

Università degli Studi di Napoli “Federico II”



Facoltà di Scienze MM. FF. NN.

PHD Thesis
XXIII cycle

The $Z \rightarrow \mu^+ \mu^-$ decay channel in the
CMS experiment at LHC: from cross section
measurement with the 2010 7 TeV collision
dataset to offline machine luminosity monitor

Relatori:
Ch.mo Prof. C.Sciacca
Dott. L.Lista

Studente:
de Gruttola Michele

a te, IGRA

Abstract (in Italiano)

Questo lavoro è stato possibile grazie ad una borsa del Ministero Italiano dell'Università e della Ricerca Scientifica, MIUR, con tematica "ICT e componentistica elettronica".

Durante il programma di dottorato si è avuta l'opportunità di lavorare e partecipare attivamente all'esperimento CMS all'acceleratore CERN di Ginevra toccando diversi punti della lunga catena che porta dalla presa dati all'analisi e produzione dei risultati.

Il rivelatore CMS, che ha preso dati di collisione all'energia di 7 TeV dal 30 marzo 2010 al CERN, è stato progettato per compiere misure di precisione delle collisioni protone-protone fornite dall'acceleratore Large Hadron Collider, LHC.

Nel primo anno del dottorato si è sviluppato un tool software per immagazzinare e riprendere i dati di condizione e calibrazione del rivelatore nei database di CMS. Si è ideato e integrato un tool di monitor delle transazioni ai database, che prevede l'utilizzo di pagine web. Negli anni successivi del programma di dottorato il lavoro è stato incentrato, invece, sullo studio del canale di decadimento in muoni del bosone Z .

La misura della sezione d'urto di produzione del bosone Z è un primo banco di prova del Modello Standard della Fisica delle particelle elementari in un regime di energia mai esplorato prima. Da subito può rivelare infatti l'insorgenza di particelle dal comportamento "esotico". In aggiunta, le proprietà del bosone Z e il suo successivo decadimento in coppie di muoni sono note con alta precisione dai passati esperimenti e quindi possono essere usate per calibrare e allineare il rivelatore. Ancora, la produzione del bosone Z rappresenta un fondo fisico per molti altri processi.

Questa tesi descrive la strategia di misura della sezione d'urto del processo $pp \rightarrow Z + X \rightarrow \mu^+ \mu^- + X$ durante la prima fase dell'esperimento CMS con bassa luminosità integrata. L'originalità del metodo risiede nell'uso di metodi basati sui dati reali e non sulle simulazioni: la sezione d'urto inclusiva del processo $Z \rightarrow \mu^+ \mu^-$ viene determinata con un fit simultaneo del numero di eventi $Z \rightarrow \mu^+ \mu^-$, le efficienze medie di ricostruzione dei muoni nel tracciatore di CMS e nel sistema di rivelatore di muoni, l'efficienza di trigger, e del taglio di isolamento.

Il numero di Z ricavato dal metodo deve poi essere corretto per l'accettazione

geometrica e per la luminosità integrata per ricavare la sezione d'urto. La misura ottenuta con i primi 2.9 pb^{-1} di dati di collisione è:

$$\sigma(\text{pp} \rightarrow \mathbf{ZX}) \times \mathbf{BF}(\mathbf{Z} \rightarrow \mu^+\mu^-) = \mathbf{0.924} \pm \mathbf{0.031}(\text{stat.}) \pm \mathbf{0.022}(\text{syst.}) \pm \mathbf{0.101}(\text{lumi.})\text{nb}, \quad (1)$$

riferita ad un intervallo di massa di $60 < m_{\mu^+\mu^-} < 120 \text{ GeV}/c^2$. Il valore misurato è in accordo con la predizione del Modello Standard ed è stato pubblicato nell'articolo [1].

La presa dati è stata seguita con costanza nel 2010 grazie a lunghi periodi trascorsi al CERN: dalla ricerca dei primi eventi $Z \rightarrow \mu^+\mu^-$ in primavera, alla prima misura usando 198 nb^{-1} di dati presentata alla conferenza ICHEP a Parigi, nel luglio 2010. L'analisi è stata estesa ad una statistica di 2.9 pb^{-1} di dati di collisione per la pubblicazione su rivista. Viene riportato infine un aggiornamento della misura con l'intero insieme di dati raccolti nel 2010, corrispondente a circa 35 pb^{-1} di dati. Inoltre vengono mostrate le prospettive dell'uso degli eventi selezionati di $Z \rightarrow \mu^+\mu^-$ e $W \rightarrow \mu\nu$ per effettuare una stima della luminosità fornita dall'acceleratore.

Abstract

This thesis has been possible thank to a bourse from Italian Minister of Education, MIUR.

During the phd program I had the opportunity to take part actively to the CMS experiments at CERN: many aspects of the chain to take and analyze the data have been studied. Indeed, in the first year of the phd program a tool to store and retrieve the CMS condition data in the CMS databases has been developed. A web monitor tool has been also deployed to trace and check the correctness of the transactions. Instead, in the second and third year of the phd program the decay of the Z boson in two muons has been analyzed.

The CMS detector, which took 7 TeV collision data for the entire 2010 at CERN, is designed to provide precise measurements of TeV proton-proton collisions at the Large Hadron Collider. The measurements of the cross section of the Z boson production at LHC provides a first test of the Standard Model in a new energy domain and may reveal exotic physics processes. Moreover, the properties of the Z boson resonance and its decay into two muons are known to very high precision from LEP experiments and hence can be used as a physics process for calibration and alignment. The Z boson production is also a common background process for many other physics analyses and must therefore be well understood.

This thesis describes the measurement strategy of the cross section for the process $pp \rightarrow Z + X \rightarrow \mu^+\mu^- + X$ at the CMS experiment during its start-up phase with low integrated luminosity available. The originality of the method relies on a fully data driven approach: the $Z \rightarrow \mu^+\mu^-$ inclusive cross-section is determined with a simultaneous fit of the yield of $Z \rightarrow \mu^+\mu^-$ events, the average reconstruction muon efficiencies in the tracker and in the muon detector, the trigger efficiency, as well as the efficiency of the cut applied to select isolated muons. The extracted Z yield has to be just corrected for the geometrical acceptance and for the integrated luminosity in order to measure the cross section. The measurements obtained with the first 2.9 pb^{-1} is

$$\sigma(\mathbf{pp} \rightarrow \mathbf{ZX}) \times \mathbf{BF}(\mathbf{Z} \rightarrow \mu^+\mu^-) = \mathbf{0.924} \pm \mathbf{0.031}(\text{stat.}) \pm \mathbf{0.022}(\text{syst.}) \pm \mathbf{0.101}(\text{lumi.})\text{nb}, \quad (2)$$

referred to the invariant mass range $60 < m_{\mu^+\mu^-} < 120 \text{ GeV}/c^2$. The value measured

is in agreement with the Standard Model prediction. That measurement has been recently published in the paper [1]. The entire 2010 collision dataset has been also analyzed a first preliminary results for the cross-section measurement with 35 pb^{-1} of data is also presented.

I followed constantly the data taking in 2010, thank to frequent travel to CERN, from the first $Z \rightarrow \mu^+\mu^-$ candidate hunting in the first month, to the very first MC driven cross-section measurement showed at ICHEP conference in Paris, in July 2010, using 198 nb^{-1} of data. The data-driven technique has been applied with the dataset of 2.9 pb^{-1} . The update of the measurement with the entire 2010 dataset corresponding to about 35 pb^{-1} is also shown, together with the prospects to use the $Z \rightarrow \mu^+\mu^-$ and $W \rightarrow \mu\nu$ selected events to perform an offline evaluation of the integrated luminosity furnished to the experiment.

Contents

| | |
|--|-----------|
| List of figures | V |
| List of Tables | XI |
| I Overview | 1 |
| Introduction | 1 |
| 1 Particle Physics as of 2010: motivation for the LHC physics program | 3 |
| 1.1 The Electroweak Theory | 5 |
| 1.2 Higgs sector | 6 |
| 1.3 Open problems of the Standard Model | 10 |
| 1.4 LHC physics program | 13 |
| 1.5 Phenomenology of proton-proton collision | 22 |
| 1.6 Monte Carlo Generator | 26 |
| 2 The LHC and the CMS experiment | 29 |
| 2.1 The LHC Machine | 30 |
| 2.1.1 Performance goal | 30 |
| 2.1.2 Injection chain | 33 |
| 2.1.3 Luminosity lifetime | 34 |
| 2.1.4 Integrated luminosity | 35 |
| 2.1.5 LHC performance in 2010 | 35 |
| 2.2 The CMS experiment | 36 |
| 2.2.1 Superconducting magnet | 38 |
| 2.2.2 Inner tracking system | 40 |
| 2.2.3 Electromagnetic calorimeter | 46 |
| 2.2.4 Hadron calorimeter | 51 |
| 2.2.5 The muon system | 58 |
| 2.2.6 Trigger | 65 |

| | | |
|------------|---|------------|
| II | CMS data processing workflow: non event data | 71 |
| 3 | Computing challenges in CMS: data processing work-flow | 73 |
| 3.1 | CMS computing model | 73 |
| 3.2 | Online system | 74 |
| 3.3 | Framework and Event Data Model | 75 |
| 3.4 | Data Formats and processing | 77 |
| 3.5 | Data quality monitor | 79 |
| 3.6 | Analysis model and tool | 80 |
| 4 | Non Event Data in CMS | 83 |
| 4.0.1 | Non-Event Data Description | 84 |
| 4.1 | The Database Architecture | 85 |
| 4.1.1 | The Online Master Database | 87 |
| 4.1.2 | Offline database | 87 |
| 4.1.3 | PopCon | 88 |
| 4.2 | First Experience in Operating the Population on the Condition Databases | 89 |
| 4.2.1 | Alignment and calibration work-flow | 91 |
| 4.2.2 | Condition Objects Written Using PopCon in 2009 | 91 |
| 4.2.3 | Central Population of the Condition Databases | 91 |
| 4.3 | PopCon Web Based Monitoring | 95 |
| III | $Z \rightarrow \mu^+\mu^-$ cross-section measurement and luminosity monitor | 99 |
| 5 | Muon and $Z \rightarrow \mu^+\mu^-$ reconstruction in CMS | 101 |
| 5.1 | Muon identification | 101 |
| 5.2 | Z Bosons Production at LHC | 108 |
| 5.3 | Reconstruction of Z boson decays in CMS | 110 |
| 6 | Z analysis | 113 |
| 6.1 | $Z \rightarrow \mu^+\mu^-$ cross-section determination | 114 |
| 6.2 | W/Z rediscovery phase | 115 |
| 6.3 | MC driven first era analysis | 115 |
| 6.3.1 | Monte Carlo data samples | 117 |
| 6.3.2 | Event selection | 117 |
| 6.3.3 | Results on data and comparison with MC | 119 |
| 6.3.4 | Muon efficiencies | 120 |
| 6.3.5 | $Z \rightarrow \mu^+\mu^-$ cross-section determination with the first 198 nb ⁻¹ of 7 TeV collision data | 122 |

| | | |
|--------------------|---|----------|
| 6.3.6 | Systematic uncertainties | 123 |
| 6.4 | Data driven analysis | 125 |
| 6.4.1 | Method description and trigger requirements | 125 |
| 6.4.2 | Correlation studies | 130 |
| 6.4.3 | Data samples | 135 |
| 6.4.4 | Event selection | 136 |
| 6.4.5 | Fit results | 141 |
| 6.4.6 | Kinematic acceptance | 142 |
| 6.4.7 | Systematic Uncertainties | 146 |
| 6.4.8 | Cross-section determination | 147 |
| 6.4.9 | Correction factors for $W \rightarrow \mu\nu$ analysis | 149 |
| 6.4.10 | Trigger efficiency estimate using $Z \rightarrow \mu^+\mu^-$ | 151 |
| 6.5 | Luminosity measurement using W and Z | 153 |
| 6.6 | Overview of all CMS EWK results in 2010 | 156 |
| 6.7 | Update of $Z \rightarrow \mu^+\mu^-$ results with the entire 2010 proton-proton dataset | 159 |
| Conclusions | | 1 |
| References | | 2 |

