JETS PRODUCED IN ASSOCIATION WITH Z-BOSONS IN CMS AT THE LHC

by

CHRISTOS LAZARIDIS

A dissertation submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY
(PHYSICS)

at the

UNIVERSITY OF WISCONSIN – MADISON

2011

JETS PRODUCED IN ASSOCIATION WITH Z-BOSONS IN CMS AT THE LHC

Christos Lazaridis

Under the supervision of Professor Wesley H. Smith

At the University of Wisconsin — Madison

A study of jet production in association with a Z boson in proton-proton collisions at a 7 TeV center-of-mass energy is presented, using data collected with the CMS detector during the 2010 LHC collisions run, corresponding to an integrated luminosity of 36.1 pb^{-1}. The measured jet multiplicity distributions are corrected for efficiency and unfolded for detector effects to be directly comparable with theoretical predictions. The leading reconstructed jet transverse energy distributions are measured, unfolded and compared with theoretical predictions. Ratios of $Z+\geq n-jets/Z+\geq (n-1)-jets$ and $Z+\geq n-jets/Z+\geq 0-jets$ versus the exclusive n reconstructed jets are also presented.

To my parents and brother

for their love, support and (a lot of) patience

and to all the friendships that lasted.

“Are you a physicist?”

“Me? I don’t know anything about science!”

“Marvellous! Ideal qualification!”

Johnny and the Dead

TERRY PRATCHETT

ACKNOWLEDGMENTS

This thesis would not have been possible without the support of many wonderful people throughout these years. First and foremost I would like to express my deep gratitude to professors Wesley Smith and Sridhara Dasu. Professor Smith was the first faculty member I talked to in UW-Madison and his deep enthusiasm about the field was a major deciding factor in the beginning and a constant source of encouragement at every step of the way. Professor Dasu's remarkable guidance and advice on physics analysis provided valuable assistance from my early steps till the full completion of my studies.

My professors at the Aristotle University of Thessaloniki Christos Eleftheriadis and Georgios Voutsas were the ones who gave me a lot of guidance during my undergraduate studies and the much needed first push to make the move and apply abroad for PhD studies.

During my presence at CERN, scientist Pamela Klabbers' guidance in learning the secrets of the L1 trigger operation was extremely helpful and she was always of assistance in any issues might come up. Scientist Monika Grothe's advice in performing data analysis proved beneficial in many occasions and helped discover and solve many analysis-related glitches. Bryan Dahmes' helping hand on building the code from scratch, the long discussions on physics and analysis strategies and of course our basketball games will never be forgotten. I would also like to thank scientist Sacha Savin for his suggestions and Ajit, Dan and Will from the UW computing staff for providing immediate assistance whenever computing problems came up. Also, Marc Magrans for his programming advice and help during my involvement with the CMS L1 trigger software.

Special thanks goes to my classmates Kira Grogg and Marc Weinberg. Through our common path from Madison to 904 and to 32-3-B16, you guys were an endless source of ideas and support. Michalis Bachtis' understanding of physics but most of all his enthusiastic temperament made life a lot more interesting at the office. I would also like to thank fellow students Ian, Mike, Jeff, Jessica, Josh and Lindsey for their suggestions and friendship.

My stay at CERN would have been a lot more boring without the amazing community of noisy Greeks around. "$E\upsilon\chi\alpha\rho\iota\sigma\tau\acute{\omega}$!" goes to Fanouria and Theodoros (with the recent addition on Panagiotis) for their friendship and support in all the ups and (especially) the downs; Konstantinos for his persistence in making me a better person and for being a true and long time friend; Eleni for all the good times and constant reminders. Alex for the afternoon coffee breaks and linux conversations; Loukas for his unique way of describing things and situations; Eirini for the looooong discussions; my friends from the Aristotle University of Thessaloniki: Giannis for all the guitar nights, Vasoula for listening to me when needed and Antreas for just being cool; Kostas for the cooking ideas; Manos for the movie nights and excursions; and Ilektra for being the postdoc showing the way ahead.

My two-and-a-half-year presence in Madison was rendered a lot better thanks to the great people I met there. My roommates and good friends Haris and Aris; Nicos and our the end-of-the-week coffee soires; Loizos with his unconventional view of reality and all the anime knowledge; Mina my compatriot and Vasilis and Liza and the opportunity they gave me to visit Puerto Rico. Last but definitely not least, my dear Indian friends Danny and Vikram who provided me with a whole new perspective to life in general (albeit a different one each).

Friends back in Greece were a valuable asset reminding me of home and keeping me on the right track from afar when needed. I owe much to my long-time high school friends Maria and Anna-Maria and the long conversations we held via phone or Skype. Vlasis for paving the way, being a constant inspiration to push myself and to be sincere, too many things to put it all down on paper. Christos, Savvas and Makis for proving that friendships can withstand the test of distance. And naturally, my Athenian friends Sofia for all the psychological analyses we did, Eleni for bringing smiles and Gitsa for the Greek movies and the cheesecake (want more!).

Finally, I would like to thank my parents, Giannis and Maria, and my brother Nikos for their unconditional love and support in my endeavors all these years.

Contents

List of Tables

List of Figures

When you have dived off a cliff, your only hope is to press for the abolition of gravity.

Unseen Academicals
TERRY PRATCHETT

Chapter 1

Introduction

The road to knowledge has no obvious path to tread on. It is in human nature to ask questions and look for possible answers; answers that rarely come easy. Starting with the ancient Greeks, a philosophical attempt was initially made to provide interpretations to topics of concern. Mathematics and geometry, particularly with the contributions of Pythagoras and Euclid, laid the foundations of describing the cosmos in a more coherent and methodical manner. It took centuries however to overcome long-standing dogmas and move on from the initial assumption of Democritus that matter constitutes of indivisible particles which he called $\acute{\alpha}\tau o\mu\alpha$, to Isaac Newton and Galileo Galilei in the 1600s to reach a more scientific way of thinking, in the sense we perceive it today.

The field of particle physics provides a description of the universe in its most fundamental level by describing the basic building blocks that matter is made of, and the ways they interact. The prevailing theory accepted today is the so-called Standard Model (SM), which has been developed throughout the 20th century and is described in detail on the next section. Many of its breakthroughs were possible only after the 1940s, when a series of advancements in accelerator technology provided fully controllable particle beams allowing physicists to probe previously uncharted territories[1]. Today's state-of-the-art particle accelerators, like the Large Hadron Collider at CERN and the Tevatron at Fermilab are enabling the scientific community to look deeper in the sub-atomic scale and cross-check theoretical predictions.

1.1 The Standard Model

1.1.1 A historical approach

The Standard Model (SM) [3, 4, 5] is the most comprehensive and widely-accepted theory available. It contains 12 fundamental *fermions* (and their respective anti-particles) which are the constituents of matter, and describes the strong, weak and electromagnetic interactions between them using mediating particles called *gauge bosons*. The only predicted SM particle not yet observed experimentally is the hypothetical Higgs boson, devised to explain the process by which particles acquire mass. The search for the Higgs boson is the main goal of the two general-purpose detectors, ATLAS and CMS, at the Large Hadron Collider.

The history of modern particle physics starts in 1897 with the discovery of the electron by J.J. Thompson and his famous *plum pudding model*, according to which the atom is composed of electrons surrounded by a mixture of positive charges to balance the negative charges of the electrons. Rutherford's scattering experiment in 1909 exposed the model's weaknesses; the bulk of the atomic mass was concentrated on a small, heavy core at the

---

[1] Betatrons [1] invented in the 1940s accelerated electrons inside a vacuum around a circular path of increasing radius by means of an increasing magnetic field; synchrotrons [2] went beyond the betatron limit (about 300 MeV); here the particles are accelerated in a ring of constant radius by using an increasing electric field which raises particle energy and an increasing magnetic field, keeping them on the same circular track.

center of the atom. Five years later, the Bohr model was the first atomic model that provided explanations to existing problems (the Rydberg formula for the spectral emission lines of atomic hydrogen, for example) and appeared to be a step in the right direction. This early "era" closed after the discovery of the proton (Rutherford, 1919) and the neutron (Chadwick, 1932), by providing a full description of the atomic structure.

It came naturally after these discoveries to wonder what holds the nucleus in place as it contains several tightly-packed positive charged protons. Yukawa assumed the existence of a force field binding the particles together, and a mediator particle being the carrier of that force. Cosmic ray experiments revealed the existence of the *pion*, initially assumed to be Yukawa's force carrier, as well as one more new particle, the *muon*.

Elastic electron–nucleus scattering experiments gave hints that there might be further internal structure inside the nucleus. In addition, the pion turned out to be the ground state in a sequence of further excited states. In an attempt to organize these and many more particles that had been discovered in the meantime, mainly through bubble chamber photographs, in 1964 Gell-Mann and Zweig independently proposed that the nucleons, belonging in a broader family of particles called *baryons*, are made of three spin-1/2 particles they named quarks, while the *mesons* (particles belonging to the same family as the pion) are made of quark–anti-quark combinations. All baryon and meson combinations are also called called hadrons. The quark model was not immediately accepted, as a series of experiments attempting to observe individual quarks failed, giving rise to the idea of *quark confinement*, meaning that quarks cannot be isolated but exist only in bound states. Another objection was that the quark model seemed to violate Pauli's exclusion principle by assuming that certain bound states were supposed to consist by three identical quarks. To overcome this problem, an extra degree of freedom called *color* was assigned to the quarks so that each one carries one out of three different colors in such bound states. The particles responsible for keeping the quarks together inside baryons and mesons were called *gluons*. The discovery of the J/ψ meson and the fact that it fitted well in the quark model along with more experimental observations in agreement with it, gave the model more credibility.

In the meantime, inconsistencies with the energy and momentum conservation in beta decay experiments in the 1930s led physicists to assume the existence of one more particle, the *neutrino*. Experimental verification of its existence came in 1956 by Cowan and Reines. The actual beta decay mechanism was explained by the assumption that this weak interaction was mediated by the exchange of some particles. The theoretical framework predicting these intermediate *vector bosons* was introduced in 1968 by Sheldon Glashow, Abdus Salam and Steven Weinberg (GSW) who provided a unified description of the electromagnetic and weak forces. Experimentally, the existence of the electroweak interactions was verified in two stages, both of them at CERN; first came the discovery of neutral currents in neutrino scattering by the Gargamelle collaboration in 1973, and second the discovery of the W and the Z gauge bosons in proton-antiproton collisions at the Super Proton Synchrotron accelerator by the UA1 and UA2 collaborations in 1983.

1.1.2 Quarks and leptons

Summing up, fundamental matter particles in the Standard Model are classified as quarks (Table 1.1) or leptons (Table 1.2), grouped in three generations of increasing mass. Particles of higher generations decay via weak interactions to particles of the first generation, explaining why everyday matter comprises of first generation particles only. All matter particles have spin 1/2 and the u, c, t quarks carry a +2/3 electrical charge, the d, s, b quarks

Quarks			
up (u)	$[1.7 - 3.3\,\text{MeV}]$	down (d)	$[4.1 - 5.8\,\text{MeV}]$
charm (c)	$[1.27^{+0.07}_{-0.09}\,\text{GeV}]$	strange (s)	$[101^{+29}_{-21}\,\text{MeV}]$
top (t)	$[172 \pm 1.6\,\text{GeV}]$	bottom (b)	$[4.19^{+0.18}_{-0.06}\,\text{GeV}]$

Table 1.1: Quarks overview. Numbers in brackets correspond to the mass of each quark[7].

Leptons			
electron (e)	$[0.511\,\text{MeV}]$	electron neutrino (ν_e)	$[< 2\,\text{eV}]$
muon (μ)	$[105.6\,\text{MeV}]$	muon neutrino (ν_μ)	$[< 2\,\text{eV}]$
tau (τ)	$[1776.8\,\text{MeV}]$	tau neutrino (ν_τ)	$[< 2\,\text{eV}]$

Table 1.2: Leptons overview. Numbers in brackets correspond to the mass of each lepton[7].

carry a -1/3 charge while the e, μ, τ leptons carry a +1 electrical charge. Neutrinos have zero charge and were initially believed to be massless. However, recent observations from *neutrino oscillation* experiments [6] point to neutrinos having nonzero mass. The respective antiparticles of all 12 particles also exist, carrying opposite charges.

1.1.3 Fundamental forces

Every fundamental interaction is associated with spin-1 mediator particles (table 1.3). In order of decreasing force strength these are: the gluon for the strong force, the photon for the electromagnetic force, the two W's and the Z for the weak force and the postulated graviton for gravity[2]. Figure 1.1 provides a visual overview of the interactions between particles in the Standard Model.

---

[2] The gravitational force has not been successfully added to the SM yet, but at this scale it is so insignificant (25 orders of magnitude smaller than the weak force) that has virtually no effect on experimental observations.

Interaction	Relative Magnitude	Mediator boson	Charge	Mass [7]
Strong	1	8 colored gluons	0	0
Electromagnetic	10^{-2}	γ	0	$< 1 \times 10^{-18}\,\mathrm{eV}$
Weak	10^{-7}	W^{+}	+1	$80.4\,\mathrm{GeV}$
	10^{-7}	W^{-}	-1	$80.4\,\mathrm{GeV}$
	10^{-7}	Z	0	$91.1\,\mathrm{GeV}$

Table 1.3: Overview of the fundamental interactions and mediator particles in the Standard model.

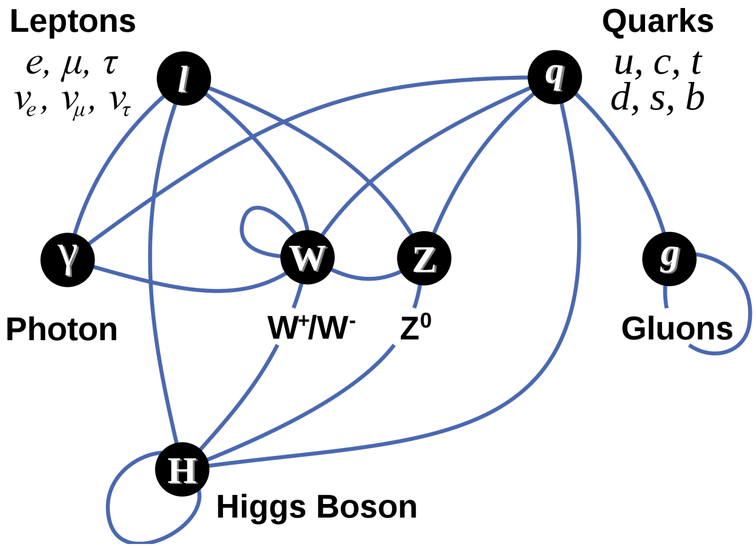


Figure 1.1: Summary of interactions between particles described by the Standard Model [8].

1.2 Creating and observing Z+Jets events

In order to study interesting processes at the Large Hadron Collider, two opposing, intense proton beams are created and intersect at four points where they cross. Each beam crossing produces proton-proton collisions and giant detectors have been built in each in-

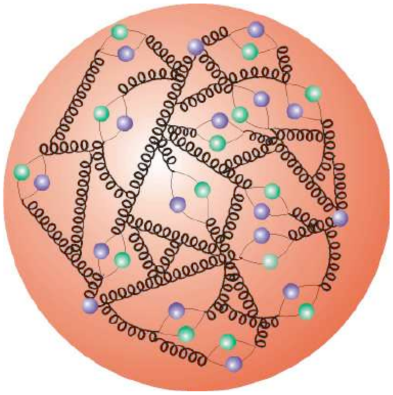


Figure 1.2: Proton structure showing gluons (spirals), *sea* quark-antiquark pairs and valence quarks.

tersection to collect information about the collision products. This information is then recombined and can give us a good picture of what happened in a proton-proton collision event.

1.2.1 Proton structure

Protons are composite particles, made of two up and a down quark, held together by the strong force, which is mediated by gluons in their simplest description. However, when a collision occurs, the real picture is more complicated; gluons inside the protons can momentarily split, giving us short-lived virtual quark-antiquark pairs, known as *sea* quarks (the initial up and down quarks are called *valence* quarks). This is shown pictorially on Figure 1.2. All particles inside a proton (real or virtual) are collectively called *partons*.

The complex picture inside the colliding protons is described by models developed by studying experimental data. *Parton distribution functions* are defined as the probability density for finding inside the proton a particle of certain longitudinal momentum fraction at a given momentum transfer.

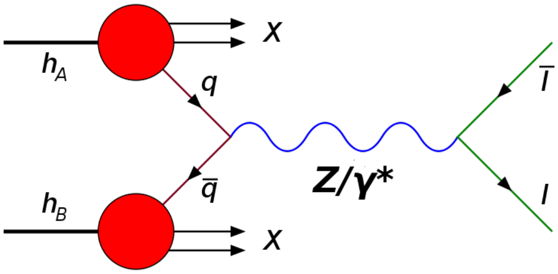


Figure 1.3: Drell-Yan process in proton-proton scattering. The incoming protons can be seen coming in from the left (drawn as blobs). An interaction occurs between a quark from one proton and an antiquark from the other creating a Z boson which then decays into a pair of leptons.

Due to the intrinsic proton structure, when two protons collide, essentially the strong collision happens only between two of the partons, as shown in Figure 1.3.

1.2.2 Z boson production

The main production mechanism in proton-proton collisions is the so-called Drell-Yan [9] process. This happens when a quark from one proton and an antiquark from another annihilate and create a Z boson which then decays into a pair of opposite-charged leptons. An example of this process can be seen on Figure 1.3. The most common Z boson production channels in the Large Hadron Collider are discussed with more detail in section 2.1.

The lepton pair (electrons in this study) are identified by the detector and by measuring their energies and momenta, we can reconstruct the mass of the originating particle:

$$M_{ee} = \sqrt{(E_1 + E_2)^2 + (\vec{p}_1 + \vec{p}_2)^2}. \tag{1.1}$$

1.2.3 Jet production

Along with the Z boson, outgoing partons can be produced; for example, in the Drell-Yan case in Figure 1.3 one (or both) of the quarks participating in the strong interaction can radiate a gluon. Due to color confinement, these outgoing partons cannot exist individually; they go through a process called *hadronization*, as a result of which they turn into combinations of colorless hadrons. The number of hadrons produced by the hadronization of a single initial parton is very large and the tight cone of outgoing particles that is formed, is which is what we call a jet.

1.3 The study of jets produced in association with Z bosons

The topic of this thesis, associated production of jet events with a Z boson decaying to electrons, is very large at the 7 TeV collision energy of CERN's Large Hadron Collider. The clean experimental signature of two high-energy leptons originating from the Z boson decay and the production of associated jets is a very useful observable for many purposes: detector calibration, tuning the Monte Carlo generators used for event simulation and as a test to pQCD calculations at this new energy regime. In addition, this process constitutes an important background to many new physics searches, such as certain Higgs decay modes.

> *The sun is simple. A sword is simple.*
> *A storm is simple. Behind everything*
> *simple is a huge tail of complicated.*
>
> I Shall Wear Midnight
> TERRY PRATCHETT

Chapter 2

Z+Jets events

2.1 Z boson production and decays

At proton-proton collisions, the main Z production channel is the Drell-Yan [9] process during which a quark and a antiquark annihilate to create a pair of leptons through the exchange of a virtual photon or Z boson: $q\bar{q} \rightarrow Z/\gamma^{*} \rightarrow l^{+}l^{-} + X$. A tree-level and a higher order Feynman diagram for the Drell-Yan process are shown in Figure 2.1. Subleading processes contributing to Z production are quark-gluon scattering $qg \rightarrow qZ$ and gluon-gluon fusion $gg \rightarrow q\bar{q}Z$.

To calculate the Z inclusive cross section, we need to take into account the intrinsic proton structure. Parton Distribution Functions (PDFs) describe the probability density to find a particle of certain longitudinal momentum fraction x at momentum transfer Q^2. Due to the perturbative QCD limitations, they are determined experimentally either from deep inelastic lepton-nucleon scattering experiments (like ZEUS and H1 at HERA) or from other hard scattering processes (eg. from Tevatron data). Figure 2.2 shows the PDF distributions at Q^2=10000 GeV^2. Combined with the LHC kinematics plane plot in Figure 2.3 we can deduce that over the measurable rapidity range $(|\eta| < 2.4)$ gluons are the dominating partons, meaning that the scattering mostly happens between *sea* quarks generated by the $g \rightarrow q\bar{q}$ splitting process.

Convoluting the PDFs with the generic process $p_i p_j \rightarrow Z + X$ where p_i, p_j are the interacting partons of the proton and summing over all partons, the cross section can be calculated by:

$$\sigma_Z = \sum_{i,j} \sigma_{p_i p_j \rightarrow Z+X} \otimes PDF\left(\chi_i, \chi_j, Q^2\right) \tag{2.1}$$

where χ_i, χ_j the proton momentum fraction that p_i, p_j are carrying. The W/Z production cross section for leptonic decays as a function of the collider energy is reported in Figure 2.4 along with experimental observations from previous experiments. The Z/γ^* inclusive production cross section as computed by the FEWZ [10] package at NNLO is 3048 ± 132 pb if the mass of the boson is restricted to be above 50 GeV and only leptonic decay channels are considered.

The improvement in precision we get when studying Z production at leading versus higher orders is demonstrated in Figure 2.5 where the Z rapidity distributions are shown at leading order, next-to-leading order and next-to-next-to-leading order (LO, NLO and NNLO respectively) with the respective error bands. Figure 2.6 shows the predicted Z rapidity distribution at three LHC collision energies ($\sqrt{s}$ =7, 10 and 14 TeV) at LO and NLO using three different PDFs.

The Z boson prefers to decay hadronically, with almost 70% of the decays being to strong-interacting particles. “Invisible” decays to neutrinos account for 20% of the decays and the remaining are leptonic decays to electrons, muons and taus in almost equal amounts. A detailed listing exists in Table 2.1. This study is focused on Zs decaying to electrons only, which account for approximately 3.4% of the total branching ratio.

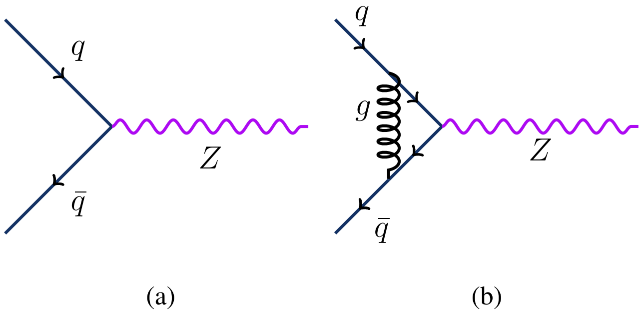


Figure 2.1: Drell-Yan Z production Feynman diagrams. (a) Shows the basic Drell-Yan process (b) is an example of NLO correction.

Mode	Fraction
e^+e^-	(3.363 ± 0.004)%
$\mu^+\mu^-$	(3.366 ± 0.004)%
$\tau^+\tau^-$	(3.370 ± 0.008)%
invisible	(20.00 ± 0.06)%
hadrons (total)	(69.91 ± 0.06)%

Table 2.1: Summary of Z decay modes [7].

2.2 Associated jet production

As a result of the Drell-Yan process, the majority of Z bosons is produced at rest or with very little momentum. Production of hard outgoing partons (jets) in association with a Z boson results in more complex and interesting events. An additional factor of interest is that Z+jets final states are common to many rare signals like top decay, or new physics such as associated production of the Higgs boson.

The number of processes contributing to these final states increases as the numbers of jets goes up. For example, while there are only 9 processes contributing to (Z+1)-jet events,

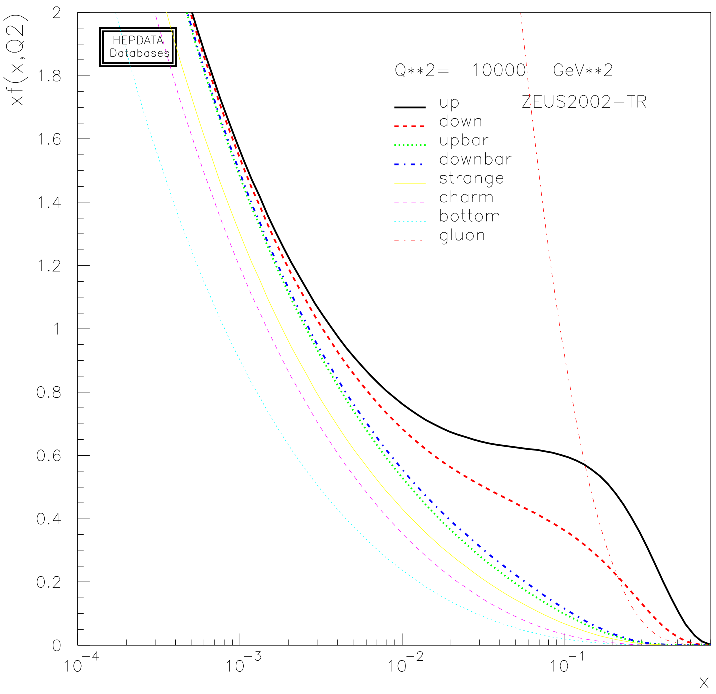


Figure 2.2: Parton Distribution Functions at $Q^2 = 10000\,\mathrm{GeV}^2$ [11]

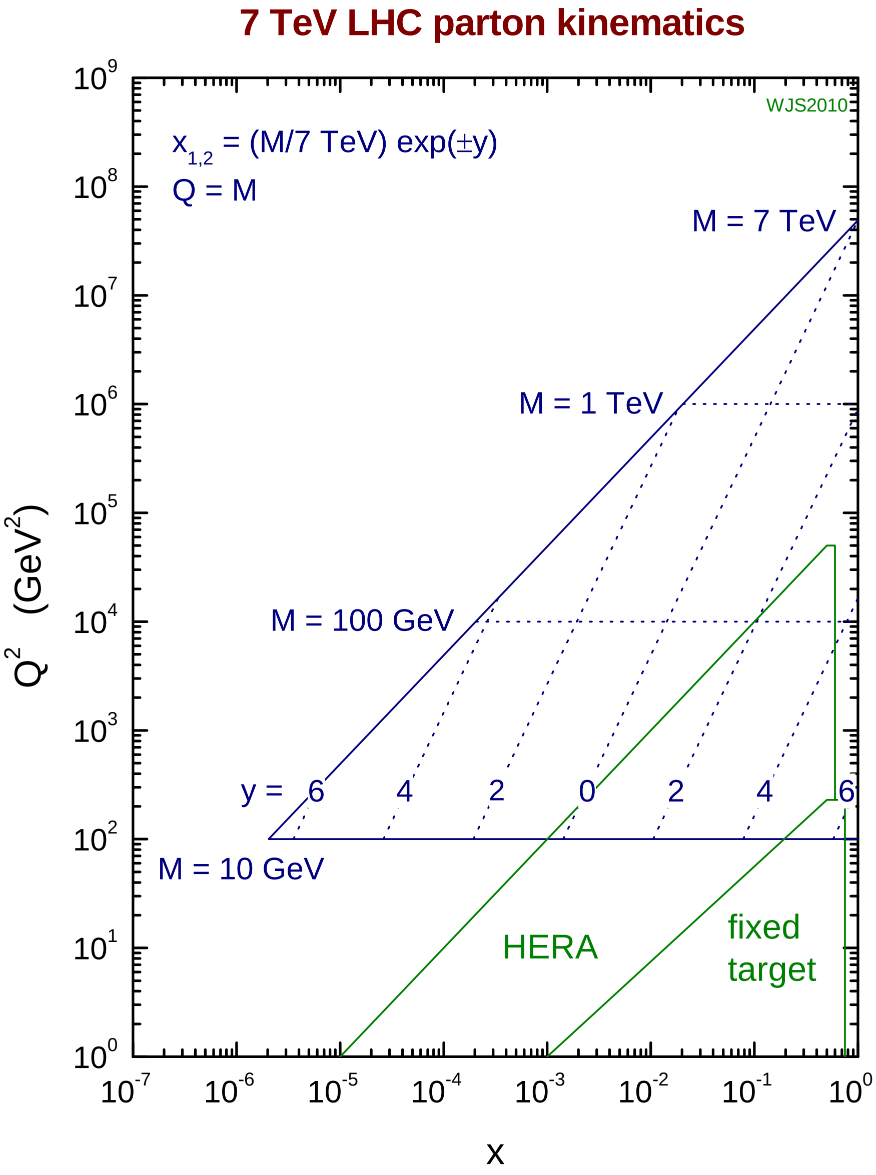


Figure 2.3: The LHC kinematic plane showing the relation between the parton variables (x, Q^2) and kinematic variables corresponding to a final state of mass M produced with rapidity y at the LHC [12] for 7 TeV collision energy.

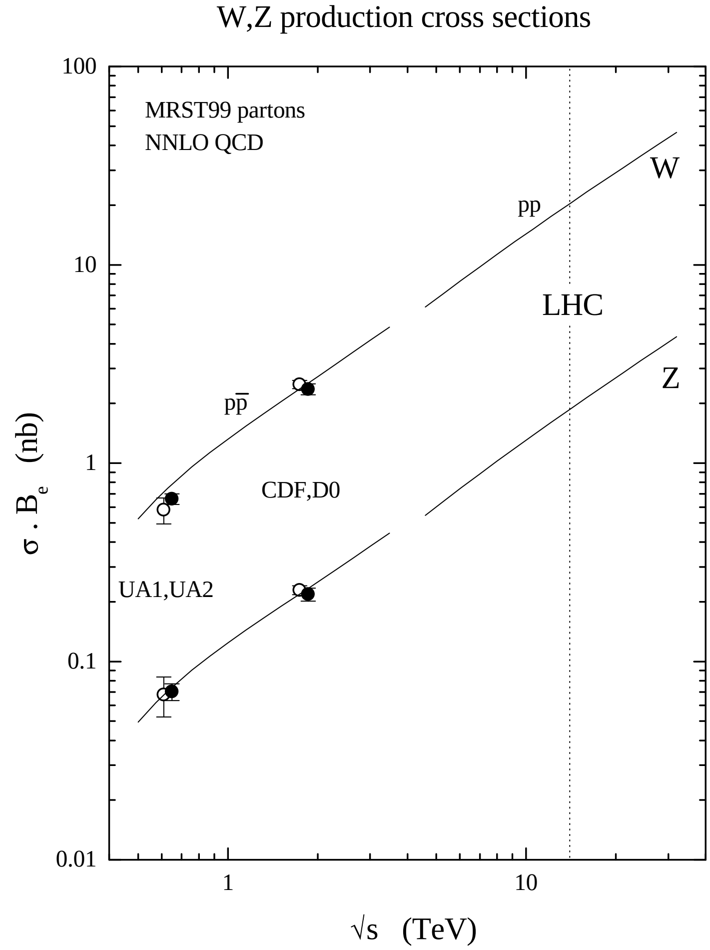


Figure 2.4: Predictions for the total W, Z production cross sections times leptonic branching ratios in $p\bar{p}$ and pp collisions as a function of the collider energy $\sqrt{s}$. Experimental measurements from the UA1, UA2, CDF and D0 experiments are also shown [13].

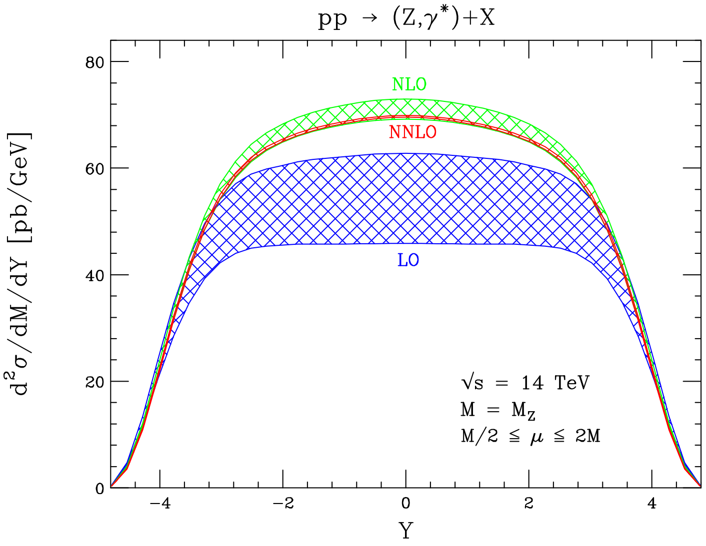


Figure 2.5: Expected Z production rapidity distribution at the LHC ($\sqrt{s} = 14\,\text{TeV}$) at LO, NLO and NNLO. [14]

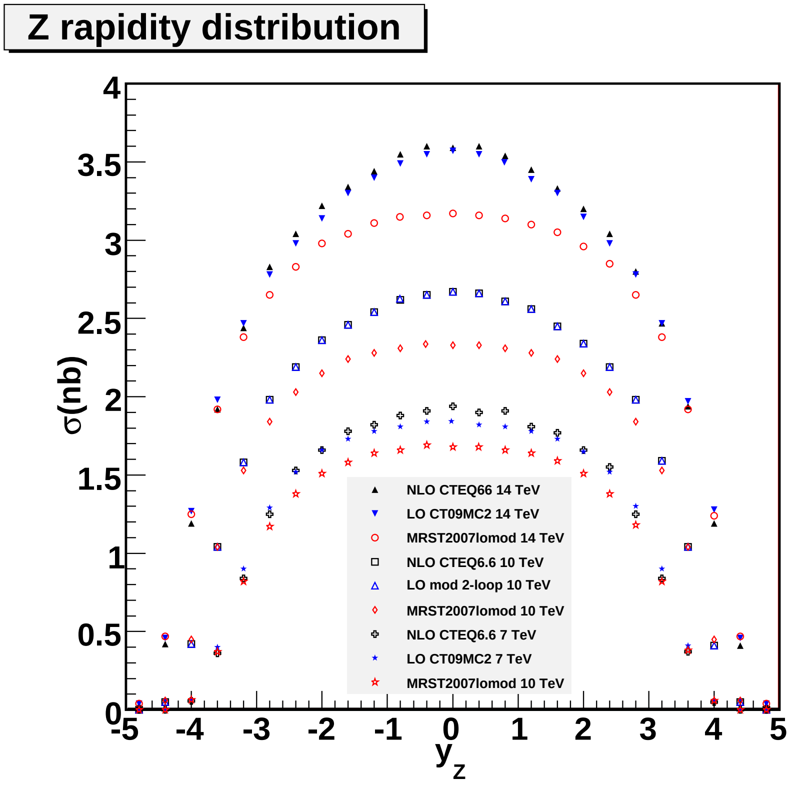


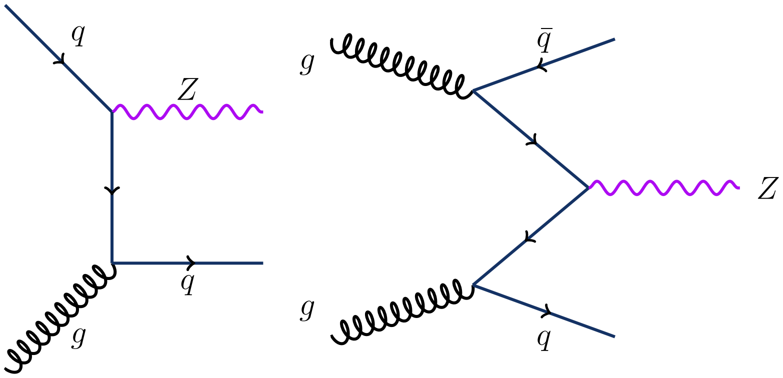
Figure 2.6: Predictions for the Z rapidity distribution at the LHC ($\sqrt{s}$ =7, 10 and 14 TeV) in Δy =0.4 bins, given at NLO using the CTEQ6.6M PDFs, and at LO using the CT09MC2 and MRST2007lomod PDFs. [15]

for (Z+4)-jet events there are up to 485 tree-level processes [16]. For high Q^2 the value of the strong coupling constant is $\alpha_s \ll 1$, therefore multiple exchanges involving terms in $a_s^2, a_s^3, ...$ have decreasing cross sections. Feynman diagrams demonstrating Z production in association with 1 and 2 outgoing jets in the final state are shown in Figure 2.7, the first originating from a quark-gluon interaction and the second from a gluon-gluon interaction. Due to the difficulty in calculating the extra loops involved in higher-order calculations, it is very difficult to calculate cross sections even at next-to-leading order. Current state-of-the-art calculations go up to Z+3 jets at NLO [17].

