W Cross Section Measurement in the Electron Channel in pp Collisions at $\sqrt{s}$= 7 TeV

by

Pieter Bruno Bart Everaerts

M.S., Universiteit Gent (2006)

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2011

Author ...
Department of Physics
May 20, 2011

Certified by ...
Prof. Steven C. Nahn
Associate Professor
Thesis Supervisor

Accepted by ...
Prof. Krishna Rajagopal
Associate Department Head for Education

W Cross Section Measurement in the Electron Channel in pp Collisions at $\sqrt{s}$= 7 TeV

by

Pieter Bruno Bart Everaerts

Submitted to the Department of Physics
on May 20, 2011, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

From March until November 2010 the Compact Muon Solenoid experiment recorded 36 pb^{-1} of pp collisions at $\sqrt{s} = 7$ TeV. One of the first precision tests of the Standard Model that can be performed with this data is the measurement of the W-production cross section and the charge asymmetry in the cross section. In this thesis, both measurements are performed in the electron decay channel. The results obtained are:

$$\begin{aligned} \sigma(W \to e\nu) &= 10.48 \pm 0.03(stat.) \pm 0.15(syst.) \pm 0.09(th.) \pm 0.42(lumi.)\text{nb} \\ \frac{\sigma(W^+ \to e^+\nu)}{\sigma(W^- \to e^-\bar{\nu})} &= 1.430 \pm 0.008(stat.) \pm 0.022(syst.) \pm 0.029(th.) \end{aligned}$$

The measurements agree with state-of-the-art NNLO QCD calculations with the latest parton distribution functions.

Thesis Supervisor: Prof. Steven C. Nahn
Title: Associate Professor

Acknowledgments

I think physicists are the Peter Pans of the human race. They never grow up and they keep their curiosity.

I love this quote of Isidor Rabi, because it reminds me of my quest in the world of particle physics. My PhD was a long and exciting adventure of challenges, possible pitfalls and interesting discoveries. But the main reason I made it to the end, was not due to my own cunningness or wit, but thanks to help of other people. Just like Peter Pan had to rely on the Lost Boys, I had to rely on bosses, colleagues, friends and family and I want to thank them all for their help.

First of all, I want to thank Steve, my supervisor. I want to thank him for taking me as a student, for his eternal enthusiasm, his love of hardware and the lessons he taught me. Probably the most important thing he taught me is that you cannot plan your life around the beam and physics conferences, but live it the way you want and fit the physics in.

A special thanks also goes to Kristian, the post-doc I worked with during most of my PhD. Kristian first guided me through the world of tracker DAQ and then initiated me in analysis work. He was always there to answer my questions, point out the flaws in my reasoning and helped me to grow as a physicist.

I also need to thank my two main colleagues, Kevin and Phil, a lot. Phil and I made a wonderful tracker commissioning team and then we both moved on to the W analysis and continued to collaborate as an awesome duo. Kevin helped me a lot with my analysis work, giving me hints on using root and being ready to jump in whenever extra manpower was needed to make a deadline.

I would like to thank all the other people in our group, for their help, their advice

and their nice company. I tried to find something more specific for most of them: I want to thank Christoph for the ice cream socials, Gerry for his eternal advice on systematic uncertainties, Roger and Matt to allow me to put my feet on their desk, Erik for his help in tracker, Josh and Si for electron-specific help, Matthew for taking over the high-rate set-up, Slawek for always having our back in tracker land, Valentina and Oliver for being nice friends and Markus for being so German, always pushing us harder to excel,...

Of course I also want to thank all the other colleagues I worked with, in the Green Barrack for the tracker and the Vector-Boson-Task-Force for the W analysis.

Next on the list are all my friends, in Geneva, Boston as well as in Belgium, for always being there, on the phone or in person, providing me with fun memories, and helping me to live the life I love so much :-)

Of course I also want to thank my family. Bedankt Mich en Roepa om me altijd te steunen om mijn dromen de wereld rond te volgen. Ook Barbara en Wouter waren altijd daar wanneer ik hen nodig had en ze zijn de beste zus en broer die ik me had kunnen wensen. Moeke, dank je om me op te hoogte te houden van de laatste Truiense weetjes met je wekelijkse brief . En natuurlijk wil ik ook Leena, Laure en Matteo bedanken omdat ze zo schattig zijn en het me altijd zielsgelukkig maakt om hen te zien :-)

Last of all I want to thank my wifey. Line, thanks for always being there, always cheering me up when I need it and marrying me the weekend before ICHEP :-) I love you.

Contents

Chapter 1

Introduction

The production of W bosons in a pp collider and their subsequent decay into electrons can be described in the Standard Model of particle physics. First an introduction to electroweak theory is given, and then the theory is applied to W production and decay in order to describe the theoretical foundations of the W$\rightarrow$eν cross section measurement.

1.1 Electroweak Theory

The origin of the weak theory started with the mystery of the electron energy spectrum in neutron β-decay, which indicated a three-body decay, but no third particle was observed. Wolfgang Pauli suggested a light neutral particle which would interact only through this reaction. In 1933 Enrico Fermi then described β-decay as a direct interaction of four fermions at one vertex[1]. In today's terminology, one of the down quarks from a neutron (udd) would change into an electron, anti-neutrino and an up-quark, so forming a proton (uud). Similarly muon decay could be described as a single point interaction for 4 particles.

The first revisions to Fermi's theory happened after Yang and Lee hypothesized parity did not have to be conserved to solve the so-called $\tau - \theta$ puzzle[2]. The τ and θ were two particles with the same mass and lifetime but different decay modes and the question was whether they actually were the same particle or two different particles.

Experiments by C.S. Wu and collaborators showed that parity indeed was violated in weak decays[3]. The surprising observation of parity violation led to the evolution of the weak theory in terms of a V-A (vector-axial) structure[4, 5] to accommodate parity violation. Further experiments confirmed the V-A theory of the weak interaction.

At the end of the 1950s this V-A theory was accepted as the basis of weak interactions. Its fundamental drawback was its divergent high-energy behavior, which prompted various ideas to cure this issue. Guided by the successful cancellation of similar divergences in quantum electrodynamics, attempts were made to construct a gauge theory of weak interactions.

By the mid-1960s all the pieces were in place for a successful construction of a theory of weak interactions by Sheldon Glashow, Abdus Salam and Steven Weinberg[6, 7, 8]. In their explanation of weak interactions using gauge theories, they unified the electromagnetic and weak interactions. The unification became the basis of today's Standard Model of particle physics. They suggested a unification under an SU(2) $\otimes$ U(1) gauge group. The corresponding gauge bosons are the three W bosons of weak isospin from SU(2) (W^+, W^0, and W^-), and the B^0 boson of weak hypercharge from U(1), all of which are massless. In the Standard Model, the physically observable W, Z and γ bosons, are produced by the spontaneous symmetry breaking of the electroweak symmetry from SU(2)$\otimes$ $U(1)_Y$ to only the $U(1)_{em}$ symmetry, caused by the Higgs mechanism[9, 10]. The spontaneous symmetry breaking causes the W^0 and B^0 bosons to mix together to form the Z boson and the photon as follows:

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B^0 \\ W^0 \end{pmatrix} \tag{1.1}$$

where θ_W is the weak mixing angle. The Higgs mechanism also gives mass to the W and Z bosons. $U(1)_Y$ is not the same as $U(1)_{em}$ since electric charge (Q) is a linear combination from hypercharge and weak isospin:$Q = Y/2 + I_3$, where Y is the the weak hypercharge and I_3 is the isospin. In 1971 't Hooft and Veltman provided the Glashow-Salam-Weinberg (GSW) electroweak model with a solid theoretical foundation by demonstrating it to be renormalizable, with finite cross section and specified

in terms of just one unknown parameter, $\sin^2\theta_W$[11]. The neutral Z partner serves to cancel the divergences in order to avoid unitarity violation.

The first experimental verification of electroweak unification theory came from the Gargamelle experiment, a huge bubble chamber at CERN (2 m diameter, 4.8 m long), which discovered weak neutral currents (compatible with Z exchange), as predicted by the theory. The first events were found in the very clean $\bar{\nu}_\mu + e \rightarrow \bar{\nu}_\mu + e$ channel where a single electron is projected at a small forward angle (about 2 degrees) to the beam at GeV energies. After 2 years of running, only 3 events were found but the background was expected to be much smaller (< 0.01 evts)[12]. The actual discovery happened in the hadronic channel, which is a much more challenging analysis because of the higher background. In the end inclusive events (102 ν, 63 $\bar{\nu}$) of the type ν + N$\rightarrow$ ν + hadrons with a large amount of hadronic energy gave large enough signal-to-background ratio to claim discovery[13].

The final proof for the GSW electroweak model came from the direct observation of W and Z events in the SPS. Physicists David Cline, Peter McIntyre and Carlo Rubbia suggested modifying CERN's biggest and newest accelerator at the time, the SPS, from a one-beam accelerator into a two-beam collider[14]. A beam of accelerated protons would collide with a beam of antiprotons, greatly increasing the available energy in comparison with a single beam colliding against a fixed target. The energy available in the collision is proportional to the energy of the beam in the case of colliding beams, while for a fixed target experiment, the collision energy scales as the square root of the beam energy. Simon Van der Meer at CERN had already invented a way of producing and storing dense beams of antiprotons[15].

The two ideas together would lead to CERN's biggest discoveries so far, done by the UA1 experiment. The W and Z were discovered in 1983, my year of birth. For selecting Ws, UA1 required an electromagnetic energy deposition of > 15 GeV, a correlated ionization track with a transverse momentum of > 7 GeV/c, missing transverse energy of > 14 GeV, and no hadronic energy deposition within 30° back-to-back in the plane transverse to the electron candidate. This selection is very similar to the one used by the Compact Muon Solenoid collaboration, as will be shown in

chapter 3. The UA1 selection yielded six events, of which one was identified as background because of large hadronic deposits[16]. A second search, using stronger electron selection but no missing energy cuts, yielded similar results (the 5 events + 1 extra in the endcap). In the next, higher intensity run, UA1 was able to measure the V-A structure of the interaction by looking at the angular dependence of the W production. In that running period, W$\rightarrow \mu\nu$ evidence was found and the first Zs decaying to muons and electrons were detected. The selection of Z $\rightarrow e^+e^-$ was much easier than the W selection. Events were selected asking for two electromagnetic high-energy clusters with isolated tracks, track-cluster spatial matching, and very little hadronic energy deposited ($<$ 1 GeV) directly behind the cluster. Using these cuts, four Z$\rightarrow e^+e^-$ events were selected with no visible experimental background. UA1 published these events as the discovery of the Z[17].

Over the past two decades the W and Z properties have been measured very precisely at LEP (for Z and W) as well as the Tevatron (for W). The LHC experiments will extend this knowledge with precision measurements at higher center-of-mass energies. This thesis will focus on the measurement of the W$\rightarrow e\nu$ cross section in pp collisions at $\sqrt{s}$=7 TeV.

1.2 W Production at pp Colliders

In proton-proton collisions the dominant mechanism for vector boson production is the quark anti-quark annihilation:

$$\mathrm{u} + \bar{\mathrm{d}} \rightarrow W^+, \bar{\mathrm{u}} + \mathrm{d} \rightarrow W^-, \tag{1.2}$$

with u and d the up and down quarks. The W production can also happen with a strange quark instead of a down quark, because the actual weak eigenstate d′ associated with the weak interaction is a combination of the down and the strange quark. The mixing can be described by the Cabibbo mixing angle η_C (13.04°)[18, 19].

Most of the proton's momentum is carried by the two u and one d valence quarks,

such that reactions with the sea antiquarks results in a predominance of W^+ production relative to W^-. The sea-quark/sea-quark interactions will not result in an asymmetry. This charge asymmetry will be measured in this thesis.

The W production cross section is theoretically well-known. Using θ_C for the Cabibbo angle, G_F for the Fermi constant and the quark distribution functions we can write the differential cross section with respect to rapidity as:

$$\frac{d\sigma}{dy}(pp \to W + X) = K\frac{2\pi G_F}{3\sqrt{2}} x_a x_b \{cos^2\theta_C[u(x_a)d(x_b) + d(x_a)u(x_b)] + sin^2\theta_C[u(x_a)s(x_b) + s(x_a)u(x_b)]\}, \quad (1.3)$$

with $x_{a,b} = \frac{M_W}{\sqrt{s}} e^{\pm y}$ the parton momentum function, $u(x)$ and $d(x)$ the quark distribution functions and K a scale factor for next-to-leading-order corrections. These next-to-leading (NLO) order corrections are due to gluon vertex correction, quark-gluon interactions and real gluon emissions (fig. 1-1). For higher-order corrections the formula becomes more complicated.

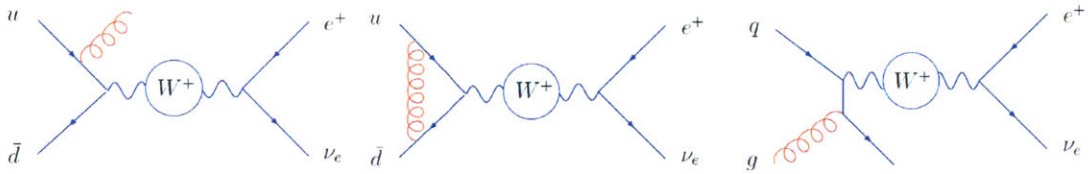

Figure 1-1: NLO corrections to the W production cross section: gluon radiation, gluon vertex correction and quark-gluon scattering[20].

W production has been calculated to NNLO in QCD and NLO in EWK (fig. 1-2). The results for the cross section versus rapidity for higher perturbative orders is shown is fig. 1-3, where the difference between W+/W- in both shape and amplitude is apparent, and the importance of higher order corrections evident. Such higher-order corrections and the tools used to calculate them will be further investigated in chapter 5, where they are used to determine the precision with which we know the fraction of the cross section which produces observable final states in CMS.

The interaction probabilities are affected by the momentum distributions of the

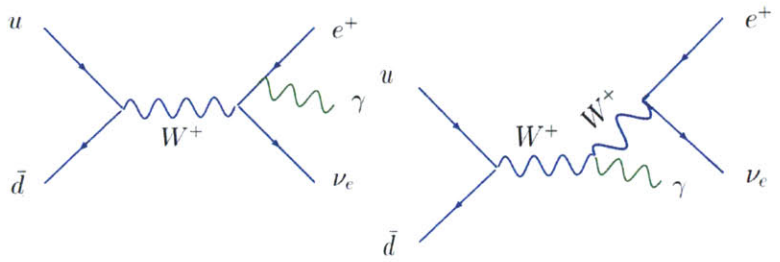

Figure 1-2: O(α) EWK corrections to the W production cross section[20].

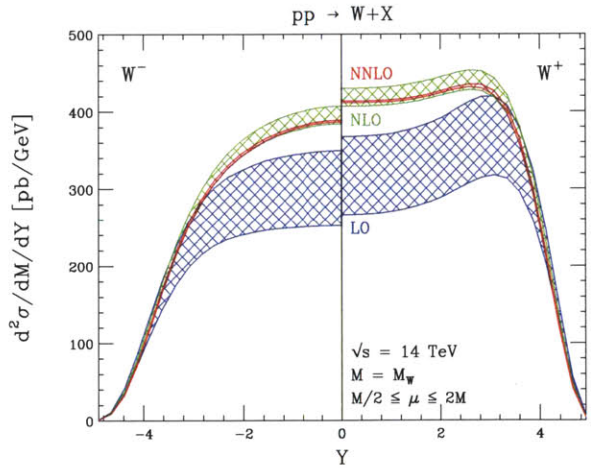

Figure 1-3: W^- and W^- cross section vs. rapidity at NNLO.Each distribution is symmetric in Y, only half the rapidity range is shown in each case. The bands indicate the common variation of the renormalization and factorization scales in the range $M_W/2 < \mu < 2M_W$. [21]

incoming partons (quarks and gluons) within the hadron. These are described by Parton Distribution Functions (PDFs), analytical fits obtained using a combination of theory and experimental data which describe the momentum distributions of partons inside a hadron. PDFs are of vital importance in calculating cross sections for such processes, and they give the largest uncertainty on the theoretical cross section prediction. Different sets of PDFs from different collaborations (MSTW[22], CTEQ[23], NNPDF[24],...) are available. All these sets use slightly different techniques for the combination and use different measurements to get their final density function. Only MSWT2008[22] has a NNLO version, this will be used in our NNLO prediction of the cross section. A mixture of different PDF sets and their error matrices will be used to quantify the PDF uncertainties. This is done so that the uncertainties within one method and the differences between the methods used by the different groups are both taken into account.

1.3 W decay

W bosons are unstable particles, and they decay to a pair of leptons or a pair of quarks. Electron and muons can be measured with much better precision than quarks, which fragment into jets of particles. Therefore, the lepton-decay mode provides a much cleaner signal. The focus of this thesis will be on the W$\rightarrow e\nu$ decay. The partial width of the decay is:

$$\Gamma(W \rightarrow e\bar{\nu}_e) = \frac{1}{48\pi} g^2 M_W = \frac{G_F}{\sqrt{2}} \frac{M_W^3}{6\pi} = 0.225\,\text{GeV}. \tag{1.4}$$

The W boson only couples to left-handed fermions (parity violation) and the coupling is always the same (universality) for every fermion. The universality holds when the fermion masses are much smaller than the W boson mass and leads to the fact that the partial widths are independent of the final states. This can be tested[25] and the muon and electron W decay modes agree within uncertainties. The only correction to this is the Cabibbo-rotation. In the following we will be using

the eigenstates of the weak force d′ and s′, which are related to the eigenstates of the weak force (down and strange quark) by a simple Cabibbo rotation[18, 19]. The branching ratios can be obtained simply by counting up the possible final states of the W decay. The quark-antiquark decays have a factor three because of the color charges which give rise to three different possible final states for every type of quark-antiquark pair. Because of the large mass of the top quark ($m_{top} > m_W$), there will not be a bottom-top quark pair. So in the end, the expected decay ratio for the W decay for $e\nu$, $\mu\nu$, $\tau\nu$, ud′, cs′ is 1:1:1:3:3, so to lowest order the W$\rightarrow e\nu$ decay is expected to have a branching ratio of 1/9. The experimental value is: 10.75$\pm$ 0.13 %[26] which is very close to this value. Higher-order corrections bring the theory estimate into full agreement with the experimental values.

1.4 Why do we measure the W Cross Section?

1.4.1 Precision test of NNLO Calculations

The W production cross section is very well predicted by theory. The available calculations include perturbative corrections up to NNLO in QCD[27, 28] and NLO in EWK[29]. The main residual uncertainties come from the parton distribution functions. A precise measurement of the cross section at 7 TeV provides a benchmark for the current state-of-the-art theoretical models and simulation tools to see whether they can accurately extrapolate to the new energy of the LHC. If not, the disagreement between theory and experiment could be hinting at new physics. At the Tevatron, which provided $p\bar{p}$ collisions at 1.96 TeV, CDF[30] and D0[31] measured the W$\rightarrow e\nu$ cross section with 2.3% and 2.6% experimental uncertainty relatively, with an extra 6% uncertainty on the luminosity. These results were consistent with NNLO predictions. We will try to reach a similar accuracy in our model to be able to probe the theory at 7 TeV.

1.4.2 Commissioning of Electrons and Missing Transverse Energy

W bosons, together with their partner the Z, are very important for calibrating the detector and commissioning leptons and missing transverse energy. The Ws provide the majority of prompt electrons that are produced in proton-proton collisions and this signal has real missing transverse energy (MET) because of the neutrino. The missing transverse energy represents the energy carried away by unmeasured particles, such as neutrinos, and given overall energy conservation is opposite the direction of the vector sum of all visible energy depositions. Because the z component of the center of mass frame is unknown in hadron-hadron collisions, the energy balance is only useful in the transverse plane. In chapter 7 we will show how we can exploit the Z signal to improve our modeling of lepton and MET scale and resolution. The precision measurement presented in this thesis is already quite sensitive to multiple interactions and mismodeling of the underlying event. For this measurement, MET and lepton energy correction techniques had to be developed[32, 25]. These techniques are already adopted by other analyses to improve their understanding of missing transverse energy and leptons. A good understanding of the signatures leptons and Ws leave in the detector plays a crucial role in new physics searches at the LHC, as discussed below.

1.4.3 Further W Studies

The W cross section measurement is the first step towards a variety of W analyses studying the Standard Model at a new, higher energy. This analysis sets the baseline to understand the W and new analyses can build on this. For example, the charge asymmetry analysis in CMS adopted some of the techniques used here[33]. The results of that analysis agree with the results presented in this thesis, but it is a differential analysis in rapidity and therefore it is much more sensitive to constrain the parton distribution functions. New PDFs based on the differential charge asymmetry analysis have already been presented[34]. Other examples are the W mass measure-

ment, which can put tighter constraints on the Higgs mass, and the W polarization measurement[35].

1.4.4 New Physics Searches

New physics searches benefit from the W cross section analysis in two ways. Some channels, like the important H→WW decay mode[36], will create real Ws and an improved understanding of the W will help them directly. It is also crucial to new physics searches because the most important backgrounds for these analyses are also composed of real Ws (e.g. diboson production). Without a proper understanding of the Standard Model backgrounds and the signature of a W in our detector, a discovery cannot be claimed. A lot of other searches, in SUSY or W′[37] do not directly look for a W, but have very similar signatures: an isolated lepton and considerable missing transverse energy. Our improved understanding of leptons and MET through analyzing the W can help these studies.

1.5 Cross Section Measurement

Cross-sections are calculated using the following formula :

$$\sigma \times BR = \frac{N_{signal}}{A\epsilon \int \mathcal{L} dt}, \tag{1.5}$$

where

- BR is the branching ratio, in our case the electron + neutrino decay mode of the W;
- N_{signal}=N_{total}-$N_{background}$, the number of signal events N_{signal} is the total number of selected events (N_{total}) minus the estimated number of background events ($N_{background}$);
- A is the kinematic acceptance, this defines how many Ws are expected to leave their decay products inside the instrumented region of the detector;

- ϵ is the efficiency, how often a real electron is reconstructed and passes the online and offline selection cuts;
- $\int \mathcal{L}dt$ is the luminosity, the amount of data accumulated, 35.9 pb^{-1} for this analysis.

The next chapters will discuss the determination of the different elements of this equation and the corresponding uncertainties for the W$\rightarrow$eν cross section. We will start with a description of the experimental set-up in chapter 2: the accelerator and the CMS detector. The luminosity determination, which is an external measurement to this thesis and provides the overall normalization, will be discussed in section 2.2.7. Then we will describe in chapter 3 the building blocks of the measurement: how do we reconstruct electrons and missing transverse energy, how do we select the events and how do we reduce our background to prepare for the final signal extraction. Chapter 5 will describe the acceptance measurement and its theoretical uncertainties, utilizing the theoretical tools described in chapter 4. Chapter 6 will discuss the efficiency measurement. The W signal extraction, based on the missing transverse energy signature will be covered in chapter 7, including the uncertainties due to energy scale and resolution. Finally, all the elements will be combined to result in a final W cross section measurement in chapter 8, which is compared to the state-of-the-art theoretical predictions.

Chapter 2

The Compact Muon Solenoid at the Large Hadron Collider

2.1 The Large Hadron Collider

The Large Hadron Collider[38] (LHC) is the world's largest and highest energy particle accelerator. The LHC is housed in the former 26.7 km long tunnel of the Large Electron Positron Collider (LEP[39]) at the border between Switzerland and France, just outside Geneva. The LHC is built at a mean depth of 100 m, with a gradient of 1.4%. The actual depth varies from 175 m under the Jura to 50 m nearest to Lake Geneva. The depth and gradient were determined due to geological constraints, trying to avoid too hard and too soft layers of stone. The tunnel, which crosses the border 4 times, consists of 8 straight sections and 8 arcs where the particles are bent.

The bending radius of the arcs in the tunnel is 2803.95 m, slightly less than expected from a circle with 26.7 km circumference, because of the straight interaction sections. Superconducting NbTi dipole magnets, cooled down to 1.9K, generate a nominal field of 8.33T used to bend the beams. The two beams with the same charge in the LHC cannot travel in opposite directions in the same beampipe. Therefore the LHC design (fig. 2-1) employs twin-bore dipole magnets in which each proton beam has its own beampipe, but share a common cryostat to save space and money. Beam focusing is done using quadrupole magnets. In the LHC ring, a total of 1232

dipole and 398 quadrupole magnets are installed. The maximum dipole magnetic field together with the bending radius of the accelerator determine the maximum energy to which a particle can be accelerated. Given the magnetic fields and the geometry of the LHC, the maximum proton beam energy is 7 TeV, corresponding to a center-of-mass pp collision energy of 14 TeV. At this moment, the LHC is not running on design energy yet, but collides protons at half this energy (7 TeV collision energy). The LHC can also collide lead ions.

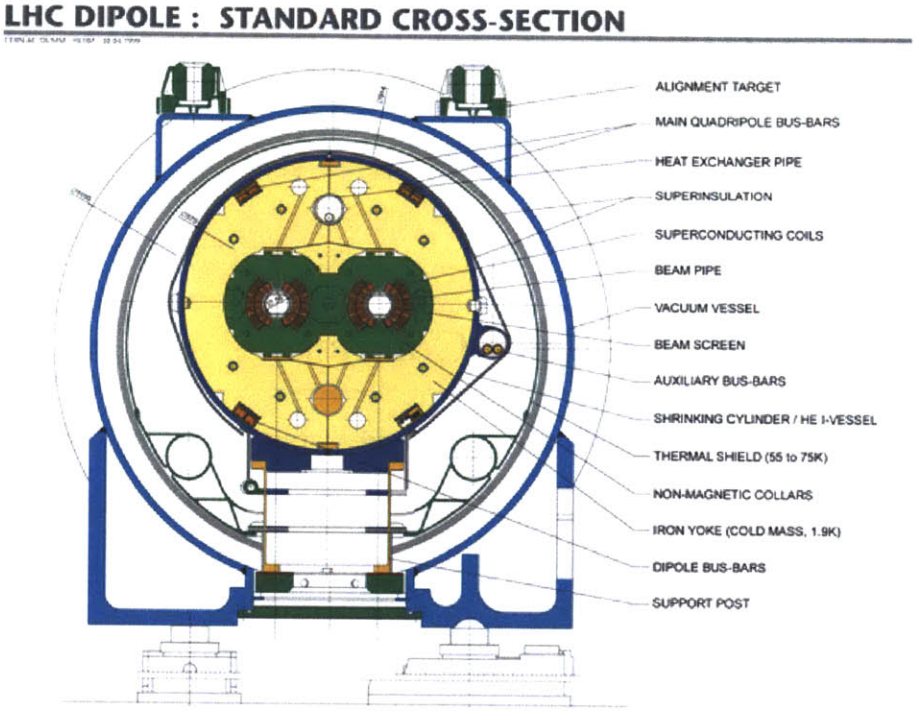

Figure 2-1: Cross-section of a LHC dipole[38].

The LHC is the last and the largest of a series of accelerators (fig. 2-2), each of which increases the proton energy incrementally. The first step is the creation of the proton beam. Electrons are stripped off hydrogen atoms provided by a hydrogen bottle. The protons are then accelerated by the linear accelerator Linac-2 to 50 MeV kinetic energy and injected into the Proton Synchroton Booster. After accelerating them to 1.4 GeV kinetic energy, the Booster feeds the protons to the Proton Synchroton(PS). The PS can take up to seven bunches of protons but only takes six to allow for an abort gap. At design luminosity the PS will split the bunches into three after

injection, and each of these is in turn split in four bunches after acceleration, allowing for 6 x 3 x 4 = 72 bunches before extraction. After the acceleration of the protons to 25 GeV, kicker magnets nudge the beam out of the PS orbit for injection in the Super Proton Synchroton (SPS) for the next stage of acceleration. An abort gap is in place to make sure none of the proton bunches reach the kicker magnets while these magnets are ramping up, which would result in a random kick off the intended orbit between the ring and the transfer line, possibly damaging the infrastructure. Up to four batches of seventy-two bunches from the PS can be injected into the SPS, which will accelerate them further up to 450 GeV per proton. Then a similar injection happens into the LHC ring. The injection can be maximally repeated 24 times to give a minimum spacing of 25 ns in each of the two beams in the LHC. Taking into account abort gaps in the different injection and extractions, the maximal allowed number of bunches is 2808 per beam in the LHC. The LHC accelerates the particles to their final energy, from 450 GeV to 3.5 GeV per beam[38].

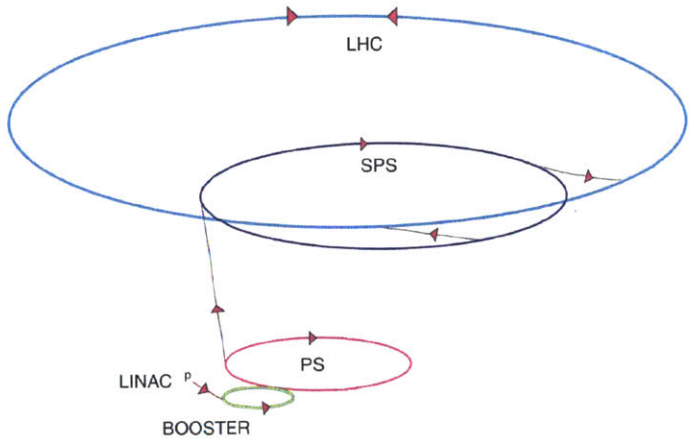

Figure 2-2: The accelerator complex feeding the Large Hadron Collider

To maximize their chance of discovering new phenomena, the experiments also need the highest number of possible interactions. The number of signal events produced can be calculated as follows: $N_{sig} = \int \mathcal{L}\sigma dt$, where σ is the cross-section of the signal process, in this case the production of a W boson decaying into an electron-neutrino pair. $\mathcal{L}$ is the instantaneous luminosity, which is a measure for the intensity

of beam and the number of collisions happening. So for rare processes (small cross-section), a high luminosity or a long period of data-taking is needed.

The luminosity is fully determined by beam parameters and can be expressed as follows:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F, \tag{2.1}$$

where

- N_b is the number of protons per bunch;
- n_b is the number of colliding bunches in the beams;
- f_{rev} is the frequency of beam circulation;
- γ_r is the relativistic γ factor for protons;
- F is a geometric luminosity reduction factor due to beam-crossing angle;
- ϵ_n is the normalized transverse beam emittance;
- β^* is the value of the betatron function at the collision point.

Together the β^* and ϵ_n parameters determine the width of the beam[38]. The LHC design luminosity is $10^{34}\,\mathrm{cm}^2/s$, but the maximum reached during 2010 was $2 \times 10^{32}\,\mathrm{cm}^2/s$. The luminosity was increased several times in different steps as can be seen in section 2.2.7 which shows the the integrated luminosity delivered by the accelerator and the one recorded by CMS. The steps up in luminosity were obtained by increasing the number of particles per bunch, the number of bunches per beam, and decreasing the transverse width of the beam at the interaction points. The other factors in the luminosity formula cannot be easily changed. Datasets are quite often referred to as "integrated luminosity" ($\int \mathcal{L} dt$), with units of inverse cross-section, since this facilitates a quick calculation of the expected number of signal events. More details about the luminosity in CMS will be given in section 2.2.7.

In the 8 straight sections, experiments and beam services can be installed. The layout of the LHC is shown in figure 2-3. Point 1 and Point 5 contain A Toriodal Large

ApparatuS (ATLAS[40]) and the Compact Muon Solenoid Experiment (CMS[41]). These are the 2 general-purpose detectors at the LHC. They are built to search for a variety of new physics, like Higgs and SUSY searches, and new phenomena in heavy ion-physics. A Large Ion Collider Experiment (ALICE[42]), located at point 2, is specifically designed for heavy ion collisions, whereas the Large Hadron Collider bottom experiment (LHCb[43], point 8) focuses on the properties of bottom quarks. At point 6 the beam can be dumped into a graphite core by pushing them off-orbit with special kicker magnets. Point 4 houses the radio-frequency cavities for the acceleration of the beam from 450 GeV to 3.5 TeV. The last 2 points, point 3 and 7, contain collimators to clean the beam.

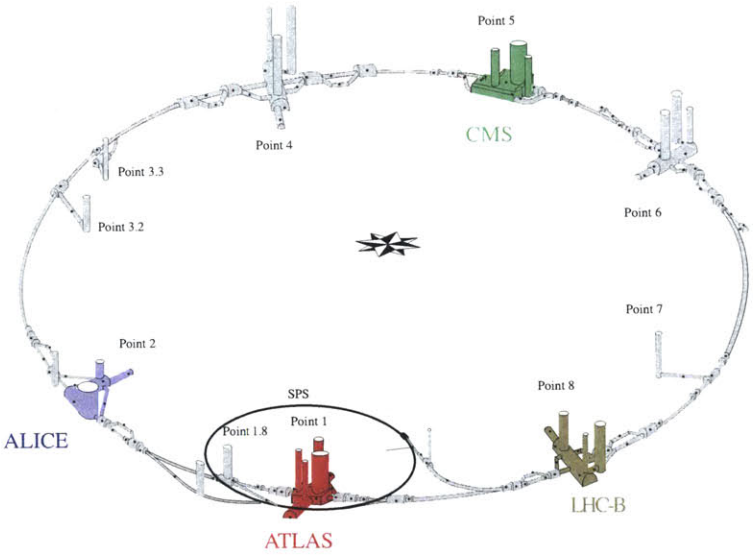

Figure 2-3: The Large Hadron Collider and its 8 interaction points

Operation of the Large Hadron Collider started the first time on 10 September 2008, but after a week of commissioning a fault in the electrical connections between two magnets caused an electrical short in the tunnel, damaging more than 50 magnets. A year later, on 23 November 2009, the first collisions were recorded, at injection energy (450 GeV per beam). The LHC was further commissioned to reach record-breaking 2.36 TeV collisions by the end of 2009. In 2010, the beam energy was increased up to 3.5 TeV, and from April until November around 47 pb^{-1} of 7 TeV

pp collisions were delivered by the accelerator to the CMS experiment. In November and December 2010 the LHC also collided lead ions with an energy of 287 TeV per beam (1.38 GeV per nucleon).

2.2 Compact Muon Solenoid

At interaction point 5 of the LHC, near the small hamlet of Cessy, France, the Compact Muon Solenoid[44] is installed in a cavern roughly 100 m underground. CMS is 21 m long and has a 15 m diameter and with these dimensions it is half the size of its chief competitor ATLAS, which justifies the adjective 'Compact'. CMS is structured like an onion with several layers of detector components. Each layer measures the track or the energy of another kind of particle coming from the collisions, as is shown in fig. 2-4. Measuring several properties of the particles, like the momentum, the energy and the charge will give us information about the particular process occurring in the event. A small introduction of the different parts provides a general view of the detector before we go into more detail about every sub-detector separately.

Particles which emerge from the collision will first traverse the silicon-based tracker[45, 46]. Here the path of the charged particles away from the interaction is mapped and some of the neutral particles, like photons, can interact with the material and then be tracked after they convert into other particles (e.g. $\gamma \rightarrow e^+e^-$). By measuring the curvature of a particle track in the magnetic field the momentum of this particle can be calculated.

The energies of the particles are measured in the calorimeters, which surround the tracker. The energy of photons and electrons are measured by the electromagnetic calorimeter (ECAL[47]), which is made up of lead-tungstate, a very dense scintillating material. The energy of hadronic jets (fragmentation of the produced quarks) is mainly measured in the hadronic calorimeter (HCAL[48]), although they will already deposit some energy in the ECAL. The HCAL is made out of brass and scintillator.

The huge superconducting 4 T magnet[49] coil lies just outside the calorimeter. The giant coil inside the cryostat, which creates the actual magnetic field, is the

solenoid which gives the experiment its last name.

The only particles passing further in the detector are the muons, the last part of the experiment's name. Muons will be measured by the muon chambers[50]. The 3 different types of muon chambers with all their own characteristics will be further described in section 2.2.5. The muon chambers are interleaved with the iron return yoke for the magnet which confines the magnetic field.

There is also dedicated hardware needed for triggering, which is used to select interesting events in real time for further study, required due to the limited bandwidth which makes it impossible to record every event to disk. The choice of which events must be kept must be done in real-time.

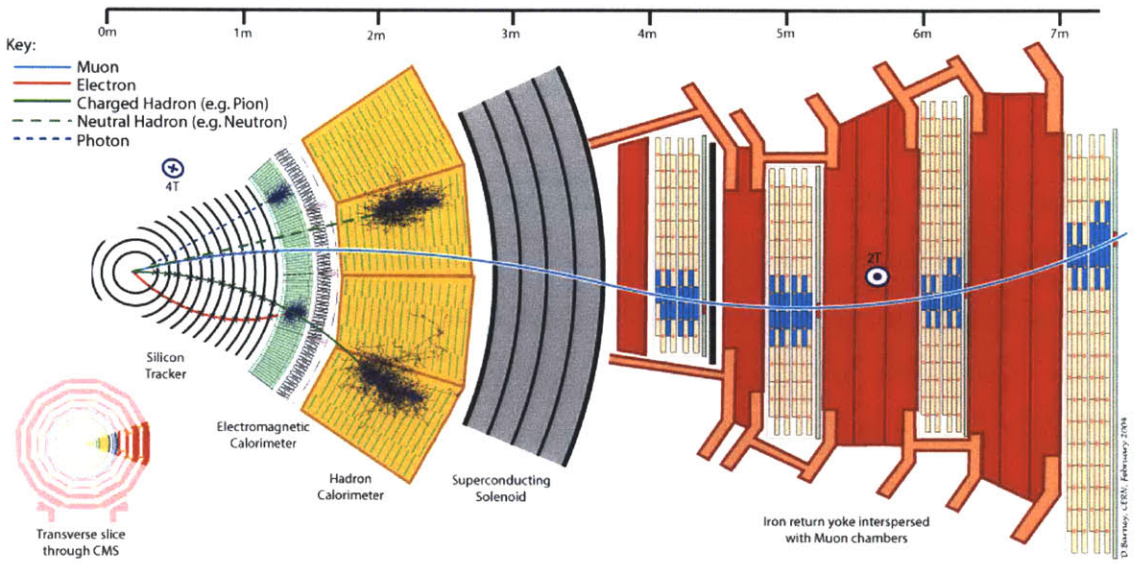

Figure 2-4: A schematic representation of how different particles interact and are detected by the various sub-detectors of CMS.

The coordinate system adopted by CMS has the origin centered at the nominal collision point inside the experiment, the y-axis pointing vertically upward, and the x-axis pointing radially inward toward the center of the LHC. Thus, the z-axis points along the beam direction toward the Jura mountains from LHC Point 5. The azimuthal angle ϕ is measured from the x-axis in the x-y plane and the radial coordinate in this plane is denoted by r. The polar angle θ is measured from the z-axis[44].

For the dimensions of the different sub-detectors, the spatial coordinate pseudorapidity, η, will also be used. It it another way of describing the angle of a particle relative to the beam axis and can be rewritten as a function of the momentum vector:

$$\eta = -ln[tan(\frac{\theta}{2})] = \frac{1}{2}ln\left(\frac{|\vec{p}| + p_L}{|\vec{p}| - p_L}\right), \tag{2.2}$$

where θ is the angle between the particle momentum $\vec{p}$ and the beam axis. p_L is the projection of the momentum along the beam axis.

2.2.1 The CMS Magnet

As already noted, CMS has a large solenoid magnet, surrounding the tracking detectors and the calorimeters. A magnetic field is needed to bend the particles so that their momentum can be measured. Particles with less momentum, bend more than particles with more momentum following the formula: $R = \frac{p_T}{0.3eB}$ with R the radius of curvature in m, B the magnetic field in T and p_T the particle momentum in GeV/c. The resolution of the transverse momentum is given approximately by the following formula : $\sigma_p = \sqrt{\frac{720}{N+4}}\frac{\sigma_y p^2}{0.3BL^2}$[26], with N the number of hits, B the magnetic field, σ_y the spatial measurement uncertainty and L the projected track length. From this formula, it is obvious that a stronger magnetic field will allow for more accurate measurements of high energy particles; the CMS design allows for measuring muon momenta up to 1 TeV with a precision better than 10%[44].

The CMS solenoid is a gigantic object in many dimensions. It is is 13 m long, with a 6 m inner diameter. The solenoid creates a 3.8 T magnetic field by running 18160 A through the superconducting niobium-titanium coils, resulting in a stored energy of 2.3 GJ. The magnet is encased in 3 iron return yokes, going out to 14 m diameter, to contain and guide the magnetic field[49]. All together, the magnet system weighs an impressive 12000 tons, making CMS 80% heavier than ATLAS.

2.2.2 The CMS Tracker

The CMS silicon tracker precisely measures the trajectories of the charged particles coming from the interaction or from secondary vertices close to the interaction. The tracker does so by measuring the position where a particle passed through the different layers. In addition, the tracker is able to measure tracks with sufficient accuracy to separate primary and secondary vertices from long-lived particles like c- and b-hadrons with high efficiency. The tracker is crucial to the W$\rightarrow$eν cross-section measurement for identifying electrons, measuring the electron tracks, their momentum, and comparing the position of the track at the calorimeter surface with the location of the deposits in the calorimeter.

The precise measurement of a charged particle trajectory involves an optimization of the trade-off between the number of measurements of the track ('hits') and the accuracy of each single measurement, and the amount of material in the tracking volume, which causes deviations from the theoretical helix due to multiple scattering and excessive energy loss. More hits will require extra layers and consequently more material. The active volume of the CMS tracker is built out of silicon, which is denser than typical gas detector technologies but provides a better single hit resolution. Also, silicon detectors are better suited to survive longer in the high-radiation environment of the LHC. Even being made out of the silicon, the tracker layer closest to the beamline only has a lifetime of around 2 years at design luminosity. The subsequent tracker layers lie further away from the interaction point and their lifetime increases to 10 years or more. Low detector occupancy is also desired to avoid degradation in position resolution and ambiguities in the track reconstructions due to multiple hits in the same sensitive volume. The occupancy increases from 10^{-4} in the innermost layers to a few percent in the outer layers, because of the different technologies used depending on the distance from the collision point.

The CMS tracker is made out of 200 m^2 of silicon and is 5.8 m long and the outermost silicon layer has a diameter of 2.4 m, making it the largest silicon detector ever built. The barrel-endcap transition of the tracker happens at $|\eta|$ = 0.9-1.5,

here the largest amount of material (> 1 radiation length) in tracker can be found because of the cabling and the support structures (fig. 2-5). The entire tracker gives coverage up to $|\eta| = 2.5$. The tracker is built up out of 4 main parts: the pixel detector (PIXEL), the tracker inner barrel and inner disks (TIB/TID), the tracker outer barrel (TOB) and the tracker endcaps (TEC) as shown in figure 2-6[46].

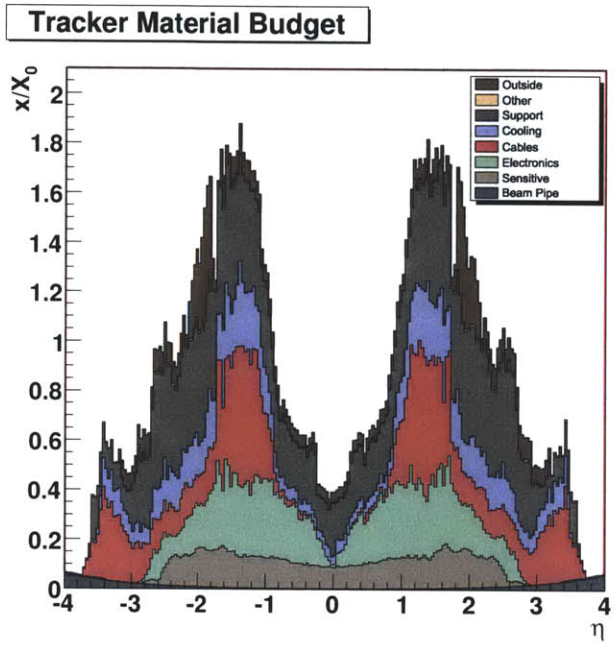

Figure 2-5: Amount of material in the tracker in units of radiation length as a function of pseudorapidity η broken down into the functional contributions[44].

The pixel detector consists of 66 million pixels, making up 1440 pixel detector modules. The pixel detector has three barrel layers at 4.4, 7.3 and 10.2 cm radius from the center of the detector and at each side there are two forward pixel layers radially extending from 6 to 15 cm diameter, positioned at z = ±34.5 cm and z = ±46.5 cm. The pixel cells are 100 μm in r-ϕ and 150 μm in z. This results in a position resolution of 10 μm in r-ϕ and 15-20 μm in z[45].