List of Figures

| | | |
|------|--|----|
| 1.1 | The three and four point-self-interactions of gauge bosons in the standard electroweak model. The momenta and charges flow into the vertexes. | 6 |
| 1.2 | The Higgs potential $V(\phi)$ for $\mu^2 > 0$ (dashed line) and $\mu^2 < 0$ (solid line). | 8 |
| 1.3 | Higgs interaction vertexes in the Standard Model. | 10 |
| 1.4 | Results of the combined fit of the EWK parameters from LEP, SLC and the TeVatron measurements (July'10). | 11 |
| 1.5 | Fit to the EWK data: $\Delta\chi^2 = \chi^2 - \chi_{min}^2$ vs. m_H curve. The line is the result of the fit using all high- Q^2 data. The band represents an estimate of the theoretical error due to missing higher order corrections. The vertical band shows the 95% CL exclusion limit on m_H from the direct searches at LEP-II (up to 114 GeV) and the TeVatron (160 GeV to 170 GeV). The dashed curve is the result obtained using the evaluation of $\Delta\alpha_{had}^{(5)}(m_Z)$. The dotted curve corresponds to a fit including also the low- Q^2 data. | 12 |
| 1.6 | Combined TeVatron limit on the Higgs mass range. The results are presented in terms of R_{lim} , the ratio of the limits obtained to the rate predicted by the SM, as a function of the Higgs boson mass. Observed and expected (median, for the background-only hypothesis) 95% C.L. upper limits on Standard Model Higgs boson production are reported. The shaded bands indicate the 68% and 95% probability regions in which R_{lim} is expected to fluctuate, in the absence of signal. The limits displayed in this figure are obtained with the Bayesian calculation. | 13 |
| 1.7 | cross-sections for various process in proton-(anti-)proton collisions as a function of center-of-mass energy. | 15 |
| 1.8 | Higgs boson production process by "gluon fusion". | 18 |
| 1.9 | Higgs boson production process by "vector boson fusion". | 18 |
| 1.10 | Higgs boson production in association with W or Z | 19 |
| 1.11 | Higgs boson production in association with a $t\bar{t}$ pair. | 19 |

| | | |
|------|--|----|
| 1.12 | SM Higgs Decay Branching Fractions. | 20 |
| 1.13 | ATLAS significance contours for different Standard Model Higgs boson masses and integrated luminosities as from simulation. The thick curve represents the 5σ discovery contour. The median significance is shown with a colour according to the legend. The hatched area below 2 fb^1 indicates the region where the approximations used in the statistical combination are not accurate, although they are expected to be conservative[32]. | 21 |
| 1.14 | Schematic illustration of three valence quarks forming a proton and interacting via the exchange of gluons. | 23 |
| 1.15 | MSTW 2008 NNLO PDFs (68% C.L.) distributions at $Q^2 = 10000 \text{ GeV}^2$ | 23 |
| 1.16 | Illustration of an hard scattering process of two partons. | 24 |
| 1.17 | Parton shower with schematic string (left) and cluster (right) hadronization models. | 25 |
| 2.1 | LHC sectors. | 30 |
| 2.2 | Overall view of the LHC experiments. | 31 |
| 2.3 | The LHC hadron injector complex. | 33 |
| 2.4 | Maximum Instantaneous luminosity per day delivered to (red) CMS during stable beams at 7 TeV centre-of-mass energy. | 36 |
| 2.5 | Integrated luminosity versus time delivered to (red), and recorded by CMS (blue) during stable beams at 7 TeV centre-of-mass energy in 2010. | 37 |
| 2.6 | Perspective view of the CMS experiment. | 39 |
| 2.7 | General artistic view of the 5 modules composing the cold mass inside the cryostat, with details of the supporting system (vertical, radial and longitudinal tie rods). | 39 |
| 2.8 | The energy over mass ratio E/M , for several detector magnets. | 40 |
| 2.9 | Schematic cross-section through the CMS tracker. Each line represents a detector module. Double lines indicate back-to-back modules which deliver stereo hits. | 42 |
| 2.10 | Geometrical layout of the pixel detector and hit coverage efficiency as a function of pseudorapidity. | 43 |
| 2.11 | Number of measurement points in the strip tracker as a function of pseudorapidity η . Filled circles show the total number (back-to-back modules count as one) while open squares show the number of stereo layers. | 45 |
| 2.12 | Material budget in units of radiation length as a function of pseudorapidity η for the different tracker sub-detectors (left panel) and broken down into the functional contributions (right panel). 30%. | 46 |

| | | |
|------|---|----|
| 2.13 | transverse momentum (left panel), transverse impact parameter (middle panel), and longitudinal impact parameter (right panel) resolution for single muons with transverse momenta of 1, 10 and 100 GeV. | 47 |
| 2.14 | Global track reconstruction efficiency for muons (left panel) and pions (right panel) of transverse momenta of 1, 10 and 100 GeV. | 48 |
| 2.15 | Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules, supermodules and endcaps, with the preshower in front. | 49 |
| 2.16 | Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters. | 52 |
| 2.17 | Isometric view of the HB wedges, showing the hermetic design of the scintillator sampling. | 54 |
| 2.18 | Hadron endcap (HE) calorimeter mounted on the endcap iron yoke. | 56 |
| 2.19 | Longitudinal and transverse views of the CMS detector showing the position of HO layers. | 57 |
| 2.20 | Location of CASTOR in the CMS forward region. | 59 |
| 2.21 | Material thickness in interaction lengths at various depths, as a function of pseudorapidity. | 60 |
| 2.22 | Layout of the CMS barrel muon DT chambers in one of the 5 wheels. The chambers in each wheel are identical with the exception of wheels -1 and +1 where the presence of cryogenic chimneys for the magnet shortens the chambers in 2 sectors. | 62 |
| 2.23 | A DT chamber in position inside the iron yoke; the view is in the $(r-\phi)$ plane. One can see the 2 SLs with wires along the beam direction and the other perpendicular to it. In between is a honeycomb plate with supports attached to the iron yoke. Not shown are the RPCs, which are attached to the DT chambers via support plates glued to the bottom and/or top faces, depending on chamber type. | 62 |
| 2.24 | Quarter-view of the CMS detector. Cathode strip chambers of the Endcap Muon system are highlighted. | 63 |
| 2.25 | Muon reconstruction efficiency as a function of pseudorapidity for selected values of p_T . Left panel: standalone reconstruction (using only hits from the muon system with a vertex constraint). Right panel: global reconstruction (using hits from both the muon system and the tracker). | 64 |
| 2.26 | Layout of a double-gap RPC. | 66 |
| 2.27 | Architecture of the Level-1 Trigger. | 67 |

| | | |
|------|--|-----|
| 2.28 | Efficiency for single muons to pass the Level-1 (solid), Level-2 (dashed), and Level-3 (dotted) triggers as a function of the generated muon pseudo-rapidity. No thresholds on p_T are applied. Note that the vertical scale starts y-axis. The dips at $ \eta $ 0.3 and 0.8 are due to gaps in the muon chamber coverage. | 69 |
| 3.1 | Modules within the CMS Application Framework. | 76 |
| 3.2 | CMS data reduction work-flow. | 77 |
| 3.3 | Sketch of the DQM system, consisting of branches for online and offline monitoring. | 80 |
| 3.4 | Illustration of the reconstruction of composite particle candidates in PAT. | 81 |
| 4.1 | Condition databases architecture. | 86 |
| 4.2 | Schema of the classes for the PopCon package. | 90 |
| 4.3 | Offline work-flow for alignment and calibration used in CRAFT and 2010 collision data period. | 92 |
| 4.4 | Schematic illustration of the central system used to populate ORCON, and of the web monitoring system. | 94 |
| 4.5 | PopCon monitoring Architecture | 95 |
| 4.6 | Screen-shot of part of ORCOFF transactions activity during a typical high-activity day of CMS 2010 data taking. | 96 |
| 4.7 | PopCon activity between March and beginning of May 2010. | 97 |
| 4.8 | Screen-shot of the web page produced in a typical high-activity 2010 collision day by the monitoring system that checks the watchdog tools for the automatic population of ORCON. Different colours helps to identify, quickly, the seriousness of the problem. | 97 |
| 5.1 | A display of a collision event with a muon crossing the DT and CSC chambers in the overlap region between barrel and endcap. | 102 |
| 5.2 | Muon ID selection variables on selected $Z \rightarrow \mu^+\mu^-$ events in data and simulation. Top left: number of tracker hits (pixel plus strip) associated to the muon track; top right: number of pixel hits associated to the muon track; bottom left: number of muon hits in the muon chambers associated to the track; bottom right: number of muon chambers with matches. | 105 |
| 5.3 | Muon isolation variables on selected $Z \rightarrow \mu^+\mu^-$ events in data and simulation. Top left: tracker isolation; top right: ecal isolation; bottom left: hcal isolation; bottom right: relative combined isolation. . . | 107 |
| 5.4 | (Left) Leading order diagram of the Drell-Yan process. (Right) One of the next to leading order diagram of the Drell-Yan process. | 108 |

| | | |
|-----|---|-----|
| 5.5 | Muon candidates classification, from left to right: tracker track, stand-alone track and global muon. | 110 |
| 5.6 | Golden $Z \rightarrow \mu^+\mu^-$ candidates variables distribution on 3 pb^{-1} 7 TeV collision data and simulation. Top left: di-muon mass; top right: di-muon p_T ; center left: di-muon y ; center right: number of di-muon candidates per event; bottom left: p_T of muons from selected $Z \rightarrow \mu^+\mu^-$ candidates; bottom right: η of muons from selected $Z \rightarrow \mu^+\mu^-$ candidates. | 112 |
| 6.1 | Scenario of the $Z \rightarrow \mu^+\mu^-$ cross-section extraction versus the integrated luminosity. | 114 |
| 6.2 | Event display with the first $W \rightarrow \mu\nu$ and $Z \rightarrow \mu^+\mu^-$ candidates detected at CMS. Top: $W \rightarrow \mu\nu$ candidate; bottom left: first $Z \rightarrow \mu^+\mu^-$ candidate; bottom right: second $Z \rightarrow \mu^+\mu^-$ candidate. | 116 |
| 6.3 | Invariant mass distribution of the selected $Z \rightarrow \mu^+\mu^-$ candidates in data superimposed to the MC expectation. (a): linear scale; (b): log scale. | 119 |
| 6.4 | (left) Transverse momentum distribution of the selected $Z \rightarrow \mu^+\mu^-$ candidates in data superimposed to the MC expectation. (right) Rapidity distribution of the selected $Z \rightarrow \mu^+\mu^-$ candidates in data superimposed to the MC expectation. | 120 |
| 6.5 | Measurements of inclusive cross-sections from CMS and experiments at lower-energy colliders. The solid symbols represent $\sigma(W \rightarrow l\nu) \times BF(W \rightarrow l\nu)$ and the hollow symbols, $\sigma(Z(\gamma^*) \rightarrow l^+l^-) \times BF(Z \rightarrow l^+l^-)$ | 125 |
| 6.6 | Left: invariant mass distribution for selected $Z_{\mu\mu}$ (red points) and $Z_{\mu t}$ (black points) candidates in signal events. The $Z_{\mu\mu}$ distribution is normalized in order to have the same number of events as the $Z_{\mu t}$ sample. Right: difference between the $Z_{\mu\mu}$ and $Z_{\mu t}$ distributions. | 128 |
| 6.7 | Left: invariant mass distribution for selected $Z_{\mu\mu}$ (red points) and $Z_{\mu\mu}^{\text{non iso}}$ (black points) candidates in signal events. The $Z_{\mu\mu}$ distribution is normalized in order to have the same number of events as the $Z_{\mu\mu}^{\text{non iso}}$ sample. Right: difference between the $Z_{\mu\mu}$ and $Z_{\mu\mu}^{\text{non iso}}$ distributions. | 128 |
| 6.8 | Left: invariant mass distribution for selected $Z_{\mu s}$ candidates in signal events (black points) superimposed to the pdf (red points) determined from $Z_{\mu\mu}$ candidates by using, for one of the muons in the pair, the momentum of the associated stand-alone muon. Right: difference between the $Z_{\mu\mu}$ and $Z_{\mu s}$ distributions. | 129 |
| 6.9 | Invariant mass distribution of $Z_{\mu\mu}^{2\text{HLT}}$ candidates for data and MC signal and background events for a luminosity of 2.9 pb^{-1} | 138 |