Performing precise measurements serve as validation tests of pQCD predictions and to help to properly model and constrain backgrounds for other searches. For this reason, in the presented analysis ratio measurements of

$$\frac{Z+\geq n-jets}{Z+\geq 0-jets} \text{ and } \frac{Z+\geq n-jets}{Z+\geq (n-1)-jets}$$

were performed as many significant sources of systematic errors (in particular from jet energy scale, integrated luminosity and lepton selection efficiency) cancel out.



(a)

(b)

Figure 2.7: Quark-gluon and gluon-gluon Z production Feynman diagrams. (a) shows a gq interaction and (d) shows Z production via gluon-gluon fusion with two jets in the final state.

> *Difficulty? What do you mean, difficulty? It's the single simplest machine in the entire Universe!*
>
> The Restaurant at the End of the Universe
> DOUGLAS ADAMS

Chapter 3

Experimental Setup

The need to probe at the TeV energy scale to seek answers to fundamental particle physics questions fueled the decision to build the Large Hadron Collider at the European Center for Nuclear Research (CERN). The beam energy and design luminosity of the Large Hadron Collider have been carefully chosen in order to study the expected phenomena at this energy regime. These conditions also require robust detectors to collect collision data; the Compact Muon Solenoid is one of the four detectors placed along the accelerator circumference.

3.1 Large Hadron Collider

The Large Hadron Collider (LHC) [18] is a two-ring particle accelerator and collider, built in the same 26.7 km tunnel as the Large Electron Positron collider. The tunnel has eight straight sections and eight arcs, and lies between 45 m and 170 m below the earth's surface. The two proton beams rotate in opposite directions with a nominal energy of 7 TeV and collide in 4 interaction points where four experiments have been built; CMS and ATLAS are two general-purpose studying general SM processes and looking for new physics, the ALICE experiment will investigate heavy-ion collisions and LHCb is dedicated to b-meson physics and will carry out precise CP-violation measurements.

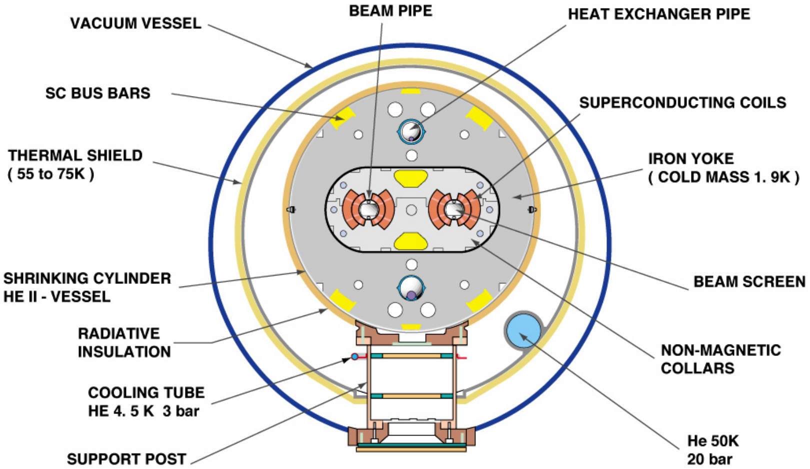


Figure 3.1: Cross section of LHC dipole magnet.

To keep the protons on track, 1232 superconducting dipole magnets operating at 1.9 K deliver the nominal 8.33 T magnetic field required to deflect them (currently operating at half the field). The LHC uses twin bore magnets that consist of two sets of coils and beam pipes within the same structure (Figure 3.1).

The LHC is supplied with protons from an accelerator chain which gradually increases proton energy (Figure 3.2). The protons are acquired by stripping electrons from hydrogen atoms; these are accelerated in a linear accelerator (LINAC2) to an energy of 50 MeV. The Proton Synchrotron Booster (PSB) increases their energy to 1.4 GeV and the beam is then fed to the Proton Synchrotron (PS) where it is accelerated to 25 GeV before they reach the Super Proton Synchrotron (SPS) and 450 GeV, the energy they carry as they are injected in the LHC.

The number of generated events per second for a particular process under study in the LHC is given by:

$$N = \sigma \times L \tag{3.1}$$

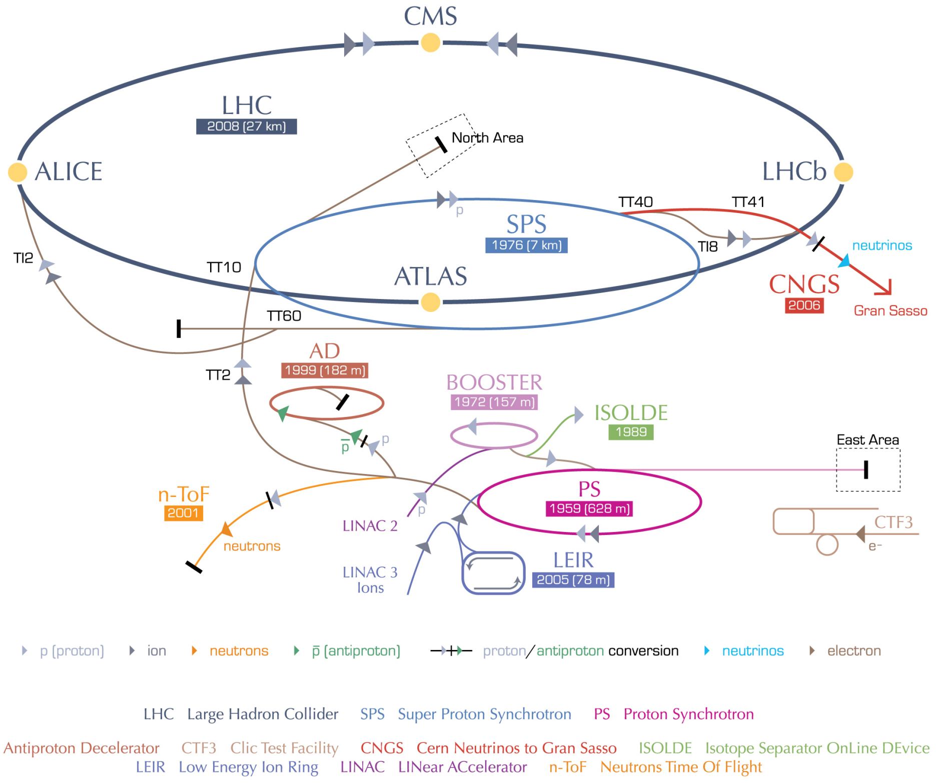


Figure 3.2: The LHC is the last ring in a complex chain of particle accelerators. It is supplied with protons from the injector chain Linac2 → Booster → PS → SPS.

where σ is the process cross section and L the luminosity of the machine. Luminosity depends on the machine parameters and for a Gaussian beam distribution is given by:

$$L = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta*} \tag{3.2}$$

where N_b is the number of particles per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency, γ_r the Lorentz factor, ϵ_n the normalized transverse beam emittance and $\beta*$ the beta function at the collision point.

The designed LHC nominal luminosity is L = 10^{34} $cm^{-2}s^{-1}$; during the 2010 running period the maximum achieved instantaneous luminosity was 2×10^{32} $cm^{-2}s^{-1}$. The protons arrive in the LHC in bunches of approximately 10^{11} protons, with a bunch spacing of 25 ns, resulting in 2808 bunches in total. For the 2010 run period, LHC achieved to collide up to 424 bunches per proton beam.

For the first years of operation, LHC is accelerating beams up to 3.5 TeV per beam [19], a decision taken after an incident during early commissioning tests [20] which delayed data taking by more than a year. During the 2010 running, the LHC delivered $47\,\mathrm{pb}^{-1}$ of collision data. The CMS experiment recorded $43\,\mathrm{pb}^{-1}$ of these, and after quality checks approximately $36\,pb^{-1}$ of collision data were approved for analyses.

3.2 The CMS Detector

The Compact Muon Solenoid [21] is a general-purpose detector installed at Interaction Point 5 of the LHC ring, near the village of Cessy in France, built to exploit the full physics reach of the LHC. At design luminosity on average 20 inelastic collisions are expected every 25 ns, the products of which need to be collected and processed. The detector should provide good spatial and time resolution in order to distinguish between the event of interest and pile-up events. This requires a large number of detector channels, which need to be well-synchronized.

In short the general requirements for such a detector are:

- Good charged particle momentum resolution and reconstruction efficiency in the inner tracker, high granularity especially near the interaction point.
- Good electromagnetic energy resolution, wide coverage, efficient photon and lepton isolation
- Hadron calorimeters with hermetic coverage, providing good missing transverse en-

ergy and dijet-mass resolution

- Good muon identification, charge determination and momentum resolution over a wide range of momenta and angles.

To achieve the above, the main feature of CMS is a large superconducting solenoid magnet, providing a 3.8 T field. Starting from the closest to the interaction point, the individual subdetectors are the Tracker, the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL) which are surrounded by the solenoid, followed by layers of muon detectors (see Figure 3.2 for a schematic overview of the detector and its components). The overall dimensions of CMS are 21.6 m in length, a diameter of 14.6 m and a total weight of 12500 t.

The coordinate frame

CMS has adopted a coordinate system with the origin centered at the nominal collision point inside the experiment, the z-axis along the beam pipe, the y-axis pointing vertically upward, and the x-axis pointing radially inward toward the center of the LHC. Since CMS is cylindrically symmetric around the beam line, the radial distance from the beam is defined as r, the azimuthal angle ϕ is measured from the x-axis in the x-y plane and the polar angle θ is measured from the z-axis. Pseudorapidity is defined as $\eta = -\ln\tan\left(\frac{\theta}{2}\right)$. Therefore, the physical observables, momentum and energy transverse to the beam (p_T and E_T respectively), are computed directly from the x and y components. The overall missing energy in the transverse plane is denoted as E_T^{miss}.

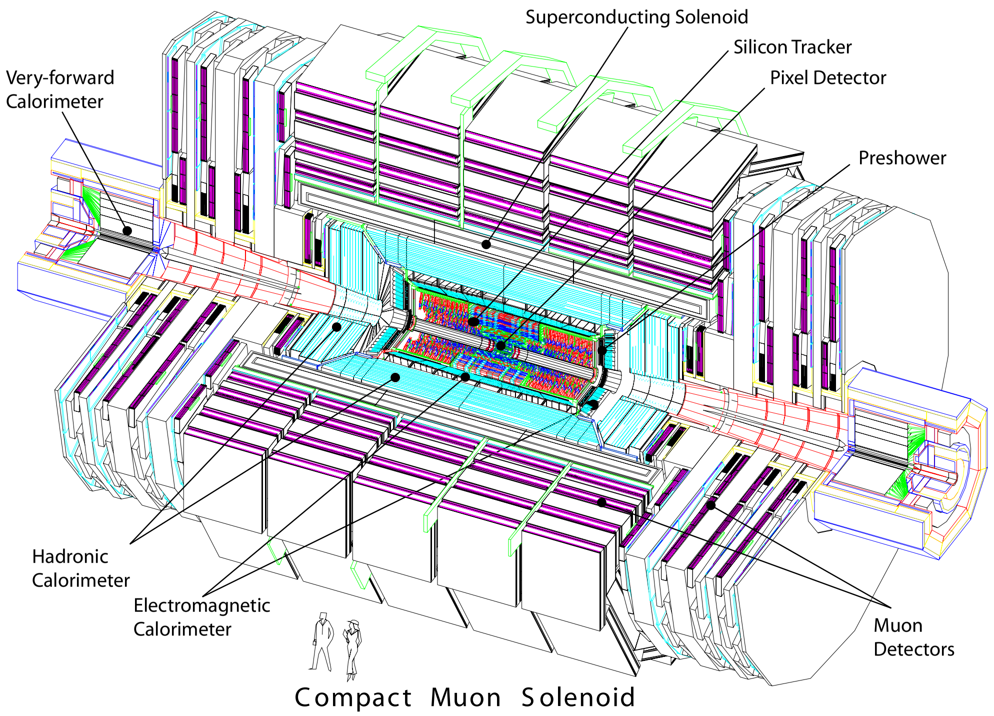


Figure 3.3: A schematic picture of the Compact Muon Solenoid detector and its components. The human figures demonstrate the size of the detector.

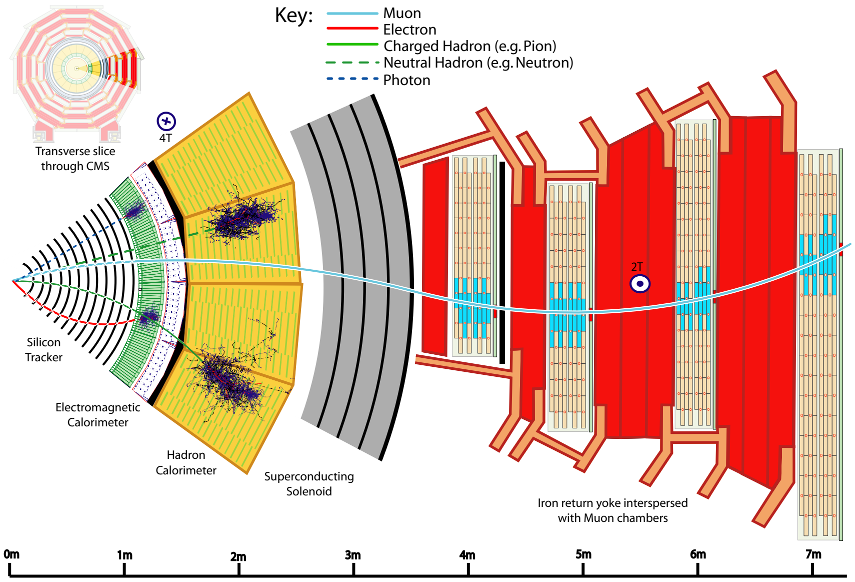


Figure 3.4: Transverse slide of the CMS detector, showing particles incident on the different subdetectors.

3.2.1 Magnet

Precise measurement of charged particle momentum at high energies requires large bending power and, as a result, a strong magnetic field. Momentum resolution measurement error depends on the projected track length in the magnetic field L and the strength of the field B,

$$\frac{\delta p}{p} \sim \frac{1}{L^2 B} \tag{3.3}$$

To improve resolution, CMS has opted for a high magnetic field leading to a compact tracking volume. The superconducting magnet for CMS has been designed to reach a 4 T field (currently operates at 3.8 T to increase the lifetime of the magnet) in a free bore of 6 m in diameter and 12.5 m in length. The coil is made from 4 layers of NbTi superconductor. The flux is returned through a 10 000 t yoke comprising of 5 wheels and 2 endcaps, composed of 3 disks each. The return field is large enough to saturate 1.5 m of iron, allowing 4 muon stations to be integrated to ensure robustness and full geometric coverage.

3.2.2 Tracking System

The CMS tracker [22, 23] is the detector that lies closest to the beam interaction point. It is designed to provide a precise and efficient measurement of charged particles trajectories emerging from the LHC collisions, as well as a precise reconstruction of secondary vertices. In order to achieve optimal pattern recognition of the particle tracks traversing through its volume, two basic properties are fundamental: low cell occupancy and large hit redundancy. Low hit occupancy is achieved thanks to the high granularity of the detector, while the large number of detector layers ensures measurement redundancy. CMS has opted for an all-silicon tracking system, composed of a pixel detector and a silicon strip tracker, providing good coverage up to a pseudorapidity of $|\eta| < 2.4$. Its dimensions are approximately 540 cm in length and 220 cm in diameter.

The material budget in the tracker is of utmost importance; more tracker material results in lower performance as energy is lost due to Bremsstrahlung in the electron case and nuclear interactions in the hadron case. Figure 3.5 shows the material budget of the CMS tracker in units of radiation length X/X_0. At $\eta \approx 0$ starts at about 0.4 X_0, reaching a maximum of about 1.8 X_0 at an $|\eta| \approx 1.4$ due to the presence of cabling and other auxiliary services to fall again to about 1 radiation length at $|\eta| \approx 2.5$.

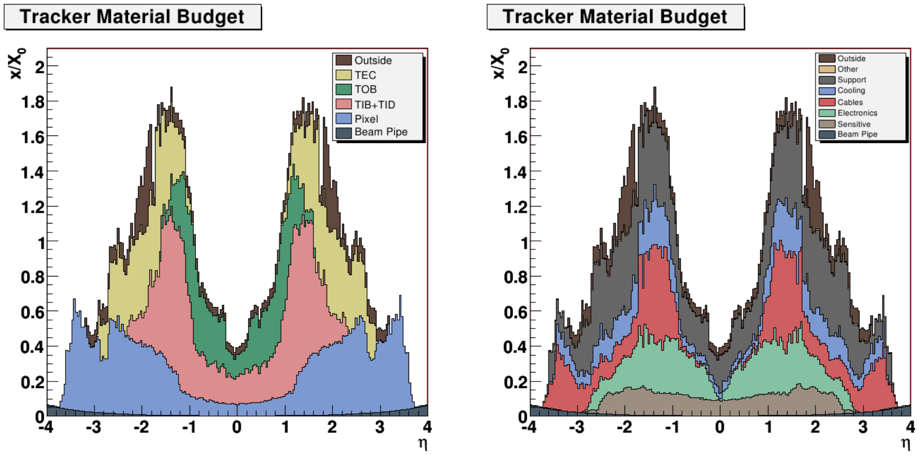


Figure 3.5: Tracker material budget in units of radiation length as a function of pseudorapidity for the different sub-detectors (left) and by functional contributions (right).

Pixel detector

The pixel detector consists of three layers of pixel modules, placed at radii of 4.4, 7.3 and 10.2 cm and two end disks extending from 6 to 15 cm in radius placed at 34.5 cm and 46.5 cm from the collision point. Each silicon sensor has a surface of $100\,\mu m \times 150\,\mu m$ and a thickness of $300\,\mu m$, with about 65 million such channels. The spatial resolution has been measured to be about $10\,\mu m$ for the (r, ϕ) measurement and about $20\,\mu m$ for the z measurement. The pseudorapidity range covered is $-2.5 < \eta < 2.5$ and the arrangement of the barrel layers and endcap discs ensures the existence of 3 tracking points over almost the full covered range (Figure 3.6).

Silicon strip tracker

At higher radii, due to the reduced particle flux, silicon microstrip detectors have been used. The CMS silicon strip tracker has about 9.3 millions of strips, arranged in ten barrel

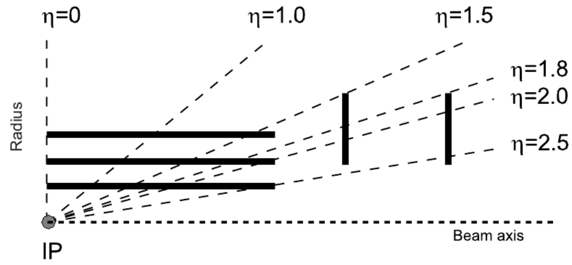


Figure 3.6: Pixel detector hit coverage.

layers extending to a radius of 1.1 m complemented by twelve endcap disks on each side, covering up to $|\eta| < 2.5$. Some layers are single sided, while some carry double sided detectors. The double sided modules, are mounted back-to-back with an angle of $100\,\mu$rad in order to provide a measurement of the second coordinate (z in the barrel and r on the disks). The tracker layout ensures at least 9 hits in the silicon strip tracker with at least 4 of them being two-dimensional measurements. The entire silicon strip tracker consists of about 15 400 modules, mounted on a carbon fiber support structure. The silicon strip module arrangement on the CMS tracker can be seen in Figure 3.7.

3.2.3 Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) CMS has created is a hermetic, homogeneous calorimeter made of lead tungstate crystals ($PbWO_4$). The high density of the crystals (8.28 g/cm^3), in combination with the small Molière radius[1] (21.9 mm) and the short radiation length (89 mm) allowed CMS to build a compact calorimeter which provides fast response (about 80% of the produced light can be collected in 25 ns, the nominal LHC bunch spacing), high granularity and good energy resolution.

---

[1] Molière radius: the radius of a cylinder containing on average 90% of the shower's energy deposition.

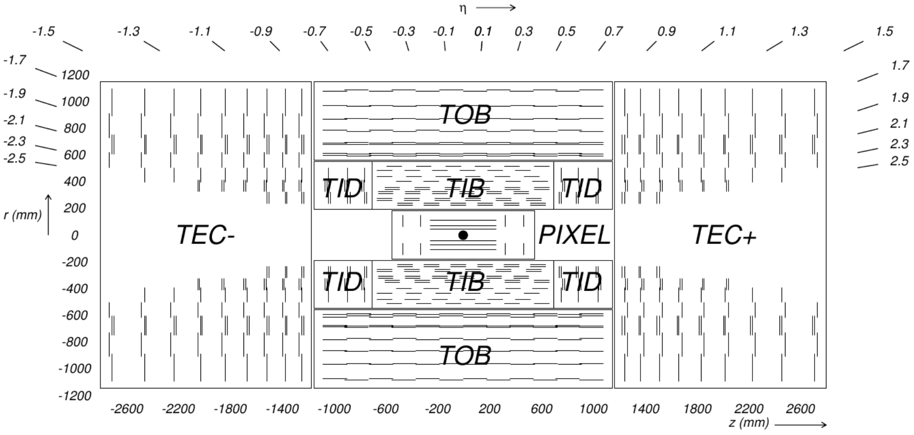


Figure 3.7: Schematic cross section of the CMS tracker. Each line represents a detector module. Double lines indicate back-to-back modules.

The ECAL barrel (EB) covers the pseudorapidity range $|\eta| < 1.479$ and is made of 61 200 lead tungstate crystals. The front face of each crystal is $22{\times}22\,\mathrm{mm}^2$ to match the Molière radius of PbWO_4. The length of each crystal is 230 mm, corresponding to 25.8 X_0. The crystals are mounted so that a 3^o angle is formed with the vector from the interaction point to avoid cracks aligned with the particle trajectory. The EB is complemented by 7 324 crystals in each side of the two endcap (EE) disks, covering the range $1.479 < |\eta| < 3$. The EE crystals have a front face of $28.62{\times}28.62\,\mathrm{mm}^2$ and a length of 220 mm (24.7 X_0). Avalanche photodiodes (APDs) in the barrel and vacuum phototriodes (VPTs) in the endcaps collect and amplify the scintillated light. The transition region between barrel and endcap is called the “crack” region where other services had to be installed (cabling etc.), thus leading to smaller reconstruction efficiency, and is usually excluded from analyses.

A preshower detector is placed in front of the endcap disks, made up of two strips of silicon strip detector and disks of lead absorber in order to identify neutral pions and improve electron discrimination against minimum ionizing particles and improve the determination

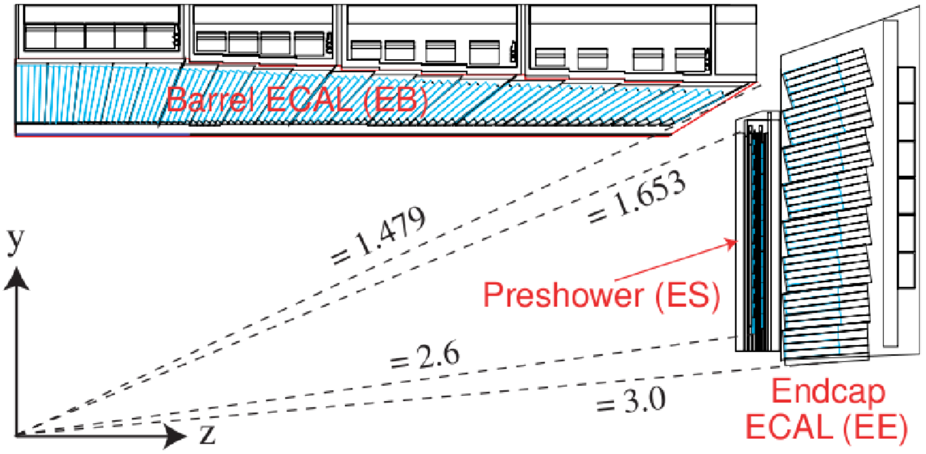


Figure 3.8: Schematic of the ECAL transverse view. The tilted placement of the crystals and the “crack” region are distinguishable.

of the position of electrons and photons.

The parametrized ECAL energy resolution is expressed as

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{3.4}$$

where S is the stochastic term, N the noise term and C the constant term. The stochastic term describes the contribution from photostatistics and event to event fluctuation in the lateral shower containment. The noise term includes the contribution from electronics, digitization and pileup noise. Contributions to the constant term come from the non-uniformity of the longitudinal light collection, inter-calibration errors and energy leakage from the back of the crystal. Test beam data [21] have shown typical energy resolution to be

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right)^2 + (0.30\%)^2 \tag{3.5}$$

3.2.4 Hadronic calorimeter

To measure the energy of hadron jets and also reconstruct the missing energy coming from neutrinos or other possible exotic particles CMS has deployed a brass/scintillator sampling calorimeter (HCAL). To reconstruct efficiently missing transverse energy, a hermetic coverage up to $|\eta| = 5$ is required.

The design of HCAL was strongly affected by the CMS magnet restrictions, as a big portion resides between ECAL and the superconducting solenoid. It is composed of four subdetectors:

- **Barrel Hadronic Calorimeter (HB)**: located between EB and the CMS magnet, covers up to $|\eta| < 1.4$
- **Endcap Hadronic Calorimeter (HE)**: also placed inside the CMS magnet, is made of two endcaps complementing HB and extends the covered range up to $|\eta| < 3$
- **Outer Hadronic Calorimeter (HO)**: sits just outside of the CMS magnet and ensures that enough sampling depth is provided in the barrel region
- **Forward Hadronic Calorimeter (HF)**: placed at $11.2\,\mathrm{m}$ from the interaction point, extends the covered pseudorapidity range to $|\eta| = 5.2$

The HB absorber consists of a 40-mm-thick front steel plate, followed by eight 50.5-mm-thick brass plates, six 56.5-mm-thick brass plates, and a 75-mm-thick steel back plate. The total absorber thickness at 90^o is 5.82 interaction lengths (λ_I). The HB effective thickness increases with the polar angle (θ) as $1/\sin\theta$, resulting in 10.6 interaction lengths at $|\eta| = 1.3$. The electromagnetic calorimeter in front of HB adds about 1.1 λ_I of material.

The HE uses the same brass absorber, but the plates have a thickness of 79 mm. Between the absorber layers, 70 000 plastic scintillator tiles are used. The $\Delta\eta \times \Delta\phi$ granularity of the HB tiles is 0.087×0.087, matching the ECAL trigger towers. The HE has a granularity of 0.087×0.087 for $|\eta| < 1.6$ and 0.17×0.17 for $|\eta| > 1.6$.

The HO utilises the solenoid coil as an additional absorber equal to $1.4/\sin\theta$ interaction lengths and is made of two scintillator layers, with the same granularity as HB; the total depth in the central region is thus extended to about $11.8\,\lambda_I$.

Due to the harsh radiation environment in the forward region (on average, 760 GeV per proton-proton interaction is deposited into the two forward calorimeters, compared to only 100 GeV for the rest of the detector), the two HFs use quartz fibres as an active material embedded in a steel absorber.

The energy resolution of the CMS HCAL is parametrized as

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{90\%}{\sqrt{E}}\right)^2 + (4.5\%)^2 \tag{3.6a}$$

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{172\%}{\sqrt{E}}\right)^2 + (9.0\%)^2 \tag{3.6b}$$

with the first relationship describing the barrel/endcap resolution and the second the HF.

A longitudinal view of CMS, showing the locations of hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters with their respective $|\eta|$ coverage is shown on Figure 3.9.

3.2.5 The Muon detectors

The precise and robust measurement of muons was of central importance from the early stages of CMS planning. The muon system is designed to reconstruct the momentum

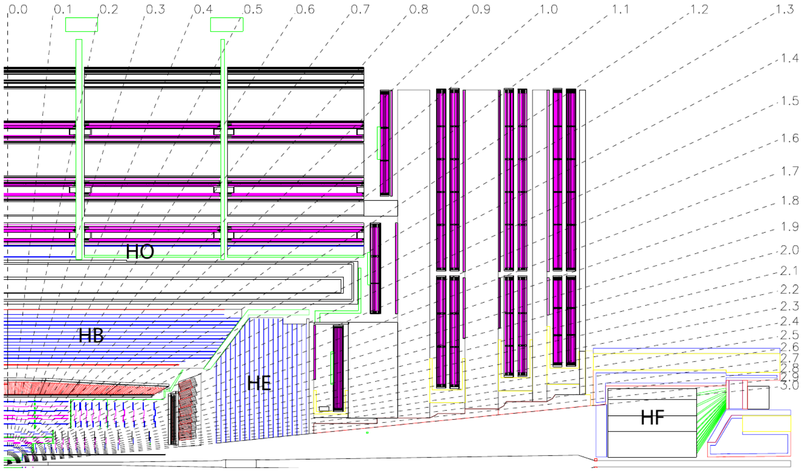


Figure 3.9: Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters and the η region they cover.

and charge of muons over the entire kinematic range of LHC. In addition, it is required to have a quick enough response to the passage of muons in order to provide information to the CMS trigger system. The muon detectors are placed outside of the CMS magnet, contained in the return yoke to exploit the return field. Due to the geometry of CMS, the muon system also has a cylindrical barrel section and two endcaps. The system consists of three independent gaseous subdetectors [24], utilizing different detection technologies which complement each other.

In the barrel region ($|\eta| < 1.2$) where the neutron-induced background is small and the muon rate is lower, Drift Tube (DT) chambers are used (Fig. 3.10). The DTs are aluminum tubes with a stainless steel wide in the middle acting as the anode. There are four stations of DTs among the layers of the return yoke plates. The first 3 stations contain 8 chambers each, which measure the muon coordinate in the $r-\phi$ plane and 4 chambers along the

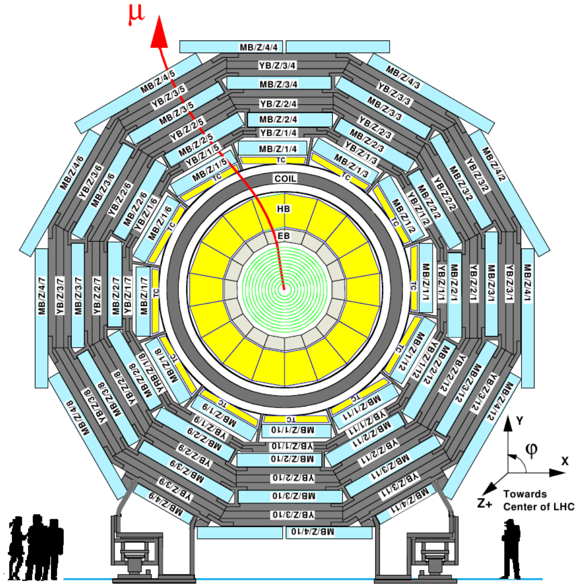


Figure 3.10: Schematic layout of the CMS barrel muon DT chambers in one of the 5 wheels.

beam line, providing a measurement in the z direction. The fourth station does not contain the z-measuring planes. The 2 sets of 4 chambers in each station are separated as much as possible to achieve the best angular resolution. The number of chambers in each station and their orientation were chosen to provide good efficiency when linking together muon hits from different stations into a single muon track and for rejecting background hits.

In the two CMS endcaps, where the muon and background rates are high, 468 cathode strip chambers (CSC) are used. The CSCs are multiwire proportional chambers comprised of 6 anode wire planes interleaved among 7 cathode panels. Their fast response time, fine segmentation, and radiation resistance makes the CSCs ideal to identify muons between $|\eta|$ values of 0.9 and 2.4. There are 4 stations of CSCs in each endcap, with chambers positioned perpendicular to the beam line.

In addition to the DTs and CSCs, the CMS muon system utilizes Resistive Plate

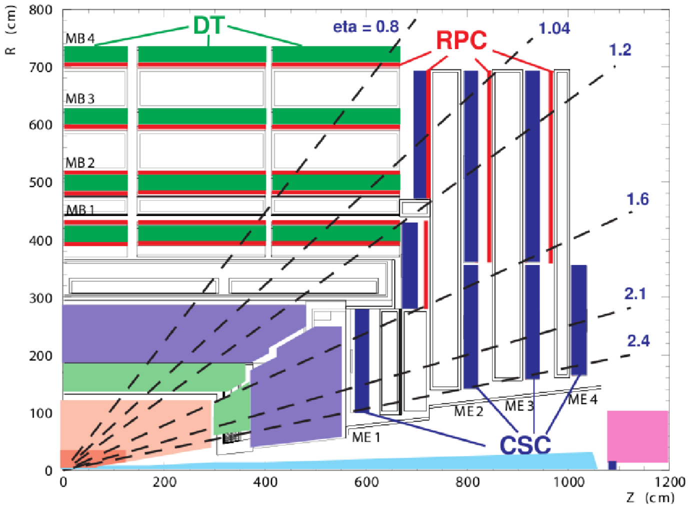


Figure 3.11: Layout of one quadrant of the CMS muon system in the r-z plane showing DT, CSC and RPC subdetectors.