At radii greater than 20 cm, the charged particle flux is reduced sufficiently to allow the use of silicon strip detectors while maintaining low occupancy. The tracker inner barrel and disks (TIB/D) extends from 20-55 cm radially and is composed of four barrel layers, supplemented by three disks at each end. The TIB/D is surrounded

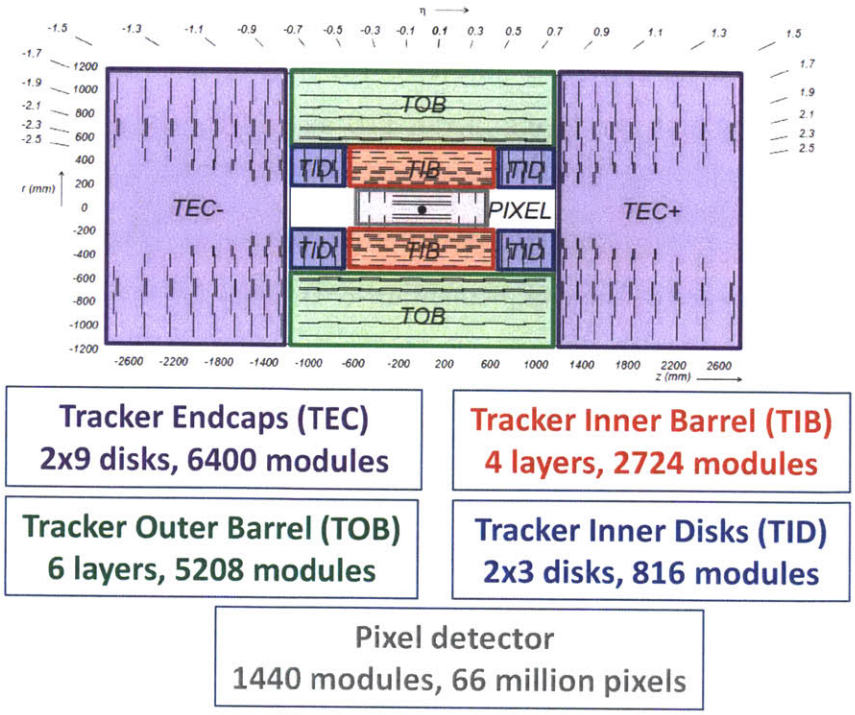

Figure 2-6: Overview of the CMS tracker.

by the tracker outer barrel which has an outer radius of 116 cm and consists of six barrel layers. The TOB extends in z up to ± 118 cm. Beyond this z range the Tracker EndCaps (TEC) cover the region 124 cm $< |z| <$ 282 cm and 22.5 cm $< r <$ 113.5 cm. Each TEC is composed of nine disks, carrying up to seven rings of silicon micro-strip detectors. The achieved single-point resolution with the silicon strip modules is 23-53 μm and a signal-to-noise ratio of 18-27, depending on the exact geometry of the module. The different layers of the detector have modules with different strip pitches, strip lengths and module thicknesses, as shown in table 2.1[46, 44].

Sub-detector	Module thickness	Strip pitch	Strip length
TIB/D	320 μm	80-120 μm	116.9 mm
TOB	500 μm	122-183 μm	183.2 mm
TEC 1-4	320 μm	81-139 μm	85.3-115.2 mm
TEC 5-7	500 μm	126-172 μm	144.4-201.8 mm

Table 2.1: Module thickness, strip pitch, and strip length for different module types.

In both TIB and TOB the strips are oriented along the beam axis, providing r-ϕ position measurements. In TID and TEC the strips point radially, providing z-ϕ measurements. The first two layers in the TIB and TOB, along with the inner two disks of the TID and rings 1, 2 and 5 of TEC have two detector modules mounted back-to-back, with a 100 mrad stereo angle between them. These stereo detectors provide a measurement of the complementary coordinate for their detector (z in the barrel and r in the endcaps). This tracker layout tries to provide at least nine hits in the silicon strip tracker in the full tracker acceptance with at least four of them being two-dimensional measurements[46].

For single muons of transverse momenta of 1, 10 and 100 GeV figure 2-7 shows the expected resolution of transverse momentum and the track efficiency as a function of pseudorapidity[41]. For high momentum tracks (100GeV) the transverse momentum resolution is around 1-2% up to $|\eta| \approx 1.6$, beyond which it degrades due to the reduced lever arm and multiple scattering in the tracker material starts playing an important role, accounting for 20-30% of the observed resolution. The track efficiency for muons is about 99% over most of the acceptance. For $|\eta| \approx 0$ the efficiency decreases slightly due to gaps between the ladders of the pixel detector at z $\approx$ 0. At high $|\eta|$ the efficiency drop is mainly due to the reduced coverage by the pixel forward disks. For pions and hadrons in general the efficiency is lower ($\approx$ 90%) because of interactions with the material in the tracker.

Figure 2-8 depicts the control flow to the front end ASICs, and the data flow back back into the electronics for processing and storage. The Frond End Controller (FEC) VME board sends trigger and clock information to the Digital OptoHybrid Module (DOHM), which performs the optical to electrical conversion and forwards the electrical signals to the Control and Communication Unit (CCU), which distributes them to all front-end ASICs. The front-end ASIC (implemented as the APV chip) samples the strip charge, does the analogue pulse shaping, stores the data locally, and upon request transfers them optically to the Front-End-Drivers (FED) VME board. There the data is digitized and under default operation clusters are formed and zerosuppression applied, dropping clusters below a preset threshold. This threshold

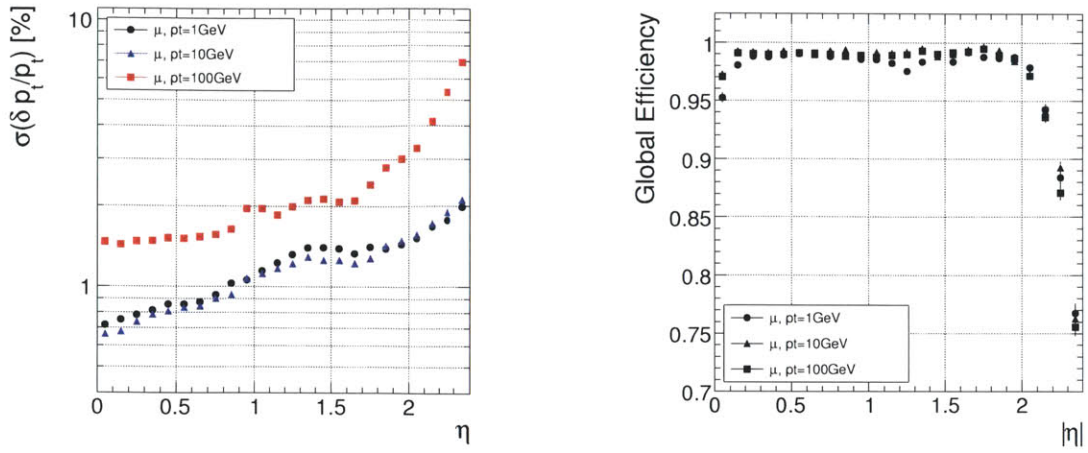

Figure 2-7: Transverse momentum resolution (left) in the tracker and track efficiency (right) for muons with transverse momenta of 1,10 and 100 GeV[44].

is determined by the amount of noise of the detector, which is measured in specific commissioning runs. Finally the FED pushes the data out to the Central CMS DAQ.

2.2.3 The Electromagnetic Calorimeter

The energy of final state particles leaving the tracker volume is measured by calorimeters. The Electromagnetic calorimeter (ECAL) is just outside the tracker volume and it measures the energy of electrons and photons. It is a homogeneous calorimeter comprised of 75848 lead tungstate ($PbWO_4$) crystals equipped with fast, radiation hard readout electronics[47]. The ECAL is divided into a central barrel (EB) and two endcap (EE) sub-detectors (Figure 2-9). A preshower detector is located in front of each endcap. Lead tungstate crystal is made primarily of metal and is heavier than stainless steel ($\rho = 8.28\, g/cm^3$), but with a touch of oxygen in this crystalline form it is highly transparent and scintillates as electrons and photons pass through it. The crystals emit blue-green scintillation light with a broad maximum at 420-430 nm. These high-density crystals have a very short radiation length (0.89 cm) and a small Moliere radius (2.10 cm) and thus allow for a compact detector with fine granularity and good shower containment for electrons and photons in the 5-500 GeV range.

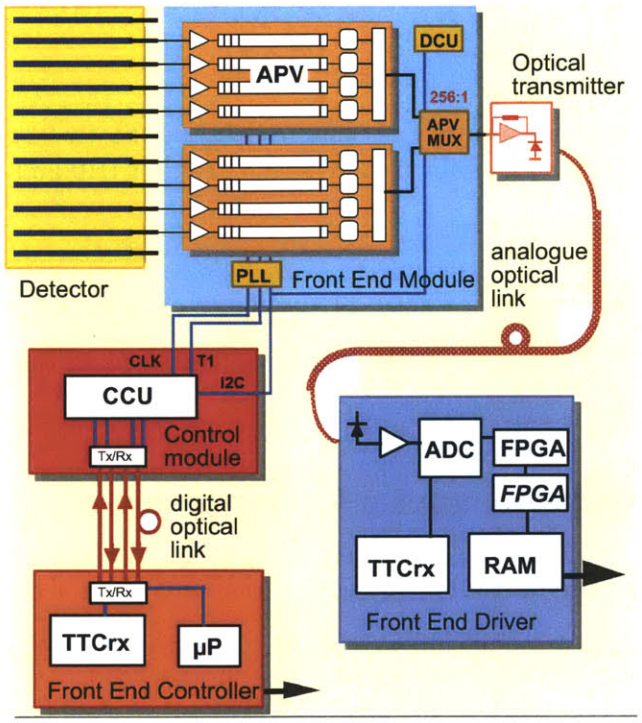

Figure 2-8: Schematic view of the CMS tracker read-out electronics[44].

The ECAL is used to measure the electron energy very precisely and to identify the electron by looking at the pattern and location of the ECAL deposits.

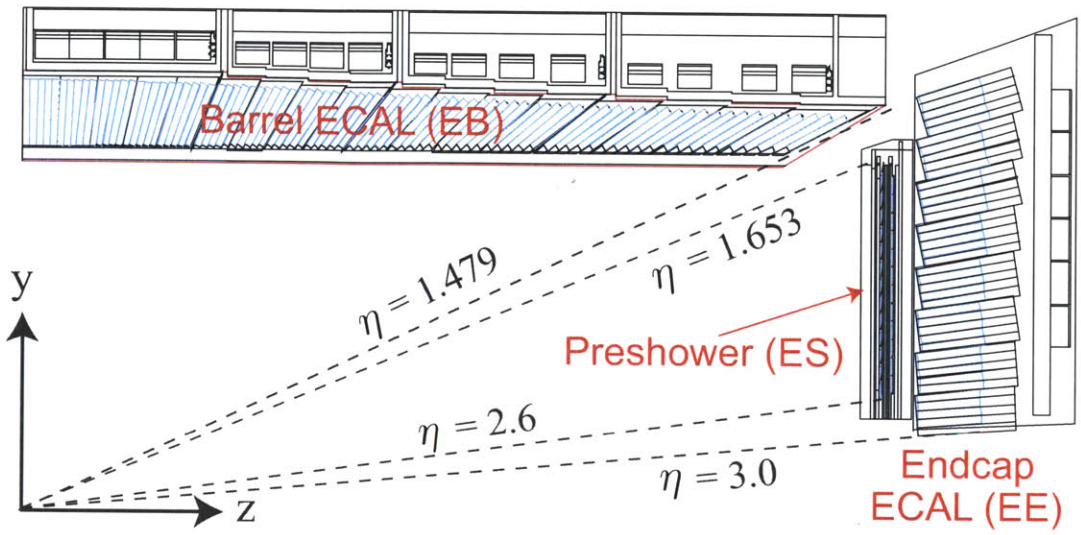

Figure 2-9: View of one quarter (r,z) of the CMS Electromagnetic Calorimeter.

In the ECAL barrel, there are 61200 crystals going up to pseudorapidity 1.479. A 3% displacement in both θ and ϕ minimizes the number of particle trajectories aligned with the gaps between crystals, reducing the longitudinal energy loss in the cracks. The crystal cross-section corresponds to approximately 0.0174x0.0174 in η-ϕ or 22x22 mm^2 at the front face of crystal, and 26x26 mm^2 at the rear face. The crystal length is 230 mm corresponding to 25.8 radiation lengths[47, 44]. The endcaps (EE) cover the rapidity range $1.479 < |\eta| < 3.0$. The longitudinal distance between the interaction point and the endcap envelope is 315.4 cm. The endcap consists of identically shaped crystals grouped in mechanical units of 5x5 crystals. The two endcaps are each made of two D-shapes, which hold 3662 crystals arranged in an x-y geometry. In the endcap the crystals are slightly larger and shorter. They have a front face cross section 28.62x28.62 mm^2 and a rear face cross-section of 30x30 mm^2, and a length of 220 mm (24.7 X_0). The crystals point at a focus 130 cm beyond the interaction point, creating angles ranging from two to eight degrees. This is done again to minimize the effects of the gaps between the crystals for particles coming

from the interaction point.

Avalanche photodiodes (APDs) are used as photodetectors in the barrel and vacuum phototriodes (VPTs) in the endcaps. Both have a low photoelectron yield, with only 4.5 photoelectrons/MeV being produced in the photodetectors. However the scintillation decays of the two crystals is fast: 80% of the light is emitted within 25 ns[44].

In the endcap regions, where the angle between the two emerging photons from the decay of a neutral pion is likely to be very small, there is a risk that neutral pions mimic high-energy photons when they decay into two closely-spaced lower energy photons that the ECAL detects as a single energy deposition. To resolve the two photons coming from a neutral pion, the preshower calorimeter is installed. The preshower detector[51] is a sampling calorimeter, consisting of two layers of lead absorber each followed by a layer of silicon strip sensors. The first absorber layer is 2 radiation lengths thick and is followed by strips oriented along y. The second absorber layer is one radiation length thick and is followed by strips oriented along x. The strips have a pitch of 1.9 mm. Due to its fine granularity the preshower detector enhances photon and electron identification in the forward region $1.653 < |\eta| < 2.6$, where it provides more precise position and shower shape measurements.

The ECAL energy resolution can be parametrized as follows:

$$(\frac{\sigma}{E})^2 = (\frac{S}{\sqrt{E}})^2 + (\frac{N}{E})^2 + C^2, \tag{2.3}$$

with S the stochastic term, N the noise term and C the constant term. The stochastic term combines the fluctuations of the energy deposits and the photodetectors and has been measured to be 2.8% during testbeam operation. The noise term comes from the electronics and was measured to be 0.12 GeV. Finally, the constant term is determined by non-uniformity of the longitudinal light collection, leakage of energy at the back of the crystal and mainly by the calibration errors. In the testbeam this term was measured to be 0.3%. This will be the value for a perfectly calibrated detector. Mis-calibration will directly affect this term and degrade the energy resolution of

the ECAL. The constant will dominate the energy resolution at high energy. The measurement of the resolution terms during a beam test[52] in shown in fig. 2-10.

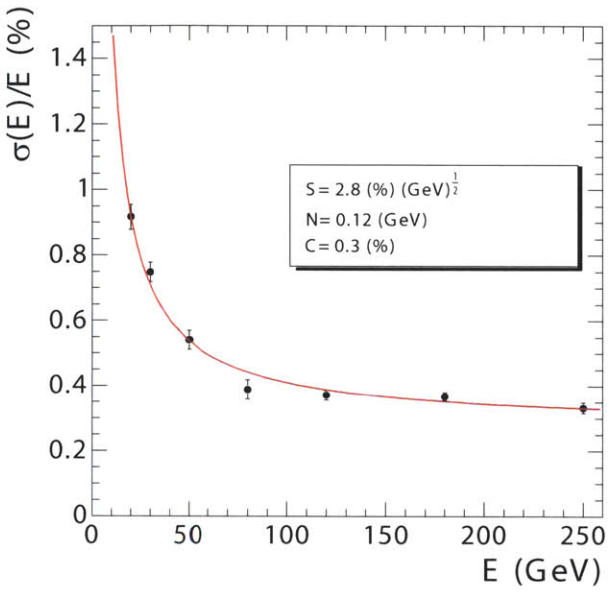

Figure 2-10: ECAL energy resolution, $\frac{\sigma}{E}$, as a function of electron energy as measured from beam test. The energy was measured in an array of 3x3 crystals with an electron impacting the central crystal. The stochastic (S), noise (N), and constant (C) terms are given[44].

The goal of the ECAL calibration is to achieve the most accurate energy measurement for electron and photons. The estimated particle energy, obtained from the ECAL, can be expressed as:

$$E_{e,\gamma} = G \times F \times \sum_{i}^{cluster} c_i \times A_i \,, \tag{2.4}$$

where G is the absolute ECAL scale and F accounts for energy losses due to bremsstrahlung and containment variations. The c_i factors are the inter-calibration coefficients while the A_i are the crystal signal amplitudes.

The main source of channel-to-channel response variation in the Barrel is the crystal-to-crystal variation of scintillation light yield which has an rms of roughly

15%. In the Endcaps, the crystal signal yield and the product of the gain, quantum efficiency and the photocathode area of the VPTs have an rms variation of almost 25%. To calibrate the inter-calibration constants, two main techniques have been used. The first method, the ϕ-symmetry method assumes that for a large sample of minimum bias events the total deposited transverse energy should be the same for all crystals in a ring at fixed pseudorapidity[53, 54]. The total energy deposited in each crystal is compared with the mean of the distribution of total energies for all crystals in a ring to calibrate the constants. Using π_0s and ηs decaying into two photons, further calibration can be used using the expected invariant mass. Fits to the invariant mass spectrum can be used to calibrate the absolute ECAL energy scale as well as further calibrate the inter-calibration constants, this time also in η[53, 54]. Improved measurements for the energy region of interest will be presented in this thesis, using the Z invariant mass spectrum in section 7.4.

Although radiation resistant, the ECAL crystals also show a limited but rapid loss of optical transmission under irradiation due to the production of color centers which absorb a fraction of the transmitted light. This will lead to dose-rate dependent fluctuations in the crystal transparency, which are expected to range from 1-2% at low luminosity in the barrel, to tens of per cent in the high η regions of the endcap at high luminosity. To correct for this, a laser system has been installed to inject laser pulses into the crystals via optical fibres. The signal of these laser pulses is measured and can be used to correct the changes in transparency[44, 47, 55]. This is also very important for the trigger, since it relies on the crystal energy measurement to select interesting events.

The CMS ECAL detector suffers from fake signals due to energy deposition by heavily ionizing particles in the avalanche photodiode. These signal shows up as a high energy deposit in one crystal with negligible deposition in neighboring crystals. It will be referred to as anomalous 'spikes' in the ECAL. Physically, such a signal is very unlikely to occur since the typical cross section of an electromagnetic shower in the crystals corresponds to the Molière radius of lead tungstate (2.19 cm). So we expect typical electromagnetic showers to extend over many crystals (crystal size:

2.4x2.4 cm^2 in barrel, 3.0x3.0cm^2 in endcap) and the number of crystals to increase with the particle energy. The same is true for hadronic showers in the ECAL. In the offline reconstruction such spikes are rejected by looking at the energy deposits in the 4 neighboring crystals of a high-energy crystal. If this energy is below a energy-dependent threshold, then we mask these crystals in order not to include them for the reconstruction. It was noticed that the spurious signals are not in phase with the beam crossing. Therefore an extra veto was applied on all deposits in excess of 2 GeV, that the measured time be within $\pm 5\sigma$ (typically a few ns) of the expected time[56]. The rate of these spikes is roughly one per 10^3 minimum-bias events.

2.2.4 The Hadronic Calorimeter

The hadronic calorimeter (HCAL) measures the energies of hadrons, particles made of quarks and gluons. The detector (fig. 2-11) also has to be as hermetic as possible, so that any hadronic particle in the detector will be detected. This directly relates to the estimate of missing transverse energy, for which a large value is a signature of the neutrino from our W decay. The barrel and endcap hadronic calorimeters completely surround the ECAL and are immersed in the magnetic field from the solenoid, maximizing the amount of interaction lengths contained inside the coil. They are hermetically joined, with the barrel part extending out to $|\eta| = 1.4$ and the endcaps covering the range $1.3 < |\eta| < 3.0$. The granularity of the calorimeters is $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 1.6$ and $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$ for $|\eta| \geq 1.6$[48, 44]. Because of the limited space inside the magnet, some of the hadronic energy in the barrel could still leak out further. Therefore, an outer hadron calorimeter (HO) or tail catcher is placed outside the solenoid completing the barrel calorimeter. The HO utilizes the solenoid coil as an additional absorber and is used to identify late starting showers and to measure the shower energy deposited after HB. This increases the number of radiation lengths in the HCAL to 10[48].

The CMS barrel and endcap hadronic calorimeters are sandwich-like sampling calorimeters, made of repeating layers of brass as a dense absorber and tiles of plastic scintillator. Brass is chosen as main absorber because of its relatively short interaction

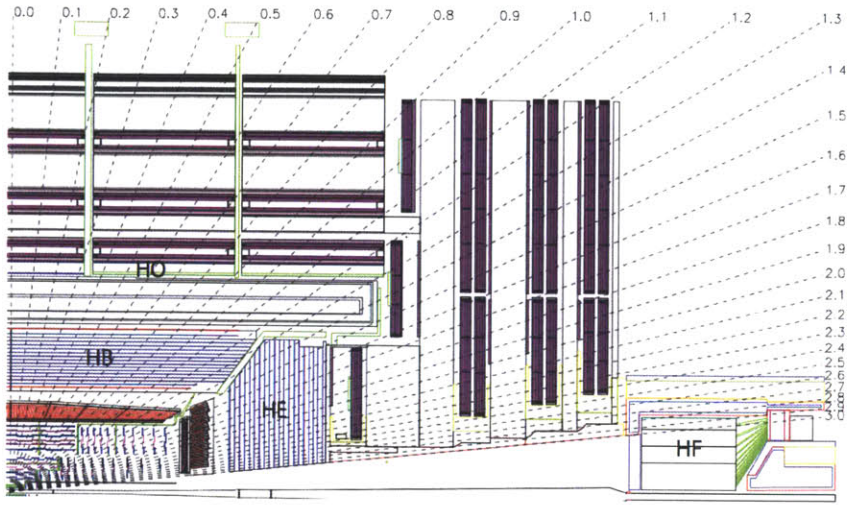

Figure 2-11: View of one quarter (r,z) of the CMS Hadronic Calorimeter[44].

length, its ease of use and because it is non-magnetic. When a hadronic particle hits a plate of brass, an interaction occurs producing numerous secondary particles. As these secondary particles flow through successive layers of absorber, they too can interact and a cascade or "shower" of particles results. As this shower develops, the particles pass through the alternating layers of active scintillation material causing them to emit blue-violet light. Within each tile optical "wavelength-shifting fibers" absorb this light and shift the blue-violet light into the green region of the spectrum, and clear optic cables then carry the green light away to the hybrid photodiodes that read them out. Every hybrid photodiode reads out up to 18 hadronic calorimeter cells.

The forward calorimeters (HF) of CMS extends calorimetric coverage to $|\eta| = 5$ to ensure excellent hermiticity. The design of the HF calorimeter was guided by the necessity to survive the challenging LHC conditions. Successful operation critically depends on the radiation-hardness of the active material. This was the principal reason why steel was chosen as absorber material and quartz fibers as the active medium. The signal is generated when particles reach velocities larger than speed of light in the quartz and send out Cerenkov radiation. This radiation is then guided

by the fibers to the photomultipliers. The fibers run parallel to the beam line, and are bundled to form 0.175x0.175 $\Delta\eta \times \Delta\phi$ towers[48, 44].

The detector is subdivided into two longitudinal segments. Half of the fibers run over the full depth of the absorber (165 cm, 10 interaction length) while the other half starts at a depth of 22 cm from the front of the detector. This allows for particle identification, distinguishing between electromagnetic and hadronic showers. Photons and electrons deposit most of their energy in the first 22 cm, while the hadronic ones deposit almost equal size signals in both parts.

The energy resolution of the HCAL is worse than that of the ECAL. Hadronic cascades typically have many fewer secondary particles than electromagnetic, leading to larger statistical fluctuations. Furthermore, the CMS HCAL is a sampling calorimeter. These factors lead to a much larger stochastic term in the energy resolution of the HCAL (3%) than in the resolution of the ECAL (0.5%). In fig. 2-12 the resolution of jets and missing transverse energy is shown. The E_T resolution for jets with E_T above 60 GeV is below 10%, while the resolution on the missing transverse energy is between 5 and 10% depending on the amount of activity in the event[57, 56].

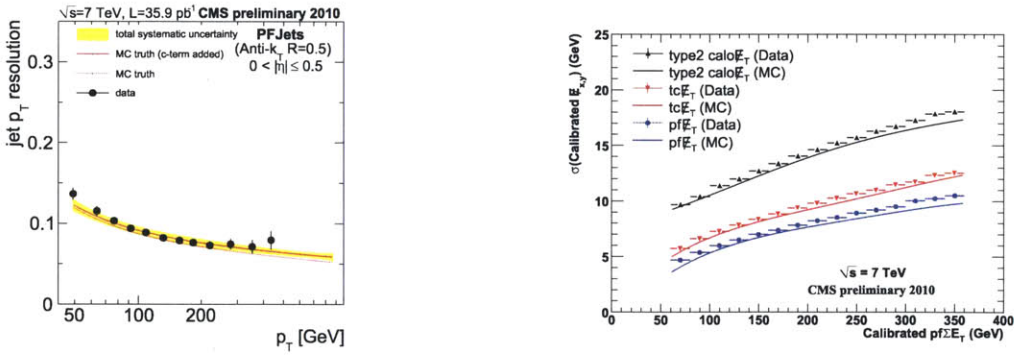

Figure 2-12: Transverse momentum resolution as a function of jet E_T (left) and missing transverse energy resolution as a function of the amount of activity in the events (right)[57, 56].

The electronic noise in the HCAL is are predominantly caused by a thermally emitted electron ionizing a gas or surface molecule in the acceleration gap of the Hybrid PhotoDiode (HPD). That ion is accelerated back to the cathode and liberates further electrons, causing a signal equivalent to many photo-electrons (ion feedback).

Also discharges in the HPDs can cause extra noise. A Hybrid PhotoDiode (HPD) is considered noisy if at least 17 HCAL cells were above 1.5 GeV threshold and then the measurement is not used in the offline reconstruction. Also the timing information can be used to discriminate: the ion feedback noise has a flat behavior in time, while the discharge happens fully in 1-2 timeslices (25 ns each). The nominal pulse shape spreads out over longer time than the discharge but less long than the ion feedback. Comparing the signal in the first 50 ns and the signal in 250 ns allows us to discriminate between real HCAL deposits and noise in the HPDs. Fake HCAL deposits coming from noisy HPDs are masked before the offline reconstruction is run[56].

In the Forward Calorimeter (HF), charged particles producing Cerenkov light in the window of the photomultipliers cause spikes similar to the ones in the ECAL. Again both timing and topological information was used to reject fake signals from noisy photomultipliers. The energy deposited in a long fiber normally has to be larger than the one deposited in a short fiber. For a hadronic shower we expect $E_{long} \approx 2E_{short}$, while for an electromagnetic shower $E_{EM} = E_{long} - E_{short}$. Therefore fibers with considerably less (factor 20) energy deposited in the long fibers than in the short fibers can be rejected. For the long fibers the algorithm is more complicated since electromagnetic showers will normally deposit most of their energy in the long fibers and very little in the short ones. It was noticed, however, that deposits in the long fibers in excess of 120GeV, which normally backed up by at least 1% of the long-fibre energy in the short fibres of the same HF tower. If this is not the case, the long fibers are rejected. Finally, a timing constraint is imposed. The measured time of HF deposits in excess of 30GeV has to be within ± 5 ns of the expected time, and the time distribution had to be compatible with that expected from a real energy deposit originating from the collision[58, 56]. Read-out signals that were flagged as noise by these algorithms, are ignored in the offline reconstruction algorithms.

2.2.5 The Muon Chambers

Because muons can penetrate several meters of iron without being stopped, unlike most particles they are not halted by any of the CMS calorimeters. Therefore, chambers to detect muons are placed at the periphery of the experiment where they are the only particles likely to register a signal. The muon system has 3 functions: muon identification, momentum measurement, and triggering. It is embedded in the the iron construction of the magnet's return yoke, such that muon momentum and charge measurements can also exploit the strong magnetic return field. CMS can measure muons up to $|\eta| = 2.4$.

The muon chambers in CMS (fig. 2-13) use 3 different kinds of gas detectors, all chosen for their specific qualities[50]. In the barrel region ($|\eta| < 1.2$) 4 stations of drift tubes are installed. The first three stations each contain eight chambers, in two groups of four, which measure the muon coordinate in the r-ϕ bending plane, and four chambers which provide a measurement in the z direction, along the beam line. The fourth station does not contain the z-measuring planes. Drift Tubes (DTs[50, 44]) are chosen for their excellent spatial resolution ($250\mu m$). The lower rates of muons in the central region compared to the forward region allows this type of detector, without substantial deadtime due to the longer drift times associated with these gas detectors. In the endcap ($0.9 < |\eta| < 2.4$), the muon rates and background levels are considerably higher. Therefore Cathode Strip Chambers (CSCs) are chosen for their fast response time, fine segmentation, and radiation resistance. Four layers of CSCs are installed in every endcap of CMS[50, 44]. Both types of muon chambers are augmented (for $|\eta| < 1.6$) with Resistive Plate Chambers (RPCs), which have inferior spatial resolution but excellent timing resolution, of order 3 ns, making them suitable for the trigger system. They also help to resolve ambiguities in attempting to make tracks from multiple hits in a chamber. In the barrel six layers of RPCs are installed, in the endcaps three layers per side[50, 44]. The final resolution of the muon chambers can be seen in fig. 2-14. It can be seen that the muon chambers start to improve the transverse momentum resolution for muons with p_T >100 GeV.

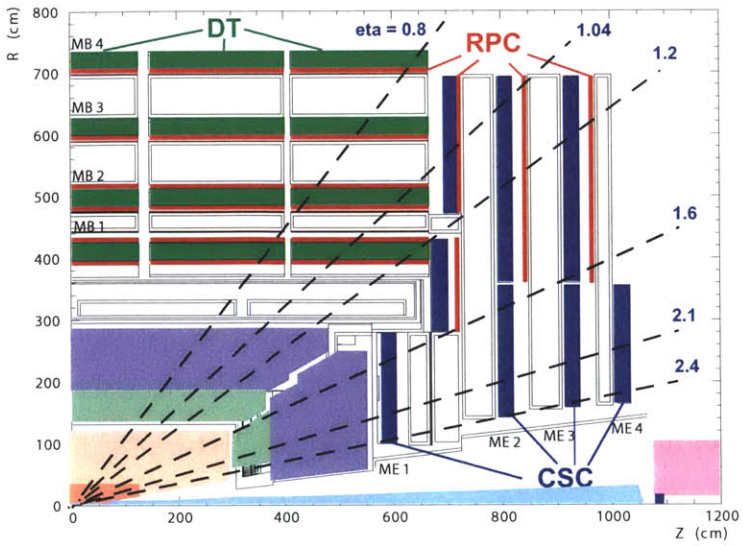

Figure 2-13: View of one quarter (r,z) of the CMS Muon Chambers[44].

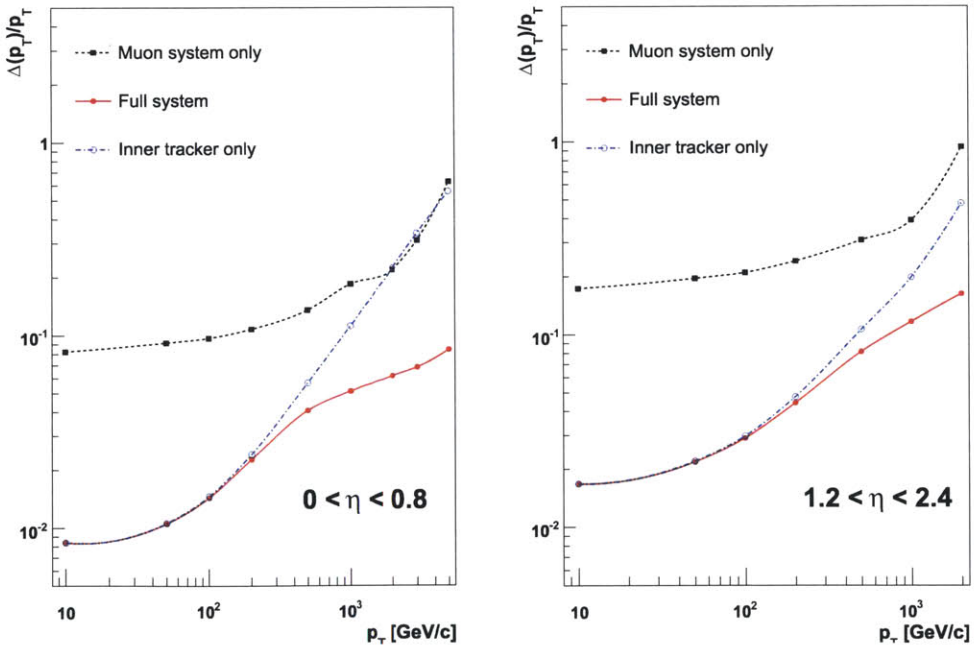

Figure 2-14: The muon transverse-momentum resolution as a function of the transverse-momentum using the muon system only, the inner tracking only, and both for $|\eta| < 0.8$ at the left and $1.2 < |\eta| < 2.4$ at the right[44].

2.2.6 CMS Trigger and Data-Acquisition

The CMS trigger and data-acquisition(DAQ) system has been designed to collect and analyze the data at the LHC bunch crossing frequency of 40 MHz[59]. Only a few hundred events can be written to tape, so the interesting events have to be selected. The various sub-detector Front-End systems store data continuously in 40-MHz pipelined buffers. The corresponding data is transferred to the DAQ via the Front-End-Drivers (FEDs) when a Level-1 trigger is sent. The L1 triggers only use calorimeter and muon chamber information to assess whether an event is interesting or not. It reduces the incoming data rate to 100kHz maximum. The Front-End Read-out Links read out the data from the FEDs and are able to merge data from two FEDs. The event builder combines the event fragments from all FEDs from one Level-1 trigger into a complete event. It is then passed on to the Filter Unit in a computer farm (Event Filter) for further processing. The High-Level-Trigger(HLT) runs on the Event Filter system and performs simple physics selections, using faster versions of the offline reconstruction software, to filter events and reduce the number of selected events with another factor 1000. More details about the exact triggers used in this analysis, can be found in section 3.3. In the data-acquisition a special system has been introduced to avoid overflow of the front-end buffers of the various sub-detectors. The information is fed back to the DAQ system which can then prevent Level-1 triggers to be sent. This will result in dead-time of the detector.

2.2.7 Luminosity Measurement in CMS

The luminosity measurement needs two steps: the absolute measurement to set the normalization and the real-time measurement to determine the differences between run periods with distinct beam conditions. CMS uses different techniques and cross-checks to ensure a reliable luminosity measurement.

Absolute Luminosity

The absolute luminosity is measured during a special run, called a 'Van der Meer'-scan, after the pioneer of such absolute luminosity measurements[15]. The two beams are scanned against each other in the transverse plane to determine the effective area A_{eff} of their overlap region. The most general luminosity formula for bunched beams can be written as:

$$L = \nu K \sum_{i=1}^{n} N_1^i N_2^i \int \rho_1^i(\vec{r}, t)\rho_2^i(\vec{r} - \Delta\vec{r}, t)d\vec{r}dt, \tag{2.5}$$

with the sum extending over the number of colliding bunch pairs, ν the beam revolution frequency, N_j^i the number of protons in bunch i of beam j, $\rho_1^i(\vec{r}, t)$ is the normalized particle density and K is a kinematic relativistic factor[60].

There are two independent methods for deriving the absolute luminosity scale. The first method assumes a double-Gaussian beam profile, which leads to a dependence of the luminosity on the beam displacements which can be measured. More details can be found in [60]. Another option is to exploit the excellent vertex resolution of CMS and integrate out the dependence of the luminosity formula on the beam separation[61]. In this way, the beam density functions can be extracted and used to calculate the effective area of the beam. As long as the beam displacements are the same in the case, the exact beam displacements do not have to be known. Both methods have been used in CMS as a cross-check and differences between them and between different runs have been used to assess a component of the systematic uncertainty in the luminosity determination. The largest systematic uncertainty comes from the measurement of the beam currents, which is used to determine the amount of protons in a bunch. The bunch currents are measured by dedicated devices, called Fast Beam Current Transformers[62]. The final uncertainty on the luminosity measurement is 4%[61].

Real-time Luminosity Measurement

The overall luminosity normalization is not sufficient, we need to know the instantaneous luminosity to measure how much data was taken in a specific run period when the whole detector was read out and a certain trigger scheme was used. To do this, a real-time luminosity measurement is added to the absolute one as described before. There are online and offline methods to do this. Online, two algorithms are used. The first one is zero-counting, where the average number of empty towers in the HF ($3 < |\eta| < 5$) is used to infer the mean number of interactions per bunch crossing[60]. The standard zero-counting method would look at the number of interactions follows a Poisson distribution with μ the mean number of interactions per event:

$$p(n, \mu) = \frac{\mu^n e^{-\mu}}{n!} . \tag{2.6}$$

The number of empty events can now be written as $p(0, \mu) = e^{-\mu}$ and thus the mean number of interactions follows the following formula:

$$\mu = -ln(p(0, \mu)) . \tag{2.7}$$

This method only works as long as there are sufficient amount of events without any interactions. At the luminosities obtained that is not the case anymore and so-called 'zero famine' sets in. To defeat the zero famine at high luminosity, zeros are counted in a much smaller solid angle, namely the HF towers. This provides 864 quasi-independent measurements of the luminosity which are averaged to give the final luminosity estimate. The second method exploits the linear relationship between the average transverse energy per tower in the HF and the luminosity. For this measurement, only towers in the range $3.5 < |\eta| < 4.2$ are used to obtain the most linear dependence[60]. Figure 2-15 shows the linear behavior between these two techniques and the real-time luminosity.

Offline two further methods are used to cross-check the luminosity. The first one is related to the online measurement. It looks at the coincidence of $\sum E_T$ deposits

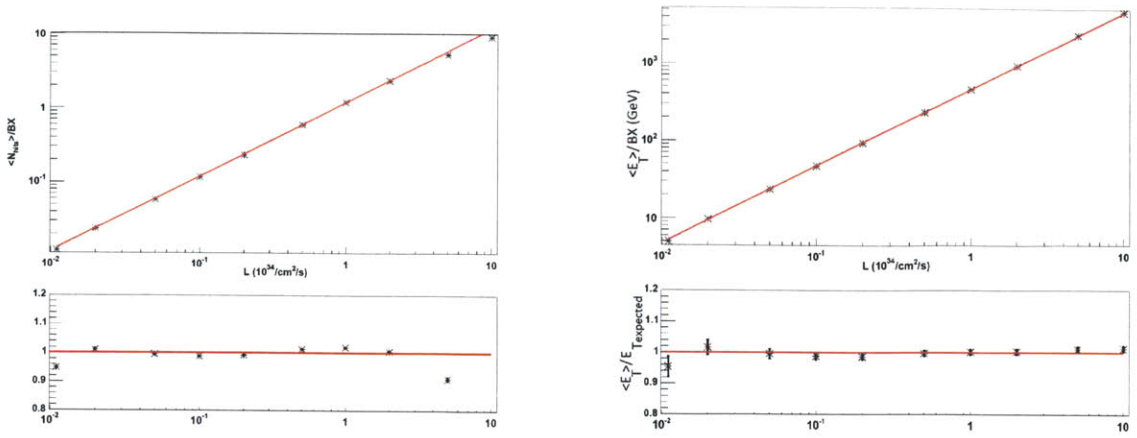

Figure 2-15: Two methods to estimate the real-time luminosity: zero-counting method(left) and linear relationship between average tower energy and luminosity (right).

above 1 GeV in the forward and backward HF. The number of deposits yields an estimate for the luminosity[60]. Timing cuts are added to eliminate non-collision backgrounds. The second method looks for vertices with at least 2 tracks. The z-position of these vertices has to be within 150 mm of the interaction region. Zero-bias triggers (coincident beam pick-up) are used to eliminate possible biases due to trigger efficiency or dead time[60]. All the methods are consistent within uncertainties.

LHC Operation in 2010

The LHC delivered 47 pb^{-1} to the CMS experiment in 2010. (fig. 2-16) CMS had a overall data-taking efficiency of 92% and recorded 42 pb^{-1} of this data to tape. The instantaneous luminosity in this period increased over 5 orders of magnitude, requiring almost continuous changes to the triggering scheme. Roughly 84% of the recorded data was declared 'good for physics', implying that all detectors ran stably in this period. This 36 pb^{-1} of good pp collision data at 7 TeV is the dataset used for this analysis. The systematic uncertainty on this luminosity measurement is 4% and this will be the dominant systematic uncertainty on the cross section measurement.

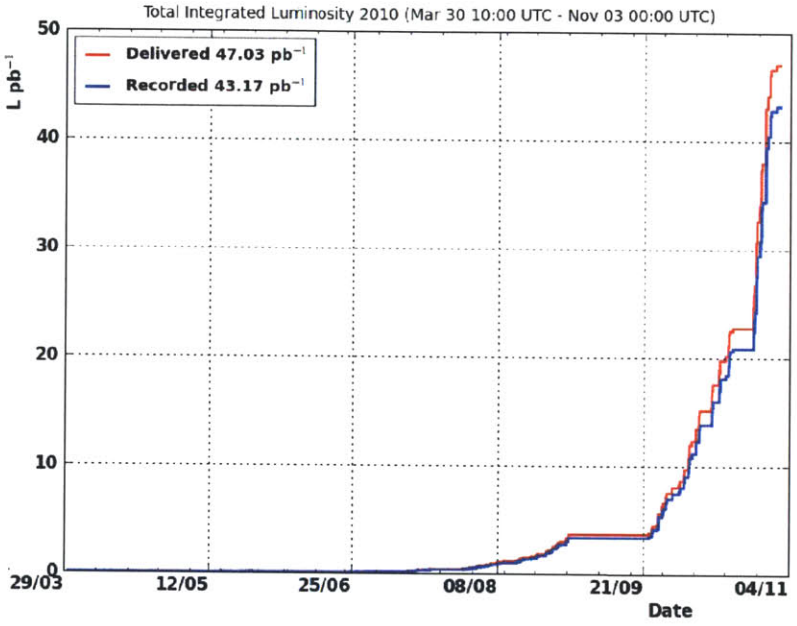

Figure 2-16: Delivered and recorded luminosity for the CMS experiment in 2010

Chapter 3

Online and offline data selection

One of the most important signatures of the W is a prompt high-p_T electron. In this chapter we will describe how this electron is reconstructed and how it is distinguished from jets. We will start with a discussion of electron reconstruction, move on to discuss the different isolation and identification cuts, and finish with the trigger selection. The online selection will use many of the cuts and variables from the offline analysis. Therefore, the online selection is only discussed after the offline selection even though the trigger is the first step in the analysis chain. Finally, the reconstruction of missing transverse energy will also be discussed.

3.1 Electron Reconstruction

To reconstruct an electron, both the electron track and its energy deposits in the calorimeter need to be well-measured. This is done using specific algorithms: super-clustering algorithms to describe the calorimeter reconstruction, and the dedicated electron track reconstruction, the 'GSF tracking'. These two components will then be matched to create an electron.

3.1.1 Superclusters

The electrons shower in the electromagnetic calorimeter (ECAL) and deposit their energy typically in a few crystals, which can then be combined in a cluster of crystals. Before reaching the ECAL, the electrons have to cross the tracker material. They will lose energy due to bremsstrahlung, which leads to a spread of the energy deposits in the ϕ direction because of the bremsstrahlung photons. Those photons will not necessarily be part of the electron cluster. Nevertheless, they have to be counted to estimate the original energy of the electron. To collect such bremsstrahlung events, clusters of clusters are formed. These are the so-called super-clusters. There are two methods being used for supercluster creation, the Hybrid Algorithm for the barrel, the Multi-5x5 Algorithm for the endcap[63].

The Hybrid algorithm exploits the knowledge of the shower shape in the η direction by making fixed η-sized domino blocks while searching dynamically for bremsstrahlung contributions in ϕ. The algorithm is illustrated in figure 3-1. The starting point is a seed crystal, which is defined as the highest transverse energy crystal (minimum energy 1 GeV) in the search region. Starting from this point adjacent crystals are grouped in 'dominoes'. These dominoes are 1x3 ($\Delta\phi$x$\Delta\eta$) crystals large if the central crystal is below a certain energy threshold, 1x5 crystals if the central crystal is above a certain threshold. At this moment, the threshold is 0 GeV, so 1x5 dominoes are always built[63].