| | | |
|------|--|-----|
| 6.10 | Invariant mass distribution of $Z_{\mu\mu}^{\text{HLT}}$ candidates for data and MC signal and background events for a luminosity of 2.9 pb^{-1} . The discrepancy one sees is due to the different ϵ_{HLT} in data and simulation. . . . | 139 |
| 6.11 | Invariant mass distribution of $Z_{\mu s}$ candidates for data and MC signal and background events for a luminosity of 2.9 pb^{-1} | 140 |
| 6.12 | Invariant mass distribution of $Z_{\mu t}$ candidates for data and MC signal and background events for a luminosity of 2.9 pb^{-1} | 140 |
| 6.13 | Invariant mass distribution of $Z_{\mu\mu}^{\text{non iso}}$ candidates for data and MC signal and background events for a luminosity of 2.9 pb^{-1} | 141 |
| 6.14 | Fit curve superimposed to the invariant mass histogram of $Z_{\mu t}$ candidates for 2.9 pb^{-1} of LHC 7 TeV collision data. | 142 |
| 6.15 | Fit curve superimposed to the invariant mass histogram of $Z_{\mu s}$ candidates for a sample corresponding to an integrated luminosity of 2.9 pb^{-1} of LHC 7 TeV collision data. | 143 |
| 6.16 | Fit curve superimposed to the invariant mass histogram of $Z_{\mu\mu}^{\text{non iso}}$ candidates for a sample corresponding to an integrated luminosity of 2.9 pb^{-1} of LHC 7 TeV collision data. | 143 |
| 6.17 | Poissonian likelihood ratio variation from the minimum versus the yield value compared with a Gaussian parabola. | 145 |
| 6.18 | Muon single muon trigger efficiency versus muon η , obtained using tag and prob method on $Z \rightarrow \mu^+ \mu^-$ selected events on 2.9 pb^{-1} of LHC 7 TeV collision data. | 152 |
| 6.19 | Muon single muon trigger efficiency versus muon p_T obtained using tag and prob method on $Z \rightarrow \mu^+ \mu^-$ selected events on 2.9 pb^{-1} of LHC 7 TeV collision data. | 153 |
| 6.20 | Snapshot of the web page produced with the offline luminosity measurement tool using W and Z bosons decaying into muons. The numbers are in nb^{-1} | 155 |
| 6.21 | Measurements of inclusive cross-sections from CMS and experiments at lower-energy colliders. The plain red symbols represent $\sigma(W \rightarrow l\nu) \times BF(W \rightarrow l\nu)$ and the empty red symbols $\sigma(Z(\gamma^*) \rightarrow l^+l^-) \times BF(Z \rightarrow l^+l^-)$ | 157 |
| 6.22 | Comparison with the theory for the CMS inclusive Z cross-section measurements. | 157 |
| 6.23 | Comparison with the theory for the the CMS inclusive Z and W cross-section measurements. | 158 |
| 6.24 | Comparison of the CMS and ATLAS measurements reported in 2010 7 TeV collision data. | 159 |

| | | |
|------|--|-----|
| 6.25 | fit with the data driven template from $Z \rightarrow \mu^+\mu^-$ to non isolated sample with the full 2010 data. Left: tracker only isolation case. Right: relative combined isolation case. Clear disagreement below the mass peak is visible for the combined isolation plot. | 161 |
| 6.26 | Invariant mass distribution of $Z \rightarrow \mu^+\mu^-$ golden candidates for data and MC signal and background events for a luminosity of 35 pb^{-1} . Left: linear scale. Right: logarithmic scale. | 162 |
| 6.27 | Fit curve superimposed to the invariant mass histogram of $Z_{\mu t}$ candidates for 35 pb^{-1} of LHC 7 TeV collision data. | 163 |
| 6.28 | Fit curve superimposed to the invariant mass histogram of $Z_{\mu s}$ candidates for a sample corresponding to an integrated luminosity of 35 pb^{-1} of LHC 7 TeV collision data. | 163 |
| 6.29 | Fit curve superimposed to the invariant mass histogram of $Z_{\mu\mu}^{\text{non iso}}$ candidates for a sample corresponding to an integrated luminosity of 35 pb^{-1} of LHC 7 TeV collision data. | 164 |

List of Tables

| | | |
|-----|---|-----|
| 1.1 | Quarks and lepton generations and their charges. Only the left handed component are shown. | 3 |
| 1.2 | Fundamental interactions, and their mediators. | 4 |
| 1.3 | Fundamental interactions, and their mediators. | 5 |
| 1.4 | Cross-sections at a centre-of-mass energy of 14 TeV and approximate numbers of expected events per second and per year for some important physics processes at the luminosity of $10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ | 14 |
| 2.1 | LHC ultimate proton beam and collision parameters. | 34 |
| 2.2 | Expected hadron fluence and radiation dose in different radial layers of the CMS tracker (barrel part) for an integrated luminosity of 500fb^{-1} (≈ 10 years). | 42 |
| 2.3 | Absorber thickness in the HB wedges. | 53 |
| 2.4 | Scintillator in the HB wedges. | 54 |
| 4.1 | 2009 CMS condition objects list: “conf.” stays for configuration, “calib.” for calibration and “cond.” for condition. | 93 |
| 5.1 | Branching fractions of the Z boson in the main allowed channels. . . | 108 |
| 5.2 | Next-to-next-to leading order prediction for $Z \rightarrow l^+l^-$ cross-section for different energies of LHC. | 109 |
| 6.1 | Summary of analyzed Monte Carlo samples for the various signal and background processes. | 117 |
| 6.2 | MC cross-sections, cross-sections times acceptance and efficiency of the full selection, and expected yields for the nominal luminosity. All the numbers have been evaluated on the reference MC samples (POWHEG). The reported errors are only the statistical uncertainties due to the MC statistics. | 121 |

| | | |
|------|---|-----|
| 6.3 | Muon efficiencies for $Z \rightarrow \mu^+\mu^-$ analysis. In the first column we report the MC efficiencies estimated from the $Z \rightarrow \mu^+\mu^-$ reference MC sample. In the second column we report the results obtained by counting di-muon candidates in different categories. | 122 |
| 6.4 | Summary of systematic uncertainties for the $Z \rightarrow \mu^+\mu^-$ channel using the first 198 nb^{-1} of 7 TeV collision data. | 124 |
| 6.5 | List of the efficiency terms to be used in Equation (6.28) for the different reconstructed Z categories. | 132 |
| 6.6 | Analyzed MC data samples. | 136 |
| 6.7 | Number of candidates expected (MC) and selected (data) in each category with an invariant mass in the range $[60-120] \text{ GeV}/c^2$ for 2.9 pb^{-1} in data and MC. Here $Z_{\mu\mu} = Z_{\mu\mu}^{\text{HLT}} + Z_{\mu\mu}^{\text{2HLT}}$. For MC the separate contributions from signal and background processes are shown. | 138 |
| 6.8 | Correlation coefficients of the fit parameters from the simultaneous fit minimization | 144 |
| 6.9 | Comparison between fit parameters results with the fit model described in this chapter performed in data with simulation signal and background scaled to the data luminosity. χ^2/NDOF and p -value are also requested. MC-truth values of the average efficiencies are also shown for comparison. | 144 |
| 6.10 | Table of systematic uncertainties for the simultaneous fit $Z \rightarrow \mu^+\mu^-$ cross-section measurement. | 148 |
| 6.11 | Final efficiency factors used in the $W \rightarrow \mu\nu$ analysis. There were obtained applying the simultaneous fit technique on a clean $Z \rightarrow \mu^+\mu^-$ sample. | 149 |
| 6.12 | Correction factors for subsets of muons | 150 |
| 6.13 | Table of muon single muon trigger efficiency versus η obtained in data using the tag and prob method on $Z \rightarrow \mu^+\mu^-$ selected event. | 152 |
| 6.14 | Table of average single muon trigger efficiency versus integrated luminosity. The highest unrescaled muon trigger is reported in each period. | 161 |
| 6.15 | Comparison between fit parameters performed in data with simulation signal and background scaled to the data luminosity, using the full 2010 data, χ^2/NDOF and p -value are also requested. Data/simulation ratio are also shown for comparison. | 164 |

Part I

Overview

Introduction

Between the border of France and Switzerland and 100 meters underground, the largest experiment in human history is currently active. Two decades of planning, developing and building came to an end in 2009, when the experiment finally started collision operation. Roughly 10,000 physicists from more than thirty nations all over the world have been working jointly to achieve this project, unified by the wish to understand the universe we live in. The core of the project is the Large Hadron Collider (LHC) and its four independent particle detectors ATLAS, CMS, ALICE and LHCb.

The CMS detector at the Large Hadron Collider will probe the modern theory of particle physics, commonly known as the Standard Model of particle physics, at energies which have never been reached under laboratory conditions before. Despite the great success of the Standard Model, it is not believed to be the final answer for the description on high energy physics. Many open questions remain, e.g. the verification of mass generation by spontaneous symmetry breaking or the origin of the so-called dark matter, which is believed to give the greatest contribution to the overall matter content of the universe. Before answering these questions at the CMS experiment, it is important to measure well known physics processes in order to understand the detector. One very interesting Standard Model process is the production of the Z boson, which has several important aspects. The measurement of its total cross-section, i.e. the probability of its production, is per se an important test of the standard model predictions. Moreover, the decay of the Z boson into two leptons is an important background process for various other physics studies. Leptons, originating from Z bosons with large transverse momenta, may fake the signature of predicted new particles, which decay into high energetic leptons. As already mentioned, an important aspect of the Z boson and its decay into leptons is its well understood theoretical and experimental description. The expected high rate of producing Z bosons at LHC therefore offers the possibility to use this channel as a calibration benchmark for the whole CMS detector, and finally as an offline measurement of the LHC luminosity delivered to experiment. This thesis focuses on the decay of the Z boson into two muons, since this provides a relative clean signature which can be clearly discriminated from other processes.