Chambers (RPC) to provide fast, independent and highly-segmented trigger information. The RPCs provide information complementary to the other two muons subdetectors over a large range ($|\eta| < 1.6$) of the muon system. The RPCs are double-gap chambers, operated in avalanche mode to ensure good operation at high rates. They produce a fast response, with good time resolution but coarser position resolution than the DTs or CSCs, and also help to resolve ambiguities in attempting to make tracks from multiple hits in a chamber.

The CMS muon system consists of about 25 000 m^2 of detection planes and about one million readout channels.

3.2.6 Trigger and Data Acquisition

As discussed (section 3.2), the LHC is expected to deliver proton-proton collisions every 25 ns, corresponding to a crossing frequency of 40 MHz. Approximately 10^9 interactions are expected to be produced every second at design luminosity. Considering that only a very small fraction of these events are hard scattering interactions containing possibly interesting events and that the size of every event is of the order of 1 MB, a drastic rate reduction is required. This task is performed by the trigger system, which filters and stores events with interesting signatures for offline analysis.

CMS has chosen a two-level trigger architecture: a Level-1 Trigger (L1) which consists of custom-designed, programmable electronics and a High-Level Trigger (HLT) which runs on a computer farm of about 1 000 commercial processors, using scaled-down versions of the CMS offline reconstruction algorithms. The combined reduction rate has been designed to be at least 10^6, with the L1 having an output rate of 100 KHz and the HLT further reducing this rate to 300-400 Hz.

3.2.7 Level 1 trigger

To achieve the required rate reduction in the limited latency window of 3.2 μs, the L1 trigger [25] uses coarsely segmented data from the calorimeters and the muon system to identify photons, electrons, muons, jets and missing transverse energy. The full detector data is kept in pipeline memories in the detector front-end electronics until the L1 decision is reached; the L1-Accept is then propagated to the various subdetectors through the Timing, Trigger and Control (TTC) system.

The L1 Trigger has local, regional and global components. Trigger Primitive Generators (TPG) identify energy deposits in calorimeter trigger towers and track segments or hit patterns in muon chambers. Regional Triggers use this information to determine ranked and sorted trigger objects. The Global Calorimeter (GCT) and Global Muon Triggers (GMT) determine the highest-rank calorimeter and muon objects across the entire experiment and transfer them to the Global Trigger (GT) which combines the information and

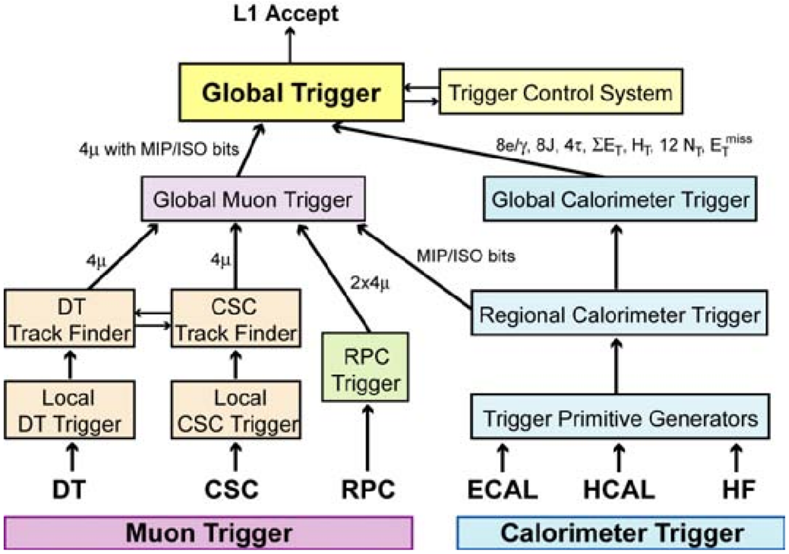


Figure 3.12: Level-1 Trigger architecture. The flow of information for the Calorimeter and Muon triggers is shown.

decides whether to keep the event or not (Fig. 3.12).

Calorimeter trigger

The calorimeter trigger TPGs sum the transverse energies measured in ECAL crystals and HCAL towers. In the region up to $|\eta| = 1.74$, each trigger tower has an (η, ϕ) coverage of 0.087×0.087 corresponding to 5×5 ECAL crystals. Beyond that boundary, the towers are larger. The TPG information from 7 000 trigger towers are transmitted through high-speed serial links to the Regional Calorimeter Trigger (RCT) which detects signatures of regional electron, photon, tau and jet candidates as well as missing and total transverse energy. The position and transverse energy of these regional candidates are then fed to the GCT which determines the top four highest-rank isolated and non-isolated calorimeter trigger objects across the entire detector, total transverse energy, missing transverse energy, jet counts, jet

E_T sums (H_T) and the missing hadronic transverse energy and sends them to the GT.

Muon trigger

All three muon subdetectors provide trigger information. The barrel DT chambers provide local trigger information in the form of track segments in the ϕ-projection and hit patterns in the η-projection. The endcap CSCs provide 3-dimensional track segments and the RPCs deliver track candidates based on regional hit patterns. The Regional Muon Trigger consists of DT and CSC Track Finders which join segments to reconstruct muon tracks. The Global Muon Trigger (GMT) then combines the information from the three subdetectors achieving better efficiency and momentum resolution than the standalone systems. The muon trigger coverage is extends up to an $|\eta| \leq 2.4$.

L1 Trigger Supervisor

Being such a complex system, the CMS L1 trigger needs a system to control and monitor its proper operation. This is achieved through the L1 Trigger Supervisor [26] (TS) framework which provides a unified interface to developers, experts and users to the online software infrastructure of the various trigger subsystems.

The TS is based on XDAQ [27], a platform for the development of distributed data acquisition systems [28] and acts as an intermediate layer between the low-level hardware operations and the users. Each trigger subsystem implements its own online TS application (cell) which can be accessed from anywhere using a web browser. A hierarchical structure exists, where all L1 subsystems can be controlled from a top-level node called the Central Cell. The communication between cells is handled via the Simple Object Access Protocol [29] (SOAP) interface and a Finite State Machine is used to coordinate the configuration of

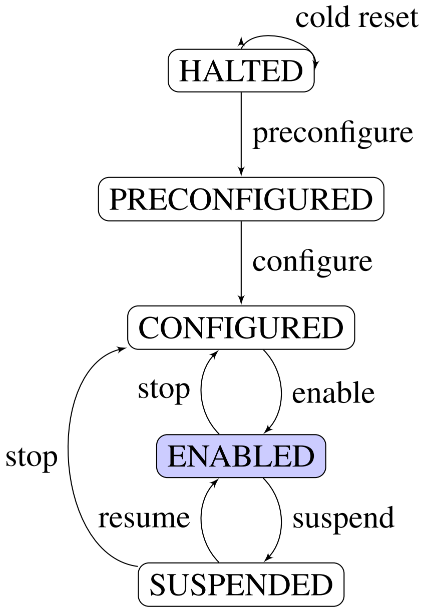


Figure 3.13: The Finite State Machine controlling the operation of the CMS L1 Trigger.

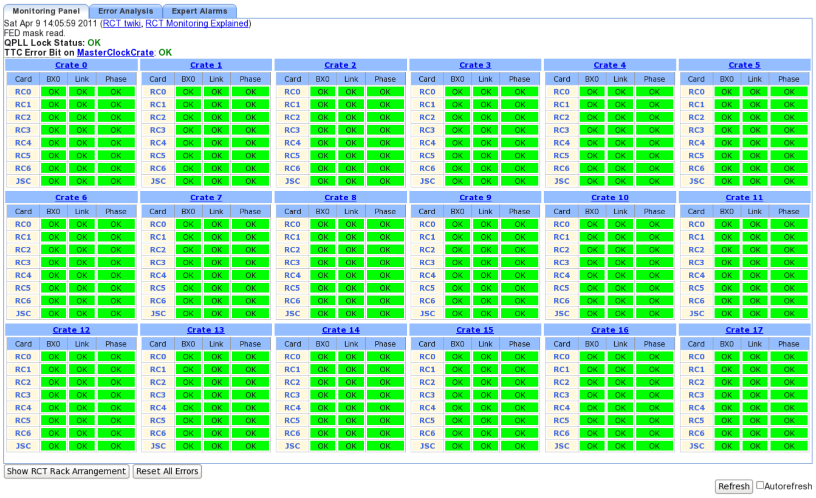
the system (Fig. 3.13)

The RCT Trigger Supervisor cell

The RCT trigger supervisor cell handles the configuration and monitor operation of the Regional Calorimeter Trigger. A supervisor cell receives requests and sends replies to the Central Trigger Cell and coordinates 20 worker cells; 18 of them corresponding to RCT hardware crates, one RCT Master Clock Crate (MCC) and the RCT TTCci[2]. When the "configure" request is received from the central cell, a proper key is also passed, corresponding to a specific set of look-up tables to be loaded on the hardware memories. At all times, a specific set of bits reporting the status of the system is polled and if a programmable threshold is reached, the experts are informed via e-mail. Channels can be masked/unmasked if required and a graphical overview of the status of all the incoming links from ECAL/HCAL

---

[2] TTCci: Timing & Control CMS Interface



Sat Apr 9 14:05:59 2011 (RCT twiki, RCT Monitoring Explained)
FED mask read.
QPLL Lock Status: OK
TTC Error Bit on MasterClockCrate: OK

Crate 0

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Crate 1

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Crate 2

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Crate 3

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Crate 4

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Crate 5

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Crate 6

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Crate 7

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Crate 8

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Crate 9

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Crate 10

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Crate 11

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Crate 12

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Crate 13

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Crate 14

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Crate 15

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Crate 16

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Crate 17

Card	BX0	Link	Phase
RC0	OK	OK	OK
RC1	OK	OK	OK
RC2	OK	OK	OK
RC3	OK	OK	OK
RC4	OK	OK	OK
RC5	OK	OK	OK
RC6	OK	OK	OK
JSC	OK	OK	OK

Figure 3.14: The RCT Trigger Supervisor monitoring panel. Detailed bit error status is provided for every RCT link.

is available (Fig. 3.14).

3.2.8 High Level trigger

Events passing the L1 trigger are then passed to the HLT [30]. Due to the L1 rate reduction, the HLT requires and average 50 ms decision window per event. As a consequence of the extra available time, the HLT has access to the complete detector read-out data, therefore it can perform complex calculations, similar to the ones made in offline data analysis. Data read from subdetectors are assembled by a builder unit and then passed to a 1 Tbit/s switching network that dispatches events to the processor farm.

In order to make efficient use of the allocated trigger decision time, the HLT code is divided in 3 virtual layers: *Level-2* where muon and calorimeter information is used, reconstructing more refined objects starting from the L1-accepted objects, *Level-2.5* additionally uses hit information from the pixel detector and *Level-3* makes use of the full detector in-

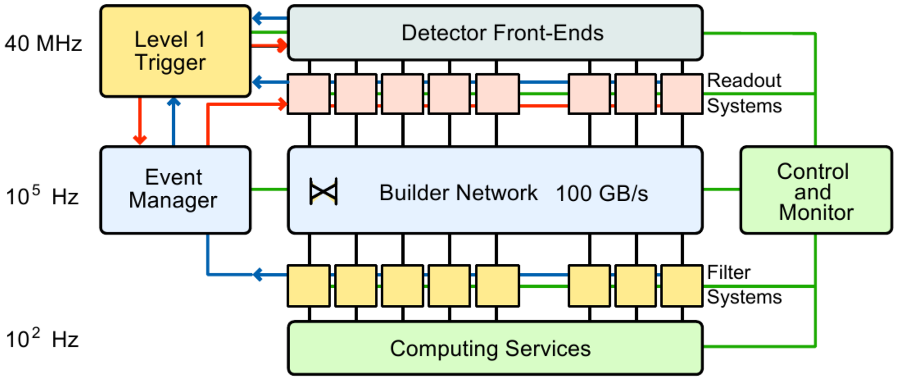


Figure 3.15: Architecture of the CMS Data Acquisition system.

formation. This way, algorithms that are computationally expensive are only executed on interesting events.

> *The last thing she wanted was to see her friend getting ideas in her head. There was such a lot of room in there for them to bounce around and do damage.*
>
> Unseen Academicals
> TERRY PRATCHETT

Chapter 4

Event Simulation and Reconstruction

The software framework is a vital component in the operation of an experiment as complex as CMS. All reconstruction algorithms, trigger paths, analysis strategies and other operational details of the experiment were studied and optimized well before the first LHC data were recorded by means of a robust simulation of the apparatus and of the expected processes using Monte Carlo (MC) simulated events.

4.1 Simulation Chain

The full simulation of a collision event at CMS, consists of three phases:

- *event generation*, where a collision is simulated, starting from the proton-proton interaction until the production of the final decay products to be observed with the CMS detector;
- *detector simulation* models the interactions of the generated final-state particles as they pass through the CMS detector materials and the responses of the different subdetectors and
- *digitization* emulates the response of the detector electronics to the detector hits.

4.2 Event Generation

Generating a Monte Carlo collision event is a complex process which the generators subdivide in several simpler, distinct steps. The *Parton Distribution Functions* (PDFs) describe the probability of finding a proton constituent with a momentum fraction *x*. The collision between two such partons is the *hard process* of the event which will generate the events of interest. As a first step, the matrix element according to the Feynman rules and subsequently the cross section of the hard process under study are calculated. The accelerated colored particles from this step produce QCD radiation, resulting in *parton showers* called initial (if coming from incoming partons) and final (from outgoing partons) state radiation (ISR and FSR respectively). Due to the proton-proton collisions in the LHC environment, initial state radiation is always present and at the same time partons existing in the final state contribute to the formation of high-p_T tight cones of particles called jets.

Outgoing partons in jets cannot exist as free particles, a result of *color confinement*. If two colored particles separate more than a certain distance, quark-antiquark pairs are created. As a consequence, partons in jets group form colorless hadrons by means of a mechanism called *hadronization*.

Along with the hard scatter which creates the process of interest, the *underlying event* contributes to what will be observed and constitutes a background to our measurement. It originates from the remaining partons of the incoming hard interaction protons, as well as additional pp interactions in the same beam crossing.

Many of the particles produced in the previous steps are unstable and decay into stable ones. Particles with short lifetimes are decayed by the generator, whereas longer-lived particles that can reach the detector are left undecayed and passed as an input to the next step.

A schematic representation of a generated event is shown in Fig. 4.1, showing the hard collision which would create our signal event, along with the secondary interactions which affect our measurement as well as the hadronization of quarks and gluons and the decays of short-lived particles.

4.2.1 Event Generators

A variety of Monte Carlo event generators exists which can be used to generate proton-proton collisions at LHC energies; some are general-purpose, others exist to cover specific processes. For this thesis, PYTHIA [31] and MadGraph [16] were the main generators used, and TAUOLA [32] was employed in the generation of samples containing $t\bar{t}$ decays.

PYTHIA

PYTHIA is a general-purpose generator. It has been used extensively at LEP, HERA and the Tevatron for e^+e^-, ep and $p\bar{p}$ physics. It contains a large subprocess library covering Standard Model physics but also SUSY, Technicolor and other Exotics processes.

To describe the hadronization process, the *Lund string model* [33] is used. This model is based on a picture where (anti)quarks are linearly confined, located at the ends of a string and gluons are energy and momentum carrying kinks on the string. Therefore, a gluon comes to be attached to two string pieces, one related to its color and the other to its anticolor, and experiences a confinement force twice that of a quark. The string breaks by producing new $q\bar{q}$ pairs; a quark coming from one break can combine with an antiquark

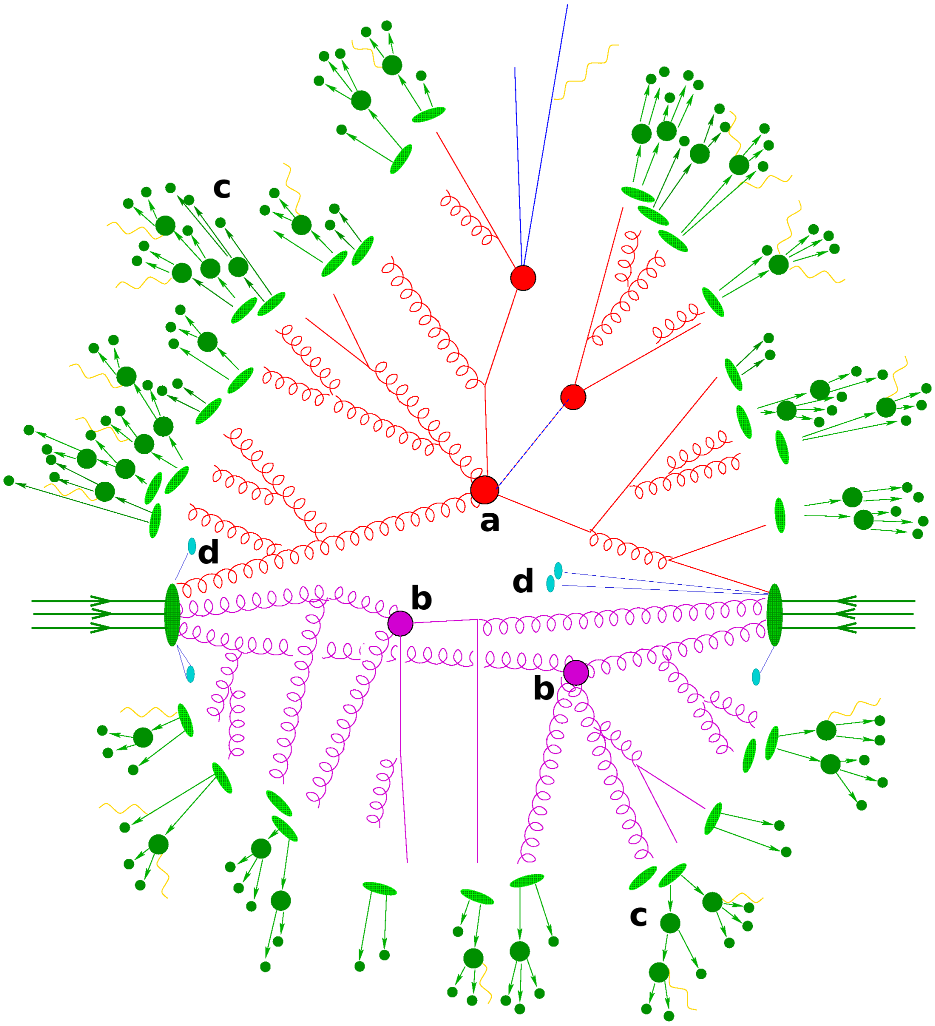


Figure 4.1: Schematic representation of a generated event. Apart from the hard collision (a), softer multiple interactions also take place (b) between the two incoming protons. Resulting quarks and gluons hadronize and decay (c). Beam remnants are marked with (d).

from a neighboring break to form a color singlet meson.

The generation of the underlying event is a complicated process. For its description several phenomenological models exist, with various degrees of sophistication. Due to the inherent uncertainties, a different number of "tunes" have been developed. At CMS, two

PYTHIA tunes have been applied to the generated MC samples and were studied: the Z2 [34] tune which seems to agree better with collected data and the D6T [35] tune which was used as complementary for systematics studies.

MadGraph/MadEvent

MadGraph [16] is a Matrix Element calculator. Given a process, it automatically creates the amplitudes for all the relevant subprocesses and produces the mappings for the integration over the phase space. This process-dependent information is passed to MadEvent which allows the calculation of event cross sections and the generation of unweighted events. The event information (particle IDs, momenta, spin etc.) is stored in the "Les Houches" format [36] and then interfaced with PYTHIA which handles the rest of the generation steps (involving parton showering, hadronization etc.).

4.3 Detector Simulation and Digitization

The detailed simulation of the CMS detector is based on the GEANT4 [37] toolkit. It relies on a detailed description of the hierarchy of the subdetector volumes and materials, and knowing which parts are "sensitive detector" (i.e., equipped with a readout sensor) as opposed to "dead materials" (cables and cooling for example). It takes as input generated particles, traces them through the simulated geometry, and models physics processes that accompany particle passage through matter. Results of each particle's interactions with matter are recorded in the form of simulated hits. An example of a simulated hit can be energy loss by a given particle within a "sensitive volume" of one of the subdetectors, stored along with several other characteristics of the interaction. Particles can be either "primary" (generated particles) or "secondary" (originating from GEANT4-modeled interactions of a primary particle with matter). These simulated hits are then used as input to emulators which model the response of the detector readout and trigger electronics and digitize this information by also taking into account noise and other factors.

4.4 Event Reconstruction

The aim of the event reconstruction process is to create analysis-level objects by combining recorded signals from the tracker, calorimeters and muon detectors. At low level, reconstructed hits are collected and through sophisticated algorithms energy clusters and segments are built. These are then used to construct higher-level objects such as electrons, photons, muons and jets.

4.5 Electron Reconstruction

Electron showers deposit their energy in several ECAL crystals. Approximately 94% of the incident energy of a single electron is contained in 3x3 crystals, and 97% in 5x5 crystals. As they cross the CMS tracker, they radiate photons through the Bremsstrahlung process, which results in a spread in ϕ of the initial electron energy as the electrons are bent from the magnetic field. This spread depends on the electron transverse momentum, leading to wider areas where the energy is deposited for low p_T electrons. It is important to properly model this phenomenon for the optimal reconstruction of low p_T electrons.

Electron showers deposit their energy in several ECAL crystals. The reconstruction of electrons [38, 39] in CMS starts with the creation of “superclusters”, groups of one or more associated clusters of energy deposits, constructed using an algorithm which takes into account the magnetic field-induced energy spread. Superclusters are matched to track seeds (pairs or triplets of hits) in the inner tracker layers and electron tracks are built from these

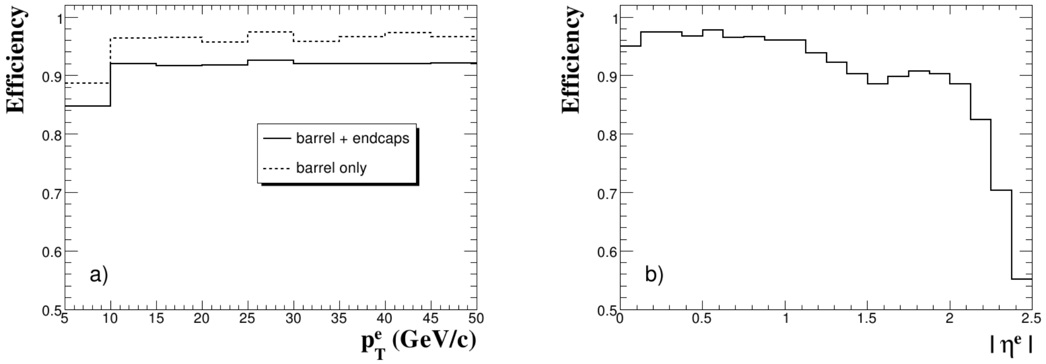


Figure 4.2: Electron track reconstruction efficiency [38] (a) as a function of p_T and (b) as a function of η, for electrons uniformly distributed in p_T between 5 and 50 GeV/c. In (a), the efficiency is shown averaged over the full ECAL barrel and endcaps η range (full line) and for the barrel only (dotted line).

seeds.

Track reconstruction begins from a segment in the pixel layer close to the beam pipe as a starting point for a combinatorial track-finding procedure [40] which outputs a number of candidates. Compatible hits on the next silicon layers are first searched for, then an extrapolation is performed, using a Bethe-Heitler [41] modeling of the electron losses and a Gaussian-Sum Filter (GSF) [42] in the forward fit. This procedure is iterated until the last tracker layer, unless no hits are found in two successive layers. If more than one compatible hit is found on a subsequent layer, many candidate trajectories are grown in parallel. A minimum of five hits is required to reconstruct a track.

The ECAL supercluster and the matching track information are then combined in order to improve overall precision. Figure 4.3 shows the effective RMS of the combined estimate, together with that of the standalone ECAL and tracker measurements. The addition of track momentum measurement improves significantly the measurement precision for

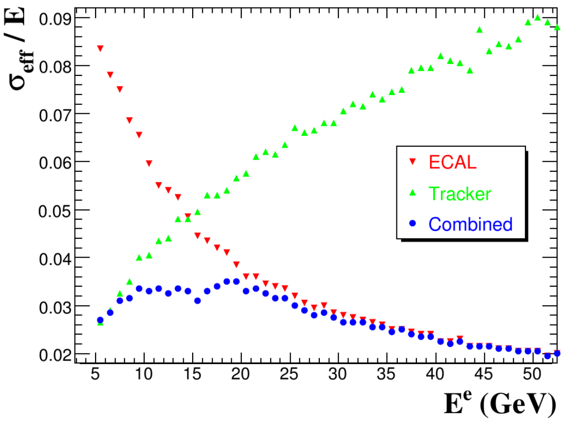


Figure 4.3: The resolutions as measured by the relative effective RMS of the corrected supercluster energy[38] (downward triangles), the reconstructed track momentum at origin (upward triangles), and of the combined electron momentum estimate (circles) as a function of the electron incident energy for electrons in the ECAL barrel.

electrons below 20 GeV.

4.6 Jet Reconstruction

As discussed (Section 4.2), due to QCD confinement the emitted partons produced in hard collisions form colorless hadrons which result in collimated particle showers moving along the direction of the originating parton, called jets. These hadrons deposit their energy on the CMS calorimeters through electromagnetic and hadronic interactions; the collected information is then used to produce detector-level jets.

4.6.1 The Particle Flow Algorithm

The Particle Flow (PF) reconstruction algorithm [43] attempts to identify all stable particles existing in an event, namely electrons, photons, charged hadrons, neutral hadrons

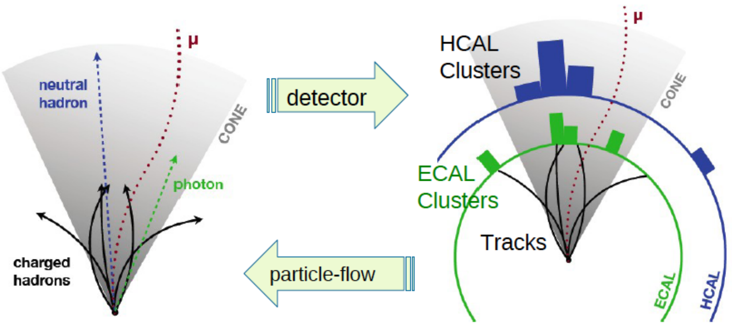


Figure 4.4: The Particle Flow algorithm. Particles in the CMS detector are seen as tracks and energy depositions. The PF algorithm attempts to fully reconstruct an event by combining information for all CMS subdetectors.

and muons by combining information from all CMS sub-detectors (Fig. 4.4). This list of particles is then used to form a more coherent picture of the event: to build jets (from which the energy and direction of the originating quark/gluon is deduced), and to determine the missing transverse energy (MET) which gives an estimate of the direction and energy of any neutrinos or other invisible particles in an event, etc. The PF building blocks are in the form of charged tracks, calorimeter clusters and muon tracks. The granularity and near-hermetic coverage of the CMS detector, together with a series of advanced tracking and clustering algorithms provide an excellent reconstruction performance. In addition, the use of the CMS tracker information improves the jet p_T resolution versus pure calorimetric jet reconstruction.

4.6.2 Jet Clustering Algorithms

Jet clustering algorithms can be classified into two major groups depending on the reconstruction approach; in the first group, cone-based algorithms take particles or calorimeter

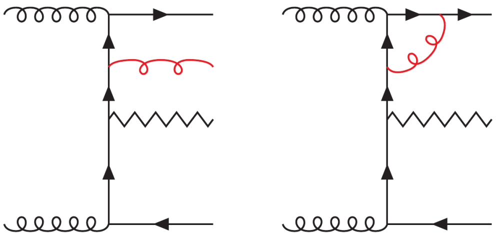


Figure 4.5: Emission of a real (left) and a virtual (right) gluon. If the jet reconstruction is not infrared safe and the clustering is different in the two cases, the result will not be consistent with pQCD computations.

towers as seeds in an iterative search for stable cones in an event. In the second group, a distance is defined between pairs of particles and successive recombinations of the pair of closest particles are performed, stopping when all resulting objects are too far apart.

The jet reconstruction algorithms must be insensitive to soft radiation in the event i.e. to be *infrared safe* (Fig. 4.5) in order to be consistent with perturbative QCD calculations. The jet-finding procedure should also find jets that are insensitive to any collinear radiation in the event, in the sense that the outcome is unchanged if the energy carried by a single particle is split among two collinear particles (*collinear safe*).

Three jet clustering algorithms are currently implemented and supported in the CMS software:

- k_t (cone size 0.4/0.6)
- seedless infrared safe (cone size 0.5/0.7)
- anti-k_t algorithm (cone size 0.5/0.7)

The anti-Kt algorithm which belongs to the second group has been used as it provides both infrared and collinear safety while being of equal performance in terms of processing speed as cone-based algorithms.

anti-k_t Algorithm

In the anti-k_t algorithm [44], one introduces distances d_{ij} between objects (particles, pseudojets) i and j and d_{iB} between object i and the beam (B). The smallest of the distances is calculated; if it is a d_{ij}, the objects i and j are recombined by adding their four-momenta, while if it is d_{iB} it is labeled as a jet and is removed from the list of objects. Distance measures are defined as:

$$d_{ij} = \min\left(k_{ti}^{2p}, k_{tj}^{2p}\right) \frac{\Delta_{ij}^2}{R^2} \tag{4.1}$$

$$d_{iB} = k_{ti}^{2p} \tag{4.2}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and k_{ti}, y_i and ϕ_i are the transverse momentum, rapidity and azimuth of particle i respectively. The parameter p governs the relative power of the energy versus geometrical (Δ_{ij}) scales. The anti-k_t algorithm is a sub case of the above set of equations where $p = -1$, while for $p = 0$ and $p = 1$ we get the so-called k_T and Cambridge/Aachen algorithms respectively.

Despite their appealing simplicity and long list of advantages, sequential clustering algorithms are not widely used at current hadron collider experiments, historically due to prohibitively growing CPU requirements for large numbers of input particles to be clustered per event. The CMS software however interfaces to the FastJet [45] package, which provides novel implementations of these algorithms, with reduced processing times.

4.6.3 Jet Energy Corrections

The calorimeters have a non-linear response to the energy deposited by particles, and as a consequence it is not straightforward to translate the measured energy to the true energy of the particle. To alleviate this problem, a set of *jet energy corrections* has been developed, which are applied to the measured jet energy depositions.

CMS has adopted a factorized solution to the problem of jet energy corrections, where each level of correction takes care of a different effect. The set of corrections used for this thesis is as follows:

L1 Offset Correction : The goal of the L1 correction is to remove the energy from pileup events. In principle, this will remove any dataset dependence on luminosity so that the following corrections are applied upon a luminosity independent sample. The data-driven FastJet [46] pileup subtraction algorithm has been used, which is based on estimating the pileup and underlying event transverse momentum density using the calculated jet area. Figure 4.6 shows the effect of the L1 pileup removal correction.

L2 Relative Jet Correction : The goal of the L2 Relative correction is to make the jet response flat versus η. Essentially, the uniformity in pseudorapidity is achieved by correcting a jet in arbitrary η relative to a jet in the central region ($|\eta| < 1.3$).

L3 Absolute Jet Correction : The goal of the L3 Absolute correction is to make the jet response flat versus p_T. Once a jet has been corrected for η dependence, it is corrected back to particle level, i.e. the corrected jet p_T is equal on average to the generator-level jet p_T.

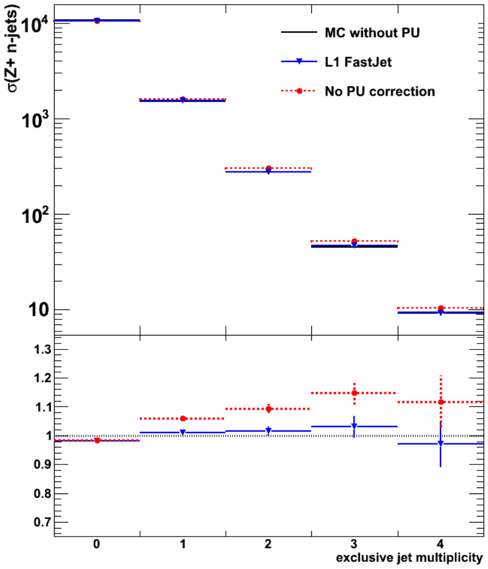


Figure 4.6: Effect of pileup corrections. With the straight line, a Z Monte Carlo sample without pileup is shown. The dashed line with points is the same sample, after the addition of pileup and the continuous line with arrows is after applying the L1 FastJet corrections for pileup removal. The plot on the bottom shows the ratios with the no pileup sample before and after L1 corrections.

4.7 Full Simulation

To perform efficiently all the steps described above, the CMS collaboration has developed a C++-based software framework called CMSSW. The framework takes care of interfacing with different event generators depending on the studied physics process, then passing the information to the detector simulation and digitization software to produce an output file in a format similar to the actual detector readout. On this information the event reconstruction code is executed and produces containers of electrons, muons, jets and other event information to be used for analysis. This chain is referred-to as “full simulation” with the detector simulation being the most time and CPU-intensive step.

Real stupidity beats artificial intelligence every time.

Hogfather
TERRY PRATCHETT

Chapter 5

Event Selection

In order to find events of interest in the recorded collision data, a series of selection requirements is applied. The CMS trigger provides a first level of filtering, followed by a series of quality cuts on the offline reconstructed objects to further improve the sample purity.

5.1 Online Selection

Throughout the 2010 running period, triggers have been changing and adapting to the higher luminosities delivered by the LHC (the CMS trigger system has been described in Section 3.2.6). To extract Z+Jets events, a set of triggers corresponding to the lowest-threshold unprescaled single electron trigger available was used. To achieve the highest possible efficiency and event yield, photon triggers were preferred for the early LHC running period. For later runs, electron triggers of increasing threshold and quality cuts were introduced.

