Sep 2007.

[11] LHCb Collaboration. *LHCb magnet: Technical Design Report.* Technical Design Report LHCb. CERN, Geneva, 1999.

[12] LHCb Collaborations. *LHCb reoptimized detector design and performance: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2003.

[13] LHCb Collaboration. *LHCb inner tracker: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2002.

[14] LHCb Collaboration. *LHCb outer tracker: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2001.

[15] O Callot and S Hansmann-Menzemer. The forward tracking: Algorithm and performance studies. Technical Report LHCb-2007-015. CERN-LHCb-2007-015, CERN, Geneva, May 2007.

[16] O Callot, M Kucharczyk, and M Witek. Velo-tt track reconstruction. Technical Report LHCb-2007-010. CERN-LHCb-2007-010, CERN, Geneva, Apr 2007.

[17] LHCb Collaboration. *LHCb RICH: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2000.

[18] LHCb Collaboration. *LHCb calorimeters: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2000.

[19] LHCb Collaboration. *LHCb muon system: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2001.

[20] M Kachtchouk et al. Design and construction of the wire chambers for the LHCb muon system. Technical Report LHCb-2001-026, CERN, Geneva, Dec 2001.

[21] LHCb Collaboration. *LHCb muon system: second addendum to the Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2005. Submitted on 9 Apr 2005.

[22] G Lanfranchi, X Cid Vidal, S Furcas, M Gandelman, J A Hernando, J H Lopez, E Polycarpo, and A Sarti. The muon identification procedure of the lhcb experiment for the first data. Technical Report LHCb-PUB-2009-013. CERN-LHCb-PUB-2009-013, CERN, Geneva, Aug 2009.

[23] LHCb Collaboration. *LHCb trigger system: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2003. revised version number 1 submitted on 2003-09-24 12:12:22.

[24] S Amato, A Satta, B Souza de Paula, and L De Paula. Hlt1 muon alley description. Technical Report LHCb-2008-058. CERN-LHCb-2008-058, CERN, Geneva, Nov 2008.

[25] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth.*, A462:152–155, 2001.

[26] J. Allison et al. Geant4 developments and applications. *IEEE Transactions on Nuclear Science*, 53:270–278, February 2006.

[27] I.J.R. Aitchison and A.J.G. Hey. *Gauge Theories in Particle Physics, Volume I: From Relativistic Quantum Mechanics to QED (Graduate Student Series in Physics)*. Taylor & Francis, 2002.

[28] Yoichiro Nambu. Axial vector current conservation in weak interactions. *Phys. Rev. Lett.*, 4(7):380–382, Apr 1960.

[29] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19(21):1264–1266, Nov 1967.

[30] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory*. Addison-Wesley, Reading, 1995.

[31] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13(16):508–509, Oct 1964.

[32] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13(9):321–323, Aug 1964.

[33] Francis Halzen and Alan D. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. Wiley, 1984.

[34] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak interactions with lepton-hadron symmetry. *Phys. Rev. D*, 2(7):1285–1292, Oct 1970.

[35] R. K. Ellis, W. J. Stirling, and B. R. Webber. *QCD and Collider Physics (Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology)*. Cambridge University Press, 2003.

[36] Sigurdur Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces (Graduate Studies in Mathematics)*. American Mathematical Society, 2001.

[37] Murray Gell-Mann. Symmetries of baryons and mesons. *Phys. Rev.*, 125(3):1067–1084, Feb 1962.

[38] David J. Gross and Frank Wilczek. Asymptotically free gauge theories. i. *Phys. Rev. D*, 8(10):3633–3652, Nov 1973.

[39] David J. Gross and Frank Wilczek. Ultraviolet behavior of non-abelian gauge theories. *Phys. Rev. Lett.*, 30(26):1343–1346, Jun 1973.

[40] H. David Politzer. Asymptotic freedom: An approach to strong interactions. *Physics Reports*, 14(4):129 – 180, 1974.

