University of Trieste

Faculty of Mathematical, Physical and Natural Sciences
Master of Science in Physics

M.Sc. Thesis

Study of the associated production of a Z boson and hadronic jets in proton-proton collisions at the CMS experiment

Supervisor:
Dr. Giuseppe Della Ricca

Candidate:
Davide Scaini

Assistant supervisor:
Dr. Matteo Marone

Academic Year 2010-2011

A Marisa e Pierluigi, esempi limpidi
in un mondo (pur)troppo mediocre.

My momma always said,
"Life is like a box of chocolates,
you never know what you're gonna get."

La mia l'ha detto una volta sola
"So che saprete arrangiarvi."

Contents

Introduzione

Questo lavoro di tesi è incentrato sullo studio della produzione associata di un bosone vettore Z e getti adronici, con particolare enfasi allo studio dei criteri di selezione degli eventi di interesse in un contesto di alta molteplicità di interazioni contemporanee (pile-up) caratteristico della fase attuale di presa dati al Large Hadron Collider.

I processi che coinvolgono la produzione di bosoni vettori (V) in associazione a getti adronici sono di grande importanza dell'ambito del programma di fisica del collisore adronico Large Hadron Collider. Lo studio dei getti adronici prodotti con W e Z costituisce un banco di prova importante nello studio della Cromodinamica Quantistica perturbativa. Inoltre, una misura precisa della sezione d'urto del processo V+getti è fondamentale perché questo processo costituisce un fondo significativo nella ricerca di "nuova fisica". I dati provenienti dagli esperimenti presso il Large Hadron Collider possono inoltre fornire delle misure interessanti per i processi di produzione di bosoni vettori con un numero di getti maggiore a tre, poiché le predizioni al secondo ordine in teoria delle perturbazioni al momento disponibili hanno una precisione variabile dal 10 al 30%.

Nel Capitolo 1 vengono presentati l'infrastruttura che compone l'acceleratore adronico Large Hadron Collider, l'esperimento Compact Muon Solenoid e le caratteristiche dei suoi sottorivelatori assieme al sistema di trigger e di acquisizione degli eventi. Nel Capitolo 2 viene esposta brevemente la cornice teorica in cui si inserisce questo studio. Nel Capitolo 3 sono descritti gli algoritmi per la ricostruzione e i criteri di selezione per elettroni e getti. Sono inoltre discussi i metodi per la correzione degli effetti di pile-up e dell'energia dei getti. Nel Capitolo 4 vengono mostrati i risultati ottenuti durante questo lavoro. In primo luogo sono riportate le operazioni preliminari all'analisi, che sono state l'oggetto principale di questo lavoro, ed in particolare: le operazioni di adattamento della distribuzione del numero di vertici d'interazione per evento alla distribuzione osservata nei dati, lo studio dei menu e delle selezioni di trigger, lo studio delle distribuzioni delle variabili d'isolamento per la selezione degli elettroni prima e dopo le procedure di correzione del pile-up e il confronto con la simulazione, e infine gli effetti su alcune distribuzioni d'interesse delle correzioni all'energia dei getti. Il dettagliato studio preliminare sulle selezioni degli elettroni in eventi ad alto numero di interazioni di pile-up ha permesso di ottenere una serie di risultati: un confronto tra simulazione e dati attraverso delle distribuzioni inclusive di test, la misura delle efficienze dei criteri di selezione, e le conseguenze della rimozione degli effetti del rivelatore (unfolding) su alcune distribuzioni di interesse, in particolare sulla distribuzione del numero di eventi di Z+getti al variare del numero di getti nell'evento.

Introduction

This thesis is focused on the study of associated production of a vector boson Z and hadronic jets, with particular emphasis on the study of the criteria for selecting the events of interest in a high multiplicity and pile-up environment characteristic of the current phase of the data taking at the Large Hadron Collider.

Processes involving vector bosons (V) in association with jets are central to the physics program of LHC. The study of jets with W and Z bosons provides a stringent and important test of perturbative QCD. In addition, a precise measurement of the V+n jets cross sections is crucial since these processes are a significant background in searches for new physics. Data coming from the experiments at the Large Hadron Collider may also provide precision measurements for the V+n jets (with n up to 4 for the W and 3 for the Z) processes, since the predictions currently available at next-to-leading order in perturbation theory are known with a precision varying from 10 up to 30%. This large range is due to uncertainties on the parton distribution functions and on the nature of perturbative calculations, which makes them dependent on the choice of renormalization and factorization scales.

In Chapter 1 the infrastructure that makes up the hadron accelerator Large Hadron Collider, the Compact Muon Solenoid experiment and the features of its subdetectors together with the trigger and data acquisition system are introduced. In Chapter 2 the theoretical framework in which this study is placed is briefly described. In Chapter 3 the reconstruction algorithms and selection criteria for electrons and jets are presented. The methods for the correction of the pile-up effect and the jets' energy are also discussed. Chapter 4 shows the results achieved during this work. First the preliminary steps of the analysis, which are the principal subject of this work are discussed, with particular detail on: the adaptation of the Monte Carlo distribution of the number of vertices to the measured in data, the study of the trigger menus, the isolation variables distributions for electrons before and after the pile-up corrections in comparison with the Monte Carlo simulation, and finally the effects of the jets' energy corrections on some specific distributions. The detailed preliminary study on electron selections in events with high pile-up produced a number of results: a comparison between simulation and data through some test inclusive distributions, the efficiency measurement for the selection criteria, and finally the removal of the detector effects (unfolding) for some distributions of interest, in particular for the distribution of the number of Z+jets events with varying number of jets in the event (Z yield).

Chapter 1

The LHC accelerator and CMS experiment

1.1 The LHC accelerator and collider

The Large Hadron Collider (LHC) is a two-ring superconducting hadron accelerator and collider built by CERN in the Geneva region.

The LHC is installed in the existing 26.7 km tunnel, that was constructed between 1984 and 1989 for the CERN LEP machine [1]. It is designed to yield head-on collisions of two protons (or ions) beams of 7 TeV (or 2.75 TeV per nucleon) each, with a luminosity of 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ (10^{27} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$).

Hadron colliders are well suited to the task of exploring new energy domains. The beam energy and the design luminosity of the LHC have been chosen in order to study physics at the TeV energy scale. A wide range of physics is potentially possible with the seven-fold increase in energy and a hundred-fold increase in integrated luminosity over the previous hadron collider experiments.

The prime motivation of the LHC is to elucidate the nature of electroweak symmetry breaking for which the Higgs mechanism is presumed to be responsible. The experimental study of the Higgs mechanism can also shed light on the mathematical consistency of the Standard Model at energy scales above 1 TeV. Various alternatives to the Standard Model invoke new symmetries, new forces or constituents. These discoveries could take the form of supersymmetry or extra dimensions or something really and surprisingly new.

In order to explore the physics at this scale, six experiments have been built along the LHC: *A Thoroidal Lhc ApparatuS* (ATLAS), *Compact Muon Solenoid* (CMS), *Large Hadron Collider beauty* (LHCb), *A Large Ion Collider Experiment* (ALICE), *LHC-forward*, *TOTal Elastic and diffractive cross-section Measurement* (TOTEM), Figure 1.1.
In two of the four interaction points are placed ATLAS and CMS, called "general purpose experiments", since their physics program is wide and various; one of their main goals is the search for the Higgs boson. LHCf and TOTEM are placed along the beam pipe after the interaction points, and are focused on studying particles produced at very small angle with respect to the beams.
In the remaining two interaction points are placed more specific experiments: ALICE is designed to study the products of the collisions between lead-lead ions and search

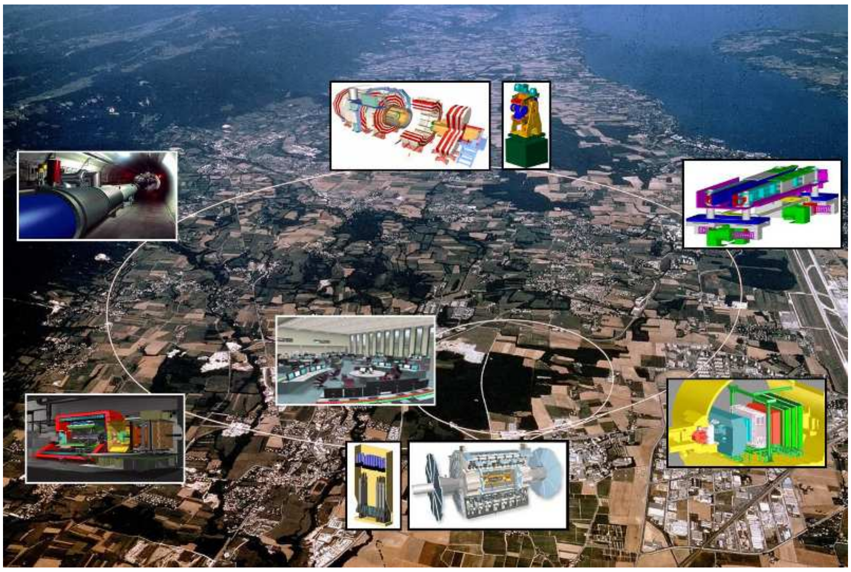


Figure 1.1: Aerial view of LHC with pictures of the experiments.

for a new state of the matter called "quark-gluon plasma"; LHCb is built to study in great detail the physics of the b-quark.

1.1.1 Accelerator structure

In this section are described some of the features of the LHC accelerator. The chain of CERN's accelerator complex that are needed to reach the energy of 7 TeV per-beam is depicted in Figure 1.2, and can be schematized as follows:

- *Linac2* it is a linear accelerator of approximate length of 36 m, that provides a beam (a bunch of varible length between 20 and 150 µs) of 50 MeV with a current of $\sim 150 - 180$ mA and a repetition rate of 1 Hz [2]
- *Proton Syncrotron Booster* it is a circular accelerator with a circumference of about 157 m which boosts the protons until 1.4 GeV
- *Proton Syncrotron* with a circumference of 628 m, increases the energy of the bunch to 26 GeV
- *Super Proton Syncrotron* has a circumference of 6.9 km and it is now used as the final injector for high-intensity proton beams into the Large Hadron Collider, accelerating protons from 26 GeV to 450 GeV.

The LHC beamline, unlike particle-antiparticle colliders that can have both beams sharing the same phase space in a single ring, is made out of two rings where the two counter-rotating beams circulate. The LHC uses twin bore magnets that consist of two sets of superconducting coils and beam channels within the same mechanical structure and cryostat, see Figure 1.3. The ring is made up of 1232 dipole bending magnets of 15 m length which maintain the bunches in the right orbit, and 392 quadrupole magnets, each 5–7m long, to focus the beams. The dipoles develop a field of 8.4 T perpendicularly with respect to the beam direction. This

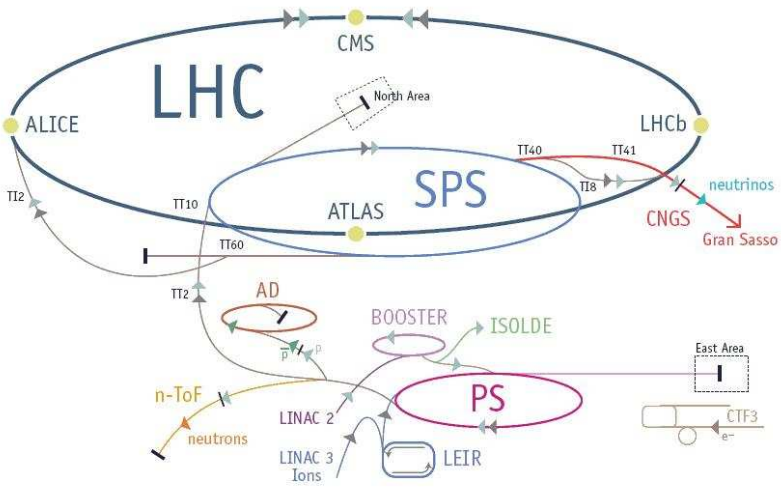
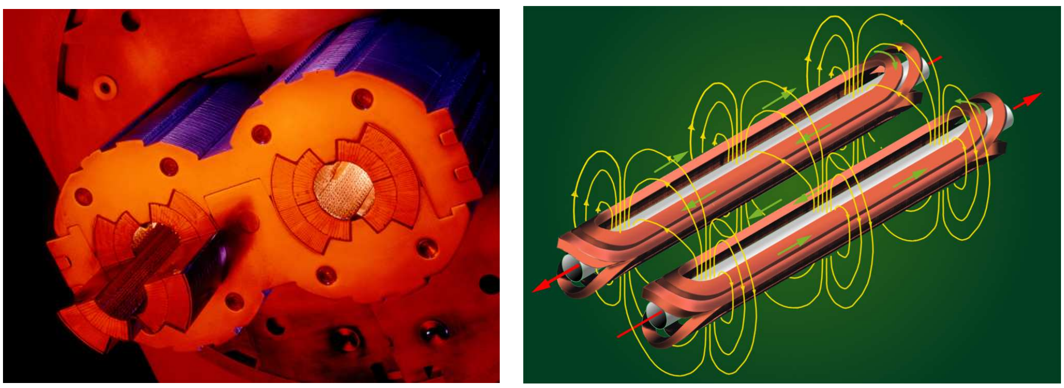


Figure 1.2: LHC accelerator's injection chain.



(a) Section of the magnet and coils **(b)** Magnetic field lines

Figure 1.3: LHC superconducting dipole coils.

field is reached by circulating the current in the superconducting regime, which is achieved by cooling every single magnet to 1.9 K with superfluid ^{3}He as coolant.

The two circulating beams cross each other in four interaction points, previously focused and collimated by special magnets (for more details see Chapter 3 in [1]) in order to increase the probability of interaction between the protons. The protons in the beams are clustered in *bunches* of about 10^{11} protons per bunch. Every beam will consist of 2808 bunches spaced from each other by 25 ns (nominal values); the expected rate of bunch-crossings is then 40 MHz. The machinery that is in charge of increasing the energy of the beam from 450 GeV to 7 TeV, keeping the bunches in sync, are eight radiofrequency cavities grouped in two modules made by four cavities each (Figure 1.4) operating at 400 MHz.

In addition to the beam energy, the other crucial parameter for a collider is the *instantaneous luminosity*:

$$\mathscr{L} = \frac{N_p^2 n_B f_{rev} \gamma}{4\pi \epsilon_n \beta^*} F$$

Figure 1.4: LHC radiofrequency cavities grouped in fours cryomodules with two cryomodules per beam.

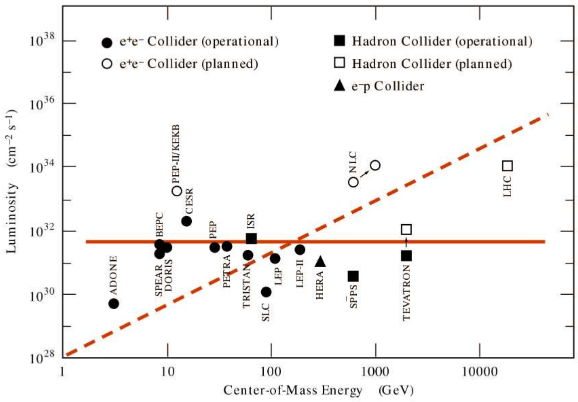
where N_p is the number of particles per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency, γ the relativistic gamma factor; ϵ_n is the normalized transverse beam emittance which describes the spread of the particles in the phase space, β^* the beta function at the collision point; lastly, F is the geometric luminosity reduction factor due to the crossing angle at the interaction point.
The LHC design instantaneous luminosity is 10^{34} cm$^{-2}s^{-1}$, see Figure 1.5. During the 2010 run LHC touched the peak luminosity of $2 \cdot 10^{32}$ cm$^{-2}s^{-1}$, while during 2011 the maximum of $5 \cdot 10^{33}$ cm$^{-2}s^{-1}$ (25 times the 2010 peak) was reached. This high luminosity has a direct influence on the number of interactions per bunch-crossing (see Figure 1.5a for comparison); the probability of having n interactions per bunch-crossing is described by a poissonian distribution

$$P(n) = \frac{(\mathscr{L} \cdot \sigma)^n}{n!} \exp^{-\mathscr{L} \cdot \sigma}$$

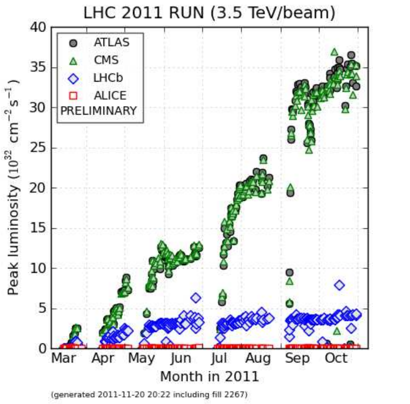
where $\mathscr{L}$ is the instantaneous luminosity and σ is the proton-proton total cross-section (σ_{tot}).

The final state of the n interactions depends on the details of the single process occurred. The probability of the single final state is given by its cross-section, in fact the total number of events for a given process is equal to the cross-section for this process times the integrated luminosity. Since the minimum-bias cross-section is much bigger than the one for any other process, the interactions will be almost entirely minimum-bias ones, see Figure 1.6. As a comparison, $\sigma_{minbias} \sim 70$ mb while the cross-section for an interesting process like the production of the Z boson is of the order of few nb [3]. The effect of having for each bunch-crossing (hopefully) one interesting vertex and then other minimum-bias vertices is called *pile-up*, and the amount of pile-up is determined essentially by the instantaneous luminosity.

The amount of interactions produced during a data taking period (called *run*) is represented by the *integrated luminosity*, that is the integral of the luminosity with respect to time. The total number of events for a given process is $N_{tot} = \sigma_{process} \int \mathscr{L} dt$. Therefore the integrated luminosity is a key measure in particle physics; during 2010 the LHC delivered around 47 pb^{-1}, while in 2011 5.7 fb^{-1}, see Table 1.1.



(a) Peak luminosities of several existent and planned machines (the dashed line indicates luminosity increasing as the square of center-of-mass energy) [4].



(b) Evolution of the instantaneous luminosity during 2011 at LHC.

Figure 1.5: Relative and absolute luminosity conditions at LHC.

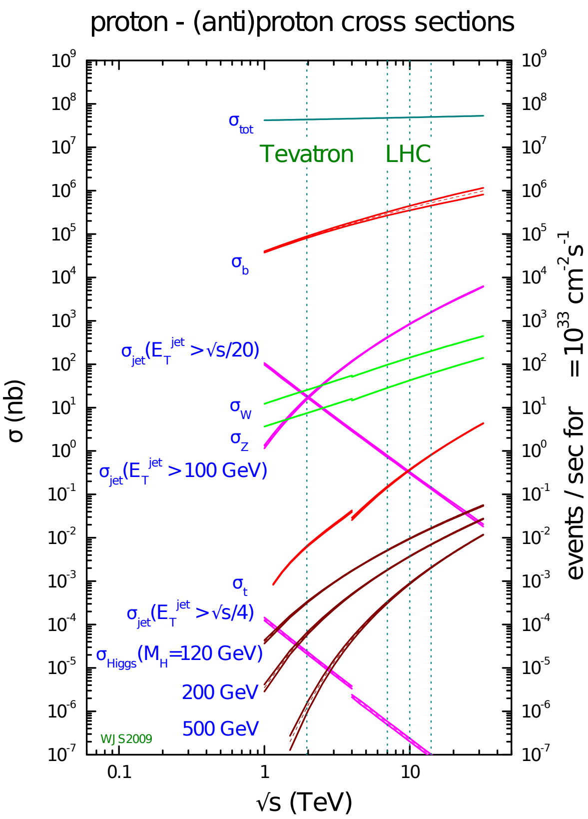


Figure 1.6: Energy dependence of some characteristic cross-section at hadron colliders.

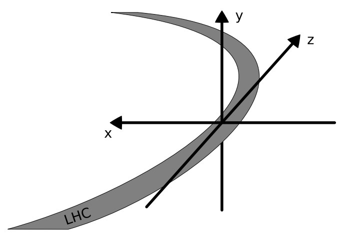
parameters	**p-p design**	**p-p 2011**
E_{beam}	7 Tev	3.5 TeV
$B_{dipoles}$	8.33 T	8.33 T
Δt_{bunch}	25 ns	50 ns
n_B	2808	changing
N_p	1.1×10^{11}	changing
$\mathscr{L}$	10^{34} cm$^{-2}s^{-1}$	5×10^{33} cm$^{-2}s^{-1}$ (peak)
pile-up	20 (mean)	6 (mean)

Table 1.1: Summary of the design and actual key parameters of the LHC machine.

1.1.2 Coordinate system and kinematic variables

There are some conventions to be clarified as far as the coordinate system and kinematic variables used in LHC are concerned.

Imagine to be at one of the four interaction points, the z-axis is the tangent to the accelerator orbit at the interaction point. The x-axis and y-axis lay on the plane perpendicular to the z-axis; the positive x-axis direction points to the center of the accelerator's ring, while the y-axis is directed upward. The coordinate system is then right-handed.



Some useful variables:

- distance from the beam axis $r = \sqrt{x^2 + y^2}$
- azimuth angle $\phi = \arctan(x/y)$
- polar angle $\theta = \arctan(y/z)$
- pseudorapidity $\eta = -\ln(\tan(\theta/2))$

As praxis, the coordinates used to locate a point in space around the interaction point are (r, ϕ, η), in particular the two (ϕ, η) identify the direction of an outgoing particle (given the interaction point).
Since it is not possible to know the exact parton momentum, the usual practice is to consider as neglegible the transverse component with respect to the longitudinal one. For this reason it is useful to define some other kinematical variables:

- particle transverse momentum $p_T = p \sin\theta$ [GeV/c]
- particle transverse energy $E_T = \sqrt{p_T^2 + m^2c^4} \simeq p_T c$ [GeV]
- missing transverse energy $MET = -\sum_{i=1}^{n} p_T^i c$ [GeV]

Since the transverse momentum (and the transverse energy) of the initial state is neglegible, the sum over the transverse momentum of all final particles should be null, because of the momentum conservation principle. For several reasons (from geometric acceptance to the weakly interacting particles, e.g. neutrinos) it is possible to measure the momentum of all particles in final state, so the missing transverse energy is frequently remarkably different from zero.

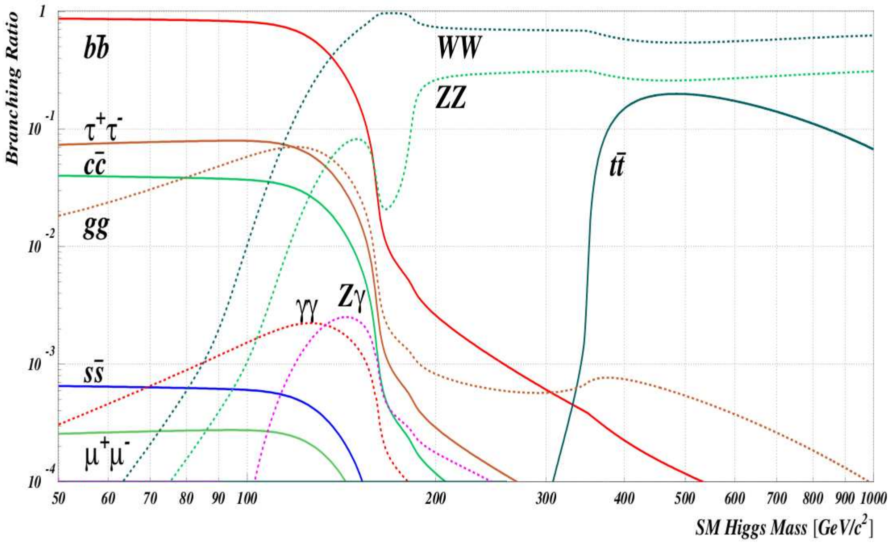


Figure 1.7: Branching Ratio for several decay channels for the Higgs boson versus its mass.

1.1.3 Scientific goals

The LHC collider has been built to study in great detail the already known physics of the Standard Model (SM), and to investigate the theoretical predictions beyond the SM. Some examples, focussed on the CMS goals, are listed below.

Tests of the Standard Model

The LHC will continue testing the SM, performing high precision measurements on electroweak and QCD physics. Some interesting aspects are for example the measure of the CKM matrix elements, the study of CP-symmetry violation and the top-quark physics.

Higgs Boson search

In SM the mechanism that gives a mass to the matter constituents makes use of the spontaneous symmetry breaking of $SU(2)_L \otimes U(1)_Y$, and it is called *Higgs Mechanism*. This model predicts the existence of a scalar particle called *Higgs boson*, whose mass is not yet known. The production cross-section of the Higgs boson is obscured by other background processes, electroweak and QCD ones, with the same final state. Figure 1.7 shows the branching-ratios of several decay channels of the Higgs boson that determines final states relevant to the search of this particle.

Although there is no theoretical prediction for the Higgs boson mass, there are some experimental and theoretical arguments to limit its mass range. In the low mass region, the experiments at LEP excluded the existence of a Higgs boson up to 115 GeV; for theoretical reasons the limit for an heavy Higgs is at around 800 GeV. In 2010 the experiments CDF and D0 at the Tevatron accelerator excluded the mass range between 100-109 GeV and 158-175 GeV with 95 % confidence level (CL) [5]. More recently CMS and ATLAS at LHC presented some improvements; quoting the results of CMS, has been excluded the region between 127-600 at 95 % CL and

128-525 at 99 % CL by combining several decay channels [6].
The non-excluded region remains between 115 and 128 GeV. In this range the clearest channels for the SM Higgs boson discovery are the decay in two photons and in four leptons; both ATLAS and CMS have the possibility to access this channel thanks to the energy resolution of their electromagnetic calorimeters.

Supersymmetry

The simplest theories that extend the SM are the supersymmetric ones (SUSY). These theories are based on a special group of transformations. Writing the SM Lagrangian and requiring to be invariant under this group of transformations, gives a theory with all the SM particles plus a supersimmetric partner (*superpartner*) for each one, that differs by half a unit of spin.

The SUSY is interesting because it has some nice features built-in, such as it:

- solves the "hierarchy problem" between weak and gravitational forces of the Standard Model (if SUSY exists at the TeV scale)
- allows for the high-energy unification of the couplig constants for weak, strong and electromagnetic interactions
- provides a candidate for Dark Matter

The Minimal Supersymmetric Standard Model is one of the best studied candidates for physics beyond the Standard Model.

Super massive gauge vector bosons

Several theoretical SM extensions predict the existence of Z and W gauge vector bosons supermassive partners, called Z′ and W′. Like their "light" partners, they can decay in two leptons, a lepton and a neutrino and quark-antiquark.

A W′ boson could be detected at hadron colliders through its decay to lepton plus neutrino or top quark plus bottom quark, after being produced in quark-antiquark annihilation. The LHC reach for W′ discovery is expected to be a few TeV. The Z′ boson would be produced by quark-antiquark annihilation and decay to an electron-positron pair or a pair of opposite-charged muons. The most stringent current limits come from ATLAS experiment, which excludes Z′ bosons up to masses of 1.83 TeV (for Sequential Standard Model) [7].

Extra-dimensions

Since the SM does not actually include the gravity, and does not unify all the known interactions, there exist some theoretical efforts to identify a mechanism that can unify the fundamental forces. One of the possible available theories uses more general geometrical spaces with dimension greater than four. Depending on the number of extra dimensions and the theoretical formalism used, CMS collaboration sets lower limits on the effective Planck scale in the range of 2.3-3.8 TeV at the 95% confidence level [8].

Heavy-ions physics

In addition to proton-proton collisions, LHC will produce lead-lead ions collisions. Heavy ions collisions are characterized by the high number of parton-parton interactions for each collision, due to the high density of the constituent partons.

Quarks and gluons are bound in color-singlet baryons and mesons, because of the non-abelian nature of QCD and the phenomenon is called color confinement, or, in a more narrow sense, quark confinement. Because of the mentioned peculiarity of heavy ions collisions, it should be possible to access a state of hadronic matter where quarks and gluons are not confined: this state is called *Quark Gluon Plasma* (QGP). To identify the QGP it is necessary to measure precisely the energy of hadronic jets, since the presence of QGP affects the hadronization process changing the transverse momentum and pseudorapidity distribution of the jets.

1.2 The CMS experiment

The total proton-proton cross-section at $\sqrt{s} = 14$ TeV is expected to roughly be 100 mb. At the design luminosity the general-purpose detectors such as CMS will therefore observe an event rate of approximately 10^9 inelastic events/s. This leads to a number of formidable experimental challenges. The online event selection process (trigger) must reduce the huge rate to about 100 events/s for storage and subsequent analysis. The short time between bunch crossings, 25 ns, has major implications for the design of the read-out and trigger systems. These conditions also require a very careful design of the detectors [9].

At the design center-of-mass energy and luminosity, an average number of about 20 inelastic collisions will be superimposed on the event of interest. This implies that around 1000 charged particles will emerge from the interaction region every 25 ns. The products of an interaction under study may be confused with those from other interactions in the same bunch crossing. This problem becomes more severe when the response time of a detector element and its electronic signal is longer than 25 ns. The effect of this pile-up can be reduced by using high-granularity detectors (ensuring low occupancy) with good time resolution. This implies a large number of detector channels. The resulting millions of detector electronic channels require very good synchronization. The large flux of particles coming from the interaction region leads to high radiation levels, demanding radiation-hard detectors and front-end electronics.