For the first $8.3\,\mathrm{pb}^{-1}$ of data the delivered instantaneous luminosity allowed for a lower L1 trigger threshold. The L1_SingleEG5 path triggers on events containing one electromagnetic object above the set $E_T > 5\,\mathrm{GeV}$ threshold. An increase in the collision rate led us to move to the higher-threshold L1_SingleEG8 path as a L1 seed for the high-level

Run range	High Lever Trigger	Level-1 seed
$<$ 138000	HLT_Photon10_L1R	L1_SingleEG5
138000 - 141882	HLT_Photon15_Cleaned_L1R	L1_SingleEG5
141882 - 144114	HLT_Ele15_SW_CaloEleId_L1R	L1_SingleEG5
146428 - 147116	HLT_Ele17_SW_CaloEleId_L1R	L1_SingleEG5
147196 - 148058	HLT_Ele17_SW_TightEleId_L1R	L1_SingleEG8
148819 - 149064	HLT_Ele17_SW_TighterEleIdIsol_L1R_v2	L1_SingleEG8
149181 - 149442	HLT_Ele17_SW_TighterEleIdIsol_L1R_v3	L1_SingleEG8

Table 5.1: Electron triggers used to extract Z+Jets events.

HLT type	H/E	$\delta\eta_{in}$	$\delta\phi_{in}$	$\sigma_{i\eta i\eta}$
CaloEleId	0.15	-	-	0.014 (0.035)
TightEleId	0.15	0.01 (0.01)	0.08 (0.08)	0.012 (0.032)
TigherEleIdIsol	0.05	0.008 (0.007)	0.1 (0.1)	0.011 (0.031)

Table 5.2: High-level trigger identification requirements by path. Thresholds for barrel (endcap, if it differs)

electron triggers we used in this analysis.

The HLT selection was also tightened as data rates were increasing in order to keep the recorded data rates manageable. Initially, the applied HLT_Photon10_L1R path required the presence of an ECAL deposit without any track matching required. A similar requirement with a higher threshold exists on the HLT_Photon15_Cleaned_L1R trigger path, with an extra precondition demanding the ratio of the E_T of the most energetic crystal over the $\Sigma\left(E_T\right)$ of a 3×3 region around it, to be less than 98%, in order to reject triggers originating from noisy ECAL channels. This was part of all subsequent triggers. After run #146428 more complex identification and isolation quality requirements were applied to the HLT objects. These are listed in Tables 5.2 and 5.3. Table 5.1 lists all High-Level triggers used, along with their respective Level-1 trigger seeds.

HLT type	Iso_{Ecal}/p_T	Iso_{Hcal}/p_T	Iso_{Track}/p_T
CaloEleId	-	-	-
TightEleId	-	-	-
TigherEleIdIsol	0.125 (0.075)	0.05	0.15 (0.1)

Table 5.3: High-level trigger isolation requirements by path. Thresholds for barrel (endcap, if it differs)

5.1.1 Trigger Efficiency

To estimate trigger efficiency two methods have been developed; one involved using a dedicated L1 trigger bit and a Tag and Probe (T&P) method (Section 6.5). The first was needed to estimate trigger efficiency in the first 100^{-1} pb of collected data when statistics for T&P were low and to cross-check the validity of early T&P results. The trigger efficiency numbers used in this analysis as well as electron reconstruction, identification and isolation, have been calculated with Tag and Probe.

L1 ECAL Activity Trigger

In the early days of LHC running, trigger efficiency was monitored through the usage of specially designed High Level triggers, part of the EGMonitor primary dataset [47]. The ECAL activity triggers HLT_Activity_Ecal_SCx, with x = 7, 15, 17 were seeded from minimum bias triggers, based on information from the beam scintillation counters. The raw data information in ECAL barrel, endcap and preshower detectors was used to reconstruct offline-quality superclusters. The trigger fired if at least one supercluster was above threshold.

For events that fired one of these ECAL activity triggers, the efficiency was calculated based on the formula:

$$\frac{\text{Electron candidates that pass event selection and fire the trigger under study}}{\text{Electron candidates that pass event selection}} \tag{5.1}$$

Electron candidates were taken from the standard offline electron collection. A series of quality cuts was applied (listed in Table 5.5) corresponding to the Working Point 80 described in detail in section 5.2.

An offline electron candidate was considered as having fired the trigger under study when it could be matched to a trigger object that passed the respective trigger E_T threshold. Matching was performed using a $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ cone. For L1 (HLT) triggers, $\Delta R = 0.5(0.2)$ was used as cone radius. Minimum Bias and a $Z \rightarrow e^+e^-$ Pythia Monte Carlo samples have been used to verify consistency between data and MC. Calculated ratios are reported in Table 5.1.1 and within statistical error are in excellent agreement with unity.

Efficiencies for the L1, HLT and HLT given a L1 trigger were calculated versus supercluster E_T (Figure 5.1), supercluster η and ϕ (Figure 5.2). Differences in efficiency between the ECAL barrel and endcaps are expected due to extra energy deposited in the endcap region originating from the proton remnants, and are studied separately. For the η and ϕ efficiency plots, the electrons were required to have a supercluster $E_T > 20$ GeV (same as the offline selection) to ensure only electrons above the turn-on curve of 5(15) GeV for L1(HLT) are considered. In the kinematic region of interest the trigger shows 100% efficiency within statistical uncertainties.

5.2 Offline Electron Selection

The selection of a clean electron collection is important in analyses with electron final states. For early LHC data taking, CMS has favored a robust, cut-based approach

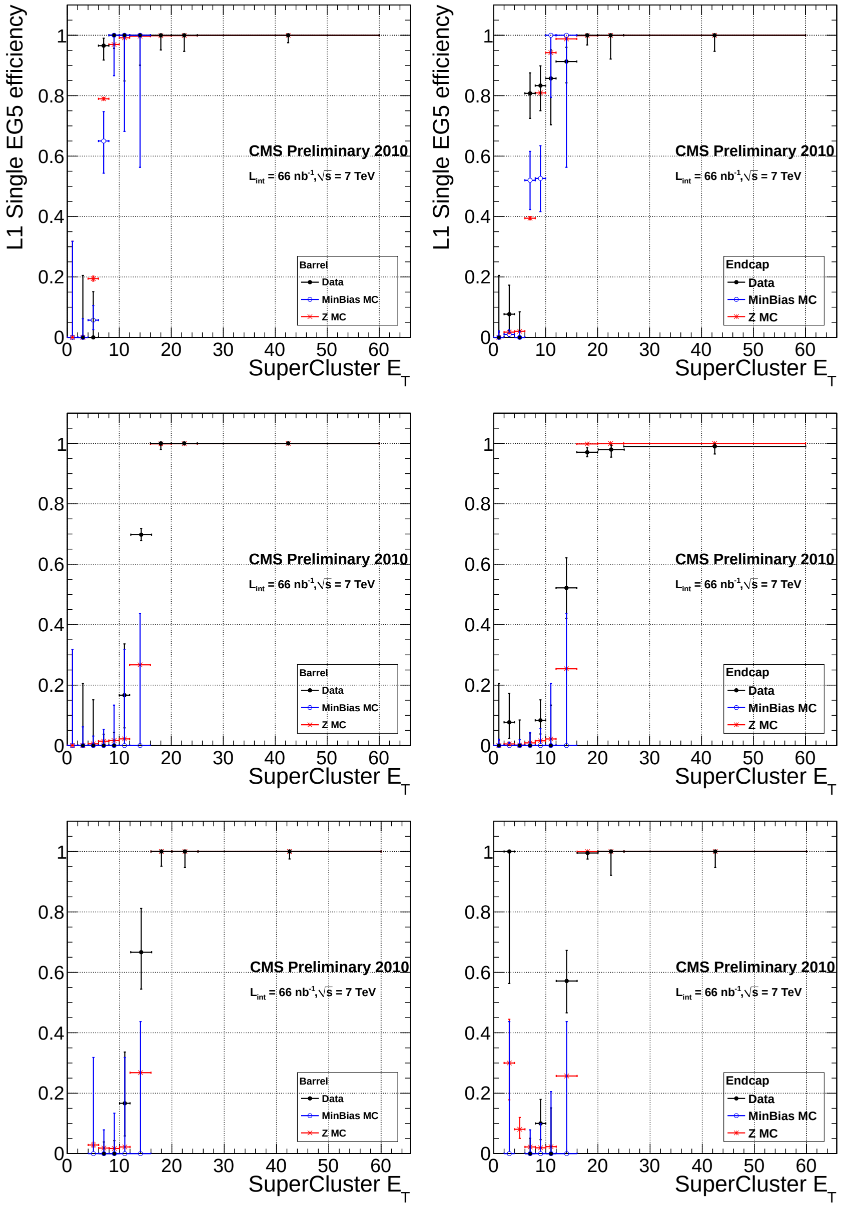


Figure 5.1: Efficiency turn-on curves as a function of supercluster E_T for electron candidates in runs 138564 - 140181, for L1_SingleEG5 (top), HLT_Photon15_Cleaned_L1R (middle), and HLT_Photon15_Cleaned_L1R with additional condition that L1_SingleEG5 has fired (bottom). The left (right) column shows candidates in the ECAL barrel (endcap) only.

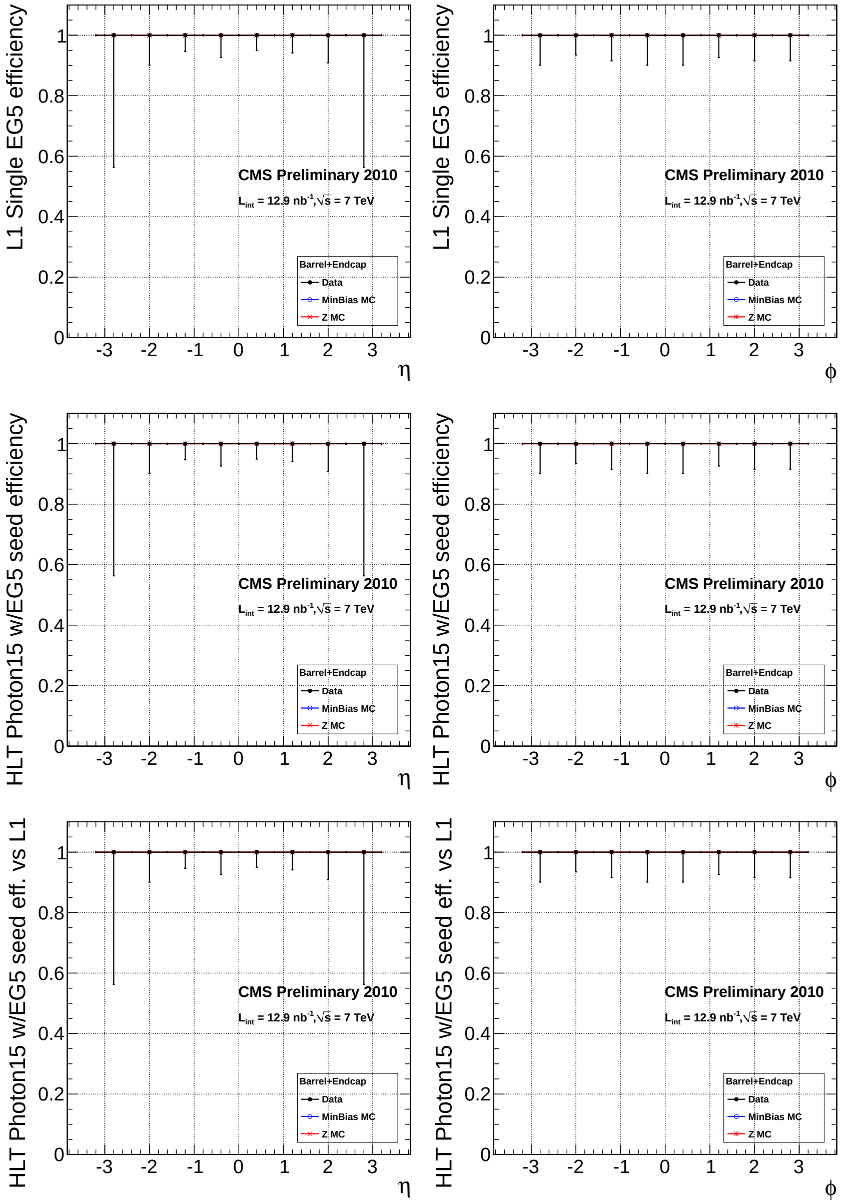


Figure 5.2: Efficiency turn-on curves as a function of supercluster η (left) and ϕ (right) 132440 - 137028, for L1_SingleEG5 (top), for HLT_Photon10_L1R with additional condition that offline a supercluster be found with $E_T > 15$ GeV (middle), for HLT_Photon10_L1R with additional conditions that offline a supercluster be found with $E_T > 15$ GeV and that L1_SingleEG5 has fired (bottom). Only electron candidates with $E_T > 20$ GeV of their corresponding supercluster are considered.

	c_ϵ(L1)	c_ϵ(HLT|L1)	c_ϵ(L1) × c_ϵ(HLT|L1)	c_ϵ(L1+HLT)
EB, 20-25 GeV	$1.000^{+0.}_{-0.009}$	$1.000^{+0.}_{-0.019}$	$1.000^{+0.}_{-0.021}$	$1.000^{+0.}_{-0.015}$
EB, 25+ GeV	$1.000^{+0.}_{-0.004}$	$1.000^{+0.}_{-0.006}$	$1.000^{+0.}_{-0.007}$	$1.000^{+0.}_{-0.006}$
EE, 20-25 GeV	$0.962^{+0.018}_{-0.026}$	$1.000^{+0.}_{-0.015}$	$0.962^{+0.018}_{-0.030}$	$0.982^{+0.013}_{-0.026}$
EE, 25+ GeV	$0.986^{+0.008}_{-0.013}$	$0.991^{+0.006}_{-0.013}$	$0.977^{+0.010}_{-0.018}$	$0.990^{+0.008}_{-0.015}$
EB+EE, 20-25 GeV	$0.985^{+0.007}_{-0.010}$	$1.000^{+0.}_{-0.008}$	$0.985^{+0.007}_{-0.013}$	$0.992^{+0.006}_{-0.011}$
EB+EE, 25+ GeV	$0.995^{+0.003}_{-0.004}$	$0.997^{+0.002}_{-0.005}$	$0.992^{+0.004}_{-0.006}$	$0.996^{+0.003}_{-0.005}$

Table 5.4: Data/MC ratios for the efficiency of L1_SingleEG5 (L1), HLT_Photon15_Cleaned_L1R (L1+HLT) and for HLT_Photon15_Cleaned_L1R with additional condition that L1_SingleEG5 has fired $(\mathrm{HLT}|\mathrm{L1})$, as a function of the transverse energy of the electron supercluster and of the detector region. Individual HLT correction factors were obtained from a 64 nb^{-1} subset of the collision data collected between runs 138046 and 139459. Individual L1 and combined L1+HLT corrections are computed using roughly 71 nb^{-1} available between runs 138046 and 140174.

due to its simplicity and good efficiency. Ultimately, multi-variate techniques may provide higher performance, but a cut-based selection can be a useful tool to understand the data and compare them directly with the Monte Carlo. These cuts have been optimized in order to retain the signal on inclusive $W \rightarrow e\nu$ events in Monte Carlo samples [48] with various degrees of efficiency, called Working Points. For this analysis an asymmetric selection is applied on the two electrons originating from the Z boson; Working Point 80 is used for the tight leg and the looser Working Point 95 requirements are applied to the second electron.

The selection variables (Table 5.5) are categorized in three groups and in general have different values for barrel ($|\eta| < 1.4442$) and endcap ($1.566 < |\eta| < 2.5$) electrons. The region $1.4442 < \eta < 1.566$ is the EB-EE transition region and is excluded due to decreased reconstruction efficiency as explained in Section 3.2.3. In the next few paragraphs, the details of each variable are presented, as well as plots showing the effect of applying the WP80 selection to signal and background Monte Carlo electrons. The plots presented show the effect of each cut after applying all previous ones and demonstrate the gradual reduction of the various background processes.

5.2.1 Electron identification

The electron identification [49] makes use of a set of variables in order to distinguish between electrons coming from a Z decay from background electrons. The variables used are:

- The calorimeter shower shape variable $\sigma_{i\eta i\eta}$ which describes the width of the ECAL cluster along the η direction computed for all the crystals in the 5×5 block of crystals centered on the highest energy crystal of the seed cluster. Electrons coming from a Z boson have a small spread in η.

	WP95		WP80	
	Barrel	Endcap	Barrel	Endcap
$\sigma_{i\eta i\eta}$	0.01	0.03	0.01	0.03
$\Delta\phi_{in}$	n/a	n/a	0.06	0.03
$\Delta\eta_{in}$	0.007	0.01	0.004	0.007
H/E	0.15	0.07	0.04	0.025
Track iso	0.15	0.08	0.09	0.04
ECAL iso	2.0	0.06	0.07	0.05
HCAL iso	0.12	0.05	0.10	0.025
Missing hits $\leq$	1	1	0	0
$\Delta cot(\theta)$	n/a	n/a	0.02	0.02
Dist	n/a	n/a	0.02	0.02

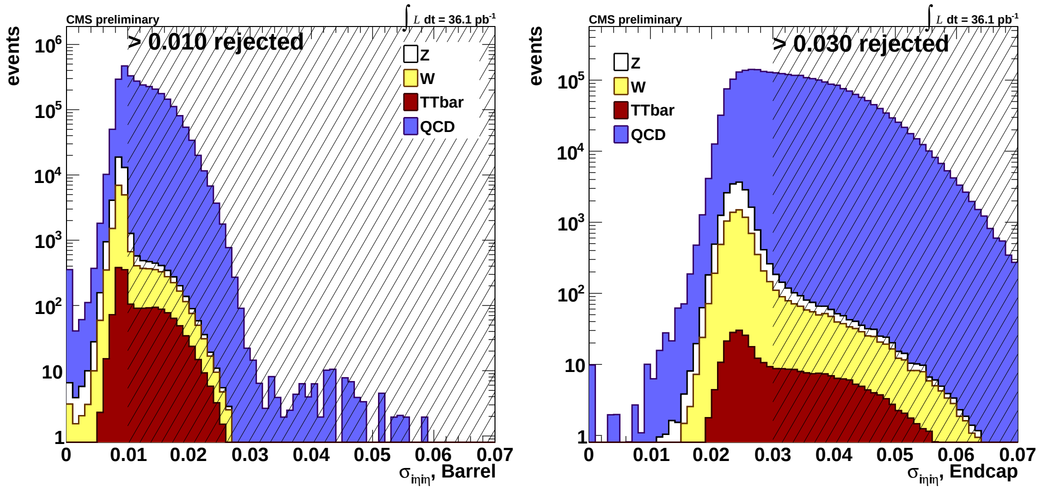
Table 5.5: Electron selection variables

- The $\Delta\phi_{in}/\Delta\eta_{in}$ variables describe the geometrical matching between the electron track direction at the vertex extrapolated to the supercluster and the measured supercluster position. There should be a good agreement between the two.
- Finally, the fraction of the electron energy deposited in the hadronic calorimeter (H/E) is expected to be small. The steep falloff at $(H/E) = 0.15$ at Figure 5.6 comes for generation-level requirements during MC production in order to reduce sample size, as the offline cut is much tighter.

5.2.2 Electron isolation

Electrons coming from a Z boson decay are expected to be fairly isolated. Three isolation variables are used to evaluate this:

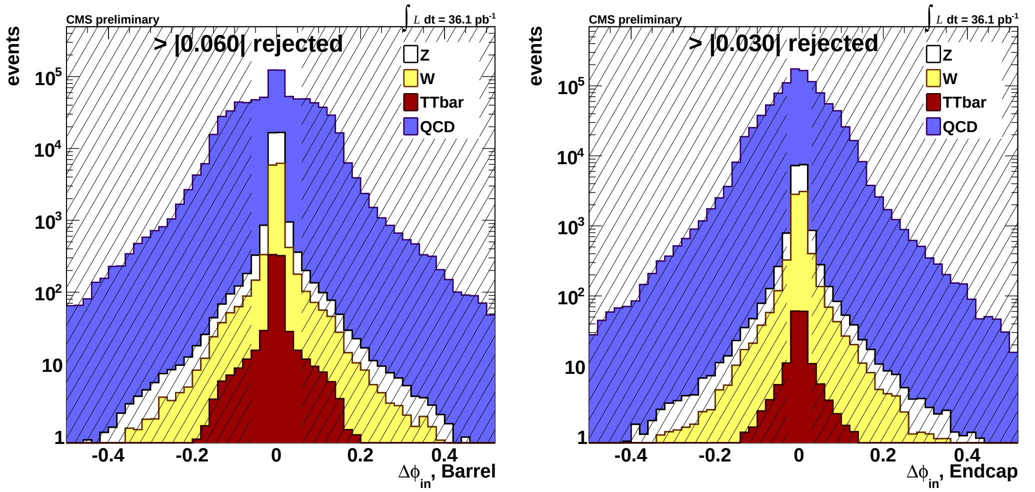
Tracker isolation : The sum of the transverse momentum of tracks with $p_T > 1.0\,\text{GeV}$ in a $0.04 < \Delta R < 0.3$ annulus centered around the track direction at the vertex.



(a) Barrel

(b) Endcap

Figure 5.3: Electron Identification : $\sigma_{i\eta i\eta}$

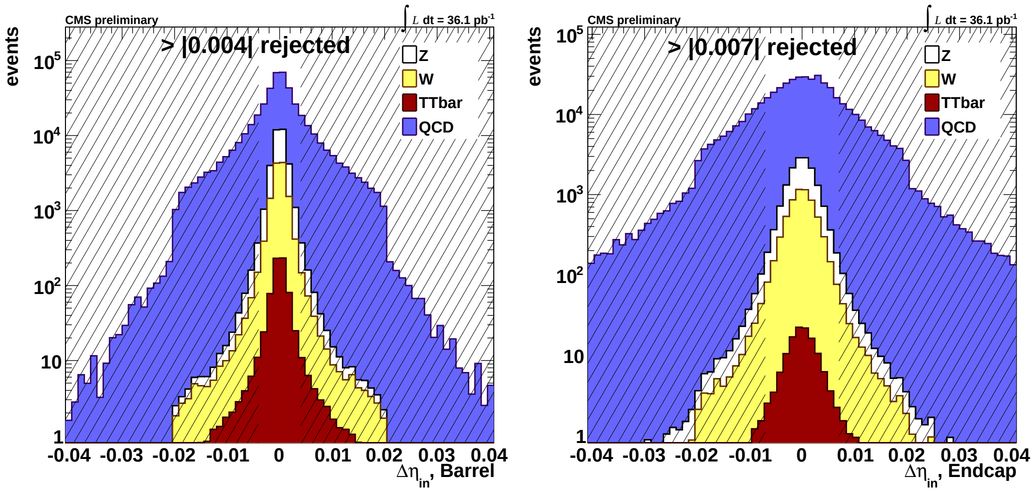


(a) Barrel

(b) Endcap

Figure 5.4: Electron Identification : $\Delta\phi_{in}$

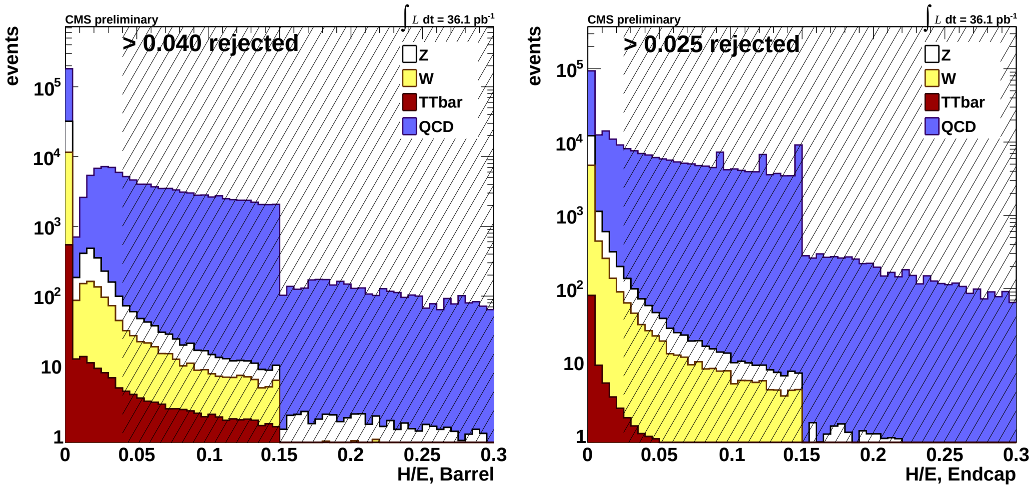
ECAL isolation : is the sum of the ECAL reconstructed hits inside a a $\Delta R < 0.3$ cone around the position of the ECAL supercluster. The energy of the electron energy is



(a) Barrel

(b) Endcap

Figure 5.5: Electron Identification : $\Delta\eta_{in}$



(a) Barrel

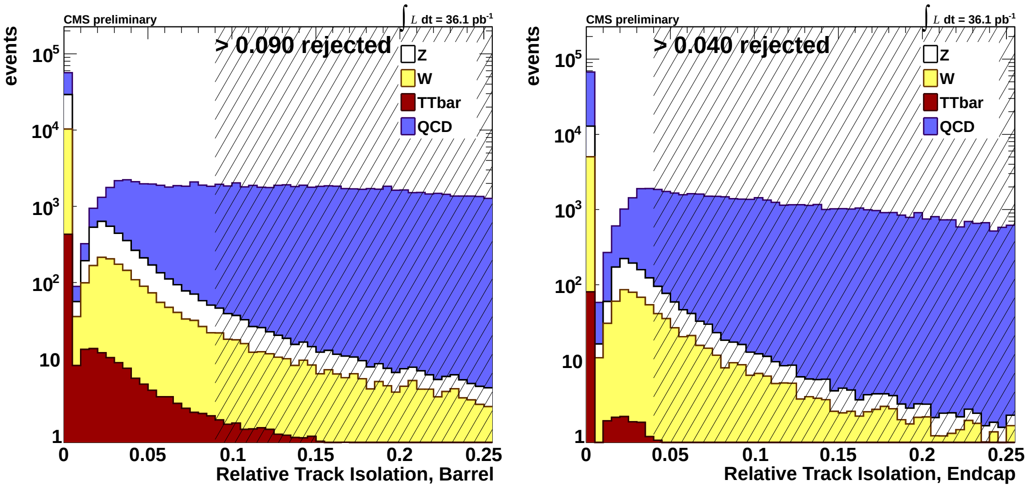
(b) Endcap

Figure 5.6: Electron Identification : (H/E)

removed by applying an inner veto cone of three crystals and a strip of three crystals width in η. Hits of energy < 0.08 GeV in the barrel, or $E_T < 0.14$ GeV in the endcap, are not considered in the sum to reduce noise contributions.

HCAL isolation : The E_T sum of HCAL towers surrounding the electron in an annulus of $0.15 < \Delta R < 0.3$.

The cut is applied on the respective *relative isolations* by dividing each one with the electron transverse energy.



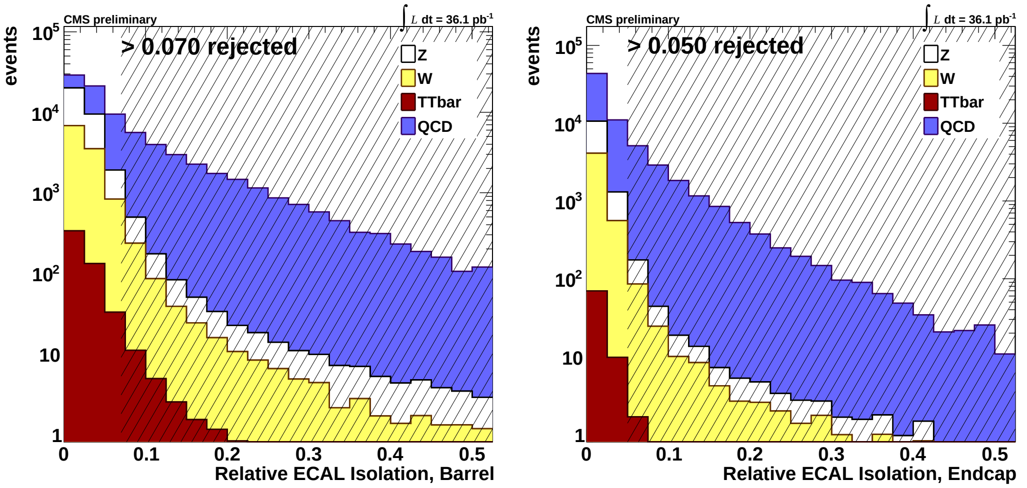
(a) Barrel

(b) Endcap

Figure 5.7: Electron Isolation : Track

5.2.3 Conversion rejection

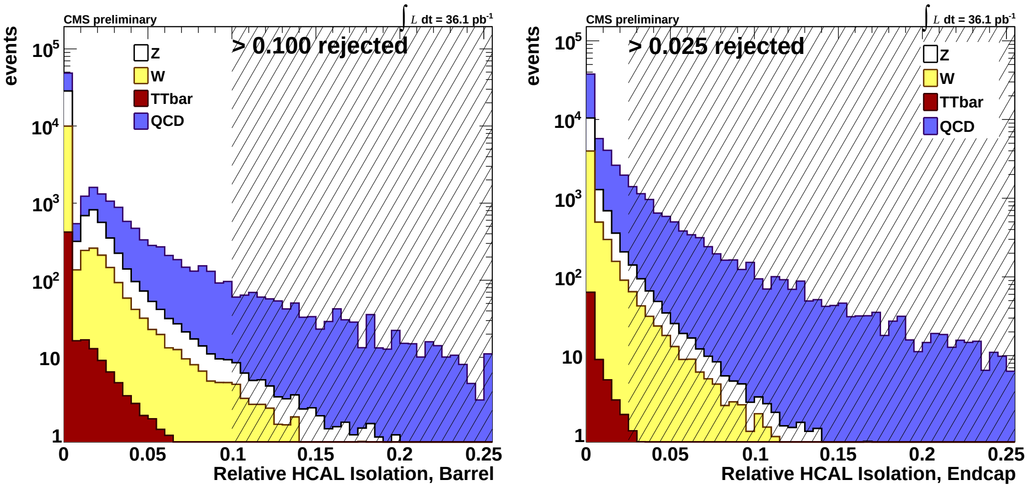
Electrons from converted photons are a non-negligible background to prompt electrons coming from the Z boson decays [48]. In order to reject such backgrounds, electrons with missing expected hits in front of the innermost valid track hit are rejected as originating from a conversion that occurred in the tracker material. The general track collection is also inspected to locate possible conversion partner tracks. To be identified as a conversion partner, the track must:



(a) Barrel

(b) Endcap

Figure 5.8: Electron Isolation : ECAL



(a) Barrel

(b) Endcap

Figure 5.9: Electron Isolation : HCAL

- have opposite sign as the electron track
- approximately the same $\Delta cot\left(\theta\right)$ as the electron track

- small distance (“Dist”) in the $R - \phi$ plane, where “Dist” is defined as the distance in the x-y plane between the two tracks when the track in question and the electron GSF track would be parallel when extrapolated. All neighboring tracks in a $\Delta R = 0.3$ cone are considered.

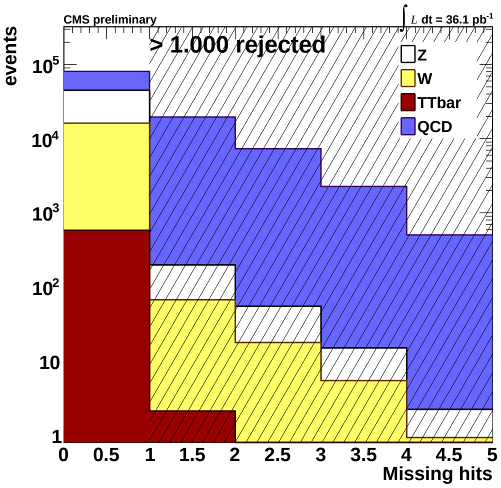
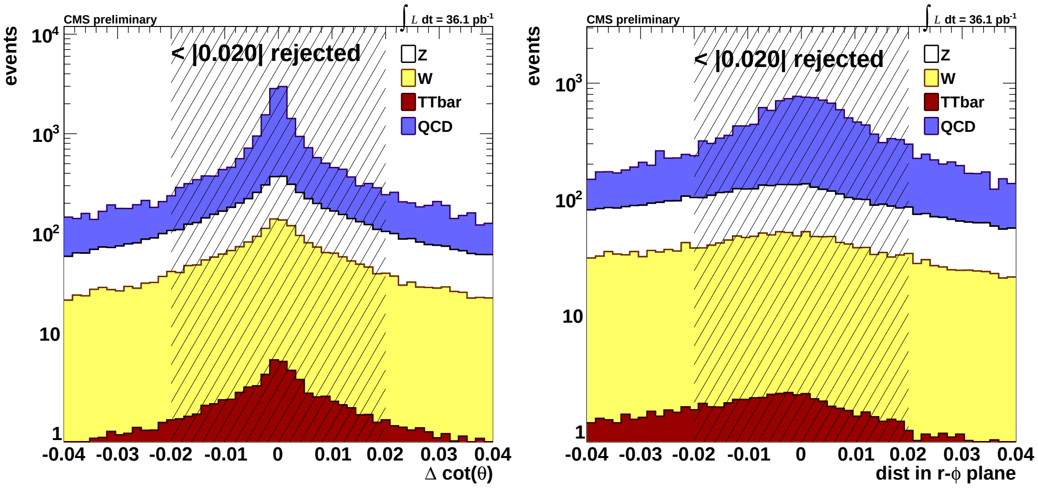


Figure 5.10: Conversion Rejection : Missing Hits

5.2.4 Electron distributions after cuts

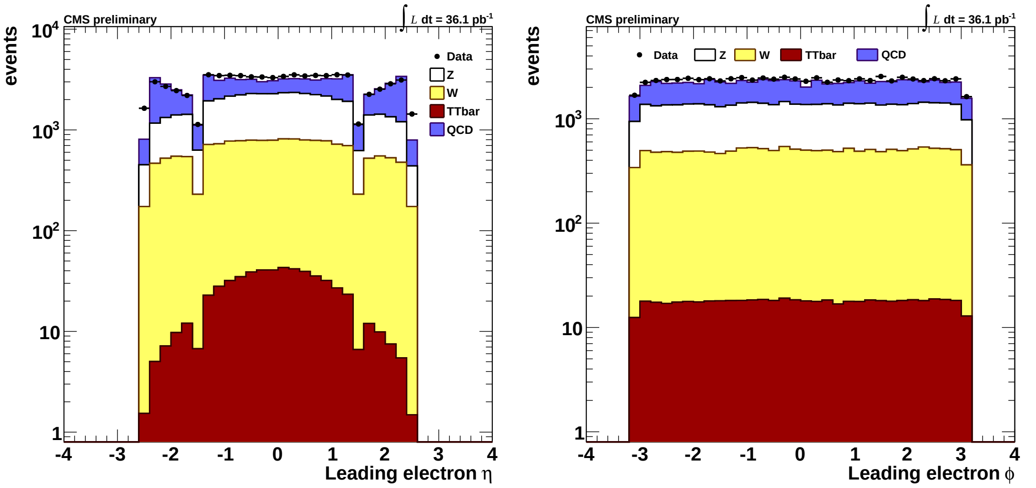
The application of the full Working Point 80 selection leaves a set of events where most of the QCD background which was dominating by a few orders of magnitude has been suppressed, while most of the signal has been retained. The η, ϕ and p_T distributions of the leading WP80 electron are plotted in Figures 5.12 and 5.13. The agreement between the data points and Monte Carlo predictions are very good in the ϕ case while for η more centrally-produced electrons are favored in the data. The electron p_T plot shows good agreement at the kinematic region between 30-60 GeV where we expect to find most electrons coming from Z decays (around half the Z mass). Below and above that region we see some disagreement



(a) $\Delta cot\left(\theta\right)$ (b) "Dist"

Figure 5.11: Conversion Rejection : $\Delta cot\left(\theta\right)$ and "Dist"

between data and Monte Carlo predictions, attributed to the poor QCD modeling. These disappear once the invariant mass of the dielectron pair is required to be close to the Z mass.