[41] Siegfried Bethke. The 2009 world average of α_s. *The European Physical Journal C - Particles and Fields*, 64:689–703, 2009.

[42] J. D. Bjorken. Asymptotic Sum Rules at Infinite Momentum. *Phys. Rev.*, 179:1547–1553, 1969.

[43] G. Parisi. Bjorken scaling and the parton model. *Physics Letters B*, 42(1):114 – 116, 1972.

[44] J I Friedman and H W Kendall. Deep inelastic electron scattering. *Annual Review of Nuclear Science*, 22(1):203–254, 1972.

[45] H1 and Zeus Collaborations. Combined measurement and qcd analysis of the inclusive $e^{\pm}p$ scattering cross sections at hera. *Journal of High Energy Physics*, 2010:1–63, 2010. 10.1007/JHEP01(2010)109.

[46] Ian Hinchliffe and Axel Kwiatkowski. Parton-model sum rules1. *Annual Review of Nuclear and Particle Science*, 46(1):609–645, 1996.

[47] *Measurement of the fractional momentum carried by the gluons in the low Q squared photon-gluon fusion events at the ep HERA collider*, July 1989.

[48] Sau Lan Wu. e^+e^- physics at petra–the first five years. *Physics Reports*, 107(2-5):59 – 324, 1984.

[49] K Nakamura and Particle Data Group. Review of particle physics. *Journal of Physics G: Nuclear and Particle Physics*, 37(7A):075021, 2010.

[50] TASSO Collaboration. Evidence for planar events in e+e- annihilation at high energies. *Physics Letters B*, 86(2):243 – 249, 1979.

[51] PETRA collaboration. Discovery of three-jet events and a test of quantum chromodynamics at petra. *Phys. Rev. Lett.*, 43(12):830–833, 1979.

[52] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys.*, B126:298, 1977.

[53] V. N. Gribov and L. N. Lipatov. Deep inelastic e p scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15:438–450, 1972.

[54] Fabio Maltoni, Thomas McElmurry, Robert Putman, and Scott Willenbrock. Choosing the factorization scale in perturbative qcd. 2007.

[55] Sidney D. Drell and Tung-Mow Yan. Partons and their applications at high energies. *Annals of Physics*, 66(2):578 – 623, 1971.

[56] Marzani S. *High Energy Resummation in Quantum Chromo-Dynamics.* PhD thesis, University of Edimburgh, Edimburgh, 2009.

[57] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. Pythia 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026, 2006.

[58] Marchesini et al. Herwig: A monte carlo event generator for simulating hadronemission reactions with interfering gluons. *Comput. Phys. Commun.*, 67, 1992.

[59] Thomas D. Gottschalk. An improved description of hadronization in the qcd cluster model for e+e- annihilation. *Nuclear Physics B*, 239(2):349 – 381, 1984.

[60] G.J. Alner et al. The ua5 high energy simulation program. *Nuclear Physics B*, 291:445 – 502, 1987.

[61] T. Aaltonen et al. Studying the Underlying Event in Drell-Yan and High Transverse Momentum Jet Production at the Tevatron. *Phys. Rev.*, D82:034001, 2010.

[62] D. Acosta et al. Soft and hard interactions in $p\bar{p}$ collisions at $\sqrt{s} = 1800$ and 630 gev. *Phys. Rev. D*, 65(7):072005, Apr 2002.

[63] J. Butterworth, J. Forshaw, and M. Seymour. Multiparton interactions in photoproduction at HERA. *Zeitschrift fur Physik C Particles and Fields*, 72:637–646, 1996.

[64] George Sterman. Summation of large corrections to short-distance hadronic cross sections. *Nuclear Physics B*, 281(1-2):310 – 364, 1987.

[65] S. Catani and L. Trentadue. Resummation of the qcd perturbative series for hard processes. *Nuclear Physics B*, 327(2):323 – 352, 1989.