The detector requirements for CMS to meet the goals of the LHC physics programme can be summarised as follows:

- high charged-particle momentum resolution and reconstruction efficiency in the inner tracker, efficient triggering and offline tagging of τ leptons and b-jets, requires pixel detectors close to the interaction region (see Section 1.2.2)
- high electromagnetic energy resolution, good diphoton and dielectron mass resolution ($\approx 1\%$ at 100 GeV), wide geometric coverage, π^0 rejection, and efficient photon and lepton isolation at high luminosities (see Section 1.2.3)
- high missing-transverse-energy and dijet-mass resolution, requiring hadron calorimeters with a large hermetic geometric coverage and with fine lateral

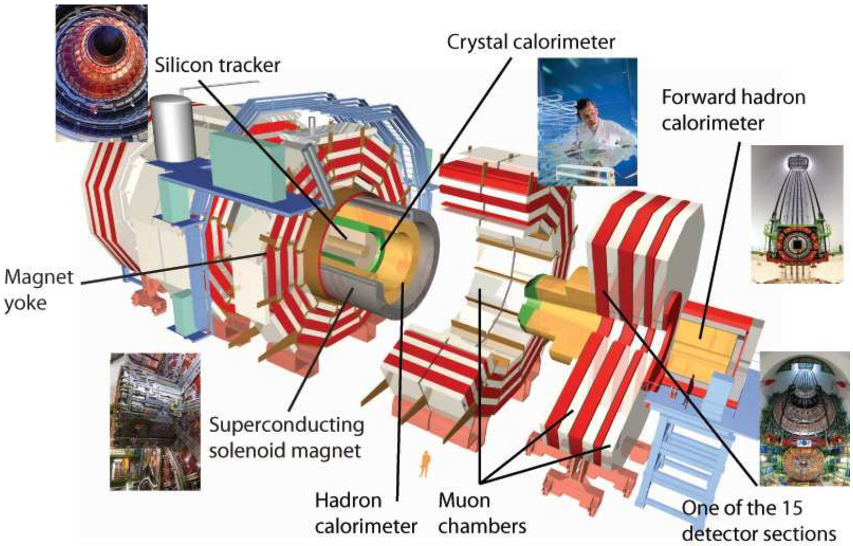


Figure 1.8: CMS cutaway with its detectors.

segmentation (see Section 1.2.4)

- high muon identification and momentum resolution over a wide range of momenta and angles, good dimuon mass resolution ($\approx 1\%$ at 100 GeV), and the ability to determine unambiguously the electric charge of muons with momentum $p < 1$ TeV (see Section 1.2.5)

The overall dimensions of the CMS experiment are a length of 21.6 m, a diameter of 14.6 m and a total weight of 12500 t, see Figure 1.8 for a cutaway of CMS with its detectors. Its structure is typical of detectors installed at colliders; it is composed of different cylindrical devices coaxial with the beam pipe in the central region, referred to as barrel. At both ends of the cylinder detector planes are installed, perpendicular to the beam, the so called endcaps ensuring a very high detector hermeticity. At the core of the CMS detector sits a high magnetic field and a large-bore superconducting solenoid surrounding an all-silicon pixel and strip tracker, a lead-tungstate scintillating-crystals electromagnetic calorimeter, and a brass-scintillator sampling hadron calorimeter. The iron yoke of the flux-return is instrumented with four stations of muon detectors covering most of the 4π solid angle. Forward sampling calorimeters extend the pseudo-rapidity coverage to high values ($|\eta| \leqslant 5$) assuring very good hermeticity.

Schematically (see Figures 1.8 and 1.9 as reference):

- **Tracker** $r < 1.2$ m, $|\eta| < 2.5$: the silicon tracker consists of a pixel vertex detector and an outer Silicon Strip Tracker (SST) to reconstruct charged particle tracks and individuate primary and secondary vertices.

- **ECAL** $1.2 < r < 1.8$ m, $|\eta| < 3$: the Electromagnetic CALorimeter precisely measures electrons and photons energy, and is constituted of $PbWO_4$ scintillating crystals and a forward preshower detector.

- **HCAL** $1.8 < r < 2.9$ m, $|\eta| < 5$: the hadron calorimeter system is dedicated

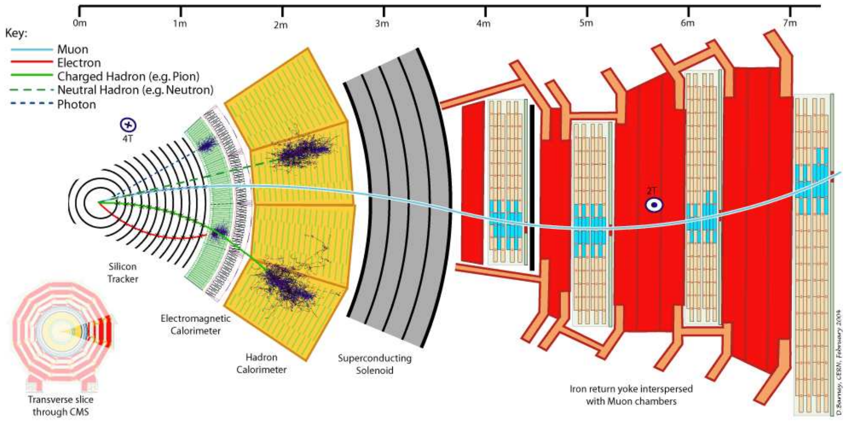


Figure 1.9: CMS experiment slice, where the sub-detectors are depicted along the radius.

to jet position and transverse energy measurements; it is extended in the forward region $3 < |\eta| < 5$ with a very forward hadron calorimeter (HF).

- **Magnet** $2.9 < r < 3.8$ m, $|\eta| < 1.5$: the magnet is large enough to accommodate the calorimeters and the inner tracker, providing a 4 T longitudinal magnetic field supplied by a superconducting solenoid.
- **Muon system** $4 < r < 7.4$ m, $|\eta| < 2.4$: the muon chambers are merged inside the magnet return yoke to detect and reconstruct muon tracks, and are composed of Drift Tubes (DT) in the barrel and Cathode Strip Chambers (CSC) in the endcaps and complemented overall up to $|\eta| < 2.1$ by Resistive Plate Chambers (RPC).

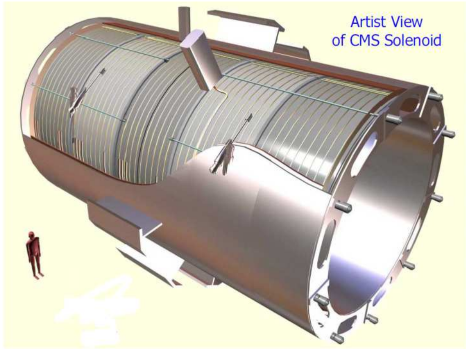
In the following sections are described in more detail the detectors that compose the CMS experiment, focussing on their characteristics that make them suitable for the targeted physics.

1.2.1 Superconducting solenoidal magnet

An important aspect of the overall detector design is the magnetic field configuration. Large bending power is required to precisely measure high-momentum muons and other charged particles. The choice of the magnet structure strongly influences the remaining detector design.

The CMS collaboration has chosen a long superconducting solenoid ($L = 12.5$ m) with an inner diameter of 5.9 m generating a uniform magnetic field of 4 T, see Figures 1.8 and 1.9. These characteristics make this superconducting solenoid the largest and most powerful one ever designed, see Figure 1.10.

The favourable dimensional ratio (length/radius) of the solenoid and the high magnetic field provide the necessary bending power for precision charged particle tracking, and efficient muon detection and measurement up to rapidities of 2.5. Thus the muon spectrometer uses a single magnet, simplifying the detector design. The inner coil diameter is large enough to accommodate the tracker and the calorimeters. The magnetic flux is returned through a 1.8 m thick saturated iron yoke (with a 1.8 T return field) instrumented with four layers of muon chambers.



(a) Superconducting magnet, artistic view.

(b) Magnet section, it is possible to see the magnet spires and the Al thermal stabilizer.

(c) Magnet mounted in CMS.

Figure 1.10: Different pictures and details of the CMS solenoidal magnet.

The magnet conductor consists of three concentric parts: a central flat superconducting cable, a high purity aluminium stabiliser and an external aluminium-alloy to reinforce the sheath. The superconducting cable is a Rutherford type with 40 NiTb strands sourrounded by pure Al as a thermal stabilizer, and is kept cooled at 1.9 K by a liquid helium cryogenic system. This structure is depicted in Figure 1.10. The current circulating in the magnet is around 20 kA, with a stored energy of 2.7 GJ.

1.2.2 Tracking system

The silicon tracker is the inner detector of CMS and represents an essential detector to address the large number of LHC physics goals. It extends in the region $|\eta| < 2.5$, $r < 1.2$ m, $|z| < 2.7$ m (see Figures 1.8 and 1.9) and it is completely based on semiconductor detectors made of silicon, covering the largest ever designed Si detector surface of 198 m^2, [10] [11]. Using the tracker, vertices and charged particle tracks must be reconstructed in the highly congested LHC environment.
In the absence of a magnetic field, the charged track density simply falls off as the inverse of the squared distance from the interaction point. Under the effect of the 4 T field, the decrease is initially more gradual and then significatively more pronounced than $1/r^2$.

To better solve the pattern recognition problem, the tracker is designed to fulfil two basic properties: low cell occupancy and large hit redundancy. For these reasons it is structured into an *inner silicon pixel detector* surrounded by *several layers*

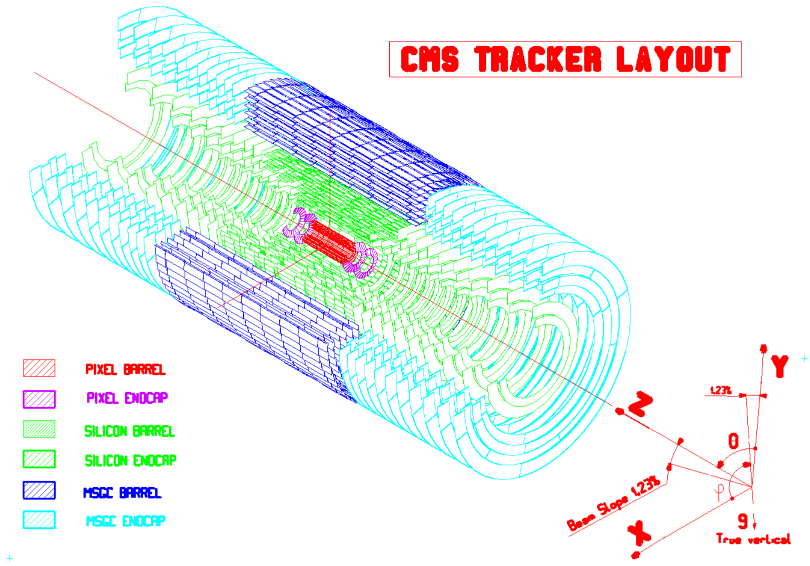


Figure 1.11: Prospective view of the CMS tracker layout.

of silicon microstrip detectors of different size and pitch between the strips, see Figure 1.11. The low occupancy is obtained by working with: high granularity detectors (mainly the ones closer to the interaction point because they have to cope with higher particle fluxes), and fast primary charge collection (obtained with thin detectors and overdepleting the silicon bulks).

Another strong constraint imposed on the tracker comes from the material budget that should be reduced in order not to degrade the electromagnetic calorimeter performance, in particular for the detection of the Higgs boson in the channel $H \to \gamma\gamma$. This stringent requirement limits the allowed total number of active layers and determines both the amount and type of material and the cable routing layout.

Silicon pixel detector

The CMS pixel system is the detector closest to the interaction point ($r < 10$ cm) and consists of two barrel layers and two disks for each endcap. The layers are composed of modular detector units, each one consisting of a thin silicon sensor segmented with n^+-pixels on a n-type substrate.

Since the main task required to the pixel system is to measure track impact parameters, both $r\phi$ and z coordinate will be important, depending on the topology of the vertex. For this reason, a square pixel shape has been chosen so as to optimize the spatial resolution measurement in both coordinates simultaneously. Moreover, in the region covered by this detector the particle flux can reach the frequency of 10^7 cm^{-2}s^{-1}, so it is necssary to have a great space resolution and granularity to be able to distinguish the main vertex from the pile-up ones.

The resolution can be much improved by exploiting charge sharing effects, together with an analog signal readout, to profit from position interpolation between pixels sharing the hit signal.

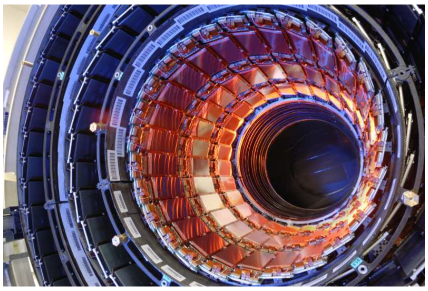


Figure 1.12: The CMS microstrip silicon tracker.

In the barrel, with a sensitive detector thickness between 200 and 250 µm and by deliberately not tilting the detectors, the natural pixel size will be around 150 µm × 150 µm giving optimal resolution and cluster size conditions. In the endcap disks (up to $|\eta| < 2.4$), the electric and magnetic fields are parallel and most tracks are close to normal incidence. For these reasons, the detectors will be rotated by 20 degrees around their radial symmetry axis, such that both the tilt angle and the induced Lorentz effects increase the charge sharing among neighbouring pixels in both the r and $r\phi$ directions.
In this way, in spite of pixel dimensions of 150 µm, a spatial resolution of 10 µm is expected in the $r\phi$ direction and for inclined tracks in the barrel, while 15 µm are expected in both r and $r\phi$ coordinate in the endcaps.

Microstrip silicon tracker

The silicon tracker covers the radial region between 20 and 60 cm. The modules are arranged in a spiral structure in the barrel and in disks in the forward regions, and are supported by carbon fibre wheels, Figure 1.12. The silicon tracker is essentially composed by four sections:

- **TIB** *(tracker inner barrel)*: it is the central and most internal part of the detector ($|z| < 65$ cm), and is composed by four layers of micro-strips. The strips width varies from 80 µm to 120 µm, while the ratio of the strip pitch to the strip width is 0.25 for all types [12]. The attained resolution is between 23 and 34 µm in r and ϕ, and around 230 µm along the z-axis.

- **TOB** *(tracker outer barrel)*: it is the external part of the barrel (having the same length along z-axis of the TIB), and it is made of six layers of micro-strips with different properties from the TIB ones. The achieved resolution varies from 35 to 52 µm in r and ϕ, while of around 530 µm along the z-axis.

- **TEC** *(tracker endcaps)* and **TID** *(tracker inner discs)*: the endcap detector complements the barrel at both ends with detector planes perpendicular to the beam. Each endcap consists of 10 disks, whose position along the z-axis have been adjusted to obtain full coverage and to optimize the track reconstruction performance over the pseudo-rapidity region $1.2 \leq |\eta| \leq 2.5$. Beyond $|\eta| \approx 2.5$ the radiation level and the track density becomes too high to operate silicon microstrip detectors reliably.

This layout leads to a number of detection channels as large as 5.4×10^6, equally distributed between barrel and endcaps, and a total active surface close to 75 m^2 covering the region with $\eta < 2.4$.
The size of the CMS silicon tracker is nearly two orders of magnitude larger than previous silicon microstrip vertex detectors used at LEP; in addition to this, it must operate in the hostile environment of LHC and use very fast electronics.

Track reconstruction

The main algorithm used in the CMS experiment in order to reconstruct charged tracks is the Kalman Filter, which is a recursive formulation of the least-squares method of fitting a set of measurements to a track model [13] [14].

The combinatorial Kalman Filter (CKF) is a recursive procedure which starts from an initial trajectory estimate, and integrates pattern recognition and track fitting.
The track hits that are given by the Tracker are, at best, in number of:

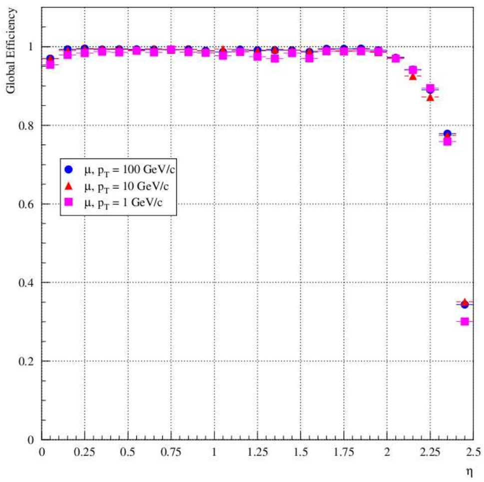
- 2 points in space from the pixel tracker, with 3 coordinates each and a spatial resolution of around 10 μm
- 14 points in space from the micro-strip detector, with 2 coordinates each and a spatial resolution of around 30 μm

In the default strategy for track reconstruction in CMS, the information given by the Tracker are used as a starting point to screen all the possible combinations between the collected hits, picking out the ones that constitute with higher probability a track.

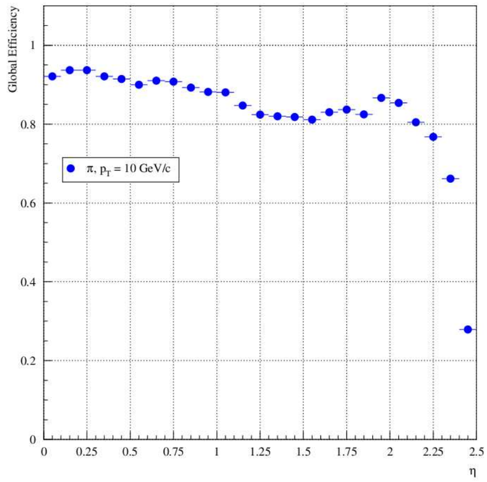
The initial estimate of the trajectory (seed) consists of a track segment made of a set of compatible measurements in the pixel layers close to the beam pipe.
The Kalman Filter proceeds by an alternating sequence of propagation and update steps. In the propagation step, the trajectory at the current surface is extrapolated to the next tracker layer. The covariance matrix is propagated as well taking multiple scattering and energy loss into account. In the update step, the extrapolated trajectory is combined with the observation.
For each seed, a tree of possible track candidates is generated. The number of trajectories is limited according to their chi-squared, χ^2, and the number of missing hits. Since only the estimate in the last layer is based on the full track candidate information, the surviving track candidates are submitted to a Kalman smoothing procedure to obtain the optimal estimates at every measurement point.

The efficiency to reconstruct tracks with the combinatorial Kalman Filter is shown in Figure 1.13a and 1.13b as a function of the pseudorapidity for single muons and pions respectively. Single muons are reconstructed with an efficiency close to 100% up to $\eta = 2$, while single pions are reconstructed with an efficiency of about 85% due to their nuclear interaction with the Tracker material (about 20% of pions with 1 GeV/c do not reach the outer layer).

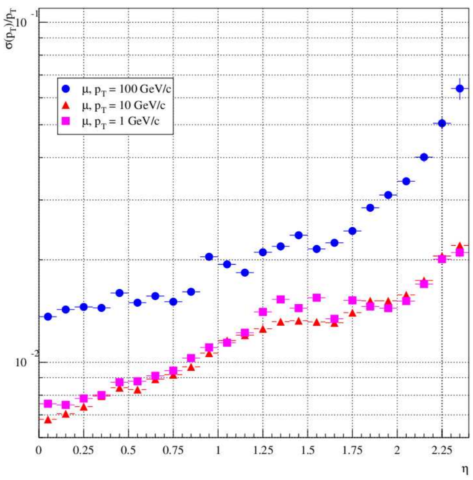
Transverse momentum resolution as a function of $|\eta|$ is shown in Figure 1.13c, for single muons with different transverse momentum p_T values. The resolution of the transverse momentum is dominated by multiple scattering for p_T values below 10 GeV/c, by the detector spatial resolution for p_T values above 100 GeV/c, and it degrades for $|\eta| = 1.7$ as long as the track exits from the Tracker detector.



(a) Track reconstruction efficiency for single muons with different p_T as a function of the pseudorapidity $|\eta|$



(b) Track reconstruction efficiency for single pions with $p_T = 100$ GeV/c as a function of the pseudorapidity $|\eta|$



(c) Relative transverse momentum resolution for single muons with different p_T values as a function of the pseudorapidity $|\eta|$

Figure 1.13: Track reconstruction efficiency and momentum resolution.

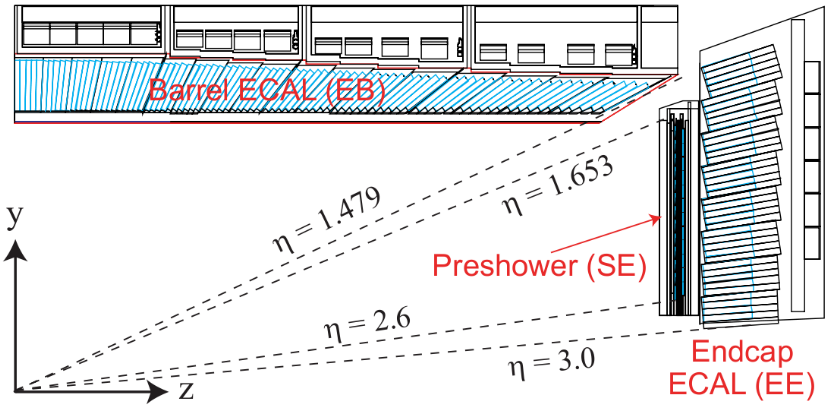


Figure 1.14: Longitudinal view of a quadrant of CMS electromagnetic calorimeter.

1.2.3 Electromagnetic calorimeter

A high performance electromagnetic calorimeter is a fundamental requirement for any general purpose LHC experiment for precise measurements of electrons and photons. The design of the CMS electromagnetic calorimeter (ECAL) has been prompted by the possibility to observe the decay of a light Higgs boson $m_H <$ 140 GeV/c^2 into a pair of photons. Since in this mass range the intrinsic Higgs width Γ_H is less than 100 MeV, the $\gamma\gamma$ invariant mass resolution is dominated by experimental resolution, which should be of the order of 1% to enhance the significance of a possible signal.

The CMS collaboration has created a hermetic, homogeneous calorimeter composed with finely segmented crystals of lead tungstate ($PbWO_4$), which is a radiation resistant and chemically inert scintillator suited to work in the LHC high dose environment. Moreover, the lead tungstate has also a short scintillation decay time $\tau = 10$ ns that allows to collect approximately 85% of the light in the 25 ns interval between two pp collisions. The small Molière radius of 21.9 mm and radiation length $X_0 = 8.9$ mm allow the shower containment in a limited space resulting in a compact calorimeter design. Figure 1.14 shows a longitudinal view of a quarter of the electromagnetic calorimeter: it is organised in a barrel region $|\eta| < 1.48$ (EB) and a forward region to cover the pseudorapidity area up to $|\eta| \leq 3.0$ (EE).

The barrel granularity is 360-fold in ϕ and (2×85)-fold in η, and results in a total of 61200 crystals. The ECAL detector has two endcaps; each endcap is divided into 2 halves, or *Dees*. Each Dee contains 3662 crystals.

Lead tungstate crystals

In addition to the already mentioned features, the crystals have a trapezoidal shape with squared front faces, and are slightly different in the two ECAL regions. In the barrel they are 230 mm long with a total radiation length $X = 25.8\ X_0$ and a 22×22 mm^2 front section, as wide as the Molière radius; the granularity is $\Delta\eta\times\Delta\phi = 0.0175\times0.0175$, high enough to achieve an efficient $\pi^0-\gamma$ separation. In the endcaps, the crystals have 28.8×26.8 mm^2 square front sections and smaller length (220 mm) and hence a smaller radiation length $X/X_0 = 24.7$, and a larger granularity at higher

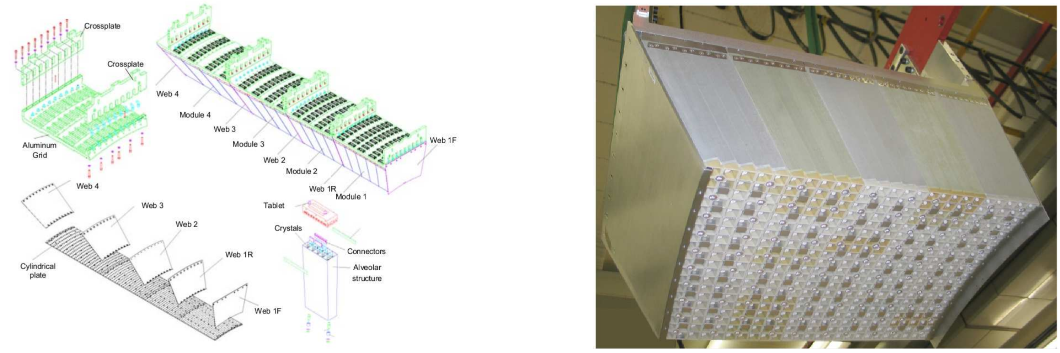


Figure 1.15: Layout of the ECAL barrel mechanics and the front view of a module equipped with the crystals, respectively.

$|\eta|$, with a maximum value of $\Delta\eta \times \Delta\phi = 0.05 \times 0.05$ in the very forward crystals, [15].

The $PbWO_4$ crystals have good transparency ($\sim$ 70 %) to light in the main emission band of scintillation light (420-430 nm). Crystals transparency is critically dependent on temperature; for this reason ECAL is equipped with a temperature monitoring and a cooling system to dissipate the heat produced by the electronics and steadily keep the detector at $18 \pm 0.1°$ C. Since the measured energy is proportional to the total scintillation light produced by the showering particle, transparency stability is critical in order to guarantee high precision in measuring particle's energy. Moreover, the transparency of the crystal is influenced negatively by the amount of radiation dose absorbed. To measure reliably the energy during data taking, the transparency is monitored through a laser calibration system.

Calorimeter structure

Barrel

The crystals are contained in a thin-walled alveolar structure (*submodule*). The alveolar wall is 0.1 mm thick and is made of an aluminium layer, facing the crystal, and two layers of glass fibre-epoxy resin. To avoid oxidation, a special coating is applied to the aluminium surface. The nominal crystal-to-crystal distance is 0.35 mm inside a submodule, and 0.5 mm between submodules. Each submodule contains 5×2 crystals.

The submodules are assembled into *modules* of different types, according to the position in η, each containing 400 or 500 crystals. Four modules, separated by aluminium conical webs 4-mm thick, are assembled in a *supermodule*, which contains 1700 crystals, see Figure 1.15. In each module, the submodules are held in partial cantilever by an aluminium grid, which supports their weight from the rear. The entire CMS ECAL Barrel (EB) is made of 36 supermodules, see Figure 1.16 left panel.

The crystal axes are inclined at an angle of 3 degrees relative to the direction of the nominal interaction point, in both the azimuthal (ϕ) and η projections, to avoid the aluminium supports to be aligned with the particle trajectory.

Endcap

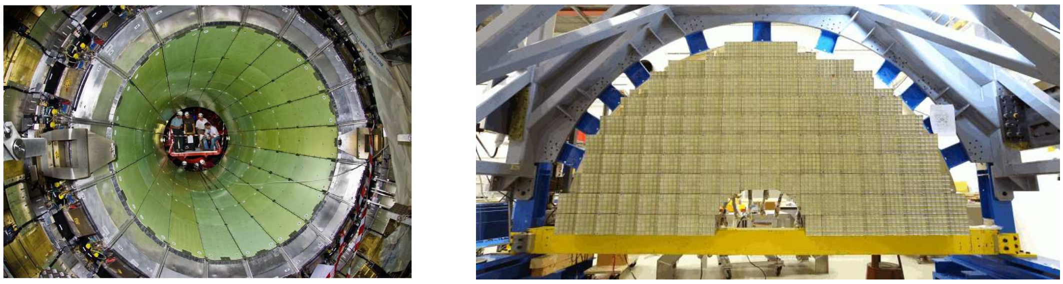


Figure 1.16: On the left is shown the ECAL barrel with equipped with all 36 supermodules. On the right is represented an assembled dee, with the super-crystals containing 5×5 crystals each.

The two ECAL endcaps (EE) are constructed from four half-disk (dees), each consisting of 3662 tapered crystals, arranged in a quasi-projective geometry. The crystals are focussed at a point 1.3 m farther than the nominal interaction point along the beam line, with off-pointing angles between 2 and 8 degrees. The crystals in each dee are organised into 138 standard 5×5 supercrystal units, and 18 specially shaped supercrystals that are located at the inner and outer radii.

Figure 1.16 showns how supermodules are mounted in the barrel (left panel) and a dee of an endcap (right panel).

Preshower

A preshower detector is placed in front of the endcap disks. ECAL Preshower is a sampling calorimeter made up of alternated layers of: two layers of silicon strip detector and two disks of lead absorber. This particular detector is dedicated to identify neutral pions, improve electron discrimination against minimum ionizing particles and improve the determination of the position of electrons and photons. The detector acceptance is between $1.653 < |\eta| < 2.6$, and its length is around two radiation lengths.

Readout electronics

The readout electronics of the electromagnetic calorimeter meets the following requirements: fast time response, high radiation hardness, be not much affected by the 4 T magnetic field generated by the superconducting solenoid. With these indications, the Avalanche Photo Diodes (APDs) have been chosen for the barrel, and the Vacuum Photo Triodes (VPTs) for the endcaps.

Scintillation light from every single crystal in EB is detected by two Hamamatsu S8148 APDs (for redundancy), with an active area of 5×5 mm^2 each, much smaller than the crystal surface; the APDs are characterized by an excellent signal-to-noise ratio (SNR) and a very high gain (in ECAL are operated at gain 50, giving approximately 4.5 photoelectrons per MeV at 18° C). These are connected in parallel to the on-detector readout electronics, which are organised in units of 5×5 crystals, each unit corresponding to a trigger tower.

Scintillation light in the EE is detected by VPTs (type PMT188) produced by NRIE, which have generally lower performances (e.g., gain $8 - 10$), but are suited to operate in very high particle flux conditions, such that close to the beam pipe at high $|\eta|$. The lower VPT performances are partially counterbalanced by a bigger active area, 280 mm^2.

The signals produced by these photomultipliers are then amplified and digitized by custom cards, called Very Front End (VFE) cards, each accepting data from 5 APD pairs. The APD signals are pre-amplified and shaped by Multiple Gain Pre-Amplifier (MGPA) ASICs located on the VFE boards, which consist of three parallel amplification stages (gains 1, 6 and 12). The output is digitised by a 12-bit ADC running at 40 MHz, which samples the pulse ten times for each channel and selects the gain with the highest non-saturated signal. Thus it is possible to measure the pedestal value for the specific photomultiplier and the shape of the energy deposit in time, limiting the noise from electronics. The data from five VFE cards are transferred to a single front-end card, which generates the trigger primitive data, and transmits it to the dedicated off-detector trigger electronics.