(a) (b)

Figure 5.12: Leading electron η, ϕ plots after application of WP80 requirements

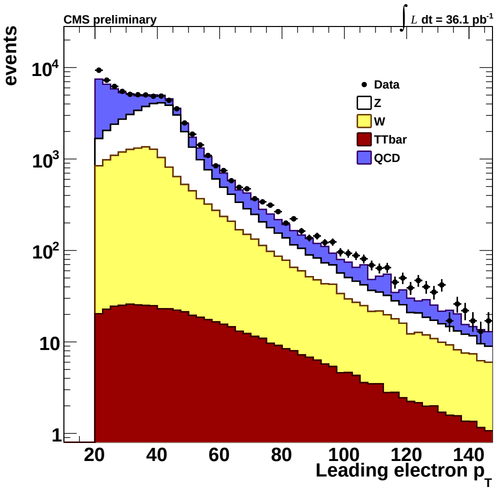


Figure 5.13: Leading electron p_T plot after application of WP80 requirements

5.3 Offline Jet Selection

Jet clusterization algorithms as well as the jet energy corrections applied have been already described in section 4.6.2. After corrections, the jet collection may still contain fake, badly reconstructed jets or jets originating from noisy areas of the detector. To discard these unwanted jets, while retaining most of the real ones, a set of jet identification criteria has been defined which keep $> 99\%$ of the real jets while removing a significant fraction of unwanted ones [50]. To populate the jet collection used in this analysis, the CMS-recommended loose set of criteria has been used by applying the following requirements:

- $p_{T,corr.} > 10\,\text{GeV}$, $|\eta| < 2.4$
- Charged hadron fraction > 0.0

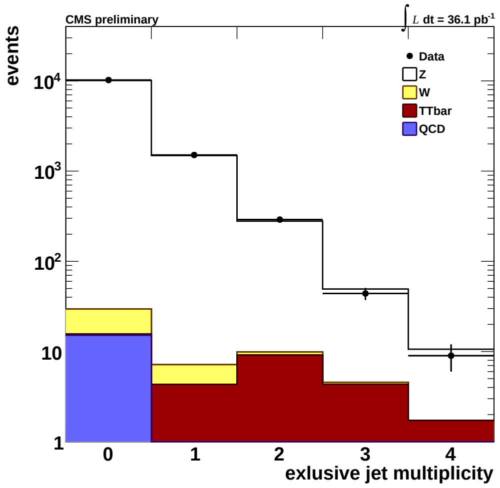


Figure 5.14: Exclusive jet multiplicity for $p_{T,corr.} > 30\,\mathrm{GeV}$ jets. Jet energy corrections (L1+L2+L3) and loose jet ID requirements have been applied.

- Neutral hadron fraction < 0.99
- Charged multiplicity > 0.0
- Charged electromagnetic fraction < 0.99
- Neutral electromagnetic fraction < 0.99

Electrons are also included in the jet list provided by the framework, therefore it has to be “cleaned” to remove the electrons identified as originating from the Z boson decay. This is achieved by means of requiring a minimum ΔR separation of 0.5 between the jets added in the final collection and the Z electrons.

For the cross section ratio calculations, only $p_{T,corr.} > 30\,\mathrm{GeV}$ jets are considered; in Figure 5.14 we can see how the exclusive jet multiplicity for data compares with the Monte Carlo predictions after applying the the loose jet ID requirements as described above. Agreement is very good for up to 2-jet events and within statistical uncertainty for the higher jet multiplicity bins. Although not shown in the plots above, jets of lower p_T are also kept in the collection as they are needed for the unfolding procedure which will be described later.

Anything that happens, happens.
Anything that, in happening, cause
something else to happen, causes
something else to happen. Anything
that, in happening, causes itself to
happen again, happens again. It
doesn't necessarily do it in
chronological order though.

Mostly Harmless
DOUGLAS ADAMS

Chapter 6

Analysis

6.1 Data

The first LHC proton-proton collisions at 7 TeV took place on March 30th, 2010 at 13:06 local Geneva time, marking the start of the LHC research program. Since then and until October 29th, 2010 the LHC produced collisions with increasing luminosity. A short lead ion collisions run followed until the end of the year. During the 7-month running period, the LHC delivered approximately 47 pb^{-1} of p-p collision data with CMS recording 43 pb^{-1} (Fig. 6.1). From the full dataset, 36 pb^{-1} were certified for electron analyses by the CMS data quality group. The rest was discarded for various reasons, like subdetectors being faulty or noisy and unstable beam conditions.

6.2 Monte Carlo Samples

To model the data, a set of MadGraph v. 4.4.12 [16] and PYTHIA v. 6.4 [31] Monte Carlo samples provided by the CMS generator group was used. The signal is modeled with a MadGraph Drell-Yan [9] plus up to four generated jets sample where the mass of the generated boson is required to be above 50 GeV/c^2 and it is restricted to leptonic de-

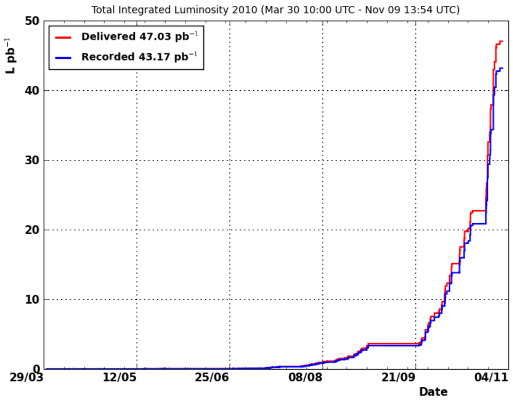


Figure 6.1: CMS recorded data for 2010. The higher line represents data delivered by the LHC and the lower line data recorded by the CMS experiment.

cays only. The inclusive Drell-Yan to e^+e^- PYTHIA sample was also evaluated. For the MadGraph samples both Z2 [34] and D6T [35] tunes were available and studied, whereas for the PYTHIA sample the Z2 tune was used. Electroweak backgrounds are modeled by MadGraph W+jets and $t\bar{t}$+jets samples, accompanied by binned PYTHIA samples modeling the expected QCD background. Unless otherwise noted, the samples used include pileup corresponding to the expected pileup in 2010 collision data. For the electroweak samples, next-to-leading order cross sections calculated either with FEWZ (W/Z) [10] or MCFM ($t\bar{t}$) [51] were used to scale them to the data luminosity. The full list of Monte Carlo samples and their cross-sections is summarized on Table 6.1.

6.3 Reconstructing Z candidates

Starting with 15041836 selected by the trigger, we apply a robust selection in order to maximize event yield. Requiring the presence of two reconstructed electrons with p_T >10 GeV and within the tracker acceptance, 40672 events remain. Next, an asymmetric selection approach is used to reconstruct the Z bosons. To achieve this, two electron collections are created; a "tight" collection with electrons fulfilling the WP80 criteria as defined on Section 5.2 and $p_T > 20\,\text{GeV/c}$ and a collection of "loose" electrons satisfying the WP95 requirements and $p_T > 10\,\text{GeV/c}$. Applying electron ID and requiring at least one candidate in each collection leaves us with 15649 events, the application of electron isolation allows 15001 events to survive and the conversion rejection discards 502 more events, leaving 14499 events. The choice was made to restrict electrons to ones coming from the ECAL-driven reconstruction algorithm (Sec. 4.5) as they are currently better understood than the tracker-driven electrons provided by Particle Flow (see section 4.6.1). This discards a minimal number of events found only by the PF algorithm, with 14286 remaining events.

The primary vertex (PV) of the event is assumed to be the one closest to the tight electron of the pair. To ensure that the two electrons originate from the same vertex, a maximum transverse impact parameter requirement of 0.035 mm is applied. This is defined as the distance in the z-plane of the primary vertex, and the vertex pointed by each of the electrons used to construct the dielectron candidate:

$$\delta_{xy} = \frac{(V_{y,el} - PV_y)P_{x,el} - (V_{x,el} - PV_x)P_{y,el}}{\sqrt{P_{x,el}^2 + P_{y,el}^2}} \tag{6.1}$$

where $V_{x,el}$, $V_{y,el}$ the vertex pointed by the track associated to the electron, $P_{x,el}$, $P_{y,el}$ the electron momentum in the respective axes and PV_x, PV_y the primary vertex position. The PV requirement leaves 13665 events pass through.

In order to keep the efficiency calculations simple, only the tight electron is used for trigger purposes and is required to match the HLT object that triggered the event within $\Delta R < 0.2$, with 13394 events.

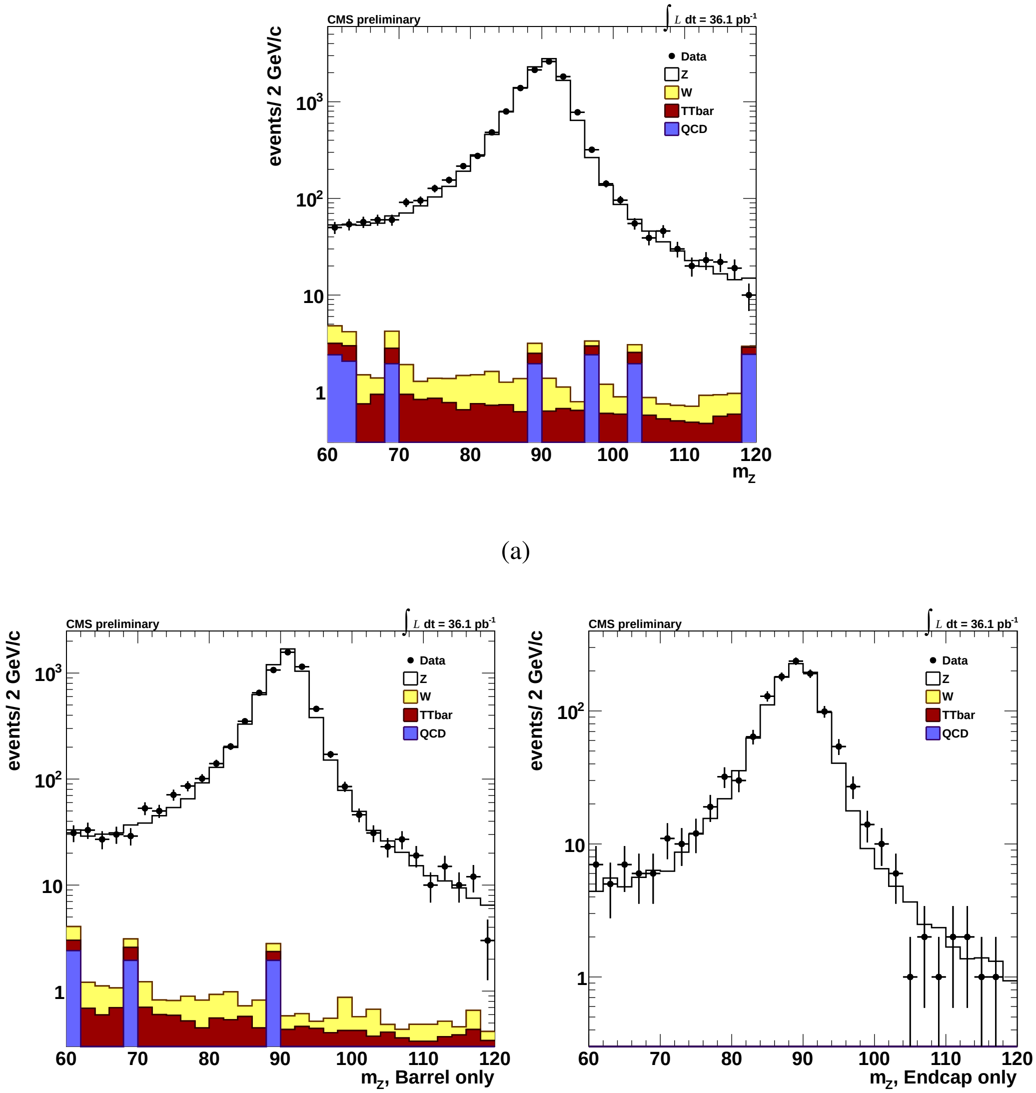
To ensure the absense of muon contamination, events containing a reconstructed muon of $p_T > 15\,\mathrm{GeV}/c$ within $|\eta| < 2.4$ are discarded. This requirement has no effect at all, discarding no events, and was put in place to keep the Z+Jets analysis in sync with the W+Jets selection as the same event yields were also used to calculate cross section ratios of W/Z+jets.

In the end, the invariant mass of the electron pair is calculated and the event is kept if it is found to be between 60 and 120 GeV/c^2. The number of cases where more than one dilepton candidates are found is very small ($<< 1\%$). In the rare case where this happens, the one with the highest scalar electron $\sum p_T$ is kept. The full selection gives us 12071 Z events with the obtained m_{ee} spectra after selection shown in Fig. 6.2 without any jet requirements applied. In the neighboring figures, the barrel-only and endcap-only electron contributions to the total number of events are shown. In Figure 6.3 the p_T and rapidity of the reconstructed Z boson are plotted. In all cases where statistics are sufficient, agreement with the Monte Carlo predictions is very good. The effect of the applied requirements and the number of events surviving after each step for the data and Monte Carlo samples is summarized on Table 6.2 for inclusive Z production. For the numbers quoted in the table for the MC background samples, a preselection of requiring the presence of two reconstructed electrons was applied in order to reduce sample size and facilitate analysis.

The final raw event counts after the full Z selection in inclusive jet bins are shown on Table 6.3 for events with Particle Flow jets of $p_T > 30\,\mathrm{GeV}/c$.

6.4 Fitting method

To fit the Z mass peak, a Z Monte Carlo lineshape as given by PYTHIA was convolved with a Crystal Ball function enhanced by the addition of an extra Gaussian and an exponen-



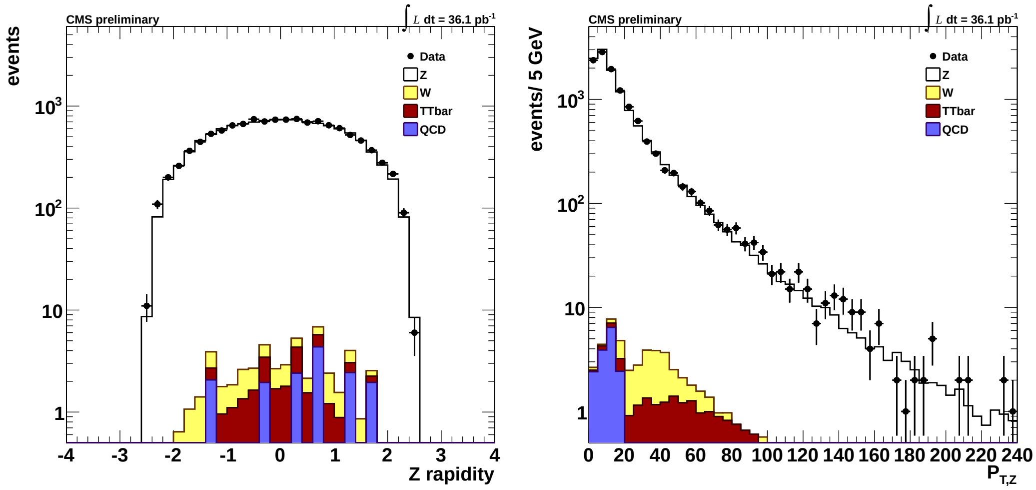
(a)

(b)

(c)

Figure 6.2: Di-electron invariant mass after full selection for the inclusive Z case (a). Plots (b) and (c) show the respective contributions of barrel-barrel and endcap-endcap electrons.

tial function. The Crystal Ball consists of a Gaussian portion and a power-law tail at the low



(a)

(b)

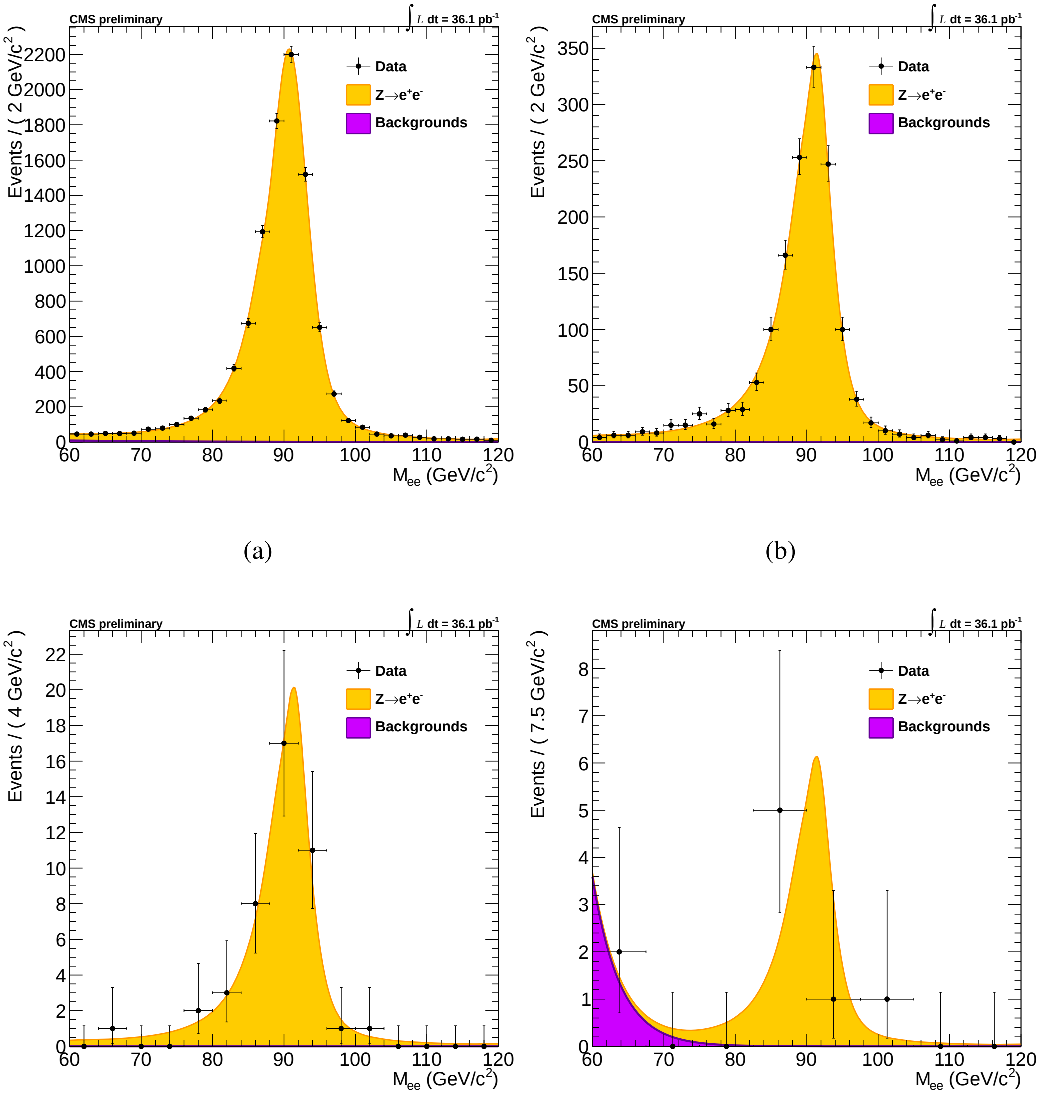
Figure 6.3: (a) p_T and (b) rapidity distributions after full selection in the inclusive $Z \rightarrow e^+e^-$ case.

end. Its analytical form is:

$$f(x;\alpha,n,\bar{x},\sigma) = \begin{cases} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right) & \frac{x-\bar{x}}{\sigma} > -\alpha \\ A\left(B - \frac{x-\bar{x}}{\sigma}\right)^{-n} & \frac{x-\bar{x}}{\sigma} \leq -\alpha \end{cases} \tag{6.2}$$

where $A = \left(\frac{n}{|\alpha|}\right)^n \exp\left(-\frac{|\alpha|^2}{2}\right)$ and $B = \frac{n}{|\alpha|} - |\alpha|$. The parameters $\alpha, n, \bar{x}, \sigma$ are floating parameters.

The fit is performed independently on the Z+0-jet and the Z+1-jet bins where statistics are sufficient. For the higher jet multiplicity bins, the shape of the fit is fixed from the Z+1-jet case and is used to fit the reconstructed data (Figure 6.4).



(a)

(b)

(c)

(d)

Figure 6.4: Z invariant mass fits for the 1 (a), 2 (b), 3 (c) and ≥4 (d) jet multiplicity bins.

6.5 Event Selection Efficiency

To properly calculate the cross section of Z+Jets events as a function of jet multiplicity, we need to correct the event yields for measurement efficiency. The formula used to calculate the total efficiency is broken down to 5 independently calculated efficiencies:

$$\epsilon_{event} = \epsilon_{Reco,20} \times \epsilon_{WP80} \times \epsilon_{trig} \times \epsilon_{Reco,10} \times \epsilon_{WP95} \tag{6.3}$$

the first three parameters correspond to the total *tight* electron efficiency and the last two to the total *loose* electron efficiency. In more detail,

- $\epsilon_{Reco,(20/10)}$ is the electron reconstruction efficiency relative to an ECAL supercluster and is calculated separately for tight (20 GeV) and loose (10 GeV) electrons,
- $\epsilon_{WP(80/95)}$ is the selection efficiency for each respective working point and
- ϵ_{trig} is the trigger efficiency for the tight WP80 electron.

The increased activity in events with extra jets is expected to affect the electron efficiency, therefore each one of the efficiencies is calculated as a function of jet multiplicity in exclusive jet bins. All standalone efficiencies were calculated with the official CMS Tag and Probe package (method described below). Jet p_T thresholds of 15 and 30 GeV$/c$ were examined with very little differences between the two results as described in [52]. For the final efficiency calculations, the results of the lower 15 GeV threshold were preferred by virtue of improved statistics.

Tag And Probe

The Tag and Probe [53] method utilizes a known mass resonance (in this case $Z \rightarrow e^+e^-$) to select particles of the desired type, and probes the efficiency of a particular selection criterion on those particles. The "tag" is an object that passes a set of very tight selection criteria designed to isolate the required particle type. The fake rate for passing tag selection criteria is expected to be very small ($\ll 1\%$). A generic set of the desired particle type under study (with potentially very loose selection criteria applied) known as “probes” is selected by pairing these objects with tags such that the invariant mass of the combination is consistent with the mass of the Z resonance between the 60 and 120 GeV mass window. The definition of the probe object depends on the specifics of the selection criterion being examined. The efficiency itself is measured by counting the number of “probe” particles that pass the desired selection criteria:

$$\epsilon = \frac{P_{pass}}{P_{all}} \tag{6.4}$$

where P_{pass} is the number of probes that fulfill the selection criteria under study and P_{all} is the total number of probes counted using the Z resonance. In the case that a probe object also passes the tag selection, it will appear in both the tag and probe lists and produce a double pair in the same event. The efficiency formula as written above accounts for these double pairs, since independent fits are performed on the combinations of passing and failing probes and the final efficiency is determined by the signal level on the passing divided by the failing sample. A Breigt-Wigner convolved with a Crystal Ball is used as fit function.

The final efficiency corrections applied to the data are truth-level Monte Carlo efficiencies, scaled by the Tag and Probe data to Monte Carlo ratio. The truth-level MC efficiencies are determined by comparing the number of events with generated electrons in the detector acceptance with the events remaining after applying the full selection. Table 6.4 contains the final event reconstruction efficiencies for each jet multiplicity.

6.6 Unfolding

Due to detection efficiency, measurement resolutions and systematic biases, a measured distribution can differ from the true distribution in the sense that some events may

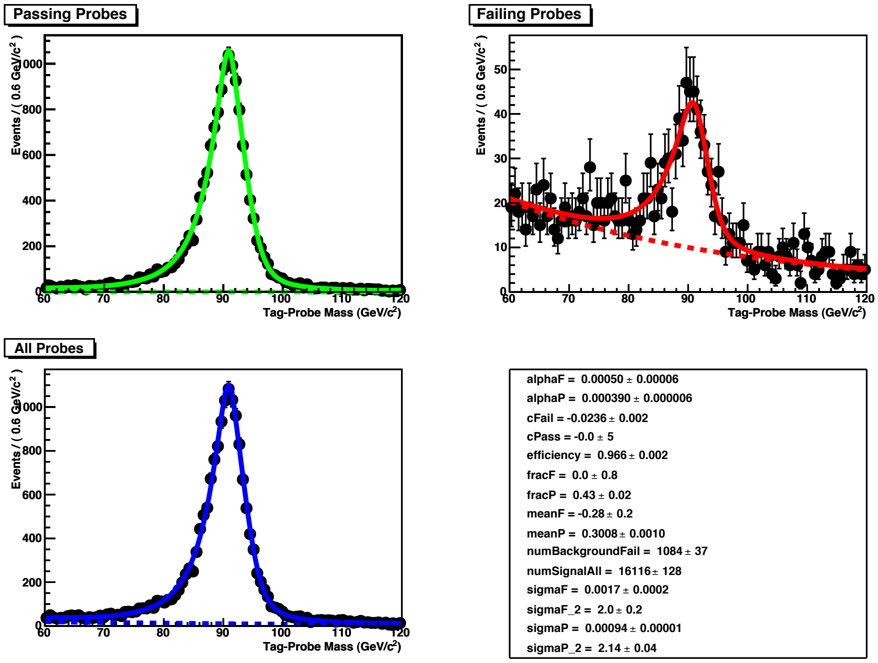


Figure 6.5: Example Tag and Probe fit results for WP80 selection. The dielectron mass distributions require one tag electron and a GSF electron with $p_T > 20\,\mathrm{GeV}$, for events with no jets with of $E_T > 15\,\mathrm{GeV}$. “Passing probes” pass the WP80 selection.

never get reconstructed or migrate to neighboring bins. To correct for these effects, an *unfolding* procedure has been applied where we attempt to retrieve the true data distribution of two measured quantities: the transverse momentum of the leading jet and the exclusive number or reconstructed jets which is used for the cross section ratio measurement.

6.6.1 Method

The RooUnfold [54] package was used to apply unfolding. RooUnfold provides a common framework to evaluate and apply several unfolding algorithms. For this analysis two of the available methods were studied:

- Singular Value Decomposition (SVD) [55] was the default unfolding method used and

- an iterative, Bayesian [56] approach was used for systematic checks.

Both require a regularization parameter to prevent the statistical fluctuations being interpreted as structure in the true distribution. Unfolding tests with different Monte Carlo samples and parameter values were performed to reach an optimal decision.

Unfolding a distribution is a two-step process. First, a response matrix R mapping the binned true distribution onto the measured one is created. For 1-dimensional true and measured distribution bins T_j and M_i, the response matrix element R_{ij} gives the fraction of events from bin T_j that are measured in bin M_i. The response matrix is determined from a Monte Carlo sample by associating true and reconstructed values. The Z+Jets MadGraph Z2 Monte Carlo sample was used to perform the first step and the resulting response matrices are shown on Figures 6.6 and 6.7. In the second step of the unfolding procedure, this response matrix is applied to the measured data, taking into account the measurement uncertainties due to statistical fluctuations in the finite measured sample.

Both unfolding techniques considered require a regularization parameter to prevent the statistical fluctuations in the data from being interpreted as structure in the true distribution. Therefore, the regularisation parameter determines the relative weight placed on the data, compared to the training sample truth.

For the iterative algorithm, the regularisation parameter specifies the number of iterations, starting with the training sample truth (zero iterations). Up to 6 iterations were evaluated for both jet multiplicity and leading jet p_T unfolding. As suggested by the RooUnfold authors, “the optimal regularisation parameter can be selected by finding the largest value up to which the errors remain reasonable (ie. do not become much larger than previous

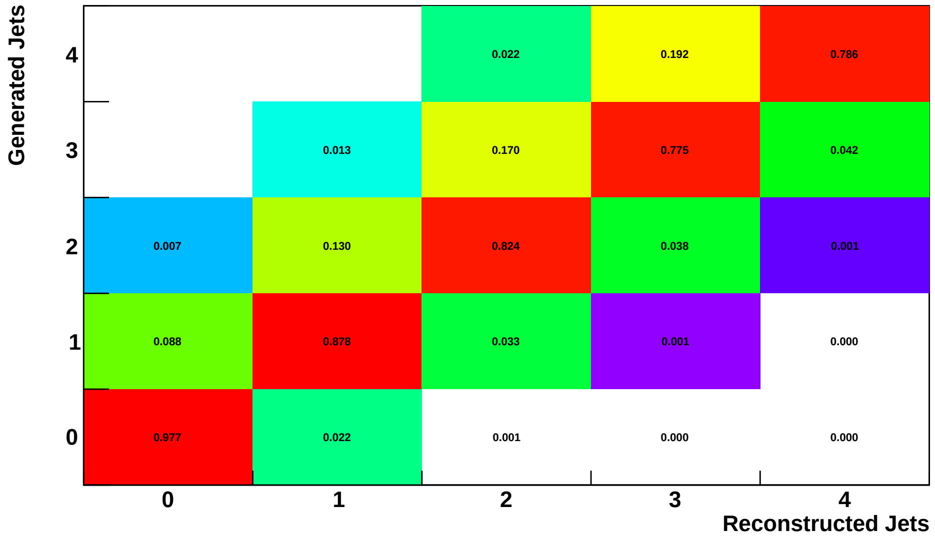


Figure 6.6: Response matrix for jet multiplicity unfolding

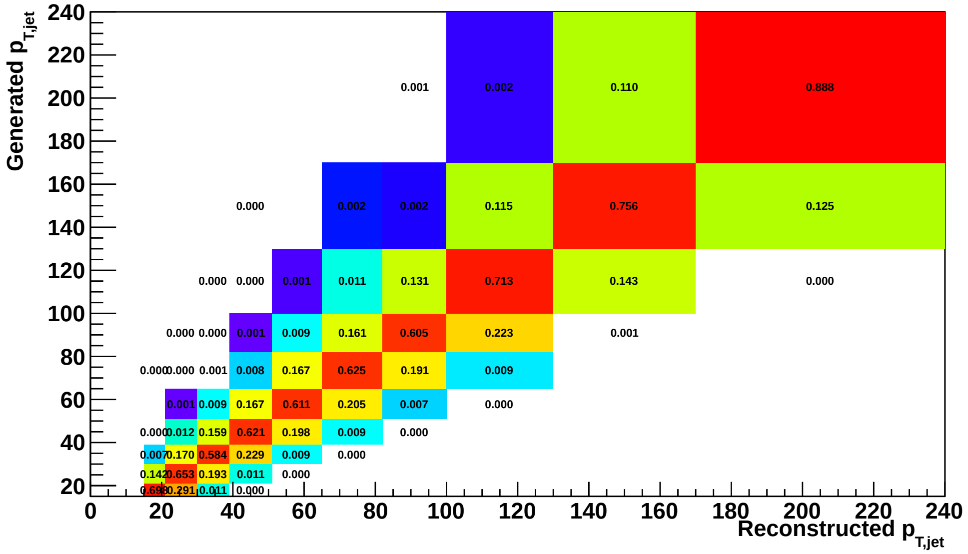


Figure 6.7: Response matrix for leading jet p_T unfolding

values)". It was observed that 4 iterations provided the best results for both unfolding cases.

For SVD, unfolding is similar to a Fourier expansion in "result to be obtained" vs "MC truth input". Here, low frequencies are treated as systematic differences between the training MC sample and the data, which should be retained in the output. High frequencies are assumed to originate from statistical fluctuations in the data which are numerically enhanced by the unfolding algorithms in lack of proper regularization. Selecting the regularization parameter value effectively determines up to which frequencies the terms in the expansion are kept. Therefore a correct choice of k_{term} is of particular importance. A small value will bias the unfolding result towards the MC truth input while a large value will give a result dominated by unphysically enhanced statistical fluctuations. The k_{term} values can lie anywhere between 2 and the number of bins; a value of 1 would return the true distribution. The case where the regularization parameter equals the number of bins, corresponds to a full inversion of the response matrix. The optimal output was observed when using $k_{term} = 5$ for both jet multiplicity and leading jet p_T unfolding.