Superclusters are then formed in the following way: seventeen steps of one crystal are taken in each direction in ϕ and new 1×5 ($\Delta\phi\times\Delta\eta$) dominoes are created in every step. The central crystals of these new dominoes are at the same η-position as the the seed crystal. Dominoes will be considered to be included in a cluster if they are above a 0.1 GeV threshold. The selected dominoes are grouped together around local maxima: dominoes which have a larger energy than their neighbors. Every group of dominoes around such a local maximum forms a cluster. If the local maximum domino has an energy below 0.35 GeV, the corresponding cluster is discarded. The remaining groups of dominoes form a cluster of clusters, a so-called supercluster.

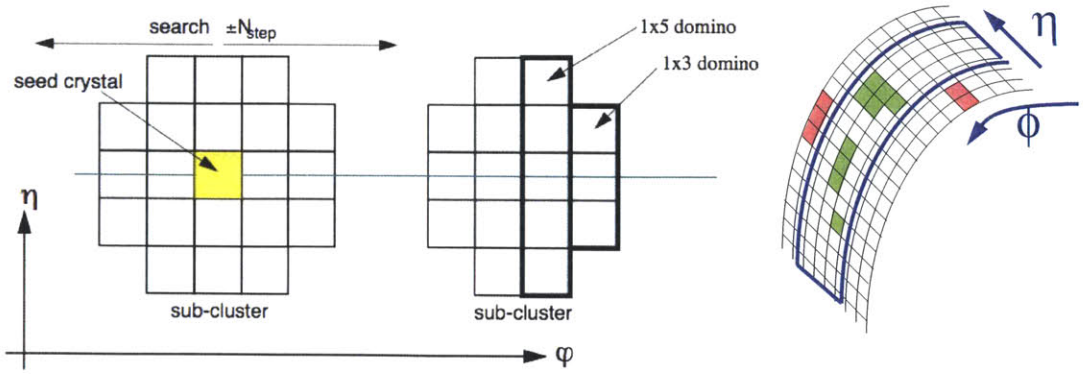

Figure 3-1: Hybrid clustering algorithm used in ECAL barrel: combining 1x5 dominoes into clusters around local maxima shown at the left and the investigated area along ϕ with narrow η-width (5 crystals) at the right[63].

In the endcap the crystals are not arranged in a $\eta - \phi$ geometry as in the barrel and thus the hybrid algorithm cannot be used. The Multi-5x5 algorithm (fig. 3-2) tries to collect energy deposits in an $\eta - \phi$ window in a different way. It starts from a list of crystals in descending order of E_T. If the next crystal in the list does not belong yet to a cluster, then it is subjected to a minimum seeding requirement (E_T >0.18 GeV) and compared to its 4 adjacent neighbors (Swiss cross pattern). If it is a local maximum, a 5x5 matrix of clusters around the seed is constructed, using only the crystals which do not belong yet to another cluster. If the crystal is not a local maximum compared to the 4 adjacent crystals, the process restarts with the next crystal in the E_T-ordered list. To allow for closely overlapping bremsstrahlung-clusters, the outer 16 crystals of a 5x5 matrix can seed a new cluster and the matrices can overlap, but crystals can only belong to one cluster[63].

Creation of superclusters occurs after the above clustering procedure is carried out. Superclusters start with seed clusters with an energy above 1 GeV, around which a rectangular window of 0.6x0.14 ($\Delta\phi \times \Delta\eta$) is built. Other clusters within this window are then added to form the supercluster. This search is done as a function of descending E_T and a cluster can only belong to 1 supercluster.

In the endcap region CMS has a preshower detector (1.6< $|\eta|$ < 2.6). Electrons

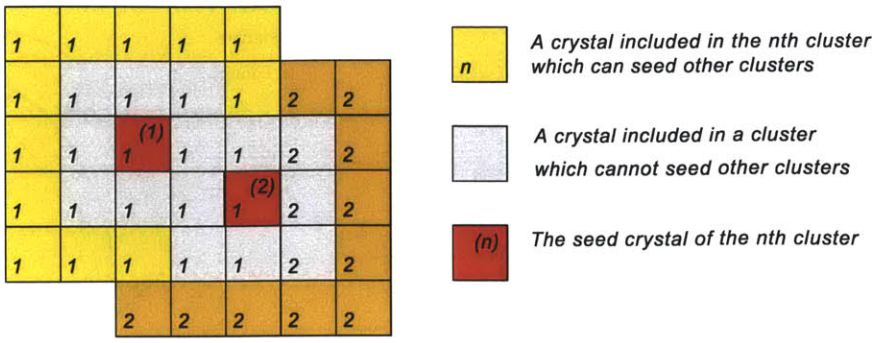

Figure 3-2: Multi 5x5 clustering algorithm used in ECAL endcap.

will deposit part of their energy in this detector so this should be added to the ECAL deposits to create the correct supercluster energy. This is done by extrapolating between the position of each deposit in the calorimeter and the primary vertex and summing up the energy of the preshower strips on the intersection of this trajectory with the preshower. The summed-up energy is then added to the corresponding supercluster before any corrections are applied.

Even after the superclustering procedure, the ECAL energy measurements do not return the true energy of the electron yet, due to leakage out of the crystals and energy lost in the tracker. Therefore, energy scale corrections are applied. There are three types of corrections. The first correction compensates for lateral energy leakage from the exposed faces of the ECAL barrel crystals. It is only applied for barrel clusters. The second correction alters the response of the clustering algorithm to the shower, this correction will depend on the dimensions of the cluster. This correction determines the fraction of energy depositions found by the algorithm will be used to assess the final energy, based on the ϕ-width of the cluster. Finally, a third correction is added for the non-linear distribution of matter in the detector and of the energy dependence of the calorimeter response. All these corrections are estimated in simulation. The last correction factor, that addresses the non-linear material distribution and the energy dependence, can also be calculated from data

using $Z \rightarrow e^+e^-$ events, where the known invariant mass of the Z provides a calibration for ECAL energy corrections. Up to now $Z \rightarrow e^+e^-$ events were only to cross-check the overall scale of the correction.

At this moment we have the calorimeter measurement of the electron, in the form of a supercluster with corrected energy. The supercluster position is the energy-weighted mean of the positions of its constituent clusters. These clusters should combine all the energy radiated by the electron and so this position corresponds to that of a non-radiating electron.

3.1.2 Gaussian-Sum-Filter (GSF) tracking

The supercluster can now be used to select to select seeds for starting the electron tracking. The supercluster position is extrapolated backwards to the primary vertex on a helix whose bending is calculated from the supercluster E_T. This is done twice, once for each charge hypothesis as can be seen in fig. 3-3. In a large window around the interpolated helix, a compatible hit is searched for in the first or second pixel layer. If this first hit is found, then the helical trajectory can be better refined and the search for a second hit is performed is a smaller window. In the endcaps, also the first layer of the strip detector can be used to look for a second hit if it has not been found in the pixel endcap disks. This is done to maintain high efficiency at large η even though there are only two pixel endcap disks. The windows used can be found in table 3.1. For the first hit, an E_T-dependent ϕ-window is used to further reduce the contamination from fake electrons from jets.

E_T	first window		second window			
	δz or δr_T	$\delta\phi$	δz	δr_T (PIX)	δr_T (TEC)	$\delta\phi$
10 GeV	$\pm 5\sigma_z$	[-0.14,0.08] rad	$\pm$0.09 cm	$\pm$0.15 cm	$\pm$0.2 cm	$\pm$4 mrad
35 GeV	$\pm 5\sigma_z$	[-0.05,0.03] rad	$\pm$0.09 cm	$\pm$0.15 cm	$\pm$0.2 cm	$\pm$4 mrad

Table 3.1: Definition of the seed matching windows. The E_T-dependent first window extension is given for 10 and 35 GeV. σ_z is the beam spot width along the z-axis. Asymmetric windows are shown for the positive charge assumption.

Electron seeds are then used to initiate a dedicated electron track building and

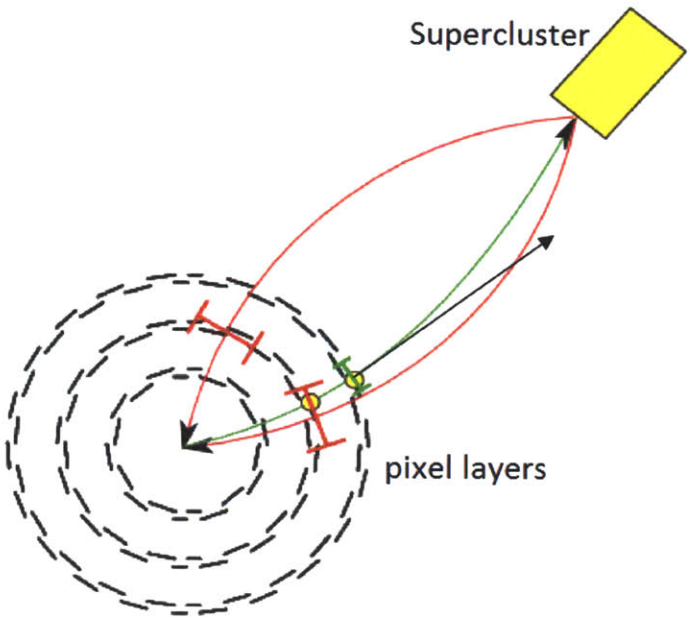

Figure 3-3: Propagating helix back from the supercluster to primary vertex to find track seeds.

fitting procedure in order to best handle the effect of bremsstrahlung energy loss[64]. The track finding is based on a combinatorial Kalman Filter[65], with a dedicated Bethe-Heitler modeling[66] of the electron energy losses. In order to preserve efficiency and to follow electron trajectories in case of bremsstrahlung emission, a very loose χ^2-compatibility is required in the building steps of the electron tracking. Only the best five candidate hits (giving the smallest χ^2) on every tracker layer are kept and only one layer with missing hits is allowed to reduce the combinatorics. The hits collected in this way are passed to a Gaussian Sum Filter (GSF) for the final estimation of the track parameters. The Gaussian Sum Filter is a modified version of the standard CMS Kalman Filter[65] where the Gaussian modeling of the energy loss in each layer is replaced by a weighted sum of Gaussian distributions. The GSF leads to multi-component trajectory states for each measurement point, with weights for each component describing the associated probability[64]. The Gaussian Sum Filter allows to model the non-Gaussian scattering of the electrons when propagating through the tracker layers. Changes in the track direction or in its curvature are correctly taken into account. This Gaussian Sum Filter is based on a multi-Gaussian approximation

of the Bethe-Heitler model of bremsstrahlung emission[67]. To calculate the track parameters, the weighted mean or the mode —the highest-weight component— of the Gaussian mixture could be used. The mode has been shown[64, 63] to give a more precise estimate of the track parameters, since it less affected by bremsstrahlung. This GSF procedure is too slow to perform on every pixel seed, which is why the supercluster-matching is used to reduce the combinatorics[64].

3.1.3 GSF-Electrons

To combine the GSF-track and the supercluster into a GSF-electron, some extra constraints need to be fulfilled. The applied selection asks for a loose geometrical matching between the supercluster and the track's projected interpolation, and imposes some requirements on the energy depositions to make the object more compatible to what one expects from an electron. The cuts imposed are the following:

1. $E_T > 4\,GeV$, an energy requirement on the supercluster transverse energy;
2. $H/E < 0.15$, the ratio of the HCAL energy deposits in a cone of ΔR of 0.15 around the supercluster to the supercluster energy;
3. $|\Delta\eta_{in}| = |\eta_{sc} - \eta_{in}^{extr}| < 0.02$, matching in η;
4. $|\Delta\phi_{in}| = |\phi_{sc} - \phi_{in}^{extr}| < 0.1$, matching in ϕ,

where η_{sc} and ϕ_{sc} are the energy-weighted position of the supercluster and η_{in}^{extr} and ϕ_{in}^{extr} the position of closest approach to the supercluster, extrapolated from the track position and direction at the vertex. The supercluster folds in the bremsstrahlung of the electron and therefore the final position of the supercluster corresponds to the electron at the vertex, before any bremsstrahlung[64].

For our analysis, we will be exclusively using the energy of the supercluster, corrected with the primary vertex position, because this will allow us to use a consistent energy definition throughout the whole reconstruction chain. The electrons in the gap region ($1.4442 < |\eta| < 1.566$) will not be used, so that the supercluster energy measurement is reliable. The energy of the electron object can be calculated using a

weighted sum of the supercluster energy and the track momentum, with weights defined as the normalized inverse of the variance of each measurement. This is ignored in a few known situations where the two measurements disagree strongly and then one of the two measurements is taken. As an example, electrons in the ECAL gap region usually do not deposit all their energy in the ECAL and therefore only use the track momentum to estimate their magnitude. They are recovered using the more complicated energy definition[64].

3.2 Electron Identification and Isolation

In a pp collider, QCD backgrounds are around a million times more frequent than the isolated electrons of the W signal, as can be seen in figure 3-4. So even a small percentage of jets faking real electrons can give a large background sample. The main categories of single-electron backgrounds are:

1. Inelastic charge exchange of charged pions and kaons: A charged pion or kaon interacts with the protons and neutrons in the nuclei of the electromagnetic calorimeter. This creates a neutral pion which decays to two photons. The entire energy of the charged pion is thus deposited in the calorimeter. The result is a track and an ECAL deposit where the track momentum and energy agree (E≈p) and an impact parameter consistent with the primary vertex. The pion or kaon can also just start showering early on in the ECAL.

2. Converted photons: the photon can come from a pion decay, a prompt photon, or a hard bremsstrahlung photon. The photons convert in the tracker material into two electrons. The conversion can then be identified using the second electron coming from the conversion. If it converts in an asymmetric fashion such that one electron carries most of the photon energy, then it can be difficult to identify the second leg of the conversion.

3. Semileptonic decays of heavy flavor particles: a B or D hadron decays semileptonically into an electron carrying the majority of the momentum.

4. matching ambiguity: a photon cluster in the electromagnetic calorimeter is matched to an unrelated track, e.g. neutral pion photon together with a track from a charged pion in a jet.

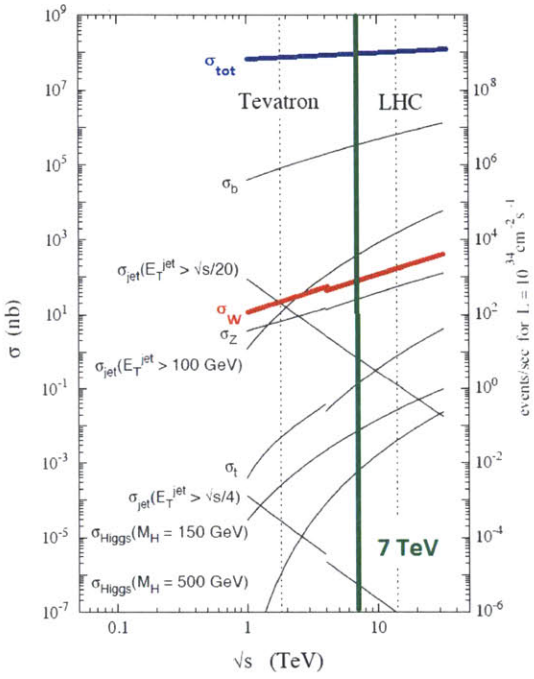

Figure 3-4: Cross-sections for QCD, EWK and Higgs processes at Tevatron and LHC.

The signal electrons from the decay of prompt Ws are separated from QCD and photon backgrounds by exploiting the differences in the signals these sources leave in both the tracker and the calorimeters.

An electron can be identified by its GSF track and supercluster. For a real electron, these two should be well-matched since they come from the same object. π^0's on the other hand only lead to deposits in the ECAL, the GSF track associated with it is usually from a charged pion and thus the match criteria are often not satisfied. To discriminate between real electrons and fakes a geometric matching between the track and the supercluster is used. The track is extrapolated using the track parameters at the vertex to the supercluster and compared to the supercluster position. The matching is done both in the azimuthal and the pseudo-rapidity plane: the variables used are: $|\Delta\eta_{in}|=|\eta_{SC} - \eta_{in}^{extr}|$ and $|\Delta\phi_{in}|=|\phi_{SC} - \phi_{in}^{extr}|$.

A second property of the electron that can be used is its shower shape. An electron shower is wide in the azimuthal plane (ϕ) because of the bremsstrahlung and the electron's curved trajectory in the magnetic field, but in general the shower is very narrow and well-confined in pseudorapidity (η). Jets that shower early, on the other hand, cause wide showers in both η and ϕ. Thus the width of a shower in η can be used as a discriminant. The variable used in CMS is the root-mean square of the shower width in η for a 5x5 crystal array centered on the highest-energy crystal (seed crystal): $\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_i^{5x5} w_i \, (0.0175\,\eta_i^{cryst} + \eta^{seed} - \overline{\eta}^{5\text{x}5})}{\sum_i^{5\text{x}5} w_i}}$ with weight factor $w_i = 4.2 + ln(\frac{E_i}{E_{5\text{x}5}})$[68]. This variable gives an estimate of the width of the shower in η in units of crystal widths. Another characteristic of electrons is that they tend to deposit the majority (if not all) of their energy in the ECAL and very little in the HCAL, while jets will lose the majority of their energy in the HCAL. This effect is exploited by using an H/E cut and thus asking for a large energy deposit in the ECAL and a small one in the HCAL.

Electrons from W decays are also isolated. Thus requiring extra energy in the tracker, ECAL or HCAL around the electron object can point towards jets. Jets will contain a large number of extra soft tracks (e.g. from charged pions/kaons) and extra deposits in the calorimeter (from all pions/kaons). Also electrons from heavy quark decays will not be isolated since they are also contained in a jet with a lot extra activity around the real electron. In this analysis, tracker, ECAL and HCAL isolation are used with the following definitions:

- Track isolation: sum of transverse momenta of tracks with momenta over 0.7 GeV in a cone of $0.015 < \Delta\text{R}(=\sqrt{\Delta\phi^2 + \Delta\eta^2}) < 0.3$. Also a strip of $\Delta\eta$ of width 0.015 is excluded. The tracks are required to match to the electron's primary vertex within 0.2 cm.

- ECAL isolation: sum of transverse energy of crystals in a cone of $\Delta\text{R}<0.3$ around the supercluster position. A strip along ϕ with an η width of 3 crystals is vetoed from the calculation to exclude the electron deposits. There is also a noise cut-off: only crystals with an energy above 0.08 GeV in the barrel and a transverse energy above 0.1 GeV in the endcap are used in the calculation.

- HCAL isolation: sum of the transverse energy of CaloTowers (clustered energy deposits with energy>0.7 GeV(barrel)or 0.8 GeV(endcap)) in a cone of 0.015 $< \Delta R < 0.3$.

The isolation variables were used separately in the selection and not combined in one variable, since this gives a better insight into what each cut does exactly. For applying the isolation cuts, relative isolation was used: the isolation variables are divided by the p_T of the electron, which accounts for more leakage of electron energy into the isolation cone for higher-p_T leptons.

A visual interpretation of the identification and isolation variables can be seen in fig. 3-5. N-1 plots show the distribution of a specific variable after applying all the other selection cuts. Such plots are useful to look at the efficacity of a specific cut and the possible improvement by strengthening or loosening the cut. N-1 plots for the identification variables are shown in fig. 3-6, for the isolation variables in fig. 3-7.

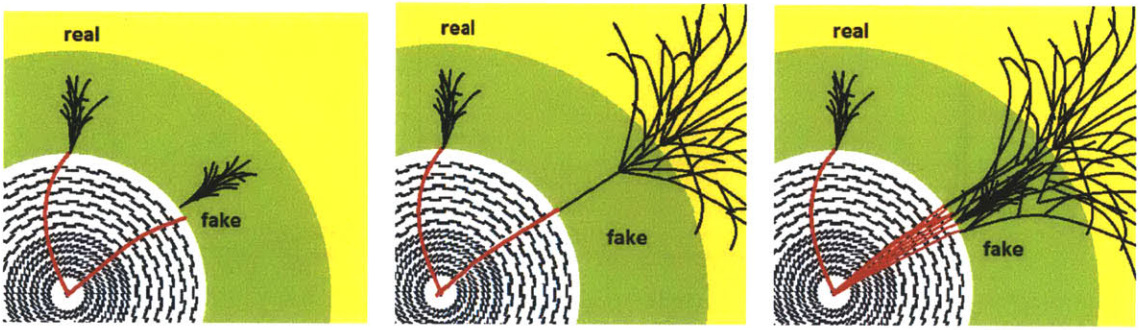

Figure 3-5: Real electrons from W decays and fakes can be distinguished by looking at track-cluster matching (left), the lateral and longitudinal profile of the shower(middle) and the isolation of the electron (right). The ECAL is shown in green and the HCAL in yellow.

Electrons from photon conversions are also an important background, because they lead to real, isolated electrons. Therefore specific cuts are applied to remove conversions. For the first cut the hit pattern of the track is investigated. The track is extrapolated back to the primary vertex and if it crosses the first pixel layer but no hit can be found in this layer, then the electron is rejected as a conversion. Conversions do not happen at the primary vertex but later in the detector volume when crossing

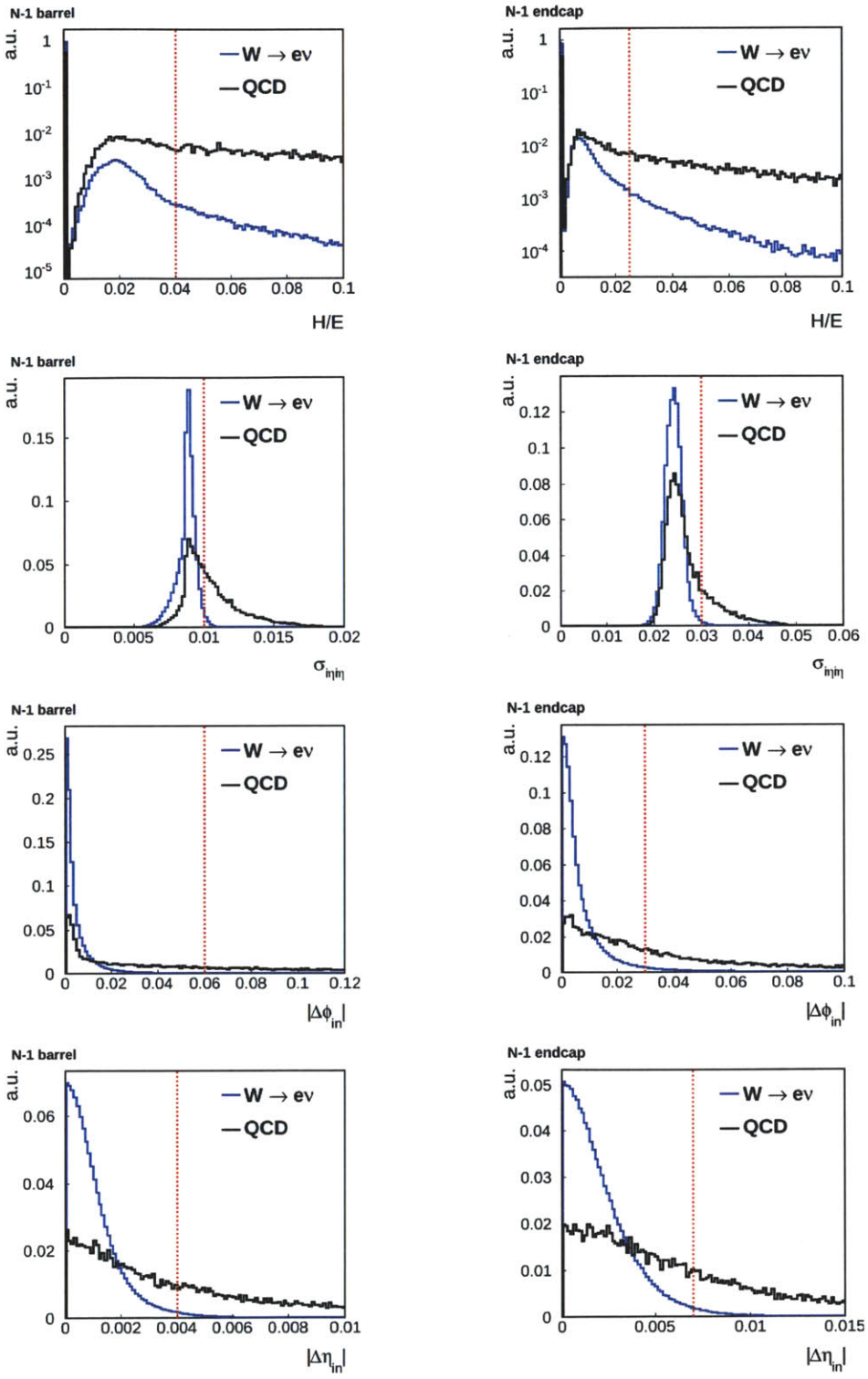

Figure 3-6: N-1 plots for the electron identification variables in barrel (left) and endcap (right). All identification and isolation cuts are applied except for the cut on the variable under investigation. The cut values can be found in table3.2. The red dotted line indicates the cut value used in the final selection.

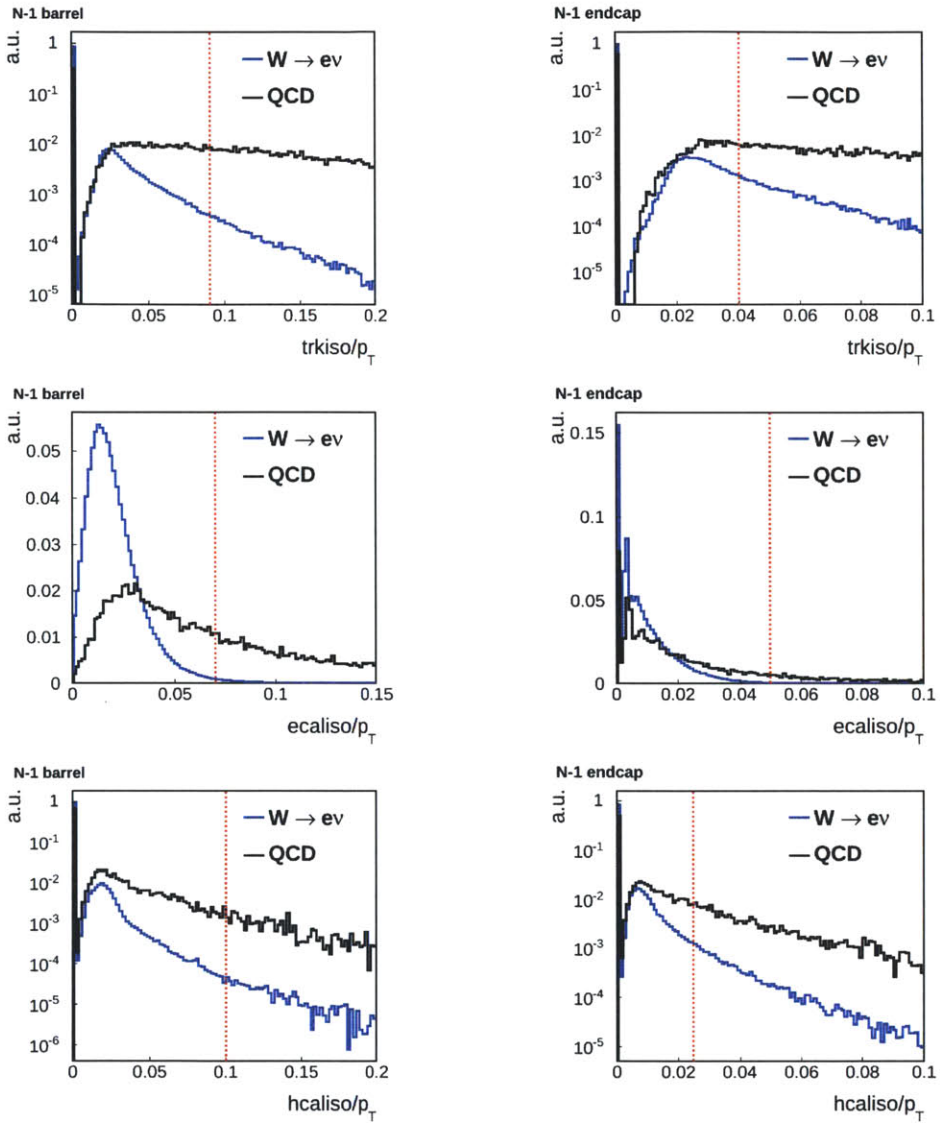

Figure 3-7: N-1 plots for the electron isolation variables in barrel(left) and endcap(right). All identification and isolation cuts are applied except for the cut on the variable under investigation. The cut values can be found in table 3.2. The red dotted line indicates the cut value used in the final selection.

material. Thus the first valid hit of a conversion electron track might not be in the first tracker layer. This cut already removes the majority of the conversions. The algorithm takes into account the presence of non-active layers to minimize the loss in efficiency for real electrons.

The second approach is to look for the partner leg of the conversion. The two tracks of a conversion pair are parallel to each other at the decay point and stay so in the r-z plane. The standard tracks in a cone of 0.3 around the GSF track of the electron are used for discrimination, making sure the standard track is not due to the electron itself. The $\Delta cot(\Theta)$ and distance between the tracks are investigated. For the $\Delta cot(\Theta)$ the difference in $cot(\Theta)$ for the standard track and the GSF track at the vertex is used. θ is the polar angle of the track with respect to the positive z-axis in CMS. The 'track-distance' parameter is the distance in x-y plane between the two tracks at the point where they would be parallel in that plane if they are extrapolated, that is for the best estimate of the decay point. If both quantities are small (<0.02), then it is likely to be a conversion and the electron is rejected.

Using all these cuts, an optimization was performed, trying to keep the highest efficiency possible while rejecting the most jets. This resulted in several so-called Working Points, these give the best jet background rejection for a certain efficiency. For the W analysis we will be using WP80, a set of selection cuts that should give roughly an 80% signal efficiency for the isolation and identifation criteria. This selection rejects a lot of background while keeping the efficiency sufficiently high. WP80 cuts are listed in table 3.2 and shown as a red line in figures 3-6 and 3-7. An extra cut was added to reduce the amount of $Z \rightarrow e^+e^-$ background: events with a second electron above 20 GeV, where the electron passes a looser set of identification and isolation cuts, are discarded. This looser set of identification and isolation values is the WP95 as described in table 3.3.

Variable	Barrel	Endcap
$\sigma_{i\eta i\eta}$	< 0.01	< 0.03
$\Delta\phi_{\text{in}}$	< 0.06	< 0.03
$\Delta\eta_{\text{in}}$	< 0.004	< 0.007
H/E	< 0.04	< 0.025
Isolation		
I_{track}/p_T	< 0.09	< 0.04
I_{ECAL}/p_T	< 0.07	< 0.05
I_{HCAL}/p_T	< 0.10	< 0.025
Missing expected hits	$= 0$	$= 0$
Partner track conversion veto		
track-distance	> 0.02	> 0.02
OR $\Delta \cot\theta$	> 0.02	> 0.02

Table 3.2: WP80 identification requirements for an ECAL-driven GSF-electron. For the partner track conversion veto, the electron candidate is kept when for its partner track the track-distance or the $\Delta \cot\theta$ is larger than 0.02.

Variable	Barrel	Endcap
$\sigma_{i\eta i\eta}$	< 0.01	< 0.03
$\Delta\phi_{\text{in}}$	< 0.8	< 0.7
$\Delta\eta_{\text{in}}$	< 0.007	< 0.01
H/E	< 0.15	< 0.07
Isolation		
I_{track}/p_T	< 0.15	< 0.08
I_{ECAL}/p_T	< 2.00	< 0.06
I_{HCAL}/p_T	< 0.12	< 0.05
Missing expected hits	≤ 1	≤ 1

Table 3.3: WP95 identification requirements for an ECAL-driven GSF-electron.

3.3 Trigger Selection and Event Skimming

During the first months of LHC data-taking, the running conditions changed dramatically, with the luminosity increasing by several orders of magnitude. This was reflected in the triggers used for the analysis, which had to change often to adapt to the new conditions. At the beginning, we could afford more inclusive triggers and tighten the selection offline, because the overall event rate was rather low, but with higher and higher luminosity, we had to be more selective in order not to saturate the bandwidth of the data-acquisition system. Starting with photon triggers, the triggers became more complex and closer to the offline selection with increasing luminosity and also the p_T thresholds had to be raised. The old triggers would then be pre-scaled (only a fraction of the events triggered would be kept) or even completely abandoned. This was possible because our knowledge of the detector improved and the planned triggers were studied during the low-rate data periods, so the new triggers could be optimized before the old triggers had to be pre-scaled.

3.3.1 Level-1 trigger

The Level-1 trigger consists of custom-designed, largely programmable electronics. The Level-1 trigger uses coarsely segmented data from the calorimeters and muon chambers, while holding the high-resolution data in pipelined memories in the front-end electronics. The Level-1 will only allow very crude energy and momentum measurements with reduced resolution. The Level-1 trigger has to analyze every bunch crossing. The allowed Level-1 trigger latency, between a given bunch crossing and the distribution of the trigger decision to the detector front-end electronics, is 3.2 μs. The sub-detectors have front-end pipeline buffers which allow to store this 3.2 μs of data and after a positive Level-1 decision this data will be pushed downstream to the data-acquisition system. This sets the scale for how fast the Level-1 trigger has to make its decisions.

The Level-1 trigger algorithm (fig. 3-8) for electromagnetic objects is based on electromagnetic trigger towers at fixed positions and works the same for electrons

and photons[59]. In the ECAL barrel trigger, trigger towers are 5×5 crystals large ($\Delta\eta \times \Delta\phi = 0.87 \times 0.87$). In the endcap where the crystals are arranged in a x-y geometry, the number of crystals clustered together in a trigger tower is not fixed and does not follow $(\eta - \phi)$ boundaries (fig. 3-9). In the endcap he number of crystals in a trigger tower depends on rapidity; the size of the towers is decided based on the background rates in the different rapidity regions. In the barrel and in the endcap, the boundaries of ECAL and HCAL trigger towers follow each other. Each trigger tower corresponds to the $(\eta - \phi)$ size of an HCAL physical tower, except for $|\eta|$ >1.74 where the HCAL tower has twice the ϕ-dimension of the trigger tower. In this region, the HCAL tower energy is divided in equal amount and assigned to two trigger towers that are contained in it.

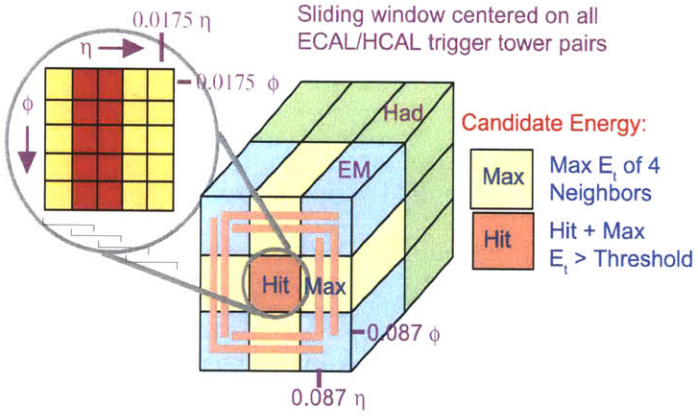

Figure 3-8: Electron L1 trigger algorithm.

Summing the E_T in the central hit tower with the maximum-E_T tower of its four neighbors, gives the transverse energy of the trigger. For a non-isolated trigger, as the one used in this analysis, two extra requirements are added. The first one is based on the longitudinal shower profile. The ratio of the E_T deposits in the HCAL and ECAL in the central trigger tower (H/E) should be smaller than 0.2. This cut is not made if the energy in HCAL or ECAL is small (< 3 GeV) or the ECAL energy is very large (> 60 GeV). The second veto utilizes the more finely grained ECAL crystal energy profile by examining the lateral extension of the shower. In the barrel a trigger tower

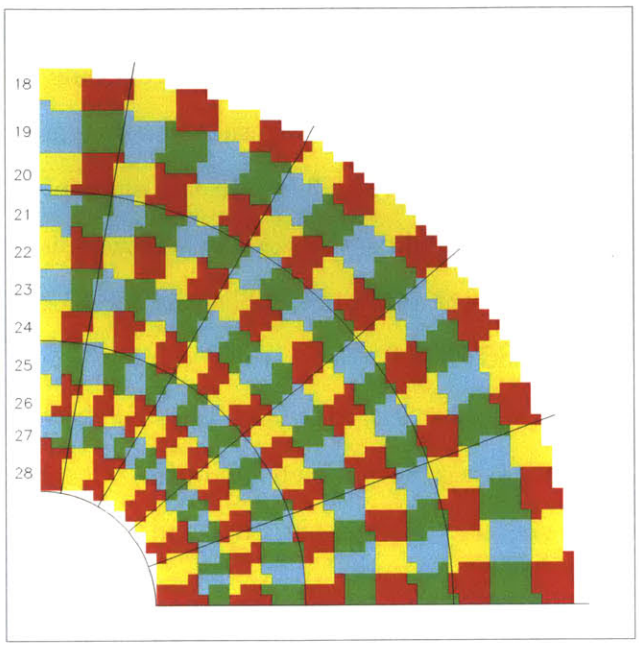

Figure 3-9: Calorimeter trigger tower lay-out in ECAL endcap.

is virtually divided in sub-regions called strips which are sets of crystals aligned along the magnetic field bending direction. The strips are 1 crystal wide in η and 5 in ϕ, to account for the energy spread in the magnetic field. In a trigger tower all 4 pairs of 2 strips are made. The strip pair with the highest energy is selected and the energy is compared with that of the full trigger tower. If this ratio R is smaller than 0.9, then the trigger is vetoed. For the low p_T domain a more stringent cut is used (0.95). In the endcap, the trigger towers cannot be split up into strips because of its more irregular shape. Here the highest energy crystal in the trigger is selected and its 4 neighbors are investigated. The sum of the energy of the highest energy crystal and its highest energy neighbor is compared with the total energy released in the tower. If the ratio of the two energies is less than 0.80, then the trigger is vetoed.

The kinematic turn-on the Level-1 triggers used is shown in fig. 3-10. The turn-on is rather slow, especially in the endcap. For the L1 e/gamma trigger with a threshold of 8 GeV (L1_EG8), we are only 95% efficient for electrons of $\geq$14 GeV in the barrel and $\geq$17 GeV in the endcap. This degradation of the energy resolution in the Level-1

trigger is due to the limited number of bits for integer representation of the various parameters used in the computations and the approximations in the algorithms. The trigger efficiency improves when adding more quality cuts to the electrons, so that the efficiency will be higher for the electrons used in this analysis (section 6.2.1).

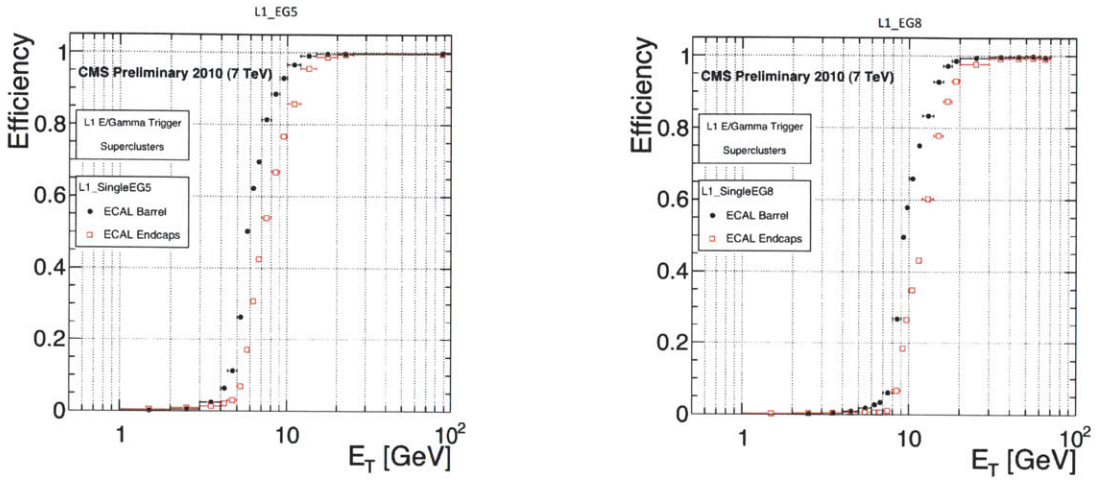

Figure 3-10: L1 trigger efficiency versus electron E_T for L1_EG5(left) and L1_EG8 trigger(right)[68].

3.3.2 High Level Trigger

The High Level Trigger is a software system implemented in a filter farm of about one thousand commercial processors. The HLT has access to the complete read-out data and can therefore perform complex calculations similar to those made in the the analysis off-line software if required for specially interesting events.

The High Level trigger for electrons has three layers of logic:

- At Stage 2 only calorimeter information is used;
- Stage 2.5 combines the calorimeter information with pixel hits;
- Full detector information and reconstructed charged tracks are available in Stage-3.

The first step of the HLT —at Stage 2— will be to apply the energy thresholds on the High Level Trigger. This is done by creating superclusters from the ECAL energy deposits using the same algorithms as used offline. Superclusters are groups of clusters along a road in the ϕ-direction built in order to collect bremsstrahlung radiated from electrons. It is also asked that the superclusters are matched with the Level-1 trigger towers in a $\eta - \phi$ window. In this analysis, the transverse energy threshold on the HLT varies from 15 GeV for the photon and lowest-energy electron triggers up to 22 GeV for the highest-energy electron triggers used in the analysis. Electrons also have an implicit $|\eta| < 2.5$ cut because the forward calorimeter is not used for the electron and photon triggers. A turn-on curve for this electron transverse energy cut is shown in fig. 3-11 for a 17 GeV electron trigger. The turn-on is very fast and the trigger is already fully efficient for electrons of 18 GeV. This shows that the energy resolution is much better than for the Level-1 trigger.

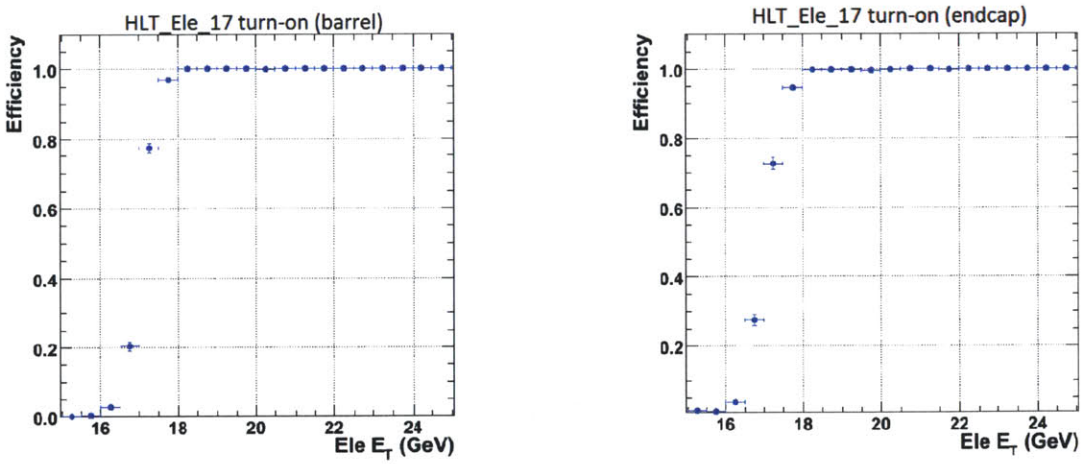

Figure 3-11: HLT trigger efficiency versus electron E_T for a HLT_Ele17 trigger in the barrel(left) and endcap(right)[68].

At level 2.5 the matching of superclusters with pixel hits is performed. A helix is propagated backward from the energy-weighted position of the supercluster to the beamspot position, using the supercluster energy measurement to determine the bending of the trajectory in the magnetic field. The estimated hit position on the pixel layers is calculated by looking at the intersection of this trajectory with the pixel layers. Then the search for a compatible hit on the first pixel layer takes place,

using the start-up matching window. There are no cuts on z, but the ϕ of the hit in the CMS axis system should be between -0.030 rad and +0.025 rad from the ϕ of the estimated hit for negatively-charged electrons. The window is changed to -0.025 rad to +0.030 rad for the positron hypothesis. If such a compatible hit is found, then it is used to make a better estimate of the longitudinal vertex position. This is done by refining the trajectory using this first hit. The vertex is still assumed to be at the beamspot. In case no hit was found on the first layer, then the search will be done on the next one. The search window for finding the second hit is in this case much smaller. The *phi* of the hit with respect to the beamspot position has to be within $\pm$0.005 rad of the one from the estimated hit position. In the barrel the second requirement is that the actual z position is within $\pm$0.05 cm from the estimated one, in the endcap ΔR <0.05 cm. The search can be repeated on the outermost pixel layer if no compatible hit is found on the middle one or if the first hit is at the second pixel layer. In the endcap, the search is extended to the first layer of the strip tracker. The search is made twice, once for each charge with appropriate modifications of the search window.