Energy resolution

The electromagnetic energy resolution of ECAL can be parameterised as a function of the incident electron/photon energy, E (in GeV), as

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} + \frac{b}{E} + c \qquad (1.2.1)$$

where a represents the stochastic term, which depends on event to event fluctuations in lateral shower containment, photo-statistics and photodetector gain; b represents the noise term, which depends on the level of electronic noise and event pile-up; and c represents the constant term, which depends on non-uniformity of the longitudinal light collection, leakage of energy from the rear face of the crystal and the accuracy of the detector inter-calibration constants. The target value for the constant term, which dominates the resolution at high energies, is 0.5% for the barrel and 1.0% for the endcaps [16].

Measurements taken with test beam electrons with energies between 20 and 250 GeV have shown that the EM energy resolution and noise performance of the ECAL barrel meets the design goals for the detector. For the barrel, the mean values of the stochastic and constant terms, computed using the energy summed over 3×3 crystal arrays, are 2.8%/E(GeV) and 0.3% respectively. The mean single channel noise, computed for 1175 crystals, is 41.5 MeV energy equivalent [17].

1.2.4 Hadronic calorimeter

The hadron calorimeter is used together with the electromagnetic one to measure the energy and direction of jets, their transverse energy E_T and the imbalance of missing transverse energy E_T^{miss}. To fulfil these requirements, it has to be thick enough to contain the whole hadron shower and have high hermeticity. Since it is placed inside the magnet, see Figure 1.17, it cannot be made with ferromagnetic materials.

The CMS hadronic calorimeter (HCAL) has a structure very similar to the ECAL one, because it is made of a cylindrical central section called barrel (HB), and two endcaps (HE) constituted by two circular discs. This part of the calorimeter covers a range up to $|\eta| \leq 3$. Both HB and HE are placed inside the CMS magnet.

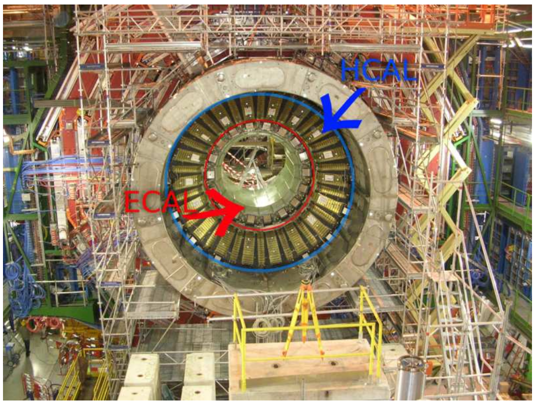


Figure 1.17: CMS electromagnetic calorimeter (ECAL) and hadronic calorimeter (HCAL).

In addition to HB and HE, there are two more calorimeter sections: the Outer Hadronic Calorimeter (HO) placed just outside of the CMS magnet, that ensures that enough sampling depth is provided in the barrel region, and the Forward Hadronic Calorimeter (HF) placed at 11.2 m from the interaction point, very close to the beam line, that extends the covered pseudorapidity range up to $|\eta| < 5.2$.

Since the HCAL is a sampling calorimeter, it is made of alternate layers of absorber (to spark the hadronic shower) and scintillator (to detect and measure the signal of the traversing particles). For example, the HB is made of 3.7 mm thick active layers of plastic scintillators alternated with 5 cm thick brass plate absorbers. The number of layers is variable along the detector, but is sufficient to assure a depth from 5 to 10 radiation lengths. The signal is read out with wavelength-shift fibres. The granularity $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ is fine enough to allow an efficient di-jet separation. Moreover, HB granularity matches ECAL trigger towers (made of 5×5 crystals), so it is possible to combine the information coming from the two calorimeters into a so called *calotower* (calorimetric tower).
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 HE has a granularity of 0.087×0.087 for $|\eta| < 1.6$ and 0.17×0.17 for $|\eta| > 1.6$.

The scintillation light produced in each scintillator layer is collected by a wavelength-shifting (WLS) fiber and guided to the readout electronics, made out of multichannel hybrid photo diodes (HPD). The amplification and digitization system is completely different from the ECAL one; the HCAL readout does not sample the signal in time, but integrates the signal over different measurements in a window between two consecutive collisions of 25 ns.
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.
Due to the harsh radiation environment in the forward region (on the 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 5 mm steel absorber. The signal produced in quarz fibres is Cherenkov light, transferred by optical fibers to some specific photomultipliers.

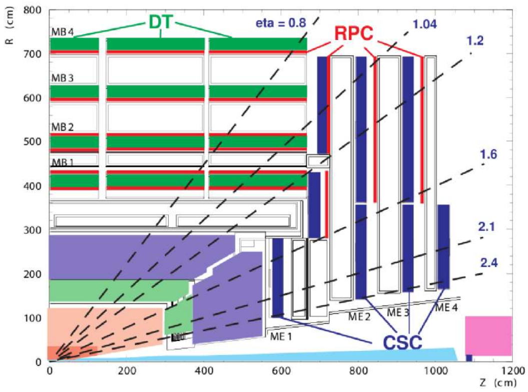


Figure 1.18: Longitudinal view of a quarter of the muon system, subdivided into barrel with drift tubes (DT) and resistive plate chambers (RPC) and endcap with cathode strip chambers (CSC) and RPCs.

1.2.5 Muon detector system

A first measure of the muon momentum is carried out by the tracker, in the region closest to the interaction point. This measure has poor resolution for muons with transverse momentum higher than 200 GeV/c, beacuse the trajectories are almost rectilinear and then the tracker is not able to extract the curvature/radius of the tracks with the required precision. For this reason a specific detector has been developed to track muons, making use of the 1.8 T magnetic field escaping from the solenoid.

The muon system is placed outside the magnet, embedded in the iron return yoke to make full use of the magnetic return flux. It plays an essential role in the CMS trigger system, because high transverse momentum muons are clear signatures of many physics processes. The main goal of this system is to identify muons and measure, when combined with the tracker, their transverse momentum.

Looking at Figure 1.18, it is possible to notice that the muon system is organised into three independent subsystems: in the barrel where drift tubes (DT) detectors are installed (because the track occupancy is relatively low < 10 Hz/cm^2), while in the endcaps cathode strip chambers (CSC) are favoured to work with higher particle rates (> 100 Hz/cm^2) and a larger residual magnetic field within the yoke plates. These two subsystems cover the $|\eta| < 2.4$ region and are arranged in a multi-layer structure to efficiently reject single hits produced by low range particles.
In the region $|\eta| < 2.1$ redundancy is provided by resistive plate chambers (RPC), which have a limited spatial resolution, but a faster response and an excellent time resolution, less than 3 ns. They are used mainly to unambiguously identify the bunch crossing and to complement the DT+CSC measurement of the muon p_T for triggering, because RPCs can be finely segmented since they do not demand a costly readout system.

- **Drift Tubes** are composed with parallel aluminium plates insulated from perpendicular "I" shaped aluminium cathodes by a polycarbonate plastic profile. The anodes are 50 µm diameter stainless steel wires placed between the "I" cathodes. The internal volume is filled with a binary mixture of 80% Ar and 20% CO_2 at atmospheric pressure, because this gas is non-flammable and can

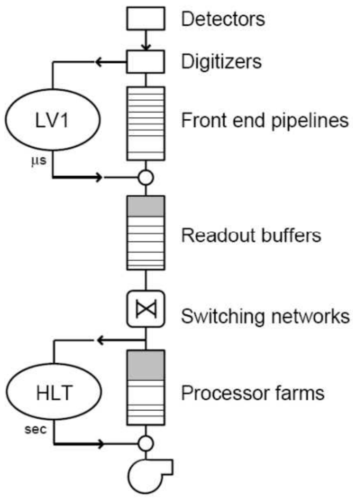


Figure 1.19: The CMS Trigger/DAQ data flow, with a single filtering step (HLT) after the Level-1 trigger.

be safely used in underground operations in large volumes, as required in CMS. The resolution is about 100 µm both in $r\phi$ and rz views.

- **Cathode Strip Chambers** are composed with arrays of anode wires between pair of cathode planes, segmented into strips perpendicular to the wires. Gaps are filled with a gas mixture of 30% Ar, 50% CO_2 and 20% CF_4. The interpolation of the signal of neighbouring strips allows a precise spatial measurement of the ϕ coordinate with 50 µm resolution.

- **Resistive Plate Chambers** are made of planes of a phenolic resin (bakelite) with a bulk resistivity of $10^{10} - 10^{11}$ Ωcm, separated from aluminium strips by an insulating film. The gaps are filled with a non-flammable gas mixture of 94.5% freon ($C_2H_2F_4$) and 4.5% isobutane (C_4H_{10}), which operates in avalanche mode to sustain the high rates.

1.2.6 Trigger and Data Acquisition System

CMS has chosen a two-level trigger system, consisting of a Level-1 Trigger (L1) and a High Level Trigger (HLT).
The Level-1 Trigger runs on dedicated processors, and accesses rough level granularity information from calorimetry and muon system. A Level-1 Trigger decision has to be taken for each bunch crossing within 3.2 µs. The Level-1 Trigger task is to reduce the data flow from 40 MHz to 100 kHz.
The High Level Trigger is responsible for reducing the L1 output rate down to the target of 100 Hz. The HLT code runs on a farm of commercial processors and can access the full granularity information of all the subdetectors [18] [19].

Level-1 trigger

The Level-1 trigger is responsible for the identification of electrons, muons, photons, jets and missing transverse energy. Its selection is based exclusively on calorimeter

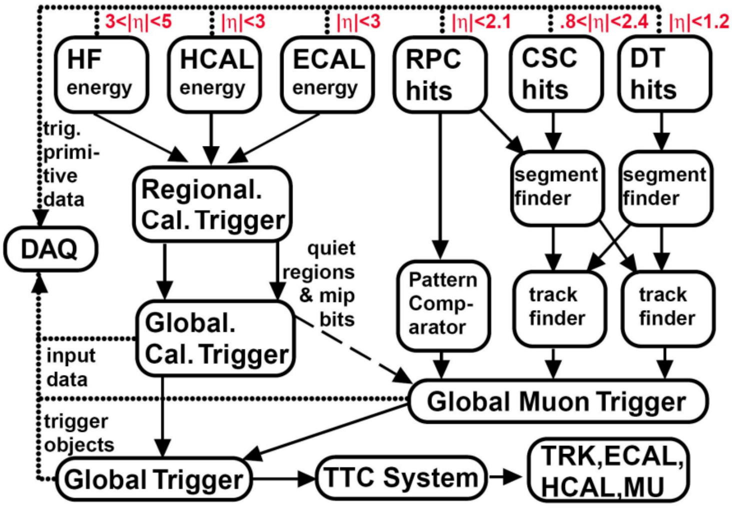


Figure 1.20: Level-1 trigger components.

and muon chamber information, processed with coarse granularity by fast hardware logical circuits. The two elements of the Level-1 trigger, calorimetry and muon system, work in parallel and analyse the data locally, combine the information and produce the output passed to the Data Acquisition System, see Figure 1.20 for the operating scheme.

The time interval taken for the L1 decision over accepting or rejecting the event, lasts approximately 3.2 µs. During this time all the recorded information is put in a pipeline waiting to be saved or deleted after the L1 answer. The type of decisions taken by the L1 is simple and based on few logic operations, this beacause of speed requirements. For example, it searches for some objects like electrons, photons, muons or missing transverse energy, requiring a minimum energy threshold. The reconstruction of these objects is made by using a partial information from the subdetectors; for example, the ECAL information is not used at full granularity, but only at trigger tower's level.

High Level Trigger

The High Level Trigger code runs on commercial processors and performs reconstruction by using the information from all subdetectors. Data read from subdetectors are assembled by a builder unit and then assigned to a switching network that dispatches events to the processor farm. The CMS switching network has a bandwidth of 1 Tbit/s.

This simple design ensures maximum flexibility, the only limitation is the total bandwidth and the number of processors. The system can easily be upgraded adding new processors or replacing the existing ones with faster ones as they become available. Since the algorithm implementation is fully software, improvements in the algorithms can easily be implemented and do not require any hardware intervention.

Event by event, the HLT code is run on a single processor, and the time available to take a decision is about 300 ms. In order to efficiently process events the HLT code has to be able to reject uninteresting events as soon as possible: computationally expensive algorithms must be run only on good candidates for interesting events. To meet this requirement the HLT code is organized in a virtually layered structure, where each step reduces the number of events to be processed in the following step:

- Level 2: uses only muon and calorimetry information;
- Level 2.5: uses also the pixel information;
- Level 3: makes use of the full information from all the tracking detectors.

The rate of events is then reduced by a 10^3 factor, from the Level-1 output rate of 100 kHz to the goal of 100 events/s that are going to be written to mass storage, saturating the CMS bandwidth capacity.

Chapter 2

Theoretical background for Z+jets production

In the first part of this chapter the Standard Model (SM) of electroweak and strong interactions based on the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ local gauge symmetry, together with its fundamental building blocks is briefly introduced. Then the associated production of Z and W bosons with hadronic jets in proton-proton collisions will be discussed in more detail. The description of this process is, at tree level, connected with the electroweak coupling constant providing a classic test for the SM physics and for the parton model for "hard" hadrons' interactions. The larger contributions to the Z and W production cross-section at the Next-to-Leading-Order (NLO) level come from (virtual or real) gluon radiation and gluon scattering processes. They have a complex signature consisting of one or more jets in association with the decay products of W and Z in the final state.

In the second part some aspects of the experimental searches for the Higgs mechanism are briefly described, since some decay channels of the predicted particle mimic the final states of W and Z +jets.
For this reason the production of vector bosons in association with hadronic jets is of crucial importance being the main background for the discovery of physics beyond the SM.

2.1 Elementary particles

In the Standard Model the elementary spin-$\frac{1}{2}$ fermions are the leptons, the quarks and their anti-particles. They are classified in three families (generations):

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}$$

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}$$

Matter consists of leptons and hadrons, the latter are divided into mesons (made up with one quark and one anti-quark) and baryons (made up with combinations of three quarks). All the elements belonging to the three families have been directly or indirectly observed (and no experimental evidence of the existence of a fourth

generation so far).
Elementary particles interact with each other via the four fundamental interactions of Nature, which are mediated by bosons. The **four forces** and the respective *force carriers* are:

- **gravitational** *graviton (G)*
- **strong** *8 gluons (g)*
- **weak** W^+, W^-, Z^0
- **electromagnetic** *photon* (γ)

The weak and electromagnetic forces are two aspects originated from the same source, the electroweak interaction.

2.2 Electroweak interactions

As already mentioned, the Standard Model of electroweak interactions is a quantum field theory whose Lagrangian is invariant under local $SU(2)_L \otimes U(1)_Y$ gauge transformations. The symmetry that arises from this invariance requirement generates interactions mediated by four vector gauge fields, that can be identified with the photon γ, two massive charged vector bosons W^+ and W^- and a massive neutral vector boson Z^0 [20]. Spontaneous breaking of this gauge symmetry generates the masses of W and Z (the γ remaining massless) through the so called "Higgs mechanism". The reminiscence of this mechanism is the predicted existende of the neutral, spin 0, Higgs particle currently searched for.

Similar to the case of gauge bosons, this symmetry makes it impossible to include in the lagrangian mass terms for the fermion fields, as a consequence the helicity is a good quantum number. In this way, Dirac spinors are introduced with well defined helicity, right-handed (R) or left-handed (L). For the lepton families, L_L and L_R states are defined as doublets and singlets of the symmetry:

$$L_L = \begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix}_L \qquad L_R = (\ell)_R$$

with $\ell = (e, \mu, \tau)$. The right-handed neutrino $(\nu_l)_R$ singlet is not included because they are considered as massless and have negative helicity from experiment, hence are pure left-handed. For the quark families the definition of the helicity states is the same, the only difference is in the right-handed singlets, that are two, one for each quark of the generation [21]. Fermion masses are generated by the same Higgs mechanism generating $M_{W,Z}$, while keeping neutrinos masses zero.

2.3 Strong interaction among quarks

The Lagrangian which describes strong interactions is invariant under local $SU(3)_C$ (colour) gauge transformations; this implies the introduction of eight gauge fields, the gluons (massless spin-1 vector bosons), that carry the strong interaction. The

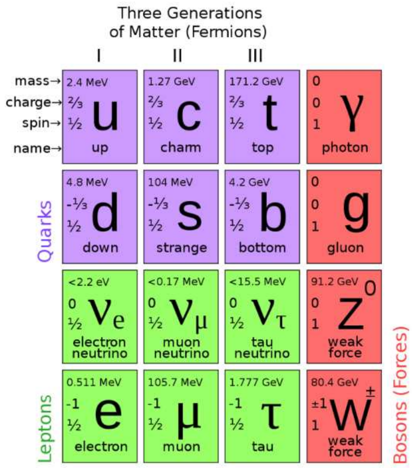


Figure 2.1: The SM predicted particles, experimentally detected.

charge associated with the strong interaction is called colour charge, and the associated interaction is called Quantum Chromodynamic theory (QCD). Unlike the electromagnetic interaction where the photon has no electric charge, the gluons do have colour charge and then they can interact both with each other and with quarks. Only quarks have colour charge, while leptons do not carry any colour charge and therefore they do not interact via the strong force.

The value of QCD coupling constant changes when moving from hard, i.e. large momentum transfer processes, to soft processes. Its value decreases logarithmically as the momentum transfer increases. This effect called *asymptotic freedom*, and justifies the use of perturbation theory when describing "hard processes" among quark and gluons (short-distance hadronic interactions). At small energy and momentum transfer (large distances), where the value of the coupling constant becomes much larger than unity, the theory behaves in a non-perturbative way. In such a regime quarks or gluons are subject to bound state dynamics. This effect is known as *confinement* and is consistent with the non-observation of free quarks and gluons [21]. Confinement effects of quarks and gluons are experimentally measurable in hadronic interactions at a particle collider: every quark and gluon produced in the final state of a collision is subject to a non-perturbative hadronization process resulting into a multitude of hadrons (see Section 2.3.1).

The table shown in Figure 2.1 sums up all the experimentally detected elementary particles.

2.3.1 Hadronization process

In a collision an event like the one represented in Figure 2.6 might be generated. In this collision the outcoming partons cannot survive as free particles because of the postulated colour confinement. The partons produced in the collision then are subjected to a non-perturbative process which forms hadrons (colour singlets). This phenomenon of hadronization gives rise to the measured objects: hadrons in the final state group in hadronic jets. From the hadrons measurement one then tries to reconstruct the "elementary" quark and gluon process governing it. In a jet, hadrons

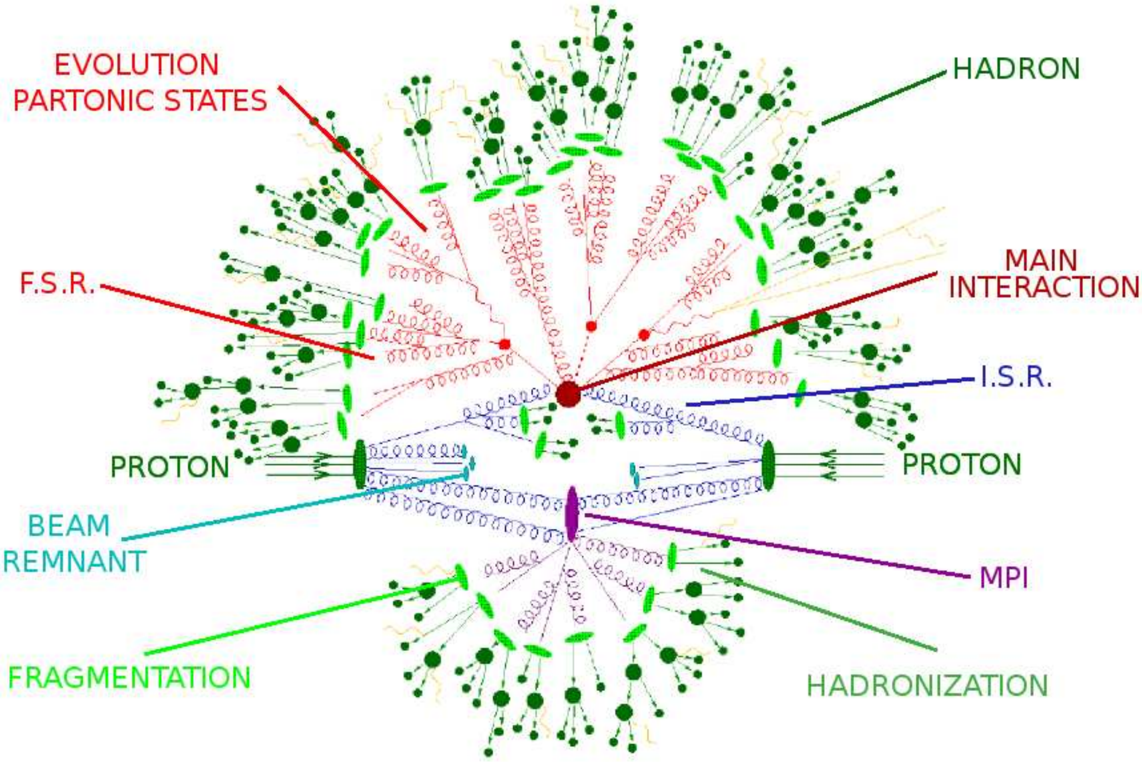


Figure 2.2: A typical proton-proton interaction is depicted. Many different processes occur together with the main interaction. For example ISR, FSR or other parton interactions (called Multi Parton Interactions MPI). The partons that interact are subjected to fragmentation and hadronization. The partons that not interact propagate almost unperturbed in some bound state, establishing colour singlets (Beam Remnants). All the processes that are not involved in the main interaction are collected into the Underlying Event notion. During a single bunch-crossing many pp collisions may occur (pile-up).

are collimated in a cone, impinging in a narrow area in the η and ϕ coordinates of the detector. The sum over all the energies of those hadrons gives the energy of the original parton and the cone axis indicates the parton direction.
The energy scale at which the hadronization starts is of the order of a mass scale Λ_{QCD} which is found around 200 MeV; this gives a lower limit to the applicability of QCD perturbation theory. The QCD dynamics of the hadronization process are not yet fully understood (as being non-perturbative) and amendable to a perturbative expansion in terms of the coupling constant α_S; for this reason hadronization is modeled and parameterized in a number of phenomenological QCD-inspired approaches (e.g., the Lund string model) to match with experimental outcomes. Thus, some properties of the original parton can be inferred from the characteristics of the jets, but these may unavoidably depend to some extent from the hadronization model adopted in the data analysis.

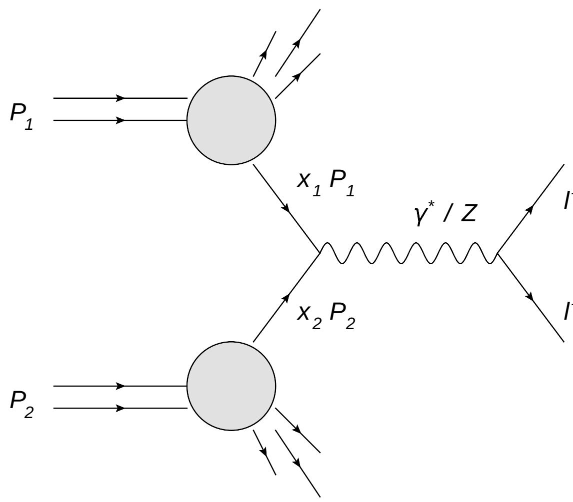


Figure 2.3: The Drell-Yan process $pp \to l^+l^- + \mathcal{X}$.

2.3.2 Underlying event and pile-up

In a proton-proton collision many subprocesses occur (and overlap) together with the main interaction, see Figure 2.2. Their effect on the event final state is to modify the hadronic jets multiplicity and momentum distributions. Two manifestations of these subprocesses are Multi Parton Interactions (MPIs) and Beam Remnants (BR). The MPIs are events where more than one (the hardest) interaction occur; the BR are the remaining pieces of the proton after the collision, forming colour singlets and propagating almost collinear to the beam axis. These processes together make up the so called Underlying Event (UE). The UE generally adds to the event final state low-p_T jets.
It is possible to remove the contributions of the soft jets for example by applying some kinematic cuts during the jets reconstruction phase (see Section 3.2).

In addition to this, there is another type of contamination to the main parton interaction: since in a single bunch there are $\sim 10^{11}$ protons, the probability of having more than one proton-proton interaction per bunch crossing is real. The average number of interactions per bunch-crossing in the 2011 data taking period was around $\langle n \rangle \sim 6$. These collisions, that add to the main one, are called pile-up (PU) interactions. The charged contribution of such interactions is detectable, for example, in the tracker, hence it is possible to isolate each vertex and remove the contributions in the limit of the vertex reconstruction resolution; the difficult task is to remove this contribution from neutral particles that leave their energy in ECAL (photons) and HCAL (neutral hadrons). In Section 3.2.4 an algorithm to remove pile-up effects during jet reconstruction is described .

2.4 W+jets and Z+jets associated production

In this section the approach to calculate the cross-section for the production of a vector boson in association with hadronic jets is described.

Drell-Yan process

It is possible to calculate this inclusive cross-section at Leading Order in QCD, starting from the Drell-Yann process (see Figure 2.3), where a lepton pair l^+l^- with large invariant mass-squared, $M^2 = (p_{l^+} + p_{l^-})^2 \gg 1\ \text{GeV}^2$, is produced by quark-antiquark annihilation:

$$pp \to q\bar{q} \to l^+l^- + \mathcal{X}$$

where $\mathcal{X}$ is any generic hadronic final state consistent with energy and momentum conservation.

In the parton model, the inclusive cross section $\sigma_{pp\to l^+l^-+\mathcal{X}}$ for producing l^+l^- in a proton-proton collision is obtained by simply weighting the subprocess cross-section $\hat{\sigma}_{q\bar{q}\to l^+l^-}$ for $q\bar{q} \to l^+l^-$ with the Parton Distribution Functions (PDF) $f_q(x)$ and $f_{\bar{q}}(x)$ (where x is the parton momentum fraction) and summing over all quark-antiquark combinations in the protons. Symbolically, one can write:

$$\sigma^{DY} = \sum_q \int dx_1 dx_2 f_q(x_1) f_{\bar{q}}(x_2) \hat{\sigma}_{q\bar{q}\to l^+l^-} \tag{2.4.1}$$

$f_q(x)$ and $f_{\bar{q}}(x)$ are extracted from deep inelatic electron scattering experiments. The formal domain of this equation is the asymptotic “scaling” limit where M^2,$s \to \infty$ and $\tau = M^2/s$ is fixed.

The lowest-order total cross section for quark-antiquark annihilation into a lepton pair via an intermediate off-mass-shell photon γ^* is given by:

$$\hat{\sigma}\left(q(p_1)\bar{q}(p_2) \to l^+l^-\right) = \frac{4\pi\alpha^2}{3\hat{s}} \frac{1}{N} Q_q^2 \tag{2.4.2}$$

where $\hat{s} = (p_1 + p_2)^2 = x_1 x_2 s$, p_1 and p_2 the parton's four-momentum, N the number of colours, Q_q quark fractional charge and α is the electromagnetic coupling constant.

In general the incoming quark and antiquark have a spectrum of collision energies $\sqrt{\hat{s}}$, so it is more appropriate to consider the differential lepton pair mass distribution. Thus the subprocess cross section for producing a lepton pair of mass M is

$$\frac{d\hat{\sigma}}{dM^2} = \frac{4\pi\alpha^2}{3M^2N} Q_q^2 \delta(\hat{s} - M^2) \tag{2.4.3}$$

Substituting eq. 2.4.3 into eq. 2.4.1 gives the parton model cross section for this process (at LO):

$$\begin{aligned}
\frac{d\sigma}{dM^2}^{DY} &= \int_0^1 dx_1 dx_2 \sum_q \{f_q(x_1) f_{\bar{q}}(x_2) + (q \leftrightarrow \bar{q})\} \times \frac{d\hat{\sigma}}{dM^2}(q\bar{q} \to l^+l^-) \\
&= \frac{4\pi\alpha^2}{3M^2N} \int_0^1 dx_1 dx_2 \delta(x_1 x_2 s - M^2) \\
&\quad \times \left[\sum_q Q_q^2 \{f_q(x_1) f_{\bar{q}}(x_2) + (q \leftrightarrow \bar{q})\}\right] \quad .
\end{aligned} \tag{2.4.4}$$

This cross section is calculated at leading order (LO), and hence it is possible to calculate perturbative QCD corrections. The calculations are not shown here, but in Figure 2.4 the diagrams wich contribute at $\mathcal{O}(\alpha_s)$. There are three classes of contributions: (a) virtual gluon corrections to the LO contribution, (b) real gluon corrections, (c) quark-gluon scattering process together with the corresponding $\bar{q}g$ contribution.

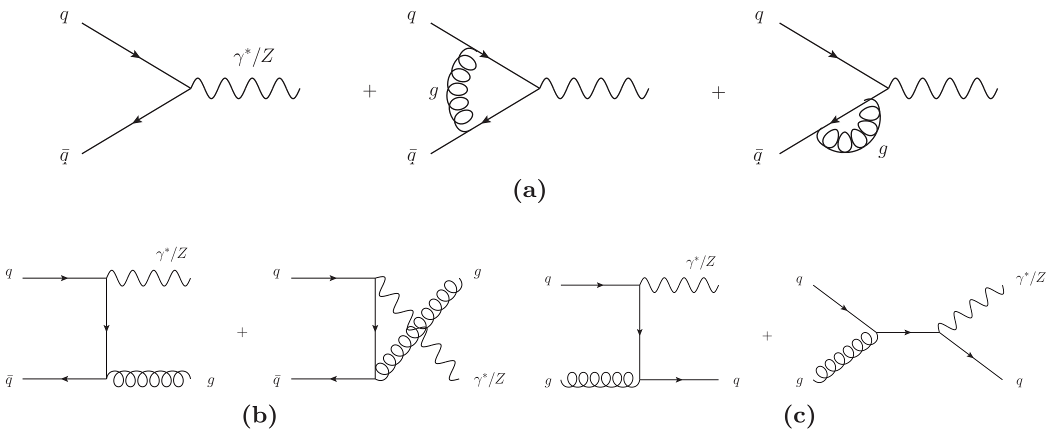


Figure 2.4: The leading- and next-to-leading-order diagrams for the Drell-Yan process.