[66] Stefano Forte and Giovanni Ridolfi. Renormalization group approach to soft gluon resummation. *Nuclear Physics B*, 650(1-2):229 – 270, 2003.

[67] A De Roeck and H Jung. HERA and the LHC : A Workshop on the Implications of HERA for LHC Physics. Resummation. Geneva, 2005.

[68] V. S. Fadin, E. A. Kuraev, and L. N. Lipatov. On the pomeranchuk singularity in asymptotically free theories. *Physics Letters B*, 60(1):50 – 52, 1975.

[69] George Sterman et al. Handbook of perturbative qcd. *Rev. Mod. Phys.*, 67(1):157–248, Jan 1995.

[70] R Vogt. What is the real k factor? *Acta Phys. Hung. New Ser. Heavy Ion Phys.*, 17(hep-ph/0207359. LBNL-51206):75. 25 p, Jul 2002.

[71] A.D. Martin, R.G. Roberts, W.J. Stirling, and R.S. Thorne. Uncertainties of predictions from parton distributions I: Experimental errors. *The European Physical Journal C - Particles and Fields*, 28:455–473, 2003.

[72] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur.Phys.*, C63:189–285, 2009.

[73] J. Pumplin at al. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP07*, 012, 2002.

[74] S. I. Alekhin. Parton distributions from deep-inelastic-scattering data. *Phys. Rev. D*, 68(1):014002, Jul 2003.

[75] The H1 and ZEUS Collaborations. Combined measurement and QCD analysis of the inclusive *ep* scattering cross sections at HERA. *Journal of High Energy Physics*, 2010:1–63, 2010.

[76] J. Pumplin at al. *Phys Rev*, D65:014013, 2002.

[77] J M Campbell, J W Huston, and W J Stirling. Hard interactions of quarks and gluons: a primer for lhc physics. *Reports on Progress in Physics*, 70(1):89, 2007.

[78] R.S. Thorne et al. *J. Phys.*, G28:2717, 2002.

[79] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung. Uncertainties of predictions from parton distribution functions. I. the lagrange multiplier method. *Phys. Rev. D*, 65(1):014012, Dec 2001.

[80] Richard D. Ball, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, José I. Latorre, Andrea Piccione, Juan Rojo, and Maria Ubiali. A determination of parton distributions with faithful uncertainty estimation. *Nuclear Physics B*, 809(1-2):1 – 63, 2009.

[81] The NNPDF Collaboration, Luigi Del Debbio, Stefano Forte, José I. Latorre, Andrea Piccione, and Joan Rojo. Neural network determination of parton distributions: the nonsinglet case. *Journal of High Energy Physics*, 2007(03):039, 2007.

[82] BCDMS Collaboration. *Phys. Lett*, B223(485), 1989.

[83] BCDMS Collaboration. *Phys. Lett*, B237(592), 1990.

[84] M. Arneodo et al. Measurement of the proton and deuteron structure functions, F2(p) and F2(d), and of the ratio sigma(L)/sigma(T). *Nucl. Phys.*, B483:3–43, 1997.

[85] M. Arneodo et al. Accurate measurement of F2(d)/F2(p) and R(d)-R(p). *Nucl. Phys.*, B487:3–26, 1997.

[86] M. R. Adams et al. Extraction of the ratio f_2^n/f_2^p from muon-deuteron and muon-proton scattering at small x and q^2. *Phys. Rev. Lett.*, 75(8):1466–1470, Aug 1995.

[87] C. Adloff et al. Deep-inelastic inclusive e p scattering at low x and a determination of alpha(s). *Eur. Phys. J.*, C21:33–61, 2001.

[88] C. Adloff et al. Measurement and QCD analysis of neutral and charged current cross sections at HERA. *Eur. Phys. J.*, C30:1–32, 2003.