At Stage 3 full event information is available, including tracking information. The first step is the electron track-finding, seeded by the Stage-2.5 pixel pairs. The standard Kalman-filter is used for the patter recognition and track finding, since the Gaussian-Sum-Filter approach described in section 3.1.2 is too time-consuming to be run at HLT. In order not to lose too many electron tracks because of bremsstrahlung effects, the track-finding is done with very loose cut parameters. With increasing luminosity, the triggers got more complicated and started implementing more and more of the offline electron identification and isolation cuts to reduce the trigger rate of background while rejecting as little signal as possible. The isolation and identification variables are calculated at Stage 3 and the cuts are applied. The anomalous spikes in the ECAL (section 2.2.3) also affect us at the trigger level, that is why spike cleaning is applied and this is also done at Stage 3. The spike cleaning is done investigating the adjacent crystals and checking whether only the central crystal has a large energy deposit or the energy is shared between neighboring crystals.

Table 3.4 lists the different triggers in the HLT for different run ranges. The cuts applied for the different trigger names are given in table 3.5. Except for the HLT_Photon10_L1R trigger, all triggers have ECAL spike cleaning applied to them. The efficiency of these different triggers will be determined in section 6.2.1.

Run Range	Lumi (pb^{-1})	HLT Trigger	L1 Trigger
132440-137028	0.005	HLT_Photon10_L1R	L1_SingleEG5
138564-140401	0.26	HLT_Photon15_Cleaned_l1R	L1_SingleEG5
141956-144114	2.9	HLT_Ele15_SW_CaloEleId_L1R	L1_SingleEG5
144115-147145	5.1	HLT_Ele17_SW_CaloEleId_L1R	L1_SingleEG8
147146-148058	9.5	HLT_Ele17_SW_TightEleId_L1R	L1_SingleEG8
148103-149065	10.3	HLT_Ele22_SW_TighterCaloIdIsol_L1R_v1	L1_SingleEG8
149180-149442	8.1	HLT_Ele22_SW_TighterCaloIdIsol_L1R_v2	L1_SingleEG8

Table 3.4: Summary of triggers used. The name of the HLT triggers mentions first the object which is reconstructed with its transverse energy threshold. For the electron triggers the SW indicates the start-up window, the pixel-supercluster matching window used in the HLT reconstruction. Then the cuts applied are mentioned, the explanation is in table 3.5. The Level-1 trigger name has specifies whether one or more e/gamma triggers need to be found to pass this Level-1 selection and what the transverse energy threshold for the e/gamma triggers is.

Trigger-extension	H/E	$\Delta\eta_{in}$	$\Delta\phi_{in}$	$\sigma_{i\eta i\eta}$	ecaliso/pt	hcaliso/pt	trkiso/pt
CaloEleId	0.15	NA	NA	0.014/0.035	NA	NA	NA
TightEleId	0.15	0.01	0.08	0.012/0.032	NA	NA	NA
TighterCaloIdIsol	0.05	NA	NA	0.011/0.031	0.125/0.075	NA/0.05	0.15/0.1

Table 3.5: Cuts (barrel/endcap) applied to different triggers.

3.4 Particle Flow Missing Transverse Energy

In a pp-collider, the longitudinal energy is not well-constrained by the beam energy since the fractions of energies carried by the different quarks and gluons are not known but determined using particle density functions. On the other hand, the collisions happen head-on and thus, the net transverse energy of the event has to be zero. The missing transverse energy of an event is the negative vector sum of all the transverse energy deposits in the detector. This missing transverse energy is a measure of how

many particles leave the detector undetected. This can be due to the limited coverage of the detector, or new physics like a SUSY particle. In the W decay, it is the neutrino which leaves the detector undetected, and the presence of this neutrino is a very important discriminant between W events and its main backgrounds. The missing transverse energy used in this analysis is calculated with the particle-flow algorithm.

3.4.1 Particle-flow algorithm

Particle-flow aims at reconstructing and identifying all stable particles in the event: electrons, muons, charged and neutral hadrons, and photons[58]. In order to achieve the optimum resolution in terms of energy, direction and particle type, it combines the information from all the CMS sub-detectors. The transverse energy vectors of all the identified particles are then added up, to determine the missing transverse energy($\not\!\!E_T$).

The CMS silicon tracker plays an important role for particle-flow. It measures the charged hadrons with superior angular and energy resolution than what would be done with the calorimeter alone and it also seeds the particle-flow electron reconstruction. Roughly 2/3 of the jet energy is in the form of charged particles, so it is important that this component is measured with optimum precision. The tracking algorithm is designed to exploit the high granularity and precise position measurements of the tracking, leading to both high efficiency and low fake-track rates. For this an iterative tracking procedure is used, starting with very strict seeding and reconstruction criteria and then loosening them up iteratively. This can be done because the hits which are unambiguously connected to a track in a previous step are removed and thus less possibilities are left for random combinatorics to play a role. The tracking algorithm also allows for displaced vertices to reconstruct conversions and long-lived particles.

The electron reconstruction in particle-flow comes from tracker driven seeds. For this, it is important that tracks with very few hits (minimum 3) can be reconstructed. Normally such tracks will be discarded because of the low quality but they can be used to seed the electron reconstruction. If the electron does not brem much, the normal tracking will still describe the track well and the track can be reconstructed

up to the ECAL surface and then the matching with a cluster can be done. The comparison between the track momentum and the cluster energy determines whether the full electron reconstruction will be run. In the case of a high-bremsstrahlung electrons, two things can happen: the track has very few hits or the χ^2 of the track is large. If this is the case, then a partial GSF refit is done where a mixture of maximum 5 Gaussian distributions is used to model the energy loss. After this, a multivariate estimate of the electron compatibility is made, based on the quality parameters of the GSF track and the matching of the track with the ECAL supercluster. This will ultimately determine whether the full GSF-tracking and electron reconstruction is run.

Particle-flow also has its own clustering in the calorimeter in order to allow detection of low-energy particles and to separate nearly overlapping energy deposits. The clustering algorithm is run in every part of the calorimeter separately. The clustering starts with cluster seeds, which are local calorimeter cells above a certain threshold (0.23 GeV in barrel, 0.80 GeV in endcap). Topological clusters are formed by adding cells to the cluster which have at least one side in common with one of the cells already making up the cluster and which have an energy above a different threshold (80 GeV in barrel, 300 MeV in endcap, to avoid adding noise). A topological cluster which has multiple cluster seeds (local maxima above threshold) can give rise to as many particle-flow clusters as the number of cluster seeds it contains. The energy of each cell in the topological cluster will be shared among all particle-flow clusters according to the cell-cluster distance, with an iterative determination of the cluster energies and positions. In the first iteration, the particle-flow cluster starts from the energy and position of the seed cell. Each cell in the topological cluster will share its energy between particle-flow clusters proportional to $exp(-d_i^2/R^2)$ with d_i the distance of the cell to particle-flow cluster i (originally the position of the seed cell). The position of the particle-flow cluster is then recalculated as the center of gravity of nine central cells and then the procedure is repeated until it converges[58].

A particle may give rise to multiple particle-flow elements: a track, deposits in the calorimeters and/or a muon track. To make sure all the information from the

different detectors is taken into account, the different elements need to be linked correctly. For example, for an electron, its track and ECAL cluster need to be matched correctly to give one particle, otherwise the electron is identified as the combination of a charged pion and a photon by the particle-flow algorithm and the energy of the particle is counted more than once. The first step for doing such linking is extrapolating the track to three positions: the expected depth in the ECAL for an electromagnetic shower, one interaction length in the HCAL, and the first layers of the pre-shower in the endcap region. This triple extrapolation allows for different particle hypotheses. Linking is done if the extrapolated position of the track is within the cluster boundaries. For bremsstrahlung, tangents to the track are taken and if they can be extrapolated to a cluster in the ECAL, that deposit is counted as a part of bremsstrahlung. For muon tracks in the tracker and muon chambers, a global fit is done. For electrons, as explained before, a multivariate variable based on track- and supercluster variables is used to decide whether the full electron reconstruction is performed. Otherwise the matched cluster and track are regarded as a photon/pion pair and the energy of the supercluster and momentum of the track will be counted separately[58]. Particle-flow was not used for the standard electron reconstruction because the reconstruction efficiency is harder to measure in data than for the ECAL-driven electrons. Finally some remaining bad tracks (very few hits, high χ^2) are not used in the particle-flow reconstruction, because the calorimeter deposits matched to these bad tracks provide a more accurate measurement. The transverse energy vectors of all the particles reconstructed by the particle-flow algorithm will be added up to calculate the missing transverse missing energy in the event.

3.5 Conclusion

In this chapter the tools for reconstructing and selecting electrons were described. This knowledge was used to describe the triggers used to record the W data. The particle-flow algorithm was used to calculate missing transverse energy. Using this knowledge of electrons and missing transverse energy, we are ready to select and

identify Ws decaying to electrons in the CMS detector.

Chapter 4

Monte Carlo Event Generators and Tools.

Simulation is an important part of a current-day physics analysis. In this study, it will be used to calculate the acceptance, to give a first rough idea of the efficiency, as a baseline for the modeling of the missing transverse energy, among others. Different Monte-Carlo (MC) event generators and integrators are used for all these steps and to quantify the uncertainties associated to them. In this chapter a small summary of the MC tools used will be given.

There are two kinds of MC tools that will be used:

1. Monte Carlo (MC) event generators that produce unweighted events distributed according to the theoretical prediction.

2. cross section integrators that effectively sum over the kinematic distributions of the W decay products.

The two techniques differ in flexibility and accuracy. Event generators are general purpose tools designed to describe a variety of physics processes. They produce exclusive events and can be interfaced with a detailed detector simulation(using Geant4 for CMS) to realistically model the W$\rightarrow$eν process as it is observed in the detector. The reconstruction code can also be run on these generated events. Pythia[69] provides predictions that are formally only leading order (LO), with higher order effects are

approximated via parton showers (PS) and matrix element (ME) corrections. Newer event generators like POWHEG[70] include full next-to-leading-order (NLO) matrix elements, however these programs continue to rely on parton showering for an account of higher order soft QCD effects. Event generators can also be leading-order in perturbative QCD, but next-to-loading order in the electroweak corrections to allow for an improved QED modeling.

The treatment of QCD radiation in dedicated integrators is typically more complete. Such programs (e.g.: ResBos[28], FEWZ[27]) also extend the cross section calculation to higher perturbative order and thus more accurately predict the distribution of W decay products. Although technically more precise, integrators produce inclusive events that, in their lack of secondary detail, do not truly represent what is seen in the CMS detector.

Pythia and POWHEG will be used in CMS as the standard event generators and the other generators will be used in chapter 5 to study the theoretical uncertainties associated with the missing higher-order QCD and QED effects. FEWZ is also used to calculate the best theoretical prediction of the W$\rightarrow e\nu$ cross section.

4.1 Pythia

Pythia [69] is used as a leading order (LO) event generator for QCD processes. Although the Pythia matrix elements are strictly leading order, it approximates higher order corrections via parton showers and matrix element corrections. These corrections lead to distributions that are more 'NLO-like', which should translate to a more realistic acceptance estimate.

Parton showers (PS)

Initial and final state QCD radiation arises from the splitting of partons with branchings $q \rightarrow qg$, $g \rightarrow gg$ and $g \rightarrow q\bar{q}$. The probability of a branching occurring is related to splitting functions, which are evolved in Q^2 by the DGLAP equation [71, 72, 73]. Parton showering algorithms use these functions in the implementation of leading-log

(LL) resummation. The resummation procedure becomes process-dependent beyond LL, so the PS technique implements resummation only to this order so as to maintain general applicability.

Some higher order corrections, however, are not process dependent and can be included in the PS algorithms used by LO MCs. Angular ordering is one example. In angular ordered showers, the opening angles of consecutive branchings are required to decrease, as occurs in soft gluon radiation. A soft gluon emitted from the daughters of a $q \to qg$ branching, for example, can only resolve the individual quark and gluon if the opening angle is smaller than that of the original $q \to qg$ branching. Otherwise it can only see the color charge of the parent quark (or, equivalently, the coherent sum of the charges of the daughters). Showers are p_T-ordered in newer versions of Pythia (≥ 6.34), which implicitly imposes angular ordering.

Initial-state QCD radiation (ISR) influences the entire boson transverse momentum spectrum. Figure 4-1 shows that ISR effects reach to rather large p_T for W^+. When ISR is disabled, W p_T can only result from the intrinsic momentum of the parton in the proton. The rightmost plot in figure 4-1 shows that, as expected, the W boson is produced more centrally when including ISR effects.

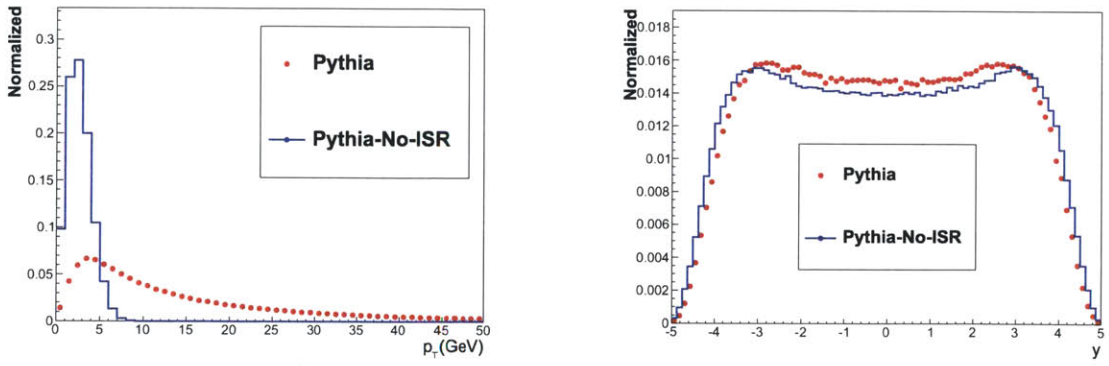

Figure 4-1: W^+ transverse momentum (left) and rapidity (right) distributions for Pythia with ISR switched on and off.

Matrix Element Corrections (MEC)

Parton showers provide a reasonable description of soft QCD radiation, but this approach breaks down at high p_T where hard radiation dominates. Behavior in this regime is better described with higher order matrix elements. Pythia includes a matrix element that describes one real emission. The generators transition between the parton shower and matrix-element (ME) descriptions at some intermediate scale above which parton showers are reweighted to reproduce the higher order matrix element effects. The matrix element corrections are these reweightings, they are applied in Pythia by modifying the branching probabilities in the parton shower. For this, the first splitting of the parton shower is weighted with a different factor depending on whether it is $q\bar{q} \rightarrow Wg$ or a $qg \rightarrow Wq$. The weight factor corresponds to the ratio of the ME and PS cross-sections for such splittings and is p_T and rapidity dependent. Figure 4-2 shows the impact of this correction on the W^+ distributions from Pythia. As expected, the corrections lead to a harder high p_T tail and a slightly more central rapidity distribution. MECs influence the p_T spectrum, but to a lesser extent than ISR alone.

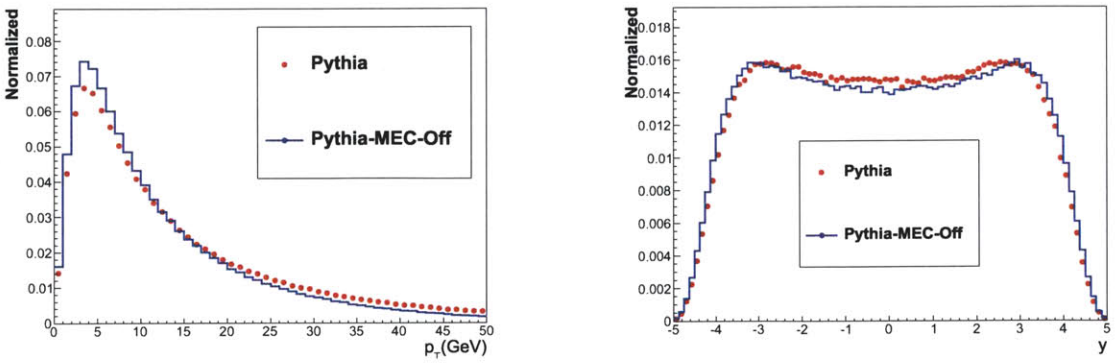

Figure 4-2: W^+ transverse momentum (left) and rapidity (right) distributions for Pythia with MEC switched on and off.

4.2 NLO event generators

The Next-to-Leading-Order (NLO) event generators combine a NLO matrix element with the normal leading-log parton shower from the LO event generators. This is non-trivial since the parton shower contains terms of the same order as the matrix element and double-counting them needs to be avoided. The standard NLO event generator used by CMS for simulating electroweak physics is POWHEG [70] with Pythia used for the parton shower. POWHEG avoids double-counting by generating the hardest emission first. Subsequent emissions in the parton shower are guaranteed to be softer than the hardest emission by requiring p_T-ordered showers or an explicit p_T veto. Pythia implements p_T-ordered showering and the parton shower starts at a scale below that of the NLO matrix-element. The interface of a NLO matrix element with a leading-log parton shower can also be done by removing the shower-based approximation of these terms, this is the case in other next-to-leading order event generators like MC@NLO [74]. This can lead to negative weights of certain events, such as when the parton-shower approximation is larger than the actual NLO contribution. Negative weights events are unphysical, but the calculations will lead to physical observables if one keeps correctly track of both negative and positive weight events. In POWHEG no subtraction is done, therefore all event weights are guaranteed to be positive. POWHEG was chosen as the default generator for CMS because it uses Pythia for its showering and CMS focused its underlying-event tuning for the parton showers on Pythia. The positive weights also allow for simpler book-keeping in the analysis.

4.3 Horace

Horace[29] is a Monte Carlo event generator for W and Z production and leptonic decay at hadron colliders. It includes the exact 1-loop electroweak radiative corrections matched with a QED Parton Shower to take into account also higher-order QED leading effects. It is NLO accurate in electroweak effects and LO in QCD effects. Similar

to POWHEG, Horace can be interfaced with Pythia's showering, allowing the final output to be used in CMS detector simulation and reconstruction. Radiated photons can be recovered in the electron reconstruction and this will influence the effects of the electroweak corrections.

4.4 FEWZ

FEWZ [27] is an integrator that computes the differential production cross sections of gauge bosons in perturbative QCD up to next-to-next-to-leading order (NNLO). At leading order, quark-antiquark annihilation is the only process that contributes to the cross section. Quark-gluon scattering is included at NLO, and all $O(\alpha_S^2)$ diagrams, including both gluon-gluon fusion and quark-quark scattering, contribute to the cross section at NNLO. FEWZ implements a fully differential description of electroweak gauge boson production and includes spin correlations and finite width effects. It does not include parton showering or resummed soft QCD effects.

4.5 Resbos

In the region of low W p_T the W p_T distribution cannot be well-described by the next-to-leading-order (NLO) perturbative calculation. The calculations are dominated by a large amount of collinear radiation. As a consequence the lepton transverse momentum cannot be well-modeled in the region of low W transverse momentum, which is where the bulk of the distribution lies. These event generators use the parton shower technique to describe soft/collinear radiation, which is only formally accurate to leading logarithmic order (LL).

Resbos[28] is a integrator for hadron collisions that resums the effects of the initial state multiple soft-gluon emission to predict the distributions of the leptons from the decay of Ws produced. The resummed cross section is a combination of the fixed-order NLO cross section with an all-order sum of large logarithmic terms originating from multiple soft gluon emission, the calculation is process-dependent[75]. This

yields a Sudakov form factor and cures divergences as W $p_T \rightarrow 0$. Second, ResBos uses so-called grid files that contain resummed, fixed order cross section results for a number of kinematic points. ResBos interpolates between these points and calculates various kinematic distributions for the W boson and its decay products. Using such a grid of NNLO scale factors, the high-p_T part of the Resbos spectrum is basically NNLO. This can be checked by comparing the Resbos results with FEWZ, which is a full NNLO calculation. Figure 4-3 shows that the high-p_T tail agrees between Resbos and FEWZ but there are descrepancies at low p_T due to the lack of resummation in FEWZ. Resbos can also be interfaced with PHOTOS[76] to simulate electromagnetic final-state-radiation effects. Past studies have demonstrated remarkable agreement between ResBos predictions for the W p_T distribution in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV [77].

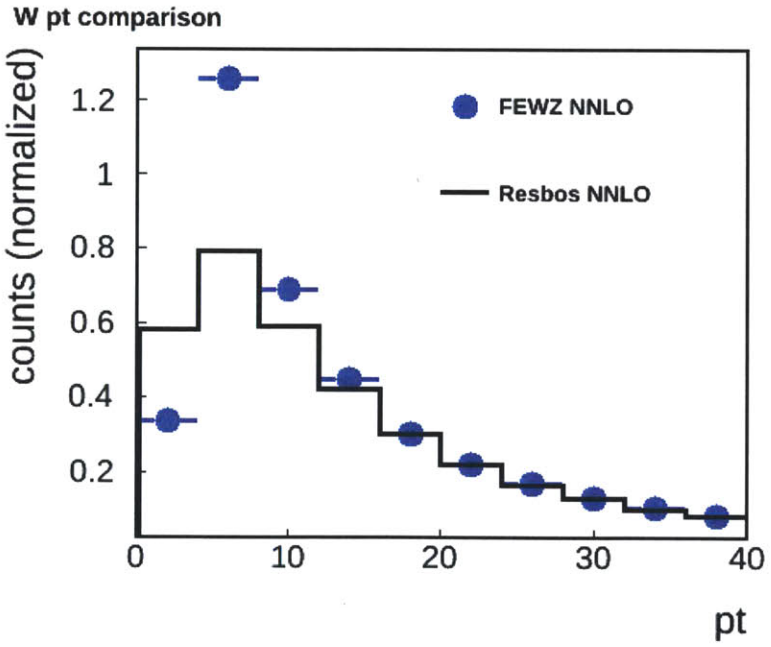

Figure 4-3: W p_T spectrum at NNLO, calculated with FEWZ and Resbos. FEWZ and ResBos essentially agree on asymptotic NNLO behavior. The FEWZ spectrum at low p_T diverges because of lack of resummation. (dip at very low p_T due to Q^2 cut-off in the calculation)

Chapter 5

W$\rightarrow$ eν acceptance

The acceptance is the fraction of all the electrons produced from W decay that have transverse momentum above a 25 GeV p_T-threshold and end up in the well-instrumented region of CMS. This number is fully taken from simulation, since it cannot be estimated from data. There is no single event generator that gives the best calculation for both electroweak and QCD effects. Therefore the acceptance is estimated using our baseline Monte Carlo simulation (POWHEG) and then the influence of different effects is investigated using specific simulation tools, from which systematic uncertainties are derived. These uncertainties can rise from higher-order corrections as well as different models (for final-state-radiation or parton distribution functions (PDF), to name some examples). Table 5.1 shows the different effects that will be studied and the event generators and integrators used to study them.

Effect	simulation tool
Soft QCD	Resbos
Missing NNLO	Resbos
> NNLO	FEWZ
EWK	Horace
FSR	Horace
PDF uncertainties	POWHEG

Table 5.1: Higher-order effects on the acceptance and simulation tools used to study them.

5.1 Baseline Calculation

The baseline MC generator for our signal samples is POWHEG with CT10[23] PDF set. As described before, POWHEG is next-to-leading order in perturbative QCD and can be interfaced with Pythia. Pythia will perform the hadronization and showering, initial and final-state radiation. Underlying event and Pile-Up (multiple interactions) are also modeled by Pythia.

Two definitions of the acceptance will be used. The so-called "generator acceptance" only considers the effects of the kinematic cuts of E_T >25 GeV and $|\eta| < 2.5$, computed on the generator quantities, which are not yet affected by detector resolution. This acceptance is easy to compare with the other simulation tools and used to extract our systematic uncertainties for the QCD effects. For the final calculation of the cross section we will be using "ECAL acceptance". Here the fiducial cuts, also excluding the gap (E_T >25 GeV and $|\eta| < 1.4442$ OR $(1.566 < |\eta| < 2.5)$) are applied to the supercluster quantities. This the gap between the ECAL barrel and the ECAL endcap is cut out because it is poorly instrumented and the electron reconstruction will be worse there. The ECAL acceptance is calculated using supercluster quantities, because the efficiency of the first step in the electron reconstruction, the clustering, cannot be easily probed in data. Cross-checks can be made, but an efficiency measurement parallel to the other reconstruction and identification steps is not possible. The electroweak and PDF uncertainties are evaluated on the fully simulated and reconstructed samples, since the reconstruction can change the effects considerably, for example by trying to combine electrons with radiated photons in the superclustering procedure.

The acceptance for the $W \to e\nu$ analysis is summarized in table 5.2. The acceptance for the inclusive W cross-section is calculated by weighting the W^+ and W^- acceptances with the cross-section predictions at NNLO (6.15 ± 0.29 nb for W^+, 4.29 ± 0.23 nb for W^-[25]).

Sample	A_{GEN}	A_{ECAL}
$W^+ \to e^+\nu$	0.5279	0.5018
$W^- \to e^-\bar{\nu}$	0.5043	0.4808
$W \to e\nu$	0.5182	0.4931
W^+/W^-	1.0467	1.0436

Table 5.2: Baseline acceptances for the W analysis. The acceptance for the inclusive W cross-section is calculated by using the cross section weighted average of the W^+ and W^- acceptances.

5.2 QCD Resummation and NNLO Corrections

POWHEG uses Pythia's parton shower to model the soft, non-perturbative QCD. Therefore POWHEG is only accurate to leading logarithmic (LL) order for the soft QCD effects, while Resbos's resummation procedure gives a next-to-next-to-leading log (NNLL) description. The hard matrix elements in POWHEG are also only accurate up to NLO in perturbative QCD, while Resbos allows the use of a K-factor grid to get an effective NNLO description. So Resbos can be used to look at both of these effects.

Since Resbos is only an integrator and not an event generator, it cannot be used for simulation and reconstruction. However, the acceptances calculated with Resbos can be compared with those of the baseline POWHEG MC. This comparison shows that the parton shower model used in POWHEG offers a sufficient description of soft QCD radiation for acceptance estimates. The acceptance differences discussed in table 5.3 are small enough that no correction is needed for missing soft QCD effects. The remaining differences will be used the determine a systematic uncertainty due to these two effects: missing higher orders(> NLO) in the perturbative QCD, and missing higher orders (> LL) in the description of the soft QCD effects. The central values of the differences in acceptance, summarized in table 5.3, are taken as systematic uncertainties due to resummation and NNLO QCD effects, yielding a 0.5% systematic uncertainty on the acceptance calculation.

	$W^+ \to e^+\nu$	$W^- \to e^-\bar{\nu}$	$W \to e\nu$	W^+/W^-
ΔA/A	$-0.63 \pm 0.08\%$	$-0.31 \pm 0.09\%$	$-0.53 \pm 0.06\%$	$-0.32 \pm 0.12\%$

Table 5.3: Changes in acceptance due to resummation and NNLO QCD effects. The central values of these differences are used as systematic uncertainties.

5.3 Higher-Order QCD Corrections

There are no calculations available that determine the W cross-section and acceptance to an order higher than NNLO. To estimate an uncertainty for these missing higher order terms, the influence of the renormalization (μ_R) and factorization (μ_F) scales on the acceptance calculation is investigated. In infra-red factorization, the bare parton densities absorb all long-distance physics and acquire a factorization scale dependence, while in ultra-violet renormalization the bare strong coupling absorbs all very-short time physics and acquires a renormalization scale dependence. Fixed order calculations depend on these two scales, while the actual physics should be independent of these scales. Therefore the variations due to changing the scales provide an estimate of the higher-order effects. Resbos does not allow modification of the scales, but because Resbos and FEWZ agree at high p_T, where the NNL0 calculation is dominant, FEWZ, which does allow such changes, can be used to assess the higher order contributions.

FEWZ is an integrator that allows the users to cut on the daughter lepton kinematics. This allows us to extract acceptances, since both the cross-section with and without lepton cuts can be calculated. This is done for three values of the renormalization and factorization scales: $\mu_R = \mu_F = \mu$, where $\mu = M_W$, $2M_W$, $M_W/2$, as was done in [78]. The main problem with FEWZ is that it is computationally intensive, which leads to large uncertainties on the extracted numbers. Increasing the computing power by running the calculation on a computer farm already reduced the problem but the remaining uncertainties are still large. The final uncertainty for the scale dependence is calculated by taking half of the largest pair-wise combination

of the acceptances:

$$\delta_{scale} = \frac{1}{2} max[|Acc_{M_W} - Acc_{2M_W}|, |Acc_{2M_W} - Acc_{M_W/2}|, |Acc_{M_W/2} - Acc_{M_W}|]. \quad (5.1)$$

The missing higher-order corrections give an uncertainty less than 1%, the specific values are shown in table 5.4.

	$W^+ \to e^+\nu$	$W^- \to e^-\bar{\nu}$	$W \to e\nu$	W^+/W^-
δ_{scale}	0.77 ± 0.40%	0.50 ± 0.37%	0.34 ± 0.28%	1.14 ± 0.54%

Table 5.4: Systematic uncertainty on the acceptance coming from scale variation.

5.4 Electroweak Corrections

There are two sources of systematic uncertainties associated with electroweak corrections. Firstly, there are the missing NLO EWK effects (virtual corrections and radiation from W) which can be quantified by comparing HORACE with all its corrections switched on to HORACE with only FSR modeling on. Secondly, the systematic uncertainty on the FSR modeling needs to be determined. Since FSR is described by PYTHIA in the baseline sample, the difference in acceptances between HORACE (FSR only) and PYTHIA are taken to estimate the systematic uncertainty. The comparisons have to be done for fully reconstructed and selected events, since some of the photon radiation can be recovered by the GSF tracking and superclustering procedures. The effect of QED initial-state-radiation (ISR) on the acceptance was found to be insignificant, by comparing PYTHIA samples with QED ISR enabled and disabled. Hence, the uncertainty on QED ISR modeling is ignored. Table 5.5 gives the results of the differences in acceptances due to electroweak NLO corrections and FSR modeling. The small differences (<0.3%) are taken as systematic uncertainties on the acceptance calculation.

Effect	$W^+ \to e^+\nu$	$W^- \to e^-\bar{\nu}$	$W \to e\nu$	W^+/W^-
FSR	+0.12 ± 0.17 %	+0.05 ± 0.20 %	+0.09 ± 0.13 %	+0.07 ± 0.26 %
EWK	+0.05 ± 0.14 %	+0.29 ± 0.21 %	+0.14 ± 0.12 %	-0.24 ± 0.25 %

Table 5.5: Acceptance differences due to final-state-radiation and missing higher-order EWK effects.

5.5 Parton Distribution Function (PDF) Uncertainties

Parton distribution functions (PDFs) encapsulate the distribution of quark and gluon momenta within the proton, and are obtained from global fits to data from deep-inelastic scattering, W-production, Drell-Yan and jet data. A number of groups[79] have produced publicly available PDFs using different data sets and analysis frameworks. These PDF sets contain a central PDF and then error PDFs to assess all the associated uncertainties that go into the PDF calculation. For determining the error PDFs, first the eigenvectors in the parameter space of the PDF description are determined. Then the 1 σ up and down variation of every eigenvalue creates a new error PDF. The final number of error PDFs is twice the final number of independent parameters that determine the PDF. NNPDF[24] does not have a set of error PDFs but provides replicas: a set of PDFs, created by Monte-Carlo sampling of the probability on the space of physical observables. In this way, also this set of replicas provides the uncertainty associated with the PDF set. There are many differences in the PDF analyses between these groups: different input data, different values of α_s, different treatments of heavy quarks, different values of heavy quark masses, different ways of parameterizing PDFs, etc., but in general there is good agreement among them. The official recommendations from the PDF4LHC working group[80] are followed here to choose which PDF sets to use and how to assign the PDF uncertainties.

For the PDF uncertainties, we have to calculate the acceptances up to full selection since the reconstruction can influence the results. Therefore, the Monte Carlo output is fed into the detector simulation, the simulated events are reconstructed and the W

selection is applied. Rather than rerunning the full chain for every PDF and error PDF, the simulation is reweighted event-by-event reweighting by the ratio of the initial quark-distribution for the two PDF choices. For every PDF set the uncertainties on the acceptance are calculated by looking at the results of all the error PDFs (68% CL) and also including the uncertainty on α_s. It has been chosen to consider $\Delta\alpha_s$ = 0.0012 as the 1 σ variation of the strong coupling constant. The combination of both errors will give the "PDF+α_s" uncertainty for every set. The final systematic uncertainty is given by the envelope provided by the central values and PDF+α_ss errors from the MSTW2008[22], CT10[23] and NNPDF2.0[24] PDFs. Every one of these PDF sets has a slightly different way of calculating the uncertainties and the details will be given in the following subsections.

5.5.1 CT10

The PDF uncertainties on the acceptance for CT10 are calculated following the instructions from the CTEQ collaboration[23]:

$$\Delta Acc^{+} = \sqrt{\sum_{i=1}^{2} 6[max(Acc_i^{+} - Acc_0, Acc_i^{-} - Acc_0, 0)]^2}, \tag{5.2}$$

$$\Delta Acc^{-} = \sqrt{\sum_{i=1}^{2} 6[max(Acc_0 - Acc_i^{+}, Acc_0 - A_i^{-}, 0)]^2}, \tag{5.3}$$

with Acc_0 the acceptance calculated with the central PDF and Acc_i^{+} and Acc_i^{-} the acceptances using the positive and negative error PDF for the ith eigenvector in parameter space. This uncertainty needs to be scaled down by a factor 1.645 to obtain the 68% confidence level uncertainties, since the standard CT10 error PDFs are given for 90% CL.

The α_s uncertainty has to be calculated separately. This is done by using the complementary CT10as PDF set and calculating the acceptances for α_s=0.116, 0.118 and 0.120. This is the 90% confidence level interval around central value of 0.118 ($\Delta\alpha_s$ = 0.0012), so the uncertainty has to be rescaled again by 1.645 to get the 68%

CL uncertainty.

The final PDF+α_s uncertainty for CT10 is obtained by adding the two uncertainties in quadrature. The values can be found in table 5.7.

5.5.2 MSTW2008

MSTW[22] provides PDF error sets for 5 different values of α_s: α_s^0,$\alpha_s^0 \pm 0.5\sigma$,$\alpha_s^0 \pm \sigma$ where σ is the 68% confidence level uncertainty on α_s^0, the standard α_s value in MSTW ($\alpha_s(M_Z)$=0.12018). The PDF+α_s uncertainty can now be estimated by calculating the envelope of the uncertainties for the error sets of the different α_s values ($\Delta\alpha_s$ = 0.0012).

First, for every value of α_s the PDF uncertainties from the error set are calculated using equation 5.2. The central values and uncertainties calculated like this for the 3 different α_s-values, can now be combined to calculate the final uncertainty due the PDF error sets as well as due to the α_s uncertainty as follows:

$$\Delta Acc^{+}_{PDF+\alpha_s} = \max_{\alpha_s} Acc_{\alpha_s} + \Delta Acc^{+}_{\alpha_s} - Acc_0, \tag{5.4}$$

$$\Delta Acc^{-}_{PDF+\alpha_s} = Acc_0 - \min_{\alpha_s} Acc_{\alpha_s} - \Delta Acc^{+}_{\alpha_s}, \tag{5.5}$$

with Acc_0 the central value PDF with α_s^0, Acc_{α_s} the central value PDF with α_s and $\Delta Acc^{\pm}_{\alpha_s}$ the positive and negative PDF uncertainty calculated with the error sets using formula 5.2 for this value of α_s. The corresponding central values and uncertainties are tabulated in table 5.7.

5.5.3 NNPDF2.0

The PDF+α_s uncertainty for NNPDF[24] is calculated very differently. The uncertainty is calculated by considering a set of replicas from the NNPDF sets with different α_s values corresponding to a sample of gaussian distributed values of α_s around the mean of 0.119. Table 5.6 shows the number of replicas needed for different values of α_s. NNPDF does not provide error sets, but a Monte Carlo sample of replicas of

the function, so that any statistical property of the underlying distribution can be derived from the given sample.

α_s	0.116	0.117	0.118	0.119	0.120	0.121	0.122
N_{rep}	5	27	72	100	72	27	5

Table 5.6: Number of replicas for different α_s values to calculate the PDF+α_s uncertainty.

The final uncertainty for NNPDF is then calculated as:

$$\Delta Acc^{+} = \sqrt{\frac{1}{N^{+}-1}\sum_{i=1}^{N^{+}}(A_i - A_0)^2}, \tag{5.6}$$

$$\Delta Acc^{-} = \sqrt{\frac{1}{N^{-}-1}\sum_{j=1}^{N^{-}}(A_j - A_0)^2}, \tag{5.7}$$

with A_0 the central value PDF with α_s=0.119, i runs over the N^+ replicas with $A_i > A_0$ and j over the N^- replicas with $A_j < A_0$. Table 5.7 shows the acceptance values and uncertainties associated with them using a NNPDF.

5.5.4 Results

The final PDF uncertainty is calculated by taking the envelope of the acceptance values and uncertainties by the different PDF sets:

$$Acc_{max} = \max_i(Acc_i^0 + \Delta Acc_i^{+}), \tag{5.8}$$

$$Acc_{min} = \min_i(Acc_i^0 - \Delta Acc_i^{-}), \tag{5.9}$$

$$\text{PDF+}\alpha_s \text{ uncertainty} = \frac{1}{2}(Acc_{max} - Acc_{min}), \tag{5.10}$$

where i runs over the PDF sets (CT10, MSTW2008, NNPDF2.0). Acc_i^0 is the central value calculated with this PDF set and ΔAcc_i^{+} the PDF+α_s uncertainty. This will give the 68% CL uncertainties for the PDF uncertainties. A summary of all the values related to the PDF uncertainties is given in table 5.7. The PDF systematic uncertainty is the largest for the W^+/W^- ratio (1.65%) and smaller for the inclusive

and charged yields (0.60-1.34%).

PDF set	$W^+ \to e^+\nu$	$W^- \to e^-\bar{\nu}$	$W \to e\nu$	W^+/W^-
CT10	$0.3796^{+0.0018}_{-0.0022}$	$0.3675^{+0.0032}_{-0.0032}$	$0.3747^{+0.0022}_{-0.0017}$	$1.0329^{+0.0106}_{-0.0097}$
MSTW2008	$0.3793^{+0.0017}_{-0.0013}$	$0.3711^{+0.0021}_{-0.0018}$	$0.3760^{+0.0015}_{-0.0015}$	$1.0223^{+0.0046}_{-0.0055}$
NNPDF2.0	$0.3819^{+0.0015}_{-0.0016}$	$0.3659^{+0.0027}_{-0.0026}$	$0.3755^{+0.0016}_{-0.0018}$	$1.0437^{+0.0072}_{-0.0065}$
Systematic uncert.	**0.79%**	**1.34%**	**0.60%**	**1.65%**

Table 5.7: Acceptance values and systematic uncertainties for the different PDF sets.

5.6 Summary of Theoretical Uncertainties on Acceptance

Table 5.8 combines the different effects that contribute to the systematic uncertainty on the acceptance. In a few cases, the uncertainty on the influence of an effect was larger than the bias observed due to this effect. In those cases, we took the uncertainty on the effect and not the bias as the systematic uncertainty. The systematic uncertainties for the acceptance are dominated by the uncertainty due to perturbative QCD effects of order higher than NNLO and PDF uncertainties. The largest systematic uncertainty is 2% on the acceptance for the W^+/W^- ratio, the acceptances for the cross-section measurements all have uncertainties less than 1.5%.

Effect	$W^+ \to e^+\nu$	$W^- \to e^-\bar{\nu}$	$W \to e\nu$	W^+/W^-
ISR+NNLO	0.63%	0.31%	0.53%	0.32%
>NNLO	0.77%	0.50%	0.34%	1.14%
EWK	0.17%	0.20%	0.13%	0.26%
FSR	0.14%	0.29 %	0.1%	0.25%
PDF	0.79%	1.34%	0.60%	1.65%
Total	**1.29%**	**1.51%**	**0.89%**	**2.06%**

Table 5.8: Systematic uncertainties on the acceptance calculation.

Chapter 6

Efficiencies

The electron efficiencies are estimated in simulation and then corrected by the differences between data and simulation. These differences are determined using the tag-and-probe technique, a technique exploiting the pure Z signal to get an unbiased sample of high-p_T leptons similar to those created in W decay. The total efficiency can be factorized as following, with the total efficiency defined with respect to a reconstructed supercluster passing the acceptance cuts:

$$\epsilon_{total} = \epsilon_{reco} \times \epsilon_{WP80} \times \epsilon_{trigger}, \tag{6.1}$$

with ϵ_{reco} the efficiency to create an ECAL-driven Gsf-Electron starting from a supercluster, ϵ_{WP80} the efficiency for an ECAL-driven Gsf-Electron to pass the WP80 identification and isolation criteria (table 3.2) and $\epsilon_{trigger}$ the efficiency for a fully identified and isolated electron to pass the trigger (HLT and Level-1) requirements.

In this chapter, we will first explain the tag-and-probe method and the associated definitions before showing the results of the efficiency measurements in data. Finally, we will determine the systematic uncertainties of the tag-and-probe method.

6.1 Tag-and-Probe Method

The tag-and-probe method exploits the very pure Z sample. The Z $\rightarrow e^+e^-$ sample will give an unbiased high-purity electron sample to measure the different efficiencies. This is done by putting very stringent identification and isolation criteria on one of the electrons (the 'tag'). This tag electron together with the Z mass constraint allows us to find the other electron coming from the Z decay. This electron is used as an unbiased, high p_T 'probe' to measure the efficiency. In this measurement, we will ask that the tag-electron passes the full W selection: the single-photon or single-electron triggers as defined in table 3.4 and the WP80 identification and isolation cuts (table 3.2). The sample is cleaned up further by requiring an invariant mass window (60 GeV$<M_{e^+e^-}<$120 GeV) for the combination of the tag and probe electrons. The amount of background remaining can be estimated by performing a fit to the di-electron invariant mass spectrum. The probe definition depends on the measured efficiency: the selection criteria applied are always the passing criteria of the previous efficiency step. The identification and isolation efficiency will use reconstructed electrons as probes and the trigger efficiency identified and isolated electrons. This is done to take the correlations between the different efficiencies correctly into account.

The exact probe requirements and pass/fail criteria are:

- reconstruction efficiency (ϵ_{reco}):
  - **probe:** supercluster with E_T >25 GeV in the fiducial region of the detector ($|\eta| < 1.4442$ OR $(1.566 < |\eta| < 2.5)$
  - **pass/fail:** ECAL-driven Gsf-Electron with E_T >25 GeV in the fiducial region of the detector ($|\eta| < 1.4442$ OR $(1.566 < |\eta| < 2.5)$)
- Identification and isolation efficiency (ϵ_{ID_ISO}):
  - **probe:** ECAL-driven Gsf-Electron with E_T >25GeV in the fiducial region of the detector ($|\eta| < 1.4442$ OR $(1.566 < |\eta| < 2.5)$)
  - **pass/fail:** ECAL-driven Gsf-Electron with E_T >25 GeV in the fiducial

region of the detector passing WP80 (table 3.2) isolation and identication cuts.

- Trigger efficiency ($\epsilon_{trigger}$):
  - **probe:** ECAL-driven Gsf-Electron with E_T >25 GeV in the fiducial region of the detector passing WP80 (table 3.2) isolation and identication cuts.
  - **pass/fail:** ECAL-driven Gsf-Electron with E_T >25 GeV in the fiducial region of the detector passing WP80 (table 3.2) isolation and identication cuts and matched in a cone of 0.2 to the unprescaled single-electron/photon trigger used in this period of data-taking (table 3.4).

6.1.1 Basic Method

The tag-and-probe candidates are split up into two independent samples: one with the probes passing the selection and one with the failing probes. Using two independent samples simplifies the uncertainty calculation, since the statistical uncertainties of the two samples are uncorrelated. The efficiency can then be extracted as:

$$\epsilon = \frac{N_{pass}}{N_{pass} + N_{fail}}, \tag{6.2}$$

where N_{pass} and N_{fail} are the number of passing probes and the number of failing probes, respectively. If the samples contain a non-negligible amount of background, a fit to the mass spectrum is performed to differentiate between the Z electrons and the background.

The kinematics of production and decay of the Z and W are different and this will influence the efficiency measurement. To correct for the differences in the kinematic spectrum between the Z tag-and-probe sample and the W candidates, the efficiencies will be measured in kinematic bins (η,p_T). The measured data efficiencies are also not used directly, but they are used to determine scale factors between data and simulation. This allows to correct for remaining kinematic dependencies after applying

the limited η-p_T binning. The tag-and-probe efficiency is also estimated in simulation and then the ratio of the tag-and-probe efficiencies in data and MC is applied to the efficiencies from the W simulation to extract the final data W efficiency:

$$\epsilon_{W,data}(\eta, p_T) = \epsilon_{W,MC}(\eta, p_T)\frac{\epsilon_{T\&P,data}(\eta, p_T)}{\epsilon_{T\&P,MC}(\eta, p_T)}. \tag{6.3}$$

$\epsilon_{W,data}$ is the estimated efficiency for the W electrons in data, $\epsilon_{W,MC}$ the calculated efficiency for the W electrons in MC, $\epsilon_{T\&P,data}$ the measured tag-and-probe efficiency in data and $\epsilon_{T\&P,MC}$ the calculated tag-and-probe efficiency in MC. The tag-and-probe efficiency in simulation can just be calculated by counting the events in the different categories. No fit has to be performed since we ensure that the probes used are real electrons coming from the Z.