At lowest order in perturbation theory represented in Figure 2.3, and assuming $M \ll M_Z$, the subprocess cross-section is given by the annihilation process $q\bar{q} \to \gamma^* \to l^+l^-$.
The effect of such corrections is that the PDF acquire a (calculable) logarithmic M-dependence, and $\mathcal{O}(\alpha_s)$ "K-factor" appears, so that eq. 2.4.1 becomes

$$\sigma^{DY} = K \sum_q \int dx_1 dx_2 f_q(x_1, M^2) f_{\bar{q}}(x_2, M^2) \hat{\sigma}_{q\bar{q}\to l^+l^-} \quad . \tag{2.4.5}$$

The M-dependent PDF extracted from analysis of inclusive inelastic electron scattering are available, so the theory is completely predictive. Also, calculations at the higher next-to-next to leading order (NNLO) are available.

To calculate the cross-sections for Z and W boson production and their identification through their leptonic final state decays, it is sufficient to substitute the $q\bar{q} \to \gamma^* \to l^+l^-$ subprocess cross-section (eq. 2.4.2) with the cross-sections $q\bar{q} \to Z \to l^+l^-$ and $q\bar{q} \to W \to l\nu$:

$$\begin{aligned} \sigma_{q\bar{q}\to Z\to l^+l^-} &= \sigma_{q\bar{q}\to Z} \cdot BR\left(Z \to l^+l^-\right) \\ \sigma_{q\bar{q}\to W\to l\nu} &= \sigma_{q\bar{q}\to W} \cdot BR\left(W \to l\nu\right) \end{aligned}$$

where $\sigma_{q\bar{q}\to Z}$ and $\sigma_{q\bar{q}\to W}$ are the production cross-sections for Z and W boson respectively, and $BR\left(Z \to l^+l^-\right)$ $BR\left(W \to l\nu\right)$ their branching-ratios to leptons. Since the decay rate of Z and W bosons are small (Γ_Z = 2.50 GeV and Γ_W = 2.08 GeV, respectively) compared to their masses, it is sufficient to consider the production of effectively stable particles [21]. The production cross-sections can therefore be approximated as Z and W were on-shell:

$$
\begin{aligned}
\sigma_{q\bar{q}\to Z} &= \frac{\pi}{3}\sqrt{2}G_F M_Z^2\left(V_q^2 + A_q^2\right)\delta\left(\hat{s} - M_Z^2\right) \\
\sigma_{q\bar{q}\to W} &= \frac{\pi}{3}\sqrt{2}G_F M_W^2 |V_{q\bar{q}}|^2 \delta\left(\hat{s} - M_W^2\right)
\end{aligned}
$$

where V_q^2 and A_q^2 are associated with the vector and axial vector coupling constants of the neutral current interaction, and $|V_{q\bar{q}}|^2$ is the appropriate CKM matrix element.

Multijet production

The total W and Z production cross sections can be decomposed in multijet cross sections of increasing arder in α_s, and summed over. For example for the W boson:

$$\sigma_W = \sigma_{W+0jets} + \sigma_{W+1jet} + \sigma_{W+2jets} + ...$$

where, in turn, each cross-section can be expanded as follows:

$$
\begin{aligned}
\sigma_{W+0jets} &= a_0 + \alpha_s a_1 + \alpha_s^2 a_2 + ... \\
\sigma_{W+1jet} &= \alpha_s b_1 + \alpha_s^2 b_2 + ... \\
\sigma_{W+2jets} &= \alpha_s^2 c_2 + ... \\
... & .
\end{aligned}
$$

The coefficients a_i, b_i, c_i, ... in these expansions are in general functions of the jet-definition parameters, for example the cone size used to cluster the partons into jets, and the transverse momentum, rapidity and separation cuts imposed on jets/clusters. The sum at each order in perturbation theory, i.e.

$$
\begin{aligned}
a_0 &= \delta_0 \\
a_1 + b_1 &= \delta_1 \\
a_2 + b_2 + c_2 &= \delta_2 \\
&...
\end{aligned}
$$

is independent from the jets parameters and corresponds simply to the perturbative expansion in powers of α_s the total cross-section.

The largest contribution to the vector boson plus jets cross-section is given by the sum of the very first coefficients of every "exclusive" multijet cross-section a_i, b_i, c_i, ... It is possible to calculate these coefficients from the Feynman diagrams of partonic processes $xy \to V + j_1...j_n$, where x, y, j_i are quarks and gluons, and V is a vector boson, Z or W. The explicit calculations of these coefficients were carried out by Berends and Giele for a multiplicity of jets up to $n_{jets} \leq 4$ [22]. Berends and Giele also studied the ratio between the cross section with n jets and the cross section with $n-1$ jets in the final state

$$f_n(V) = \frac{\sigma_{(V+n_{jets})}}{\sigma_{\left(V+(n-1)_{jets}\right)}}$$

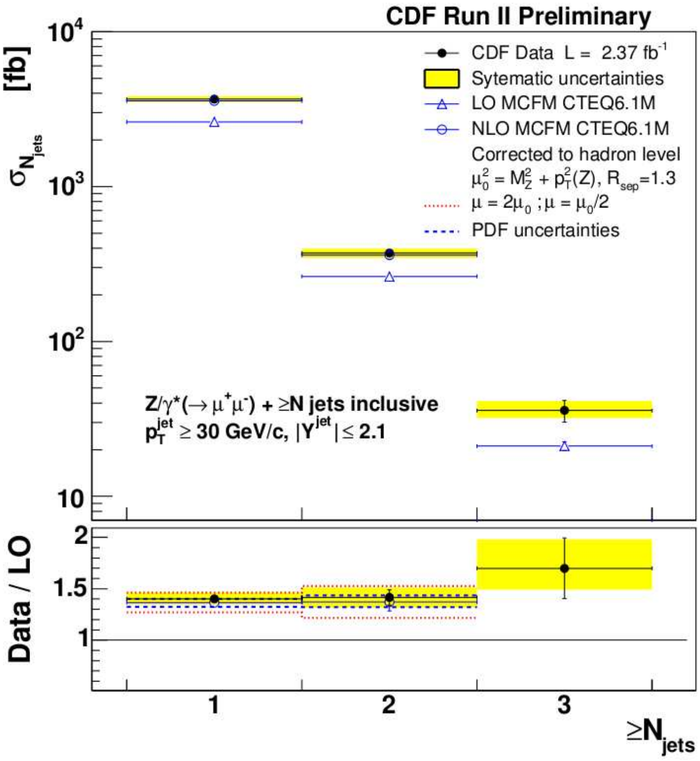


Figure 2.5: Measured total cross section for inclusive jet production in $Z/\gamma^* \rightarrow \mu\mu$ events, as a function of the number of jets. The data are compared to LO and NLO perturbative QCD predictions.

and found that this ratio should be constant. Therefore $f_n(V)$ can be parametrized as

$$f_n(V) = \alpha + \beta n_{jets} \quad .$$

This scaling has been tested in various experiments, for example Figure 2.5 shows the cross-section for the production of a Z boson in association with n jets ($\sigma_{N_{jets}}$) versus the number of jets as measured by the CDF collaboration at Tevatron, for a center-of-mass energy of 1.8 TeV [23].

2.4.1 Study of W+jets and Z+jets associated production at LHC

With the current LHC center-of-mass energy of 7 TeV, che production cross-section for Z and W bosons is about 1 nb and 10 nb respectively. During 2011, about 5 fb^{-1} of data were collected so that, roughly, we expect to measure order of 10^6 Z and 10^7 W candidates.

The study of the production of a Z or W boson with accompanying hadronic jets is one of the most important Standard Model processes in high-energy hadron-hadron collisions. Most "new physics" processes, for example the production of heavy quarks, Higgs boson and supersymmetric particles, can be mimicked by the production of vector bosons in association with jets. It is therefore important to estimate these backgrounds accurately. In addition, quantitative tests of QCD are possible: the rate of multijet production is dependent on the strong coupling constant

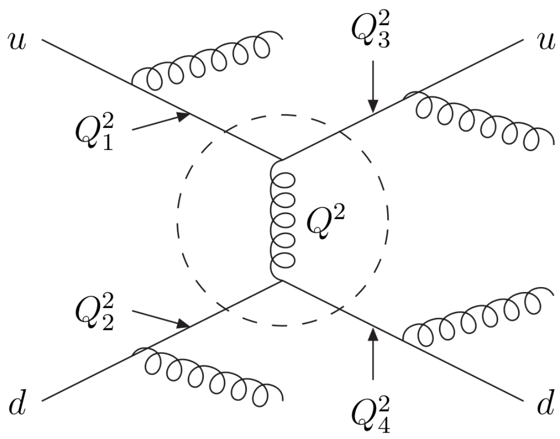


Figure 2.6: The "factorization" of a $2 \to n$ process.

and the kinematic distributions of the jets probe the undelying scattering matrix elements.

Direct measurements of the production cross sections of $W+n$ jets or $Z+n$ jets suffer from inherent theoretical and experimental uncertainties associated with the definition and counting of jets. Among the dominant experimental uncertainties are the energy response of the detector to a jet (energy scale), additional energy contributions from the underlying event, backgrounds from misidentified non-electroweak events, and jet acceptance. Such effects can modify the measured number of jets. Uncertainties in the theoretical SM predictions are dominated by the choice of the parton distribution functions (there are several possibilities in literature) and of the scale variable (identified by M in the previous example) on which they depend, initial/final state radiation (ISR/FSR), and the non-peturbative evolution of partons into on-shell particles that could be detected as jets.

2.5 Monte Carlo events generation with Parton Shower

The most direct method to estimate cross-sections in the perturbative QCD regime is to make the calculations of the transition matrix elements (ME) expressed by the Feynman diagrams of the particular process/perturbative order; combining the matrix element with the process phase-space it is possible to get the desidered cross-section. Unfortunately, for some processes the exact calculation of the matrix elements leads to divergences. However, it is possible to obtain an approximate description of the cross-section for processes with several gluons and quarks in the final state, using the *Parton Shower* approach [24].

Consider for example the process of quark scattering in Figure 2.6, where many radiative processes occur. The starting point is to decompose a complex $2 \to n$ process, where n represents a large number of partons in the final state, into a simple core process convoluted with showers. In the hard process at Leading Order, for example $ud \to ud$, the incoming and outgoing quarks must be on the mass shell at long timescales (i.e. satisfy $p^2 = E^2 - \mathbf{p}^2 = m^2 \sim 0$ as $t \to \infty$). By the uncertainty principle, however, the closer one comes to the hard interaction, or the shorter the

timescales considered, the more off-shell the partons may be. This implies that the incoming quarks may radiate a succession of harder and harder gluons, while the outgoing ones radiate softer and softer gluons.
Thus the hard subprocess (Q^2) cross-section is calculated first at LO with matrix element expression, with the approximation that all the other processess (Q_i^2) can be neglected ($Q_i^2 \ll Q^2$). In other words, first the hard process can be picked without any reference to showers, and only thereafter are showers added with unit probability, branching quarks lines with emitted quarks and gluons. This techinque is then well suited to generate events with Monte Carlo methods. It is possible to do an approximate calculation of the contributions from the emission of quarks and gluons both for Initial State Radiation (ISR) and Final State Radiation (FSR) (the details are not discussed here). One important thing to note is that this branching process should end at some time, because perturbation theory will cease to be meaningful at so small Q^2 scales where $\alpha_s(Q^2)$ diverges; at that point confinement effects and hadronization phenomena take over. Typically some lower cutoff at around 1 GeV is used to regulate both soft and collinear divergences: below such scale no further branchings are generated.

Mixing Matrix Element and Parton Shower approaches

To summarize, the explicit calculation of the Matrix Element gives the cross-sections for final states with jets well separated in the phase-space and with sufficient energy such that they fall in the perturbative regime of the theory. The Parton Shower method is well suited to describe processes within the perturbative regime with small transferred momentum Q^2, which consist in the emission of soft quarks and gluons almost collinear with respect to the initial parton.
The two methods are essentially complementary, because they cover (almost always) separate regions in the phase space. The procedure of merging without double counting or gaps in the phase space coverage is not trivial, and several alternative approaches have been developed. Therefore the best results in calculating hadronic cross-sections are nowadays obtained combining ME exact calculations at LO with PS simulations. The next frontier in MC generators is to combine ME NLO calculations with PS.

Chapter 3

Electrons and jets selection and 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 in single sub-detector channels are collected and, through several specialized algorithms, energy clusters and track segments are built. These are then used to construct higher-level objects such as electrons, photons, muons and jets, consequently used for physics analysis.

The most interesting particles for the LHC physics program have a lifetime too short to travel the distance that separates the collision point from the sensitive elements of the detector. The unstable particles produced, after a process of decay of varying complexity, originate a final state containing a diverse number of light and stable particles (aforementioned). The example that most closely interests the present work is the decay of a vector boson Z into a pair of leptons or quarks; the decay channels are quoted here with their Branching Ratio (BR) [25]:

$$Z^0 \rightarrow \begin{cases} e^+e^-, \mu^+\mu^-, \tau^+\tau^- & BR: 10.096 \pm 0.019\ \% \\ \nu_e\bar{\nu}_e, \nu_\mu\bar{\nu}_\mu, \nu_\tau\bar{\nu}_\tau & BR: 20.00 \pm 0.06\ \% \\ u\bar{u}, d\bar{d}, c\bar{c}, s\bar{s}, b\bar{b} & BR: 69.91 \pm 0.06\ \% \end{cases}$$

Even if the BR is larger for the hadronic decay channels, the leptonic ones (e^+e^-, $\mu^+\mu^-$) have a much clearer signature, enabling to separate the Z decay products from the hadronic jets, in particular in the LHC environment where there is a large background from pure QCD events.

In this work the decay channel considered is $Z \rightarrow e^+e^-$. Hence in the next Section, 3.1, the procedure to reconstruct electrons in CMS is presented. As already stated, the perturbative corrections to the Born level matrix element include the emission of jets (see Section 2.4), therefore Section 3.2 is dedicated to the jet reconstruction and related problems.

3.1 Electrons selection and reconstruction

Electron selection and reconstruction at CMS, for the decay $Z \rightarrow e^+e^-$ analysis can be schematized as follows: the interesting events for the analysis are *selected* on-line by a suitable trigger menu (a double electron trigger) implemented at the HLT level (see Section 4.2.1). The selected events are then *reconstructed* off-line, creating analysis-level objects by combining recorded signals from the detectors; during reconstruction the tracks and energy deposits in calorimeters of all candidate electrons are built-up. During *analysis* some additional selection criteria are applied to reject the background (see Section 3.1.4); the main background for primary electrons in CMS comes from "fake" electrons from hadron decays in jets, but also prompt electrons from semi-leptonic decays, and possibly electrons from early photon conversions in the tracker material.

3.1.1 Electron clustering

Electron showers deposit their energy in several ECAL crystals. Approximately 94% of the incident energy of a single electron is contained in 3×3 crystals, and 97% in 5×5 crystals.

As they cross the material in front of the calorimeter, they may radiate photons through the Bremsstrahlung process, which results in a spread in ϕ of the initial electron energy because of the bending induced by 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. About 35% of the electrons radiate more than 70% of their initial energy before reaching the ECAL. In about 10% of the cases, more than 95% of the initial energy is radiated[1] [26]. Thus, to obtain a measurement of the electron energy at primary vertex and minimize the cluster containment variations, it is essential to collect bremsstrahlung photons. This is the purpose of the superclustering algorithms. A "supercluster" (SC) groups one or more associated clusters of energy deposits, and it is constructed by using an algorithm which takes the magnetic field-induced energy spread into account. More in detail, "hybrid" and "island" clustering algorithms are used for electrons in the ECAL barrel and endcaps respectively.

The Hybrid algorithm takes advantage from the simple geometry of the Barrel and exploits the properties of the lateral shower shape in the η direction, while dynamically searching for bremsstrahlung energy in the ϕ direction. In the language of the hybrid super-clustering, what is considered as a "seed" cluster is a collection over ϕ of contiguous clusters made of 3 to 5 crystals in η, and separated by other such collections by a valley where less than 100 MeV is observed in a cluster.

The Island algorithm in the Endcaps starts with a seed crystal with local energy maximum above a defined threshold. Starting from the seed position, it builds clusters by connecting rows of crystals containing energies decreasing monotonically when moving away from a seed crystal. Then, superclusters are built by collecting other island clusters in a ϕ road in both directions around each island clusters, starting from a list of clusters ordered in E_T, in a procedure called bremsstrahlung recovery [27].

[1]This study [26] was carried out using a Monte Carlo simulation of back-to-back e^+e^- pairs in absence of pile-up. Electron samples with different fixed energy (10, 30 and 50 GeV) or fixed p_T (10, 30 and 50 GeV/c) have been considered.

3.1.2 Electron track reconstruction

The reconstruction of electrons in CMS starts with the creation of superclusters. Electrons are then characterized by the presence of a charged track pointing to the electromagnetic cluster. Starting from the reconstructed clusters, roads can be built to search for hits in the innermost part of the tracker. This procedure is used to define electron objects at the trigger level as well as for the offline reconstruction, where looser settings and thresholds are used to efficiently reconstruct electrons down to $p_t = 10$ GeV/c.

Practically, superclusters are matched to track seeds (pairs or triplets of hits) in the inner tracker layers and electron tracks are built from these 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 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 modeling [28] of the electron losses and a Gaussian-Sum Filter (GSF) [29] in the forward fit, see Section 3.1.5. This procedure is iterated until the last tracker layer, unless no hits are found in two successive layers. If more than one compatible hits are found on a subsequent layer, many candidate trajectories are grown in parallel. A minimum of five hits is required to reconstruct a track.

Such a cluster-driven pixel seed finding strategy for the tagging of primary electron-like objects has been successfully developed for robust applications at the High Level Trigger level, where fast and drastic reduction of fake background rates is a key issue [26].

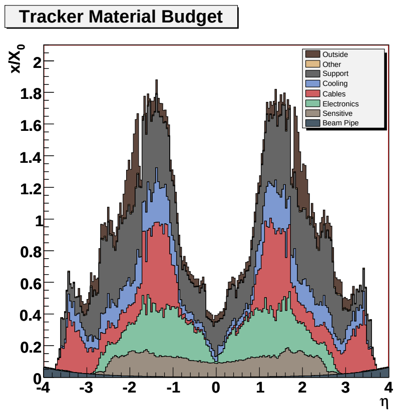
3.1.3 Electron classification

In addition to the problem of bremsstrahlung collection, the measurement of primary electron energies is affected by energy losses in the tracker material. As expected the mean energy loss increases with η, that is together with the increasing amount of tracker material, see Figure 3.1.

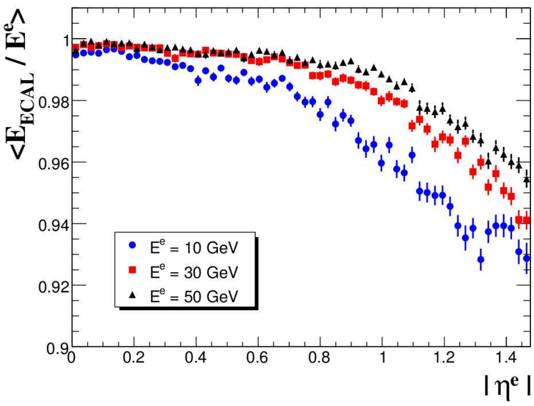
It is possible to divide the reconstructed electrons in categories, reflecting the amount of energy lost through bremsstrahlung radiation. These electron categories allow to take into account for the large differences in signal to background ratio and to put events of similar characteristics together. These classes, from “good” to “bad” electrons, are: golden, big-brem, narrow and showering electrons. Figure 3.2 shows the population of the different classes as a function of η. The shape of the distribution for the showering class clearly reflects the distribution of the material thickness. On the contrary, the observed distribution for the golden electrons is, as expected, anti-correlated with the material thickness [30].

3.1.4 Electrons isolation and selection

During data analysis some more filters are applied to the recorded and reconstructed events. These filters constitute the selection criteria which give the ultimate sample to be analyzed. The selection criteria are optimized to maximize the signal to background ratio and to meet some fixed electron selection efficiency. The resulting efficiency with the given cuts set is called Working Point (WP), then for example:



(a) Tracker material budget in units of radiation length versus pseudorapidity.



(b) The true energy reaching the ECAL barrel front face normalized to the initial energy as a function of the pseudorapidity for electrons of 10, 30 and 50 GeV energy. To compute the true energy reaching the ECAL, only photons with energy greater than 10 GeV are considered.

Figure 3.1: Comparison between tracker material budget and mean energy loss for electron versus η.

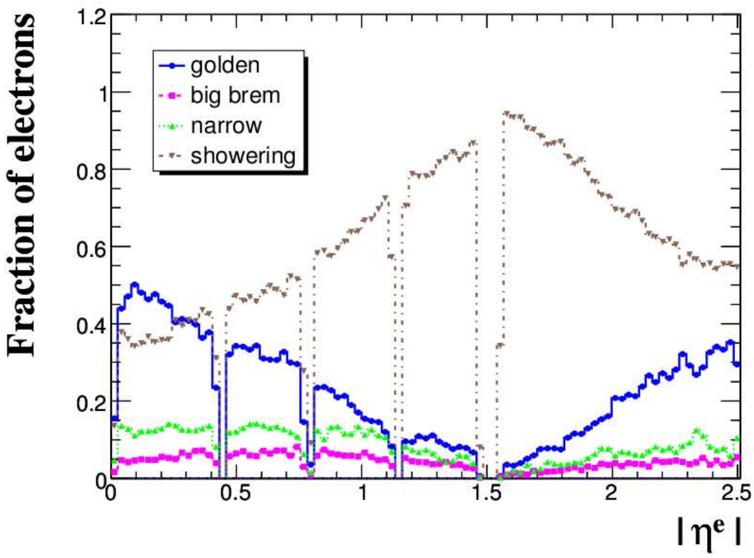


Figure 3.2: Fraction of electron population in the different classes as a function of the generated pseudorapidity for initial energies between 5 and 100 GeV. The Monte Carlo dataset used in this study [30] is a sample of nearly 2 million $Z \to e^+e^-$ events, generated with PYTHIA 6.223. The dips reveal the separation between EB and EE (at $\eta \sim 1.5$) and the submodules of EB (for $\eta < 1.5$).

WP80 refers to a set of cuts which gives a sample of electrons selected with efficiency of the 80%.

The aforementioned selection criteria are intended to: exclude background signals produced by particles that have a signature similar to the electron, reject secondary electrons not produced by the main p-p interaction that pass the HLT selection requirements. The point is to reduce the background without affecting the reconstruction efficiency of the interesting electrons.

The cuts are implemented by a set of variables which can be grouped by their function: electron identification, electron isolation and rejection of converted photons.

Electron identification

The cuts for electron identification are based on four variables that describe the transverse spread of the energy deposit, the alignment between the SC and the reconstructed track and the hadronic energy fraction of the SC.
These variables are listed below:

- $\sigma_{i\eta i\eta}$: as already stated, the energy left by the electron is distributed along ϕ due to the Bremsstrahlung effect and the presence of the magnetic field. This dispersion is not present over η, therefore the distribution of the energy released along pseudorapidity should be contained in few cristals. By cutting on the standard deviation of $\sigma_{i\eta i\eta}$ it is possible to exclude the contributions from particles that create a larger shower in the detector.

- Δ_η and Δ_ϕ: these variables describe the difference between the center of gravity of the SC and the track (reconstructed by the tracker) extrapolated to the ECAL surface. The request is then to have small values for these variables Δ_η, $\Delta_\phi \sim 0$.

- H/E: is the hadronic fraction which describes the ratio between the fraction of energy released in HCAL over the energy released in ECAL. If the interacting particle is an electron then its energy should be all contained in ECAL and the ratio should be H/E close to zero.

The applied cuts are different for Barrel and Encap, and are listed in Table 3.1.

variable	barrel		endcap	
	WP80	WP95	WP80	WP95
$\sigma_{i\eta i\eta}$	<0.01	<0.01	<0.03	<0.03
Δ_ϕ	<0.06	<0.80	<0.03	<0.70
Δ_η	<0.004	<0.007	<0.007	<0.010
H/E	<0.04	<0.15	<0.15	<0.15

Table 3.1: WP80 and WP95 electron identification cuts for 2011 data.

Electron isolation

This set of variables is tuned to reject the electrons coming from the decay of hadronization products of quarks and gluons. Since these electrons are surrounded

by many other particles coming from hadronization, it is possible to exclude them requiring the electron to be isolated. The isolation is defined in a cone of size $\Delta R = 0.3$ centered on the electron candidate track, where $R = \sqrt{\eta^2 + \phi^2}$; the value of $\Delta R = 0.3$ was determined by studying samples of simulated events [31].
Within the isolation cone, the following variables are defined:

- for the tracking system

$$IsoTrk = \frac{\sum_i p_T(i_{cone})}{p_T(e)}$$

  as the sum of all p_T tracks reconstructed in a cone around the electron (the latter excluded) divided by the momentum of the electron associated with the same track[2];

- for the electromagnetic calorimeter

$$IsoEcal = \frac{\sum_i E_T(i_{cone})}{E_T(e)}$$

- and for the hadronic calorimeter

$$IsoHcal = \frac{\sum_i E_T(i_{cone})}{E_T(e)}$$

  as the sum of all energy deposits in the cone around the deposit of the electron candidate (the latter excluded) divide by the energy deposit of the electron candidate.

For the list of cuts used in this work, see Table 3.2.

variable	barrel		endcap	
	WP80	WP95	WP80	WP95
$IsoTrk$	<0.09	<0.15	<0.04	<0.08
$IsoEcal$	<0.07	<2.00	<0.05	<0.06
$IsoHcal$	<0.10	<0.12	<0.025	<0.05

Table 3.2: WP80 and WP95 electron isolation cuts for 2011 data.

Conversion Rejection

Another source of background for the electrons reconstruction comes from the conversion of photons to positron-electron pairs, through a pair production process. Therefore, the presence of a positron and an electron coplanar in the $r - \phi$ plane with the same vertex in the final state is the pair production signature. Hence it is possible to define two variables to filter the electrons from photon conversion:

[2] If $IsoTrk \to 0^+$ then the track (and hence the electron) is perfectly isolated. Same meaning for $IsoEcal$ and $IsoHcal$.

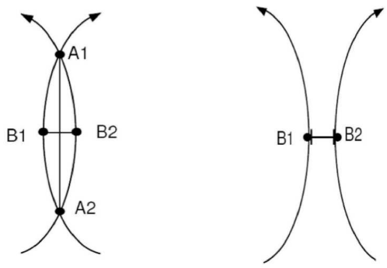


Figure 3.3: The D_{ist} variable is defined as the distance between the points B_1 and B_2, where the two particle trajectories become parallel.

- D_{cot}: defined as the difference between the polar angle cotangent of the selected electron track, $\cot\theta_{\text{gsf}}$, and the polar angle cotangent of the coversion positron partner $\cot\theta_{\text{ctf}}$ (or viceversa):

$$D_{cot} = \cot\theta_{\text{gsf}} - \cot\theta_{\text{ctf}}$$

- D_{ist}: defined as the distance between the two particle trajectories at the point where they become parallel, see Figure 3.3.

It is possible to further improve this selection by rejecting all the tracks without hits in the first layers of the tracker. This is because the photon, being neutral, does not leave signal in the tracker and converts only after interacting with the tracker material. Cutting on the tracker missing hits then makes the selection more efficient. For the list of cuts used in this work, see Table 3.3.

variable	WP80	WP95
D_{cot}	>0.02	none
D_{ist}	>0.02	none
missing hits	0	$\leq$1

Table 3.3: WP80 and WP95 photon conversion rejection cuts for 2011 data.

3.1.5 Gaussian Sum Filter electrons

The baseline for track reconstruction in the tracker is the Kalman filter, which was introduced to the high-energy physics community in the DELPHI experiment at the LEP collider at CERN. The Kalman filter [32] is a recursive formulation of the least-squares method of fitting a set of measurements to a track model. It alternates between propagation steps and update steps. In the propagation step, the track state is extrapolated to the next observation layer, and material effects are added to the covariance matrix. In the update step, the extrapolated state is combined with the observation. Due to its recursive character, the Kalman filter is well suited for doing track finding, also called pattern recognition, and track fitting concurrently.

Kalman filter is optimal when all probability densities encountered during the track reconstruction procedure are Gaussian, but not necessarily in other cases. The implicit approximation of the energy loss distribution by a single Gaussian in the

Kalman filter is quite crude. It is therefore plausible that a non-linear estimator which takes the actual shape of the distribution into account would yield better results. A non-linear generalisation of the Kalman filter (KF), the Gaussian-sum filter (GSF) [29], has therefore been implemented in the reconstruction software of the CMS tracker. The algorithm requires that all probability densities involved in track reconstruction are described by Gaussian mixtures.

A key component in the application of the GSF to electrons is the approximation of the Bethe-Heitler energy loss distribution[3] by a Gaussian mixture. By adding together different Gaussians it is possible to model different degrees of hardness of the bremsstrahlung in the layer under consideration. In the GSF, the distributions of all state vectors are allowed to be Gaussian mixtures, i.e. weighted sums of Gaussians, instead of single Gaussians as in the Kalman filter. Still, the basic structure of the GSF is very similar to the standard Kalman filter, having the same alternating sequence of propagation and update steps. In each propagation step, a weighted mixture of state vectors and corresponding covariance matrices is propagated to the nearest surface, yielding a predicted mixture.

3.2 Jet reconstruction

The CMS software framework includes several analysis tools, for instance many different algorithms for particle identification and reconstruction of complex objects such as hadronic jets or missing transverse energy. In the following sections the jet clustering algorithm anti-k_t, the event reconstruction algorithm *Particle Flow* and the pile-up removal system that makes use of *Fastjet* are presented.

3.2.1 Particle-Flow algorithm

The particle-flow reconstruction algorithm aims at providing a global (i.e., complete and unique) event description at the level of individually reconstructed particles, with an optimal combination of the information coming from all CMS subdetectors [33]. The reconstructed and identified individual particle list includes muons, electrons (with individual reconstruction and identification of all Bremsstrahlung photons), photons (either unconverted or converted), charged hadrons (without or with a nuclear interaction in the tracker material), as well as stable and unstable neutral hadrons. These particles can be non isolated, and even originated from a intricate overlap of reconstructed charged particles, ECAL and HCAL energy clusters, and signals in the muon chambers. The particle-flow algorithm starts with reconstruction performed independently within each CMS sub-detector: clustering is conducted separately in each of the electromagnetic (ECAL) and hadronic (HCAL) calorimeters and track reconstruction takes place in the combined silicon and pixel tracker system.