The response matrices used are derived from MadGraph, Tune Z2 Monte Carlo samples (Table 6.1):

- containing signal-only events including pile-up and with L1 FastJet [45] pile-up subtraction applied,
- with the generated lepton within the detector acceptance in η and
- the leading (subleading) lepton $p_T^{gen} > 20$ GeV (>10 GeV) and the invariant mass between 60 GeV and 120 GeV;
- the generated jet collection is cleaned of jets that lie within $\Delta R = 0.3$ of generated

electrons (as explained in Section 5.3) and

- includes generated jets of $p_T^{gen} > 30\,\mathrm{GeV}$ (for the jet multiplicity unfolding) or $p_T^{gen} > 15\,\mathrm{GeV}$ (for the leading jet p_T unfolding)
- the reconstructed jet collection is cleaned of jets that lie within $\Delta R = 0.3$ of generated electrons and
- jets pass the loose Jet ID requirements (described in Section 5.3) and
- includes reconstructed jets of $p_T^{reco} > 30\,\mathrm{GeV}$ (for the jet multiplicity unfolding) or $p_T^{reco} > 15\,\mathrm{GeV}$ (for the leading jet p_T unfolding);
- in addition, for the leading jet p_T unfolding the reconstructed and the generated jet are required to match within a $\Delta R < 0.2$ cone.

Applying these requirements ensures the unfolding procedure will give correct results when applied to data in which background has been subtracted, pile-up has been removed and lepton efficiencies have been corrected.

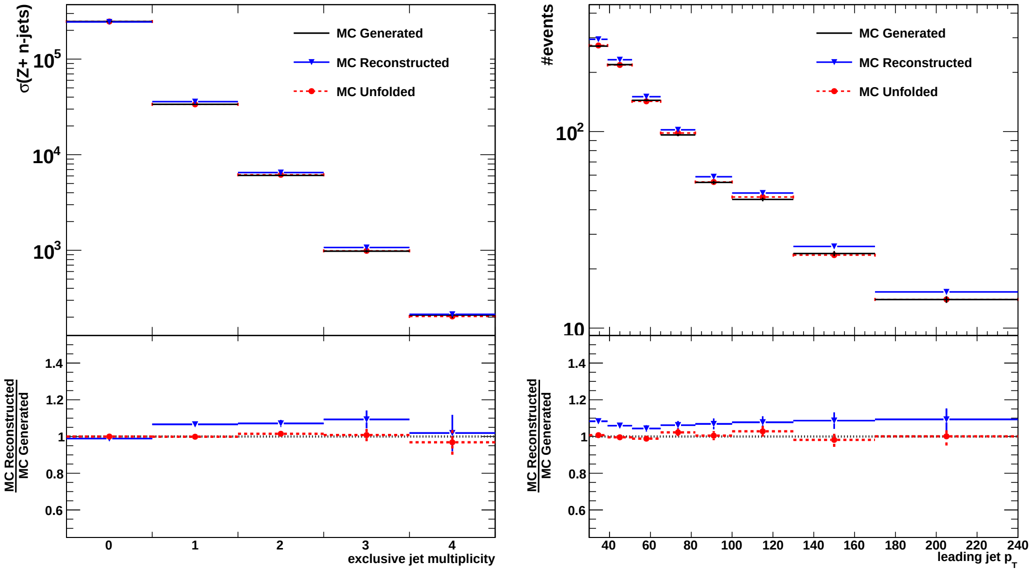
6.6.2 Validating unfolding with Monte Carlo

Before unfolding the data, a series of tests is performed to validate that the procedure works as expected using simulated samples. Three different types of checks are performed:

- Unfolding distributions using the same signal MC as the one used to create the response matrix; this was performed on the MadGraph Z2 tune MC. The sample was split into two equal parts; the first half was used for creating the response matrix while the second for testing the unfolding (Figure 6.8).

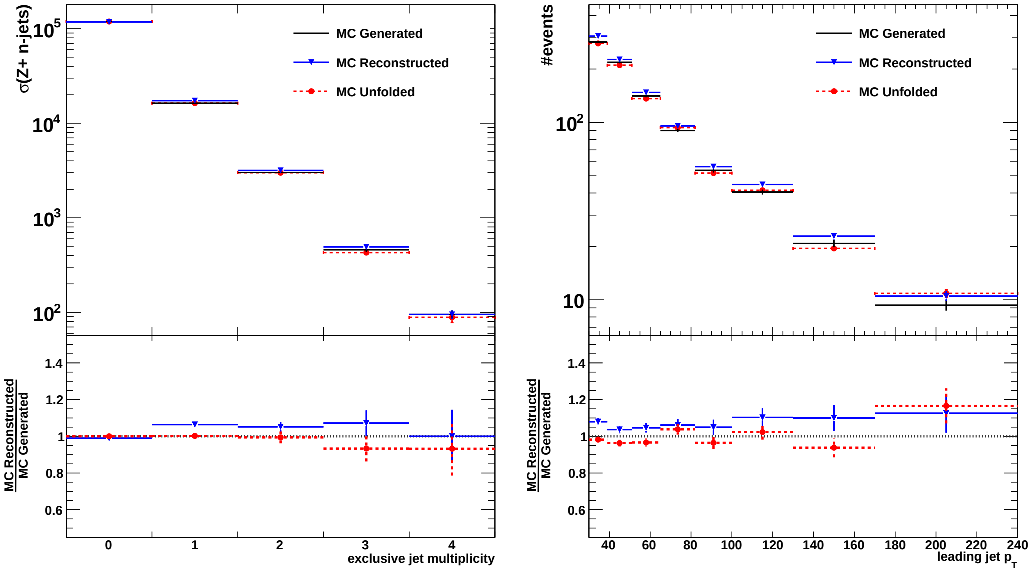
The improved agreement is demonstrated by the fact that the resulting distributions after unfolding are in better agreement with MC truth (Tables 6.5.A and 6.6.A).

- Unfolding distributions using different MC samples from the ones used to create the response matrix; the MadGraph Z2 tune response matrix was used to unfold the MadGraph D6T tune and PYTHIA Z2 tune samples with the MadGraph D6T tune results shown in Figure 6.9 and numerically on Tables 6.5.B and 6.6.B.
- Unfolding with slightly different regularization parameters to verify stability of unfolding; examples in Figure 6.10) and detailed results on Tables 6.5.C and 6.6.C.



(a) (b)

Figure 6.8: Unfolding closure test : Using the same sample for training and unfolding. Plots for (a) exclusive number of jets and (b) leading jet p_T



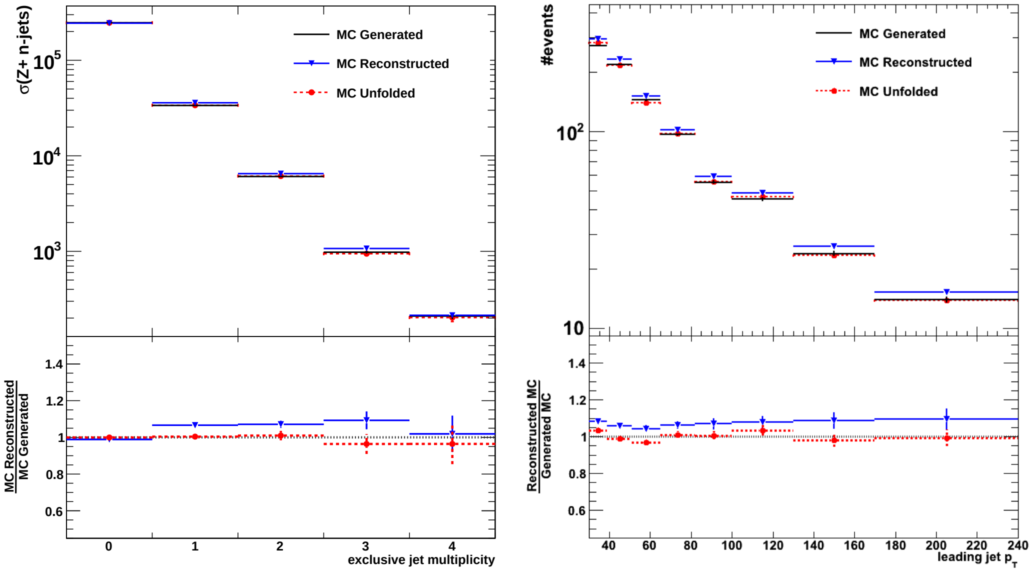
(a) (b)

Figure 6.9: Unfolding closure test : Using different sample for training and unfolding exclusive jet multiplicity. Response matrix generated from MadGraph Z2 tune Monte Carlo. Unfolding exclusive jet multiplicity and leading jet p_T for the MadGraph D6T tune sample

6.6.3 Unfolding data

After event extraction, fitting and correcting for reconstruction efficiency, the resulting distributions are unfolded using the SVD method with a regularization parameter $k_{term} = 5$ and the chosen response matrix obtained from the MadGraph, Z2 tune Monte Carlo. The results of the unfolded exclusive jet multiplicity and leading jet p_T as well as comparison with the MC predictions are in Figure 6.11.

For the jet multiplicity unfolding a closer agreement with MC truth is observed for events with up to two reconstructed jets after unfolding, with an average improvement of



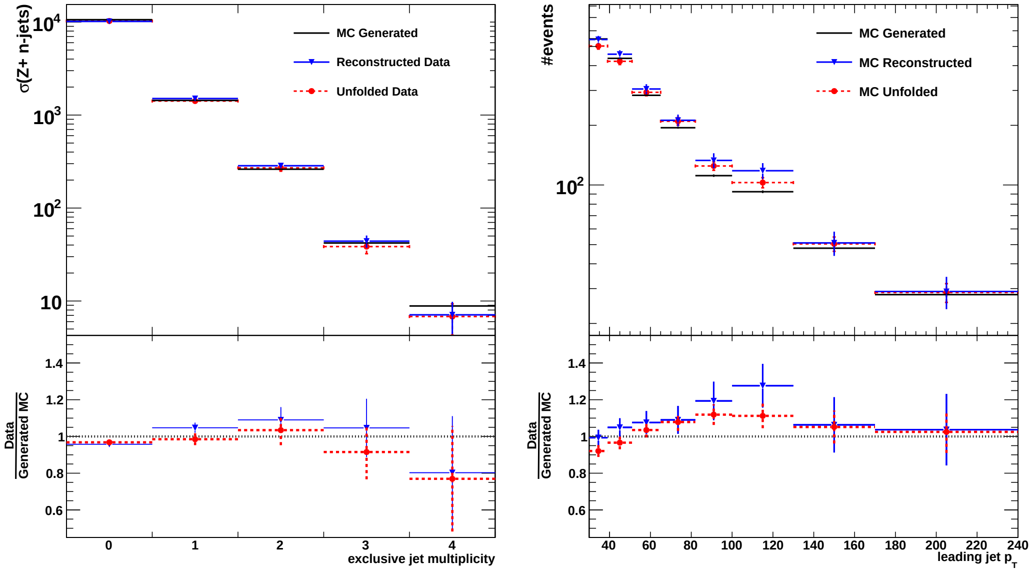
(a) (b)

Figure 6.10: Unfolding closure test : with different regularization parameters. Response matrix generated from MadGraph Z2 tune Monte Carlo. Unfolding (a) exclusive number of jets with $k_{term} = 4$ and (b) leading jet p_T with $k_{term} = 7$ for the same sample.

approximately 7%, whereas for the 3- and 4-jet cases the unfolded data diverge from MC truth more than the reconstructed, but still in agreement with unity within uncertainties. For the leading jet p_T unfolding, the first two bins hint for possible mismodeling of the leading jet p_t for low transverse momenta, whereas in the higher p_T region, there is better agreement with MC truth, seeing shifts up to 20% in the 100-120 GeV region.

6.7 Systematic uncertainties

In this section the different sources of systematic uncertainties are described, from the event reconstruction efficiency and jet energy corrections to the studied unfolding system-



(a) (b)

Figure 6.11: Unfolding (a) exclusive jet multiplicity and (b) leading jet p_T in data. The top plots show the Monte Carlo truth and data before and after unfolding. The bottom plots show the ratio of data and MC truth before and after unfolding.

atics.

6.7.1 Event reconstruction

The leading source of systematic biases influencing event reconstruction originate from the line shape choice used to fit the Z signal in the Tag and Probe fits. To estimate this effect, the efficiencies were re-calculated with a Cruijff function (modified Gaussian, see appendix B) lineshape and a systematic error was assigned corresponding to the difference between the two numbers in each jet multiplicity bin [52].

6.7.2 Jet energy scale

The uncertainty in jet energy scale (JES) affects directly the jet multiplicity spectrum, therefore is an important factor of the ratio measurement. It is estimated on pure Monte Carlo samples due to limited data statistics in the high multiplicity jet bins by shifting the jet E_T by $\pm 1\sigma_{JES}$ uncertainty.

The largest contribution in the jet energy uncertainty is coming from the standard L2+L3 η and p_T-dependent corrections (explained in detail in Section 4.6.3) and are about 2-3%. Pileup uncertainty is added as a function of jet E_T, and is around 1.2% for a 30 GeV jet. A flat 1.5% uncertainty is assigned for changes in software release and calibrations as well as a 2% flavor uncertainty as suggested by the CMS jet energy corrections group. All of these uncertainties are added in quadrature to calculate σ_{JES}.

The corrected energy of each jet is varied by $\pm 1\sigma_{JES}$, the Z event yields are recalculated and the result is used as the total JES uncertainty. An additional pile-up uncertainty has been added on top of the JES uncertainties to account for imperfect corrections using the L1FastJet algorithm. This has been found to lie at the level of 1.5% for events with zero jets above 30 GeV, 2% for events with exactly one jet, 3% for two reconstructed jets above threshold and 5% for events with more jets.

Figure 6.12 shows the data n-jet distribution with systematic error bands and the ratio of the $\pm 1\sigma_{JES}$ scaling to the unscaled.

6.7.3 Unfolding

The optimal unfolding method was chosen as the one that returned the best agreement with the truth-level Monte Carlo sample, which naturally can introduce biases. To estimate

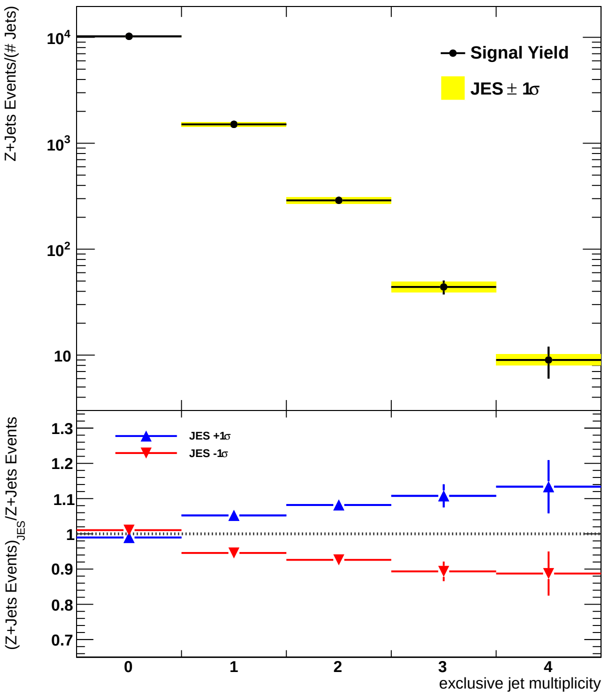


Figure 6.12: Jet energy scale systematic uncertainties. The shaded areas on the top plot show the associated systematic uncertainties and their effect on data. The bottom plot shows the shifting when applying $+1\sigma_{JES}$ (up arrow) and $-1\sigma_{JES}$ (down arrow)

the systematic uncertainty the following cases were considered:

Varying unfolding method : unfolding the MadGraph Z2 tune sample, using the Bayes method with 4 iterations

Varying MC tune : unfolding the MadGraph D6T tune sample with SVD and $k_{term} = 5$

Varying MC generator : unfolding the Pythia Z2 tune sample with SVD and $k_{term} = 5$

For the second and third case the response matrices from the respective samples are used as input.

Table 6.1: List of MC signal and background samples.

CMS Dataset name is as /*/Winter10-START39_V8-v1/AODSIM			
Dataset name	Process	Kinematic cuts	σ (pb)
DYJetsToLL_TuneZ2_M-50_7TeV-madgraph-tauola	$Z \rightarrow \ell\ell$+Jets	$M_{ll} > 50$ GeV	3048
WJetsToLNu_TuneZ2_7TeV-madgraph-tauola	$W \rightarrow \ell\nu$+Jets	None	31314
TTJets_TuneZ2_7TeV-madgraph-tauola	$t\bar{t}$ + Jets (tauola)	None	157
QCD_Pt-20to30_EMEnriched_TuneZ2_7TeV-pythia6	EM enriched QCD	p_T [20-30] GeV	1719150
QCD_Pt-30to80_EMEnriched_TuneZ2_7TeV-pythia6	EM enriched QCD	p_T [30-80] GeV	3498700
QCD_Pt-80to170_EMEnriched_TuneZ2_7TeV-pythia6	EM enriched QCD	p_T [80-170] GeV	134088
QCD_Pt-20to30_BCtoE_TuneZ2_7TeV-pythia6	BCtoE QCD	p_T [20-30] GeV	108330
QCD_Pt-30to80_BCtoE_TuneZ2_7TeV-pythia6	BCtoE QCD	p_T [30-80] GeV	138762
QCD_Pt-80to170_BCtoE_TuneZ2_7TeV-pythia6	BCtoE QCD	p_T [80-10] GeV	9422
Additional Z+Jets MCs used			
DYJetsToLL_TuneD6T_M-50_7TeV-madgraph-tauola	$Z \rightarrow \ell\ell$+Jets	$M_{ll} > 50$ GeV	3048
DYToEE_M-20_TuneZ2_7TeV-pythia6	$Z \rightarrow \ell\ell$+Jets	$M_{ll} > 20$ GeV	3048

requirement	Z+jets	W+jets	top	QCD	Data
initial number of events	106893 (100%)	24761 (100%)	1789 (100%)	8195724 (100%)	
HLT	21416 (20%)	24761 (100%)	1789 (100%)	8195724 (100%)	15041836
two reconstructed electrons	20616 (96%)	24761 (100%)	1789 (100%)	8195724 (100%)	218503 (2%)
two reconstructed electrons in acceptance	16964 (82%)	5241 (21%)	891 (50%)	1334721 (16%)	40672 (19%)
Electron ID	14577 (86%)	386 (7%)	100 (11%)	7250 (1%)	15649 (38%)
Electron Isolation	13893 (95%)	109 (28%)	50 (50%)	224 (3%)	15001 (96%)
Conversion Rejection	13755 (99%)	76 (70%)	49 (98%)	95 (43%)	14499 (97%)
ECAL Driven	13609 (99%)	67 (87%)	49 (100%)	73 (76%)	14286 (99%)
$\|\delta_{xy}\| \leq 350\mu m$	13175 (97%)	56 (84%)	47 (97%)	61 (83%)	13665 (96%)
HLT Match	12353 (94%)	55 (98%)	46 (98%)	56 (93%)	13394 (98%)
μ veto	12353 (100%)	55 (100%)	46 (100%)	56 (100%)	13394 (100%)
$60 < m(e^+e^-) < 120$ GeV/c^2	11887 (96%)	18 (32%)	20 (43%)	15 (27%)	12071 (90%)

Table 6.2: Breakdown of events at different steps of selection for the inclusive Z case for signal and background events after the full Z+jets selection is applied as described in Section 6.3. MC Events are normalized to the data integrated luminosity of 36.1 pb^{-1}. For the Monte Carlo backgrounds, a preselection has been applied to reduce sample size. The numbers is parentheses correspond to the percentage of events surviving the cut with respect to the previous one.

jets	Z+jets	W+jets	top	QCD	Data
≥ 0	11887 ± 16	18 ± 1	20 ± 0.3	15 ± 6.1	12071 ± 110
≥ 1	1830 ± 6	13 ± 0.9	19 ± 0.3	13 ± 4.3	1852 ± 43
≥ 2	334 ± 3	3 ± 0.3	15 ± 0.3	8.2 ± 3	343 ± 19
≥ 3	55 ± 1	0.8 ± 0.1	6 ± 0.2	0 ± 0	53 ± 7
≥ 4	9 ± 0.5	0.2 ± 0	2 ± 0.1	0 ± 0	9 ± 3

Table 6.3: Breakdown of events for the different jet multiplicities of signal and background events after the full Z+jets selection is applied. Events are normalized to the data integrated luminosity of 36.1 pb^{-1}. Particle Flow jets with $E_T > 30$ GeV are counted as jets as described in Section 5.3.

exclusive # jets	0	1	2	3	≥ 4
T&P Data	0.691 ± 0.004	0.667 ± 0.008	0.662 ± 0.014	0.684 ± 0.039	0.529 ± 0.062
T&P MC	0.662 ± 0.000	0.660 ± 0.001	0.653 ± 0.001	0.639 ± 0.002	0.635 ± 0.004
MC Truth	0.697 ± 0.001	0.681 ± 0.003	0.671 ± 0.008	0.625 ± 0.018	0.599 ± 0.046
ϵ_Z	0.728 ± 0.004	0.688 ± 0.017	0.680 ± 0.017	0.669 ± 0.043	0.499 ± 0.070

Table 6.4: Event selection efficiency. The final efficiency is calculated by $\epsilon_Z = \epsilon_{MC,truth} \times \frac{\epsilon_{T\&P,data}}{\epsilon_{T\&P,MC}}$

A - MadGraph Z2 tune MC for response matrix & unfolding using SVD method with k_{term}=5					
$\frac{\mathit{Reconstructed}}{\mathit{Generated}}$	0 jets	1 jet	2 jets	3 jets	4 jets
Before unfolding	0.989 ± 0.003	1.066 ± 0.008	1.071 ± 0.019	1.093 ± 0.048	1.019 ± 0.099
After Unfolding					
B - MadGraph Z2 tune MC for response matrix & unfolding MadGraph D6T tune sample using SVD method with k_{term}=5					
Before unfolding	0.990 ± 0.004	1.064 ± 0.012	1.052 ± 0.027	1.072 ± 0.070	1.000 ± 0.145
After Unfolding	1.000 ± 0.005	1.002 ± 0.014	0.994 ± 0.032	0.933 ± 0.070	0.932 ± 0.146
C - MadGraph Z2 tune MC for response matrix & unfolding using SVD method with k_{term}=4					
Before unfolding	0.989 ± 0.003	1.066 ± 0.008	1.071 ± 0.019	1.093 ± 0.048	1.019 ± 0.099
After Unfolding	1.000 ± 0.004	1.001 ± 0.009	1.011 ± 0.018	0.995 ± 0.038	0.953 ± 0.070

Table 6.5: Unfolding jet multiplicity closure tests.

A - MadGraph Z2 tune MC for response matrix & unfolding using SVD method with k_{term}=5								
$\frac{Reconstructed}{Generated}$	30 - 39	39- 51	51- 65	65-82	82-100	100- 13	130-170	170- 240
Before unfolding	1.083 ± 0.013	1.059 ± 0.015	1.044 ± 0.018	1.062 ± 0.022	1.069 ± 0.029	1.078 ± 0.033	1.087 ± 0.045	1.093 ± 0.059
After Unfolding	1.007 ± 0.013	0.996 ± 0.014	0.988 ± 0.016	1.023 ± 0.020	1.005 ± 0.025	1.028 ± 0.029	0.982 ± 0.038	1.001 ± 0.050
B - MadGraph Z2 tune MC for response matrix & unfolding MadGraph D6T tune sample using SVD method with k_{term}=5								
Before unfolding	1.080 ± 0.019	1.037 ± 0.020	1.047 ± 0.026	1.061 ± 0.032	1.049 ± 0.042	1.103 ± 0.050	1.100 ± 0.069	1.126 ± 0.105
After Unfolding	0.982 ± 0.017	0.963 ± 0.018	0.966 ± 0.022	1.038 ± 0.029	0.965 ± 0.033	1.023 ± 0.041	0.938 ± 0.053	1.166 ± 0.096
A - MadGraph Z2 tune MC for response matrix & unfolding using SVD method with k_{term}=5								
$\frac{Reconstructed}{Generated}$	30 - 39	39- 51	51- 65	65-82	82-100	100- 13	130-170	170- 240
Before unfolding	1.083 ± 0.013	1.059 ± 0.015	1.044 ± 0.018	1.062 ± 0.022	1.069 ± 0.029	1.078 ± 0.033	1.087 ± 0.045	1.093 ± 0.059
After Unfolding	0.993 ± 0.018	1.022 ± 0.018	0.998 ± 0.021	1.025 ± 0.025	0.989 ± 0.029	1.014 ± 0.033	0.989 ± 0.042	1.033 ± 0.061

Table 6.6: Unfolding leading jet p_T closure tests.

exclusive # jets	0	1	2	3	≥ 4
Jet energy scale	$\pm 1.8\%$	$^{+5.6}/_{-5.5}\%$	$\pm 8.5\%$	$^{+13.5}/_{-11.6}\%$	$\pm 12.2\%$
Event reconstruction	$\pm 1\%$	$\pm 1\%$	$\pm 2\%$	$\pm 2.3\%$	$\pm 5.5\%$
Total	$\pm 2.1\%$	$^{+5.7}/_{-5.6}\%$	$\pm 8.8\%$	$^{+13.7}/_{-11.8}\%$	$\pm 13.4\%$

Table 6.7: Sources of systematic uncertainty in exclusive jet multiplicity bins.

An experiment is a question which science poses to Nature, and a measurement is the recording of Nature's answer.

Scientific Autobiography and
Other Papers
MAX PLANCK

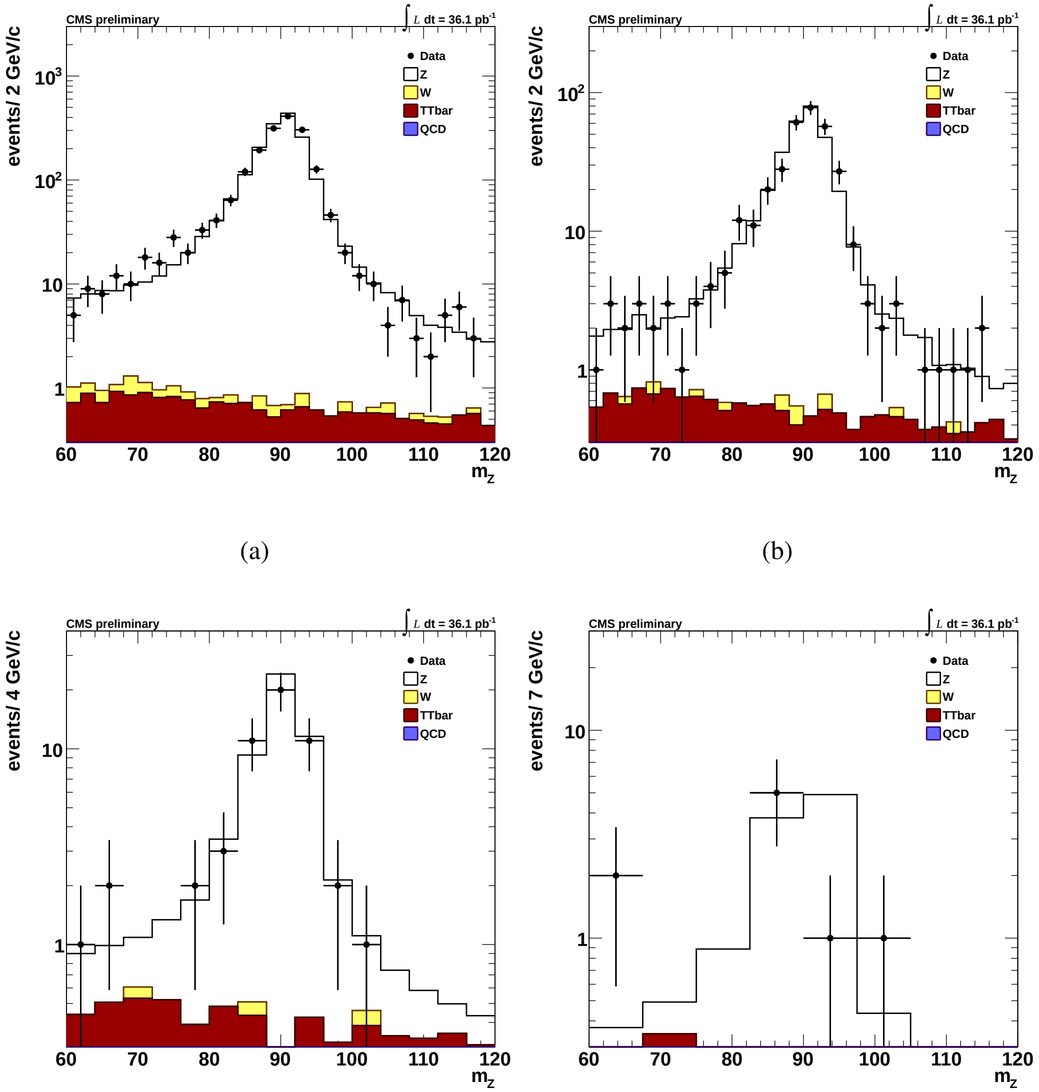
Chapter 7

Results

After applying the full selection presented in the previous chapters, we are in a position to cross-check our results with theoretical predictions given by the Monte Carlos. We compare data yields with MadGraph+PYTHIA (Section 4.2) generator calculations for Z production in association with up to 4 reconstructed jets above 30 GeV. Also, unfolded cross section ratios are compared with MadGraph+PYTHIA [16, 31] and pure PYTHIA [31] predictions.

7.1 Comparisons to Theoretical Predictions

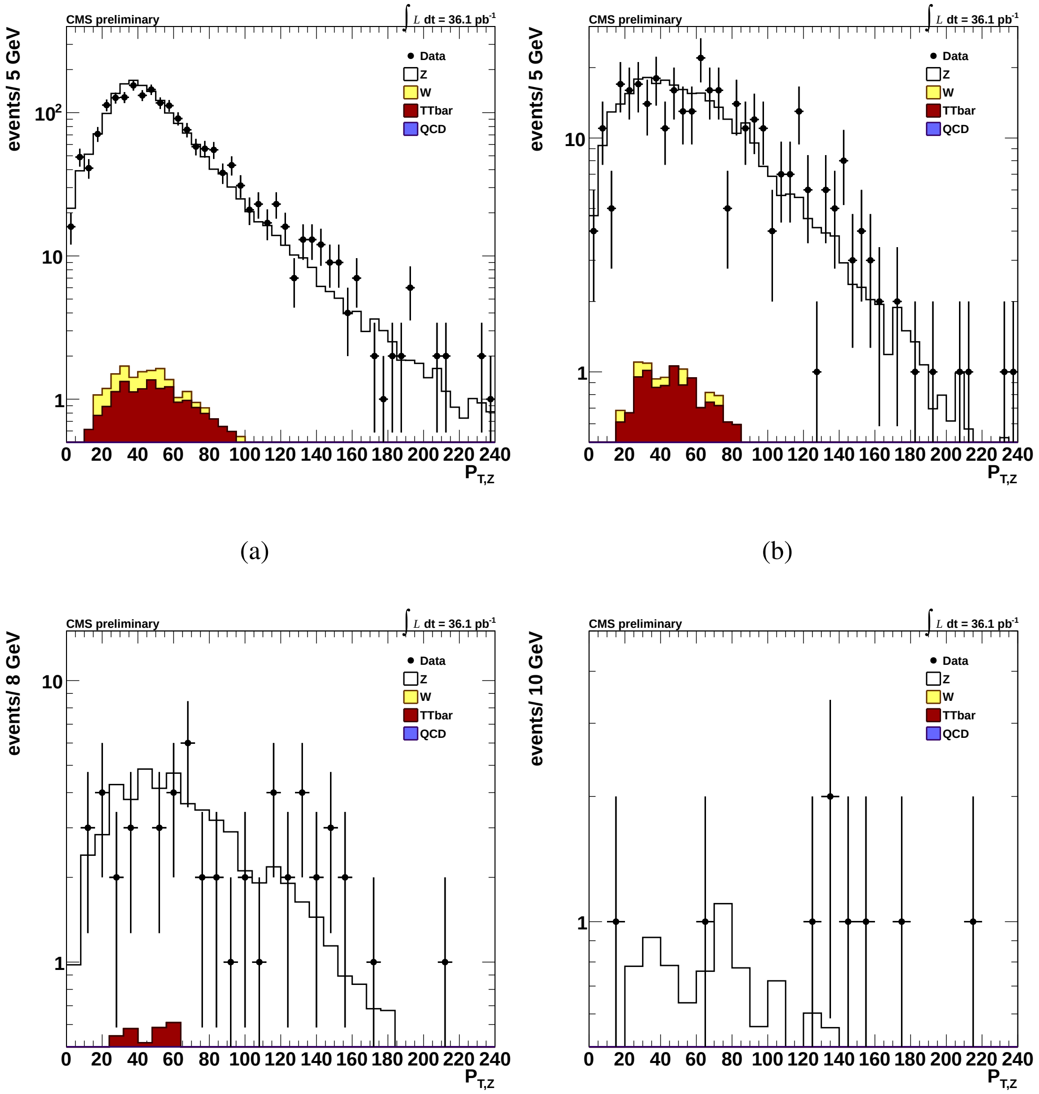
We start by presenting the distributions of the Z invariant mass, its transverse momentum and the reconstructed leading jet p_T for up to 4 reconstructed jets. Figures 7.1 and 7.2 show the Z mass spectrum; the distributions up to three jets agree well with Monte Carlo predictions, whereas the small number of events with four reconstructed jets does not give a clear picture about its shape; the integrated number of events agrees well between data and MC (Table 6.3). The same comments hold for the transverse momentum of the reconstructed Z boson (Figure 7.2) and the leading jet p_T (Figure 7.3), but here due to the wide momentum spectrum the distributions agree well up for up to two jets before the statistics are too low to reach a conclusion.