The outline of the thesis is as follows. In Chapter 1 a brief overview of the actual theoretical understanding of elementary particles and their interactions is given. This is followed by a short introduction to the Large Hadron Collider and the CMS experiment in Chapter 2. A crucial aspect for the smooth running of CMS operation is the correct matching of the collision data with the conditions and constants of calibration of the detector: an overview of the CMS computing infrastructure is given in Chapter 3 while a detailed description of the non-event data produced by CMS and needed for the analysis is given in the fourth Chapter. The fifth Chapter is dedicated to the muon and di-muon reconstruction performance in CMS.

The final sixth Chapter of this work is then dedicated to the description of the method to measure the $Z \rightarrow \mu^+\mu^-$ cross section in 2010, and the evolution of the strategy to measure it with the increase of the integrated luminosity: from the $Z \rightarrow \mu^+\mu^-$ rediscovery to the first cross-section measurement with a “cut and count”, and finally the use of fully data driven technique.

Various parts of this thesis have been presented at conferences and published in various journals [1] -[4] or CMS¹ notes [5] -[11].

¹CMS notes are internal documents of the CMS collaboration. Some of them are refereed by an internal committee and accessible for the public [9] - [11].

Chapter 1

Particle Physics as of 2010: motivation for the LHC physics program

Particle Physics studies the elementary particles that constitute matter and describes the interactions between them. This dualism matter-forces has been widely overcome in the context of modern quantum field theory. The main concept of the quantum field theory is the unified description of particles and forces in terms of quantum fields, the former have semi-integer spin (fermions), the latter have integer spin (bosons).

According to the current knowledge we can distinguish two types of elementary particles: quarks and leptons, both with spin $1/2$ (in \hbar unit) and arranged in three generations. The fundamental interactions discovered and studied in detail so far in

| | Generation | | | |
|---------|--|--|--|-------------|
| | 1 | 2 | 3 | Q/e |
| Leptons | $\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L$ | $\begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L$ | $\begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L$ | 0 -1 |
| Quarks | $\begin{pmatrix} u \\ d \end{pmatrix}_L$ | $\begin{pmatrix} c \\ s \end{pmatrix}_L$ | $\begin{pmatrix} t \\ b \end{pmatrix}_L$ | 2/3 -1/3 |

Table 1.1: Quarks and lepton generations and their charges. Only the left handed component are shown.

Particle Physics are four: the electromagnetic interaction, the weak interaction, the strong interaction and the gravitational interaction. The past experimental observations show that the first three rule the microscopic world while the latter describes the behavior of macroscopic objects. Leptons are subject to weak and electromag-

netic forces (except neutrinos which are neutrals and so are pure probes of the weak interaction), while quarks may experience also strong forces between each others.

The Standard Model is the theory, verified by the experimental data achieved so far, which describes the weak, electromagnetic and strong interactions in terms of a local gauge symmetry group:

$$SU(3)_{col} \otimes SU(2)_L \otimes U(1)_Y \quad (1.1)$$

where:

- $SU(3)_{col}$ is a non abelian gauge symmetry group which describes the strong interactions. Such a structure involves eight independent matrices, which are the generators of the group. This geometrical relation reflect the fact that the strong interactions is carried by eight massless particles, the gluons. The gluons have a strong charge, known as "color". Gluons and quarks strong interaction are well described by the quantum chromodynamic (QCD) theory[12].
- $SU(2)_L \otimes U(1)_Y$ is the weak isospin symmetry group, introduced by Glashow-Weinberg-Salam [13][14][15], which describes the unified electromagnetic and weak (EWK) interaction. The EWK force is carried by three massive vector bosons W^+, W^- and Z , plus the massless photon, γ .

The symmetry group in (1.1) extends the gauge invariance $U(1)$ of the QED to the electroweak and strong interactions. Table 1.2 lists the four interactions together with the corresponding boson mediators.

| Interaction | Mediator | Symbol |
|-----------------|----------|---------------|
| Electromagnetic | photon | γ |
| Weak | VB | W^+, W^-, Z |
| Strong | 8 gluons | g |
| Gravitational | graviton | G |

Table 1.2: Fundamental interactions, and their mediators.

The Standard Model comprises the *spontaneous symmetry breaking* mechanism (see Sec. 1.2) which explain how the vector boson of the EWK forces and all the Standard Model particles can acquire their masses. This mechanism foresee the existence of a scalar particle, the Higgs boson. The Higgs mass is a free parameter of the theory, but up to now it has escaped any direct observation. But the range of the Higgs mass values is constrained by indirect measurements. The direct detection of the Higgs bosons is one of the most important physics goal of the modern high energy physics.

1.1 The Electroweak Theory

The electroweak theory [13] [14] [15] is based on the $SU(2)_L \times U(1)_Y$ Lagrangian:

$$\mathcal{L}_{SU(2) \times U(1)} = \mathcal{L}_{gauge} + \mathcal{L}_\phi + \mathcal{L}_f + \mathcal{L}_{Yuk}. \quad (1.2)$$

The gauge part is

$$\mathcal{L}_{gauge} = \frac{1}{4} W_{\mu\nu}^i W^{\mu\nu i} - \frac{1}{4_{\mu\nu}} B^{\mu\nu}, \quad (1.3)$$

where W_μ^i ; $i = 1, 2, 3$ and B_μ are respectively the $SU(2)_L$ and $U(1)_Y$ gauge fields, with field strength tensors

$$\begin{aligned} B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu, \\ W_{\mu\nu}^i &= \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g \epsilon_{ijk} W_\mu^j W_\nu^k, \end{aligned} \quad (1.4)$$

where $g(g')$ is the $SU(2)_L(U(1)_Y)$ gauge coupling constants and ϵ_{ijk} is the totally antisymmetric tensor. The structures of the gauge fields reflects the presence of three and four-point self-interactions in the theory, as illustrated in Fig. 1.1. B is the $U(1)$ field associated with the weak hyper-charge $Y = Q - T_3$, where Q and T_3 are respectively the electric charge operator and the third component of weak $SU(2)$. Their eigenvalues will be denoted by y , Q , and t_3 , respectively. Table 1.3 list the assignments of y , Q , and t_3 for the three generation particles.

| | t | t_3 | y | Q |
|---------------------------------------|-----|-------|------|------|
| $\nu_{eL}, \nu_{\mu L}, \nu_{\tau L}$ | 1/2 | 1/2 | -1 | 0 |
| $\nu_{eR}, \nu_{\mu R}, \nu_{\tau R}$ | 0 | 0 | 0 | 0 |
| e_L, μ_L, τ_L | 1/2 | -1/2 | -1 | -1 |
| e_R, μ_R, τ_R | 0 | 0 | -2 | -1 |
| u_L, c_L, t_L | 1/2 | 1/2 | 1/3 | 2/3 |
| u_R, c_R, t_R | 0 | 0 | 4/3 | 2/3 |
| d'_L, s'_L, b'_L | 1/2 | -1/2 | 1/3 | -1/3 |
| d'_R, s'_R, b'_R | 0 | 0 | -2/3 | -1/3 |

Table 1.3: Fundamental interactions, and their mediators.

B has no self-interactions because the group is abelian. The B and W_3 fields will eventually mix to form the photon and Z boson after the symmetry breaking as will be described later in the next section.

The scalar part of the Lagrangian is

$$\mathcal{L}_\phi = (D^\mu \phi)^\dagger D^\mu \phi - V(\phi); \quad (1.5)$$

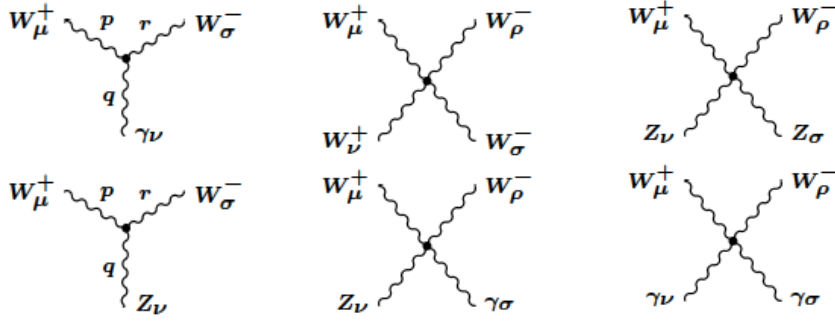


Figure 1.1: The three and four point-self-interactions of gauge bosons in the standard electroweak model. The momenta and charges flow into the vertexes.

where $\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$ is the complex Higgs scalar field, which is a doublet under $SU(2)_L$ with $U(1)$ charge $y_\phi = +1/2$. The gauge covariant derivative is defined as:

$$D_\mu \phi = \left(\partial_\mu + ig \frac{\tau^i}{2} W_\mu^i + \frac{ig'}{2} B_\mu \right) \phi, \quad (1.6)$$

where the τ_i are the Pauli matrices. The square of the covariant derivative leads to three and four-point interactions between the gauge and scalar fields (see Fig. 1.3 and discussion later).

$V(\phi)$ is the Higgs potential. The combination of $SU(2)_L \times U(1)_Y$ invariance and renormalizability restricts V to the form

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \quad (1.7)$$

For μ^2 there will be spontaneous symmetry breaking as will be discussed in the next Section. The λ term describes a quartic self-interaction between the scalar fields. Vacuum stability requires $\lambda > 0$.

\mathcal{L}_ϕ and \mathcal{L}_{Yuk} describe respectively the free leptons lagrangians and the Yukawa couplings between the single Higgs doublets, ϕ , and the various flavors of quarks and leptons. These couplings are needed to make possible the quark and leptons mass term generations in the SM.

1.2 Higgs sector

Gauge invariance (and therefore renormalizability) does not allow mass terms in the Lagrangian for the gauge bosons or for chiral fermions. Massless gauge bosons are not acceptable for the weak interactions, which are known to be short-ranged.

Hence, the gauge invariance must be broken spontaneously, in order to preserve the renormalizability. The idea [16] is that the lowest energy (vacuum) state does not respect the gauge symmetry and induces effective masses for particles propagating through it. Let us introduce the complex vector

$$v = \langle 0|\phi|0\rangle = \text{constant}; \quad (1.8)$$

which has components that are the vacuum expectation values of the various complex scalar fields. v is determined by rewriting the Higgs potential as a function of v , $V(\phi) \rightarrow V(v)$, and choosing v such that V is minimized. We interpret v as the lowest energy solution of the classical equation of motion. The quantum theory is obtained by considering fluctuations around this classical minimum, $\phi = v + \phi'$.

The single complex Higgs doublet in the Standard Model can be rewritten in a Hermitian basis as:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}(\phi_1 - i\phi_2) \\ \frac{1}{\sqrt{2}}(\phi_3 - i\phi_4) \end{pmatrix}, \quad (1.9)$$

where $\phi_i = \phi = \phi^\dagger$ represent four Hermitian fields. In this new basis the Higgs potential becomes

$$V(\phi) = \frac{1}{2}\mu^2 \left(\sum_{i=1}^4 \phi_i^2 \right) + \frac{1}{4}\lambda \left(\sum_{i=1}^4 \phi_i^2 \right)^2, \quad (1.10)$$

which is clearly $O(4)$ invariant. Without loss of generality we can choose the axis in this four-dimensional space so that $\langle 0|\phi_i|0\rangle = 0$, $i = 1, 2, 4$ and $\langle 0|\phi_3|0\rangle = \nu$. Thus,

$$V(\phi) \rightarrow V(v) = \frac{1}{2}\mu^2\nu^2 + \frac{1}{4}\lambda\nu^4, \quad (1.11)$$

which must be minimized with respect to ν . Two important cases are illustrated in Fig. 1.2. For $\mu^2 > 0$ the minimum occurs at $\nu = 0$. That is $SU(2)_L \times U(1)_Y$ is unbroken at the minimum. On the other hand, for $\mu^2 < 0$ the $\nu = 0$ symmetric point is unstable, and the minimum occurs at some non-zero value of ν which breaks the $SU(2)_L \times U(1)_Y$ symmetry.