Therefore this algorithm is a generalization of the electron reconstruction process already described in Section 3.1.2. The objects that constitute the starting points for the PF are:

[3] $f(z) = (-\ln z)^{c-1}/\Gamma(c)$ with $c = t/\ln 2$, where t is the path length in the material (in units of radiation length) and z is the fraction of energy remaining after the material is traversed.

- *PFTracks*: all the reconstructed hits in the tracker (rechits) are combined to evaluate the track candidates
- *PFCluster*: similarly, the energy deposits in the calorimeters are clustered with a dedicated algorithm to measure the energy of each particle
- *Global Muons*: muons are reconstructed by using the hits from the muon chambers and from the tracker

The algorithm relies on an efficient and pure track reconstruction, on a clustering algorithm able to disentangle overlapping showers, and on an efficient link procedure to connect these blocks together [34]. The list of particles that interact with the sub-detectors (and thus reconstructed), called *PFCandidates*, are:

- *PFElectrons*
- *PFMuons*
- *PFPhotons*
- *PFHadrons*

This list of individual particles is then used to build jets (*PFJets*), to determine the missing transverse energy E_T^{miss} (*PFMet*) (defined in Section 1.1.2), to reconstruct and identify taus from their decay products, to quantify charged lepton isolation with respect to other particles (*PFTau*), to tag *b*-jets, etc [33] [35].
The resulting list of objects (reconstructed particles, jets and met) constitutes a global description of each event, available for subsequent physics analysis.

3.2.2 Comparison of GSF and PF algorithms

In the CMS software framework there are two algorithms for the electron reconstruction which produce different objects: the GSF electrons (See Section 3.1.5) and the Particle Flow ones (see Section 3.2.1). The basic steps of the algorithms and their main differences are summarized in the following list:

- the selection of the seed from which the procedure starts:

  GSF uses ECAL energy clusters

  PF uses simple tracks

- the SuperCluster (SC) reconstruction:

  GSF searches for other ECAL energy clusters in a window around the seed,

  PF searches clusters that match tangent trajectories to the main track

- the track reconstruction:

  GSF the hits in the tracker are combined with a filter suitable for reconstruction of tracks with kinks (Gaussian-Sum-Filter)

  PF uses a simplified version of the full GSF filter

- the link of the two objects:

  GSF from ECAL SC to tracks

  PF from tracks to ECAL deposits

- the determination of the electron momentum, combining track and energy information or using only one of them:

  GSF depends in particular on the E/p ratio

  PF depends on the number of track hits, checks the sensitivity to Bremsstrahlung.

Up to now many analyses used Particle Flow jets and electrons from the GSF reconstruction, but the two algorithms are not fully integrated. In the context of a precision measurement it is necessary to have a consistent reconstruction of the whole event in order to avoid potential double counting ECAL deposits or generation of artificial missing energy. A possible solution is to use the reconstruction entirely based on the PF algorithm. A comparison test of consistency and performance between GSF and PF electron has been carried out in [36], concluding that there is no particular difference between the two descriptions.

Up to now our study is based on GSF electrons: the final version of the analysis is expected to be totally PF oriented.

3.2.3 Anti-$\mathrm{k_t}$ algorithm

In an hadron collider with high center-of-mass energy and luminosity as the LHC, the jet multiplicity can reach 7-8 jets per event. The purpose of measuring the jets' properties, i.e. measure the observables of its constituents, is to investigate some characteristics of the original parton. Since every jet is made of several particles the resulting area on the detector can become sizeable and the jet areas can overlap. Therefore, correctly reconstructing each jet by associating the energy of the original parton can be very difficult, but is of crucial importance.

Jet clustering algorithms are among the main tools for analysing the data from hadronic collisions, and are essentially grouped into two categories: sequential recombination (e.g. k_t and Cambridge/Aachen) and cone algorithms [37]. The main differences are related to the regularity of the boundaries of the resulting jets, which are associated to the sensitivity of the algorithms to non-perturbative effects like hadronisation and underlying event contamination.

Among the wide variety of algorithms, the CMS collaboration has chosen the anti-k_t algorithm, after extensive comparative studies (see as reference [31]). The anti-k_t is a sequential recombination algorithm which is infrared and collinear safe; these are essentially consequences of the resilience of jet boundaries with respect to soft radiation. The algorithm is formulated as follows [37]:

1. for each pair of particles i, j work out the distances d_{ij} and d_{iB}, according to the following equations

$$
\begin{aligned}
d_{ij} &= \min\left(p_{ti}^{2\mathfrak{p}}, p_{tj}^{2\mathfrak{p}}\right)\frac{\Delta R_{ij}^2}{R^2} \qquad \text{where} \qquad \Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \\
d_{iB} &= p_{ti}^{2p}
\end{aligned}
$$

where $R = \sqrt{\eta^2 + \phi^2}$ defines the size of jets to be clustered and $\mathfrak{p}$ is a parameter that is $\mathfrak{p} = -1$ for the anti-k_t algorithm (1 for the k_t algorithm, and 0 for Cambridge/Aachen).

2. find the minimum of the d_{ij} and d_{iB}

3. if the minimum is the d_{ij}, recombine i and j into a single new particle and return to step 1.

4. otherwise, if it is a d_{iB}, declare i to be a [final-state] jet, and remove it from the list of particles. Return to step 1.

5. stop when no particles remain.

Less schematically: the two "distances" d_{ij}, d_{iB} are calculated initially for two particles[4] i and j: if $d_{ij} < d_{iB}$ then the two particles are recombined to form a single new particle, otherwise the second particle j is rejected. This procedure is iterated for all the particles in the event. Those which satisfy $d_{ij} < d_{iB}$ are clustered in the same jet, and are removed from the list of particles to be grouped. The procedure starts again for the particles left out from the cluster, and stops when every particle is a "final-state jet".
The optimal value for R is obtained through detailed studies on Monte Carlo samples, and is fixed to $R = 0.5$ then used for this analysis [31].

3.2.4 Pile-Up removal with Fastjet

At each bunch crossing in 2011 there were order of 10 minimum bias p-p interactions, which pollute any interesting hard events with many soft particles. The beams at LHC have a longitudinal spread, and it is possible experimentally to associate each charged particle with a distinct primary vertex that corresponds to a single p-p interaction and so eliminate some fraction of the soft contamination. However for neutral particles this is not possible, and therefore kinematic measurements for jets are adversely affected by pile-up, with resolution and absolute energy measurements suffering significantly.

The method used by the CMS collaboration is an event-by-event, jet-by-jet pileup subtraction approach that is suitable for any infrared safe jet algorithm, and which performs the corrections after the jet finding has been carried out, so as to ensure independence on the specific detector setup.
It is based on two ingredients: (1) the measurement of each jet's susceptibility to contamination from diffuse noise and (2) an essentially parameter-free technique to measure the level, ρ, of this diffuse noise[5] in any given event.

1. The practical measurement of the jet areas is carried out using the *FastJet* package [38] and relies on the fast computational strategies for jet clustering.

[4]What is identified as a "particle" is a task fulfilled by the Particle Flow algorithm, see Section 3.2.1.

[5]By noise (or also "background"), we refer to any form of diffuse contamination in the event, usually due to minimum bias pile-up and to some extent the underlying event.

Given a suitable definition of jet area, the modification of a jet's transverse momentum p_t by diffuse noise can be shown to be [39] (neglecting the fluctuations of noise and other effects connected with clustering algorithms):

$$p_{tj}^{(sub)} = p_{tj} - A_j \cdot \rho$$

where A_j is the jet's area and ρ, the level of diffuse noise, corresponds to the amount of transverse momentum added to the event per unit area, for example by minimum bias particles.

Two main definitions of the jet area can be introduced. A first one, the passive area, involves scanning a single infinitely soft particle (a "ghost") over the rapidity-azimuth cylinder and determining the region over which it is clustered within a given jet. It can be understood as a measure of the susceptibility of the jet to contamination from an UE with pointlike structure. A second definition, the active area, involves adding a dense coverage of infinitely soft "ghosts" and counting how many are clustered inside a given jet. It can be understood as a measure of the susceptibility of the jet to contamination from an UE with uniform, diffuse structure. The passive and active areas differ because in the latter case the ghosts can cluster among themselves as well as with the actual event particles, thus playing a more active role in the clustering.

At high luminosity at LHC ρ, is expected to be $\sim 10 - 20$ GeV per unit area. The second term in the equation above is therefore the geometrical contamination of the jet.

2. In order to estimate ρ, Cacciari and Salam [40] suggest

$$\rho = \text{median}\left[\left\{\frac{p_{tj}}{A_j}\right\}\right]$$

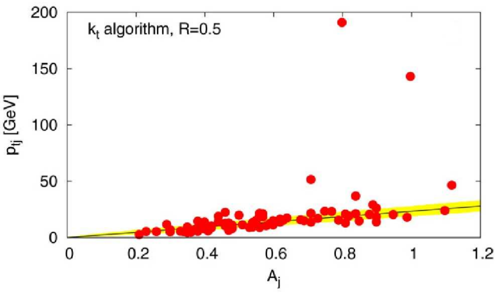
Note that the above pile-up subtraction procedure is valid as long as three conditions hold:

- The pileup noise should be independent of rapidity and azimuth.
- The radius R of the jet algorithm should be bigger than the typical distance between minimum bias particles, otherwise the extraction of ρ from the median will be biased by the large amount of empty area not associated with jets.
- The number of pileup jets should be much larger than the number of jets from the hard interaction that are above scale $\langle A \rangle \rho$. If this is not the case, then the median will be significantly biased by the hard jets.

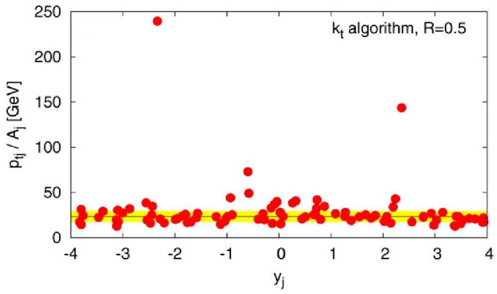
In Figure 3.4 is quoted the calculation of ρ for a dijet event with 22 pile-up interactions as an illuminating example.

3.2.5 Jet Energy Corrections

The measured jet energy has not a linear relationship with the true corresponding energy and therefore it is not straightforward to translate the measured jet energy to the true particle or parton energy. The Jet Energy Corrections (JEC) are a set of



(a) Scatter plot of the jet transverse momentum p_{tj} versus its area A_j, for an LHC dijet event with a pileup of 22 minimum bias interactions (simulated with the default tune of Pythia 6.325). The line is given by ρA_j.



(b) The ratio p_{tj}/A_j as a function of the rapidity, y_j , for the same event; the line and band are given by ρ.

Figure 3.4: Evaluating ρ for a dijet simulated event with 22 pile-up interactions.

tools that allows the proper mapping of the measured jet energy deposition to the analysis desired level (particle or parton).

CMS has adopted a factorized solution to the problem of jet energy corrections, where each level of correction takes care of a different effect.
Two important points must be clarified in the JEC's flow: each level of correction is essentially a scaling of the jet four momentum with a scale factor (correction) which depends on various jet related quantities (p_t, η, flavour, ...); the levels of correction are applied sequentially (the output of each step is the input to the next) and with fixed order.

The steps and goals of corrections implemented are:

L1 *pile-up*: remove the energy coming from pile-up events. In principle this will remove any dataset dependence on luminosity, so that the following corrections are applied on a luminosity independent sample.

L2 *relative jet correction*: make the jet response flat along η. Essentially, the uniformity in pseudorapidity is achieved by correcting a jet placed at arbitrary η relative to a jet in the central region ($|\eta| < 1.3$). The derivation of the Relative correction is done either by using MC truth or by employing a data driven method (dijet balance).

L3 *absolute jet correction*: make the jet response flat over p_t. The derivation of the absolute correction is done either by using MC truth information or by employing data driven techniques (Z/γ+jet balance).

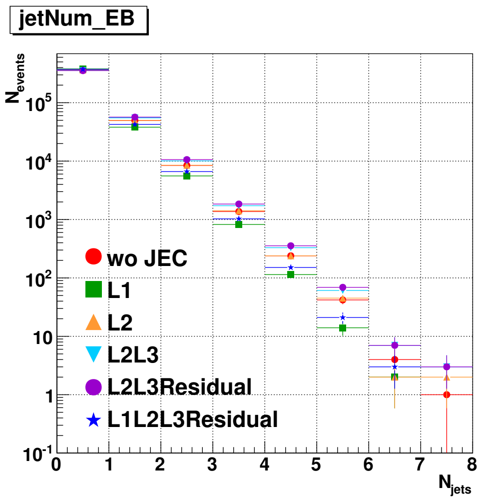
L4 *electromagnetic energy fraction*: make the jet response uniform versus the electromagnetic energy fraction (it is optional). It is a residual correction on top of the default L2+L3 and it has been shown to improve the jet resolution.

L5 *jet flavour correction*: correct for the jet flavor dependence (optional). The L2+L3 corrections scale the energy of an "average QCD jet" back to the energy of the corresponding generator level particle jet. However, an analysis of the individual jet flavors (u, d, s, c, b, gluon) shows that different corrections are needed for different jet flavours. L5 corrections provide corrections only relative to the "typical" mix of flavours present in the sample that they have been derived from.

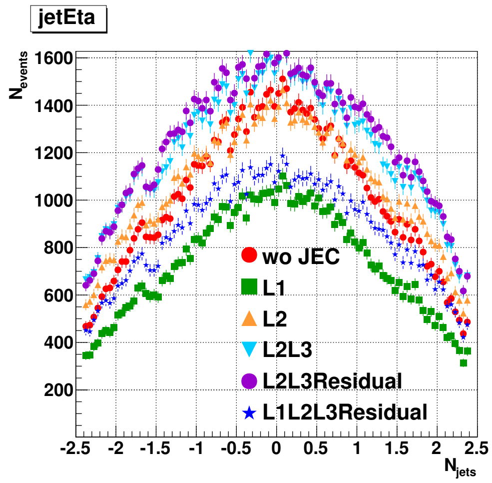
L7 *parton jet correction*: applied on top of the default L2+L3 correction and corrects back to the parton level, which means that the corrected p_t measured by calorimeters (*CaloJet*) is equal to the originating parton p_t on average.

L2L3 Residuals *residual calibration of data*: this correction takes care of fixing the small differences between data and MC (in η and p_t). By definition, the residual calibration is applied to data only.

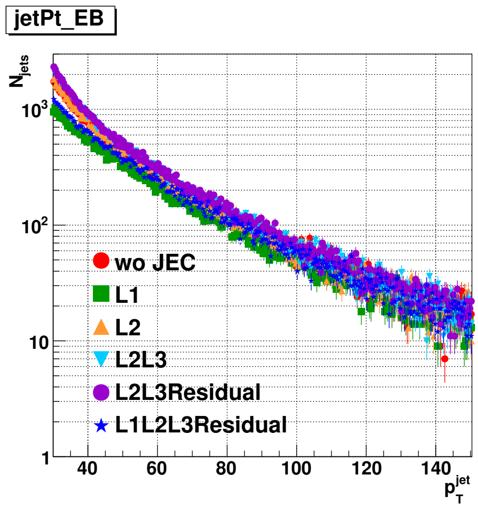
The mandatory corrections at CMS are: for Monte Carlo L1+L2+L3, while for Data L1+L2+L3+L2L3Residuals. See Figure 3.5 for the effect of JECs.



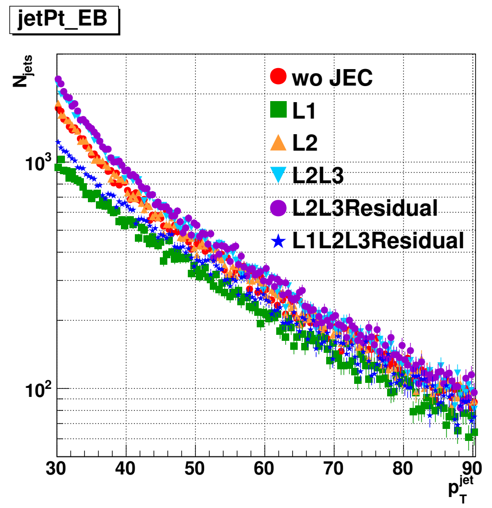
(a) Jet multiplicity in EB



(b) Jet pseudorapidity



(c) Jet transverse momentum in EB, from 30 to 150 GeV



(d) Jet transverse momentum in EB, from 30 to 90 GeV

Figure 3.5: Here are shown the JEC effects, in red the results without any correction, in green with L1, in yellow with L2, in ciano with L2 and L3, in violet with L2, L3 and the residual, in blu with all the JEC active.

Chapter 4

Study of the associated production of a Z boson and hadronic jets

In this chapter the main steps of the study of the associated production of a Z boson and hadronic jets are presented, with particular attention to the pile-up issue which is of crucial importance in 2011 data taking and relevant in the early steps of the analysis. The very first stage is the study of the unprescaled triggers at lowest energy available, together with an estimate of their constancy in time. Secondly, a study on how the distribution of the isolation variables is affected by pile-up has been carried out. Since the selection cuts are applied after the pile-up correction, the results of the analysis are directly affected by this procedure. The method for efficiency measurements, called "Tag&Probe", is then illustrated in Section 4.4.1. Finally the unfolding procedure is accomplished, obtaining (for example) the Z yield unfolded distribution ready to be compared with theoretical predictions.

4.1 Data and Monte Carlo samples

Data Samples

The analysis is performed using proton-proton collision data collected between March and August 2011 by the CMS experiment at a center-of-mass energy of 7 TeV. This period of data taking is called "RunA" (and is composed by several runs), and took place before the August technical stop. The following data taking run, "RunB" (September-December), has an higher instantaneous luminosity and pile-up. In Figure 4.1 the delivered and the recorded luminosity during 2011 data taking are plotted. The certified luminosity by the CMS experiment for RunA is about 2.2 fb^{-1}, from which are roughly expected 5×10^5 Z bosons with up to $7 \div 8$ jets, while the 2010 integrated luminosity was 36 pb^{-1} with up to 4 jets produced. The datasets used, containing double electron trigger events (see Section 4.2.1), are reported in Table 4.1. The events are reconstructed and analyzed with the CMSSW_4_2_4 version of CMS software.

data sets	**run range**
/DoubleElectron/Run2011A-ZElectron-05Jul2011ReReco-ECAL-v1	160329-168437
/DoubleElectron/Run2011A-ZElectron-05Aug2011-v1	170842-172619
/DoubleElectron/Run2011A-ZElectron-03Oct2011-v1	172620-175770

Table 4.1: The data sets used in the analysis with the relative run ranges.

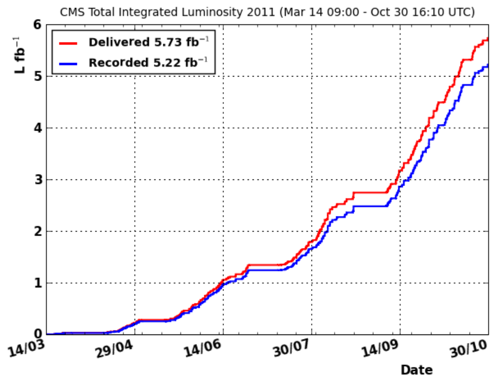


Figure 4.1: Total luminosity delivered and recorded in 2011.

Monte Carlo samples

The signal is modeled with a sample of Drell-Yan plus up to four generated jets events (Z+jets), restricted to leptonic decays generated with `MadGraph` [41]. The events produced with `MadGraph` are then passed to the shower Monte Carlo program `Pythia` [42] (with the parameters set called *tune Z2* [43]) where partons are perturbatevely evolved through the emission of QCD radiation, and turned into physical states. The total equivalent luminosity for this sample is about 12 fb^{-1}.

Electroweak backgrounds are modeled by `MadGraph` plus `Pythia` (like for Z+jets sample) and are W+jets and $t\bar{t}$+jets samples, with an equivalent luminosity of 2.6 fb^{-1} and 24 fb^{-1} respectively.

The QCD background is described by two types of samples, both simulated with `Pythia`:

- the `QCD_EMenr_XtoY` are QCD events, where at least one real, slightly isolated, electron or photon has been produced as part of a QCD diagram.
- the `QCD_BCtoE_XtoY` is the same as QCD_EM, but where the parent of the electron or photon is a bottom or charm quark.

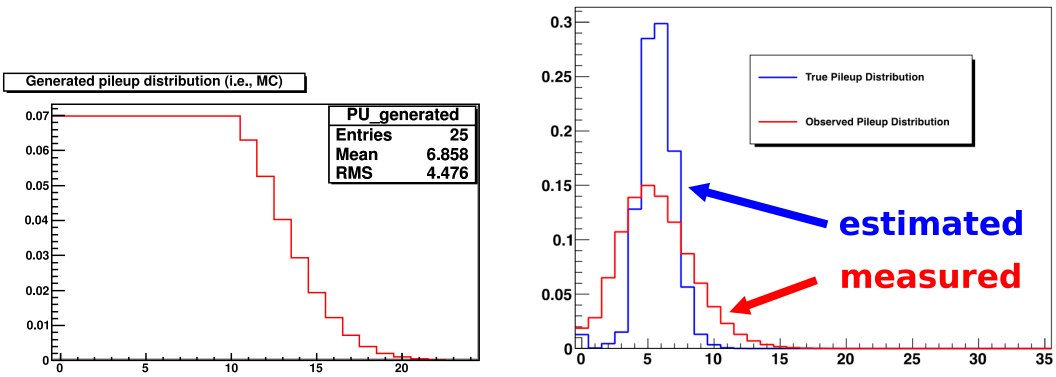
All these generated events are then passed through the detector simulation performed by the `Geant4` software [44]. For the full list of simulated samples used in this analysis see Table 4.2.

4.1.1 Monte Carlo reweighting

The instantaneous luminosity grew sistematically throughout the 2011 data taking (see Figure 1.5b). An "a-priori" knowledge of the pile-up final distribution was not accessible at the time of the production of MC samples. For this reason they could

Signal
/DYJetsToLL_TuneZ2_M-50_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1
Backgrounds
/TTJets_TuneZ2_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1
/WJetsToLNu_TuneZ2_7TeV-madgraph-tauola/Summer11-PU_S4_START42_V11-v1
/QCD_Pt-20to30_EMEnriched_TuneZ2_7TeV-pythia6/Summer11-PU_S4_START42_V11-v1
/QCD_Pt-20to30_BCtoE_TuneZ2_7TeV-pythia6/Summer11-PU_S4_START42_V11-v2
/QCD_Pt-30to80_EMEnriched_TuneZ2_7TeV-pythia6/Summer11-PU S4_S4_START42_V11-v1
/QCD_Pt-30to80_BCtoE_TuneZ2_7TeV-pythia6/Summer11-PU_S4_START42_V11-v1
/QCD_Pt-80to170_EMEnriched_TuneZ2_7TeV-pythia6/Summer11-PU_S4_START42_V11-v1
/QCD_Pt-80to170_BCtoE_TuneZ2_7TeV-pythia6/Summer11-PU_S4_START42_V11-v1

Table 4.2: Monte Carlo samples.



(a) Flat plus poissonian tail distribution used to generate Spring11 and Summer11 Monte Carlo vertices distribution.

(b) Measured (red) and estimated (blue) normalized pile-up distribution in data.

Figure 4.2: Comparison between Monte Carlo (on the left) pile-up distribution and data (on the right) PU distribution.

not be generated with the right PU contribution. To overcome this problem, all the MC samples were generated according to a more general distribution, and then reweighted to match the real distribution of the number of vertices measured in data. In particular, the MC samples of interest have been simulated with a "flat 10+Poisson tail" distribution for the number of pileup interactions which was meant to roughly cover the conditions expected for 2011 data-taking. The flat10+Poisson assumption has the peculiarity to be flat from 0 to 10 pile-up vertices and for higher values of PU to fall down with a poissonian tail, see Figure 4.2a.

The official CMS reweighting tool adapts the simulated number of vertices in MC to the "real" pile-up distribution in data. It is necessary to specify what "real" means: it is possible to reconstruct the number of vertices per bunch-crossing or to estimate the number of vertices from the known instantaneous luminosity and the minimum-bias cross section. In the first case the number of measured vertices does not always reflect the number of interactions occurred in the event, as shown in Figure 4.3 from a study performed on simulated events. In the second case the

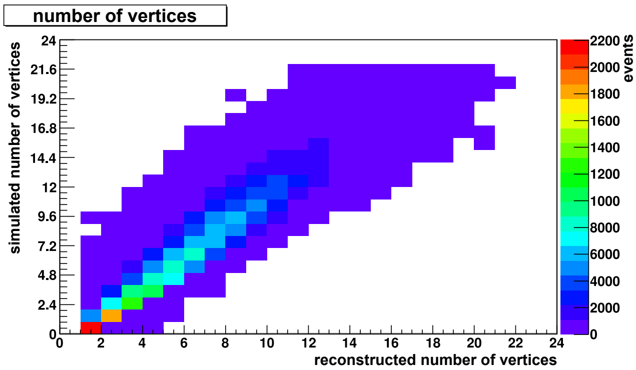
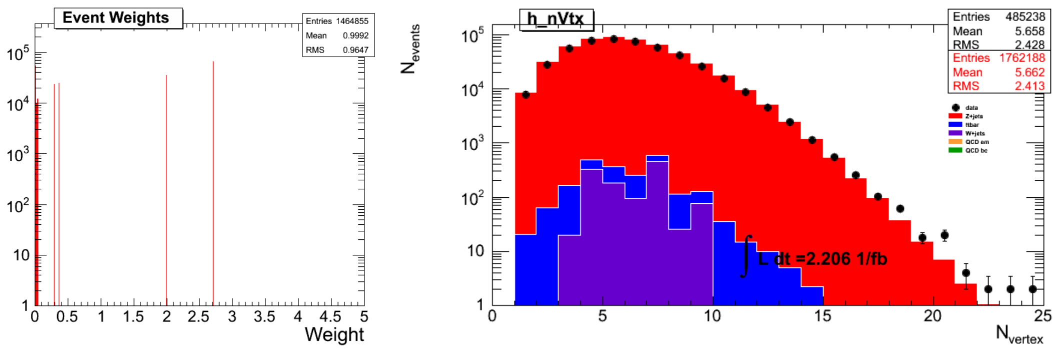


Figure 4.3: MC pile-up generated vertices versus measured vertices. The colour scale indicates the number (red high, blue low) of events.



(a) Set of weights used to reweight the pile-up distribution of Monte Carlo samples.

(b) Monte Carlo reweighted distribution (coloured area) plotted in comparison with data (black dots). Red area is the Z+jets (main contribution), then blue $t\bar{t}$+jets and violet W+jets Monte Carlo samples.

Figure 4.4: Result of the reweighting procedure: a set of weights is obtained and is applied to the Monte Carlo, the reweighted distribution of the number vertices then matches the data.

number of vertices is simply given by

$$\#PUvtx = \mathscr{L} \cdot \sigma_{minbias} \quad .$$

The two distributions are quite different and are reported in Figure 4.2b. The reweighting recipe adapts the simulated number of vertices in MC to the data (estimated distribution) by assigning to each event a weight, either greater or lower than 1. For example, a set of weights used to reweight the samples are shown in Figure 4.4a, while Figure 4.4b shows the result of the procedure.

4.2 Event selection criteria

In this ection the event selection approach is reviewed. The HLT selection methods and paths, the offline electron selections (identification and isolation) together with the pile-up removal and finally the jet energy corrections are described.

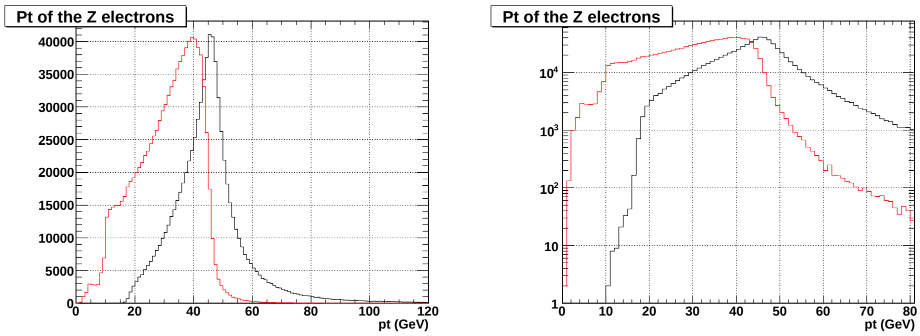


Figure 4.5: Electrons spectra: the leding electron is plotted with black line (higher mean p_t) and the second in p_t ranking is drawn with red line (lower mean p_t). On left with regular scale, and on right with log scale on y axis.

4.2.1 High Level Trigger

Due to the rapidly changing instantaneous luminosity during 2011, triggering data was one of the most challenging aspects. The physics requirements are translated in some trigger conditions called "paths" (e.g. HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL_v1); a list of paths is called "trigger menu". Thus paths and menus evolved during 2011 to cope with increasing luminosity and simultaneously to fulfill the analysis needs. From the analysis point of view, to extract the interesting events, a set of paths corresponding to the lowest-threshold unprescaled double electron trigger available has been used.

In our analysis, double electrons triggers are formed requiring asymmetric thresholds on the two electrons. The first electron is required to have a transverse energy E_T above 17 GeV, while the second one has a threshold of 8 GeV. The spectra of the two electrons are shown in Figure 4.5.