(a) (b)

(c) (d)

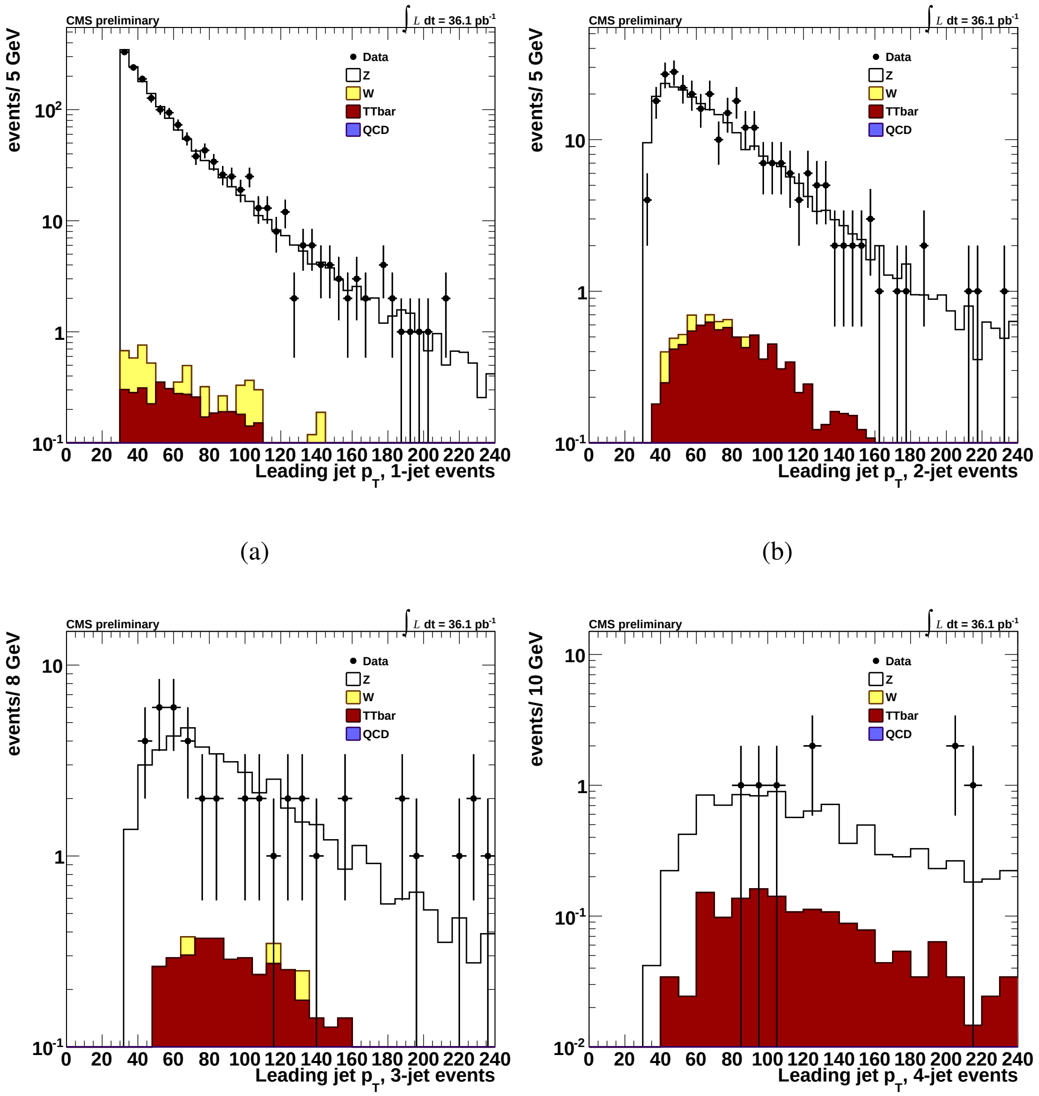
Figure 7.1: Di-electron invariant mass for Z+≥ 1 (a), ≥2 (b), ≥3 (c) and ≥4 (d) Particle Flow jets.



(a) (b)

(c) (d)

Figure 7.2: Di-electron transverse momentum for events with $\geq$ 1 (a), $\geq$2 (b), $\geq$3 (c) and $\geq$4 (d) Particle Flow jets.



(a)

(b)

(c)

(d)

Figure 7.3: Leading jet p_T for events with $\geq$ 1 (a), $\geq$2 (b), $\geq$3 (c) and $\geq$4 (d) jets after applying the full $Z \rightarrow e^+e^-$ selection

7.2 Cross section ratios

Unfolded cross section ratios of $\frac{Z+\geq n-jets}{Z+\geq (n-1)-jets}$ and $\frac{Z+\geq n-jets}{Z+\geq 0-jets}$ as well as the unfolded leading jet p_T spectrum are shown here after applying the full selection, correcting the resulting cross sections for efficiency (Section 6.5) and applying unfolding to account for detector effects (Section 6.6). Results are not corrected for acceptance in order to keep the results independent of Monte Carlo generators that would be used to make such corrections and the added uncertainties this would introduce.

Measurement of the cross section ratios has the added benefit that several important uncertainties cancel out. Luminosity, one of the most important uncertainties at about 4%, cancels out completely. Jet energy scale uncertainties also partially cancel out, as well as possible uncertainties in the electron efficiency that are independent from the jet presence in the events.

Table 7.1 has the final data cross sections after efficiency corrections and unfolding compared with the official CMS analysis [57]. The agreement is good with small differences. These variations originate from the fact that the CMS results, in an attempt to avoid double-counting, resulted in being less efficient in finding Z candidates. This effect is mediated after efficiency corrections at the expense of requiring slightly larger correction factors. Tables 7.2 and 7.3 report the resulting cross section ratios and the associated systematic errors from jet and unfolding uncertainties. The same PYTHIA [31] is presented in Figures 7.4 and 7.5 compared with Monte Carlo predictions from the MadGraph Z2 and D6T tune samples as well as the PYTHIA Z2 tune sample. Agreement with the MadGraph samples is very good. In contrast, we observe increasing disagreement with PYTHIA as the number of

Exclusive jet bin	0-jet	1-jet	2-jet	3-jet	4-jet
Fit yield	10161 ± 113	1505 ± 39	285 ± 18	44 ± 7	7 ± 3
Efficiency-corrected yield	13957 ± 173	2188 ± 62	419 ± 28	66 ± 11	14 ± 6
Final unfolded yield	14097 ± 182	2077 ± 74	398 ± 34	59 ± 11	13 ± 6
CMS Result	14189 ± 285	2053 ± 99	401 ± 45	53 ± 13	12 ± 6

Table 7.1: Final event yields in exclusive jet bins. Errors are statistical only.

	ratio with stat. error	JES syst error (±)		Unfolding		
				SVD/Bayes	Generator	Tune
$\geq 1/\geq 0$ jets	0.151 ± 0.004	0.0134	0.0098	-0.0089	0.00892	6.84e-04
$\geq 2/\geq 0$ jets	0.0285 ± 0.00182	0.00234	0.00231	-0.0014	0.00769	-2.43e-04
$\geq 3/\geq 0$ jets	0.00424 ± 7.08e-04	0.00139	3.023-04	-4.8e-04	0.00183	1.21e-04
$\geq 4/\geq 0$ jets	7.61e-04 ± 3.12e-04	1.8e-04	7.62e-05	3.24e-05	4.66e-04	6.96e-05

Table 7.2: Results for cross section ratio $\sigma\left(Z+\geq n\,\text{jets}\right)/\sigma\left(Z+\geq 0\,\text{jets}\right)$.

jets increases since PYTHIA is a fixed-order tree-level matrix element calculator, and thus does not handle the presence of extra jets.

7.3 Leading jet p_T unfolding

The result of unfolding the leading jet transverse momentum distribution is shown in Figure 7.6 for data, MadGraph and PYTHIA Monte Carlo generators. The contents of each bin have been scaled by the bin width to account for the variable bin size used to minimize bin-to-bin migrations. As with previous observations, the MadGraph samples provide better

	ratio with stat. error	JES syst error (±)		Unfolding		
				SVD/Bayes	Generator	Tune
$\geq 1/\geq 0$ jets	0.151 ± 0.004	0.0134	0.0098	-0.0089	0.00892	6.84e-04
$\geq 2/\geq 1$ jets	0.189 ± 0.020	-0.00663	7.59e-04	-0.0013	0.0445	-8.17e-04
$\geq 3/\geq 2$ jets	0.149 ± 0.034	-0.0155	0.00425	-0.0102	0.0311	0.00605
$\geq 4/\geq 3$ jets	0.179 ± 0.117	0.008	0.0113	0.0328	0.0488	0.0113

Table 7.3: Results for cross section ratio $\sigma\left(Z+\geq n\,\text{jets}\right)/\sigma\left(Z+\geq (n-1)\ \text{jets}\right)$.

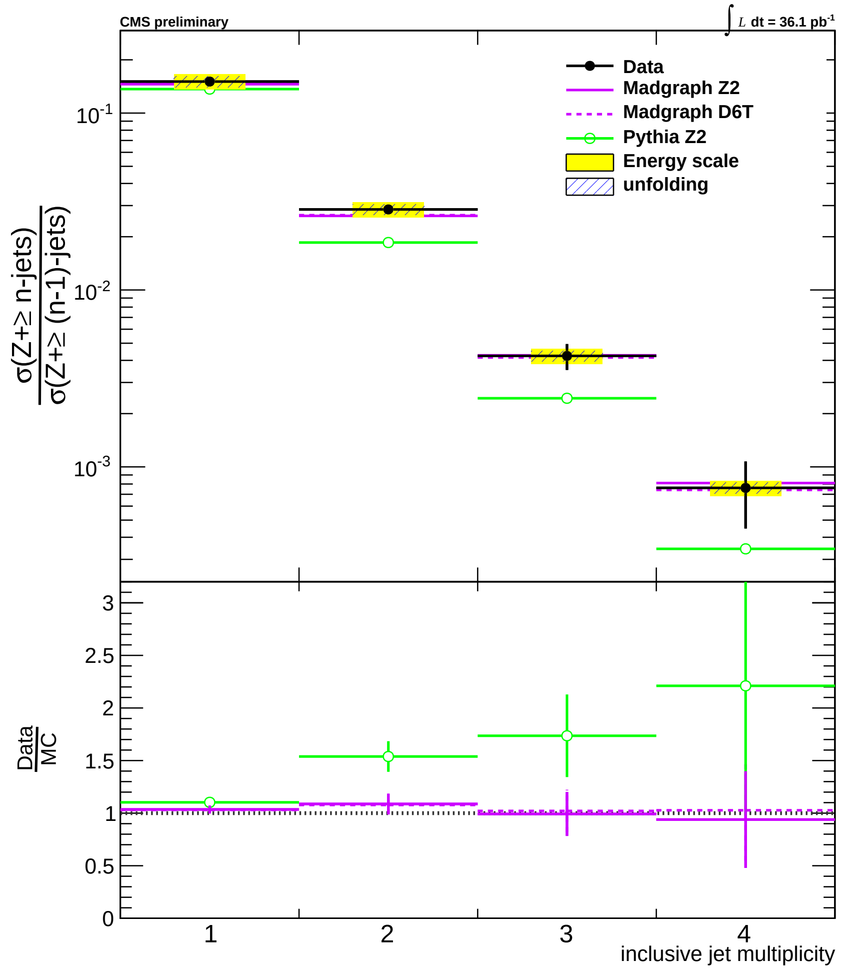


Figure 7.4: Plot for cross section ratio $\sigma\left(Z+\geq n\,\text{jets}\right)/\sigma\left(Z+\geq 0\,\text{jets}\right)$ vs. number of jets compared with theoretical predictions from MadGraph and PYTHIA.

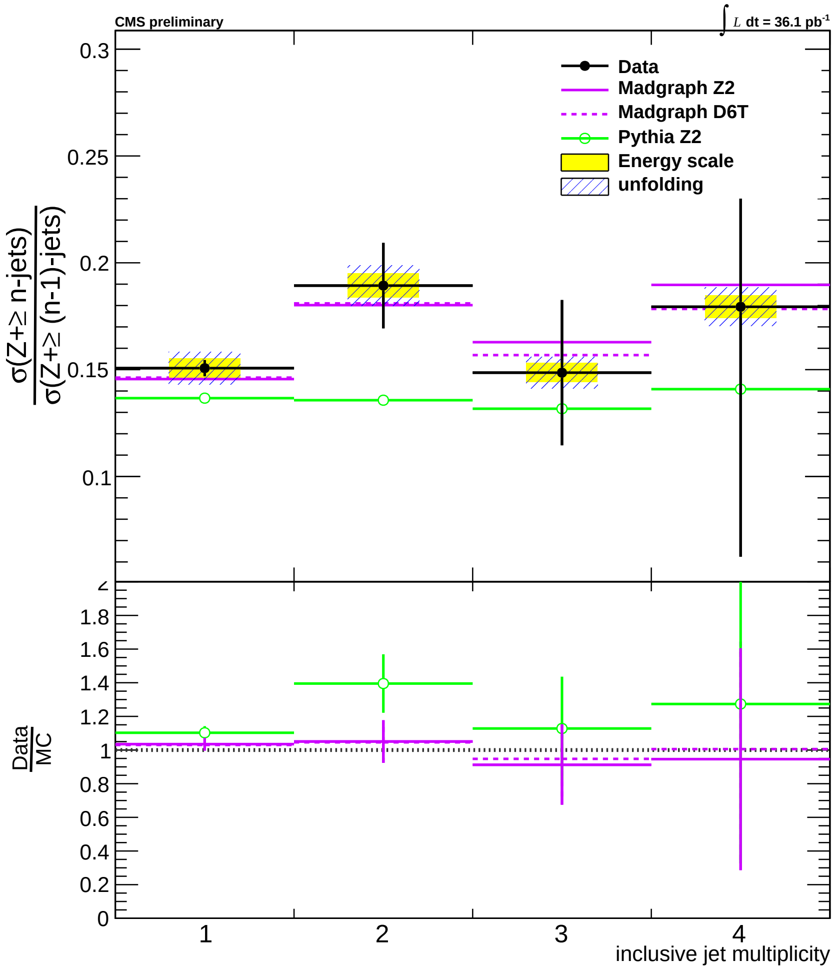


Figure 7.5: Plot for cross section ratios $\sigma\left(Z+\geq n\,\text{jets}\right)/\sigma\left(Z+\geq (n-1)\text{ jets}\right)$ vs. number of jets compared with theoretical predictions from MadGraph and PYTHIA.

agreement; the better description of the leading jet p_T provided by the Z2 tune compared to the D6T tune is also exhibited.

7.4 Previous Z+Jets measurements

The Z boson, since its initial discovery at the SPS collider in 1983 has been studied in great detail. The Large Electron-Positron collider (LEP) was built at CERN to perform accurate studies of the W/Z bosons properties. LEP detectors determined the mass of the W and Z bosons with an accuracy of 1‰, and even today 11 years after final shutdown, the four LEP experiments (ALEPH, Delphi, OPAL and L3) have provided the most accurate Z mass measurements so far [7].

The most recent results on Z+Jets production come from the two Tevatron experiments (CDF and D0). At the Tevatron, being a $p\bar{p}$ collider, the production mechanisms are different. At LO in pQCD the Z+Jets events are generated by the $qg \rightarrow Z + q$ and $p\bar{p} \rightarrow Z + g$ processes. The collision energy during the Run II Tevatron running period is $\sqrt{s}$ =1.96 TeV, about 3.5 times less than the current LHC collision energy.

7.4.1 CDF

The CDF collaboration has published results [58, 59] on inclusive $Z/\gamma^* \rightarrow e^+e^-$+jets production with a total of 2.5 fb^{-1} of $p\bar{p}$ collision data. To extract these results, two electrons of $E_T > 25$ GeV were required:

- at least one central electron $|\eta| < 1.0$ passing a tight selection and
- the second electron fulfilling a looser set or criteria can be either central or forward: $|\eta| < 1.0$ or $1.2 < |\eta| < 2.8$.

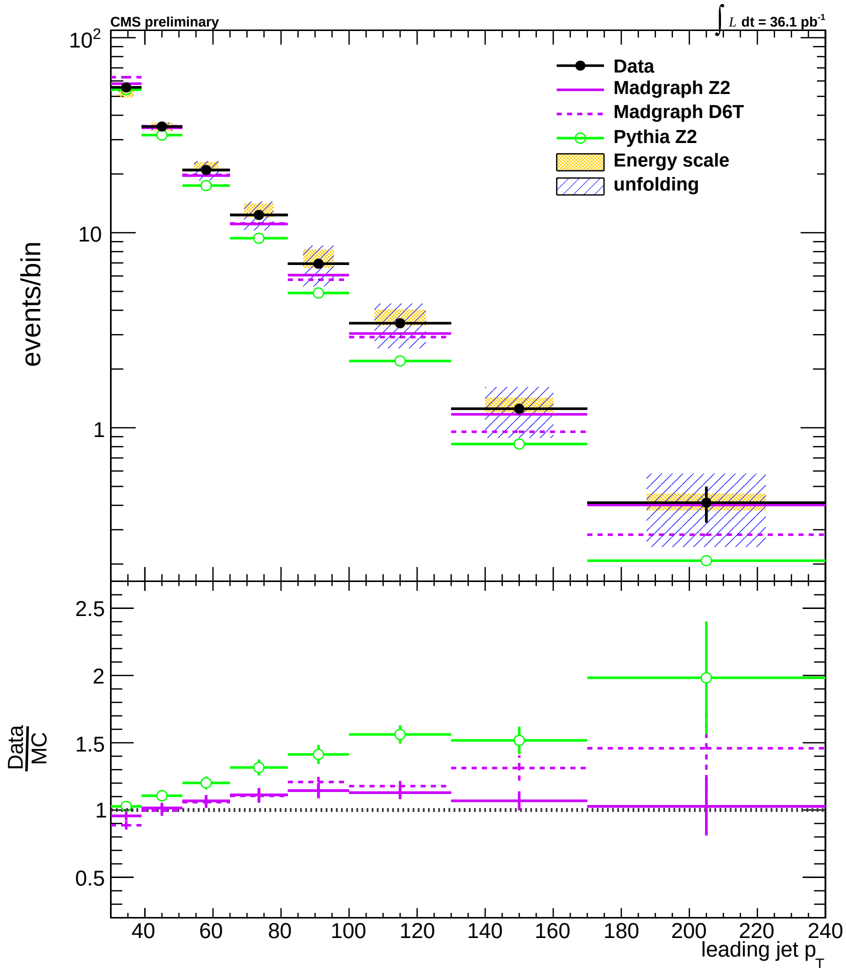


Figure 7.6: Unfolding leading jet p_T. Comparisons with theoretical prediction from MadGraph and PYTHIA generators. Bin contents have been scaled by the bin size to account for the variable bin widths.

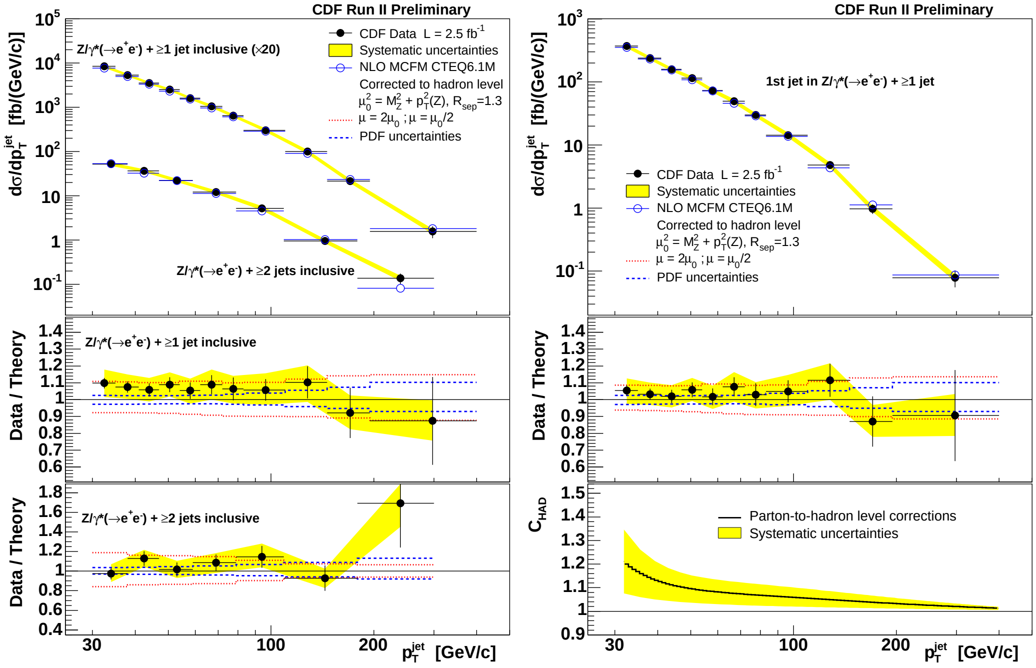
Events are required to have a reconstructed primary vertex within 60 cm around the nominal interaction point. Jets were reconstructed using the MidPoint algorithm with $R = 0.7$ and only jets with $p_T > 30\,\mathrm{GeV/c}$ and $|y| < 2.1$ are considered. Jet counting is done in inclusive jet bins and ratios up to 3-jet events are calculated. The Z mass window was set to $66 < M_{ee} < 116\,\mathrm{GeV/c^2}$ and a minimum separation of $\Delta R > 0.7$ between jets and each of the two electrons identified as coming from the vector boson was required.

Figure 7.7 shows on the left the inclusive jet cross section as a function of the jet transverse momentum for Z production in association with ≥1-jet and ≥2-jet events and how they compare to theoretical predictions; on the right, the leading jet cross section in Z/γ^* ≥1-jet events is shown and how it compares with theory. The bottom right plot shows the additional C_{had} factor that was applied, accounting for non-pQCD contributions.

Figure 7.8 on the left shows the total inclusive cross section vs inclusive jet multiplicity and how it compares with theory and on the right, the $\frac{Z+\geq n-jets}{Z+\geq(n-1)-jets}$ cross section ratio is plotted. Agreement with NLO Monte Carlo is good and within statistical uncertainties.

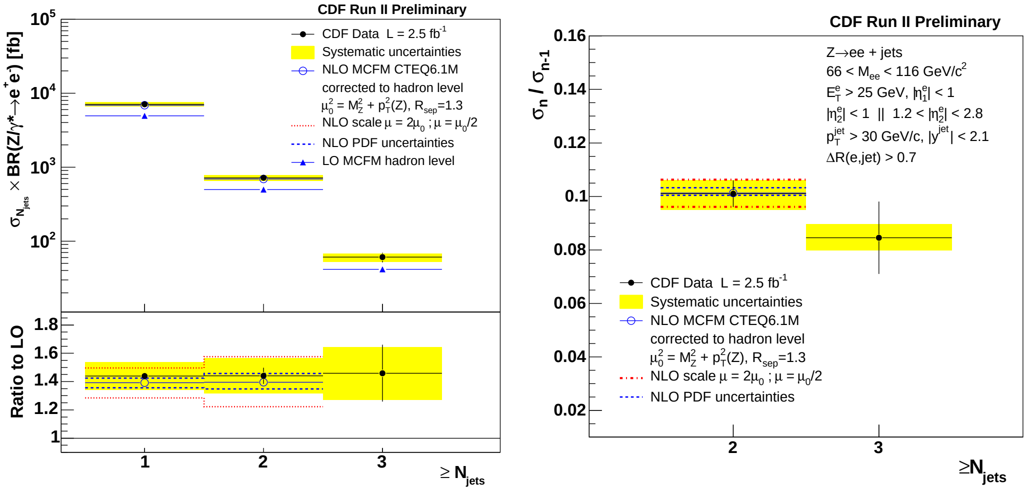
7.4.2 D0

The D0 experiment has published results with $0.4\,\mathrm{fb}^{-1}$ on the observation of $Z \rightarrow e^+e^-$ in association with jets. For their analysis, two central ($|\eta| < 1.1$) electrons with $p_T > 25\,\mathrm{GeV}$ were required to reconstruct the Z boson candidates. Same quality and isolation cuts were applied to both electrons and additionally at least one of was required to have a spatially matched track associated with the reconstructed calorimeter cluster, and the track momentum had to be consistent with the energy of the electromagnetic cluster. The invariant mass of the pair was limited to $75 < M_{ee} < 105\,\mathrm{GeV}$ in order to match the Z mass. Jets were reconstructed using the "Run II cone algorithm" with a cone of radius $R_{cone} = 0.5$.



(a) (b)

Figure 7.7: CDF results : (a) Inclusive jet cross section in Z/γ^* $\geq$1-jet and Z/γ^* $\geq$2-jet production as a function of $p_{T,jet}$ and comparison with theoretical predictions. (b) leading jet cross section in Z/γ^* $\geq$1-jet events with theoretical predictions and applied pQCD correction factors in the bottom plots. For both plots black dots are data, open circles are NLO pQCD predictions. Shaded band is systematic uncertainty, excluding luminosity uncertainty. The dashed and dotted lines indicate PDF uncertainty and variation with μ of the NLO pQCD predictions respectively.



(a) (b)

Figure 7.8: CDF results : (a) Total inclusive cross section vs inclusive jet multiplicity and ratio with theoretical predictions; (b) $\frac{Z+\geq n-jets}{Z+\geq(n-1)-jets}$ cross section ratio. For both plots black dots are data, open circles are NLO pQCD predictions. Shaded band is systematic uncertainty, excluding luminosity uncertainty. The dashed and dotted lines indicate PDF uncertainty and variation with μ of the NLO pQCD predictions respectively.

Jets were required to have $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$ and were discarded if they overlapped with the Z leptons inside a $\Delta R < 0.4$ cone. Finally, the identified primary vertex was vetoed to be at most 60 cm away from the interaction point.

Figure 7.9 shows the computed ratios of the $Z/\gamma^{*}+ \geq n$ jet cross sections to the total inclusive Z/γ^{*} cross section for up to 4 jets per event in inclusive jet bins and comparisons with PYTHIA and MCFM Monte Carlo generators. There seems to be good agreement with Monte Carlo predictions up to 3 jets but hints of divergence of the PYTHIA Monte Carlo exist in the 3-jet bin which become clearer in the 4-jet bin.

Figure 7.10 shows the comparison between data and Monte Carlo of the highest, second highest and third highest jet p_T jet in the $Z/\gamma^{*}+$ ≥1, 2 and 3-jet samples respectively, with good agreement.

7.5 In a nutshell

Results for Z+jets cross section ratios versus inclusive number of reconstructed jets were presented in this chapter. For the $\sigma\left(Z+ \geq n\,\text{jets}\right)/\sigma\left(Z+ \geq (n-1)\,\text{jets}\right)$ ratios, as the jet multiplicity increases the PYTHIA predictions fail increasingly to describe what we observe in the data while the MadGraph simulations agree quite well. In the $\sigma(Z+ \geq n\,\text{jets})/\sigma(Z+ \geq 0\,\text{jets})$ case, again both tunes of the MadGraph samples match well and PYTHIA does not match well with data points; in some cases it appears to be within uncertainty, but this happens for the three and four reconstructed jets cases where statistics are limited.

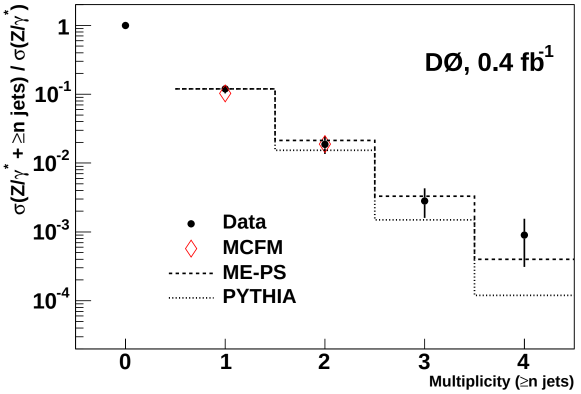


Figure 7.9: D0 results : Ratios of the $Z/\gamma^{*}+\geq n$ jet cross sections to the total inclusive Z/γ^{*} cross section versus jet multiplicity. Dark circles are data uncertainties. The dashed line represents predictions of LO Matrix Element (ME) calculations using PYTHIA for parton showering (PS) and hadronization. The dotted line represents the predictions of PYTHIA normalized to the measured $Z/\gamma^{*}+\geq 1$ jet cross-section ratio and the open diamonds represent MCFM predictions.

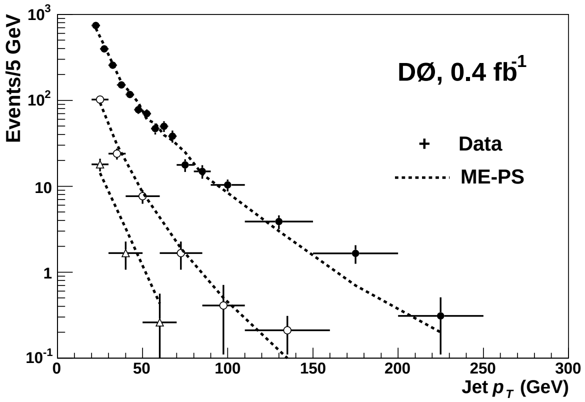


Figure 7.10: D0 results : Comparison between data and theory for the highest (dark circles), second highest (open circles) and third highest (open triangles) p_T jet distribution in the $Z/\gamma^{*}+\geq$1, 2 and 3-jet samples respectively.

I may not have gone where I intended to go, but I think I have ended up where I needed to be.

The Long Dark Tea-Time of the Soul
DOUGLAS ADAMS

Chapter 8

Synopsis

8.1 Summary and Conclusions

In the preceding pages, the full $Z \rightarrow (e^{+}e-)$+jets analysis was described. We presented the apparatus that made this analysis possible, the CMS detector, and the reconstruction algorithms giving us access to higher level objects built from the raw energy deposits recorded by the various subdetectors. We then focus on selecting a dielectron sample aiming for high event yield. Jet reconstruction was performed using *Particle Flow*, an algorithm designed to exploit information from all the subdetectors of CMS optimizing particle reconstruction and identification. Events were corrected for pileup and efficiency using data driven techniques, and were unfolded to correct for bin migrations due to detection efficiency, measurement resolutions and systematic biases.

In order to maximize yields an asymmetric selection approach was preferred, requiring a well-identified and isolated electron as the primary leg of the Z boson and keeping a looser second leg. One more reason for making the particular selection for the tight leg, was that this analysis was also part of a W/Z ratio measurement therefore the tight leg had to be kept synchronized with the W electron selection. Despite the much looser requirements on the second leg, assisted by the clean Z production signature we end up with an almost background-free sample.

This inclusive sample was then studied in exclusive jet bins on 0-4 jets, providing a good comparison of the jet associated Z production and how it compares with theoretical predictions. The cross sections as a number of jets were then unfolded and used to estimate the $\frac{Z+\geq n-jets}{Z+\geq(n-1)-jets}$ and $\frac{Z+\geq n-jets}{Z+\geq 0-jets}$ ratios and compared to different Monte Carlo generators showing very good agreement with Matrix Element (provided by MadGraph) plus Parton Shower (PYTHIA) Monte Carlo generator predictions.

The fact that the pure PYTHIA [31] samples do not model the data well is expected as it does not handle multiple jets in the matrix element and as a result PYTHIA predictions diverge increasingly from the data as the number of reconstructed jets increases. Using MadGraph [16] allows us to calculate matrix elements for events with a Z boson and up to four generated jets, originating from the hard scatter event in the final state, demonstrating good agreement with the data.

8.2 Outlook

As the LHC continues to collect data at full speed (the 2011 target integrated luminosity of 1 fm^{-1} was reached early June 2011) the measurement of Z+Jets event can become more precise and higher jet multiplicity bins can be studied. With the current integrated luminosity, there should be sufficient statistics to extend the measurement to 5 or more inclusive jets. Also, as the NLO event generators, with $2 \rightarrow 3, 4, 5...$ processes are becoming available more detailed comparisons will be possible. As more statistics also mean a more accurate jet p_T measurement, the unfolding procedure can be performed on the subleading jet as well.

Due to the high accuracy with which the Z boson can be detected and measured, along with the good knowledge we have of its production rate, it can always be used as a luminosity monitor of the LHC, as well as a standard candle for precise detector calibration. But most importantly, looking at the dielectron invariant mass spectrum might open a window to exciting new physics such as the hypothesized Z' boson, which is expected to decay in similar ways as the Z.

Bibliography

[1] D. W. Kerst. Acceleration of Electrons by Magnetic Induction. *Physical Review*, 58:841–841, November 1940.

[2] E. D. Courant, M. S. Livingston, and H. S. Snyder. The Strong-Focusing Synchroton-A New High Energy Accelerator. *Physical Review*, 88:1190–1196, December 1952.

[3] Donald H. Perkins. *Introduction to High Energy Physics*. Cambridge University Press, 4th edition, 2000.

[4] David Griffiths. *Introduction to Elementary Particles*. John Wiley & Sons, Inc, 1987.

[5] I J R Aitchison and A J G Hey. *Gauge Theories In Particle Physics, Volume I: From Relativistic Quantum Mechanics to QED*. Taylor & Francis Group, 2003.

[6] G. Karagiorgi, A. Aguilar-Arevalo, J. M. Conrad, M. H. Shaevitz, K. Whisnant, M. Sorel, and V. Barger. Leptonic CP violation studies at MiniBooNE in the $(3+2)$ sterile neutrino oscillation hypothesis. *Phys. Rev. D*, 75(1):013011, Jan 2007.

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

[8] Wikipedia: Article on Standard Model.

[9] S. D. Drell and T.-M. Yan. Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies. *Physical Review Letters*, 25:316–320, August 1970.

[10] Ryan Gavin, Ye Li, Frank Petriello, and Seth Quackenbush. FEWZ 2.0: A code for hadronic Z production at next-to- next-to-leading order. 2010.

[11] Alessandro Tricoli, Amanda M. Cooper-Sarkar, and Claire Gwenlan. Uncertainties on W and Z production at the LHC. 2005.

[12] W.J. Stirling, private communication.

[13] Alan D. Martin, R. G. Roberts, W. James Stirling, and R. S. Thorne. Parton distributions and the LHC: W and Z production. *Eur. Phys. J.*, C14:133–145, 2000.

[14] John M. Campbell, J. W. Huston, and W. J. Stirling. Hard Interactions of Quarks and Gluons: A Primer for LHC Physics. *Rept. Prog. Phys.*, 70:89, 2007.

[15] Hung-Liang Lai et al. Parton Distributions for Event Generators. *JHEP*, 04:035, 2010.

[16] Johan Alwall, Pavel Demin, Simon de Visscher, Rikkert Frederix, Michel Herquet, Fabio Maltoni, Tilman Plehn, David L. Rainwater, and Tim Stelzer. MadGraph/MadEvent v4: the new web generation. *Journal of High Energy Physics*, 2007(09):028, 2007.

[17] C. F. Berger et al. Vector Boson + Jets with BlackHat and Sherpa. *Nucl. Phys. Proc. Suppl.*, 205-206:92–97, 2010.

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

[19] CERN Press Release. LHC to run at 3.5 TeV for early part of 2009-2010 run rising later. 6 August 2009.