We are interested in the case $\mu^2 < 0$, for which the Higgs doublet is replaced, in first approximation, by its classical value $\phi \rightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu \end{pmatrix} \equiv v$. The generators L^1 , L^2 , and $L^3 - Y$ are spontaneously broken (e.g., $L^1 v \neq 0$). On the other hand, the vacuum carries no electric charge ($Qv = (L^3 + Y)v = 0$), so the $U(1)_Q$ of electromagnetism is not broken. Thus, the electroweak $SU(2)_L \times U(1)_Y$ group is spontaneously broken to the $U(1)_Q$ subgroup, $SU(2)_L \times U(1)_Y \rightarrow U(1)_Q$. To quantize around the classical vacuum, we write $\phi = v + \phi'$, where ϕ' are quantum fields with zero vacuum

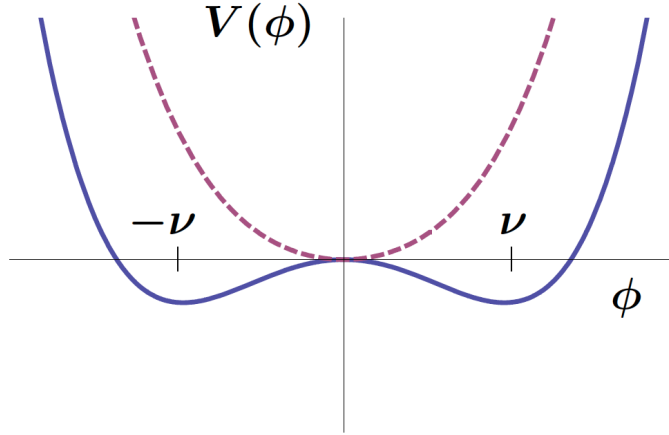


Figure 1.2: The Higgs potential $V(\phi)$ for $\mu^2 > 0$ (dashed line) and $\mu^2 < 0$ (solid line).

expectation value. To display the physical particle content it is useful to rewrite the four Hermitian components of ϕ' as:

$$\phi = \frac{1}{\sqrt{2}} e^{i \sum \xi^i L^i} \begin{pmatrix} 0 \\ \nu + H \end{pmatrix}, \quad (1.12)$$

H is a Hermitian field which will turn out to be the physical Higgs scalar. If we had been dealing with a spontaneously broken global symmetry the three Hermitian fields ξ^i would be the massless pseudo-scalar Nambu-Goldstone bosons[17] that are necessarily associated with broken symmetry generators. However, in a gauge theory they do not interfere with the other degrees of freedom and so they disappear from the physical spectrum. To see this it is useful to go to the unitary gauge:

$$\phi \rightarrow \phi' = e^{-i \sum \xi^i L^i} \phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + H \end{pmatrix}, \quad (1.13)$$

in which the Goldstone bosons disappear. In this gauge, the scalar covariant kinetic energy term takes the simple form

$$\begin{aligned} (D_\mu \phi)^\dagger D^\mu \phi &= \frac{1}{2} (0 \ \nu) \left[\frac{g}{2} \tau^i W_\mu^i + \frac{g'}{2} B + \mu \right]^2 \begin{pmatrix} 0 \\ \nu \end{pmatrix} + H \text{ terms} \\ &\rightarrow m_W^2 W^{+\mu} W_\mu^- + \frac{m_Z^2}{2} Z^\mu Z_\mu + H \text{ terms}. \end{aligned} \quad (1.14)$$

Thus, spontaneous symmetry breaking generates mass terms for the W and Z gauge bosons

$$W^\pm = \frac{1}{\sqrt{2}} (W^1 \mp iW^2)$$

$$Z = -\sin\theta_W B + \cos\theta_W W^3. \quad (1.15)$$

The photon fields:

$$A = \cos\theta_W B + \sin\theta_W W^3 \quad (1.16)$$

remains massless. The masses are

$$m_W = \frac{g\nu}{2} \quad (1.17)$$

and

$$m_Z = \sqrt{g^2 + g'^2} \frac{\nu}{2} = \frac{m_W}{\cos\theta_W}, \quad (1.18)$$

where the Weinberg angle θ_W is defined by:

$$\tan\theta_W \equiv \frac{g'}{g} \Rightarrow \sin^2\theta_W = 1 - \frac{m_W^2}{m_Z^2}. \quad (1.19)$$

One can think of the generation of masses as due to the fact that the W and Z interact constantly with the ensemble of the scalar fields and therefore acquire masses, in analogy with a photon propagating through a plasma. The Goldstone boson has disappeared from the theory but has re-emerged as the longitudinal degree of freedom of a massive vector particle.

We can define the Fermi constant G_F as $G_F/\sqrt{2} \sim g^2/8M_W^2$, where $G_F = 1.16637(5) \times 10^{-5} \text{ GeV}^{-2}$ which is determined by the muon lifetime. The weak scale ν is therefore

$$\nu = 2M_W/g \simeq (\sqrt{2}G_F)^{-1/2} \simeq 246 \text{ GeV} \quad (1.20)$$

Similarly, $g = e/\sin\theta_W$, where e is the electric charge of the positron. Hence, to lowest order

$$m_W = m_Z \cos\theta_W \sim \frac{(\pi\alpha/\sqrt{2}G_F)^{1/2}}{\sin\theta_W} \quad (1.21)$$

where $\alpha \sim 1/137.036$ is the fine structure constant. Using $\sin^2\theta_W \sim 0.23$ from neutral current scattering (see[18] for a review of the first discovery of neutral current with the Gargamelle bubble chamber in 1972 at CERN and interpretation of the processes with the introduction of the Weinberg angle), one expects $M_W \sim 78 \text{ GeV}/c^2$, and $M_Z \sim 89 \text{ GeV}/c^2$. (These predictions are increased by $\sim (2 - 3) \text{ GeV}$ by loop corrections.) The W and Z were discovered at CERN by the UA1[19] and UA2[20] groups in 1983. Subsequent measurements of their masses and other properties have been in excellent agreement with the standard model expectations (including the higher-order corrections). The current values are

$$m_W = 80.398 \pm 0.025 \text{ GeV}, \quad m_Z = 91.1876 \pm 0.0021 \text{ GeV}. \quad (1.22)$$

Figures 1.3 summarises all the Higgs interaction vertexes resulting from the Standard Model lagrangian.

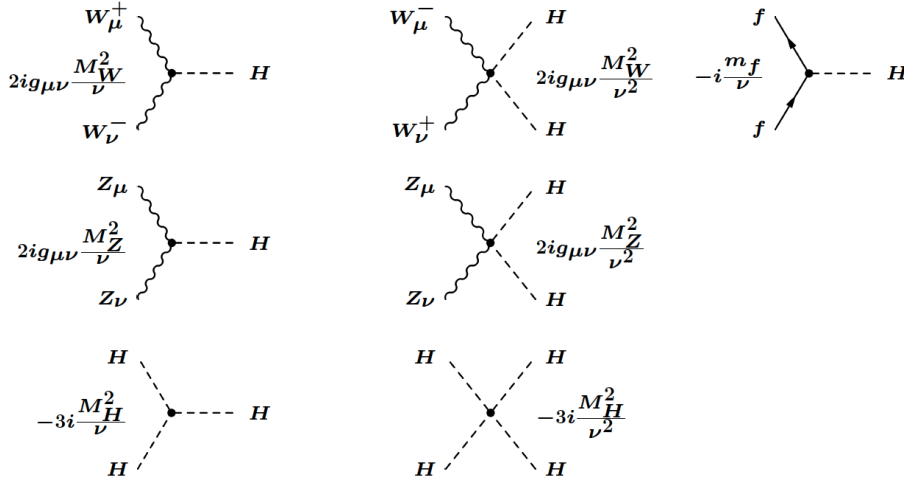


Figure 1.3: Higgs interaction vertexes in the Standard Model.

1.3 Open problems of the Standard Model

The Higgs sector brings many open problems in the Standard Model description. On the one hand, the experimental verification of the Standard Model cannot be considered complete until the physics of the Higgs sector is not established by experiments. On the other hand, the Higgs is directly related to most of the major open problems of particle physics, like the flavour problem or the hierarchy problem, the latter strongly suggesting the need for new physics near the weak scale (which could possibly clarify the dark matter identity). For a detailed description of all the open problems see Reference [21].

The present experimental information on the Higgs sector, mainly obtained from LEP, is surprisingly limited. The Higgs particle has not been found but, in the SM, its mass can well be larger than the present direct lower limit $m_H \geq 114.4$ GeV (at 95% C.L.) obtained from searches at LEP-II[22]. The radiative corrections computed in the Standard Model when compared to the data lead to a clear indication of a light Higgs, not too far from the present lower bound. The experimental upper limit on m_H , obtained from fitting the data (see table 1.4 and figure 1.5), depends on the value of the top quark mass m_t (the one-loop radiative corrections are quadratic in m_t and logarithmic in m_H). The CDF and D0[23] combined value after Run II is at present (updated for summer 2010: final Run-I results published and preliminary Run-II results) $m_t = 173.3 \pm 1.1$ GeV. As a consequence the present limit on m_H is very stringent: $m_H < 190$ GeV (at 95% C.L., after including the information from the 114.4 GeV direct bound).

In addition the recent combined results for CDF and D0 allow with a 95% C.L.

the direct exclusion for SM Higgs with mass m_H between 162 and 166 GeV/c^2 as also shown in Fig. 1.6

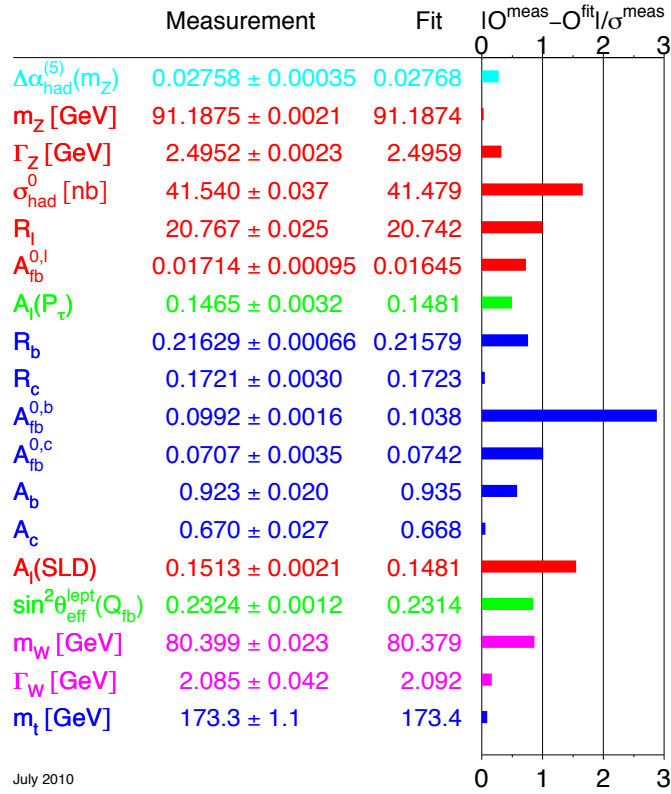


Figure 1.4: Results of the combined fit of the EWK parameters from LEP, SLC and the TeVatron measurements (July'10).