6.1.2 Efficiency Fits

When significant amounts of backgrounds are present, it is important to be able to distinguish signal and background. This is done through a fit to the invariant mass of the di-electron pair. For the signal shape all kinds of effects can cause biases, for example insufficient knowledge of energy scale, resolution or final-state radiation. In order to assess a systematic uncertainty from the fitting model, two different signal models will be used:

- Breit-Wigner distribution[1], convoluted with Crystal-Ball resolution function[2], symbolically CB⊗BW. The Breit-Wigner shape is fixed to the PDG values for the Z width and mass. The Crystal-Ball function has free parameters intended to account for energy scale shifts, mass resolution, and low mass tails from FSR and bremsstrahlung. The model has 4 floating parameters.

- Shape from MC simulation, convoluted with gaussian resolution function, symbolically MC⊗Gauss. The shape from simulation used here has the tag-and-

[1] Breit-Wigner: $f(x; x_0, \gamma) = \frac{1}{\pi}\left[\frac{\gamma}{(x-x_0)^2+\gamma^2}\right]$

[2] Crystal-Ball: $f(x; \alpha, n, \bar{x}, \sigma) = N \cdot \exp(-\frac{(x-\bar{x})^2}{2\sigma^2})$, for $\frac{x-\bar{x}}{\sigma} > -\alpha$,
$f(x; \alpha, n, \bar{x}, \sigma) = N \cdot A \cdot (B - \frac{x-\bar{x}}{\sigma})^{-n}$, for $\frac{x-\bar{x}}{\sigma} \leq -\alpha$, where $A = \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right)$ and $B = \frac{n}{|\alpha|} - |\alpha|$

probe selection applied to it. In the MC simulation final-state-radiation, bremsstrahlung, and detector effects are modeled by Pythia[69] and Geant4[81]. The gaussian convolution addresses energy scale shifts and resolution degradation in data relative to the simulation. This model has 2 floating parameters.

The advantage of the CB⊗BW model is the added freedom compared to the MC⊗Gauss model, but sometimes this can also give too much freedom and a fit convergence in a unphysical part of phase space can be obtained. Also more peculiar effects of the detector response cannot be modeled correctly as will be shown in section 6.2.3. The MC⊗Gauss model is more robust, but has the risk that the final-state radiation, bremsstrahlung and tails of the resolution function might not be perfectly modeled by the simulation.

For the background model, a polynomial as well as an exponential function could be used with similar performance. Here the exponential function will be the baseline and the polynomial option is used as a cross-check and to determine our systematic uncertainty. For the reconstruction efficiency, the turn-on in the invariant mass spectrum is slower than for the other efficiency measurements since the energy resolution on superclusters failing the electron reconstruction is worse than for reconstructed electrons. Therefore the model had to be adapted to allow for a turn-on. In the end, two models were used for the background:

- Exponential function : the model is simple (only 1 free parameter) and should be sufficient when the background is small compared to the signal and the energy resolution is good.

- Error function × gaussian distribution: more complicated function that falls off like an exponential but can model a kinematic turn-on in the spectrum. This is needed for poorly measured energy values, as well as binned efficiencies. This function is referred to as the RooCMS function[3].

[3] f(x)=$exp(\gamma(x - M_Z))erf(\beta(\alpha - x))$[78]

6.2 Results

6.2.1 Trigger Efficiency

The trigger efficiency is measured by asking that the probe electron fires the corresponding trigger (table 3.4) and is matched to the HLT object associated to this trigger in a cone with $\Delta R = 0.2$. The probes used are well-identified and isolated electrons. As a result, the background levels are very low ($\approx 0.5\%$) and mainly composed of real electrons coming from electroweak processes (WW, WZ, ZZ,Z$\rightarrow \tau\tau$,...). Consequently, this background can be ignored and the efficiencies can just be calculated by counting the number of passing and failing probes. In simulation, the same triggers as in data are not available and the time-dependence would be difficult to implement correctly. As a consequence, a generic electron trigger with a p_T threshold of 17 GeV and no additional cuts is used (HLT_Ele17_SW_L1R) in simulation.

The use of a different trigger in data and simulation leads to scale-factors that are strongly dependent on the kinematic variables p_T and η. Since no fitting is required for this step, a finer binning can be applied to the probe sample. To balance statistical precision and granularity, 60 bins were used (6 in p_T, 10 in η) to probe the kinematic dependency of the trigger. The results are shown in table 6.1. Overall, the trigger efficiency is 98% in the barrel and 97% in the endcap. Separate values for the efficiencies for e^+ and e^- can be found in appendix A.

The dependency of the triggers on η and p_T is visualized in the 1D-plots of fig. 6-1. The scale-factors between data and MC vary considerably: they vary 5% as a function of η and 3% as a function of p_T. In fig. 6-2 the trigger efficiency during data-taking can be seen. One can see a clear jump around 5 pb^{-1} of data collected, which is due to the change in Level-1 trigger threshold, as described in table 3.4.

6.2.2 Identification and Isolation Efficiency

Starting from an ECAL-driven Gsf-Electron in the fiducial region of the detector, the efficiency for it to pass the identification and isolation criteria is measured. The

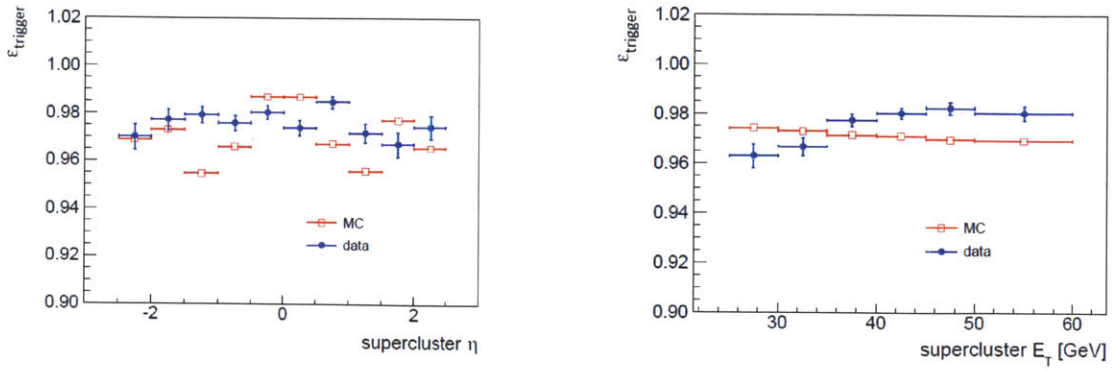

Figure 6-1: η and E_T dependence of the electron trigger efficiency.

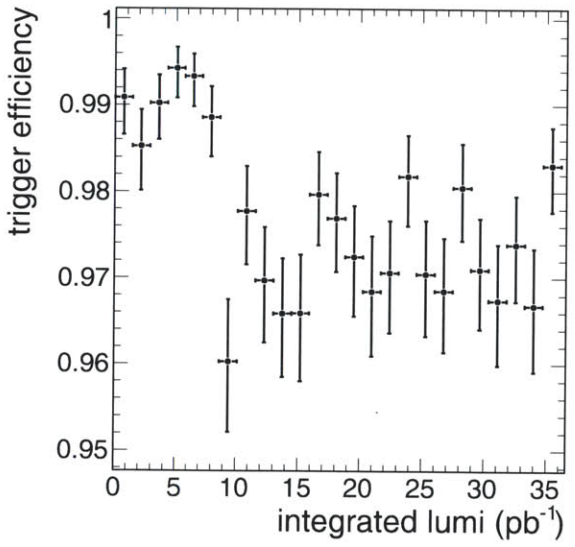

Figure 6-2: Trigger efficiency as a function of the integrated luminosity.

$25 < E_T < 30$	MC-truth	MC T&P	Data T&P	Data/MC
$-2.5 < \eta < -2.0$	0.9692 ± 0.0026	0.9687 ± 0.0027	0.9344 ± 0.0249	0.9646 ± 0.0258
$-2.0 < \eta < -1.5$	0.9733 ± 0.0026	0.9741 ± 0.0026	0.9798 ± 0.0182	1.0059 ± 0.0188
$-1.5 < \eta < -1.0$	0.9504 ± 0.0029	0.9638 ± 0.0027	0.9645 ± 0.0180	1.0008 ± 0.0189
$-1.0 < \eta < -0.5$	0.9589 ± 0.0022	0.9715 ± 0.0021	0.9830 ± 0.0119	1.0118 ± 0.0125
$-0.5 < \eta < 0$	0.9808 ± 0.0014	0.9842 ± 0.0014	0.9641 ± 0.0140	0.9796 ± 0.0142
$0 < \eta < 0.5$	0.9824 ± 0.0013	0.9864 ± 0.0013	0.9515 ± 0.0156	0.9647 ± 0.0159
$0.5 < \eta < 1.0$	0.9620 ± 0.0022	0.9716 ± 0.0021	0.9834 ± 0.0116	1.0122 ± 0.0122
$1.0 < \eta < 1.5$	0.9529 ± 0.0028	0.9652 ± 0.0026	0.9663 ± 0.0154	1.0011 ± 0.0162
$1.5 < \eta < 2.0$	0.9732 ± 0.0025	0.9773 ± 0.0024	0.9483 ± 0.0233	0.9704 ± 0.0240
$2.0 < \eta < 2.5$	0.9674 ± 0.0026	0.9676 ± 0.0027	0.9493 ± 0.0210	0.9811 ± 0.0219
$30 < E_T < 35$	MC-truth	MC T&P	Data T&P	Data/MC
$-2.5 < \eta < -2.0$	0.9664 ± 0.0022	0.9652 ± 0.0024	0.9581 ± 0.0175	0.9926 ± 0.0183
$-2.0 < \eta < -1.5$	0.9707 ± 0.0021	0.9721 ± 0.0022	0.9747 ± 0.0148	1.0027 ± 0.0154
$-1.5 < \eta < -1.0$	0.9511 ± 0.0022	0.9587 ± 0.0022	0.9835 ± 0.0098	1.0258 ± 0.0104
$-1.0 < \eta < -0.5$	0.9606 ± 0.0016	0.9700 ± 0.0016	0.9488 ± 0.0130	0.9782 ± 0.0135
$-0.5 < \eta < 0$	0.9819 ± 0.0011	0.9862 ± 0.0011	0.9806 ± 0.0090	0.9943 ± 0.0092
$0 < \eta < 0.5$	0.9818 ± 0.0011	0.9872 ± 0.0010	0.9632 ± 0.0110	0.9756 ± 0.0111
$0.5 < \eta < 1.0$	0.9617 ± 0.0016	0.9732 ± 0.0015	0.9720 ± 0.0103	0.9987 ± 0.0107
$1.0 < \eta < 1.5$	0.9538 ± 0.0021	0.9603 ± 0.0022	0.9669 ± 0.0129	1.0069 ± 0.0136
$1.5 < \eta < 2.0$	0.9786 ± 0.0018	0.9803 ± 0.0019	0.9409 ± 0.0189	0.9598 ± 0.0194
$2.0 < \eta < 2.5$	0.9630 ± 0.0023	0.9608 ± 0.0025	0.9791 ± 0.0123	1.0190 ± 0.0131
$35 < E_T < 40$	MC-truth	MC T&P	Data T&P	Data/MC
$-2.5 < \eta < -2.0$	0.9675 ± 0.0019	0.9668 ± 0.0021	0.9545 ± 0.0164	0.9874 ± 0.0171
$-2.0 < \eta < -1.5$	0.9747 ± 0.0016	0.9759 ± 0.0017	0.9818 ± 0.0107	1.0060 ± 0.0111
$-1.5 < \eta < -1.0$	0.9520 ± 0.0017	0.9544 ± 0.0019	0.9761 ± 0.0088	1.0227 ± 0.0094
$-1.0 < \eta < -0.5$	0.9581 ± 0.0014	0.9678 ± 0.0014	0.9773 ± 0.0075	1.0098 ± 0.0079
$-0.5 < \eta < 0$	0.9797 ± 0.0010	0.9881 ± 0.0009	0.9782 ± 0.0076	0.9900 ± 0.0077
$0 < \eta < 0.5$	0.9797 ± 0.0010	0.9866 ± 0.0009	0.9743 ± 0.0081	0.9876 ± 0.0082
$0.5 < \eta < 1.0$	0.9590 ± 0.0014	0.9683 ± 0.0014	0.9916 ± 0.0050	1.0241 ± 0.0054
$1.0 < \eta < 1.5$	0.9540 ± 0.0017	0.9565 ± 0.0019	0.9777 ± 0.0088	1.0221 ± 0.0094
$1.5 < \eta < 2.0$	0.9765 ± 0.0015	0.9770 ± 0.0017	0.9761 ± 0.0123	0.9991 ± 0.0127
$2.0 < \eta < 2.5$	0.9629 ± 0.0020	0.9634 ± 0.0022	0.9736 ± 0.0122	1.0106 ± 0.0129
$40 < E_T < 45$	MC-truth	MC T&P	Data T&P	Data/MC
$-2.5 < \eta < -2.0$	0.9707 ± 0.0016	0.9697 ± 0.0020	0.9960 ± 0.0059	1.0272 ± 0.0064
$-2.0 < \eta < -1.5$	0.9718 ± 0.0015	0.9722 ± 0.0017	0.9824 ± 0.0091	1.0104 ± 0.0095
$-1.5 < \eta < -1.0$	0.9512 ± 0.0016	0.9533 ± 0.0017	0.9825 ± 0.0069	1.0307 ± 0.0075
$-1.0 < \eta < -0.5$	0.9606 ± 0.0012	0.9659 ± 0.0013	0.9829 ± 0.0060	1.0176 ± 0.0063
$-0.5 < \eta < 0$	0.9803 ± 0.0009	0.9877 ± 0.0008	0.9874 ± 0.0054	0.9997 ± 0.0055
$0 < \eta < 0.5$	0.9810 ± 0.0009	0.9882 ± 0.0008	0.9752 ± 0.0069	0.9868 ± 0.0070
$0.5 < \eta < 1.0$	0.9611 ± 0.0012	0.9656 ± 0.0013	0.9797 ± 0.0067	1.0146 ± 0.0071
$1.0 < \eta < 1.5$	0.9539 ± 0.0015	0.9547 ± 0.0017	0.9674 ± 0.0093	1.0134 ± 0.0099
$1.5 < \eta < 2.0$	0.9761 ± 0.0014	0.9760 ± 0.0016	0.9724 ± 0.0108	0.9964 ± 0.0112
$2.0 < \eta < 2.5$	0.9675 ± 0.0017	0.9691 ± 0.0019	0.9806 ± 0.0100	1.0119 ± 0.0105
$45 < E_T < 50$	MC-truth	MC T&P	Data T&P	Data/MC
$-2.5 < \eta < -2.0$	0.9720 ± 0.0020	0.9732 ± 0.0024	0.9936 ± 0.0095	1.0210 ± 0.0100
$-2.0 < \eta < -1.5$	0.9728 ± 0.0018	0.9739 ± 0.0021	0.9692 ± 0.0141	0.9952 ± 0.0147
$-1.5 < \eta < -1.0$	0.9519 ± 0.0019	0.9536 ± 0.0022	0.9763 ± 0.0110	1.0238 ± 0.0117
$-1.0 < \eta < -0.5$	0.9588 ± 0.0016	0.9614 ± 0.0017	0.9760 ± 0.0088	1.0152 ± 0.0094
$-0.5 < \eta < 0$	0.9802 ± 0.0011	0.9869 ± 0.0010	0.9853 ± 0.0068	0.9984 ± 0.0070
$0 < \eta < 0.5$	0.9806 ± 0.0011	0.9860 ± 0.0010	0.9858 ± 0.0066	0.9998 ± 0.0068
$0.5 < \eta < 1.0$	0.9609 ± 0.0015	0.9646 ± 0.0016	0.9919 ± 0.0057	1.0283 ± 0.0062
$1.0 < \eta < 1.5$	0.9507 ± 0.0019	0.9515 ± 0.0022	0.9763 ± 0.0101	1.0260 ± 0.0108
$1.5 < \eta < 2.0$	0.9766 ± 0.0017	0.9760 ± 0.0020	0.9797 ± 0.0119	1.0038 ± 0.0124
$2.0 < \eta < 2.5$	0.9642 ± 0.0022	0.9657 ± 0.0026	0.9929 ± 0.0105	1.0281 ± 0.0112
$E_T > 50$	MC-truth	MC T&P	Data T&P	Data/MC
$-2.5 < \eta < -2.0$	0.9740 ± 0.0022	0.9725 ± 0.0028	0.9646 ± 0.0205	0.9919 ± 0.0212
$-2.0 < \eta < -1.5$	0.9691 ± 0.0021	0.9693 ± 0.0024	0.9735 ± 0.0155	1.0044 ± 0.0161
$-1.5 < \eta < -1.0$	0.9476 ± 0.0022	0.9493 ± 0.0024	0.9871 ± 0.0091	1.0398 ± 0.0099
$-1.0 < \eta < -0.5$	0.9576 ± 0.0017	0.9614 ± 0.0018	0.9857 ± 0.0074	1.0253 ± 0.0080
$-0.5 < \eta < 0$	0.9809 ± 0.0012	0.9868 ± 0.0011	0.9778 ± 0.0087	0.9908 ± 0.0089
$0 < \eta < 0.5$	0.9806 ± 0.0012	0.9870 ± 0.0010	0.9846 ± 0.0080	0.9976 ± 0.0081
$0.5 < \eta < 1.0$	0.9599 ± 0.0017	0.9647 ± 0.0017	0.9906 ± 0.0067	1.0268 ± 0.0072
$1.0 < \eta < 1.5$	0.9516 ± 0.0021	0.9527 ± 0.0023	0.9769 ± 0.0119	1.0254 ± 0.0127
$1.5 < \eta < 2.0$	0.9780 ± 0.0018	0.9777 ± 0.0021	0.9762 ± 0.0139	0.9984 ± 0.0144
$2.0 < \eta < 2.5$	0.9670 ± 0.0025	0.9668 ± 0.0030	0.9650 ± 0.0178	0.9982 ± 0.0186

Table 6.1: Trigger efficiencies for $e^{\pm}$ with 6x10 bins in η-p_T.

passing sample has no significant background contribution and thus no background component is included in the fit model for the passing sample. For the failing probes, there is still considerable contamination by QCD and an exponential model is used for the background shape in the failing sample. As signal shape, the Breit-Wigner convoluted with a Crystal-Ball function is used as baseline, to allow the signal shape to adjust its final-state-radiation and bremsstrahlung model. Alternative signal and background models will be used to assess the systematic uncertainties. The fits to the barrel and endcap for all probes with p_T >25 GeV are shown in fig. 6-3.

The kinematic dependency is well-modeled by the simulation, even though the agreement is not perfect. Based on the kinematic dependency of data/simulation scale-factors, the binning chosen for the identification and isolation efficiency measurement is 3 (in p_T) x 4 (in η). The 4 bins in η are based on the different detector parts (barrel with $\eta < 0$, barrel with $\eta > 0$, endcap with $\eta < 0$, endcap with $\eta < 0$). The results for this binning are shown in table 6.2, the efficiencies extracted with the CB⊗BW are the baseline, the MC⊗Gauss model is used as a cross-check. The selection efficiency is 84% in the barrel and 73% in the endcap. Separate values for the efficiencies for e^+ and e^- can be found in appendix A.

The efficiency and scale-factor dependency on η and p_T is visualized in 1D-plots in fig. 6-4 with a finer binning than the one used for the calculation. Results with both signal models (CB⊗BW and MC⊗Gauss) are shown. The dependence of the efficiency on the electron transverse momentum is well-modeled by the simulation, and the data/MC scale-factor is fairly flat as a function of p_T. The efficiency variations as a function of pseudorapdity are more poorly modeled in the MC and the data/MC scale-factors show variations up to 5%.

The main differences between simulation and data can be read off from table 6.3, which shows the 'N-1 efficiencies' in data and simulation for all the WP80 cuts. 'N-1 efficiency' means that all cuts except for the one under study are applied and the efficiency of this cut is then investigated. It turns out that the shower shape in the barrel and the track-cluster matching in the endcap are responsible for most of the discrepancy in efficiency between data and simulation.

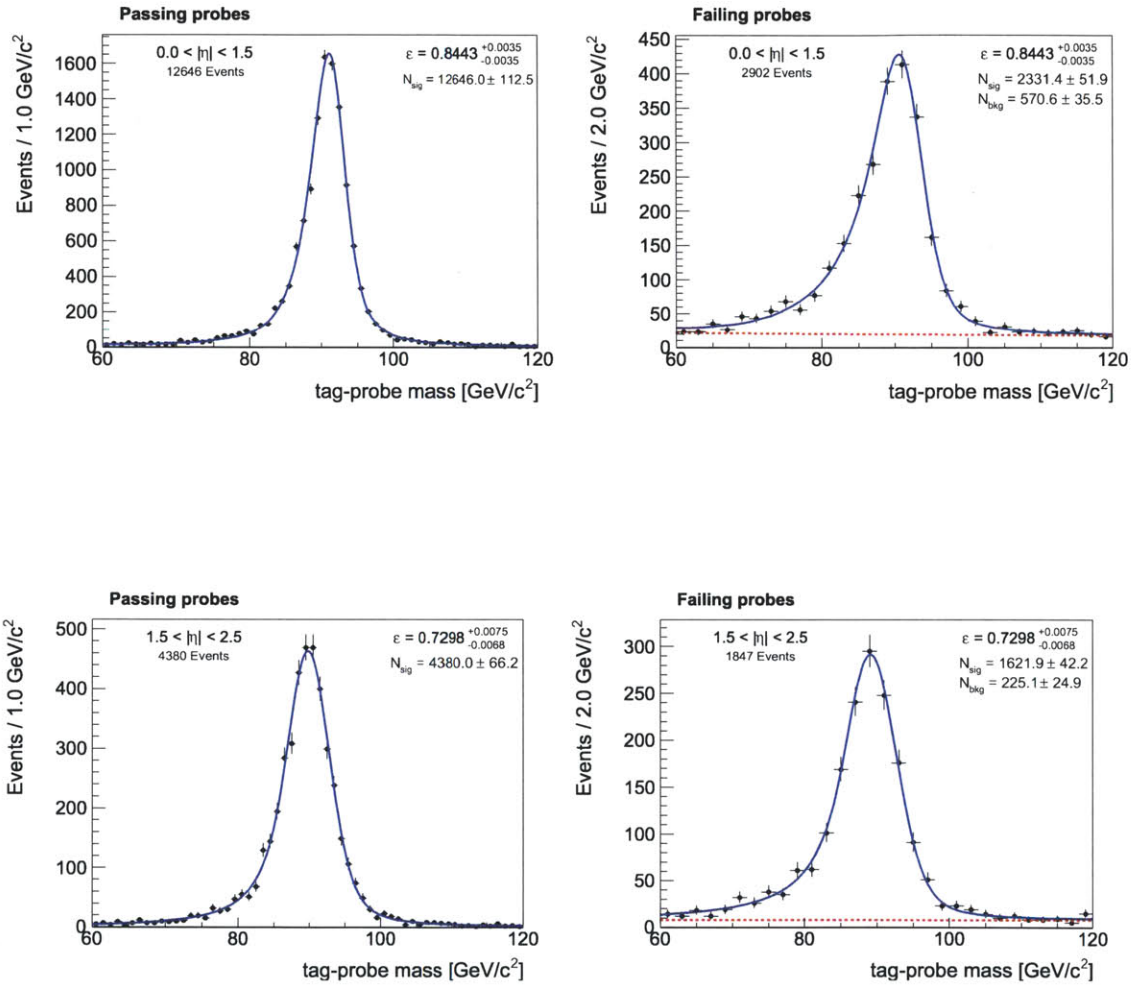

Figure 6-3: Mass fits for the passing and failing sample for the identification and isolation efficiency measurement. The top plots are for the barrel, the bottom ones for the endcap.

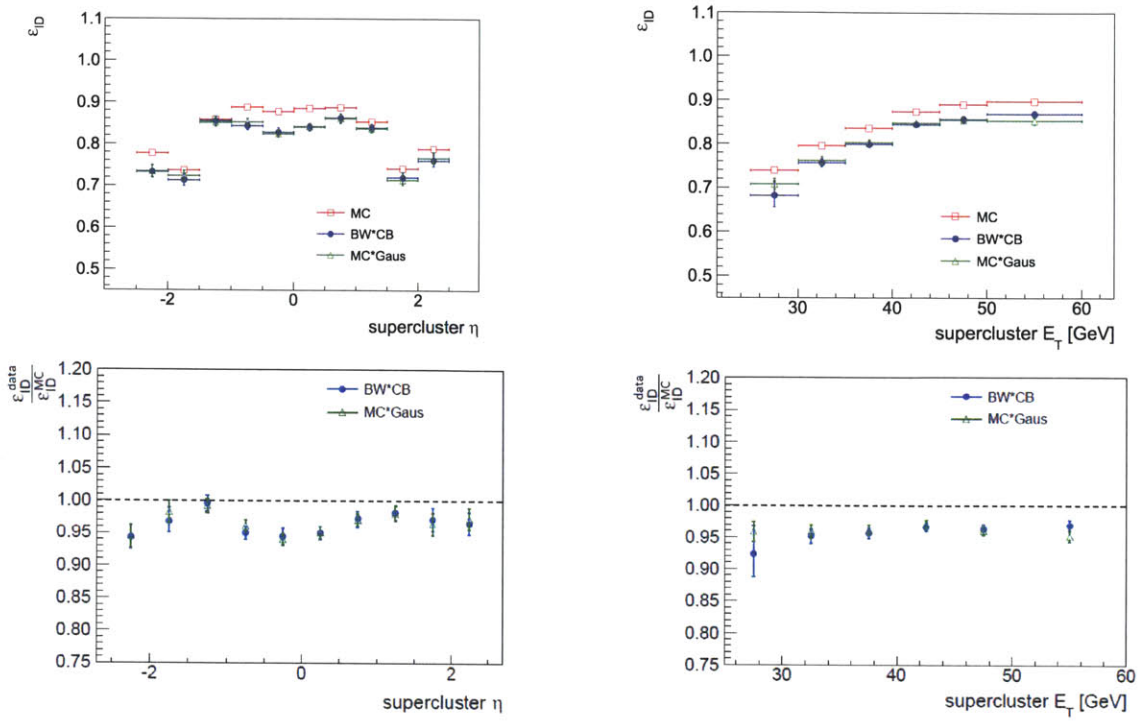

Figure 6-4: η and E_T dependence of the efficiency in data and MC (top) and the MC/data scale factor (bottom) for the electron identification and isolation efficiency. Results with both signal models (CB⊗BW and MC⊗Gauss) are shown.

$25 < E_T < 35$	MC-truth	MC T&P	*BW*CB*	Data/MC	*MC*Gauss*	Data/MC
EE-	0.6774 ± 0.0026	0.6814 ± 0.0027	0.6121 ± 0.0333	0.8983 ± 0.0490	0.6517 ± 0.0190	0.9565 ± 0.0281
EB-	0.8091 ± 0.0014	0.8097 ± 0.0015	0.7515 ± 0.0350	0.9280 ± 0.0433	0.7707 ± 0.0128	0.9518 ± 0.0159
EB+	0.8112 ± 0.0014	0.8108 ± 0.0015	0.7759 ± 0.0212	0.9570 ± 0.0262	0.7900 ± 0.0122	0.9744 ± 0.0152
EE+	0.6897 ± 0.0026	0.6944 ± 0.0027	0.6519 ± 0.0263	0.9388 ± 0.0380	0.6594 ± 0.0185	0.9496 ± 0.0268
$35 < E_T < 45$	MC-truth	MC T&P	*BW*CB*	Data/MC	*MC*Gauss*	Data/MC
EE-	0.7678 ± 0.0018	0.7692 ± 0.0020	0.7344 ± 0.0132	0.9547 ± 0.0174	0.7361 ± 0.0130	0.9570 ± 0.0171
EB-	0.8854 ± 0.0008	0.8870 ± 0.0009	0.8546 ± 0.0067	0.9635 ± 0.0076	0.8593 ± 0.0066	0.9687 ± 0.0075
EB+	0.8851 ± 0.0008	0.8872 ± 0.0009	0.8564 ± 0.0068	0.9653 ± 0.0078	0.8614 ± 0.0067	0.9709 ± 0.0077
EE+	0.7712 ± 0.0018	0.7739 ± 0.0020	0.7432 ± 0.0134	0.9603 ± 0.0174	0.7451 ± 0.0129	0.9627 ± 0.0168
$E_T > 45$	MC-truth	MC T&P	*BW*CB*	Data/MC	*MC*Gauss*	Data/MC
EE-	0.8190 ± 0.0021	0.8180 ± 0.0025	0.8066 ± 0.0165	0.9861 ± 0.0204	0.7981 ± 0.0164	0.9757 ± 0.0203
EB-	0.9144 ± 0.0009	0.9153 ± 0.0010	0.8775 ± 0.0079	0.9587 ± 0.0087	0.8718 ± 0.0080	0.9524 ± 0.0088
EB+	0.9160 ± 0.0009	0.9184 ± 0.0010	0.8769 ± 0.0259	0.9548 ± 0.0282	0.8691 ± 0.0083	0.9462 ± 0.0091
EE+	0.8253 ± 0.0021	0.8256 ± 0.0024	0.8317 ± 0.0161	1.0074 ± 0.0197	0.8211 ± 0.0157	0.9945 ± 0.0192

Table 6.2: Electron identification efficiencies for $e^{\pm}$ with 4x3 binning in η-p_T. The efficiencies extracted with the CB⊗BW-model are the baseline, the MC⊗Gauss model is used as a cross-check.

Barrel	MC T&P	Data T&P	Data/MC
I_{track}/p_T	0.993787	0.991843	0.998044
I_{ECAL}/p_T	0.977562	0.972675	0.995001
I_{HCAL}/p_T	0.998718	0.998131	0.999412
Missing expected hits	0.998182	0.992484	0.994292
H/E	0.990670	0.989924	0.999248
$\sigma_{i\eta i\eta}$	0.977772	0.964729	0.986660
$\Delta\phi_{\text{in}}$	0.982943	0.983267	1.000330
$\Delta\eta_{\text{in}}$	0.981266	0.982716	1.001480
Endcap	MC T&P	Data T&P	Data/MC
I_{track}/p_T	0.974856	0.971125	0.996172
I_{ECAL}/p_T	0.987425	0.985005	0.997549
I_{HCAL}/p_T	0.971389	0.965327	0.993759
Missing expected hits	0.995239	0.989642	0.994376
H/E	0.969803	0.973148	1.003450
$\sigma_{i\eta i\eta}$	0.992981	0.994559	1.001590
$\Delta\phi_{\text{in}}$	0.928469	0.928445	0.999974
$\Delta\eta_{\text{in}}$	0.980699	0.970452	0.989552

Table 6.3: Efficiency of each WP80 requirement after '$(N-1)$-cuts'.

6.2.3 Reconstruction Efficiency

For measuring the reconstruction efficiency, a supercluster is the starting point. This is the most background-contaminated sample, and both the passing and failing samples must be fit with a background contribution. This is especially an issue for the failing sample, where we have a low signal-to-background ratio because of the fairly impure probe sample as well as the low reconstruction inefficiency.

For the background, an exponential shape is used for the passing probes, while the turn-on effect in the mass spectrum for the failing probes forces us to use the more complicated RooCMSShape here[78]. The low signal-to-background ratio for the failing sample requires lower granularity for sufficient statistical precision and the fit is only performed once for the barrel and endcap for the whole p_T range. This is acceptable, since there does not seem to be a strong kinematic dependence on the reconstruction efficiency which is not well-modeled by simulation.

In the failing probe sample, the mass spectrum shows a small shoulder at the low mass tail due to the signal component as can be seen in fig. 6-5. This behavior is even

seen in the Monte-Carlo signal sample with the supercluster tightly matched to the generated electron. This shoulder in fig. 6-5 is caused by the energy mismeasurement of electrons traversing the gaps between the ECAL crystals. Since these electrons hit the gaps of the ECAL, the deposit in the ECAL will be lower and these electrons will have a relatively large energy deposit in the HCAL before they are stopped, thereby leading to a larger H/E. The sample can be enriched in these gap electrons by requiring a larger H/E. Figure 6-5 shows the mass distribution for the failing probes in MC. Asking for H/E >0.15 clearly leads to a more significant bump. The $\eta - \phi$ distribution of these events on the left of fig. 6-5 with H/E >0.15 is shown on the right panel, where the ECAL gaps are clearly visible.

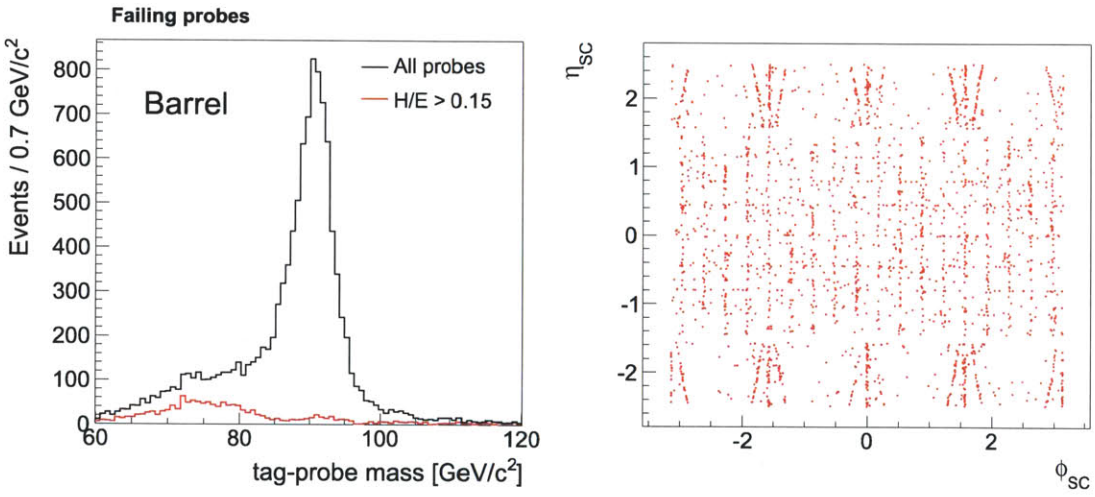

Figure 6-5: Mass distribution for the failing barrel supercluster probes and the contribution with H/E > 0.15 (left). The super-cluster $\eta - \phi$ map for failing probes with H/E > 0.15 (right).

The effect becomes even stronger when only looking at low p_T probes. Figure 6-6 shows the probes with 25GeV< p_T <30 GeV. The effect is much more pronounced because of a threshold effect. The electron energy must be incorrectly measured by a larger amount in order to fall into this low p_T bin, and this creates a distinct low mass peak. Looking at even lower p_T bins would lead to peaks further away from the probes.

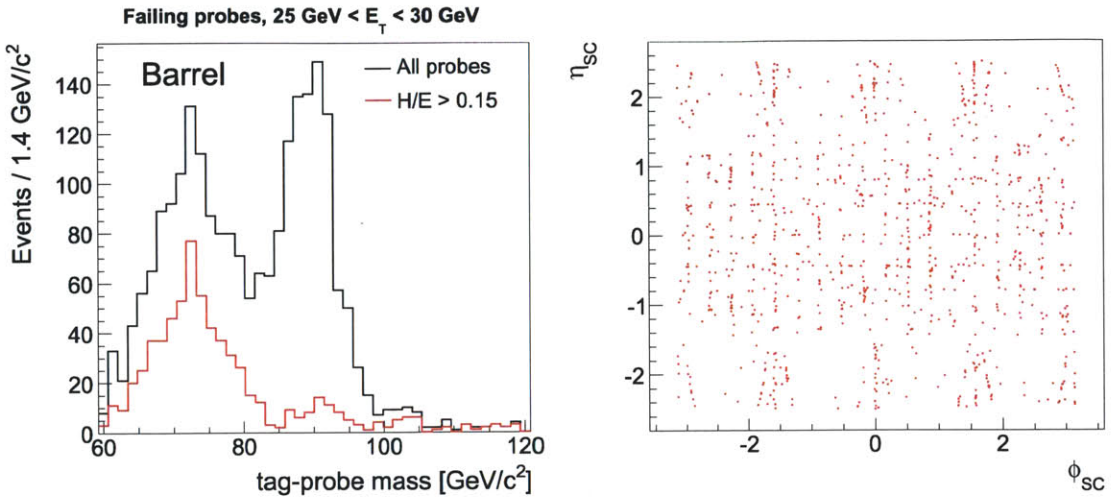

Figure 6-6: Mass distribution for the failing barrel supercluster probes with 25 GeV $< E_T <$ 30 GeV and contribution with H/E > 0.15 (left) and the corresponding low E_T super-cluster $\eta - \phi$ map for failing probes with H/E > 0.15 (right). The effect is more pronounced because the electrons have been more mismeasured.

This shoulder can clearly not be modeled well by a CB⊗BW model. Therefore, the MC shape, convoluted with a gaussian resolution function and mass shift is used for the signal model in the fit. To facilitate convergence the resolution of the gaussian smearing for the failing sample is taken to be the same as the resolution of the passing sample. A systematic uncertainty is assigned to asses the maximum error that could be introduced by using the same resolution for both samples.

Cross-checks have been made using different types of cleaned probes (shower shape cut, photons with isolation cuts,...) to check the obtained values and check whether additional kinematic dependence would be required. The data/simulation scale factors are largely independent with respect to η and p_T. Therefore only one bin in the barrel and one in the endcap are sufficient to cover the dependence. The fits for the reconstruction efficiency are shown in fig. 6-7 and the final electron reconstruction efficiencies are 97.00% $\pm$ 0.35% in the barrel and 95.16% $\pm$ 1.06% in the endcap. The values for e^+ and e^- separately can be found in appendix A.

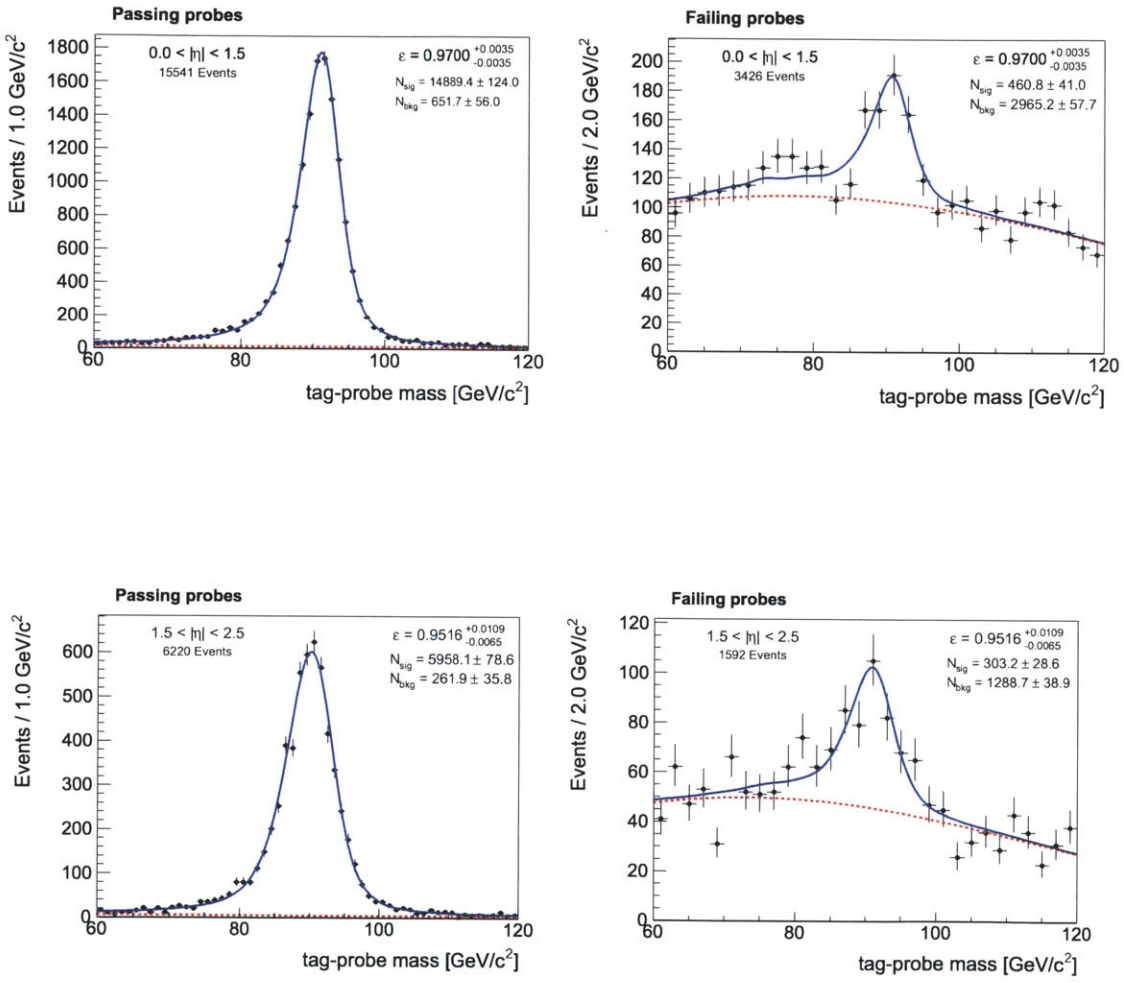

Figure 6-7: Mass fits for the passing and failing sample for the electron reconstruction efficiency measurement. The top plots are for the barrel, the bottom ones for the endcap.

6.3 Systematic Uncertainties

To calculate the systematic uncertainties, we assess the possible biases in our results because of mismodeling or limited knowledge. For example, the final-state-radiation (FSR) and bremsstrahlung may not be perfectly described by the Crystal-Ball function. An alternative model for these effects, using Pythia and Geant, is used to assess how much this can affect the final results. Extra flexibility has been built into the fit to deal with mismodeling of energy scale and resolution. The systematic uncertainties associated with the background modeling are assessed using alternative background models. For the reconstruction efficiency, one of the main uncertainties is also how well the electrons traversing the ECAL gaps are modeled. It was also checked that the tag definition does not introduce additional uncertainties.

The final systematic uncertainties are calculated using the'pseudo-experiment'-technique . The standard signal and background model are, one at a time, replaced by alternative models, where one specific effect is altered. This alternative model is then used as a template to generate pseudo-data, which is then fitted with the nominal fit model, this is a pseudo-experiment. The pull distribution for these pseudo-experiments is constructed as follows: pull=$\frac{\epsilon_{meas}-\epsilon_{true}}{\sigma_{meas}}$ with ϵ_{meas} the efficiency extracted from every pseudo-experiment, ϵ_{true} the efficiency used to generate pseudo-data from and σ_{meas} the asymmetric (MINOS[82]) fit uncertainty on the efficiency from every pseudo-experiment. The average bias in the pull translated to an average bias in the extracted efficiency from these pseudo-experiments and this is taken as the systematic uncertainty associated with this effect. Fig. 6-8 shows an example of how the systematic uncertainty is calculated using pseudo-experiments.

6.3.1 Electron Identification and Isolation Efficiency

Signal Shape

The main systematic uncertainty on the signal shape comes from the modeling of the final-state radiation and especially of the bremsstrahlung. As an alternative model compared to the CB⊗BW, the simulation, convoluted with a gaussian resolution

function is used. Here the FSR and bremsstrahlung model comes from Pythia and the Geant simulation. The systematic uncertainty is extracted as follows:

1. Fit the CB⊗BW and exponential background model to data to extract the background and signal parameters.

2. Fit the MC⊗Gauss and exponential background model to data to get reasonable parameters for the gaussian smearing.

3. Replace in the fit from step 1 the CB⊗BW with the MC⊗Gauss from Step 2, but keep the background shape and yield from step 1 and generate 1000 sets of pseudo-data from this.

4. Fit the pseudo-data with the nominal CB⊗BW and exponential background model.

5. Extract the systematic uncertainty by looking at the bias in the efficiency extracted from the pseudo-experiments.

The different steps are represented graphically in figure 6-8. The signal shape uncertainty is 0.06% in the barrel and 0.18% in the endcap.