[89] C. Adloff et al. Measurement of neutral and charged current cross-sections in positron proton collisions at large momentum transfer. *Eur. Phys. J.*, C13:609–639, 2000.

[90] C. Adloff et al. Measurement of neutral and charged current cross-sections in electron - proton collisions at high Q^2. *Eur. Phys. J.*, C19:269–288, 2001.

[91] S. Chekanov et al. Measurement of the neutral current cross section and F2 structure function for deep inelastic e+ p scattering at HERA. *Eur. Phys. J.*, C21:443–471, 2001.

[92] J. Breitweg et al. Measurement of high-Q**2 charged-current e+ p deep inelastic scattering cross sections at HERA. *Eur. Phys. J.*, C12:411–428, 2000.

[93] S. Chekanov et al. Measurement of high-Q**2 e- p neutral current cross sections at HERA and the extraction of xF3. *Eur. Phys. J.*, C28:175, 2003.

[94] S. Chekanov et al. High-Q**2 neutral current cross sections in e+ p deep inelastic scattering at s**(1/2) = 318-GeV. *Phys. Rev.*, D70:052001, 2004.

[95] S. Chekanov et al. Measurement of high-Q**2 charged current cross sections in e+ p deep inelastic scattering at HERA. *Eur. Phys. J.*, C32:1–16, 2003.

[96] F.D. Aaron et al. Measurement of the proton structure function fl(x,q2) at low x. *Physics Letters B*, 665(4):139 – 146, 2008.

[97] J. T. Londergan and A. W. Thomas. Additional corrections to the gross–llewellyn smith sum rule. *Phys. Rev. D*, 82(11):113001, Dec 2010.

[98] Richard D. Ball, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, José I. Latorre, Andrea Piccione, Juan Rojo, and Maria Ubiali. Precision determination of electroweak parameters and the strange content of the proton from neutrino deep-inelastic scattering. *Nuclear Physics B*, 823(1-2):195 – 233, 2009.

[99] David Alexander Mason. *Measurement of the strange - antistrange asymmetry at NLO in QCD from NuTeV dimuon data.* PhD thesis. FERMILAB-THESIS-2006-01.

[100] G. Moreno et al. Dimuon production in proton-copper collisions at $\sqrt{s} = 38.8$ gev. *Phys. Rev. D*, 43(9):2815–2835, May 1991.

[101] J. C. Webb et al. Absolute Drell-Yan dimuon cross sections in 800-GeV/c p p and p d collisions. 2003.

[102] J. C. Webb. *Measurement Of Continuum Dimuon Production In 800-GeV/C Proton-Nucleon Collisions.* PhD thesis. (2003), hep-ex/0301031.

[103] R. S. Towell et al. Improved measurement of the anti-d/anti-u asymmetry in the nucleon sea. *Phys. Rev.*, D64:052002, 2001.

[104] V. M. Abazov et al. Measurement of the shape of the boson rapidity distribution for $p\bar{p} \to Z/gamma^* \to e^+e^- + X$ events produced at $\sqrt{s}$ of 1.96-TeV. *Phys. Rev.*, D76:012003, 2007.

[105] Timo Antero Aaltonen et al. Measurement of $d\sigma/dy$ of Drell-Yan e^+e^- pairs in the Z Mass Region from $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Lett.*, B692:232–239, 2010.

[106] T. Aaltonen et al. Direct measurement of the w production charge asymmetry in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ *tev*. *Phys. Rev. Lett.*, 102(18):181801, May 2009.

[107] T. Aaltonen et al. Measurement of the inclusive jet cross section at the fermilab tevatron $p\bar{p}$ collider using a cone-based jet algorithm. *Phys. Rev. D*, 78(5):052006, Sep 2008.

[108] A. Abulencia et al. Measurement of the Inclusive Jet Cross Section using the $\boldsymbol{k_T}$ algorithm in $\boldsymbol{p\overline{p}}$ Collisions at $\boldsymbol{\sqrt{s}} = 1.96$ TeV with the CDF II Detector. *Phys. Rev.*, D75:092006, 2007.