LHC will address the following questions : do the Higgs particles actually exist? How many: one doublet, several doublets, additional singlets? SM Higgs or SUSY[24] Higgses? Fundamental or composite (of fermions, of WW...)? Pseudo-Goldstone boson of an enlarged symmetry? A manifestation of large extra dimensions (5th component of a gauge boson, an effect of orbifolding or of boundary conditions...[25])? Or some combination of the above or something so far unthought of?

Many of these questions will be addressed soon by LHC experiments.

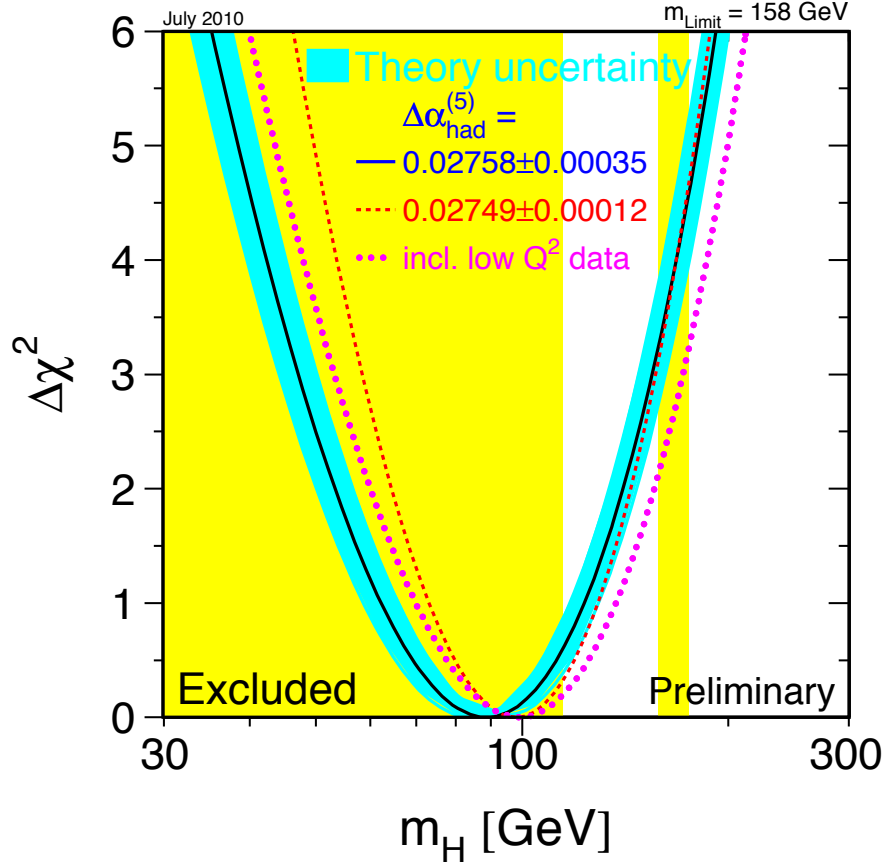


Figure 1.5: Fit to the EWK data: $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ vs. m_H curve. The line is the result of the fit using all high- Q^2 data. The band represents an estimate of the theoretical error due to missing higher order corrections. The vertical band shows the 95% CL exclusion limit on m_H from the direct searches at LEP-II (up to 114 GeV) and the Tevatron (160 GeV to 170 GeV). The dashed curve is the result obtained using the evaluation of $\Delta\alpha_{had}^{(5)}(m_Z)$. The dotted curve corresponds to a fit including also the low- Q^2 data.

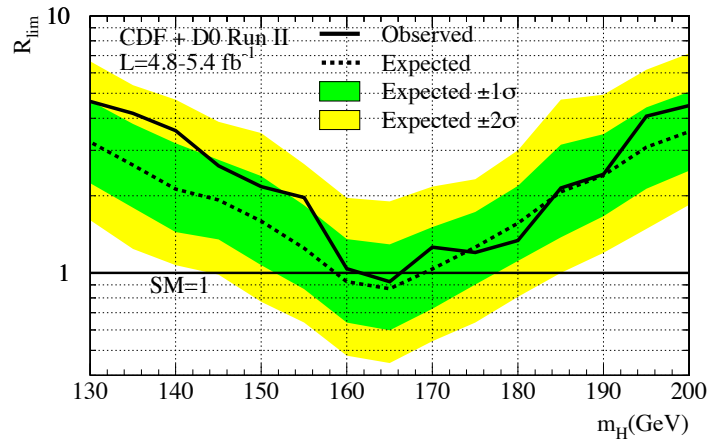


Figure 1.6: Combined TeVatron limit on the Higgs mass range. The results are presented in terms of R_{lim} , the ratio of the limits obtained to the rate predicted by the SM, as a function of the Higgs boson mass. Observed and expected (median, for the background-only hypothesis) 95% C.L. upper limits on Standard Model Higgs boson production are reported. The shaded bands indicate the 68% and 95% probability regions in which R_{lim} is expected to fluctuate, in the absence of signal. The limits displayed in this figure are obtained with the Bayesian calculation.

1.4 LHC physics program

After a construction time of more than 10 years, the Large Hadron Collider (LHC) delivered the first 7 TeV proton-proton collisions this year. After running in an initial phase at a centre-of-mass energy of 7 TeV, it is anticipated that the design energy of 14 TeV and the design luminosity of $10^{34} \text{cm}^{-2} \text{s}^{-1}$ will be reached. This machine will open up the possibility to explore the TeV energy range, which plays a key role in the investigation of the electroweak symmetry breaking and in the search for physics beyond the Standard Model. The experiments ATLAS and CMS have been designed and optimized to cover a large spectrum of possible physics signatures. The main focus will be on the search for the Higgs boson as well as for particles predicted by supersymmetry or theory like technicolor[21], new gauge bosons and searches for composite quarks and leptons. Besides the discovery potential for new physics the experiments also have a large potential to perform precision Standard Model measurements, like measurements of the W and top quark masses or triple gauge boson couplings. Many of the Standard Model measurements, in particular at the beginning of the data-taking period, will be used for a detailed understanding of the detector performance and for an optimization of reconstruction algorithms.

It is expected that after the first run in 2010/11 an initial luminosity of $10^{33} \text{cm}^{-2} \text{s}^{-1}$

can be achieved. During the further years of operation, this value should rise to the design luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. Integrated luminosities of 1 and 10 fb^{-1} per year should therefore be collected at those luminosity respectively. The expected cross sections at a center-of-mass energy of 14 TeV and the corresponding event rates at low luminosity are given in Table 1.4 and figure 1.7.

| Process | $\sigma(\text{pb})$ | Events/s | Events/year | Other machines |
|--|---------------------|-----------------|---------------|-------------------------|
| $W \rightarrow e\nu$ | 1.5×10^4 | 15 | $\times 10^8$ | 10^4 LEP, 10^7 TeV. |
| $Z \rightarrow ee$ | 1.5×10^3 | 1.5 | $\times 10^6$ | 10^7 LEP |
| $t\bar{t}$ | 800 | 0.8 | $\times 10^7$ | 10^6 TeV. |
| $b\bar{b}$ | 5×10^8 | 5×10^5 | 10^{12} | 10^8 Belle/BaBar |
| $\tilde{g}\tilde{g}$ ($m_j = 1 \text{ TeV}$) | 1 | 0.001 | $\times 10^4$ | |
| Higgs ($m_H = 120 \text{ GeV}$) | 10 | 0.01 | $\times 10^5$ | |
| Inclusive jets with $p_T > 200 \text{ GeV}/c$ | 10^5 | 100 | 10^9 | 10^7 (all) |

Table 1.4: Cross sections at a centre-of-mass energy of 14 TeV and approximate numbers of expected events per second and per year for some important physics processes at the luminosity of $10^{33} \text{ cm}^{-2} \text{ s}^{-1}$.

For the reduced center-of-mass energy of 7 TeV the reduction in cross sections depends on the energy scale of the final state and is largest for heavy objects with masses in the TeV range.

For the pair production of top quarks, for example, a reduction factor of about four is found; for W and Z bosons is approximately a factor of two. At 14 TeV with a luminosity of $10^{33} \text{ cm}^{-2} \text{ s}^{-1}$, almost 50 W and five Z bosons decaying to lepton pairs will be produced every second, as well as one $t\bar{t}$ pair and 500,000 $b\bar{b}$ pairs. The copious $t\bar{t}$ production will constitute a significant background for many searches of new physics signals since it may lead to characteristic final states with leptons, jets and missing transverse energy \cancel{E}_T . On the other hand it can be used to determine important detector performance parameters in the first luminosity running.

LHC discovery potential during the first luminosity phase

Before going on, it is necessary to define the most relevant and used kinematic variables and definitions.

The transverse momentum p_T is defined as the component of a particle's (or jet's) momentum \vec{p} perpendicular to the beam line, i.e., $p_T = p \sin \theta$, where $p = |\vec{p}|$ and θ is the angle w.r.t. the beam line. Dealing with energy (E) deposits in calorimeters, the transverse energy is introduced, $E_T = E \sin \theta$.

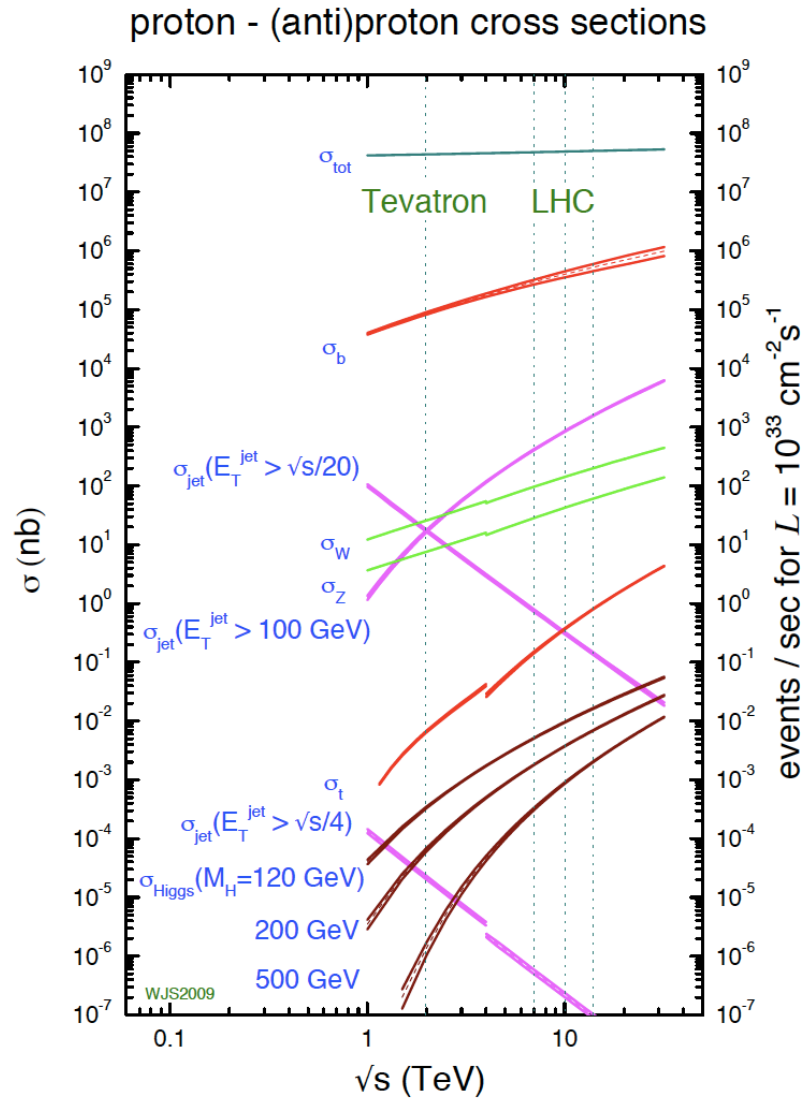


Figure 1.7: cross sections for various process in proton-(anti-)proton collisions as a function of center-of-mass energy.