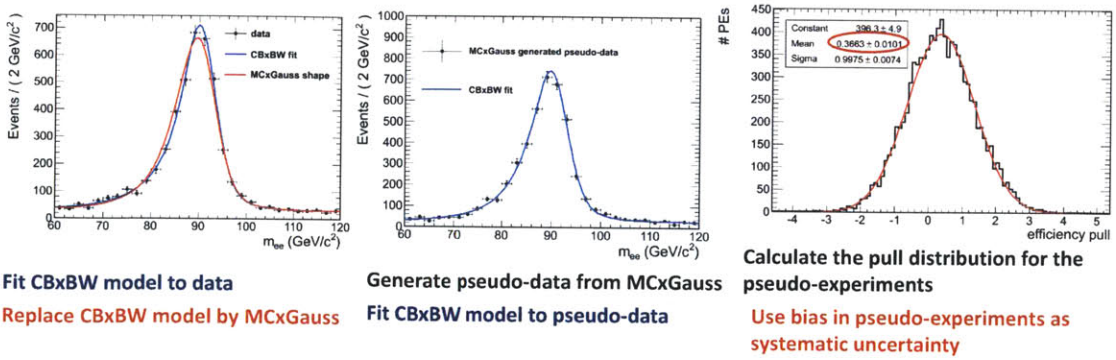

Figure 6-8: Procedure to extract a systematic uncertainty for the signal shape assumption in the electron identification and isolation efficiency.

Background Shape

A realistic alternative background model is an inverse power-law ($f(m) = m^{-\alpha}$). The systematic uncertainty will be calculated as follows:

1. Fit the CB⊗BW and exponential background model to data to extract the background and signal parameters.

2. Determine the order of the power-law by fitting CB⊗BW and power-law to the data, where we fit this time from 60-180 GeV. The larger window allows for a more precise determination of the order of the power-law.

3. Replace in the fit from step 1 the exponential model with the power law from Step 2 and generate 1000 sets of pseudo-data from this.

4. Fit the pseudo-data with the nominal CB⊗BW and exponential background model.

5. Extract the systematic uncertainty by looking at the bias in the pseudo-experiments.

We find the values of α determined from the fits to be approximately 1. (0.97 in barrel and 1.06 in endcap) The uncertainty on the efficiency associated with the background model is 0.06% on the in the barrel and 0.12% in the endcap.

These systematic uncertainties have been re-calculated in the 12 bins used in the efficiency measurements. Propagating the separate systematic uncertainties through, gives a 0.22% systematic uncertainty on the efficiency due to signal shape and 0.20% due to background modeling.

6.3.2 Electron Reconstruction Efficiency

Signal Shape

The low mass shoulder coming from electrons traversing the ECAL gaps gives a large systematic uncertainty (see section 6.2.3). To assess it, we try to change the position

and the magnitude of this shoulder, determining how the signal shape would change if the effect on the energy measurement or the amount of electrons going through the gaps was different than expected. To isolate the shoulder from the rest of the distribution we use the H/E cut that was shown in fig. 6-5 to be able to single out the effect to a large extent. So two mass templates are created, one with $H/E < 0.15$ which remains unaltered, and one with $H/E > 0.15$ which will be used to investigate the uncertainties. The procedure used to extract the systematic uncertainties associated with the shoulder is:

1. Fit the MC⊗Gauss and RooCMS background model to data to extract the background and signal parameters.

2. Create the following alternative MC⊗Gauss model depending on the effect being studied:

   (a) Split up the Z MC sample according to H/E. Weight the amount of $H/E > 0.15$ with a factor 0.5 for the down fluctuation and a factor 1.5 for the up fluctuation. Add the weighted $H/E > 0.15$ component to the unchanged $H/E < 0.15$ component and convolve it with a gaussian resolution to get two modified MC⊗Gauss models.

   (b) Split up the Z MC sample according to H/E. Move the $H/E > 0.15$ distribution up and down with 7.5 GeV (half the distance between the $H/E > 0.15$ and $H/E < 0.15$ peaks). Add the shifted $H/E > 0.15$ component to the unchanged $H/E < 0.15$ component and convolve it with a gaussian resolution to get two modified MC⊗Gauss models.

3. Replace in the fit from step 1 the MC⊗Gauss model with the modified MC⊗Gauss models from Step 2 and generate 1000 sets of pseudo-data from this. For every effect, there are two models, based on whether the effect was increased or decreased.

4. Fit the pseudo-data with the nominal MC⊗Gauss and RooCMS background model background model.

5. Extract the systematic uncertainty by looking at the bias in the pseudo-experiments.

We combine in quadrature the efficiency uncertainties due the normalization and the position of the low-mass shoulder. The efficiency uncertainty from the normalization is found to be 0.43% in the barrel and 0.48% in the endcap. The efficiency uncertainty due to the position is found to be 0.25% in the barrel and 0.54% in the endcap.

The next effect under study is the FSR and bremsstrahlung model, which could not be well-described in the simulation. To check this the results can be compared with an alternative model that has the freedom to adjust the low-mass tail: the MC⊗Gauss model can be replaced by the CB⊗BW model, which has such freedom. The shoulder at low mass will make it impossible though to fit a CB⊗BW to the full mass spectrum in data, therefore it is only fitted to the H/E < 0.15 part. The uncertainties associated with the H/E > 0.15 have been determined before. To calculate the systematic uncertainty on the efficiency due to imperfect FSR and bremsstrahlung modeling, we first need to determine realistic parameters for our CB⊗BW:

1. Fit the MC⊗Gauss (H/E<0.15) and RooCMS background model to H/E $<$ 0.15 data to extract the background and signal parameters for this subsample.

2. Fit the CB⊗BW and RooCMS background model to H/E < 0.15 data while keeping the background model fixed to the parameters extracted in step 1 (otherwise signal and background models are combined). Extract realistic CB⊗BW parameters from this fit.

3. Fit the MC⊗Gauss and RooCMS background model to **all** the data with extract the background and signal parameters for the full sample.

4. Split up the Z MC according to H/E. Replace the H/E < 0.15 component with the CB⊗BW model extracted in Step 2. Add this model to the Z MC with H/E > 0.15 convoluted with the gaussian resolution derived in step 3 get a modified MC⊗Gauss model.

5. Replace in the fit from Step 3 the MC⊗Gauss model with the modified model created in Step 4 and generate 1000 sets of pseudo-data from this.

6. Fit the pseudo-data with the nominal MC⊗Gauss and RooCMS background model background model.

7. Extract the systematic uncertainty by looking at the bias in the pseudo- experiments.

The systematic uncertainty on the efficiency due to the FSR and bremsstrahlung modeling is found to be 0.50% in the barrel and 0.22% in the endcap.

Finally, the resolution of the gaussian smearing of the signal shape in the failing sample is fixed to be the same as the resolution of the signal shape in the passing sample. Therefore the resolution determination is completely dominated by the resolution in the passing sample. To assess an uncertainty for this assumption, the simulation without any extra smearing and with twice the smearing of the passing sample will be used as a template for the signal shape in the failing sample:

1. Fit the MC⊗Gauss and RooCMS background model to all the data with to extract the background and signal parameters for the full sample.

2. Change the smearing in the failing sample from being the same as the passing sample to twice the passing sample and remove all gaussian smearing altogether to get two alternative models.

3. Replace in the fit from step 1 the MC⊗Gauss model with the modified models created in Step 2 and generate 1000 sets of pseudo-data from this.

4. Fit the pseudo-data with the nominal MC⊗Gauss and RooCMS background model background model.

5. Extract the systematic uncertainty by looking at the bias in the pseudo- experiments.

The systematic uncertainty on the efficiency due to resolution effects is 0.30% in the barrel and 0.50% in the endcap. The systematic uncertainties from the resolution

effects, the low mass shoulder as well as the FSR and bremsstrahlung effects, are then combined in quadrature to give the final systematic uncertainty on the signal shape for the reco-efficiency.

Background Shape

To estimate the background-shape systematic, we rely on simulation to provide an alternative background model from which to generate pseudo-data. A background template is constructed from QCD and W simulation samples, since these are expected to be the dominant backgrounds to the Z mass spectrum. The systematic uncertainty is calculated as follows:

1. Fit the MC⊗Gauss and RooCMS background model to all the data to extract the background and signal parameters for the full sample.

2. Replace in the fit from step 1 the RooCMS background model with the alternative shape provided by the QCD and W Monte Carlo samples and generate 1000 sets of pseudo-data from this.

3. Fit the pseudo-data with the nominal MC⊗Gauss and RooCMS background model background model.

4. Extract the systematic uncertainty by looking at the bias in the pseudo- experiments.

The systematic uncertainty due to the background modeling is 0.70% in the barrel and 0.50% in the endcap. To make sure the high level of background does not bias our efficiency measurement, cross-checks have been made using different probes, with some quality criteria applied to them: photons with tight track-isolation cuts, photons with calorimeter isolation cuts and and superclusters with shower shape cuts. All these results were consistent, with a maximal variation of 0.4% in the barrel and 0.6% in the endcap.

Summary

All the systematic uncertainties for the efficiency measurement are summarized in table 6.4. Summed up, there is a 1% systematic uncertainty on the efficiency measurement both in the barrel and the endcap, which is comparable to the statistical uncertainty on the measurement.

	$\varepsilon_{\mathrm{ID}}$		$\varepsilon_{\mathrm{RECO}}$			
	Sig. shape	Bkg. shape	Sig. shape	Bkg. shape	Resolution	**total**
Barrel	0.06%	0.06%	0.71%	0.70%	0.30%	**1.04%**
Endcap	0.18%	0.12%	0.76%	0.50%	0.50%	**1.06%**

Table 6.4: Summary of systematic uncertainties for electron reconstruction efficiencies.

Chapter 7

W signal extraction

7.1 Introduction

After applying the full electron identification and isolation selection, there is still a substantial amount of background left. For further distinguishing between the W signal and the remaining QCD background, we exploit the other leg of the W decay. A W decays to an isolated lepton and a neutrino. Because of the presence of a neutrino, W events will have real missing transverse energy, in contrast to the most of the remaining background. Therefore, a fit of the missing transverse energy spectrum for the selected data sample is performed to extract the W signal yield. There are two possibilities to fit the $\not\!\!E_T$: using fixed shape templates or analytical functions. For the signal and the electroweak backgrounds fixed shapes will be used, which are derived from MC simulation with some corrections derived from data. This can be done since the electroweak processes are fairly well described by simulation. For the QCD background, on the other hand, we do not want to use simulated shapes because we do not expect the simulation to describe the QCD background well enough. Therefore a data-driven method must be used for which the $\not\!\!E_T$ shape is selected from a background-enriched sample or is described by an analytical function that adjusts to the data.

A fixed shape template for the QCD background can be created by inverting some of the selection requirements. Inverting means that the cut variables are required to

be larger instead of smaller than their cut value, which enriches the sample in background. For example, inverting the GSF-track–supercluster matching requirements or the isolation requirement provides a fairly pure background sample. Preferably, the cuts that are inverted are minimally correlated with the variable of interest (in this case, $\not\!\!E_T$), the remaining correlation is estimated in data and the $\not\!\!E_T$ is corrected using the knowledge of this correlation. The main advantages for a fixed shape template are that it does not suffer from poor convergence due to a lack of events in the selected data sample and that it allows to compare other kinematic distributions (p_T,M_T) between the background-enriched sample and the selected-data sample. In practice, it is very difficult to correct for all the correlations between the $\not\!\!E_T$ and the identification and isolation variables used to create the background-enriched sample. For electrons, at least two identificationvariables need to used to create such a control sample and there are an insuffient number of events in the sample to allow for a 2D-extrapolation of the correlations. Therefore, simulation is used to chose the pair of cut variables which is least correlated with $\not\!\!E_T$. However this suffers from the assumption that the $\not\!\!E_T$ dependency on the cuts used to select the background sample is similar in data and MC or can be derived from data. To give an example, if the resolution degrades more between the anti-selected and selected sample in data than in simulation, then this method cannot compensate for that effect. It even becomes almost impossible to correct for all the correlations when using particle-flow $\not\!\!E_T$, because it has a holistic handling of the event and correlates all the available information.

An analytical function gives more flexibility and allows for adjustments between selected and anti-selected (background-enriched) samples. But the function has to describe the selected and anti-selected shapes well. This can be checked by looking at a variety of shapes in simulation as well as in data (using different anti-selections) and checking how well the function described this spectrum of shapes. Because of this increased flexibility, the analytical function method is used, as further describes in section 7.2.

For the description of the W signal, a fixed template for $\not\!\!E_T$ is used. This template is based on the shape from POWHEG MC, but corrected for $\not\!\!E_T$ scale and resolution

effects based on investigations with the Zs. This method will be explained in detail in section 7.3. For the main electroweak backgrounds, such as Z→ee and W→ $\tau\nu$, templates were extracted and corrected in a similar way and the normalization was determined by a fixed ratio of the Z→ee and W→ $\tau\nu$ yield to the $W \to e\nu$ yield. This ratio was taken from theoretical calculations. For the other EWK backgrounds, the yield was also coupled to the W yield , but the shape was taken from MC simulation. Because the EWK backgrounds can be reliably calculated, we use the W yield to establish their normalization, which also simplifies the fit.

7.2 Background Model

Our analytical fit model of the QCD background is based on the expectation of negligible real missing energy in QCD events, such that the observed missing energy stems from random fluctuations in both x and y components due to finite resolution, which follow a gaussian distribution.

We start by assuming a perfect gaussian distribution for the $\not\!\!E_T$ x and y components:

$$f(\not\!\!E_{x,y}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\not\!\!E_{x,y}^2}{2\sigma^2}} . \tag{7.1}$$

Then we combine the two components into the $\not\!\!E_T$, the magnitude of the missing transverse energy vector, dropping the normalization and we get the Rayleigh distribution:

$$f(\not\!\!E_T) = \not\!\!E_T e^{-\frac{\not\!\!E_T^2}{2\sigma^2}} . \tag{7.2}$$

This is only a crude approximation. The heavy-flavor QCD backgrounds have real missing transverse energy because of the presence of neutrino's in their semi-leptonic decays. The variance of $\not\!\!E_T$ also depends on the $\sum E_T$ of the event through stochastic noise, caused by the statistical sampling nature of the energy deposits in individual calorimeter crystal or tile. Events with more $\sum E_T$ have worse absolute resolution. The resolution roughly scales as $\sqrt{\sum E_T}$. To account for this, we will use a linear function for the resolution σ: $\sigma_{new} = \sigma_0 + \sigma_1 \not\!\!E_T$. This covers most of the effects.

The missing higher-order terms will be used to assign a systematic uncertainty to the background model. So the final fitting function for the QCD missing transverse energy, is:

$$f(\not\!\!E_T) = \not\!\!E_T e^{-\frac{\not\!\!E_T^2}{2(\sigma_0+\sigma_1\not\!\!E_T)^2}} . \tag{7.3}$$

This model describes the simulation extremely well.

To check whether the model is able to describe the data sample if it were different than the MC, we also checked the fits in different types of anti-selected MC, and anti-selected data. The track-matching requirements were inverted and thus it was required that the supercluster and extrapolated track position had more than a certain distance between them, the exact cuts can be found in table 7.1. The $\not\!\!E_T$ distribution of this sample, was shown to look very similar to the $\not\!\!E_T$ shape of the background component in the selected sample[25]. The only problem is the remaining signal contamination. For this, the shape from simulation is used with floating normalization. Fig. 7-1 demonstrates the good agreement between the fit and the background-enriched control sample. The blue line is the background shape plus the estimate for signal contamination, which affects the high MET tail and the red dotted line is the background-only estimate. The systematic uncertainties from the analytical function method are discussed in section 7.8.5.

Variable	Barrel	Endcap
$\Delta\phi_{\text{in}}$	> 0.06	> 0.04
$\Delta\eta_{\text{in}}$	> 0.007	> 0.009
H/E	< 0.04	< 0.025
Isolation		
I_{track}/p_T	< 0.09	< 0.04
I_{ECAL}/p_T	< 0.07	< 0.05
I_{HCAL}/p_T	< 0.10	< 0.025

Table 7.1: Cuts for the anti-selected sample in data

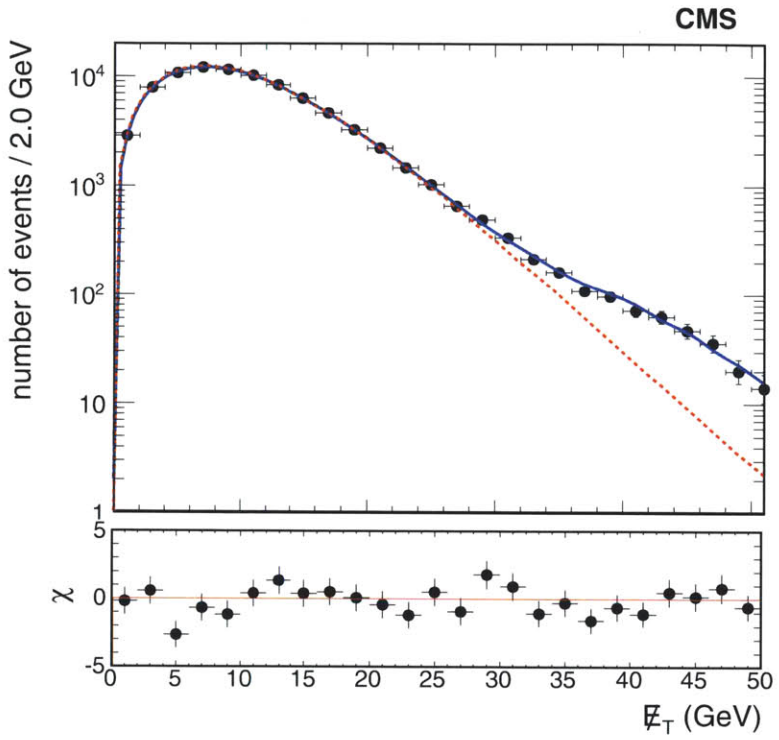

Figure 7-1: Fit to the anti-selected background sample (we invert cuts on the track-match variables while maintaining the rest of the selection, table 7.1). The blue line represents the model used to fit the inverted-selection data. This is a Rayleigh function plus a floating-yield signal template that counts for the signal contamination in the inverted region. The red dotted line shows the background-only part of the template. At the bottom the difference between the fit and the data is shown, in units of the statistical uncertainties on the data points (χ).

7.3 Z Recoil Method

The missing transverse energy distribution for the W signal is different in data than in the POWHEG MC sample. Mismodeling of detector defects and multiple interactions in the simulation and imperfect calibration in the data are just a few effects that cause this discrepancy. Because of this, the simulated shape of missing transverse energy is not able to describe the shape in data to sufficient precision to be used in the fit, and must be corrected for such effects through a data-driven approach. The Z recoil method uses the information about the calorimeter response and resolution and the underlying event from the Z events to recalculate the missing transverse energy spectrum for the W boson, resulting in a more compatible description.

The missing transverse energy for the W can be calculated from the energy of the lepton and the hadronic and electromagnetic recoil to the W boson in the event. The energy of the lepton is well-measured, but the recoil is poorly measured and needs to be modeled. Since the W and Z share a common production mechanism and are close in mass, the recoil model is based mainly on $Z \rightarrow e^+e^-$ decays. The transverse recoil ($\vec{u}_T$) for W events can be written as:

$$\vec{u}_T = -\vec{\not{E}}_T - \vec{p}_T, \tag{7.4}$$

where $\vec{\not{E}}_T$ is the missing transverse energy vector and $\vec{p}_T$ the electron transverse momentum. For Z events on the other hand, it is

$$\vec{u}_T = -\vec{\not{E}}_T - \vec{p}_{T,1} - \vec{e}_{T,2}, \tag{7.5}$$

with $\vec{p}_{T,1}$ and $\vec{p}_{T,2}$ the p_T of the first and second lepton respectively. The recoil vector can now be split into two components, one of them (u_1) parallel to the boson direction and the other one (u_2) perpendicular to the boson direction. The u_1 component is the calorimeter response balancing the boson and is mainly sensitive to calorimeter response and resolution, while the u_2 component is predominantly determined by the underlying event and the multiple interactions.

Z bosons provide a very pure sample and the direction of the Z boson is very well-determined by the two leptons. For the Zs the recoil in both data and simulation can be determined and the differences in recoil response and resolution between MC simulation and data can be calculated. The differences in response and resolution for the Z recoil get applied to the recoil in simulated events with W bosons as a function of boson transverse momentum as described in section 7.3.6. For the events with W bosons, MC simulation has to be used because the exact boson direction cannot be determined in data, since the exact direction and energy of the neutrino is unknown in data. Scale factor functions between data and MC as a function of Z boson transverse momentum are used to apply the corrections to the W recoil. Using the p_T of the vector bosons allows to account for the kinematic differences between W and Z bosons.

7.3.1 Recoil Calibration

The first step in the recoil method is to calibrate the response and resolution using Z events in both simulation and data, for subsequent application to simulated W events. The method uses a double-gaussian for the recoil as a function of boson p_T and the fit is done unbinned to exploit optimally the limited number of Z events. This method is rather complicated and to explain it we will be using a stepwise description: first the gaussian binned method will be explained, then the extrapolation to an unbinned version is made and finally the double-gaussian components are added. This is also the way the method grew historically, with the binned gaussian method used for the early data with only 3 pb^{-1}[32], after which it was refined to allow for better precision[25].

7.3.2 Binned Gaussian Method

To calibrate the recoil, first the recoil has to be fully described and understood in data as well as in MC. The response and resolution of the two components u_1 and u_2 need to be extracted.

For the response, the expectation is that u_1 is closely related to the boson p_T since calorimeter response plays an important role here. The u_2-component is expected to be independent of the vector boson p_T and centered around zero, since it is mainly determined by the underlying event and multiple interactions, which have no missing energy on average. To fit the response function a linear function is used, since it is found to describe the data well.

The resolution functions are also accepted to be different for u_1 and u_2. It is expected that u_1 again shows a stronger dependence on the boson p_T. Empirically, a second order polynomial is chosen for the description of the resolution function. Both u_1 and u_2 are described well by a second order polynomial, but u_2 has a lesser dependence on the p_T, as expected.

The description of the response and resolution for the different recoil components is extracted in data and MC using the binned gaussian method in the following way[83]:

1. The recoil components u_1 and u_2 are binned as a function of Z p_T.

2. In every bin a gaussian distribution is fitted to the distribution of the recoil components. The mean value of the fitted gaussian distribution is the recoil response of the components, the width the recoil resolution. Fig.7-2 shows an example single gaussian fit to the recoil component u_1.

3. A polynomial fit is done to the means of the gaussians as a function of Z p_T, which defines the recoil response function. (fig. 7-3) For u_2 the parameters are consistent with zero as was expected. Therefore the parametrization for u_2 response is dropped, and the value assumed to be 0.

4. A polynomial fit is done to the widths of the gaussians as a function of Z p_T to define the recoil resolution function. (fig. 7-4)

The Z p_T bins are chosen dynamically, so that they contain at least 50 events. This gives a sufficiently large sample to allow good convergence of the gaussian fit, while

it also allows for enough bins to determine the polynomial that will describe the response and resolution curves.

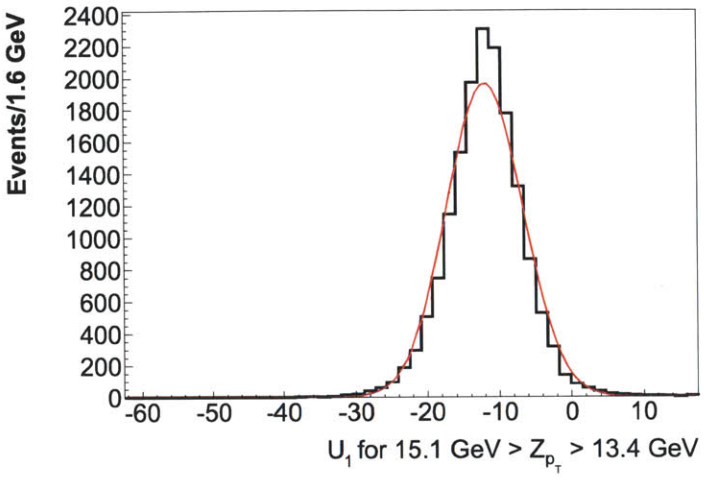

Figure 7-2: Gaussian fit to the recoil component u_1 for 15.4 GeV >Z p_T > 11.1 GeV bin[83].

7.3.3 Unbinned Gaussian Model

To go from a binned gaussian model to an unbinned gaussian version is not difficult. To parameterize the response without binning the data in Z p_T[83] a linear polynomial is fitted to the 2D-plot of $u_i(p_T)$ (with i=1,2 the two components of the recoil) as a function of Z p_T, as shown in fig. 7-5. This linear fit will give the evolution of u_i response as a function of boson p_T. Extracting the resolution is slightly more difficult, but can be done by fitting the spread in the u_i values. If we plot $\sqrt{\frac{\pi}{2}}|u_i(p_T) - u_i^{mean}(p_T)|$ as a function of Z p_T with $u_i^{mean}(p_T)$ the response value extracted from the linearly fitted response function, then we can extract the average resolution from this by fitting this 2D spectrum with a quadratic polynomial. It is important that the absolute value is taken, otherwise this curve would be flat at zero. $\sqrt{\frac{\pi}{2}}$ is a normalization factor that can be derived by looking at the expectation value of $|x-x'|$

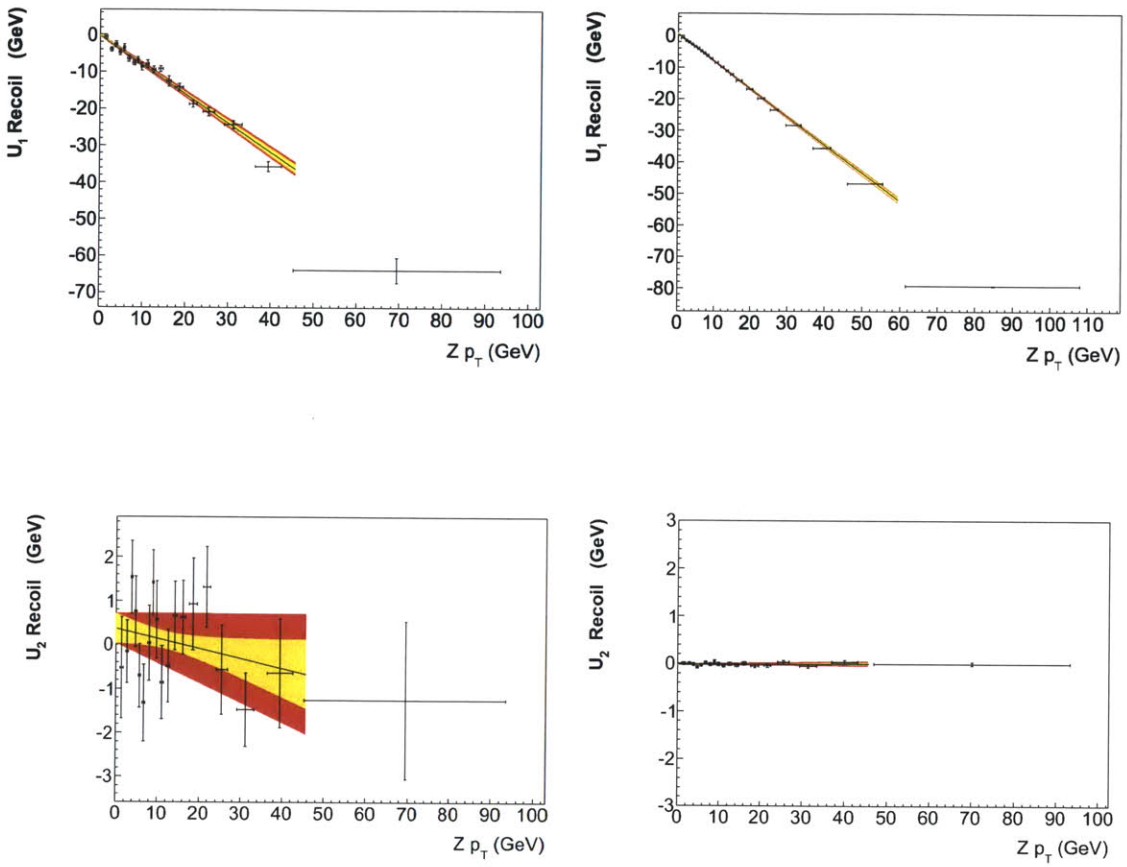

Figure 7-3: Polynomial fit of the u_1 (top) and u_2 (bottom) response in data (left) and MC(right). The yellow band shows the 1σ uncertainty from propagation of the full covariance matrix, and the red band is the 1σ uncertainty assuming maximal correlation[83]. The last point is not taken into account, because the bin spans a large Z p_T range because of the low number of events at high boson p_T and therefore the recoil components cannot be described anymore by a Gaussian distribution.

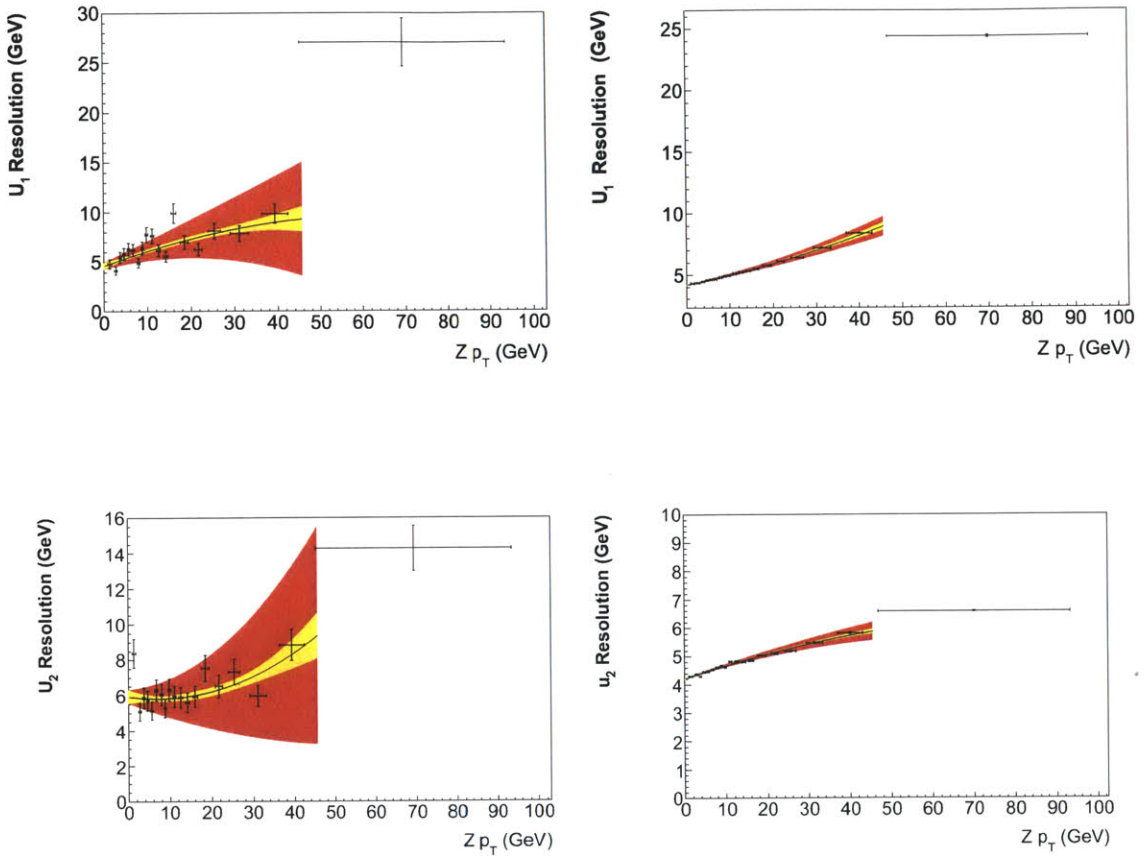

Figure 7-4: Polynomial fit of the u_1 and u_2 (bottom) resolution in data (left) and MC(right). The yellow band shows the 1σ uncertainty from propagation of the full covariance matrix, and the red band is the 1σ uncertainty assuming maximal correlation[83]. The last point is not taken into account, because the bin spans a large Z p_T range because of the low number of events at high boson p_T and therefore the recoil components cannot be described anymore by a Gaussian distribution.

with x' the mean of x for a gaussian distribution $\frac{1}{\sqrt{2\pi\sigma^2}}exp(\frac{(x-x')^2}{2\sigma^2}$:

$$< |x - x'| >= \int_{-\infty}^{+\infty} dx'|x - x'| \frac{1}{\sqrt{2\pi\sigma^2}} exp(\frac{(x - x')^2}{2\sigma^2}) = \frac{\sqrt{2}\sigma}{\sqrt{\pi}}. \quad (7.6)$$

$\sigma(\mathrm{p}_T)$ is the u_i resolution assuming a gaussian distribution of the recoil for every Z p_T value.

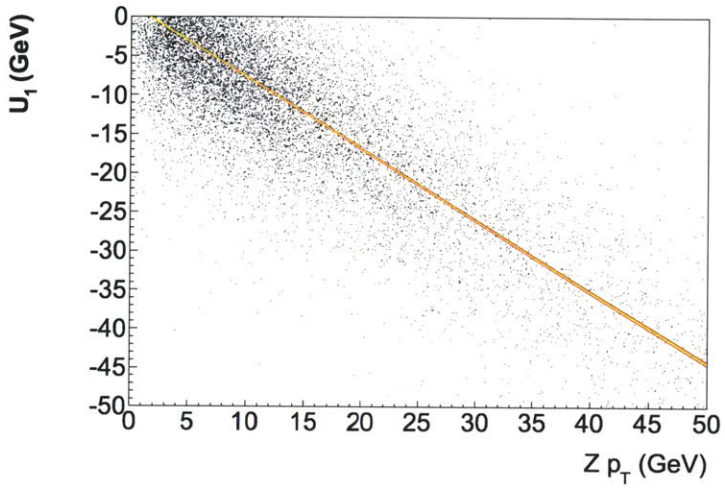

Figure 7-5: Unbinned fit of the u_1 response versus Z p_T in data[83].

7.3.4 Unbinned Double-Gaussian Model

A a significant deviation of the used single-gaussian resolution $\not\!\!E_T$ model is visible in fig. 7-6, in which the single- and double-gaussian fits to the pull $\frac{u_1(\mathrm{p}_T)-u_1^{mean}(\mathrm{p}_T)}{\sigma_1(\mathrm{p}_T)}$ for the $Z \rightarrow e^+e^-$ data is shown. The double-gaussian describes the data clearly better; the bad modeling using a single-gaussian is caused by the strong $\sum E_T$-dependence of the $\not\!\!E_T$.

The unbinned double-gaussian method[83] uses the same technique for extracting the response as the unbinned gaussian method, fitting a mean polynomial distribution to the individual recoil components. But to extract the two resolution factors σ_a and σ_b for every recoil component, more work has to be done. The resolution as

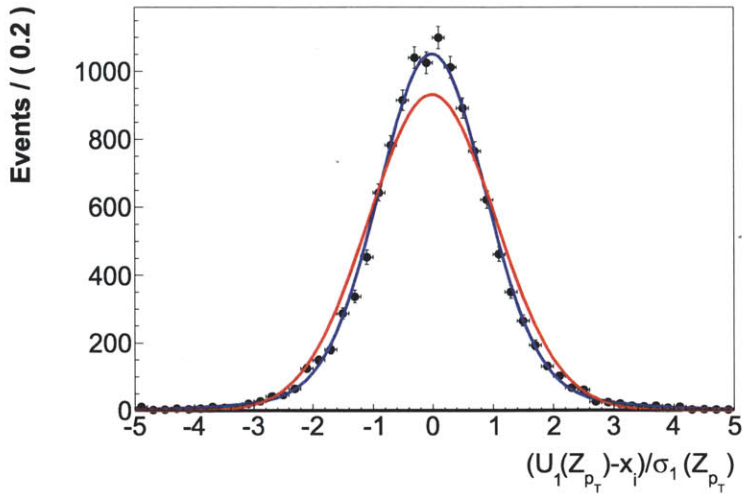

Figure 7-6: Single-gaussian (red) and double-gaussian fit to the pull distribution for the u_1 component. x_1 is the recoil response for u_1 component: $u_1^mean(p_T)$[83].

calculated in section 7.3.3 for the unbinned single-gaussian technique, provides us with the weighted mean of the two resolutions of the double-gaussian: $\sigma_{mean} = f\sigma_a + (1-f)\sigma_b$ with σ_a the smallest and σ_b the largest resolution in the double-gaussian and f the fraction of σ_a-component in the data.

To extract the average σ_a and σ_b, a fit to the pull of the recoil components $\frac{u_i(p_T)-u_i^{mean}(p_T)}{\sigma_{mean}(p_T)}$ is performed, as shown in fig. 7-7. In this formula $\sigma_{mean}(p_T)$ is the weighted mean of the two double-gaussian resolutions as a function of boson p_T and $u_i^{mean}(p_T)$ is the response for u_i-component of the recoil as a function of p_T. Fitting this distribution with a double-gaussian, yields $< f >$, $< \sigma_a >$ and $< \sigma_b >$, but this does not provide the evolution of f, σ_a and σ_b as a function of Zp_T.

For the larger resolution parameter σ_b, an approximation of the Zp_T dependence can be derived. This is the case because σ_a is typically 0.7-0.9 σ_{mean}, while σ_b is 1.5-2.5 σ_{mean}. As a consequence, the region with $|u_i(p_T) - u_i^{mean}(p_T)| > 3\sigma_{mean}$ is dominated by σ_b, with σ_a contributing less than 1%. Using this sample, and looking at the residual distribution $|u_i(p_T) - u_i^{mean}(p_T)|$, the evolution of σ_b vs. Z p_T can be determined[83], as shown in fig. 7-8.

The Z p_T dependence of σ_a is still unknown and even with the knowledge of

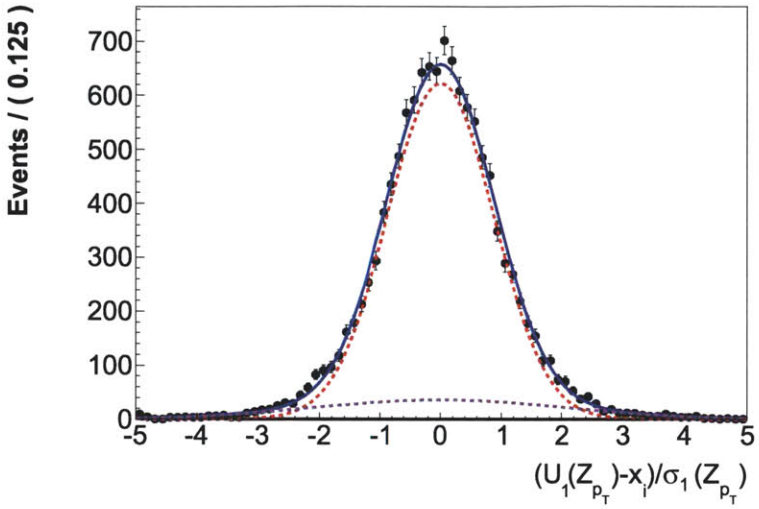

Figure 7-7: Double-gaussian fit to the residuals of the recoil in data for u_1. The fit is shown in blue, the two gaussian components in pink (σ_2) and red (σ_1)[83].

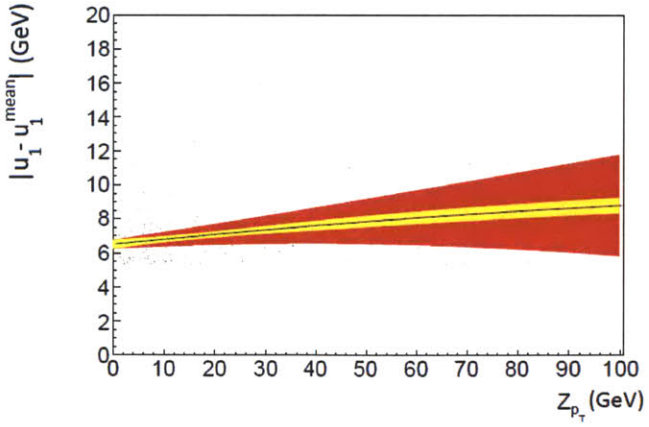

Figure 7-8: Polynomial fit to the residual distribution $|u_1(\mathrm{p}_T) - u_1^{mean}(\mathrm{p}_T)|$ as a function of Z p_T for $|u_1(\mathrm{p}_T) - u_1^{mean}(\mathrm{p}_T)| > 3\sigma_{mean}$. The yellow band shows the 1σ uncertainty from propagation of the full covariance matrix, and the red band is the 1σ uncertainty assuming maximal correlation[83].

$\sigma_b(\mathrm{p}_T)$ and $\sigma_mean(\mathrm{p}_T)$, it cannot be derived without making at least one additional assumption. Two such possible were investigated:

1. $\frac{\sigma_a(\mathrm{p}_T)}{\sigma_{mean}(\mathrm{p}_T)}$ is kept constant, and the fraction $f(\mathrm{p}_T)$ of $\sigma_a(\mathrm{p}_T)$ is recomputed using the formula for σ_{mean}: $\sigma_{mean} = f\sigma_a + (1-f)\sigma_b$

2. the fraction $f(\mathrm{p}_T)$ of $\sigma_a(\mathrm{p}_T)$ is kept constant and $\sigma_a(\mathrm{p}_T)$ is computed using the formula for σ_{mean}.

The first method is used data because the assumption that $\frac{\sigma_a}{\sigma_{mean}}$ is constant seems to uphold in MC. The second method is used to cross-check and to assign a systematic uncertainty to the assumption made using method 1[83].

So finally the information about the response and the resolution for the different recoil components is extracted as follows (the i=1,2 always points to the different recoil components u_1 and u_2):

1. Fit $u_i(\mathrm{p}_T)$ vs. Z p_T with a linear fit to extract the response function.

2. Fit $|u_i(\mathrm{p}_T) - u_i^{mean}(\mathrm{p}_T)|$ vs. Zp_T with a polynomial to extract $\frac{2}{\sqrt{2\pi}}\sigma_{i,mean}$.

3. Fit the pull $\frac{u_i(\mathrm{p}_T)-u_i^{mean}(\mathrm{p}_T)}{\sigma_{i,mean}(\mathrm{p}_T)}$ to extract $< f_i >$, $< \sigma_{i,a} >$ and $< \sigma_{i,b} >$.

4. Fit the absolute pull $|\mathrm{residual}_i| = |u_i(\mathrm{p}_T) - u_i^{mean}(\mathrm{p}_T)|$ for $|\mathrm{residual}_i| > 3$ to extract the functional dependence for $\sigma_{i,b}$ as a function of Zp_T

5. Calculated the corrected $\sigma_{i,b}(\mathrm{p}_T)$ using the information from the previous step.

6. Assuming $\frac{\sigma_{i,a}(\mathrm{p}_T)}{\sigma_{i,mean}(\mathrm{p}_T)}$ is a constant ratio, recalculate $\sigma_{i,a}(\mathrm{p}_T)$ using the formula $\sigma_{i,mean} = f\sigma_{i,a} + (1-f_i)\sigma_{i,b}$ and the results from the previous steps.

This describes fully the response and resolutions as a function of Z p_T[83]. This method is done both in Z data as well as in simulation and then scale-factors are calculated by taking the ratio, which are then applied to the W MC. More details can be found in section 7.3.6.

7.3.5 NNLO Corrections

Besides correcting for the recoil differences in data and MC, extra corrections are applied using NNLO and NNLL predictions for the W distribution. These NNLO corrections will be applied to the W MC together with the recoil calibration. Resbos predicts the boson transverse momentum distribution at NNLO level with NNLL resummation. Our default POWHEG MC is then reweighted based on the differences in the boson p_T spectrum predicted by POWHEG and Resbos. This will allow us to make NNLO corrections to the predictions. It has been shown in the Z p_T spectrum that the NNLO corrections improve the agreement between data and MC[25].

7.3.6 Applying the Recoil Calibration

As described above, the full description of the recoil is performed using $Z \rightarrow e^+e^-$ events in both data and simulation. The ratios of data to MC for the response and resolution functions as a function of the boson transverse momentum will give the calibration curves as a function of boson p_T, which are then applied to simulated W events to obtain the best description of the W data. The final $\not\!\!E_T$ templates is extracted as follows:

1. Reweight every simulated W event with a weight factor derived from the differences between POWHEG and Resbos to include higher-order QCD corrections.

2. For every simulated W event, calculate the corrected u_1 average response (u_1^{mean}) and u_1 and u_2 resolutions (σ_i) as follows with p_T^W the W boson p_T extracted from MC generator information:

$$u_{W,1}^{mean,data}(p_T^W) = \frac{u_{Z,1}^{mean,data}(p_T^W)}{u =_{Z,1}^{mean,MC}(p_T^W)} u_{W,1}^{mean,MC}(p_T^W) \tag{7.7}$$

$$\sigma_{W,i}^{data}(p_T^W) = \frac{\sigma_{Z,i}^{data}(p_T^W)}{\sigma_{Z,i}^{MC}(p_T^W)} \sigma_{W,i}^{MC}(p_T^W),\ i = 1,2\ \&\ \sigma = \sigma_a, \sigma_b. \tag{7.8}$$

3. Calculate new recoil components for the event by sampling randomly from a

gaussian with the response and resolution values determined above: u_i'=Double-Gauss(u_i^{mean},$\sigma_{i,a}$,$\sigma_{i,b}$)

4. Add the lepton vector to these recoil components to get the new $\not\!\!E_T$ for the event

The end result is a fully corrected data-driven $\not\!\!E_T$ template for the missing transverse energy in W events in data, as can be seen in fig. 7-9 which shows the MC with and without the recoil calibration applied. The $\not\!\!E_T$ signal shape template is used in the signal extraction fit discussed in section 7.7.