[20] CERN Press Release. Incident in LHC sector 3-4. 20 September 2008.

[21] The CMS Collaboration et al. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004, 2008.

[22] The CMS Collaboration. The CMS tracker system project: technical design report. *CERN-LHCC-98-006*.

[23] The CMS Collaboration. The CMS tracker: addendum to the technical design report. *CERN-LHCC-2000-016*.

[24] The CMS Collaboration. The CMS muon project, technical design report. *CERN-LHCC-97-032*.

[25] The CMS Collaboration. The TriDAS project, technical design report. Volume 1: The level-1 trigger. *CERN-LHCC-2000-038*.

[26] Ildefons Magrans de Abril, Claudia-Elisabeth Wulz, and Joao Varela. Conceptual design of the CMS Trigger Supervisor. *IEEE Trans.Nucl.Sci.*, 53(2):474–483, April 2006.

[27] XDAQ, a platform for the development of distributed data acquisition systems. `https://twiki.cern.ch/twiki/bin/view/CMS/XDaq_Main_Page`.

[28] Johannes Gutleber, Steven Murray, and Luciano Orsini. Towards a homogeneous architecture for high-energy physics data acquisition systems. *Computer Physics Communications*, 153(2):155–163, 2003.

[29] W3C standard, “Simple Object Access Protocol (SOAP)”. `http://www.w3.org/TR/soap/`.

[30] The CMS Collaboration. The TriDAS project, technical design report. Volume 2: Data acquisition and high-level trigger technical design report. *CERN-LHCC-2001-026*.

[31] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

[32] Z. Was. TAUOLA the library for tau lepton decay, and KKMC/KORALB/KORALZ/... status report. *Nucl. Phys. Proc. Suppl.*, 98:96–102, 2001.

[33] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton fragmentation and string dynamics. *Physics Reports*, 97(2-3):31–145, 1983.

[34] Rick Field. Early LHC Underlying Event Data - Findings and Surprises. 2010.

[35] Rick Field. Early LHC Underlying Event Data - Findings and Surprises. 2010.

[36] E. Boos et al. Generic user process interface for event generators.

[37] S. Agostinelli et al. G4–a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.

[38] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois. Electron reconstruction in CMS. *The European Physical Journal C - Particles and Fields*, 49(4):1099–1116, 2007. 10.1140/epjc/s10052-006-0175-5.

[39] R. Salerno. Electron reconstruction and identification in CMS at LHC. *Nuclear Physics B - Proceedings Supplements*, 197(1):230–232, 2009. 11th Topical Seminar on Innovative Particle and Radiation Detectors (IPRD08).

[40] Rainer Mankel. A concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 395(2):169–184, 1997.

[41] Hans Bethe and Walter Heitler. On the stopping of fast particles and on the creation of positive electrons. *Proceedings of the Royal Society of London, Series A*, 146(856):83–112, 1934.

[42] Jayesh H. Kotechaa and Petar M. Djuric. Gaussian Sum Particle Filtering. *IEEE Transactions on Signal Processing*, 51(10), 10 2003.

[43] The CMS Collaboration. Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. *CMS-PAS-PFT-09-001*.

[44] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- k t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, 2008.

[45] Matteo Cacciari. FastJet: A Code for fast k(t) clustering, and more. 2006.

[46] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys. Lett.*, B659:119–126, 2008.

[47] S. B. Beri, M. Grothe, L. Kaur Saini, P. Klabbers, C. Lazaridis, J. Leonard, and B. Dahmes. Trigger efficiency correction factors for inclusive W/Z cross section measurement. *CMS-AN-2010-246*.

[48] D. Barge, C. Campagnari, S. Gessner, P. Kalavase, D. Kovalskyi, V. Krutelyov, J. Ribnik, W. Andrews, D. Evans, F. Golf, J. Mulmenstadt, S. Padhi, Y. Tu, F. Wurthwein, A. Yagil, L. Bauerdick, I. Bloch, K. Burkett, I. Fisk, Y. Gao, O. Gutsche, and B. Hooberman. Study of photon conversion rejection at CMS. *CMS-AN-2009-159*.

[49] F. Beaudette, D. Benedetti, J. Branson, G. Daskalakis, E. Di Marco, C. Campagnari, C. Charlot, S. Harper, P.D. Kalavase, J.D. Keller, D. Lelas, P. Meridiani, M. Pieri, I. Puljak, R. Ranieri, N. Rompotis, C. Rovelli, R. Salerno, M. Sani, C. Seez, Y. Sirois, Y. Tu, A. Yagil, and S. Baffioni. Electron Identification in CMS. *CMS-AN-2009-178*.

[50] Niki Saoulidou. Particle Flow Jet Identification Criteria. *CMS-AN-2010-003*.

[51] John Campbell, Keith Ellis, and Ciaran Williams. MCFM - Monte Carlo for FeMtobarn processes.

[52] S. Dasu, J. Efron, K. Grogg, M. Grothe, P. Klabbers, C. Lazaridis, I. Ross, B. Dahmes, S. B. Beri, and L. Kaur Saini. Efficiency correction and unfolding for V+jets events. *CMS-AN-2011-017*.

[53] N. Adam, J. Berryhill, V. Halyo, A. Hunt, and K. Mishra. Generic Tag and Probe Tool for Measuring Efciency at CMS with Early Data. *CMS-AN-2009-111*.

[54] RooUnfold: ROOT Unfolding Framework. `http://hepunx.rl.ac.uk/~adye/software/unfold/RooUnfold.html`.

[55] Andreas Hoecker and Vakhtang Kartvelishvili. SVD approach to data unfolding. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 372(3):469–481, 1996.

[56] G. D'Agostini. A Multidimensional unfolding method based on Bayes' theorem. *Nucl. Instrum. Meth.*, A362:487–498, 1995.

[57] The CMS Collaboration. Rates of Jets Produced in Association with W and Z Bosons. *CMS-PAS-EWK-10-012*, (EWK-10-012), 2010.

[58] Measurement of the Inclusive Jet Cross Sections in $Z/\gamma^* \rightarrow (e^+e^-)$ +jets Production. `http://www-cdf.fnal.gov/physics/new/qcd/zjets_08/public.html`.

[59] T. Aaltonen et al. CDF Collaboration. Measurement of Inclusive Jet Cross Sections in $Z/\gamma * (\rightarrow e + e-) + jets$ Production in pp Collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev. Lett.*, 100(10):102001, Mar 2008.

Appendix A

Unfolding

A more comprehensive list of plots is presented here. For the case of unfolding the transverse momentum of the leading jet, a scan of the parameter space is shown:

- 1 - 5 iterations are considered for the Bayes method and
- SVD unfolding plots for k_{term} = 2 - 13.

Plots for both Monte Carlo and data are shown to demonstrate the effect of the unfolding procedure on both. The Monte Carlo plots have not been normalized to the data luminosity, hence the different scale on the y-axis between data and MC plots. The effect of jet resolution is also clear in this set of plots; we can observe it in the MC-only plots (left-hand side) where the (Unfolded MC)/(Generated MC) ratio shows an up-down effect between bins.

A.1 Bayes

Showing Monte Carlo (on the left) and data (on the right) plots to demonstrate the performance of unfolding using the Bayes method with 1 - 5 iterations.

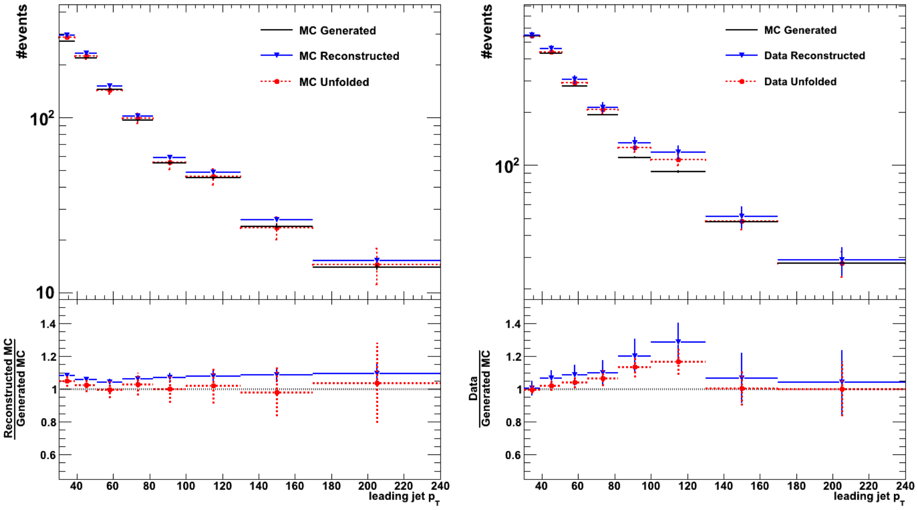


Figure A.1: Bayes unfolding : 1 iteration; Monte Carlo unfolding on the left, data on the right

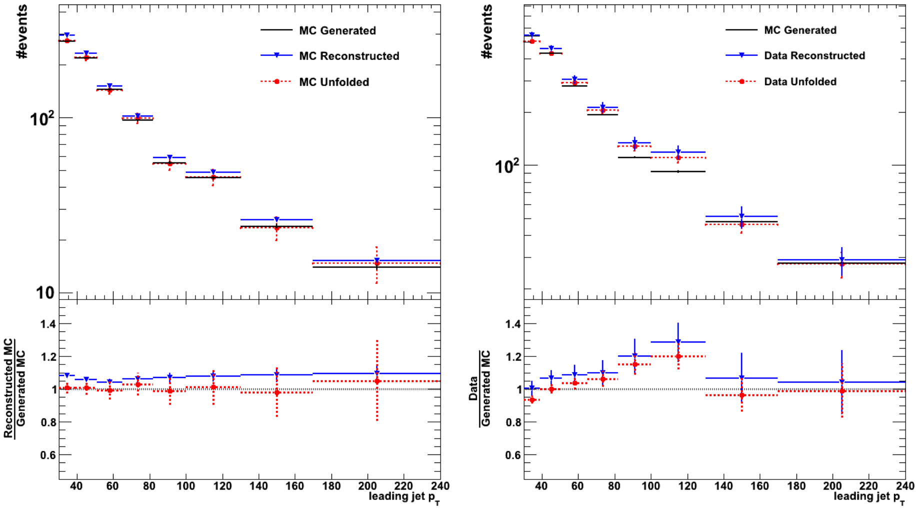


Figure A.2: Bayes unfolding : 2 iterations; Monte Carlo unfolding on the left, data on the right

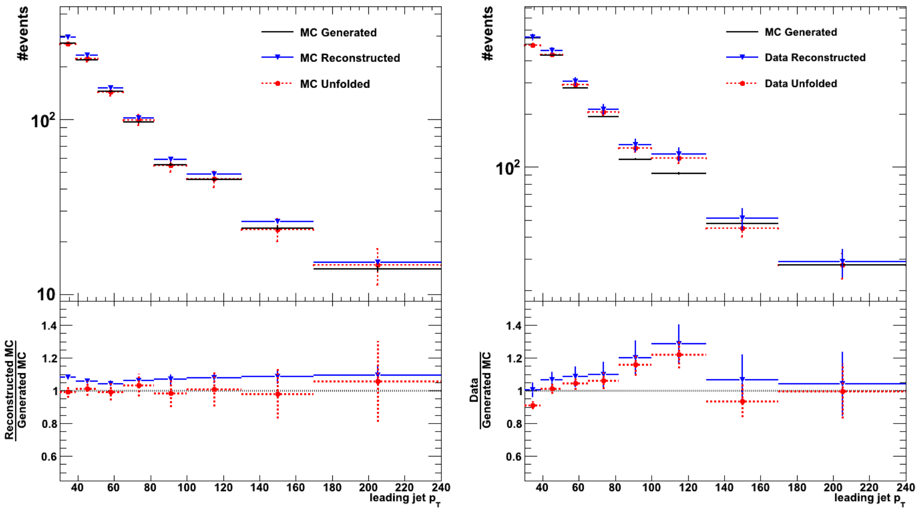


Figure A.3: Bayes unfolding : 3 iterations; Monte Carlo unfolding on the left, data on the right

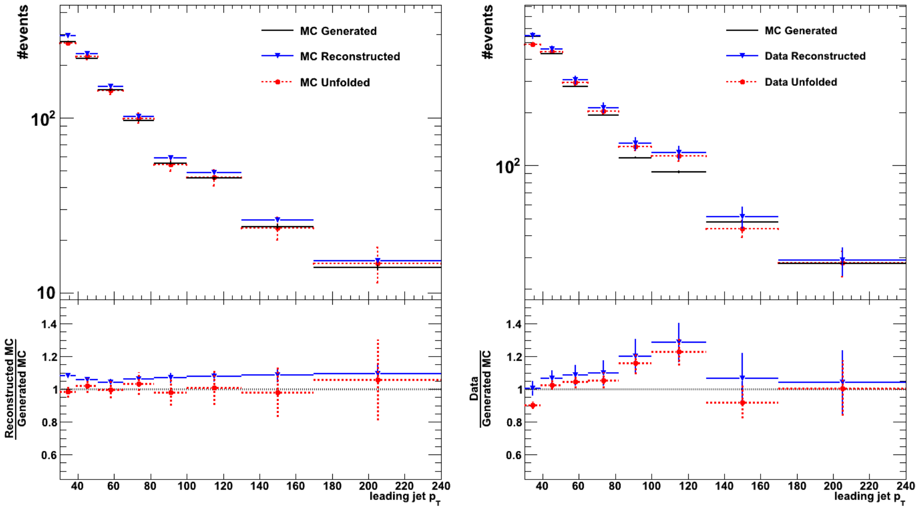


Figure A.4: Bayes unfolding : 4 iterations; Monte Carlo unfolding on the left, data on the right

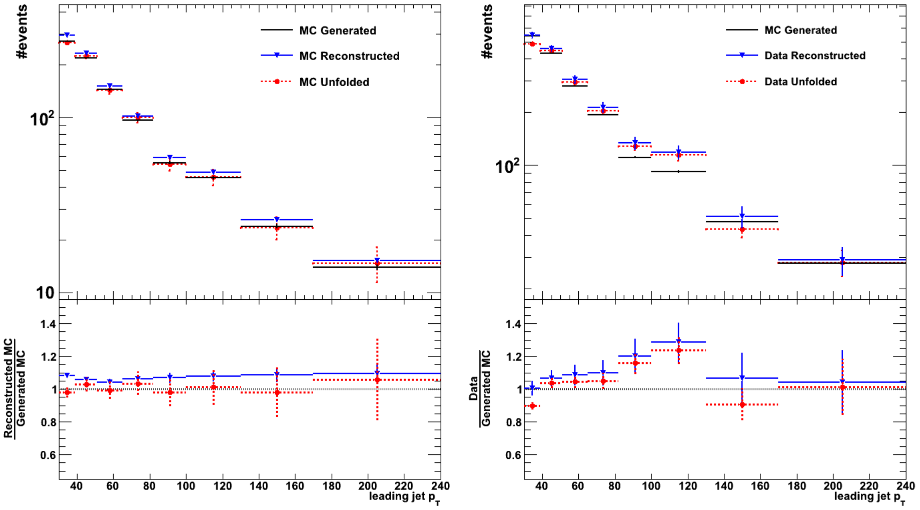


Figure A.5: Bayes unfolding : 5 iterations; Monte Carlo unfolding on the left, data on the right

A.2 SVD

Showing Monte Carlo (on the left) and data (on the right) plots to demonstrate the performance of unfolding using the SVD method with k_{term} = 2 - 13. The k_{term} = 1 case returns the training truth input without corrections. We observe the improvement up to k_{term} = 5 or 6 and the degradation at larger value.

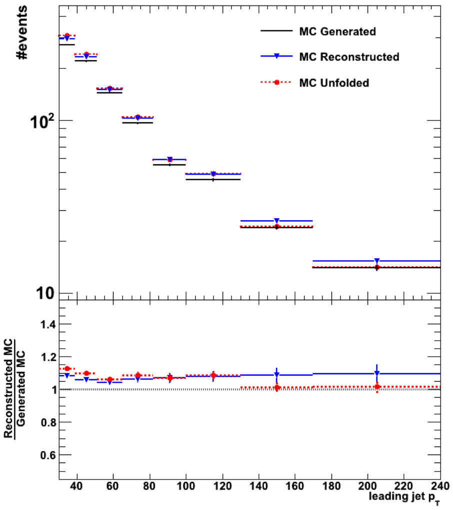


Figure A.6: SVD unfolding : k_{term} = 2; Monte Carlo unfolding on the left, data on the right

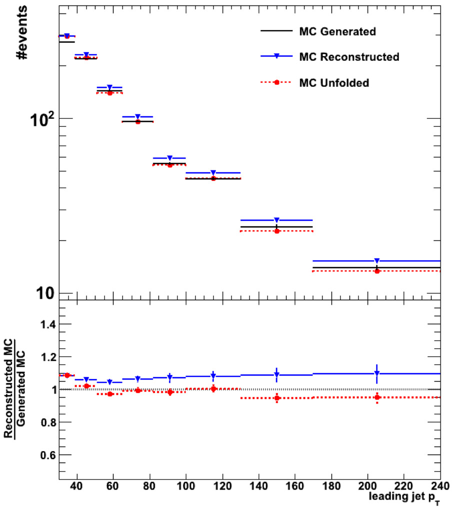


Figure A.7: SVD unfolding : k_{term} = 3; Monte Carlo unfolding on the left, data on the right

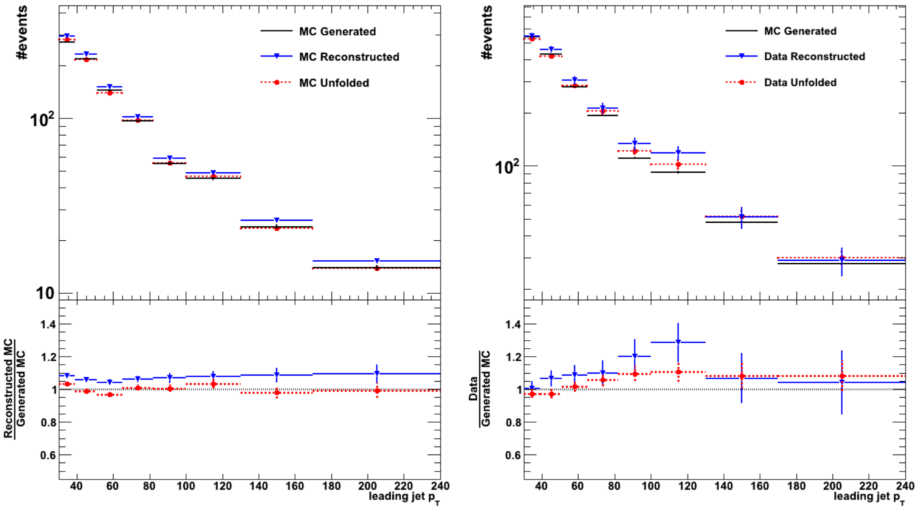


Figure A.8: SVD unfolding : k_{term} = 4; Monte Carlo unfolding on the left, data on the right

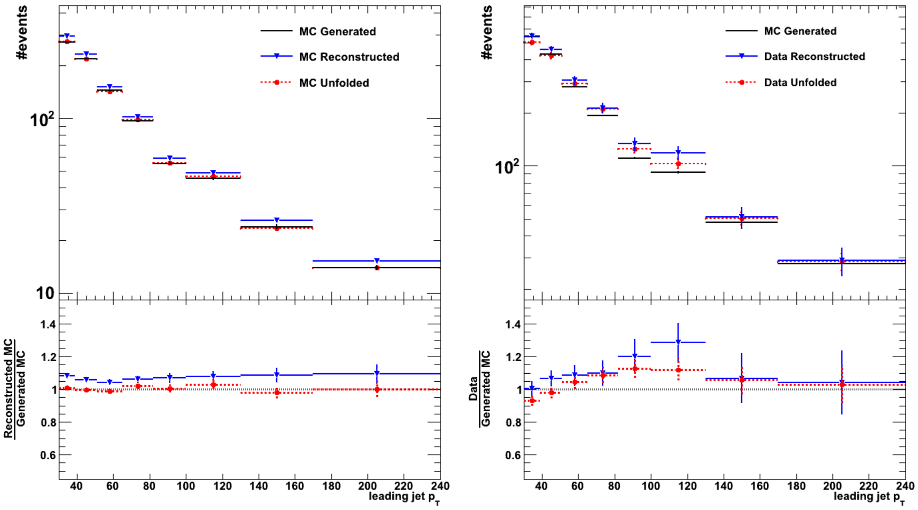


Figure A.9: SVD unfolding : k_{term} = 5; Monte Carlo unfolding on the left, data on the right

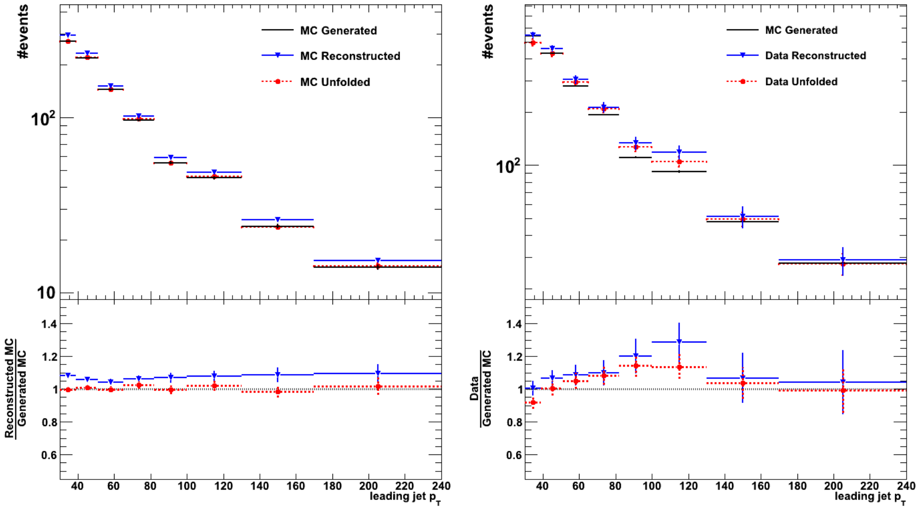


Figure A.10: SVD unfolding : k_{term} = 6; Monte Carlo unfolding on the left, data on the right

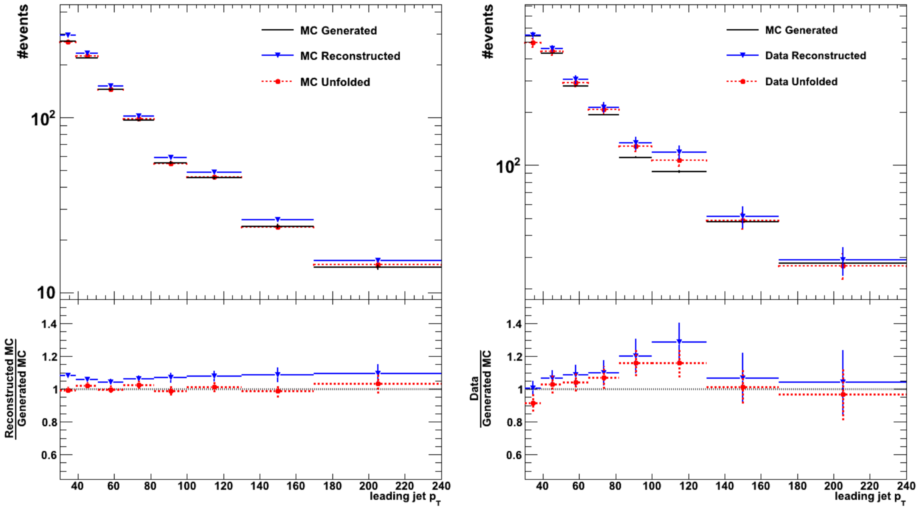


Figure A.11: SVD unfolding : k_{term} = 7; Monte Carlo unfolding on the left, data on the right

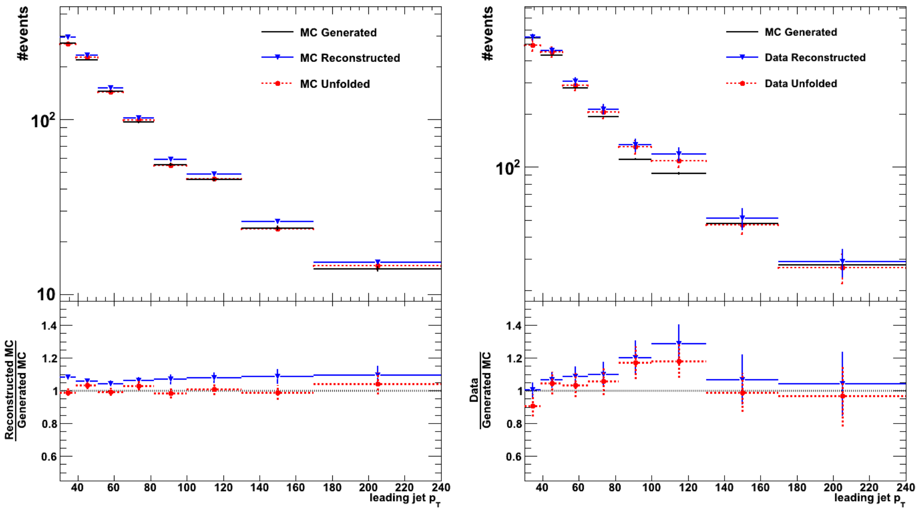


Figure A.12: SVD unfolding : k_{term} = 8; Monte Carlo unfolding on the left, data on the right

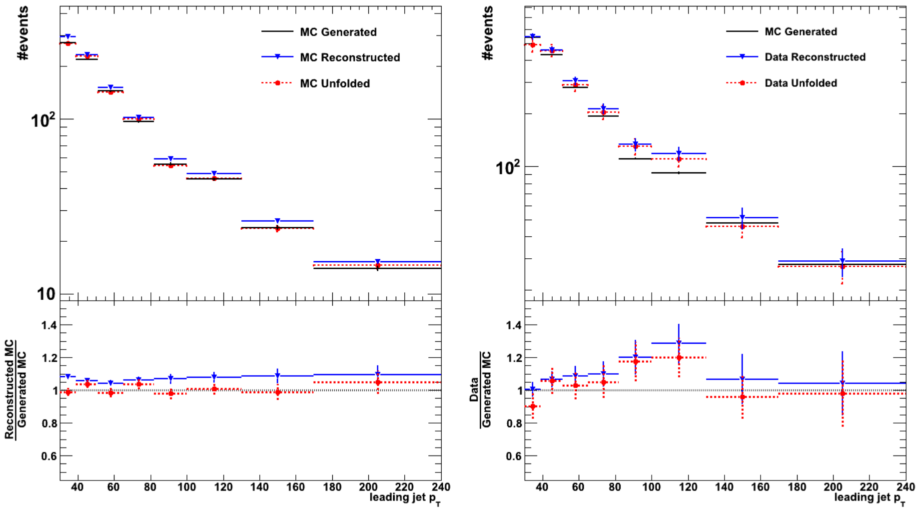


Figure A.13: SVD unfolding : k_{term} = 9; Monte Carlo unfolding on the left, data on the right

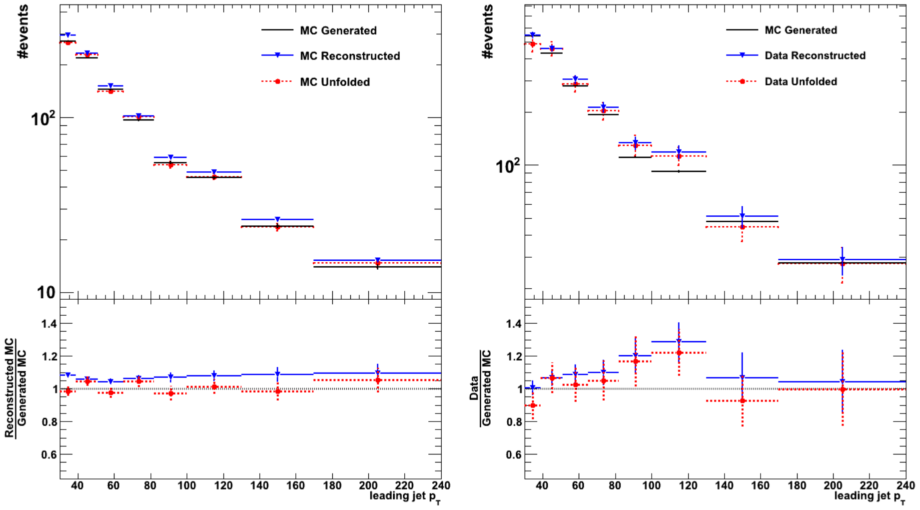


Figure A.14: SVD unfolding : k_{term} = 10; Monte Carlo unfolding on the left, data on the right

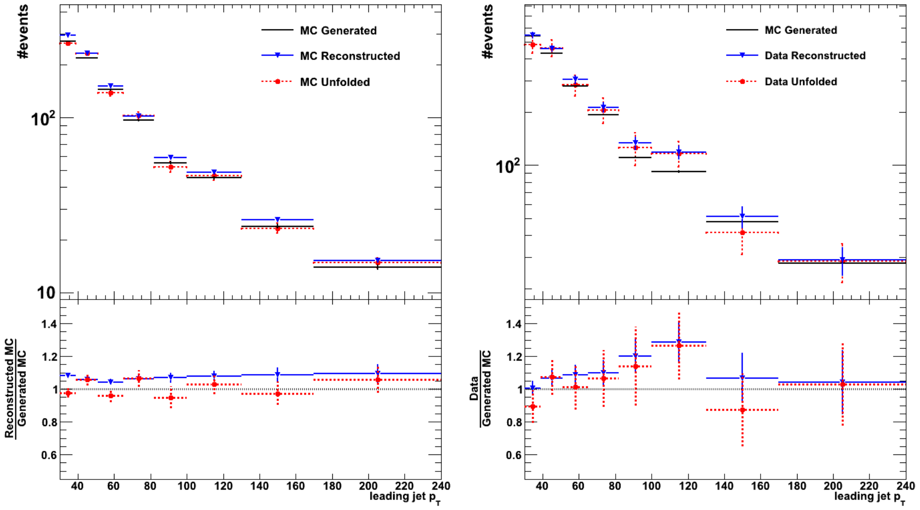


Figure A.15: SVD unfolding : k_{term} = 11; Monte Carlo unfolding on the left, data on the right

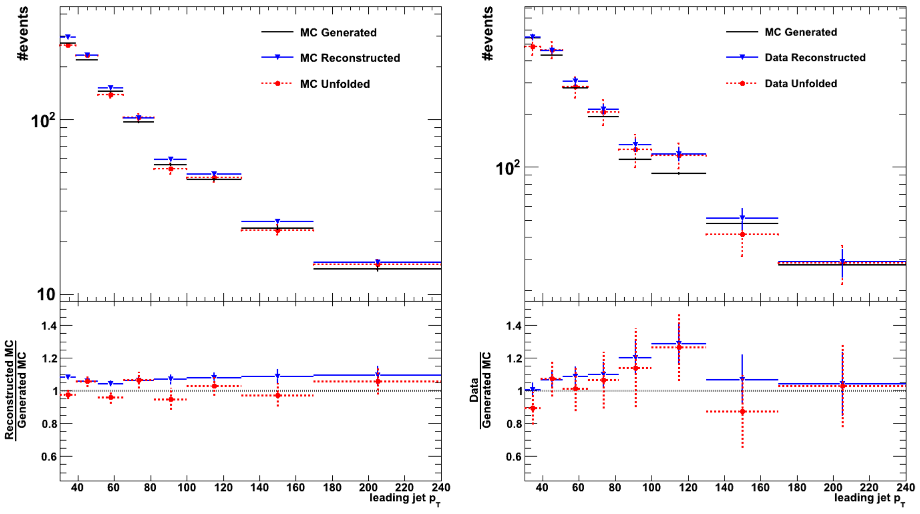


Figure A.16: SVD unfolding : k_{term} = 12; Monte Carlo unfolding on the left, data on the right

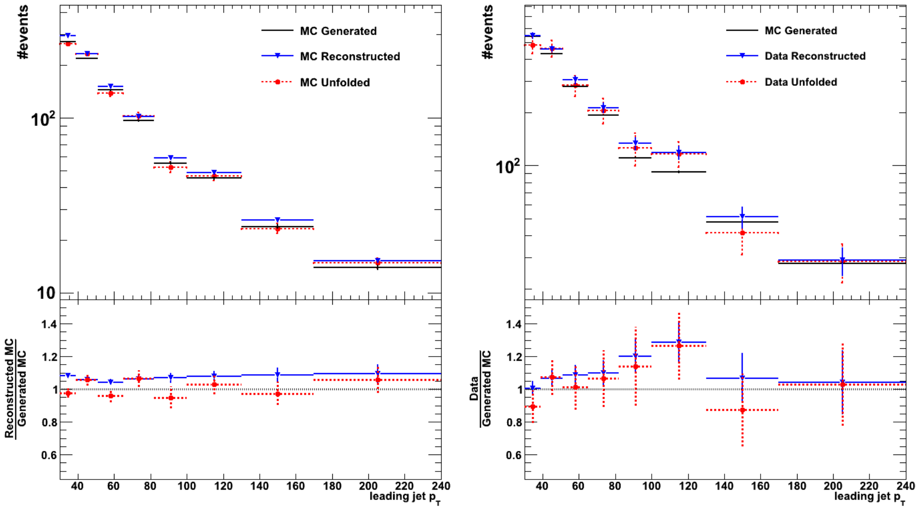


Figure A.17: SVD unfolding : k_{term} = 13; Monte Carlo unfolding on the left, data on the right

Appendix B

The Cruijff Equation

The Cruijff distribution is similar to that of a Gaussian function, with m_0 being the meanof the function, but the widths on the two sides of the mean (σ_L and σ_R) are allowed to vary, and an independent tail (α_L or α_R) is attached to each side:

$$f(x; m, \sigma_L, \sigma_R, \alpha_L, \alpha_R) = N_s \cdot e^{-\frac{(x-m_m)^2}{2\sigma^2+\alpha(x-m_0)^2}} \tag{B.1}$$

where $\sigma = \sigma_L(\sigma_R)$ for $x < m(x > m)$ and $\alpha = \alpha_L(\alpha_R)$ for $x < m(x > m)$.