A *jet* is a collimated spray of hadrons originating from quarks/gluons coming from the hard scattering (jets are experimental signatures of quarks and gluons, and unlike photons, leptons etc, jets have to be defined by an algorithm for quantitative studies).

If the energy deposits are defined as vectors (by using their directions w.r.t. the interaction point), their negative vector sum gives the missing transverse energy \cancel{E}_T . At hadron colliders the rapidity $y = 0.5 \ln [(E + p_L)/(E - p_L)]$ is a well-suited kinematic variable, because differences in rapidity are invariant under boosts along the beam direction. Here p_L denotes the momentum component along the beam line. For massless particles the rapidity is equal to the pseudo-rapidity $\eta = -\ln \tan \theta/2$. Indeed, when describing the segmentation of the detector elements at hadron colliders one usually refer to (pseudo-)rapidity intervals. Finally, the azimuthal angle around the beam direction is usually denoted as ϕ .

During the early running phase at $\sqrt{s} = 7$ TeV important Standard Model reference signals can be established. The observation of the W and Z bosons are important for the extraction of trigger and reconstruction efficiencies as well as for the understanding of the measurement of the missing transverse energy, \cancel{E}_T , which is vital for the search for physics beyond the Standard Model. The initial measurements of the W and Z boson cross sections are limited by uncertainties on the luminosity which are estimated to be at the level of 5-10% during the first 1-2 years.

Top production signal can already be established after a few months of running at a luminosity of $10^{32} \text{ cm}^{-2} \text{ s}^{-1}$. Such a signal can be identified after applying simple cuts, i.e. by requiring one lepton with $p_T > 20$ GeV/c and four jets with $p_T > 40$ GeV/c within $\eta < 2.5$ (see for example the first results from the CMS collaboration[26]). Such a sample is ideally suited to establish the jet energy scale via the $W \rightarrow qq$ and $t \rightarrow Wb$ decay chains appearing in $t\bar{t}$ events as well as to determine the *b*-tagging performance from early data[27] [28].

In addition, the experiments must be open and unbiased for early surprises and unexpected discoveries with early data. Therefore unbiased inclusive measurements of the lepton, di-lepton, jet and missing transverse energy spectra are essential. To illustrate the performance, a 5σ discovery of a new heavy vector boson Z' with Standard-Model-like couplings and with a mass of 1 (1.5) TeV/c² will be possible after collecting integrated luminosities of 70 pb⁻¹ (300 pb⁻¹) at a centre-of-mass energy of 14 (7) TeV. First data on QCD jet production will be used to test perturbative QCD and to look for deviations from the Standard Model. Given the new energy regime, higher transverse jet energies where, e.g. first signs of compositeness[29] could show up, can be rapidly probed. Even if relatively large jet energy scale uncertainties of the order of 10% for early data are assumed, compositeness scales of $\Lambda = 3$ TeV can already be probed with data corresponding to an integrated luminosity of only 100 pb⁻¹ at 7 TeV. This is close to the present TeVatron reach, where compositeness scales of 2.9 TeV are excluded at the 95% C.L.[30].

Precision measurements of standard model parameters

At the time of the LHC start-up, the W mass will be known with a precision of $\sim 25 \text{ MeV}/c^2$ from measurements at the TeVatron and LEP-II[22]. The motivation to improve this result is mainly that precise measurements of the W mass, of the top mass and of the Higgs mass will provide stringent tests of the consistency of the underlying theory. At the LHC, 60 million well-reconstructed $W \rightarrow l\nu$ decays (where $l = e$ or μ) should be collected by each experiment in one year of data taking at low luminosity. The statistical error on the W mass measurement is therefore expected to be $< 2 \text{ MeV}/c^2$. The systematic error will arise mainly from the Monte Carlo reliability in reproducing the detector performance. Uncertainties related to the physics result from the limited knowledge of the $W p_T$ spectrum, the parton density functions, the W width and from W radiative decays. Uncertainties related to the detector result from the limited knowledge of the absolute lepton energy scale and the detector energy and momentum resolution and response. Many of these uncertainties (lepton scale, detector resolution and response and the $W p_T$ spectrum) will be constrained in the experiment by using the high-statistics sample of leptonic Z decays.

Measurement of the top quark mass

At the LHC, top quark measurements will benefit from the large $t\bar{t}$ event samples, so that not only the mass and the production cross section, but also branching ratios, couplings and rare decays can be studied in detail. The best channel for the top mass measurement will most likely be $t\bar{t}$ production with one W decaying leptonically and the other one hadronically. The top mass will be determined from the hadronic part of the decay, as the invariant mass of the three jets originating from the same top ($m_t = m_{j\bar{j}b}$). The associated leptonic top decay will be used to trigger the events and to suppress backgrounds. All together, a total uncertainty in the range of $\sim 1 \text{ GeV}/c^2$ or better should be achieved.

Search for supersymmetry

In a theory with unbroken Supersymmetry (SUSY), for every type of boson there exists a corresponding type of fermion with the same mass and internal quantum numbers, and vice-versa. If SUSY[24] exists at the electroweak scale, its discovery at the LHC can be very likely. SUSY cross section is dominated by gluinos and squarks, which are strongly produced with cross sections comparable to the Standard Model backgrounds at the same Q^2 . Gluinos and squarks then decay via a series of steps into the lightest supersymmetric particle (LSP), which may itself decay, if R-parity is violated. These decay chains lead to a variety of signatures involving multiple jets, leptons, photons, heavy flavours, W and Z bosons, and missing transverse energy.

The combination of a large production cross section and distinctive signatures makes it easy to separate SUSY from the Standard Model background.

In a first step of SUSY searches at the LHC multijet events[31] with large missing transverse momentum will be studied. An excess at large E_T^{miss} would provide sensitivity to squarks and gluinos up to the TeV energy range. Already for an integrated luminosity of only 1 fb^{-1} , the reach in the jets+ \cancel{E}_T channel extends to squark and gluino masses of the order of $1.5 \text{ TeV}/c^2$. The main challenge at the LHC is not to discover SUSY but to separate the many SUSY processes that occur and to measure the masses and other properties of the SUSY particles.

The search for the Higgs boson

The main Standard Model Higgs production mechanisms are shown in figures 1.8 - 1.11.

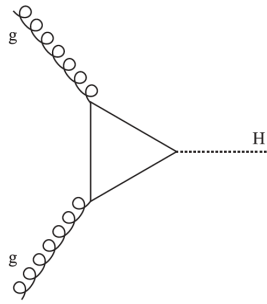


Figure 1.8: Higgs boson production process by “gluon fusion”.

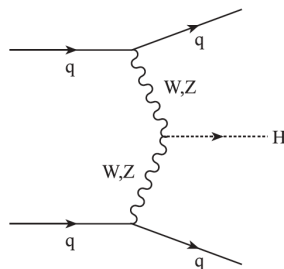


Figure 1.9: Higgs boson production process by “vector boson fusion”.

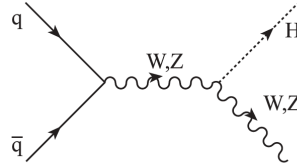


Figure 1.10: Higgs boson production in association with W or Z .

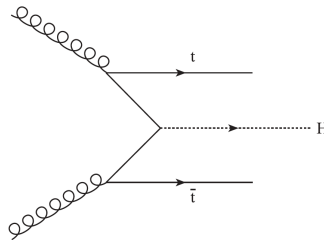


Figure 1.11: Higgs boson production in association with a $t\bar{t}$ pair.

The Standard Model Higgs boson is searched for at the LHC in various decay channels, the choice of which is given by the signal rates and the signal-to-background ratios in the various mass regions (see Fig. 1.12 for an illustrative summary).

Inclusive final states have been considered to be the most promising for the Higgs search, among them the well established $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow lll$ decay channels. In addition, more exclusive channels have been considered in the low mass region by searching for Higgs boson decays in bb or gg in association with a lepton from a decay of an accompanying W or Z boson (see Fig. 1.10) or a top quark. The search can be extended by using the vector boson fusion mode (Fig. 1.9). In vector boson fusion events the Higgs boson is accompanied by two jets in the forward regions of the detector, originating from the initial quarks that emit the vector bosons. On the other hand, central jet activity is suppressed due to the lack of colour exchange between the initial state quarks. This is in contrast to most background processes, where there is colour flow in the t-channel. Jet tagging in the forward region of the detector together with a veto of jet activity in the central region are therefore useful tools to enhance the signal-to-background ratio.

Studies have shown that the vector boson fusion production provides the only way to get access to the important $H - t\bar{t}$ mode for the Standard Model Higgs boson (see Fig.1.11). A detection of this decay mode is particularly important to extract information on the Higgs boson couplings to fermions. However, a discovery in this final state is not easy and requires integrated luminosities of the order of 30 fb^{-1} .

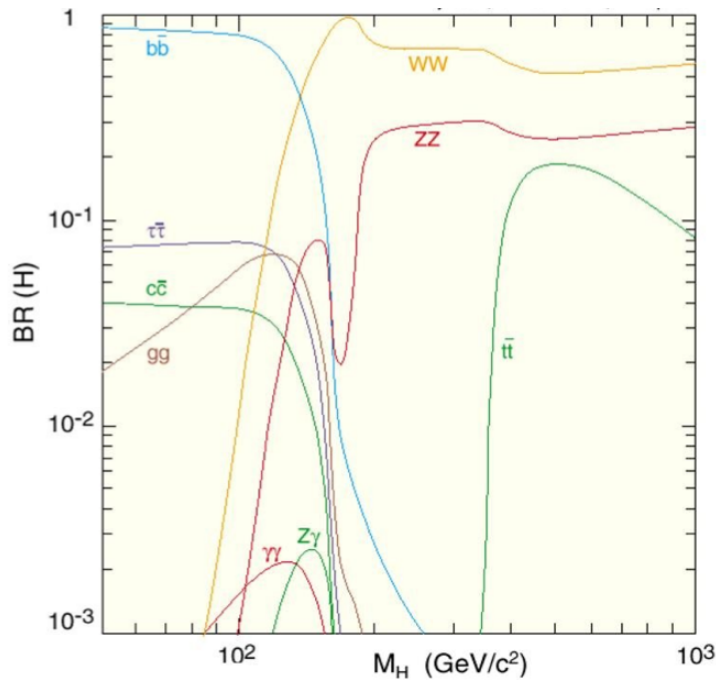


Figure 1.12: SM Higgs Decay Branching Fractions.

The sensitivity in terms of luminosity required to make a 5σ discovery in the ATLAS experiment is shown in Fig. 1.13 for the combination of a few important discovery channels. A comparable significance can be achieved in the CMS experiment. A Standard Model Higgs boson can be discovered at the LHC over the full mass range from the LEP-II lower limit up to the TeV range with a high significance. Over a large fraction of the mass range the discovery of a Standard Model Higgs boson will be possible in two or more independent channels.

Other physics beyond the Standard Model

In the absence of a scalar Higgs boson, the principal probe for the mechanism of electroweak symmetry breaking will be gauge boson (Z , W , γ) scattering at high energies. In order to explore such processes final states containing pairs of gauge bosons with large invariant mass need to be measured.