Electron candidates at the HLT stage are built by requiring a supercluster with transverse energy E_T above a certain threshold. This SC has to match an electromagnetic L1 trigger candidate and a hit in the pixel layers of the CMS detector, compatible with an electron trajectory. Due to the contamination from fake electrons coming from background, calorimetry and tracker-based electron identification and isolation criteria are applied at the HLT level, in order to reduce the respective rates. More requirements, in addition to the E_T ones, are thus added to select electron candidates (CaloId, CaloIso, TkId and TkIso), and are summarized in Table 4.3. The selected trigger paths for 2011 are:

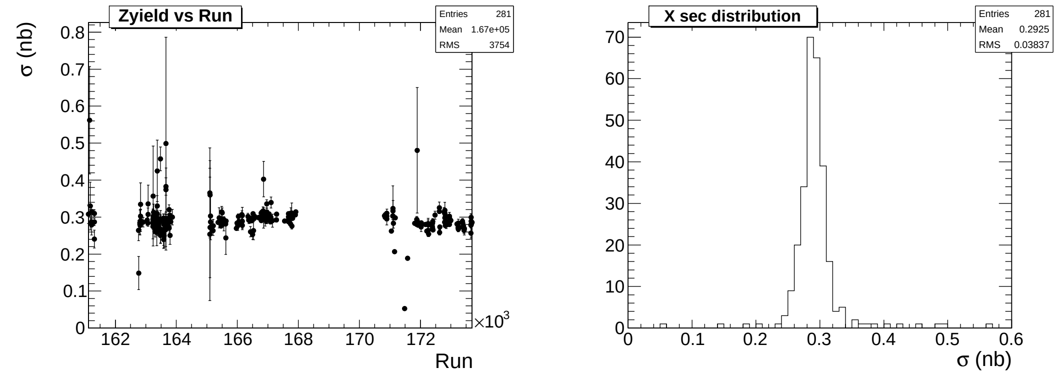
- HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL_v1 to v_6
- HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v7 to v_10
- HLT_Ele17_CaloIdVT_CaloIsoVT_TrkIdT_TrkIsoVT_Ele8_Mass30_v6 and v_7
- HLT_Ele32_CaloIdT_CaloIsoT_TrkIdT_TrkIsoT_Ele17_v1

Several tests have been carried out to check the HLT selection consistency. A good way to check whether the selected paths are stable during time or they suffer

Name	**Selections EB (EE)**	
CaloIdL	H/E	$< 0.15(0.10)$
	$\sigma_{i\eta i\eta}$	$< 0.014(0.035)$
CaloIdT	H/E	$< 0.10(0.075)$
	$\sigma_{i\eta i\eta}$	$< 0.011(0.031)$
CaloIsoVL	EcalIso/E_T	$< 0.2(0.2)$
	HcalIso/E_T	$< 0.2(0.2)$
TrkIdVL	dη	$< 0.01(0.01)$
	dϕ	$< 0.15(0.10)$
TrkIsoVL	TrkIso/E_T	$< 0.2(0.2)$

Table 4.3: Summary of the identification and isolation requirements, for EB and EE (in brackets).

for example from the increasing pile-up, is to plot the measured Z yield over the integrated luminosity per run. If the plot shows no evident trends or dips, then no gross mistakes are made. Figure 4.6 shows that the Z yield over integrated luminosity per run versus the run number is constant (within the errors) during the data taking period (RunA). This is confirms that the selected HLT paths are uniform throughout time.



(a) Z yield over integrated luminosity per run versus the run number

(b) Distribution of the Z cross section per run.

Figure 4.6: Z boson cross section per run versus the run number, without any efficiency correction.

4.2.2 Electrons selection

To identify the e^+e^- pair coming from the Z boson decay, the events must pass a series of different cuts; Table 4.4 shows the sequence of the cuts together with the percentage of passing events at each step with respect to the total number of events and to the previous cut. The first request for the event is to have triggered one of the specific double electron path; then the event is checked to have at least an electron-positron pair and the two leading electrons must be inside the η acceptance. The following cut requires the electrons to pass the WP80 cuts (see Section 3.1.4).

The distribution of WP80 isolation variables are reported in detail in Section 4.2.3. The passing electrons are ordered in decreasing order of p_t, and "tag" and "probe" electrons are identified. The tag electron is the "leading electron", the one with the highest p_t; the probe electron is the second in p_t ranking. In Figure 4.5 the tag and probe electron spectra are shown. The two electrons are then subjected to a p_t cut: the tag must have $p_t > 20$ GeV, while the probe $p_t > 10$ GeV. These values have been chosen to be above the HLT thresholds that are 17 and 8 GeV respectively. The very last requirement is that the two "electrons" must have opposite charge (so, more properly, an electron and a positron are identified).

Cut	**% total events**	**% previous cut**
events **triggered** by selected HLT paths	**48%**	–
events with ≥ 2 e	41%	85%
2 e in η acceptance	38%	92%
2 e fired specific HLT path	36%	96%
WP80	**17%**	48%
2 e with $p_t > p_{HLT}^{threshold}$	15%	85%
opposite charge	15%	$\sim 100\%$

Table 4.4: Example of the effect of the event selection cuts on the considered double electron datasets.

4.2.3 Pile-up removal in isolation variables

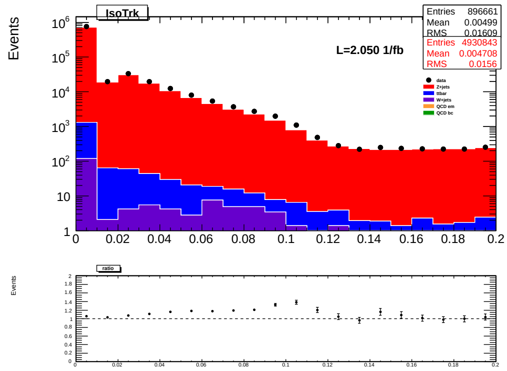
In order to estimate the effect of the PU, the distribution of the WP80 isolation variables have been studied. The study is carried out for both Barrel and Encap separately, and the effect of pile-up removal (PU-removal) corrections (see Section 3.2) is investigated.

In Figure 4.7 some distributions for *IsoTrk* variable are shown, on the left column for the Barrel and on the right for the Endcap. In the first row are reported the total distributions (i.e. with no cuts on the number of vertices of the event); in the second and third row are shown the distributions for events with one or five measured vertices. It is possible to notice that there is no particular difference in *IsoTrk* distribution for events with different number of pile-up vertices and that the distribution is well peaked at small values; therefore this variable shows no influence from PU and thus is left uncorrected.

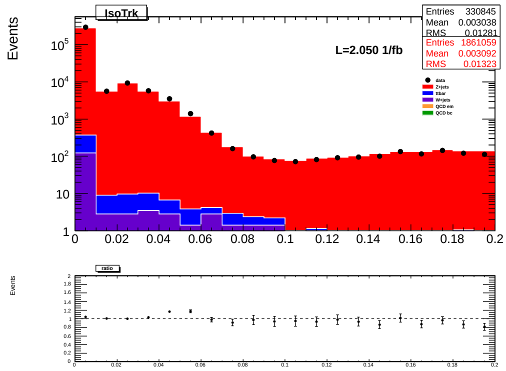
In Figure 4.8 and Figure 4.9 the distributions for *IsoEcal* variable, in Barrel and Encap respectively, are shown. In the left column the *IsoEcal* distributions with no PU-removal corrections activated are reported; this variable is not well peaked around zero, and for increasing number of pile-up vertices the distibution shifts to even larger values, deteriorating the isolation. In the right column the same plots with PU-removal corrections activated are shown; the effect of this correction is clearly visible as a shift of the distribution towards smaller values of isolation. Looking at PU-removed distributions for one and five vertices no big differences are highlighted; this shows that the procedure of pile-up removal is effective, bringing back the distributions with more than one vertex to one-vertex-like condition.

The same happens, but less visibly, for the *IsoHcal* variable, see Figure 4.10 for the Barrel and Figure 4.11 for the Endcap. This means that *IsoHcal* is less affected by the pile-up than *IsoEcal*.

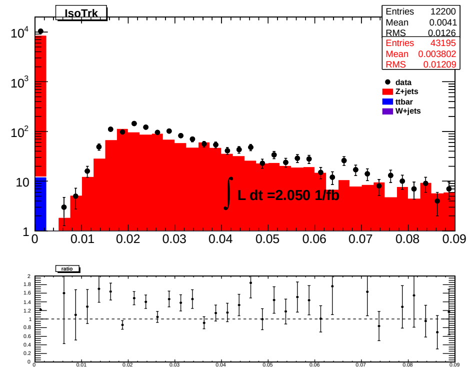
It is possible to notice that, only for the *IsoEcal* variable, there is a big difference between data and Monte Carlo in the first two bins of the distribution. This is a known problem, and is due to a non optimal treatment of the transverse shower development in the `Geant4` version used to simulate these samples. This, however, does not affect the efficiency measurements since they are extracted directly from data (see Section 4.4).



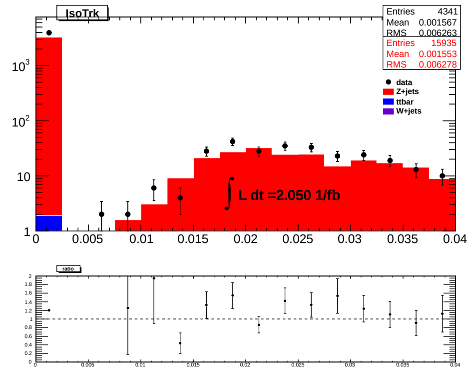
(a) Barrel total distribution plotted in a wider range than WP80.



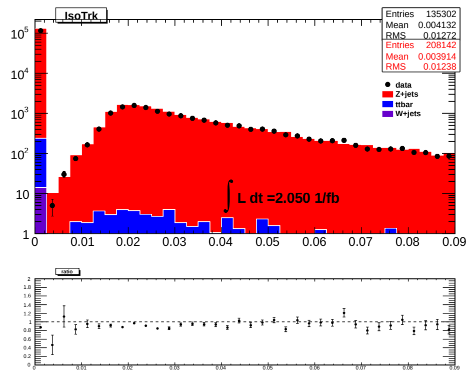
(b) Endcap total distribution.



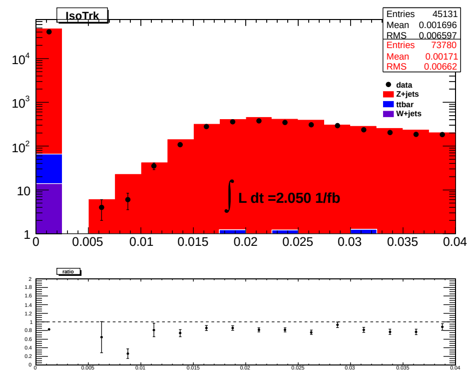
(c) Barrel distribution for events with one measured vertex, plotted in WP80 range.



(d) Endcap distribution for events with one measured vertex.

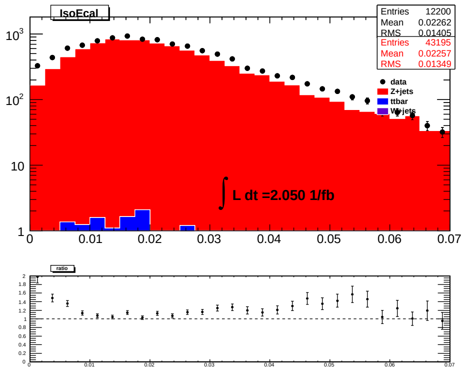


(e) Barrel distribution for events with five measured vertices, plotted in WP80 range.

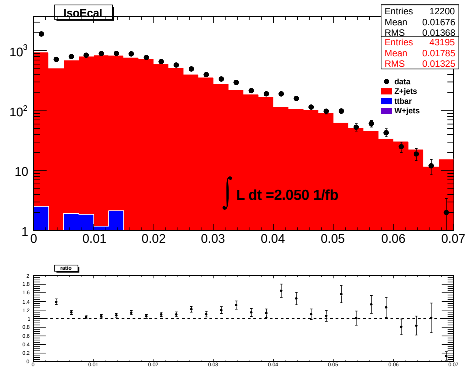


(f) Endcap distribution for events with five measured vertices.

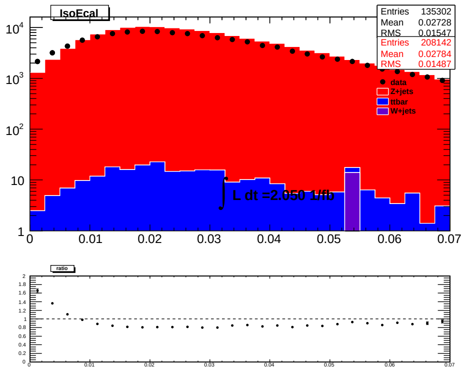
Figure 4.7: *IsoTrk* variable distributions; on the left column for the Barrel and on the right for the Endcap. The Monte Carlo is normalized to the integrated luminosity. Red area is the Z+jets (main contribution), then blue $t\bar{t}$+jets and violet W+jets Monte Carlo samples.



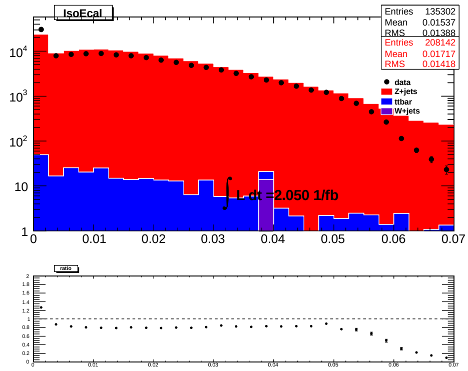
(a) Barrel distribution for events with one measured vertex, plotted in WP80 range.



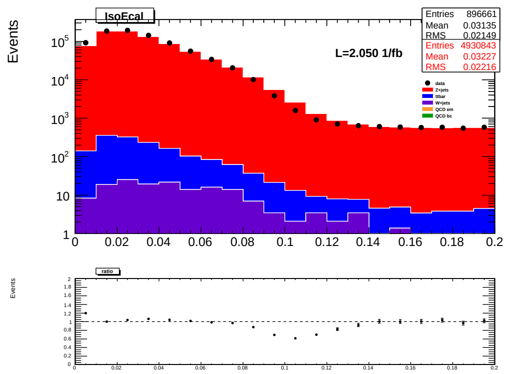
(b) Barrel distribution for events with one measured vertex, with PU-removal corrections activated.



(c) Barrel distribution for events with five measured vertices.

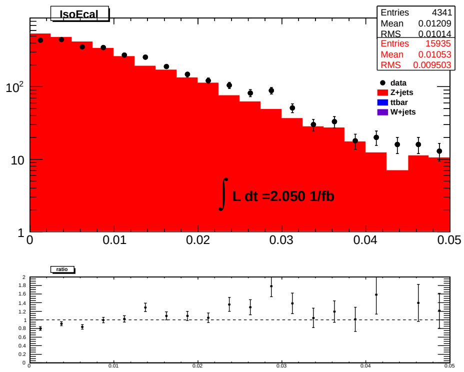


(d) Barrel distribution for events with five measured vertices.

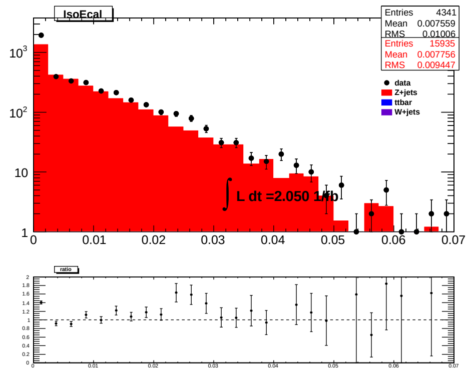


(e) Barrel total distribution plotted in a wider range than WP80.

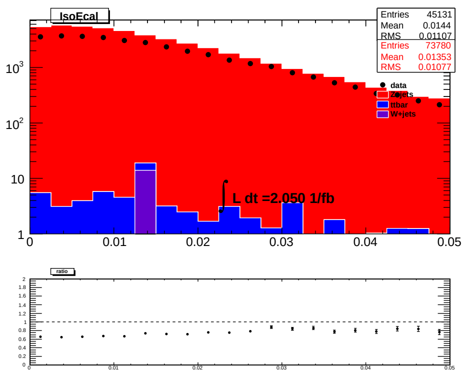
Figure 4.8: *IsoEcal* distributions for Ecal Barrel, with and without PU-removal corrections activated. The Monte Carlo is normalized to the integrated luminosity. Red area is the Z+jets (main contribution), then blue $t\bar{t}$+jets and violet W+jets Monte Carlo samples.



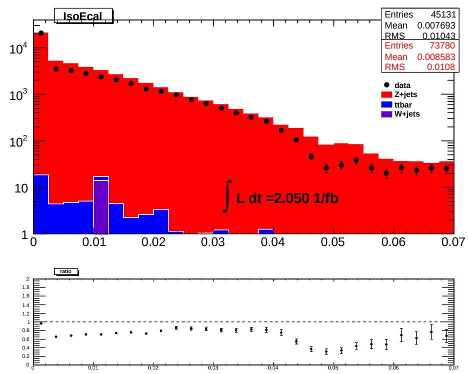
(a) Endcap distribution for events with one measured vertex, plotted in WP80 range.



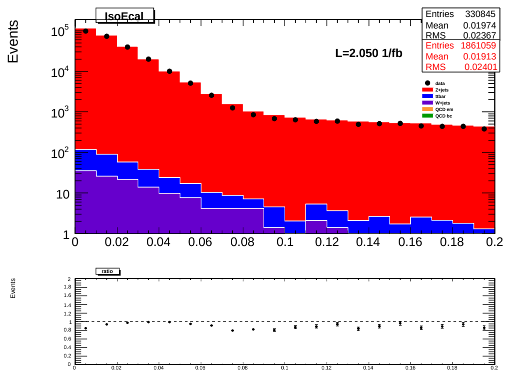
(b) Endcap distribution for events with one measured vertex, with PU-removal corrections activated.



(c) Endcap distribution for events with five measured vertices.

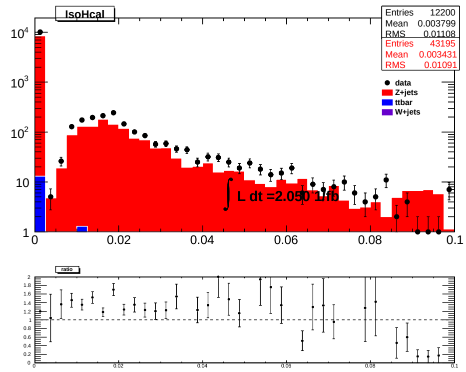


(d) Endcap distribution for events with five measured vertices.

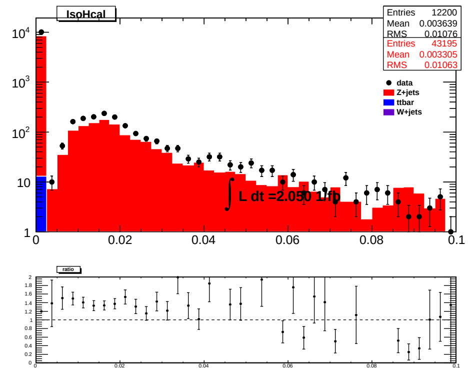


(e) Endcap total distribution plotted in a wider range than WP80.

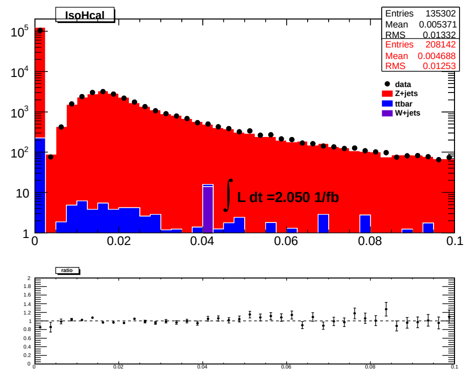
Figure 4.9: *IsoEcal* distributions for Ecal Endcap, with and without PU-removal corrections activated. The Monte Carlo is normalized to the integrated luminosity. Red area is the Z+jets (main contribution), then blue $t\bar{t}$+jets and violet W+jets Monte Carlo samples.



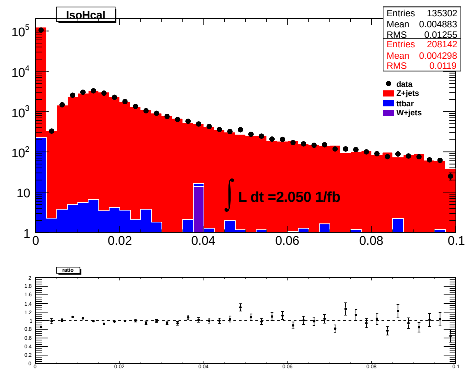
(a) Barrel distribution for events with one measured vertex, plotted in WP80 range.



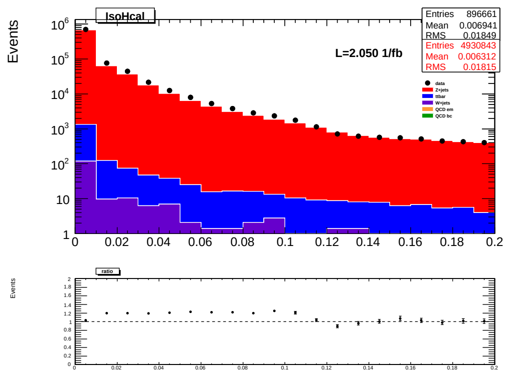
(b) Barrel distribution for events with one measured vertex, with PU-removal corrections activated.



(c) Barrel distribution for events with five measured vertices.

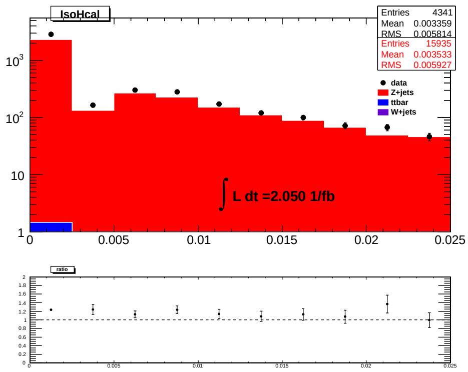


(d) Barrel distribution for events with five measured vertices.

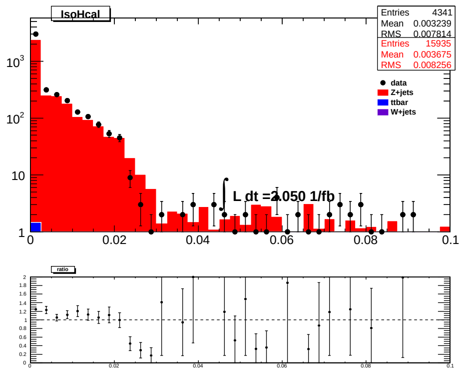


(e) Barrel total distribution plotted in a wider range than WP80.

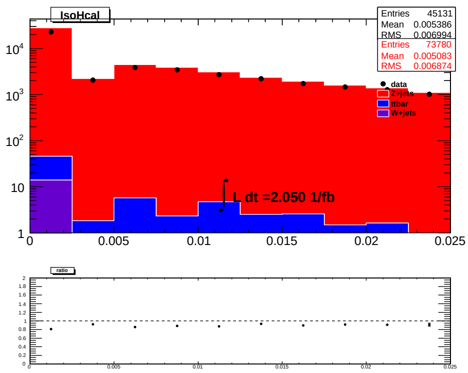
Figure 4.10: *IsoHcal* distributions for Hcal Barrel, with and without PU-removal corrections activated. The Monte Carlo is normalized to the integrated luminosity. Red area is the Z+jets (main contribution), then blue $t\bar{t}$+jets and violet W+jets Monte Carlo samples.



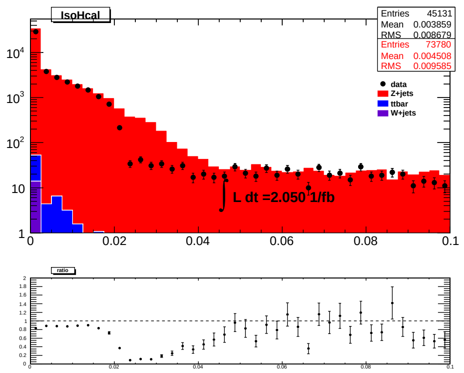
(a) Endcap distribution for events with one measured vertex, plotted in WP80 range.



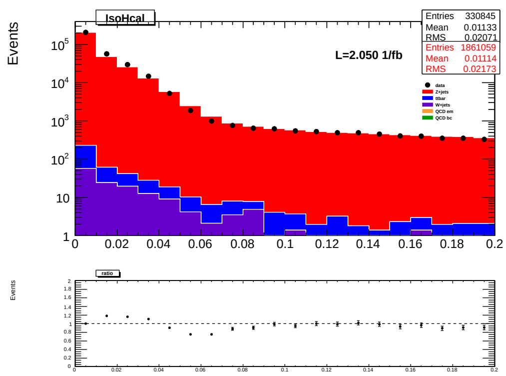
(b) Endcap distribution for events with one measured vertex, with PU-removal corrections activated.



(c) Endcap distribution for events with five measured vertices.



(d) Endcap distribution for events with five measured vertices.



(e) Endcap total distribution plotted in a wider range than WP80.

Figure 4.11: *IsoHcal* distributions for Hcal Endcap, with and without PU-removal corrections activated. The Monte Carlo is normalized to the integrated luminosity. Red area is the Z+jets (main contribution), then blue $t\bar{t}$+jets and violet W+jets Monte Carlo samples.

4.2.4 Jets

In this analysis, the jet reconstruction is performed with the Particle Flow algorithm (see Section 3.2.1). Some selection cuts are also applied in order to reject fake, badly reconstructed jets:

- jet transverse momentum $p_t > 30$ GeV
- pseudo-rapidity $|\eta| < 2.4$
- the jets have to be separated from the electrons produced by the Z decay more than $\Delta R > 0.3$ (with $\Delta R = \sqrt{(\eta_{jet} - \eta_e)^2 + (\phi_{jet} - \phi_e)^2}$ is the electron isolation cone)

Checks on rejected jets

A study has been carried out on rejected jets, to characterize the rejected jets because closer than $\Delta R < 0.3$ to the electron candidate.
When the procedure of jet clustering (Section 3.2.3) is performed on *PFCandidates* (Section 3.2.1), the algorithm clusters all the objects that may form a jet. In this collection of jets there are also the electrons from the Z decay clustered with the surrounding soft particles (coming from Underlying Event and Minimum Bias interactions). It is important to make sure to have removed these electrons from the count of jets. The required condition is that the jet have to be separated from the electron candidate more than $\Delta R > 0.3$. This study looks for the particle content of the rejected jets. In Figure 4.12 the multiplicity of the constituents, their fraction and the percentage of their energy as a function of the transverse momentum and the pseudorapidity of the jets are shown. The contributions within the isolation cone are mainly constituted by: the charged and neutral hadrons (red and blue respectively), the photons (cyan), the electrons (magenta) and finally the muons (orange - not visible).
In Table 4.5 the mean values of the distributions are summarised.

	charged hadrons	**neutral hadrons**	**photon**	***electron***	**muon**
Multiplicity (mean)	5.42	0.80	2.54	**0.98**	1.7×10^{-3}
Multiplicity Fraction [%]	50.5	8.27	24.32	14.82	4.04×10^{-2}
Energy Fraction [%]	8.2	1.76	4.22	**85.72**	9.21×10^{-3}

Table 4.5: Summary of the mean values of multiplicity, multiplicity fraction and energy fraction for different tipes of particles inside the isolation cone.

From the plots, and more clearly from the table, it is possible to see that the rejected jets contain almost always one electron. This electron carries approximately the 80% of the total energy of the jet. These observations support the hypotesis that the selections are correctly identifying isolated electrons –surrounded by particles from low scale underlying event and pile-up activity.

4.2.5 Jet Energy Corrections

In Figure 3.5 some distributions showing the effects of the different corrections used in the analysis are presented. The L1 correction subtracting the pileup effect decreases the jet energy, putting a fraction of jets below threshold, and resulting in a lowered multiplicity; this is clearly visible in all distributions, comparing red and green dots. It is possible to see that the L2 correction (red and yellow dots) change the pseudorapidity distribution but not the transverse momentum one; the opposite behaviour is observed for the L3 correction (cyan and yellow dots). The residual (violet and cyan dots) correction is needed especially for residual calibration of the relative energy scale. The total effect of all the Jet Energy Corrections is visible by comparing the red and the blue distributions. The effect of the JECs on the number of jets passing the p_t threshold is shown in Table 4.6.

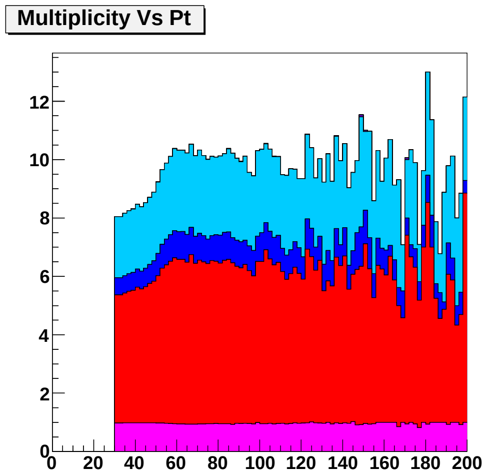
correction	%
L1	-27%
L2	+0,8%
L2+L3	+14%
L2L3Residuals	+18%
all JECs	-17%

Table 4.6: Percentage of jets passing the p_t threshold for different Jet Energy Corrections with respect to the uncorrected case.

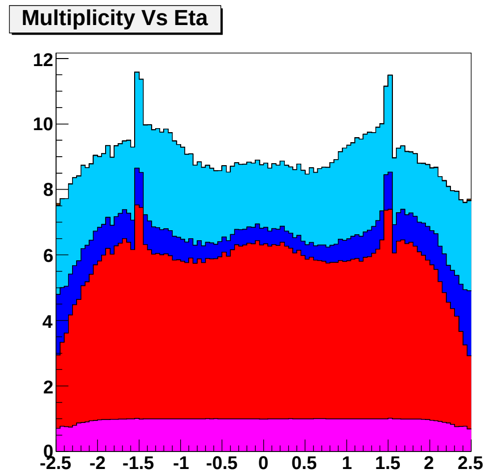
4.3 Benchmark distributions

One of the interests of this analysis is related to its role as a benchmark for Monte Carlo simulations. The Monte Carlo used in this analysis is a `MadGraph` (LO) + `Pythia` (PS) and well reproduces the inclusive distributions. In Figure 4.13 and Figure 4.14 some inclusive distributions are reported.
Black dots represent data, while coloured area shows the MC simulation: the red area is the signal Z+jets (`DYJetsToLL`), the blue area the $t\bar{t}$+jets (`TTJets`), the purple one the W+jets (`WJetsToLNu`). The QCD backgrounds are coloured in orange and green, but since the QCD contribution is negligible in the analysis (in fact no QCD event passes the selection cuts), they are not visible in the plots. The MC distributions are normalized on data luminosity.