7.3.7 Systematic Uncertainties for the Recoil Method

There are several sources of systematic uncertainties in the Z recoil method. As mentioned before, the assumption of keeping $\frac{\sigma_a}{\sigma_{mean}}$ fixed leads to a systematic uncertainty, which is assessed by using a different assumption (keeping the fraction f constant) and looking at the influence of this other assumption on the shape. A second effect which is added as a systematic uncertainty has to do with differences between Ws and Zs. The W and Z have small differences in their production mechanism, which can be seen by looking at the $\sum E_T$ distribution for W and Z data. To account for this difference, a reweighting of the Z data and MC by the ratio $\frac{\sum E_T^W}{\sum E_T^Z}$ is performed. The differences in the correction factors extracted with and without this reweighting is taken as a systematic uncertainty. The fit uncertainties on all the extracted parameters will also lead to an uncertainty on the final extracted W $\not\!\!E_T$ shape. Finally, the PDF uncertainties for the POWHEG MC are also propagated through to give an additional systematic uncertainty. The right side of fig. 7-9 shows the cumulative effect of the recoil-related systematic and statistical uncertainties on the extracted $\not\!\!E_T$ shape.

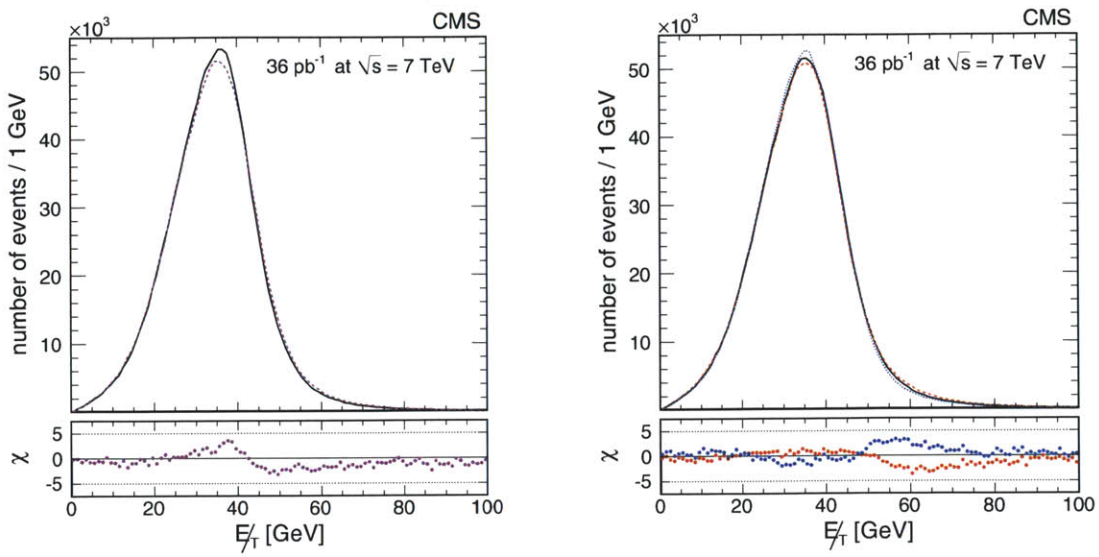

Figure 7-9: Demonstration of the effects of the recoil correction to the MC $\not{E}_T$ shape (left) and the associated systematic uncertainties (right) The residual plots at the bottom show the difference between the uncorrected and the corrected MC W $\not{E}_T$ shape (left) and the effects of the systematic uncertainties on the shape (right) in units of the statistical uncertainties on the data points (χ).

7.4 Electron Scale and Resolution

The last input needed to create the signal templates is the energy scale and resolution. The energy scale and resolution is calibrated from the Z signal in data, simultaneously for the different electron rapidity bins. The Z mass spectrum for two electrons passing all the selection cuts (table.3.2) will be used to calculate a scale factor and resolution for the electron p_T in six different η bins, amounting to twenty-one possible combinations for the dilepton system: six combinations with both electrons in same η-bin and $\frac{6\times5}{2}$ combinations with the two electrons in different η-bins. In every combination the Z mass spectrum is fitted by the POWHEG MC shape convoluted with a gaussian. In this fit, the MC mass template is scaled with the energy scale factors in data: $mass_{scale} = \frac{mass_{sim}}{\alpha_{e_1}\alpha_{e_2}}$ with α_{e_1} and α_{e_2} the energy scale factors for the electrons in the two η bins and $mass_{scale}$ and $mass_{sim}$ respectively the rescaled and the MC mass variable. The extra gaussian smearing needed for the MC mass template is the quadratic sum of the extra resolutions needed for each of the two

electrons: $\sigma_{mass} = \sigma_{e_1} \oplus \sigma_{e_2}$. A simultaneous fit over all the categories is done to extract the scale factors and resolutions, and their fit uncertainties are propagated through the analysis. The improvement of the Z mass distribution after performing this fit can be seen in fig. 7-10.

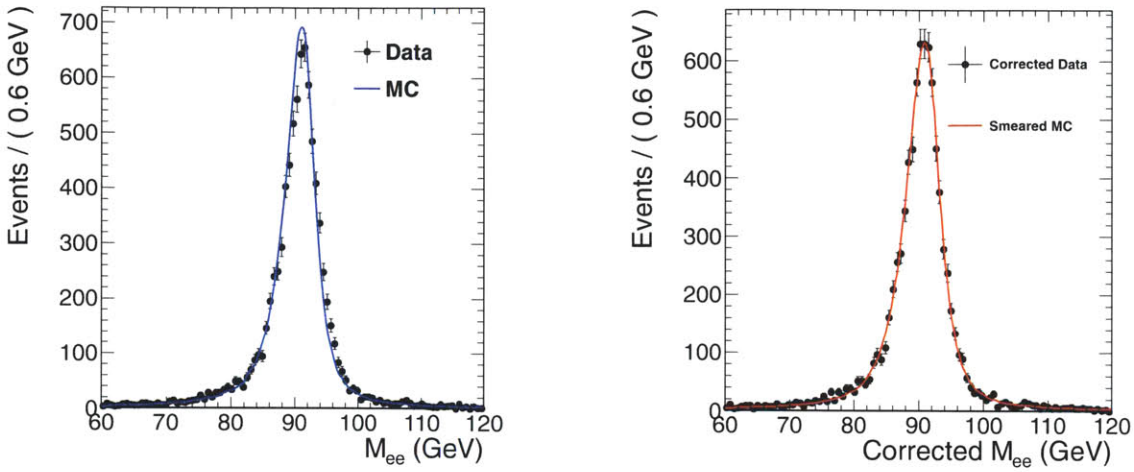

Figure 7-10: Invariant mass distribution for Z candidates in data and MC without any corrections (left). Invariant mass distribution for Z candidates in data and MC after applying the energy scale corrections to data and the necessary extra smearing to the MC (right).

The resulting electron energy scale factors and resolutions are listed in table 7.2 and plotted versus eta in figure 7-11. The η-dependency of the scale factors follows from the distribution of material of the tracker (fig. 2-5). At η=1.5 the ECAL changes from barrel to endcap, which causes the jump in the distributions. The energy scale correction is less than 1% and the extra smearing needed to make the simulation agree with the data is a maximum of 1 GeV.

The lepton-scale corrections are applied to the W lepton before making the selection, ensuring that the acceptance values are not artificially lowered. For the transverse missing energy, the corrections from table 7.2 are applied to the electron when constructing the template, since it depends on the location of the energy deposit. The energy scale factors are applied to the electrons in Z MC and additional smearing is applied to the Z MC before fitting for the recoil components, as described previously. We do not correct the electron E_T in the W Monte-Carlo directly because the effect

Region	Scale (Data→MC)	Resolution [GeV] (MC→Data)
$0 < \|\eta\| < 0.4$	0.994 ± 0.001	0.468 ± 0.135
$0.4 < \|\eta\| < 0.8$	0.995 ± 0.001	0.601 ± 0.165
$0.8 < \|\eta\| < 1.2$	0.999 ± 0.001	0.010 ± 0.538
$1.2 < \|\eta\| < 1.4442$	1.009 ± 0.002	1.014 ± 0.209
$1.566 < \|\eta\| < 2$	0.993 ± 0.002	1.063 ± 0.215
$2 < \|\eta\| < 2.5$	1.000 ± 0.002	1.022 ± 0.194

Table 7.2: Electron energy scale and resolution corrections.

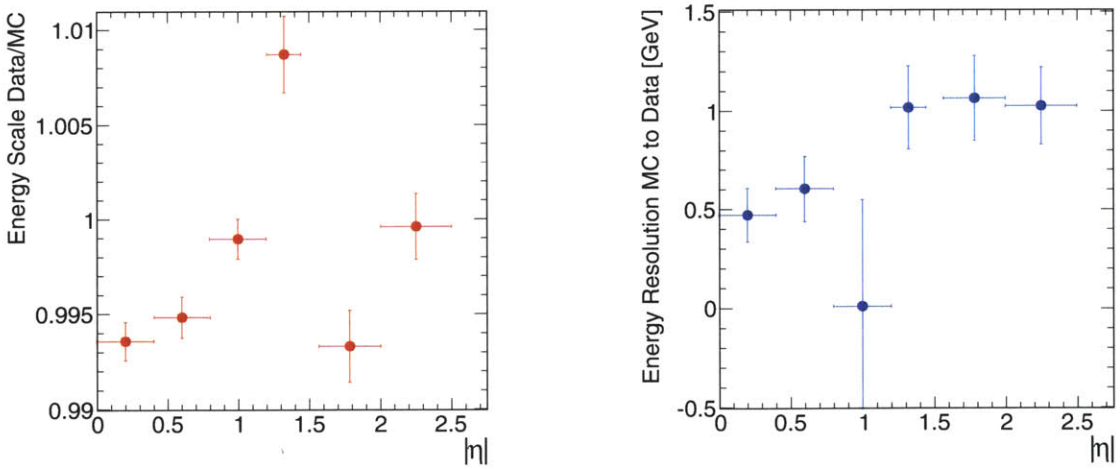

Figure 7-11: Electron energy scale (left) and resolution (right) corrections factors as a function of pseudo-rapidity.

of correction is already accounted for in the Z scale factor. The uncertainties will also be propagated through the template framework to assess the systematic uncertainties on the W yield due to changes in the $\not\!\!E_T$-shape.

7.5 Signal and Electroweak Background Templates

Using the Z recoil method and adding in the energy scale corrections and smearing, the final $W \to e\nu$ signal templates are created from recalibrated POWHEG MC and used in the fit. In the same way, the correction factors calculated will be propagated through the POWHEG MC simulation of the main electroweak backgrounds (like $Z \to ee$ and $W \to \tau\nu$) and corresponding EWK templates will be produced. The normalization of these electroweak templates is connected to the $W \to e\nu$ signal yield, through a fixed ratio of their computed cross sections. This can be done because the ratios between the production cross-sections of electroweak processes are theoretically well-understood and thus have very small theoretical uncertainties associated with them. The POWHEG samples are used and the normalization is done using the NNLO cross-section when available and the NLO estimate otherwise. For the smaller backgrounds, like $t\bar{t}$ and dibosons, the simulation is used directly without applying Z recoil corrections. The fraction of electroweak backgrounds after full selection is given in table 7.3. Drell-Yan background is part of the Z→ee estimate.

Source	N_{bkg}/N_W
Z→ee and Z→ $\tau\tau$	7.6%
W→ $\tau\nu$	3.0%
diboson production	0.14%
$t\bar{t}$	0.44%
EWK	**11.2%**

Table 7.3: Estimates of electroweak backgrounds relative to the W→eν cross section after full selection.

7.6 Charge Misidentification

This analysis concerns not only the inclusive W cross section but also the charge asymmetry between W^+ and W^- production in pp collisions. Therefore, the charge needs to be determined. One of the main systematic uncertainties on the charge ratio is the possible charge misidentification of the electrons. The electron charge confusion is mainly caused by conversions of bremsstrahlung photons close to the initial track, which confuse the gsf track reconstruction. To calculate the charge misidentification, the Z→ee sample is used again. The number of Z→ee events where the two electrons are reconstructed with the same sign (SS) and with opposite signs (OS) (fig. 7-12) can be written as a function of the number of Z→ee events (N_Z) and the charge misidentification rate for every electron:

$$OS = (1 - f_{misID_1})(1 - f_{misID_2})N_Z + f_{misID_1} f_{misID_2} N_Z \tag{7.9}$$

$$SS = (1 - f_{misID_1}) f_{misID_2} N_Z + (1 - f_{misID_2}) f_{misID_1} N_Z, \tag{7.10}$$

with f_{misID_1} and f_{misID_2} the charge misidentification rate for the first and the second electron in data. These equations can be solved simultaneously for N_Z and f_{misID}. Because of the tight WP80 cuts applied to both Z electrons, there is no background left so there is no need for mass fitting and the events can just be counted to extract the charge misidentification rate. Again, 6 lepton rapidity bins are used, leading to 21 combinations for the dilepton system. For every combination, the number of Z→ee events where both electrons are reconstructed with the same sign or with opposite sign is counted. These numbers give an over-constrained set of equations (12 unknowns for 42 constraints) and a fit is used to extract the best estimates for charge misidentification and its uncertainties. This is done in data as well as in MC simulation to calculate the correction factors in table 7.4.

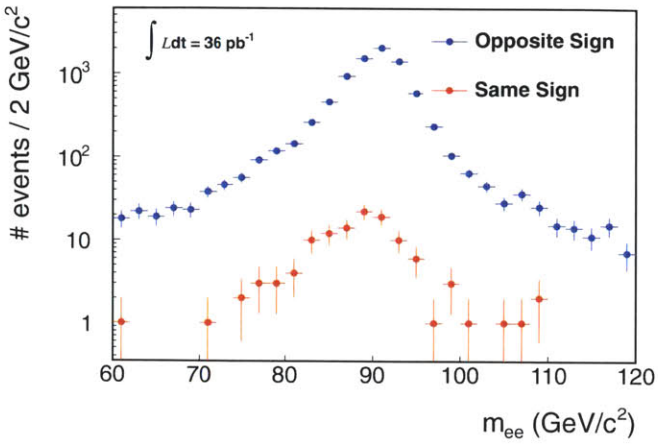

Figure 7-12: Invariant mass distribution of same sign and opposite sign Z candidates in data. These events are used to calculate the charge misidentification rate.

Region	Charge Misid (Data)	Charge Misid (MC)	Data/MC Scale Factor
$0 < \|\eta\| < 0.4$	$0.0000\ ^{+0.0005}_{0.000000}$	$0.0008\ ^{+0.0001}_{-0.0001}$	$0.00\ ^{+0.62}_{0.00}$
$0.4 < \|\eta\| < 0.8$	$0.0011\ ^{+0.0009}_{-0.0006}$	$0.0013\ ^{+0.0002}_{-0.0001}$	$0.80\ ^{+0.66}_{-0.44}$
$0.8 < \|\eta\| < 1.2$	$0.0043\ ^{+0.0016}_{-0.0013}$	$0.0027\ ^{+0.0002}_{-0.0002}$	$1.63\ ^{+0.59}_{-0.50}$
$1.2 < \|\eta\| < 1.5$	$0.0126\ ^{+0.0032}_{-0.0028}$	$0.0062\ ^{+0.0004}_{-0.0004}$	$2.05\ ^{+0.53}_{-0.47}$
$1.5 < \|\eta\| < 2$	$0.0141\ ^{+0.0032}_{-0.0028}$	$0.0145\ ^{+0.0005}_{-0.0005}$	$0.97\ ^{+0.22}_{-0.20}$
$2 < \|\eta\| < 2.5$	$0.0207\ ^{+0.0037}_{-0.0034}$	$0.0229\ ^{+0.0006}_{-0.0006}$	$0.90\ ^{+0.17}_{-0.15}$

Table 7.4: Charge misidentification rate for electrons after WP80 cuts (table 3.2).

7.7 Extracted Yields

Using the the analytical function for the QCD background and recoil-corrected templates for the signal and electroweak backgrounds, an unbinned maximum likelihood fit is performed to the missing transverse energy spectrum to extract the final W yield. The W^+ and W^- data are fitted independently in $\not\!\!E_T$, the results of these fits are shown in figure 7-13. The yield for the inclusive W cross-section is extracted by adding up the charge-specific yields. The extract signal and background shapes are combined using the extracted yields to create a model for the inclusive W $\not\!\!E_T$ distribution and the comparison of this model with data is shown on fig. 7-14. Table 7.5 gives the overall and charge specific yields, as well as the results of Kolmogorov-Smirnov statistical tests for the $\not\!\!E_T$ fits to the W^+ and W^--data, which indicate the fit reproduces the data quite well.

Sample	Extracted Yield	Kolmogorov-Smirnov probability
$W^+ \to e^+\nu$	81568 $\pm$ 297 events	0.31
$W^- \to e^-\bar{\nu}$	54760 $\pm$ 246 events	0.25
$W \to e\nu$	136328 $\pm$ 386 events	NA

Table 7.5: Yields extracted from the fits and the Kolmogorov-Smirnov probabilities of the fits. The inclusive yield is extracted by summing up the charge specific yields.

7.8 Systematic Uncertainties

The $\not\!\!E_T$ fitting procedure to extract the W yield, has several systematic uncertainties associated with it, because of the assumptions made for the signal and background templates. First a check is performed to make sure the fitting procedure itself is not biased. The Z-recoil fit also has uncertainties associated with it and these are propagated through to determine the $\not\!\!E_T$ scale and resolution. Other systematic uncertainties come from the methods used to extract the lepton scale and resolution and the charge misidentification. Finally there may exist biases due to the choice of background shapes. All these systematic uncertainties are discussed in the subsections that follow.

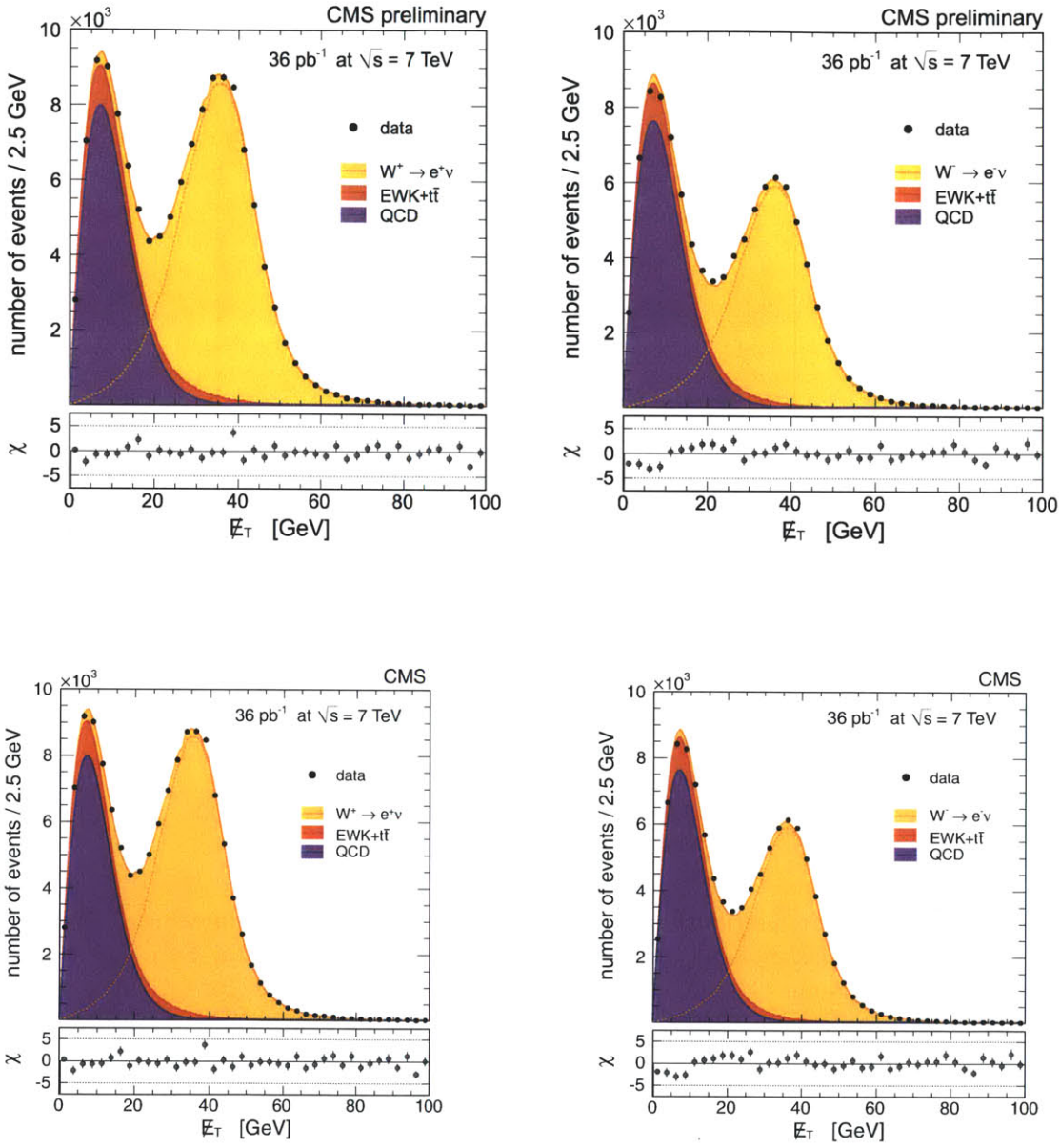

Figure 7-13: The $\not\!\!E_T$ distribution for the selected W$\to e\nu$-sample, for e^+(left) and e^- (right). The points represent the data. Superimposed are the results of the likelihood fit for QCD background (violet), all backgrounds (orange), and signal plus background (yellow). At the bottom the difference between the fit and the data is shown, in units of the statistical uncertainties on the data points (χ).

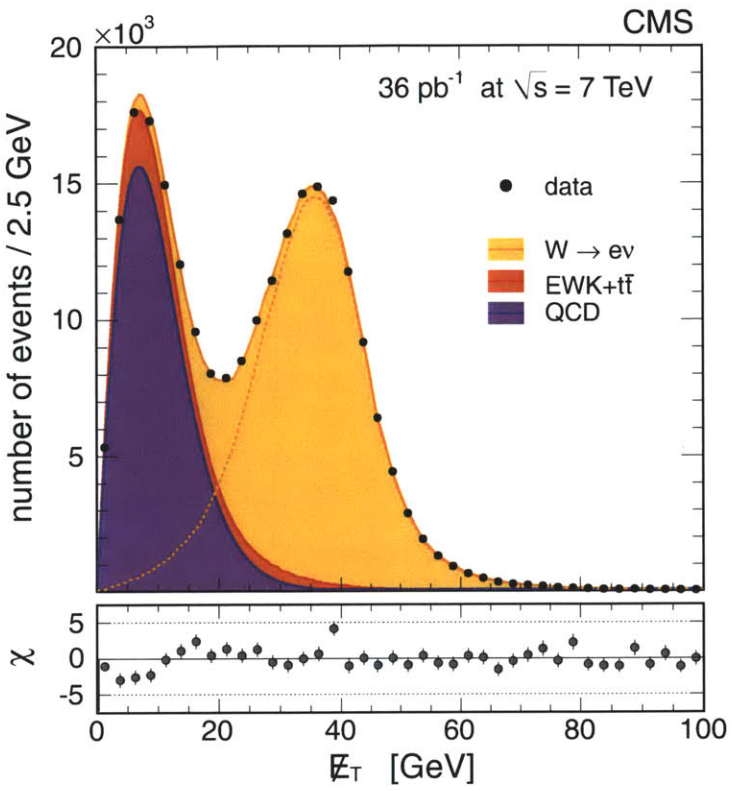

Figure 7-14: The $\not\!\!E_T$ distribution for the selected W→$e\nu$-sample, in linear scale (left) and logarithmic scale (right). The points represent the data. Superimposed are the sums of the likelihood fits to the e^+ and e^- sample for QCD background (violet), all backgrounds (orange), and signal plus background (yellow). At the bottom the difference between the model extracted from the charge-specific fits and the data is shown, in units of the statistical uncertainties on the data points (χ).

7.8.1 Internal Fit Bias

There could be an internal bias to the fitting procedure. To exclude the possibility, pseudo-experiments are generated from the best fit to data and then refitted with the same model. The results show that the method is internally unbiased, the largest bias found is 0.04% for the fit to the W^- data.

7.8.2 $\not\!\!E_T$ Scale and Resolution

In the recoil method, as described in section 7.3, there are uncertainties connected to the different parameters that go into the recoil modeling. The uncertainties on these parameters can be propagated to final template shapes. The 1σ up and down fluctuated scenarios are taken as bounding shapes, as illustrated by fig. 7-9. To determine the systematic uncertainty associated with these $\not\!\!E_T$ scale and resolution uncertainties, these shapes will be used in pseudo-experiments.

To calculate the systematic uncertainties, the starting point is always the best fit to data. The best fit model is the hybrid model with an analytical function (Rayleigh) for the jet background and fixed shape templates for the W signal and the electroweak backgrounds, which were used for the final W signal extraction fit in section 7.7. The following procedure is implemented for the $\not\!\!E_T$ scale and resolution uncertainties:

1. Fit the nominal hybrid model from section 7.7 to data and extract the best data-parameters.

2. Replace the nominal signal model by the up/down fluctuated shape, created by applying the 1 σ variations on the $\not\!\!E_T$ scale and resolution uncertainties (fig. 7-9). Do the same for the Z$\rightarrow$ee and W$\rightarrow \tau\nu$ backgrounds.

3. Generate 10000 pseudo-experiments using this modified model with the background model from step 1 and the up/down fluctuated signal and EWK background shapes.

4. Fit these pseudo-experiments with the nominal model of step 1.

5. The extracted yield bias gives the up/down uncertainty on the yield.

This method is also adopted for the other uncertainties. Alternative $\not\!\!E_T$ shapes are generated modifying a specific effect (electron energy scale, resolution, ...) and used to generate pseudo-experiments. For the $\not\!\!E_T$ scale and resolution systematic uncertainty, also the PDFs were replaced by the error PDFs to assess the PDF uncertainty on the $\not\!\!E_T$ modeling.

The systematic uncertainty due to $\not\!\!E_T$ scale and resolution is listed in table 7.6, it accounts for a 0.3% systematic uncertainty on the W signal yield.

7.8.3 Electron Energy Scale and Resolution

The uncertainties associated with the energy scale and resolution also have to be taken into account, and propagated to the final result. The uncertainties are the statistical uncertainties from the energy scale and resolution fit plus the systematic uncertainties from the fluctuations within bins. The uncertainties influence the results in two ways. First of all the energy scale uncertainties affect the acceptance. This effect can be calculated by applying the uncertainties and assessing the influence on the acceptance. This gives a 0.12% systematic uncertainty on the acceptance.

Secondly, the influence of the energy scale and resolution on the $\not\!\!E_T$ shape needs to be quantified to assess the indirect effects of the lepton scale and resolution via the recoil method on the final signal extraction. To assess this, the electron energy scale and resolution are fluctuated by $\pm$ 1σ for the electrons in the simulation before the recoil method is performed. This is done for both for the Z simulation which is used to determine the scale factors and the W simulation where these scale factors will be applied. New $\not\!\!E_T$ templates are generated based on the up and down fluctuation of every one of these effects. These templates are used in pseudo-experiments in the same procedure as described in section 7.8.2 to extract the systematic uncertainties due to electron energy scale and resolution.

A summary of the uncertainties due to energy scale and resolution can be found in table 7.6, this systematic uncertainty of around 0.5% is the largest systematic

uncertainty on the signal extraction.

7.8.4 Charge Misidentification

The uncertainties on the charge misidentification have no influence on the shapes since it is a tiny admixture of a similar shape. Therefore, the main difference comes from the extraction of the signed components. After fitting the positively charged and negatively charged spectrum separately, the charge misidentification is propagated through to get the W^+ and W^- yields. Using the uncertainties associated with the charge misidentification rate, new W^+ and W^- yields can be calculated increasing and decreasing the charge misidentification rate with its uncertainties. The systematic uncertainty on the yields due to charge misidentification can then be extracted by comparing these yields with the nominal yield extracted by propagating through the central value of the charge misidentification rate. The systematic uncertainty on the signal yields due to charge misidentification is small, only 0.15% for the W^+/W^- ratio and 0.10% for the charged yields.

7.8.5 Background Shape

The analytical background model might not describe the data perfectly and therefore lead to a bias in the measurement of the W yield. To make a credible assessment whether the model can describe the data and does not only work in simulation, a control region in data is needed. This region should exhibit $\not\!\!E_T$ behavior very similar to the selected sample but be highly background-enriched and contain as little signal as possible. It turns out that the cluster-matching variables in the electron ID are the least correlated with missing transverse energy. This can be seen intuitively by the fact that they only use position information and no energy information. Also the cluster shape variable is nearly uncorrelated with $\not\!\!E_T$, as are the conversion removal cuts. The cluster-matching cuts are not just inverted, the values are changed too. The cuts are geared to select even more QCD, thereby decreasing the signal contamination in the control sample. The cluster shape and conversion cuts are omitted, to increase

the number of events in the anti-selected sample. The final cuts for the control region are shown in table 7.1.

To determine the uncertainty introduced due to the chosen model, we will look at the effect of adding an extra high $\not\!\!E_T$ tail to the distribution. The tail determines the amount of background in the signal-rich region and uncertainty on it will directly lead to uncertainties in the measured signal yield. To assess additional effects stemming from the high $\not\!\!E_T$ tail, we add an additional second order term σ_2 to the resolution function, which is sensitive to the high tail.

$$f(\not\!\!E_T) = \frac{\not\!\!E_T}{\sqrt{2\pi}\sigma} e^{-\frac{\not\!\!E_T^2}{2(\sigma_0+\sigma_1\not\!\!E_T+\sigma_2\not\!\!E_T^2)^2}} \tag{7.11}$$

Extracting the systematic uncertainty is done in the following way.

1. Fit the model from equation 7.11 to the selected MC, anti-selected MC and anti-selected data. For the anti-selected data, the signal contamination must be taken into account. This is done in two ways. The expected signal contamination from MC is extracted and the fit is repeated and the expected signal shape from MC is taken and the yield is floated.

2. Determine the largest value of σ_2 in the fits from step 1.

3. Fit the nominal model to data and determine all the parameters.

4. Add a σ_2-tail to the nominal fit to data where the σ_2 parameter is the one from step 2.

5. Generate pseudo-experiments from this new distribution and fit it with the nominal model.

6. The bias in the signal yield is taken again as systematic uncertainty.

Figure 7-15 shows the fits to anti-selected data sample using a second order term, with the amount signal contamination floating or fixed. The values we get for σ_2 in selected MC, anti-selected MC and anti-selected data are respectively 0.0007, 0.0010

and 0.0009 GeV^{-1} for W^+. For W^- the corresponding values are 0.0008,0.0009,0.0007 GeV^{-1}. The systematic uncertainty due to incomplete modeling of the QCD tail is between 0.3% and 0.5%, the exact numbers can be found in table 7.6.

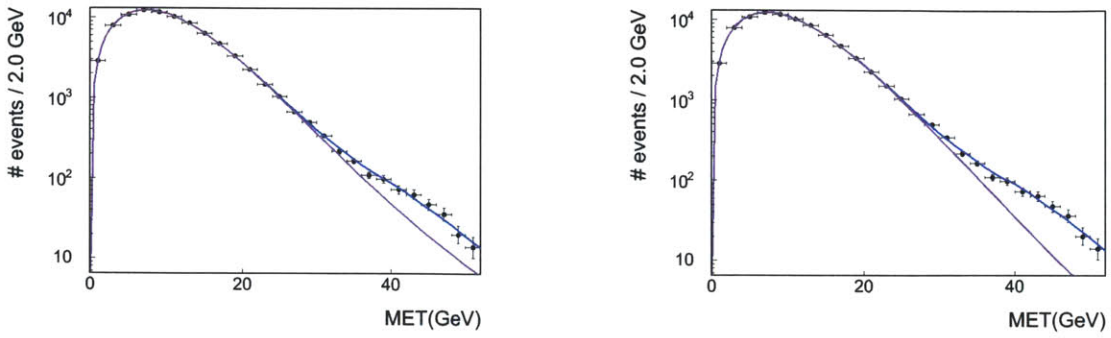

Figure 7-15: The $\not\!\!E_T$ distribution for the anti-selected data sample fitted with the second order model and a floating (left) or fixed (right) signal contamination component. The blue line is the fit to data and the purple line shows the background model only.

7.8.6 Summary of Systematic Uncertainties for the W→eν Signal Extraction

The systematic uncertainties for the W→eν signal extraction are summarized in table 7.6. For all the channels, the total systematic uncertainty associated with the $\not\!\!E_T$ fit for the signal extraction are below 1%. Thus it is not the leading systematic uncertainty, because the theoretical uncertainties on the acceptances and the systematic uncertainties from the efficiency measurements are both larger than this.

Source	$W^+ \to e^+\nu$	$W^- \to e^-\bar{\nu}$	$W \to e\nu$	W^+/W^-
Momentum scale and resolution	0.5%	0.6%	0.5%	0.1%
$\not\!\!E_T$ scale and resolution	0.3%	0.3%	0.3%	0.1%
background modeling	0.3%	0.5%	0.35%	0.4%
charge misidentification	/	0.10%	0.09%	0.15%
total	**0.7%**	**0.8%**	**0.7%**	**0.4%**

Table 7.6: Summary of the systematic uncertainties for W→eν signal extraction.

7.9 Transverse mass cross-check

Another way of distinguishing Ws from QCD background is to use the transverse mass of the lepton+$\not\!\!E_T$ system as the analysis variable. This does give comparable discriminating power as the missing transverse energy, but it introduces additional systematic effects. For these reasons the analysis was performed with $\not\!\!E_T$, however the transverse mass provides an interesting cross-check of the analysis.

Using the transverse mass as discriminator adds more correlations between electron identification variables and the fitting variable. The electron object itself is used to calculate the transverse mass and better-identified electrons will have a different transverse mass spectrum than the poorly-identified electrons, for example because the p_T resolution is better. The net result is a change in the shape for the anti-selected samples which makes it impossible to derive a model which describes both selected and anti-selected samples to be used as a template or the anti-selected sample as a control region. The QCD transverse mass is also more difficult to describe with an analytical function, and does not follow from a first principles argument as does the $\not\!\!E_T$ model.

A cross-check of the transverse mass shape was performed using the yields extracted from the hybrid $\not\!\!E_T$ fits and the corrected anti-selected sample. Fig. 7-16 shows that there is rather good agreement for the M_T distribution if we use the normalizations for signal and background as obtained in the $\not\!\!E_T$ fit. Also the p_T spectrum (fig. 7-17) is well described using this normalization. A quantitive cross-check was also performed using fixed shape templates for transverse mass, the extract yields were 81463 $\pm$ 302 events for W^+ and 54729 $\pm$ 249 for W^+[25], which both agree within 0.2% with the results of the parametrized $\not\!\!E_T$ fit.

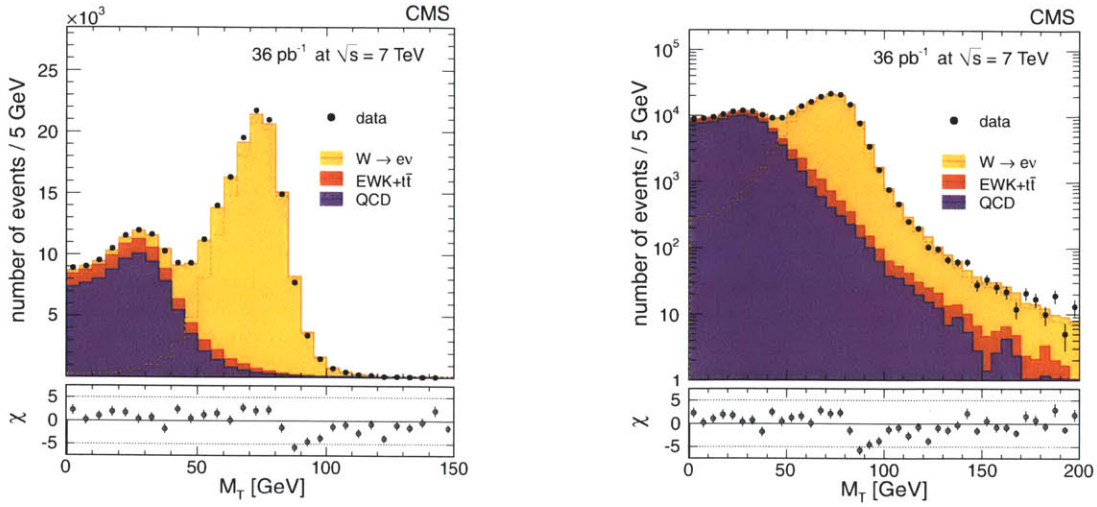

Figure 7-16: The M_T distribution for the inclusive W→$e\nu$-sample, in linear scale (left) and logarithmic scale (right).

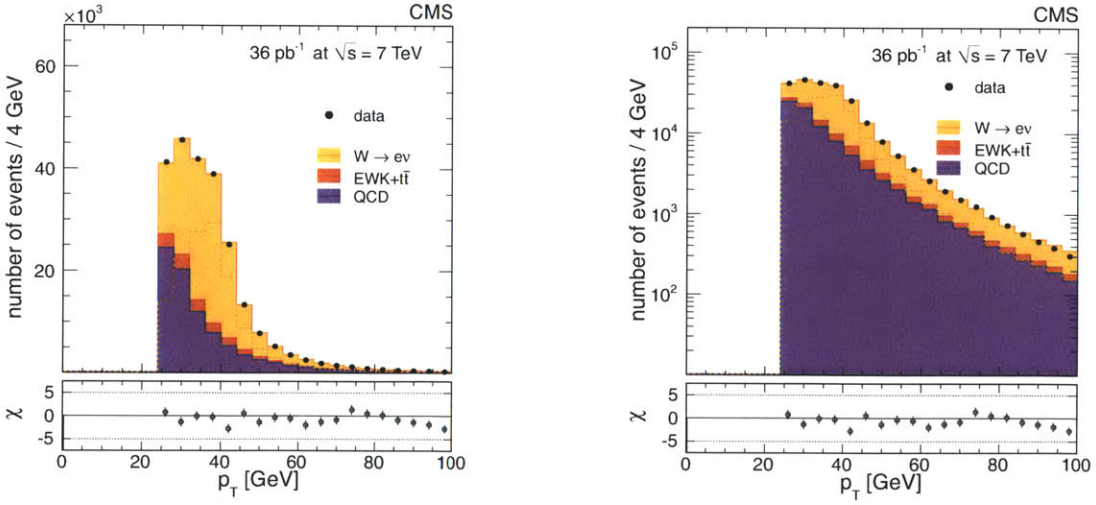

Figure 7-17: The p_T distribution for the inclusive W→$e\nu$-sample, in linear scale (left) and logarithmic scale (right).

Chapter 8

Results

Now that we determined all pieces of the puzzle, we can combine them using formula 1.5:

$$\sigma \mathrm{x} BR = \frac{N_{signal}}{A\epsilon \int \mathcal{L}dt},$$

where

- BR is the branching ratio, in our case the electron + neutrino decay mode of the W(10.75 $\pm$ 0.13 %[26]);
- N_{signal}=N_{total}-$N_{background}$. The electron selection was used to reduce e number of background events ($N_{background}$) as much as possible and then in section 7.7 an unbinned likelihood fit to the $\not\!\!E_T$ was used to select the number of signal events N_{signal} in the selected sample N_{total};
- A is the kinematic acceptance, calculated in chapter 5;
- $\epsilon = \epsilon_{MC} \times \alpha_{eff}$ is the selection efficiency, with ϵ_{MC} the MC efficiency and α_{eff} the efficiency scale factor calculated in chapter 6;
- $\int \mathcal{L}dt$ is the amount of data accumulated, 35.9 pb^{-1}.

The numbers to extract the final cross sections are given in table 8.1 and the summary of the experimental systematic uncertainties of the measurement is shown

in 8.2. The systematic uncertainty on the luminosity measurement is 4%[61]. The final results are:

$$\begin{aligned} \sigma(\mathrm{pp}\to\mathrm{WX})\times\mathrm{BR}(\mathrm{W}\to\mathrm{e}\nu) &= 10.45\pm0.03(\mathrm{stat.})\pm0.15(\mathrm{syst.})\pm0.09(\mathrm{th.})\pm0.42(\mathrm{lumi.})\mathrm{nb} \\ \text{NNLO prediction} &= 10.44\pm0.52\mathrm{nb}[25] \end{aligned} \tag{8.1}$$

$$\begin{aligned} \sigma(\mathrm{pp}\to\mathrm{W}^+\mathrm{X})\times\mathrm{BR}(\mathrm{W}^+\to\mathrm{e}^+\nu) &= 6.14\pm0.02(\mathrm{stat.})\pm0.10(\mathrm{syst.})\pm0.08(\mathrm{th.})\pm0.25(\mathrm{lumi.})\mathrm{nb} \\ \text{NNLO prediction} &= 6.15\pm0.29\mathrm{nb}[25] \end{aligned} \tag{8.2}$$

$$\begin{aligned} \sigma(\mathrm{pp}\to\mathrm{W}^-\mathrm{X})\times\mathrm{BR}(\mathrm{W}^-\to\mathrm{e}^-\nu) &= 4.32\pm0.02(\mathrm{stat.})\pm0.07(\mathrm{syst.})\pm0.07(\mathrm{th.})\pm0.17(\mathrm{lumi.})\mathrm{nb} \\ \text{NNLO prediction} &= 4.31\pm0.23\mathrm{nb}[25] \end{aligned} \tag{8.3}$$

$$\begin{aligned} \frac{\sigma(\mathrm{pp}\to\mathrm{W}^+\mathrm{X})\times\mathrm{BR}(\mathrm{W}^+\to\mathrm{e}^+\nu)}{\sigma(\mathrm{pp}\to\mathrm{W}^-\mathrm{X})\times\mathrm{BR}(\mathrm{W}^-\to\mathrm{e}^-\nu)} &= 1.424\pm0.008(\mathrm{stat.})\pm0.022(\mathrm{syst.})\pm0.029(\mathrm{th.}) \\ \text{NNLO prediction} &= 1.43\pm0.04[25] \end{aligned} \tag{8.4}$$

item	$W^+\to e^+\nu$	$W^-\to e^-\bar{\nu}$	$W\to e\nu$	W^+/W^-
Signal events	81568	54760	136328	1.490
Acceptance	0.502	0.481	0.493	1.044
MC Efficiency	0.760	0.769	0.764	0.988
Eff. scale factor	0.970	0.958	0.970	1.014
Luminosity	$35.9\mathrm{pb}^{-1}$	$35.9\mathrm{pb}^{-1}$	$35.9\mathrm{pb}^{-1}$	NA
Cross Sections	**6.14 nb**	**4.31 nb**	**10.45 nb**	**1.424**

Table 8.1: Summary of the results for W→eν cross section extraction.

Figure 8-1 shows the good agreement with previous experiments and NNLO perturbative QCD calculations. The comparison with NNLO perturbative QCD corrections already provided a stringent test of our current theoretical knowledge and the Standard Model passed with honors. The analysis is also a first step to other electroweak measurements as the differential W charge asymmetry[33], the W

Source	$W^+ \to e^+\nu$	$W^- \to e^-\bar{\nu}$	$W \to e\nu$	W^+/W^-
Lepton reco & ID	1.4%	1.4%	1.3%	1.5%
Momentum scale and resolution	0.5%	0.6%	0.5%	0.1%
MET scale and resolution	0.3%	0.3%	0.3%	0.1%
charge misidentification	0.10%	0.09%	/	0.15%
background subtraction	0.3%	0.5%	0.35%	0.4%
Total experimental	1.5%	1.6%	1.5%	1.6%
PDF uncertainties acceptance	0.8%	1.4%	0.6%	1.6%
Other theoretical uncertainties	1.0%	0.7%	0.7%	1.2%
Total theoretical	1.3%	1.6%	0.9%	2.0%
Total	**2.0%**	**2.3%**	**1.7%**	**2.5%**

Table 8.2: Summary of all the systematic uncertainties for W→$e\nu$ cross section.

polarization[35] and W+jets measurements. These measurements are extra tests for the electroweak theory and can also put constraints on parton distribution functions, as has already happened with the charge asymmetry measurement.[34] It is also a stepping stone to many new physics searches, like H→WW[36] and W′[37] searches.