[109] G Watt. *MSTW PDFs : PDF benchmarking for LHC processes, http://projects.hepforge.org/mstwpdf/pdf4lhc/* .

[110] Sergey Alekhin et al. The pdf4lhc working group interim report. Technical Report arXiv:1101.0536, Jan 2011.

[111] J. Campbell and K. Ellis. *MCFM Monte Carlo for FeMtobarn processes, http://mcfm.fnal.gov* .

[112] R.S. Thorne, A.D. Martin, W.J. Stirling, and G. Watt. Parton distributions and QCD at LHCb. In *Proceedings of the 16th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2008)*, London, 2008.

[113] Walter T. Giele and Stephane Keller. Implications of hadron collider observables on parton distribution function uncertainties. *Phys. Rev. D*, 58(9):094023, Oct 1998.

[114] Richard D. Ball, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, José I. Latorre, Andrea Piccione, Juan Rojo, and Maria Ubiali. Reweighting NNPDFs: the w lepton asymmetry. *arXiv:1012.0836v2*, 2010.

[115] Ubiali M. *A new approach in the determination of Parton Distribution Functions. Application to processes with heavy quarks in the initial state.* PhD thesis, University of Edimburgh, Edimburgh, 2010.

[116] J S Anderson and R McNulty. *Testing the electroweak sector and determining the absolute luminosity at LHCb using dimuon final states.* PhD thesis, University College Dublin, Dublin, 2008.

[117] J Keaveney and J Anderson. Measurements of muon identification efficiencies for z$\rightarrow \mu\mu$ and w $\rightarrow \mu\nu_\mu$ decays. Technical Report LHCb-INT-2011-001. CERN-LHCb-INT-2011-001, CERN, Geneva, Jan 2011.

[118] S Bifani, R McNulty, and T Shears. Preliminary analysis and identification of candidate events in the decay chain w$\rightarrow \mu\nu$. Technical Report LHCb-INT-2010-027. CERN-LHCb-INT-2010-027, CERN, Geneva, Jun 2010.

[119] J. Anderson. Probing low-x with drell-yan events at LHCb. In *Proceedings of the 17th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2009)*, Madrid, 2009.

[120] H. L. Lai et al. Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions. *Eur. Phys. J.*, C12:375–392, 2000.

[121] M. R. Whalley, D. Bourilkov, and R. C. Group. The Les Houches Accord PDFs (LHAPDF) and Lhaglue. 2005.

[122] P Nason et al. Bottom production. 1999.

[123] Stefano Frixione, Michelangelo L. Mangano, Paolo Nason, and Giovanni Ridolfi. Charm and bottom production: theoretical results versus experimental data. *Nuclear Physics B*, 431(3):453 – 483, 1994.

[124] Michelangelo L. Mangano, Paolo Nason, and Giovanni Ridolfi. Heavy-quark correlations in hadron collisions at next-to-leading order. *Nuclear Physics B*, 373(2):295 – 345, 1992.

[125] RD5 Collaboration. Measurement of hadron shower punchthrough in iron. *Zeitschrift fur Physik C Particles and Fields*, 60:1–10, 1993.

[126] LHCb electroweak group. W and Z cross-section measurements. Technical report, CERN, Geneva.

[127] K Hennesy and R McNulty. *Characterisation and commissioning of the Vertex Locator and determining the cross-section for Drell-Yan produced di-muons at LHCb.* PhD thesis, University College Dublin, Dublin, 2010.

[128] LHCb Collaboration. Measurement of $\sigma(pp \rightarrow b\bar{b}X$ at $\sqrt{s} = 7\ TeV$ in the forward region. *Physics Letters B*, 694(3):209 – 216, 2010.

[129] M Gersabeck. Lhcb tracking, alignment and physics performance. Dec 2010. LHCb-TALK-2010-052.