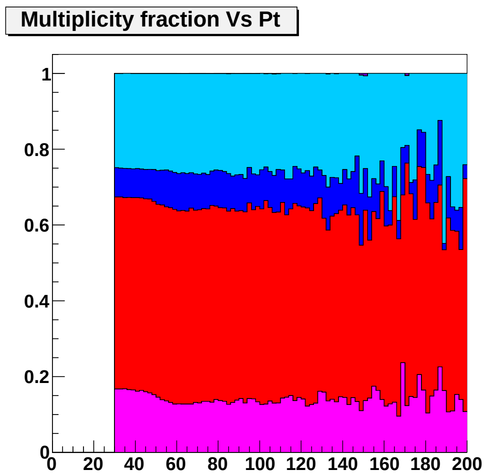
In Figure 4.13a the Z boson invariant mass is shown. The ratio plot shows that there is a shift between the data and MC distributions. In fact the mean values indicate a difference of about 250 MeV, which is going to be addressed in a coming reprocessing of the data.
In Figure 4.13b and Figure 4.13c the η and rapidity of the Z boson are shown. The agreement between data and MC is excellent.
Figure 4.13e shows the inclusive distribution of the p_t of the jets. In Figure 4.13f the p_t of the leading jet (the jet with highest p_t in the event) is reported. In



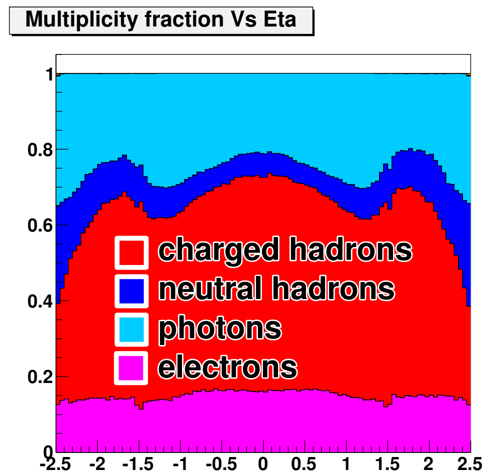
(a) Multiplicity versus the momentum of the particles found inside the isolation cone.



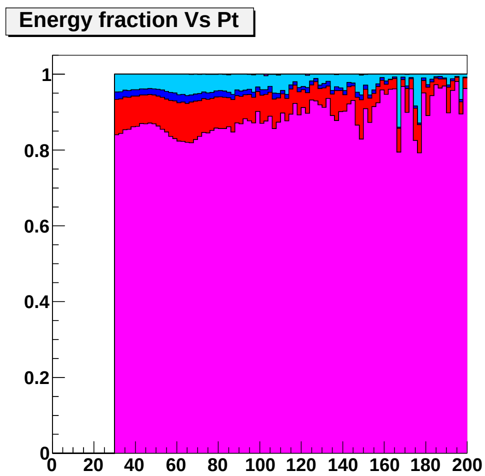
(b) Multiplicity versus η of the particles found inside the isolation cone.



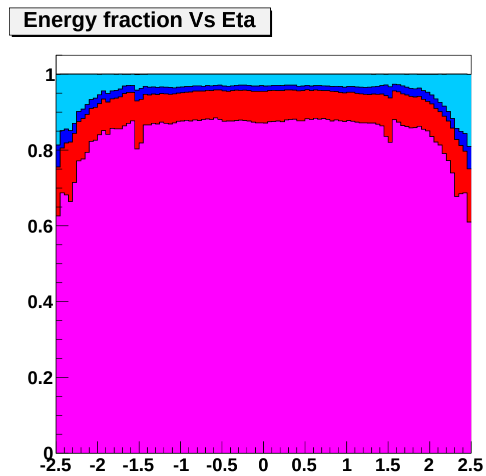
(c) Fraction of the constituents of the jet versus their momentum.



(d) Fraction of the constituents of the jet versus their η coordinate.



(e) Energy fraction of the constituents of the jet versus their momentum.



(f) Energy fraction of the constituents of the jet versus their η coordinate.

Figure 4.12: Electron isolation cone population studies.

these two plots it is possible to notice a small problem on normalization: the MC simulation seems to simulate less jets than produced in data; anyway the shape is well reproduced.
Figure 4.14a shows the Z yield and corroborates the hypothesis of a lower number of associated jets in MC: while the point at zero jets has ratio close to 1, the bins with one and two jets have less events in the simulation than in data. This distribution is also shown for events with one vertex and no particular difference is highlighted.

It is also possible to study exclusive distributions, cutting on the number of vertices of the event. This can provide more insight on the simulation details.

4.4 Efficiency

In order to calculate the cross-section for Z+jets events as a function of the jet multiplicity, the event yields must be corrected for the measurement efficiency.
The inclusive cross-section can be defined as:

$$\sigma = \frac{N}{A\epsilon L}$$

where N is the number of "signal" events, A is the acceptance of the detector, ϵ is the reconstruction and selection efficiency and L is the integrated luminosity (relative to analyzed data). The integrated luminosity is peculiar of the dataset used, while the other parameters are related to the specific analysis.

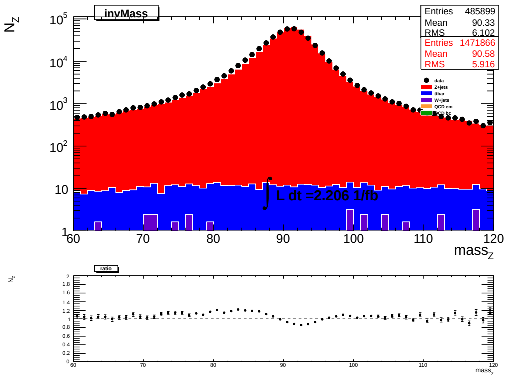
The reconstruction and selection efficiency, ϵ, can be factorized in different contributions, for each step of the selection: the double electron trigger efficiency ϵ_{HLT}, the event reconstruction efficiency ϵ_{RECO}, the selection cuts (WP80) efficiency ϵ_{WP80}. The technique used to perform these measurements is data driven, and it is very similar among the three factors. In Section 4.4.1 the efficiency measurement method is explained, and in Section 4.4.2 the results of efficiency for WP80 cuts, ϵ_{WP80}, are shown.

4.4.1 Tag and Probe method

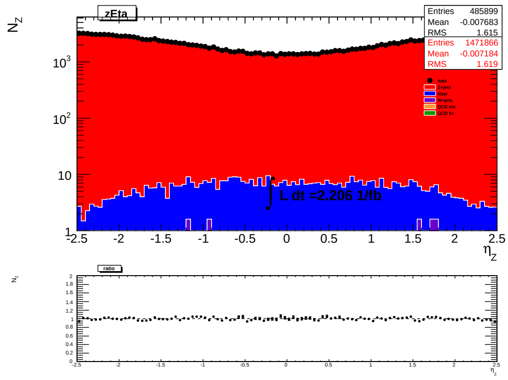
The method used to perform the efficiency measurement is data driven and is called "Tag & Probe" (T&P) [45]. In general, the selection efficiency is given by the ratio between the "signal" events that pass the selection cuts and the total number of signal events:

$$\epsilon = \frac{N_{pass}}{N_{pass} + N_{fail}} \ . \tag{4.4.1}$$

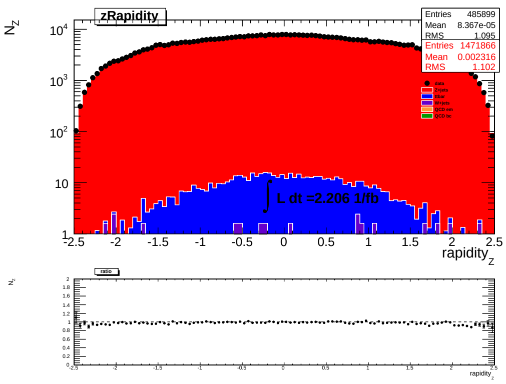
The selected signal events are the pair of electrons coming from Z decay. To perform the measurement, the electron with highest p_t and tight selection cuts is selected and labelled as "tag" electron. The sample of events containing a tag electron are mainly events where a Z boson is produced, but still some background is present. In this sample the second electron in p_t ranking is picked out, and is called "probe". The events are then divided in two samples, whether the probe passes the selection criteria or not. Both samples still contain some fraction of background events to be evaluated with a fit procedure described in the following.



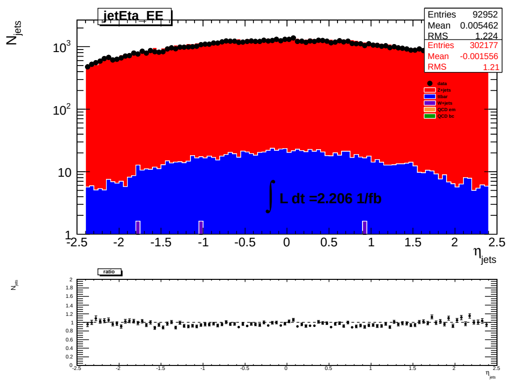
(a) Z boson invariant mass.



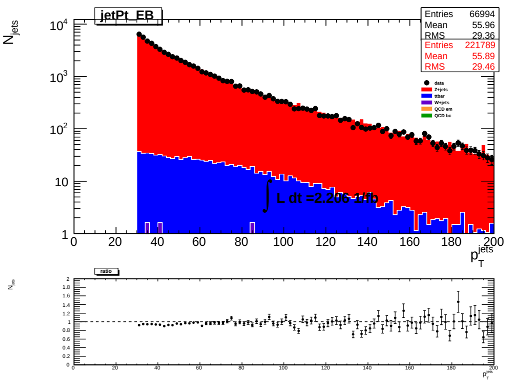
(b) Distribution of the measured η coordinate of the Z boson.



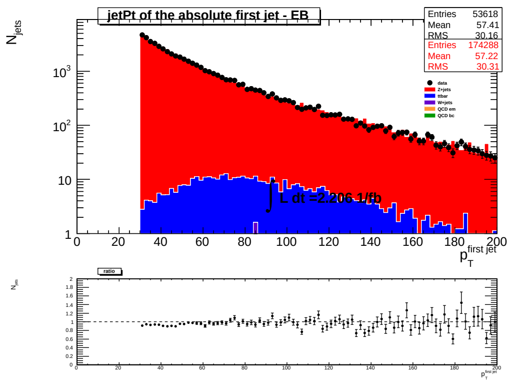
(c) Z boson rapidity distribution.



(d) Distribution of the measured η coordinate of the jets.

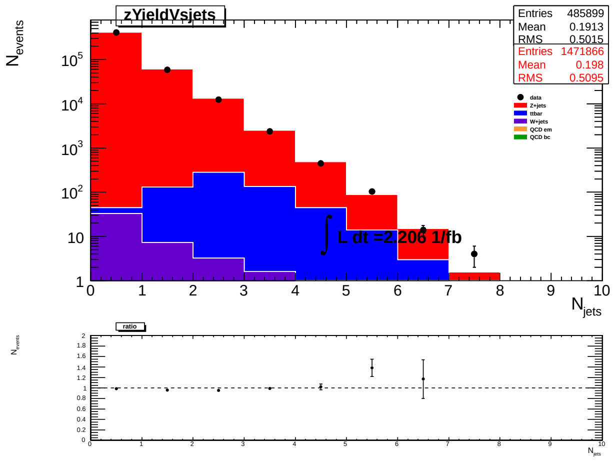


(e) Jets' p_t distribution.

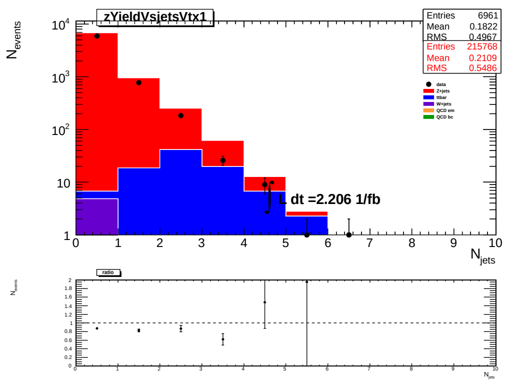


(f) Leading jets' p_t distribution.

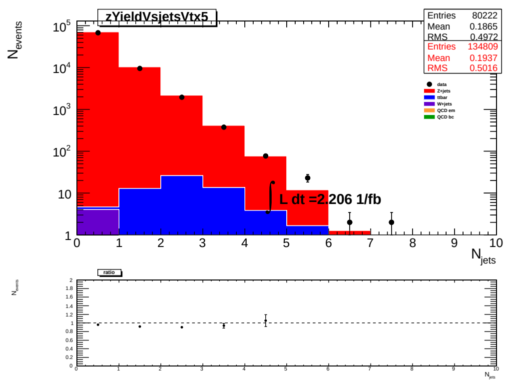
Figure 4.13: Benchmark distributions. The Monte Carlo is normalized to the integrated luminosity.



(a) Z yield (no cuts on event vertices). The Monte Carlo is normalized to the integrated luminosity.



(b) Z yield fot events with one measured vertex.



(c) Z yield fot events with five measured vertices.

Figure 4.14: Benchmark distributions. The Monte Carlo is normalized to the integrated luminosity.

The Tag&Probe efficiency will be the product of the efficiency for the first leg, the tag, times the efficiency of the second leg, the probe, $\epsilon^{TP} = \epsilon^{tag}_{lead} \cdot \epsilon^{probe}_{second}$.

Signal extraction

The fitting function is the sum of the two contributions: the signal is modeled with a Breit-Wigner function, describing the Z resonance line shape, and a Crystal Ball one describing the detector resolution, while the background is modeled with a negative exponential distribution. The Crystal ball function is a Gaussian with a tail on the low side that is traditionally used to describe the effect of radiative energy loss in an invariant mass.
The function to describe the Z boson invariant mass peak, $S(x)$, is then:

$$
\begin{aligned}
S(x) &= \frac{1}{(x-m)^2+\frac{1}{4}\Gamma^2} && \text{Breit-Wigner} \\
\otimes &\begin{cases} \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right) & \text{for } x > |a| \\ \frac{\left(\frac{n}{|a|}\right)^n \exp^{-\frac{1}{2}a^2}}{\left(\frac{n}{|a|}-|a|-x\right)^n} & \text{for } x \leq |a| \end{cases} && \text{Crystal Ball}
\end{aligned}
$$

where Γ is the decay rate, σ is the sigma of the Gaussian in the CB, a and n are two parameters that control where the CB stops to behave as a Gaussian; all the parameters are left free. The parametrization used for the background is a negative exponential distribution, $B(x) = \exp(-x)$.

From the normalization values obtained with the fitting procedure it is possible to calculate the selection efficiency as defined in equation 4.4.1.

Fitting procedure

The easiest way to obtain the number of events passing (or not) the desired selection is to directly fit the invariant mass distribution of selected probe events, and to fit the distribution of failing probes separately.
Both distributions are fitted with a binned extended maximum likelihood (where the total number of events enters as a parameter in the fit). The probability density functions (*pdf*) are defined as follows:

$$
\begin{aligned}
pdf^{(pass)}(x) &= N^{(pass)}_{tot}\left[\eta \times S^{(pass)}(x) + (1-\eta) \times B^{(pass)}(x)\right] \\
pdf^{(fail)}(x) &= N^{(fail)}_{tot}\left[\eta \times S^{(fail)}(x) + (1-\eta) \times B^{(fail)}(x)\right]
\end{aligned}
$$

where η is the fraction of events of signal with respect to the total number of events in the distribution[1]. Then the number of signal events is: $N^{(pass)}_{signal} = \eta \times N^{(pass)}_{tot}$ (this applies to both the distributions, probe pass or probe fail).

This procedure can be modified to directly extract the efficiency on a single fit over the tag distribution (regardless the probe passed or not the selection criteria). The efficiency is plugged as a parameter in the Likelihood, obtaining direcly from

[1]Notice that η here must not be confused with the “pseudorapidity” kinematical variable introduced in Section 1.1.2.

the minimization procedure the optimal value for the efficiency together with its statistical uncertainty. The *pdf* is the defined as:

$$pdf(x) = N_{tot}\left[a \times S^{(pass)}(x) + b \times S^{(fail)}(x) + c \times B^{(pass)}(x) + d \times B^{(fail)}(x)\right]$$

where the parameters a, b, c and d are defined as:

$$\begin{aligned} a &= \epsilon \times \eta \\ b &= (1-\epsilon) \times \eta \\ c &= \epsilon' \times (1-\eta) \\ d &= (1-\epsilon') \times (1-\eta) \end{aligned}$$

where ϵ' is the efficiency of the selection criteria for the background, and ϵ is the efficiency of the selection criteria for the signal (i.e. the electrons from Z boson decay). The interesting parameter is obviously the efficiency ϵ. In Figure 4.15 an example of fit in the mass range 60-120 GeV over a sample of data collected during 2011 is shown.

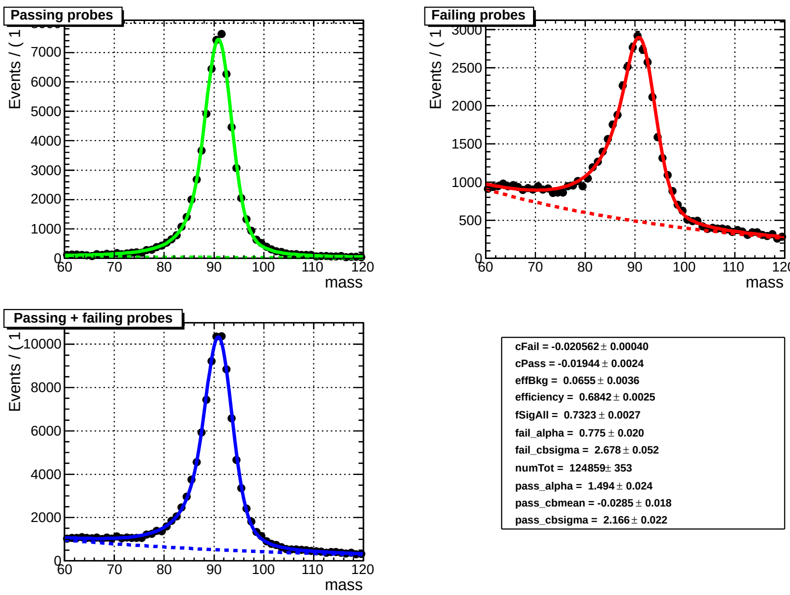
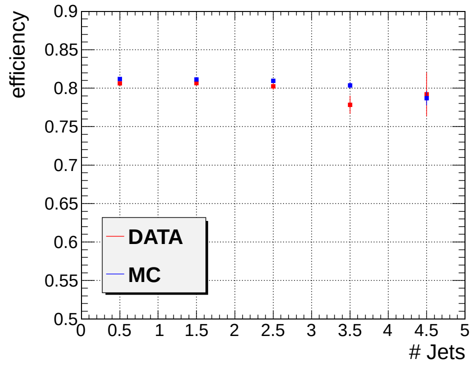


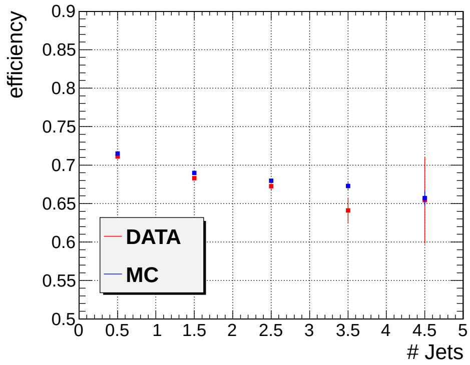
Figure 4.15: Fits over the invariant mass of the two selected electrons. This Figure shows: in the the top left plot the distribution with the passing probes (green); in the top right the distribution with the failing probes (red) and the bottom left plot shows the total distribution (blue). Dashed line shows the background. In the bottom right panel the fit values are reported.

4.4.2 WP80 selection efficiency

By applying the T&P method to the WP80 selection and identification cuts, it is possible to calculate the efficiency for the tag end the probe electron separately,



(a) Tag efficiency versus number of jets in the event.



(b) Probe efficiency versus number of jets in the event.

Figure 4.16: Efficiency for tag and probe electrons plotted versus the number of jets in the event.

both for the data and the Monte Carlo. The efficiency is measured as a function of the number of jets (with a maximum of four jets), this is necessary to separately measure the cross section for the exclusive processes $Z + 0$ jets, $Z + 1$ jet, etc.
The results for the WP80 efficiency measurements are reported in Figure 4.16. The panel (a) shows that the tag has an efficiency around $\epsilon^{tag}_{WP80} \sim 80\%$ from 0 to 4 jets, and that the MC has a slightly higher efficiency than data. The panel (b) shows the efficiencies obtained for the probe electron; as expected, ϵ^{probe}_{WP80} is smaller than ϵ^{tag}_{WP80} and is varying with the increasing number of jets from about 71% to 65%.

4.5 Unfolding procedure

In order to reach a precision measurement, a fundamental step is to correct the data distributions for the smearing effects introduced by the detector; with this correction it is possible to compare the distributions at the theoretical level. This is done using well known statistical tools, making use of different techniques to perform the unfolding. In this analysis, two robust mathematical algorithms are firstly validated and then applied:

1. Singular Value Decomposition (SVD) of the response matrix [46], based on a linear regularization algorithm through a parameter k_{SVD},

2. Bayesian method, based on the Bayes theorem as proposed by D'Agostini [47].

By performing both methods, it is possible to obtain an independent check of the physics objects unfolding. The two methods are implemented by the `RooUnfold` package [48]. The basic idea behind unfolding is to build a response matrix obtained at the generator level from the Monte Carlo simulation (thus unaffected by detector effects) and the reconstructed Monte Carlo simulation (which is affected by detector smearing effect). Then, the matrix is applied to data, obtaining in this way the unfolded data distributions. For our studies, the main observables that we want to correct –in order to be able to have a comparison with the NNLO theoretical prediction– are:

- the transverse momentum of the Z boson, p_t^Z
- the rapidity of the Z boson, y^Z
- the Z+jets multiplicity

All these distributions, prior unfolding, are not yet corrected by the efficiency and acceptance (see Section 4.4). One of the key points arising in the procedure is the selection of the best value for the regularization parameter k, both for the SVD and the Bayesian algorithm. The unfolding is done by varying this parameter, and is different for each distribution:

$$k_{SVD,Bayes} \in [2, n_{bins}]$$

where n_{bins} is the total number of bins in which is binned the particular distribution. The optimal value is selected by finding the largest value up to which the errors remain small and stable: small values will bias the unfolding result towards the MC truth input, while a too-large value will give a result that is dominated by unphysically enhanced statistical fluctuations [46].

4.5.1 Building the response matrix

The response matrix used to perform the unfolding is built up by the following criteria on the MC truth-reco distribution:

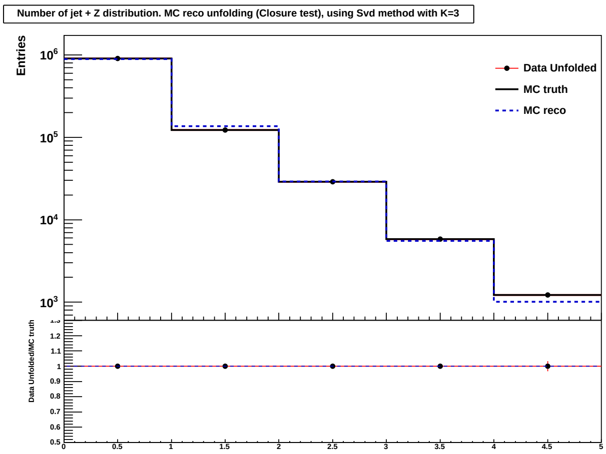
- jets are reconstructed by the anti-k_T algorithm with $\Delta R = 0.5$
- pile-up effect is subtracted, as described in Section 3.2.4
- transverse momentum of jets $p_t > 30$ GeV
- jets closer than $\Delta R = 0.3$ to the electron are removed
- electrons from Z decay has $p_t > 25$ GeV
- for generated jets is required that electromagnetic and hadronic charged and neutral energy must be < 0.99 and jet multiplicity > 0.

Data unfolding results are presented in Section 4.5.3. Some cross checks to validate the procedure have been made, as shown in the Section 4.5.2.

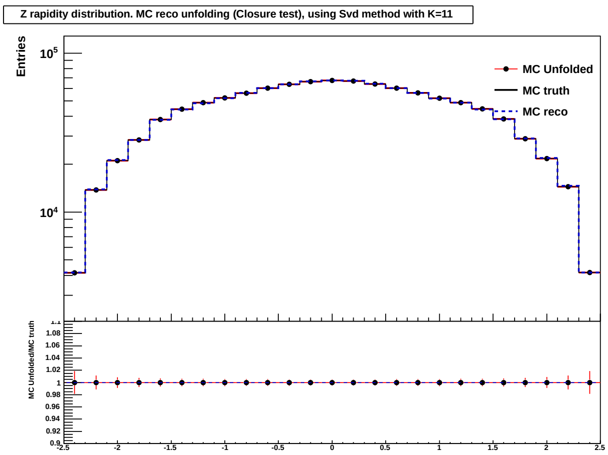
4.5.2 Validation tests

In order to demonstrate that the unfolding algorithm (either SVD or Bayesian) does not introduce any additional bias, three different tests have been carried out:

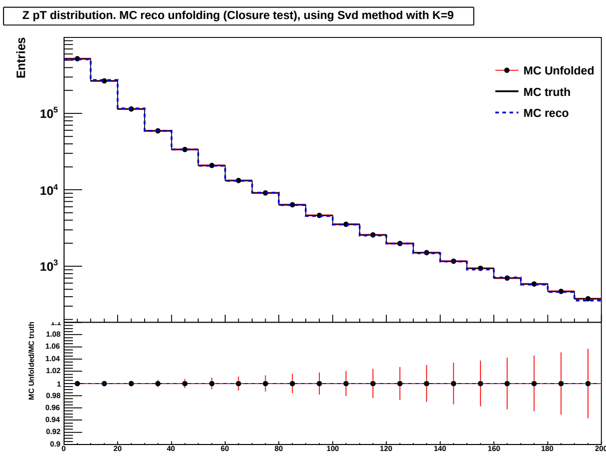
1. the response matrix is applied to the same MC used to build the matrix itself. While doing this, we expect to obtain the MC truth distribution again. This test is performed for the Z+jets yield, the rapidity and Z transverse momentum distributions, for different values of k. Some results are shown in Figure 4.17 and Figure 4.18;



(a) Z+jets multiplicity - SVD $k = 3$

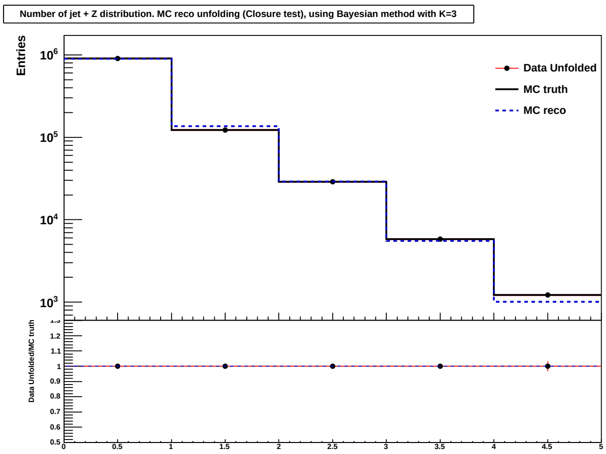


(b) Z rapidity - SVD $k = 11$

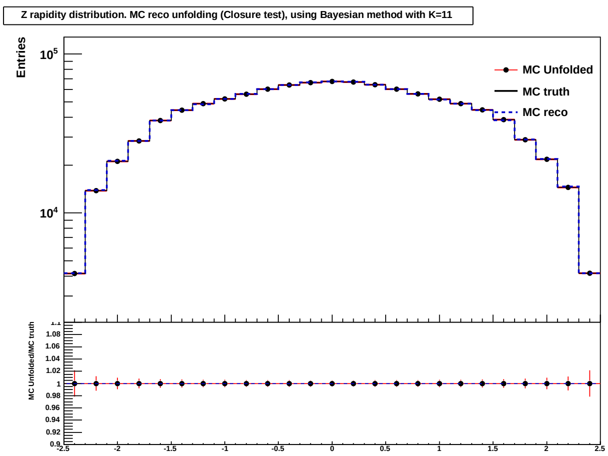


(c) Z p_t - SVD $k = 9$

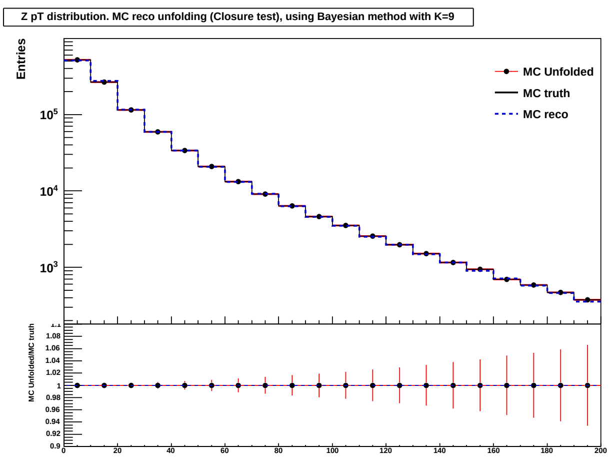
Figure 4.17: (1) Identity test: results with SVD algorithm on Z+jets multiplicity, rapidity and Z transverse momentum distributions with different k.