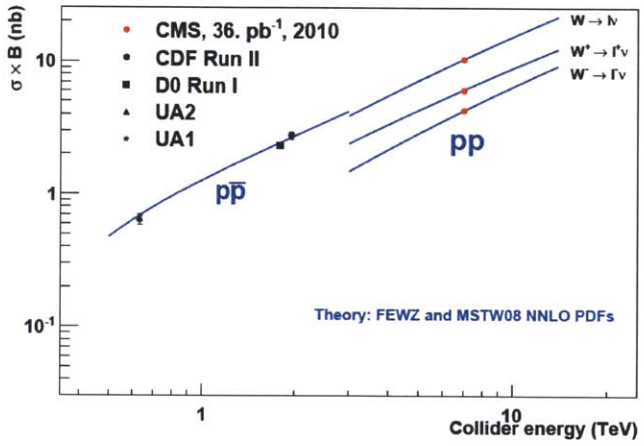

Figure 8-1: Measurements of the inclusive W cross section from CMS and lower-energy colliders. The blue lines indicate the theory prediction.

Chapter 9

Conclusions

The precision measurement of the inclusive W production cross-section and the charge asymmetry for W boson production in pp collisions at $\sqrt{s}$=7 TeV has been performed in the electron channel. Electrons were selected with roughly 80% efficiency using an optimized selection set of identification and isolation cuts. The efficiency of these cuts has been measured in the data itself, using the tag-and-probe method in $Z \rightarrow e^+e^-$ data. The final signal extraction also depends on the modeling of the transverse missing energy, the (lack of) signature that the neutrino leaves in our detector. To exploit this feature, a maximum likelihood fit to the $\not\!\!E_T$ spectrum is performed. The QCD was modeled using a parametrized function, whereas signal and Z/W backgrounds were modeled by simulation with a data-driven correction for the differences between data and simulation. Other uncertainties, connected to charge misidentification and electron energy scale and resolution, were estimated using fits to the clean $Z \rightarrow ee$ data sample. The theoretical uncertainties have been estimated using the most-advanced MC tools available: NNLO QCD, NLO EWK, NNLL resummation and QCD at orders higher than NNLO.

All the results obtained agree with Standard Model and previous experimental results. This measurement provided the foundation for understanding electrons and $\not\!\!E_T$ with the CMS detector, and established the $Z \rightarrow e^+e^-$ calibration analysis techniques. This improved understanding of the detector is a vital step on the road to the first CMS discoveries.

Appendix A

Charge-Specific Efficiency

In this appendix the measured efficiencies in data and MC are summarized for the electron and the positron separately. The reconstruction efficiency measurements are summarized in table A.1. The results for the identification and isolation efficiency can be found in table A.2 for the electrons and table A.3 for the positrons. Finally, the trigger efficiencies are tabulated in table A.4 for electrons and table A.5 for positrons.

A.1 Reconstruction efficiency

	MC-truth	MC TnP	Data TnP	Data/MC
Barrel (e^-)	0.9769 ± 0.0003	0.9782 ± 0.0003	0.9617 ± 0.0057	0.9831 ± 0.0058
Endcap (e^-)	0.9381 ± 0.0006	0.9450 ± 0.0007	0.9540 ± 0.0093	1.0095 ± 0.0099
All (e^-)	0.9656 ± 0.0003	0.9686 ± 0.0003	0.9607 ± 0.0078	0.9918 ± 0.0081
Barrel (e^+)	0.9767 ± 0.0003	0.9778 ± 0.0003	0.9780 ± 0.0054	1.0004 ± 0.0099
Endcap (e^+)	0.9438 ± 0.0007	0.9460 ± 0.0007	0.9492 ± 0.0082	1.0034 ± 0.0087
All (e^+)	0.9674 ± 0.0003	0.9689 ± 0.0003	0.9713 ± 0.0061	1.0025 ± 0.0063

Table A.1: Electron reconstruction efficiencies and scale factors for charge-specific efficiencies.

A.2 Electron identification and isolation

$25 < E_T < 35$	MC-truth	MC TnP	*BW*CB*	Data/MC	*MC*Gaus*	Data/MC
EE- (e^-)	0.6757 ± 0.0037	0.6786 ± 0.0038	0.6579 ± 0.0691	0.9694 ± 0.1020	0.6695 ± 0.0265	0.9865 ± 0.0394
EB- (e^-)	0.8097 ± 0.0020	0.8102 ± 0.0022	0.7671 ± 0.0202	0.9468 ± 0.0251	0.7598 ± 0.0159	0.9378 ± 0.0198
EB+ (e^-)	0.8073 ± 0.0020	0.8070 ± 0.0022	0.8031 ± 0.0452	0.9951 ± 0.0561	0.7951 ± 0.0172	0.9852 ± 0.0215
EE+ (e^-)	0.6921 ± 0.0036	0.6940 ± 0.0037	0.6355 ± 0.0293	0.9156 ± 0.0425	0.6562 ± 0.0250	0.9455 ± 0.0364
$35 < E_T < 45$	MC-truth	MC TnP	*BW*CB*	Data/MC	*MC*Gaus*	Data/MC
EE- (e^-)	0.7683 ± 0.0025	0.7699 ± 0.0029	0.7428 ± 0.0188	0.9648 ± 0.0247	0.7457 ± 0.0185	0.9686 ± 0.0243
EB- (e^-)	0.8872 ± 0.0012	0.8889 ± 0.0013	0.8481 ± 0.0097	0.9542 ± 0.0110	0.8561 ± 0.0095	0.9631 ± 0.0108
EB+ (e^-)	0.8840 ± 0.0012	0.8866 ± 0.0013	0.8566 ± 0.0096	0.9662 ± 0.0110	0.8633 ± 0.0095	0.9738 ± 0.0108
EE+ (e^-)	0.7741 ± 0.0025	0.7752 ± 0.0028	0.7609 ± 0.0184	0.9816 ± 0.0240	0.7602 ± 0.0183	0.9807 ± 0.0239
$E_T > 45$	MC-truth	MC TnP	*BW*CB*	Data/MC	*MC*Gaus*	Data/MC
EE- (e^-)	0.8188 ± 0.0030	0.8181 ± 0.0035	0.8109 ± 0.0226	0.9912 ± 0.0279	0.8002 ± 0.0225	0.9781 ± 0.0278
EB- (e^-)	0.9158 ± 0.0013	0.9165 ± 0.0014	0.8722 ± 0.0114	0.9517 ± 0.0126	0.8660 ± 0.0118	0.9449 ± 0.0129
EB+ (e^-)	0.9156 ± 0.0013	0.9180 ± 0.0014	0.8685 ± 0.0116	0.9461 ± 0.0128	0.8608 ± 0.0120	0.9378 ± 0.0131
EE+ (e^-)	0.8275 ± 0.0029	0.8279 ± 0.0034	0.8107 ± 0.0236	0.9793 ± 0.0288	0.8039 ± 0.0235	0.9711 ± 0.0287

Table A.2: Electron identification efficiencies for e^- with 4×3 η-p$_T$ binning.

$25 < E_T < 35$	MC-truth	MC TnP	*BW*CB*	Data/MC	*MC*Gaus*	Data/MC
EE- (e^+)	0.6792 ± 0.0037	0.6842 ± 0.0038	0.6063 ± 0.0338	0.8861 ± 0.0496	0.6261 ± 0.0278	0.9151 ± 0.0410
EB- (e^+)	0.8084 ± 0.0020	0.8093 ± 0.0022	0.7407 ± 0.0297	0.9153 ± 0.0367	0.7804 ± 0.0189	0.9642 ± 0.0235
EB+ (e^+)	0.8151 ± 0.0020	0.8145 ± 0.0022	0.7715 ± 0.0170	0.9473 ± 0.0211	0.7802 ± 0.0171	0.9579 ± 0.0211
EE+ (e^+)	0.6872 ± 0.0037	0.6948 ± 0.0038	0.6680 ± 0.0305	0.9615 ± 0.0443	0.6590 ± 0.0270	0.9485 ± 0.0392
$35 < E_T < 45$	MC-truth	MC TnP	*BW*CB*	Data/MC	*MC*Gaus*	Data/MC
EE- (e^+)	0.7672 ± 0.0026	0.7685 ± 0.0029	0.7250 ± 0.0187	0.9434 ± 0.0246	0.7264 ± 0.0185	0.9453 ± 0.0243
EB- (e^+)	0.8836 ± 0.0012	0.8851 ± 0.0013	0.8599 ± 0.0094	0.9715 ± 0.0107	0.8612 ± 0.0090	0.9730 ± 0.0103
EB+ (e^+)	0.8862 ± 0.0012	0.8879 ± 0.0013	0.8574 ± 0.0126	0.9657 ± 0.0143	0.8594 ± 0.0095	0.9680 ± 0.0108
EE+ (e^+)	0.7683 ± 0.0026	0.7726 ± 0.0029	0.7250 ± 0.0184	0.9384 ± 0.0241	0.7295 ± 0.0181	0.9442 ± 0.0237
$E_T > 45$	MC-truth	MC TnP	*BW*CB*	Data/MC	*MC*Gaus*	Data/MC
EE- (e^+)	0.8192 ± 0.0031	0.8179 ± 0.0035	0.8004 ± 0.0240	0.9787 ± 0.0297	0.7937 ± 0.0250	0.9704 ± 0.0309
EB- (e^+)	0.9129 ± 0.0013	0.9142 ± 0.0015	0.8823 ± 0.0119	0.9651 ± 0.0131	0.8784 ± 0.0111	0.9608 ± 0.0123
EB+ (e^+)	0.9164 ± 0.0013	0.9189 ± 0.0014	0.8844 ± 0.0115	0.9624 ± 0.0126	0.8789 ± 0.0116	0.9564 ± 0.0127
EE+ (e^+)	0.8231 ± 0.0030	0.8232 ± 0.0035	0.8523 ± 0.0219	1.0353 ± 0.0270	0.8355 ± 0.0220	1.0150 ± 0.0270

Table A.3: Electron identification efficiencies for e^+ with 4×3 η-p$_T$ binning.

A.3 Trigger efficiency

$25 < E_T < 30$	MC-truth	MC TnP	Data TnP	Data/MC
$-2.5 < \eta < -2.0$ (e^-)	0.9701 ± 0.0036	0.9671 ± 0.0040	0.9167 ± 0.0403	0.9478 ± 0.0418
$-2.0 < \eta < -1.5$ (e^-)	0.9728 ± 0.0037	0.9735 ± 0.0038	0.9545 ± 0.0393	0.9805 ± 0.0405
$-1.5 < \eta < -1.0$ (e^-)	0.9557 ± 0.0039	0.9656 ± 0.0037	0.9583 ± 0.0283	0.9925 ± 0.0296
$-1.0 < \eta < -0.5$ (e^-)	0.9620 ± 0.0031	0.9739 ± 0.0028	0.9792 ± 0.0187	1.0054 ± 0.0194
$-0.5 < \eta < 0$ (e^-)	0.9813 ± 0.0020	0.9845 ± 0.0020	0.9908 ± 0.0135	1.0064 ± 0.0138
$0 < \eta < 0.5$ (e^-)	0.9835 ± 0.0018	0.9873 ± 0.0019	0.9569 ± 0.0217	0.9692 ± 0.0221
$0.5 < \eta < 1.0$ (e^-)	0.9611 ± 0.0031	0.9698 ± 0.0031	0.9765 ± 0.0210	1.0069 ± 0.0219
$1.0 < \eta < 1.5$ (e^-)	0.9536 ± 0.0040	0.9658 ± 0.0036	0.9518 ± 0.0275	0.9855 ± 0.0287
$1.5 < \eta < 2.0$ (e^-)	0.9732 ± 0.0035	0.9766 ± 0.0035	0.9643 ± 0.0313	0.9874 ± 0.0323
$2.0 < \eta < 2.5$ (e^-)	0.9687 ± 0.0036	0.9688 ± 0.0037	0.9350 ± 0.0316	0.9660 ± 0.0328
$30 < E_T < 35$	MC-truth	MC TnP	Data TnP	Data/MC
$-2.5 < \eta < -2.0$ (e^-)	0.9623 ± 0.0032	0.9621 ± 0.0035	0.9677 ± 0.0222	1.0059 ± 0.0234
$-2.0 < \eta < -1.5$ (e^-)	0.9718 ± 0.0029	0.9729 ± 0.0031	0.9540 ± 0.0263	0.9806 ± 0.0272
$-1.5 < \eta < -1.0$ (e^-)	0.9536 ± 0.0030	0.9593 ± 0.0031	0.9741 ± 0.0179	1.0154 ± 0.0190
$-1.0 < \eta < -0.5$ (e^-)	0.9624 ± 0.0023	0.9703 ± 0.0023	0.9476 ± 0.0178	0.9767 ± 0.0184
$-0.5 < \eta < 0$ (e^-)	0.9808 ± 0.0016	0.9859 ± 0.0016	0.9804 ± 0.0137	0.9945 ± 0.0140
$0 < \eta < 0.5$ (e^-)	0.9817 ± 0.0015	0.9875 ± 0.0015	0.9540 ± 0.0177	0.9661 ± 0.0180
$0.5 < \eta < 1.0$ (e^-)	0.9608 ± 0.0023	0.9737 ± 0.0022	0.9747 ± 0.0148	1.0010 ± 0.0154
$1.0 < \eta < 1.5$ (e^-)	0.9529 ± 0.0030	0.9586 ± 0.0031	0.9449 ± 0.0227	0.9857 ± 0.0239
$1.5 < \eta < 2.0$ (e^-)	0.9816 ± 0.0023	0.9839 ± 0.0024	0.9167 ± 0.0311	0.9317 ± 0.0317
$2.0 < \eta < 2.5$ (e^-)	0.9616 ± 0.0032	0.9583 ± 0.0037	0.9894 ± 0.0156	1.0324 ± 0.0167
$35 < E_T < 40$	MC-truth	MC TnP	Data TnP	Data/MC
$-2.5 < \eta < -2.0$ (e^-)	0.9682 ± 0.0026	0.9684 ± 0.0029	0.9703 ± 0.0205	1.0019 ± 0.0214
$-2.0 < \eta < -1.5$ (e^-)	0.9739 ± 0.0023	0.9747 ± 0.0025	0.9725 ± 0.0190	0.9977 ± 0.0197
$-1.5 < \eta < -1.0$ (e^-)	0.9528 ± 0.0024	0.9560 ± 0.0027	0.9677 ± 0.0148	1.0123 ± 0.0157
$-1.0 < \eta < -0.5$ (e^-)	0.9578 ± 0.0020	0.9679 ± 0.0019	0.9671 ± 0.0128	0.9992 ± 0.0134
$-0.5 < \eta < 0$ (e^-)	0.9804 ± 0.0014	0.9882 ± 0.0012	0.9702 ± 0.0126	0.9818 ± 0.0128
$0 < \eta < 0.5$ (e^-)	0.9793 ± 0.0014	0.9863 ± 0.0013	0.9651 ± 0.0136	0.9785 ± 0.0138
$0.5 < \eta < 1.0$ (e^-)	0.9587 ± 0.0019	0.9683 ± 0.0019	0.9874 ± 0.0089	1.0197 ± 0.0094
$1.0 < \eta < 1.5$ (e^-)	0.9528 ± 0.0024	0.9561 ± 0.0026	0.9830 ± 0.0119	1.0281 ± 0.0128
$1.5 < \eta < 2.0$ (e^-)	0.9764 ± 0.0022	0.9765 ± 0.0025	0.9700 ± 0.0207	0.9934 ± 0.0213
$2.0 < \eta < 2.5$ (e^-)	0.9605 ± 0.0029	0.9613 ± 0.0032	0.9820 ± 0.0162	1.0215 ± 0.0172
$40 < E_T < 45$	MC-truth	MC TnP	Data TnP	Data/MC
$-2.5 < \eta < -2.0$ (e^-)	0.9724 ± 0.0022	0.9711 ± 0.0027	0.9928 ± 0.0107	1.0222 ± 0.0114
$-2.0 < \eta < -1.5$ (e^-)	0.9717 ± 0.0021	0.9739 ± 0.0024	0.9783 ± 0.0152	1.0045 ± 0.0157
$-1.5 < \eta < -1.0$ (e^-)	0.9516 ± 0.0022	0.9524 ± 0.0025	0.9817 ± 0.0108	1.0307 ± 0.0117
$-1.0 < \eta < -0.5$ (e^-)	0.9596 ± 0.0018	0.9653 ± 0.0019	0.9833 ± 0.0087	1.0186 ± 0.0092
$-0.5 < \eta < 0$ (e^-)	0.9793 ± 0.0013	0.9872 ± 0.0011	0.9891 ± 0.0077	1.0019 ± 0.0079
$0 < \eta < 0.5$ (e^-)	0.9817 ± 0.0012	0.9883 ± 0.0011	0.9652 ± 0.0120	0.9766 ± 0.0122
$0.5 < \eta < 1.0$ (e^-)	0.9605 ± 0.0018	0.9655 ± 0.0019	0.9740 ± 0.0110	1.0087 ± 0.0116
$1.0 < \eta < 1.5$ (e^-)	0.9516 ± 0.0022	0.9533 ± 0.0025	0.9667 ± 0.0140	1.0140 ± 0.0149
$1.5 < \eta < 2.0$ (e^-)	0.9777 ± 0.0019	0.9780 ± 0.0021	0.9660 ± 0.0173	0.9877 ± 0.0178
$2.0 < \eta < 2.5$ (e^-)	0.9641 ± 0.0025	0.9653 ± 0.0029	0.9769 ± 0.0161	1.0120 ± 0.0169
$45 < E_T < 50$	MC-truth	MC TnP	Data TnP	Data/MC
$-2.5 < \eta < -2.0$ (e^-)	0.9700 ± 0.0030	0.9719 ± 0.0035	1.0000 ± 0.0136	1.0289 ± 0.0144
$-2.0 < \eta < -1.5$ (e^-)	0.9732 ± 0.0025	0.9745 ± 0.0029	0.9524 ± 0.0239	0.9773 ± 0.0247
$-1.5 < \eta < -1.0$ (e^-)	0.9530 ± 0.0027	0.9571 ± 0.0030	0.9750 ± 0.0174	1.0187 ± 0.0184
$-1.0 < \eta < -0.5$ (e^-)	0.9604 ± 0.0022	0.9627 ± 0.0024	0.9706 ± 0.0150	1.0082 ± 0.0158
$-0.5 < \eta < 0$ (e^-)	0.9805 ± 0.0016	0.9874 ± 0.0014	0.9895 ± 0.0096	1.0021 ± 0.0098
$0 < \eta < 0.5$ (e^-)	0.9815 ± 0.0015	0.9867 ± 0.0014	0.9908 ± 0.0084	1.0042 ± 0.0087
$0.5 < \eta < 1.0$ (e^-)	0.9578 ± 0.0022	0.9621 ± 0.0024	0.9945 ± 0.0082	1.0336 ± 0.0089
$1.0 < \eta < 1.5$ (e^-)	0.9507 ± 0.0028	0.9527 ± 0.0031	0.9805 ± 0.0136	1.0292 ± 0.0147
$1.5 < \eta < 2.0$ (e^-)	0.9748 ± 0.0025	0.9751 ± 0.0029	0.9783 ± 0.0195	1.0032 ± 0.0202
$2.0 < \eta < 2.5$ (e^-)	0.9660 ± 0.0030	0.9674 ± 0.0035	1.0000 ± 0.0152	1.0337 ± 0.0161
$E_T > 50$	MC-truth	MC TnP	Data TnP	Data/MC
$-2.5 < \eta < -2.0$ (e^-)	0.9757 ± 0.0030	0.9766 ± 0.0036	0.9474 ± 0.0353	0.9700 ± 0.0363
$-2.0 < \eta < -1.5$ (e^-)	0.9684 ± 0.0029	0.9695 ± 0.0034	0.9753 ± 0.0220	1.0059 ± 0.0230
$-1.5 < \eta < -1.0$ (e^-)	0.9461 ± 0.0031	0.9495 ± 0.0034	0.9725 ± 0.0190	1.0241 ± 0.0204
$-1.0 < \eta < -0.5$ (e^-)	0.9588 ± 0.0024	0.9620 ± 0.0026	0.9888 ± 0.0102	1.0278 ± 0.0110
$-0.5 < \eta < 0$ (e^-)	0.9798 ± 0.0017	0.9859 ± 0.0016	0.9794 ± 0.0121	0.9934 ± 0.0124
$0 < \eta < 0.5$ (e^-)	0.9828 ± 0.0015	0.9877 ± 0.0015	0.9870 ± 0.0118	0.9994 ± 0.0120
$0.5 < \eta < 1.0$ (e^-)	0.9601 ± 0.0024	0.9660 ± 0.0024	0.9867 ± 0.0121	1.0214 ± 0.0128
$1.0 < \eta < 1.5$ (e^-)	0.9515 ± 0.0030	0.9504 ± 0.0034	0.9912 ± 0.0129	1.0429 ± 0.0141
$1.5 < \eta < 2.0$ (e^-)	0.9783 ± 0.0025	0.9780 ± 0.0029	0.9667 ± 0.0229	0.9884 ± 0.0236
$2.0 < \eta < 2.5$ (e^-)	0.9666 ± 0.0035	0.9664 ± 0.0042	0.9706 ± 0.0261	1.0043 ± 0.0273

Table A.4: Trigger efficiencies for e^- with 10×6 η-p$_T$ binning.

$25 < E_T < 30$	MC-truth	MC TnP	Data TnP	Data/MC
$-2.5 < \eta < -2.0$ (e^+)	0.9683 ± 0.0036	0.9703 ± 0.0037	0.9516 ± 0.0326	0.9807 ± 0.0338
$-2.0 < \eta < -1.5$ (e^+)	0.9738 ± 0.0036	0.9746 ± 0.0037	1.0000 ± 0.0203	1.0261 ± 0.0212
$-1.5 < \eta < -1.0$ (e^+)	0.9452 ± 0.0043	0.9619 ± 0.0039	0.9710 ± 0.0257	1.0094 ± 0.0270
$-1.0 < \eta < -0.5$ (e^+)	0.9558 ± 0.0033	0.9690 ± 0.0031	0.9875 ± 0.0182	1.0191 ± 0.0191
$-0.5 < \eta < 0$ (e^+)	0.9802 ± 0.0020	0.9840 ± 0.0020	0.9386 ± 0.0252	0.9539 ± 0.0256
$0 < \eta < 0.5$ (e^+)	0.9813 ± 0.0019	0.9855 ± 0.0019	0.9459 ± 0.0243	0.9599 ± 0.0247
$0.5 < \eta < 1.0$ (e^+)	0.9628 ± 0.0030	0.9734 ± 0.0029	0.9896 ± 0.0152	1.0167 ± 0.0159
$1.0 < \eta < 1.5$ (e^+)	0.9522 ± 0.0040	0.9646 ± 0.0037	0.9789 ± 0.0189	1.0149 ± 0.0200
$1.5 < \eta < 2.0$ (e^+)	0.9733 ± 0.0036	0.9779 ± 0.0035	0.9333 ± 0.0372	0.9544 ± 0.0382
$2.0 < \eta < 2.5$ (e^+)	0.9661 ± 0.0038	0.9662 ± 0.0040	0.9667 ± 0.0293	1.0005 ± 0.0306
$30 < E_T < 35$	MC-truth	MC TnP	Data TnP	Data/MC
$-2.5 < \eta < -2.0$ (e^+)	0.9706 ± 0.0029	0.9685 ± 0.0033	0.9459 ± 0.0306	0.9767 ± 0.0318
$-2.0 < \eta < -1.5$ (e^+)	0.9696 ± 0.0030	0.9713 ± 0.0032	1.0000 ± 0.0158	1.0295 ± 0.0166
$-1.5 < \eta < -1.0$ (e^+)	0.9487 ± 0.0032	0.9582 ± 0.0032	0.9921 ± 0.0117	1.0354 ± 0.0127
$-1.0 < \eta < -0.5$ (e^+)	0.9588 ± 0.0023	0.9697 ± 0.0023	0.9504 ± 0.0206	0.9801 ± 0.0213
$-0.5 < \eta < 0$ (e^+)	0.9830 ± 0.0015	0.9866 ± 0.0015	0.9809 ± 0.0134	0.9942 ± 0.0136
$0 < \eta < 0.5$ (e^+)	0.9819 ± 0.0015	0.9869 ± 0.0015	0.9721 ± 0.0143	0.9849 ± 0.0146
$0.5 < \eta < 1.0$ (e^+)	0.9626 ± 0.0022	0.9727 ± 0.0022	0.9693 ± 0.0157	0.9965 ± 0.0162
$1.0 < \eta < 1.5$ (e^+)	0.9548 ± 0.0030	0.9621 ± 0.0030	0.9913 ± 0.0128	1.0304 ± 0.0137
$1.5 < \eta < 2.0$ (e^+)	0.9755 ± 0.0027	0.9766 ± 0.0029	0.9667 ± 0.0229	0.9898 ± 0.0236
$2.0 < \eta < 2.5$ (e^+)	0.9644 ± 0.0032	0.9634 ± 0.0035	0.9691 ± 0.0213	1.0059 ± 0.0224
$35 < E_T < 40$	MC-truth	MC TnP	Data TnP	Data/MC
$-2.5 < \eta < -2.0$ (e^+)	0.9666 ± 0.0027	0.9650 ± 0.0031	0.9381 ± 0.0276	0.9722 ± 0.0288
$-2.0 < \eta < -1.5$ (e^+)	0.9755 ± 0.0022	0.9772 ± 0.0024	0.9910 ± 0.0132	1.0141 ± 0.0138
$-1.5 < \eta < -1.0$ (e^+)	0.9511 ± 0.0025	0.9528 ± 0.0027	0.9842 ± 0.0111	1.0329 ± 0.0120
$-1.0 < \eta < -0.5$ (e^+)	0.9585 ± 0.0019	0.9677 ± 0.0019	0.9876 ± 0.0088	1.0205 ± 0.0093
$-0.5 < \eta < 0$ (e^+)	0.9789 ± 0.0014	0.9881 ± 0.0012	0.9866 ± 0.0094	0.9985 ± 0.0096
$0 < \eta < 0.5$ (e^+)	0.9801 ± 0.0014	0.9868 ± 0.0013	0.9832 ± 0.0099	0.9963 ± 0.0101
$0.5 < \eta < 1.0$ (e^+)	0.9593 ± 0.0019	0.9683 ± 0.0019	0.9958 ± 0.0063	1.0284 ± 0.0068
$1.0 < \eta < 1.5$ (e^+)	0.9553 ± 0.0024	0.9570 ± 0.0026	0.9725 ± 0.0141	1.0162 ± 0.0150
$1.5 < \eta < 2.0$ (e^+)	0.9765 ± 0.0022	0.9774 ± 0.0024	0.9817 ± 0.0165	1.0043 ± 0.0171
$2.0 < \eta < 2.5$ (e^+)	0.9653 ± 0.0027	0.9656 ± 0.0031	0.9655 ± 0.0199	1.0000 ± 0.0209
$40 < E_T < 45$	MC-truth	MC TnP	Data TnP	Data/MC
$-2.5 < \eta < -2.0$ (e^+)	0.9690 ± 0.0024	0.9682 ± 0.0029	1.0000 ± 0.0098	1.0328 ± 0.0106
$-2.0 < \eta < -1.5$ (e^+)	0.9719 ± 0.0021	0.9707 ± 0.0025	0.9863 ± 0.0124	1.0161 ± 0.0131
$-1.5 < \eta < -1.0$ (e^+)	0.9509 ± 0.0022	0.9542 ± 0.0024	0.9833 ± 0.0098	1.0305 ± 0.0106
$-1.0 < \eta < -0.5$ (e^+)	0.9615 ± 0.0017	0.9666 ± 0.0018	0.9825 ± 0.0090	1.0165 ± 0.0095
$-0.5 < \eta < 0$ (e^+)	0.9814 ± 0.0012	0.9882 ± 0.0011	0.9858 ± 0.0084	0.9975 ± 0.0086
$0 < \eta < 0.5$ (e^+)	0.9802 ± 0.0012	0.9881 ± 0.0011	0.9842 ± 0.0082	0.9961 ± 0.0083
$0.5 < \eta < 1.0$ (e^+)	0.9617 ± 0.0017	0.9656 ± 0.0018	0.9853 ± 0.0087	1.0204 ± 0.0092
$1.0 < \eta < 1.5$ (e^+)	0.9562 ± 0.0021	0.9561 ± 0.0024	0.9682 ± 0.0134	1.0126 ± 0.0142
$1.5 < \eta < 2.0$ (e^+)	0.9743 ± 0.0020	0.9738 ± 0.0024	0.9790 ± 0.0146	1.0054 ± 0.0152
$2.0 < \eta < 2.5$ (e^+)	0.9710 ± 0.0023	0.9730 ± 0.0026	0.9844 ± 0.0141	1.0117 ± 0.0148
$45 < E_T < 50$	MC-truth	MC TnP	Data TnP	Data/MC
$-2.5 < \eta < -2.0$ (e^+)	0.9741 ± 0.0028	0.9744 ± 0.0034	0.9863 ± 0.0199	1.0122 ± 0.0207
$-2.0 < \eta < -1.5$ (e^+)	0.9723 ± 0.0026	0.9732 ± 0.0030	0.9889 ± 0.0162	1.0161 ± 0.0170
$-1.5 < \eta < -1.0$ (e^+)	0.9507 ± 0.0028	0.9500 ± 0.0032	0.9774 ± 0.0157	1.0288 ± 0.0169
$-1.0 < \eta < -0.5$ (e^+)	0.9572 ± 0.0023	0.9601 ± 0.0025	0.9805 ± 0.0115	1.0212 ± 0.0122
$-0.5 < \eta < 0$ (e^+)	0.9798 ± 0.0016	0.9864 ± 0.0015	0.9817 ± 0.0108	0.9952 ± 0.0111
$0 < \eta < 0.5$ (e^+)	0.9797 ± 0.0015	0.9853 ± 0.0015	0.9806 ± 0.0114	0.9952 ± 0.0117
$0.5 < \eta < 1.0$ (e^+)	0.9638 ± 0.0021	0.9671 ± 0.0022	0.9894 ± 0.0097	1.0231 ± 0.0103
$1.0 < \eta < 1.5$ (e^+)	0.9507 ± 0.0027	0.9503 ± 0.0031	0.9716 ± 0.0165	1.0224 ± 0.0177
$1.5 < \eta < 2.0$ (e^+)	0.9785 ± 0.0023	0.9769 ± 0.0028	0.9810 ± 0.0171	1.0041 ± 0.0178
$2.0 < \eta < 2.5$ (e^+)	0.9623 ± 0.0033	0.9639 ± 0.0039	0.9851 ± 0.0216	1.0219 ± 0.0228
$E_T > 50$	MC-truth	MC TnP	Data TnP	Data/MC
$-2.5 < \eta < -2.0$ (e^+)	0.9722 ± 0.0033	0.9682 ± 0.0043	0.9821 ± 0.0256	1.0144 ± 0.0269
$-2.0 < \eta < -1.5$ (e^+)	0.9699 ± 0.0030	0.9690 ± 0.0035	0.9714 ± 0.0253	1.0025 ± 0.0264
$-1.5 < \eta < -1.0$ (e^+)	0.9490 ± 0.0030	0.9491 ± 0.0034	1.0000 ± 0.0091	1.0536 ± 0.0104
$-1.0 < \eta < -0.5$ (e^+)	0.9564 ± 0.0025	0.9608 ± 0.0026	0.9825 ± 0.0123	1.0226 ± 0.0131
$-0.5 < \eta < 0$ (e^+)	0.9820 ± 0.0016	0.9878 ± 0.0015	0.9759 ± 0.0141	0.9880 ± 0.0144
$0 < \eta < 0.5$ (e^+)	0.9785 ± 0.0017	0.9864 ± 0.0015	0.9825 ± 0.0123	0.9960 ± 0.0126
$0.5 < \eta < 1.0$ (e^+)	0.9596 ± 0.0024	0.9634 ± 0.0025	0.9940 ± 0.0088	1.0318 ± 0.0095
$1.0 < \eta < 1.5$ (e^+)	0.9517 ± 0.0030	0.9549 ± 0.0033	0.9608 ± 0.0226	1.0062 ± 0.0239
$1.5 < \eta < 2.0$ (e^+)	0.9776 ± 0.0026	0.9774 ± 0.0031	0.9872 ± 0.0186	1.0100 ± 0.0193
$2.0 < \eta < 2.5$ (e^+)	0.9674 ± 0.0035	0.9672 ± 0.0043	0.9600 ± 0.0272	0.9926 ± 0.0285

Table A.5: Trigger efficiencies for e^+ with 10×6 η-p$_T$ binning.

Bibliography

[1] E. Fermi. Versuch einer Theorie der β-Strahlen. *Zeitschrift fr Physik*, 88:161177, 1934.

[2] T.D. Lee and C. N. Yang. Question of Parity Conservation in Weak Interactions. *Phys. Rev.*, 104:254258, 1956.

[3] C. S. Wu et al. Experimental Test of Parity Conservation in Beta Decay. *Phys. Rev.*, 105:1413, 1957.

[4] E. Sudarshan and R Marshak. Chirality invariance and the universal Fermi interaction. *Phys. Rev.*, 109:1860, 1958.

[5] R. P. Feynman and M. Gell-Mann. Theory of Fermi interaction. *Phys. Rev.*, 109:193, 1958.

[6] S.L. Glashow. Partial Symmetries of Weak Interactions. *Nucl.Phys.*, 22:579–588, 1961.

[7] A. Salam. Weak and electromagnetic interactions. In *Elementary particle theory. Relativistic groups and analyticity: Proceedings of the 8th Nobel Symposium*, page 367. Almqvist and Wiksell,Stockholm, 1968.

[8] S. Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:12641266, 1967.

[9] P. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 16:508, 1964.

[10] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 9:231, 1964.

[11] G. 't Hooft and M. Veltman. Regularization and Renormalization of Gauge Fields. *Nucl. Phys. B*, 44:189219, 1972.

[12] F.J. Hasert et al. Search For Elastic Muon-Neutrino Electron Scattering. *Phys. Lett. B*, 46:121, 1973.

[13] F.J. Hasert et al. Observation of Neutrino Like Interactions Without Muon or Electron in the Gargamelle Neutrino Experiment. *Phys. Lett. B*, 46:138, 1973.

[14] P. McIntyre C. Rubbia and D. Cline. Producing the massive intermediate vector meson with existing accelerators. In *Proc. Int. Neutrino Conf., Aachen*, page 683, 1976.

[15] S. van der Meer. Stochastic damping of betatron oscillations in the ISR. *Internal Report CERN ISR-PO/72-31*, 1972.

[16] G Arnison et al. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s}$ = 540-GeV. *Phys. Lett. B*, 122:103, 1983.

[17] G Arnison et al. Experimental Observation of Lepton Pairs of Invariant Mass Around 95-gev/c^2 at the CERN SPS Collider. *Phys. Lett. B*, 126:398, 1983.

[18] N. Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531, 1963.

[19] M. Kobayashi and K. Maskawa. CP Violation in the Renormalizable Theory of Weak Interactions. *Progr. Theor. Phys.*, 49:652, 1973.

[20] Pavel Nadolsky. Theory of W and Z boson production at the Tevatron and LHC. *AIP Conf.Proc.*, 753:158–170, 2005.

[21] Charalampos Anastasiou, Lance Dixon, Kirill Melnikov, and Frank Petriello. High-precision QCD at hadron colliders: electroweak gauge boson rapidity distributions at NNLO. *arXiv:hep-ph/0312266*, 2003.

[22] R. S. Thorne A. D. Martin, W. J. Stirling and G. Watt. Parton distributions for the LHC. *Eur. Phys. J. C*, 63:189, 2009.

[23] H. L. Lai et al. New parton distributions for collider physics. *Phys.Rev.D*, 82:074024, 2010.

[24] R. D. Ball et al. A first unbiased global NLO determination of parton distributions and their uncertainties. *Nucl. Phys. B*, 838:136, 2010.

[25] CMS colloboration. Measurements of inclusive W and Z cross sections in pp collisions at sqrt(s)=7 TeV. *CMS-PAS-EWK-10-005*, 2011.

[26] K. Nakamura et al. Review of Particle Physics. *J. Phys. G*, 37:075021, 2010.

[27] K. Melnikov and F. Petriello. The W boson production cross section at the LHC through $o(\alpha_s^2)$. *Phys. Rev. Lett*, 96:231803, 2006.

[28] C. Balazs and C.P. Yuan. Soft gluon effects on lepton pairs at hadron colliders. *Phys.Rev.D*, 56:5558–5583, 1997.

[29] C. M. Carloni Calame, G. Montagna, O. Nicrosini, and A. Vicini. Precision electroweak calculation of the production of a high transverse-momentum lepton pair at hadron colliders. *JHEP*, 10:109, 2007.

[30] The CDF collaboration. Measurements of inclusive W and Z cross sections in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV. *J. Phys. G: Nucl. Part. Phys.*, 34:24572544, 2007.

[31] The D0 Collaboration. Measurement of the Cross Section for W and Z Production to Electron Final States with the DO detector at $\sqrt{s}$ = 1.96 TeV. *D0Note 4403-CONF*, 2004.

[32] CMS colloboration. Measurements of Inclusive W and Z cross sections in pp collisions at $\sqrt{s}$=7 TeV. *JHEP*, 01:080, 2011.

[33] The CMS Collaboration. Measurement of the lepton charge asymmetry in inclusive W production in pp collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 04:50, 2011.

[34] G. Bozzi, J. Rojo, and A. Vicini. The impact of PDF uncertainties on the measurement of the W boson mass at the Tevatron and the LHC. *arXiv:hep-ph/1104.2056*, 2011.

[35] CMS collaboration. Measurement of the Polarization of W bosons with Large Transverse momenta in W+Jets Events at the LHC. *arXiv:hep-ex/1104.3829*, 2011.

[36] CMS colloboration. Measurement of W^+W^- Production and Search for the Higgs Boson in pp Collisions at $\sqrt{s} = 7$ TeV. *Phys.Lett. B.*, 699:25, 2011.

[37] CMS colloboration. Search for a heavy gauge boson W' in the final state with an electron and large missing transverse energy in pp collisions at $\sqrt{s} = 7$ TeV. *Phys.Lett. B.*, 698:21, 2011.

[38] L. Evans and P. Bryant. The LHC machine. *JINST*, 3:S08001, 2008.

[39] S. Myers and E. Picasso. The design, construction and commissioning of the CERN Large Electron-Positron collider. *Contemporary Physics*, 31:387–403, 1990.

[40] The ATLAS collaboration. ATLAS detector and physics performance. *CERN-LHCC-99-014*, 1999.

[41] The CMS collaboration. CMS Physics Technical Design Report. *CERN-LHCC-2006-001 and CERN-LHCC-2006-021*, 2006.

[42] The ALICE collaboration. ALICE Technical Design Report. *CERN/LHCC 2001-021*, 2001.

[43] The LHCb collaboration. LHCb : Technical Proposal. *CERN-LHCC-98-004*, 1998.

[44] The CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[45] The CMS Colloboration. The CMS tracker system project : Technical Design Report. *CERN-LHCC-98-006*, 1998.

[46] The CMS Colloboration. The CMS tracker : addendum to the Technical Design Report. *CERN-LHCC-2000-016*, 2000.

[47] The CMS Colloboration. The CMS electromagnetic calorimeter project : Technical Design Report. *CERN-LHCC-97-033*, 1997.

[48] The CMS Colloboration. The CMS hadron calorimeter project : Technical Design Report. *CERN-LHCC-97-031*, 1997.

[49] The CMS Colloboration. The CMS magnet project : Technical Design Report. *CERN-LHCC-97-010*, 1997.

[50] The CMS Colloboration. The CMS muon project : Technical Design Report. *CERN-LHCC-97-032*, 1997.

[51] R. Loos et al. Cms ECAL Preshower and Endcap Engineering Design Review. v.2- Preshower. *CMS-2000-054-MEETING, CERN-ECAL-EDR-4*, 2000.

[52] P. Adzic et al. Energy resolution of the barrel of the CMS electromagnetic calorimeter. *JINST*, 2:P04004, 2007.

[53] The CMS collaboration. Electromagnetic calorimeter calibration with 7 TeV data. *CMS-PAS-EGM-10-003*, 2010.

[54] G. Daskalakis. CMS ECAL Calibration Strategy. *CALORIMETRY IN HIGH ENERGY PHYSICS: XII International Conference. AIP Conference Proceedings*, 867:400–407, 2006.

[55] L. Y. Zhang X.D. Qu and R. Y. Zhu. Radiation Induced Color Centers and Light Monitoring for Lead Tungstate Crystals. *Proc. IEEE Trans. Nucl. Sci.*, 47:1741–47, 2000.

[56] CMS collaboration. Missing Transverse Energy Performance in Minimum-Bias and Jet Events from pp Collisions at $\sqrt{s}$=7 TeV. *CMS-PAS-JME-10-004*, 2010.

[57] CMS colloboration. Jet Performance in pp Collisions at $\sqrt{s}$=7 TeV. *CMS-PAS-JME-10-003*, 2010.

[58] The CMS collaboration. Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp collisions at 7 TeV. *CMS-PAS-PFT-10-002*, 2010.

[59] The CMS collaboration. CMS TriDAS project : Technical Design Report. *CERN-LHCC-2000-038*, 2000.

[60] CMS collaboration. Measurement of CMS Luminosity. *CMS-PAS-EWK-10-004*, 2010.

[61] M. Zanetti et al. Beams scan based Absolute Normalization of the CMS Luminosity Measurement. CERN-Proceedings-2011-001, 2011.

[62] D. Belorhad et al. Commissioning and First Performance of the LHC Beam Current Measurement Systems. *Proceedings of the International Particle Accelerator Conference*, page MOPE059, 2010.

[63] S. Baffioni et al. Electron reconstruction in CMS. *Eur. Phys. J. C*, 49:1099, 2007.

[64] W. Adam et al. Reconstruction of electrons with the Gaussian-Sum Filter in the CMS tracker at the LHC. *J. Phys. G: Nucl. Part. Phys.*, 31:N9–N20, 2005.

[65] The CMS collaboration. CMS Tracking Performance Results from early LHC Operation. *Eur.Phys.J.C*, 70:1165–1192, 2010.

[66] H. Bethe and W. Heitler. On the Stopping of fast particles and on the creation of positive electrons. *Proc. Royal Soc. (London) A*, 146:83, 1934.

[67] R. Fruhwirth. Application of Kalman filtering to track and vertex fitting. *NIMA*, 262:444, 1987.

[68] The CMS collaboration. Electron reconstruction and identification at $\sqrt{s} = 7$ TeV. *CMS-PAS-EGM-10-004*, 2010.

[69] T. Sjostrand et al. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006.

[70] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO vector-boson production matched with shower in POWHEG. *arXiv:hep-ph/0805.4802*, 2008.

[71] Y. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. *Sov.Phys.JETP*, 46:641, 1977.

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

[73] G. Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl.Phys.B*, 126:298, 1977.

[74] S. Frixione and B.R. Webber. Matching NLO QCD computations and parton shower simulations. *JHEP*, 0206:029, 2002.

[75] J.C.Collins, D. E. Soper, and G. Sterman. Transverse Momentum Distribution in Drell-Yan Pair and W and Z Boson Production. *Nucl.Phys. B*, 250:199, 1985.

[76] E. Barberio, B. van Eijk, and Z. Was. Photos a universal Monte Carlo for QED radiative corrections in decays. *Comput. Phys. Commun.*, 66:115, 1991.

[77] E. Nurse. W and Z properties at the Tevatron. *arXiv:hep-ex/0808.0218*, 2008.

[78] N. Adam, V. Halyo, S Yost, and W. Zhu. Evaluation of the Theoretical Uncertainties in the W to Lepton and Neutrino Cross Sections at the LHC. *JHEP*, 0809:133, 2008.

[79] HepData website. Parton Distribution Functions. ***http://durpdg.dur.ac.uk/HEPDATA/PDF***.

[80] PDF4LHC Working Group. PDF4LHC Recommendations. *http://www.hep.ucl.ac.uk/pdf4lhc/PDF4LHCrecom.pdf*, 2010.

[81] S. Agostinelli et al. Geant 4a simulation toolkit. *Nuclear Instruments and Methods in Physics Research A*, 506:250–303, 2003.

[82] F. James and M. Roos. MINUIT-a system for function minimization and analysis of the parameter errors and correlations. *Comput. Phys. Commun.*, 10:343, 1975.

[83] P. Harris et al. Modeling of W→lν MET with Boson Recoil. *CMS-AN-2010/332*, 2010.

[84] TeVI Group. Design report Tevatron 1 project. *FERMILAB-DESIGN-1984-01*, 1984.

[85] A. Sherstnev and R.S. Thorne. Parton Distributions for LO Generators. *Eur.Phys.J.C*, 55:553–575, 2008.