(a) Z+jets multiplicity - Bayesian $k = 3$

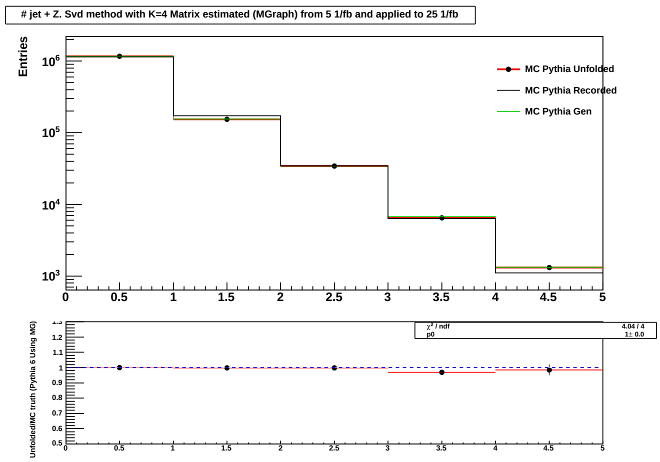


(b) Z rapidity - Bayesian $k = 11$

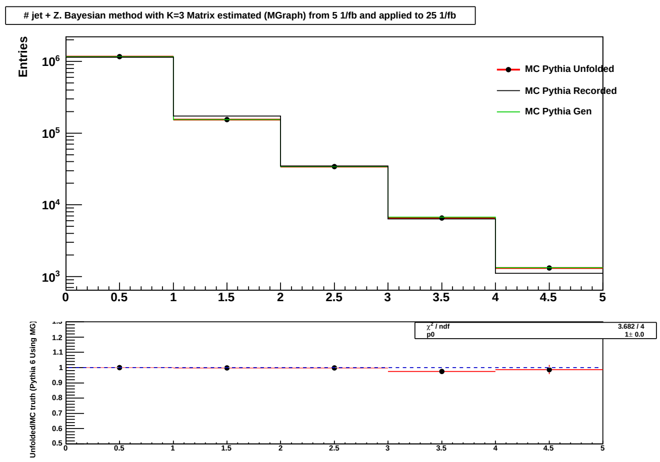


(c) Z p_t - Bayesian $k = 9$

Figure 4.18: (1) Identity test: results with Bayesian algorithm on Z+jets multiplicity, rapidity and Z transverse momentum distributions with different k.



(a) Z+jets multiplicity - SVD $k = 4$



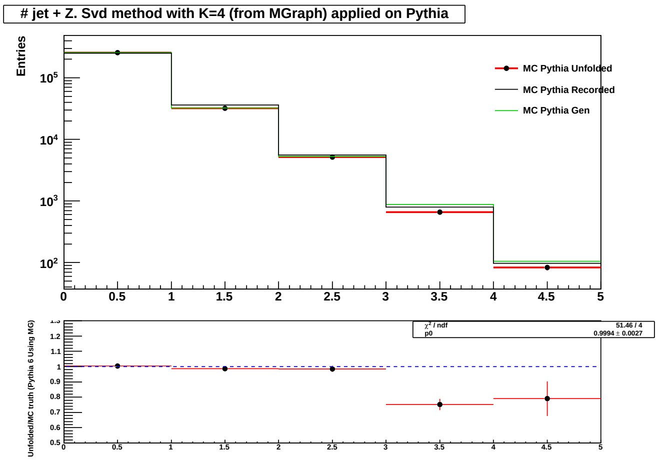
(b) Z+jets multiplicity - Bayesian $k = 3$

Figure 4.19: (2) Splitting test: results with SVD and Bayesian algorithm on Z+jets multiplicity distribution with different k.

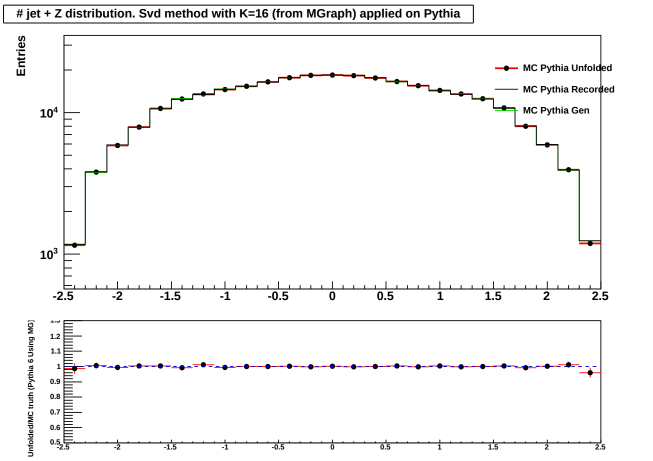
2. the MC sample used in the above test has been split into two subsamples of different size, one (subsample A) having an effective integrated luminosity approximately equal to the 2011 integrated luminosity (5 fb^{-1}) and yhe other one (subsample B) obtained from the remaining part (7 fb^{-1}). The matrix response calculated using subsample A is then applied to unfold the subsample B. In Figure 4.19 it is possible to see the good agreement between the unfolding and the truth MC;

3. the distribution obtained from a different signal Monte Carlo (`Pythia6`) has been unfolded using the response matrix derived with the `MadGraph` sample. This test has produced a very good agreement between unfolded and truth distribution, showing no explicit dependence on the type of event generator used to define the unfolding matrix (see Figure 4.20 and Figure 4.21).

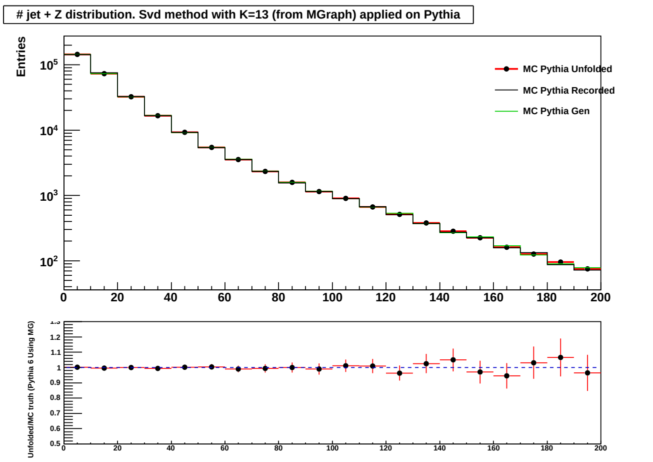
Note that each test has been made by varying the k regularization parameter from 2 to the maximum number of bins. All tests are computed for both the Bayesian



(a) Z+jets multiplicity - SVD $k = 4$

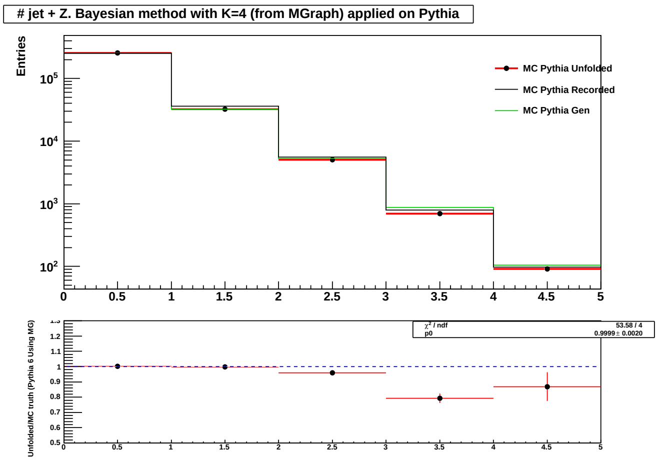


(b) Z rapidity - SVD $k = 16$

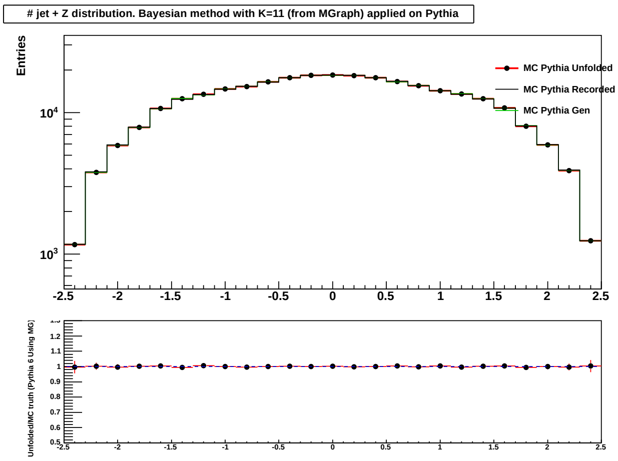


(c) Z p_t - SVD $k = 13$

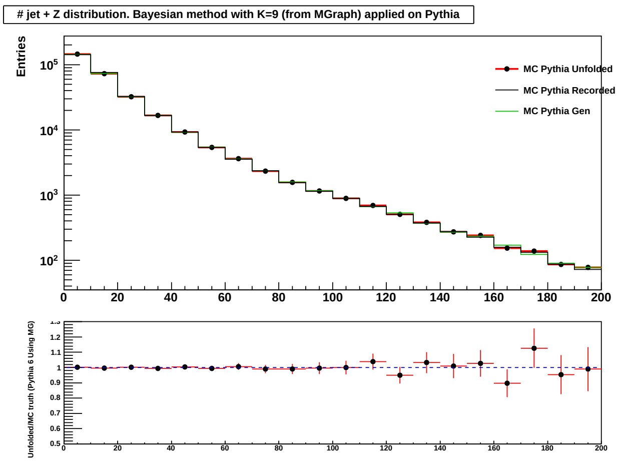
Figure 4.20: (3) Test with a different MC: results with SVD algorithm on Z+jets multiplicity, rapidity and Z transverse momentum distributions with different k.



(a) Z+jets multiplicity - Bayesian $k = 4$



(b) Z rapidity - Bayesian $k = 11$



(c) Z p_t - Bayesian $k = 9$

Figure 4.21: (3) Test with a different MC: results with Bayesian algorithm on Z+jets multiplicity, rapidity and Z transverse momentum distributions with different k.

and the SVD algorithms, so it is possible to additionally test the consistency of the different approaches. All performed tests show that the unfolding procedure is consistent with the expectations, and that the two statistically different approaches lead to compatible results.

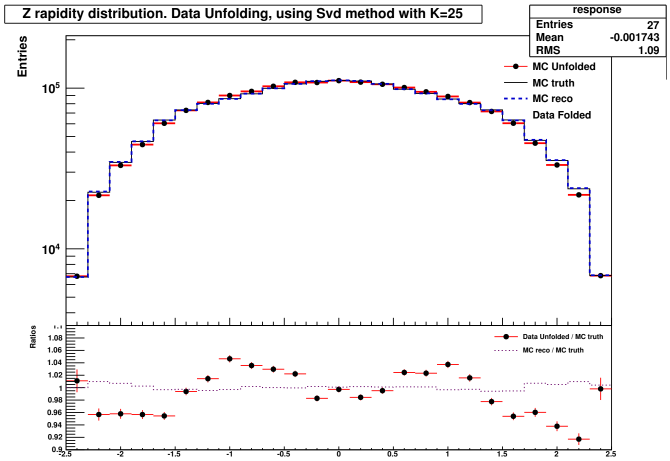
4.5.3 Unfold real data

As a last result of this analysis, the unfolding procedure has been applied to data. For this purpose, we choose the best value for the regularization parameter k:

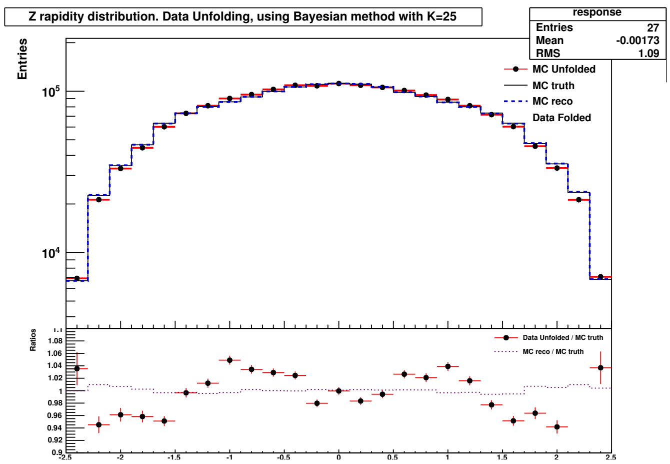
$$\begin{aligned} p_t^Z &: \; k = 20 \\ y^Z &: \; k = 25 \\ Z \text{ yield} &: \; k = 5 \end{aligned}$$

For each observable, the unfolding made with the SVD and the Bayesian methods is reported, together with its response matrix.
It is possible to see that the difference between the two MC distributions, at generator and recontructed level, is around 2% for the distributions in Figure 4.22 and in Figure 4.23, indicating that the effect of the detector is very small (dashed line in ratio plot). The unfolded data are then not much affected by this procedure. The ratio of unfolded data to MC truth then mostly reflects the differences between data and MC simulation themselves, more than the effect of the unfolding.
The effect of unfolding is that small because the energy resolution on the p_t of the Z boson is around 1-2 GeV, while the binning is made every 10 GeV making the migration between bins unlikely. (Same thing applies for the Z rapidity).

Figure 4.24 is more interesting, since it is possile to see a real effect of the detector of the order of 10% for the second bin (i.e. Z+1 jet) and Z+3 jets, while it is even larger than 20% for Z+4 jets events. The ratio unfolded data versus the MC distribution at generator level shows a difference lower that 10% for events with up to 3 jets. It is important to remember that these results are not yet corrected for efficiency and acceptance effects.

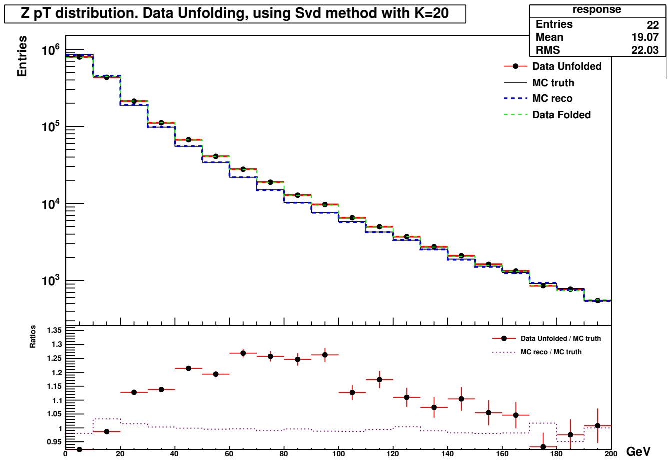


(a) Z rapidity - SVD $k = 12$

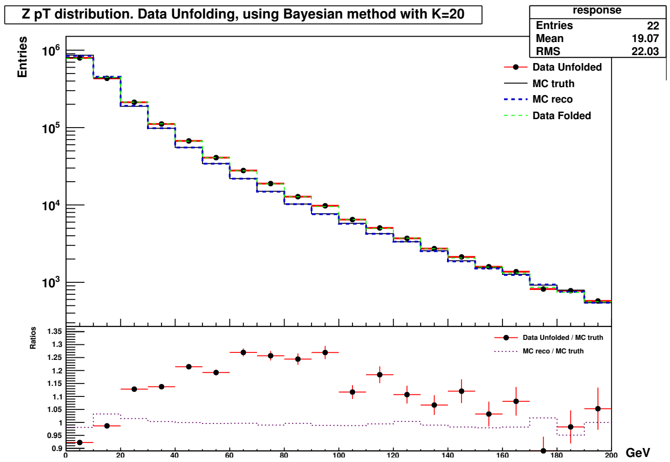


(b) Z rapidity - Bayesian $k = 12$

Figure 4.22: Unfolding procedure on data using SVD and Bayesian algorithms, for Z rapidity distribution.

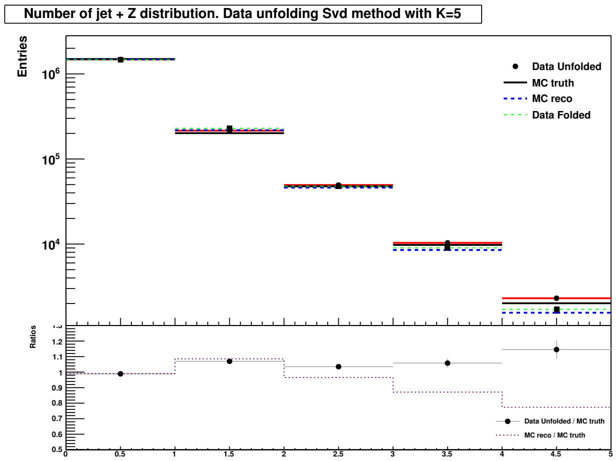


(a) Z p_t - SVD $k = 10$

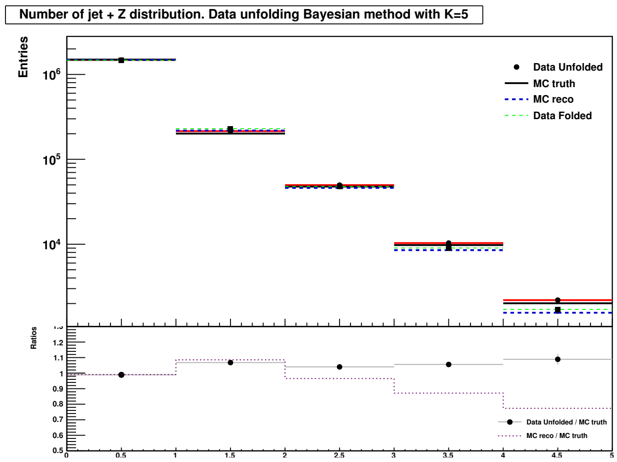


(b) Z p_t - Bayesian $k = 10$

Figure 4.23: Unfolding procedure on data using SVD and Bayesian algorithms, for p_t^Z multiplicity distribution.



(a) Z+jets multiplicity - SVD $k = 5$



(b) Z+jets multiplicity - Bayesian $k = 5$

Figure 4.24: Unfolding procedure on data using SVD and Bayesian algorithms, for Z+jets multiplicity distribution.

Conclusions

In this study all the different elements of the Z+jets analysis are discussed. The data sample considered was collected during the first half of 2011 with a center-of-mass energy of 7 TeV, corresponding to an integrated luminosity of about 2.2 fb^{-1}. The detailed study of trigger and pile-up conditions during 2011 data taking made accessible the use of 2011 data to extend the 2010 analysis [49] to much higher statistics. The results of this study are briefly summarized as follows.

The performed tests demonstrate that the pile-up removal procedure is well understood, and that is possible to perform the events selection realiably. The consistency of the Monte Carlo simulation with data was checked on some inclusive distributions, and the agreement is at $\pm 5\%$ level. The Tag&Probe method was implemented and WP80 selection efficiencies were measured on data for events with different number of jets, finalizing the first step in overall efficiency measurement. Finally, the unfolding procedure removing the detector induced effects is implemented and tested. The closure tests performed on the unfolding procedure has given some insight on the procedure mechanism itself, and behave as expected. Finally, the Z yield distribution is unfolded and ready to be finalized with acceptance and efficiency measurements and compared with theoretical predictions (e.g., Z boson cross section, Berends-Giele scaling, etc) and Monte Carlo models.

This study allows to develop a precision analysis for the Z+jets process, providing deep understanding in several areas of the CMS experiment and physics models, from detector calibration, to Monte Carlo tuning and perturbative QCD tests.

Bibliography

[1] L. Evans and P. Bryant. LHC Machine. *JINST*, 3:S08001, 2008.

[2] CERN Hadron Linacs. http://linac2.home.cern.ch/linac2/default.htm.

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

[4] K. H. P. Wolfgang. The evolution of particle accelerators & colliders, Spring 1997.

[5] The TEVNPH Working Group. Combined CDF and D0 Upper Limits on Standard Model Higgs-Boson Production with up to 6.7 fb^{-1} of Data. 2010.

[6] CMS Collaboration. Combination of CMS searches for a Standard Model Higgs boson. (CMS PAS HIG-11-032), 2011.

[7] CMS Collaboration. Search for Resonances in the Dilepton Mass Distribution in pp Collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 1105:093, 2011.

[8] CMS Collaboration. Search for signatures of extra dimensions in the diphoton mass spectrum at the Large Hadron Collider. (CMS-EXO-11-038, CERN-PH-EP-2011-173), 2011.

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

[10] CMS Collaboration. CMS, Tracker Technical Design Report. (CERN-LHCC-98-006), 1998.

[11] CMS Collaboration. The CMS Tracker: addendum to the Technical Design Report. (CERN-LHCC-2000-016), 2000.

[12] P. Azzurri. The CMS Silicon Strip Tracker. *Journal of Physics: Conference Series. JINST*, 41(1):127, 2006.

[13] E. Widl, R. Frühwirth, and W. Adam. A Kalman Filter for Track-based Alignment. (CMS-NOTE-2006-022), 2006.

[14] S. Cucciarelli. Track and Vertex Reconstruction with the CMS Detector at LHC. (CMS-CR-2005-021), 2005.

[15] CMS Collaboration. Performance and operation of the CMS electromagnetic calorimeter. *Journal of Instrumentation*, 5(03):T03010, 2010.

[16] CMS Collaboration. Time Reconstruction and Performance of the CMS Electromagnetic Calorimeter. *J. Instrum.*, 5(arXiv:0911.4044. CMS-CFT-09-006):T03011. 27 p, 2009.

[17] P. Adzic et al. Energy Resolution of the Barrel of the CMS Electromagnetic Calorimeter. (CMS-NOTE-2006-148), 2006.

[18] S. Dasu et al. CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems. (CERN-LHCC-2000-038), 2000.

[19] S. Cittolin, A. Rácz, and P. Sphicas. CMS trigger and data-acquisition project: Technical Design Report. (CERN-LHCC-2002-026), 2002.

[20] F. Halzen and A. D. Martin. *Quarks and leptons: an introductory course in modern particle physics.* John Wiley and Sons, 1984.

[21] R. K. Ellis, W. J. Stirling, and B. R. Webber. *QCD and collider physics.* Cambridge University Press, 2003.

[22] F.A. Berends, H. Kuijf, B. Tausk, and W.T. Giele. On the production of a W and jets at hadron colliders. *Nuclear Physics B*, 357(1):32 – 64, 1991.

[23] L. Cerrito. Measurements of vector boson plus jets at the Tevatron. (FERMILAB-CONF-10-311-E), 2010.

[24] T. Sjostrand. Monte Carlo Tools. (LU TP 09-31, MCnet/09/17), 2009.

[25] K. Nakamura et al. Review of particle physics. *J. Phys.*, G37:075021, 2010 and 2011 partial update for the 2012 edition.

[26] 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*, 49:1099–1116, 2007.

[27] G. L. Bayatian et al. *CMS Physics Technical Design Report Volume I: Detector Performance and Software*, volume CERN-LHCC-2006-001. CERN, 2006.

[28] H. Bethe and W. Heitler. On the Stopping of Fast Particles and on the Creation of Positive Electrons. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, Vol. 146, No. 856:pp. 83–112, Aug. 1, 1934.

[29] W. Adam, R. Frühwirth, A. Strandlie, and T. Todor. Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC. (CMS-NOTE-2005-001), 2005.

[30] P. Meridiani and R. Paramatti. Use of $Z \to e^+e^-$ events for ECAL calibration. (CMS-NOTE-2006-039), 2006.

[31] M. Cacciari, G. P. Salam, and G. Soyez. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, 2008.

[32] R. Frühwirth. Application of Kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2–3):444 – 450, 1987.

[33] CMS Collaboration. Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. (CMS PAS PFT-09-001), 2009.

[34] F. Beaudette, D. Benedetti, P. Janot, and M. Pioppi. Electron reconstruction within the Particle Flow Algorithm. (CMS-AN-2010-034), 2010.

[35] CMS Collaboration. Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV. (CMS PAS PFT-10-002), 2010.

[36] V. Candelise et al. Comparison of GSF and Particle Flow electron reconstruction performances. (CMS-AN-11-474), 2012.

[37] G. P. Salam. Towards Jetography. *Eur.Phys.J.*, C67:637–686, 2010.

[38] M. Cacciari and G. P. Salam. Dispelling the N^3 myth for the k_t jet-finder. *Phys. Lett.*, B641:57–61, 2006.

[39] M. Cacciari, G. P. Salam, and G. Soyez. The Catchment Area of Jets. *JHEP*, 04:005, 2008.

[40] M. Cacciari and G. P. Salam. Pileup subtraction using jet areas. *Phys.Lett.*, B659:119–126, 2008.

[41] F. Maltoni and T. Stelzer. MadEvent: Automatic event generation with MadGraph. *JHEP*, 0302:027, 2003.

[42] T. Sjostrand, S. Mrenna, and P. Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 0605:026, 2006.

[43] R. Field. Early LHC Underlying Event Data - Findings and Surprises. *arxiv:1010.3558*, 2010.

[44] S. Agostinelli et al. GEANT4: A simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003.

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

[46] H. Andreas and K. Vakhtang. 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.

[47] G. D'Agostini. A multidimensional unfolding method based on Bayes' theorem. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 362(2–3):487 – 498, 1995.

[48] T. Adye. Unfolding algorithms and tests using RooUnfold. *Proceedings of the PHYSTAT 2011 Workshop, CERN, Geneva, Switzerland*, (CERN-2011-006), 2011.

[49] CMS Collaboration. Rates of Jets Produced in Association with W and Z Bosons in pp Collisions at $\sqrt{s} = 7\ TeV$. (CMS PAS EWK-10-012), 2010.

Ringraziamenti

Come suggerisce (maledettamente) intelligentemente il titolo del capitolo vorrei ringraziare alcune persone. (chissà cosa verrà fuori...)

Dorme ancora la campagna, forse no,
è sveglia, mi guarda, non so.
Già l'odore della terra odor di grano,
sale adagio verso me.

Voglio sinceramente ringraziare il gruppo CMS Trieste. Giuseppe che con innegabile costanza e passione segue (e precede) tutti i suoi studenti, grazie davvero. Fabio, chiaro e puntuale faro nella fisica, con cui ho avuto la fortuna di dialogare (anche) su altri universi della Vita, grazie (anche per le sonore risate). Damiana per il supporto di varia natura ;) e la sempre estrema disponibilità, daidaidai! Vieri, grazie per aver alleggerito la mia permanenza ...ci siamo divertiti e avremmo sicuramente potuto continuare, hai ragione... Chiara, ci conosciamo da molto poco ma ho grande fiducia [...] grazie per la sopportazione. Matteo, il poeta, persona sensibile, dannatamente paziente... grazie per aver avuto la pazienza di capirmi, di capirci. Mi sono divertito, grazie per troppe cose, spero di potermi sdebitare in qualche modo. Andrea, grazie per tutto quello che mi hai passato ;) mi hai reso la vita mooolto più semplice. Grazie per avere sopportato le cose peggiori di me senza mai inca... inalberarti e per aver condiviso i primi chilometri della mia vita motociclistica, è stata una gran figata... ne abbiamo ancora molti da percorrere assieme (non solo su due ruote).

Desidero ringraziare il professor Paver per l'attenzione, la cortesia e la simpatia che ha saputo dedicarmi (senza contare le 'teoriche emozioni' che mi ha donato nei suoi corsi).

Desidero ringraziare le professoresse Gregorio e Longo per l'ospitalità.
Due persone il cui intervento mi ha permesso di arrivare alla fine: il prof. Milotti per avermi letteralmente soccorso e il prof. Camerini per aver tenuto dei corsi splendidi, che mi hanno distratto con successo da problematiche più terrene.

Ho fatto un sogno lungo la mia strada
una lacrima non mi disseterà
ma almeno i sogni vada come vada
non si piegano a nessuna volontà.
Ho sognato che la cosa più importante
non è chi sei ma quel che vuoi
e la luna sembrava un diamante
e tu dentro un mondo più libero
correvi con noi.

Grazie a mamma Marisa e papà Pierluigi, che mi hanno fatto e insospettabilmente educato. Grazie per la pazienza e la smisurata fiducia. Simone, Chiara, Eva e Marco abbiamo saputo far squadra nei momenti difficili, spero che questa esperienza ci mantenga così vicini. Siamo solo all'inizio. Manuela, come chi ti precede, dovrei ringraziarti per troppe cose. Ti ringrazio perché hai saputo conoscermi, hai saputo non fare domande quando non andavano fatte; sei una micro donna straordinaria.
Grazie a tutti per non avermi mai rotto le scatole per la mia lentezza accademica.
Grazie alla famiglia estesa: Nonna, Zie, Zii, cugini+relativi tutti, senza la quale non sarei quel che sono, è anche colpa vostra quindi.

Your time starts now
and that's the poser.
You're going to need
all the help you can get
for the ride's nearly over.

Monica e David che, come un piccolo esercito della salvezza, hanno saputo essere presenti nel momento del bisogno, con i loro "yogi" e i "teucci" e non ultimo il "giovane scussut" hanno sfogato in maniera pacifica le mie tensioni interne.
Mauro e Gabriella che ci hanno accolto a casa loro come figli, senza vivere la sfortuna di averci tra le palle ogni giorno.
Marco dedicato a svagarmi (fotografia, informatica, musica ... qualsiasi cosa atta a 'perdere tempo' cit.) nei rari momenti di libertà, ed Elisa che ha avuto la caparbietà di frequentarmi, di capirmi, di vivere con me... la capacità di ascoltare e il piacere di condividere profondissimi silenzi. Grazie per non avere fatto domande e non aver desiderato risposte, per esservi donati interamente.
Florian, grazie per la fiducia, per lo sforzo che fai a salutarmi come si deve, per la competenza anticlericale... e per aver finito il lavoro che avevo duramente iniziato su Elisa. Elisa, abbiamo sempre più cose in comune e non so se è davvero un bene (per te lo dico). Grazie per la tua spontaneità, sincerità, determinazione e il supporto incondizionato ...e le tue torte.
Marco F. grazie per aver condiviso con me lunghe studiate che hanno prodotto i nostri peggiori risultati accademici... e per le dedaliche chiaccherate sul futuro incerto, dagli tutto!
Sandro grazie per aver risolto i miei dubbi teorici, o per lo meno per avermi convinto che i tuoi conti erano giusti :) per me è stato un onore incrociare le nostre vie.

A tutti coloro che fanno in qualche modo parte della mia vita e la rendono utile, un grazie. Continuate così.

Qualunque sia la verita è il grande giorno ormai
c'è una via nelle mie mani non mi tradirà
c'è un mondo che mi attira già in questa azzurrità
io non so se il pazzo sono io o sei tu che rimani qua.

Raccontano che le stelle già parlano di noi
e che l'anima mia vicine così non le ha viste mai.

Andare per andare via all'origine di noi
dove gli astri tramontano è non c'è gravità
Andare per andare via non è solo una piazza
se io chiedo di vivere all'uomo che c'è in me.

È ora! È ora! La terra a chi lavora!