The ATLAS and CMS experiments will also be sensitive to a variety of other possible extensions of the Standard Model. Discovery limits for other phenomena are summarized below:

- One of the more noteworthy alternatives developed in the late 1970s was an entirely new type of force called a “technicolor” force. The basic idea was

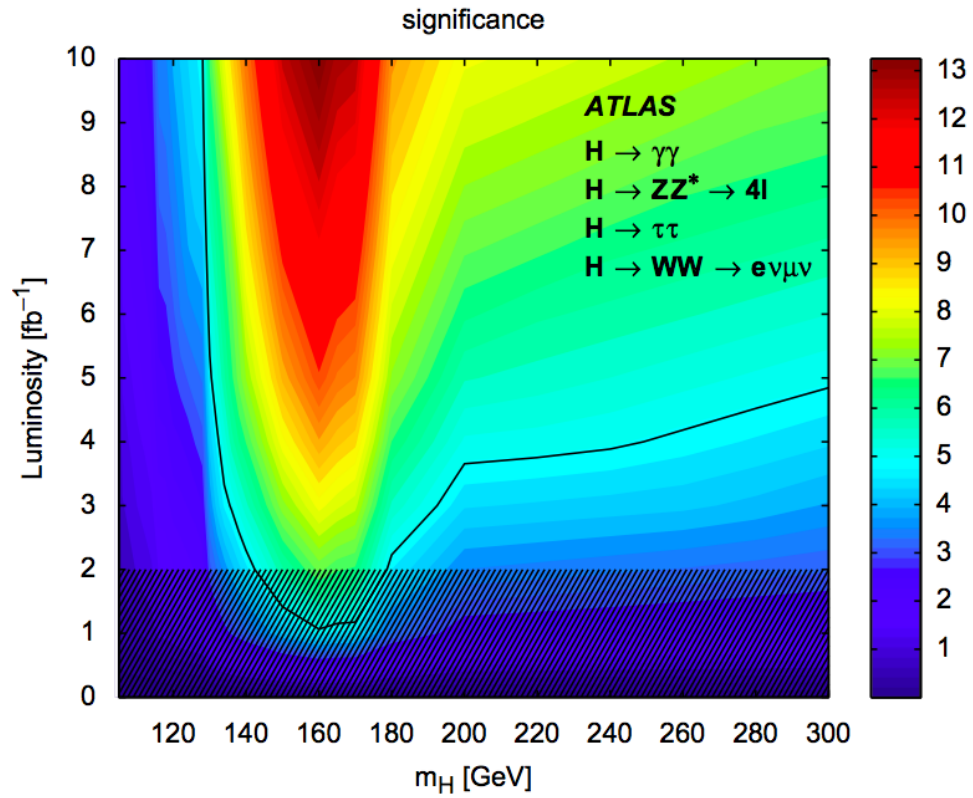


Figure 1.13: ATLAS significance contours for different Standard Model Higgs boson masses and integrated luminosities as from simulation. The thick curve represents the 5 σ discovery contour. The median significance is shown with a colour according to the legend. The hatched area below 2 fb¹ indicates the region where the approximations used in the statistical combination are not accurate, although they are expected to be conservative[32].

to construct Higgs bosons as composite particles – like mesons and hadrons – rather than assume they are elementary particles like leptons and quarks. Essentially, this idea would hypothesize a new force rather like the color force, but at a scale about a thousand times smaller. The force was called technicolor because of the analogy with the color force. Technicolor resonances can be searched for in their decays to a pair of gauge bosons, or to a techni-pion and a gauge boson. LHC sensitivity for these resonances extends up to the TeV range. Although the technicolor parameter space is very large, there is a number of potential channels which allow for combinations of signatures to help in understanding the nature of the resonances.

- According to the current knowledge quarks are not thought to be composed of anything smaller, but if one was found in an excited state, it would show this to be wrong. CMS and ATLAS can detect excited quarks up to masses in the order of 5-6 TeV/c².
- Leptoquarks are hypothetical particles that carry information between quarks and leptons given a generation and allowing quarks and leptons to interact. They are color-triplet bosons that carry both lepton and baryon numbers. They are encountered in various extensions of the Standard Model, such as technicolor theories or Grand Unification Theories[33]. The term Grand Unified Theory or GUT, refers to any of several similar models in particle physics in which at high energy scales, the three gauge interactions of the Standard Model which define the electromagnetic, weak, and strong interactions, are merged into one single interaction characterized by a larger gauge symmetry and one unified coupling constant rather than three independent ones. LHC discovery potential for first generation leptoquarks extends up to ~ 1.5 TeV/c².
- W' and Z' bosons (or W -prime and Z -prime bosons) refer to hypothetical new gauge bosons that couple to Standard Model righted-handed fermions via a right handed extension of weak isospin. They are named in analogy with the Standard Model W and Z bosons. Many GUT theories included such particles. W' and Z' should be detectable up to masses in the order of 5-6 TeV/c² with CMS and ATLAS.

1.5 Phenomenology of proton-proton collision

A proton-proton collision at very low energies can be approximated as an elastic scattering of two electrically charged extended objects. At higher energies, the structure of the proton becomes visible and plays an important role in the scattering process. So it is convenient to introduce mathematical functions of the proton, which

represent its internal structure, i.e.

$$F(x) = \sum_i f_i(x)x \quad (1.23)$$

where x is a fraction of the total momentum of the proton carried by the parton and $f_i(x)$ denotes the momentum distribution of the i^{th} type of parton, also called parton density function (PDF). The interaction between (valence) quarks is dominated by the strong interaction, which produces additional gluons and quark anti-quark pairs (see Figure 1.14). As a result the proton consists not only of three valence quarks, but also of a “sea” of further gluons and quarks. It was yet not possible to calculate the PDFs for the proton and hence the knowledge of the PDFs relies on (mainly) deep inelastic scattering experiments, such as HERA[34]. The current knowledge of the PDFs for the proton is shown in Fig. 1.15. As can be seen in the plots the sea quarks have mostly been generated by the flavour blind $g \rightarrow q\bar{q}$ splitting process. Thus the precision of our knowledge of W and Z cross sections at the LHC is crucially dependent on the uncertainty on the momentum distribution of the gluon.

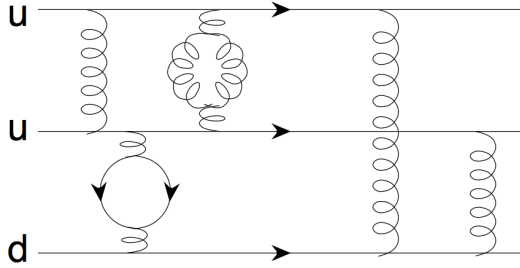


Figure 1.14: Schematic illustration of three valence quarks forming a proton and interacting via the exchange of gluons.

MSTW 2008 NNLO PDFs (68% C.L.)

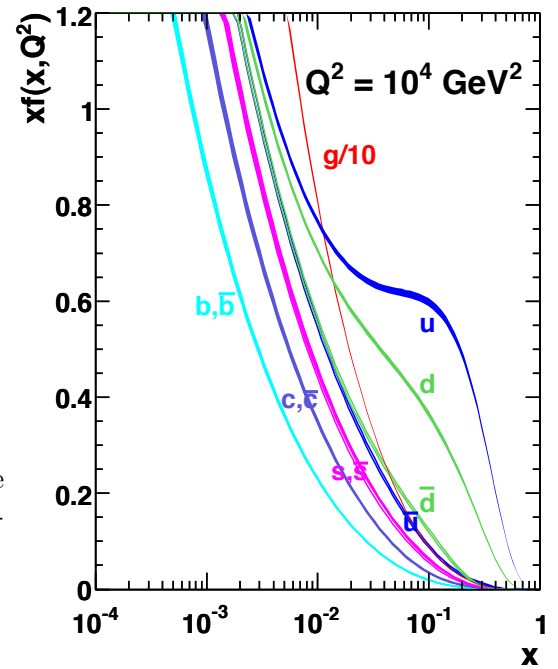


Figure 1.15: MSTW 2008 NNLO PDFs (68% C.L.) distributions at $Q^2 = 10000 \text{ GeV}^2$.

This ignorance is not the only problematic issue for the calculation of matrix elements, since higher orders in perturbation theory have not yet been calculated for all QCD processes. In fact, even the calculation of lower orders is mathematically very demanding and only a few processes have been calculated so far to second order perturbation theory. Hence, several simplifications and approximations must be applied to describe QCD interactions. The basis for a theoretical description is the so-called factorization theorem. In a naive interpretation, the theorem states that the hard QCD interaction of two protons can be split up in several stages, as illustrated in Fig. 1.16.

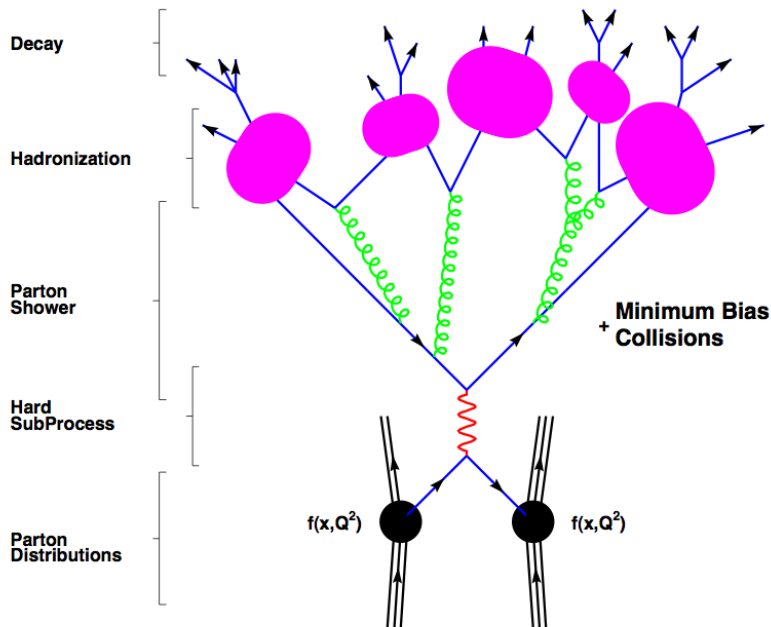


Figure 1.16: Illustration of an hard scattering process of two partons.

The hard scattering describes the actual sub-process between two partons. The calculation of this process is at least available at tree-level, i.e. in leading order perturbation theory, for few cases also for higher orders. All those orders of perturbation theory, not included in the calculation of the hard scattering process, must be approximated for the initial state and final state partons. Since the partons are electrical and/or color charged, they are accompanied by the emissions of gluons and photons. The emission of gluons dominates for hadronic interactions. These perturbative corrections are approximated by the so called parton shower approach. Each radiation process is simulated by the branching of a parton into a parton with lower

energy plus an emitted gluon, i.e. $q \rightarrow qg$. The energy distribution between the two daughter particles can be modeled with the Altarelli-Parisi splitting function[35]. The two newly created particles might again undergo a branching. The shower evolution is stopped when reaching a fixed energy scale of the branching parton, which is usually around 1 GeV, i.e. well above the confinement regime. So far, the above description of the scattering process leaves colored objects in the final state, which cannot be the final state particles. The process in which colorless hadrons are formed from colored objects is called hadronization and involves non-perturbative processes which cannot be described analytically. Two different phenomenological approaches are used to describe the hadronization and thus the confinement of colored objects (see also Figure 1.17).

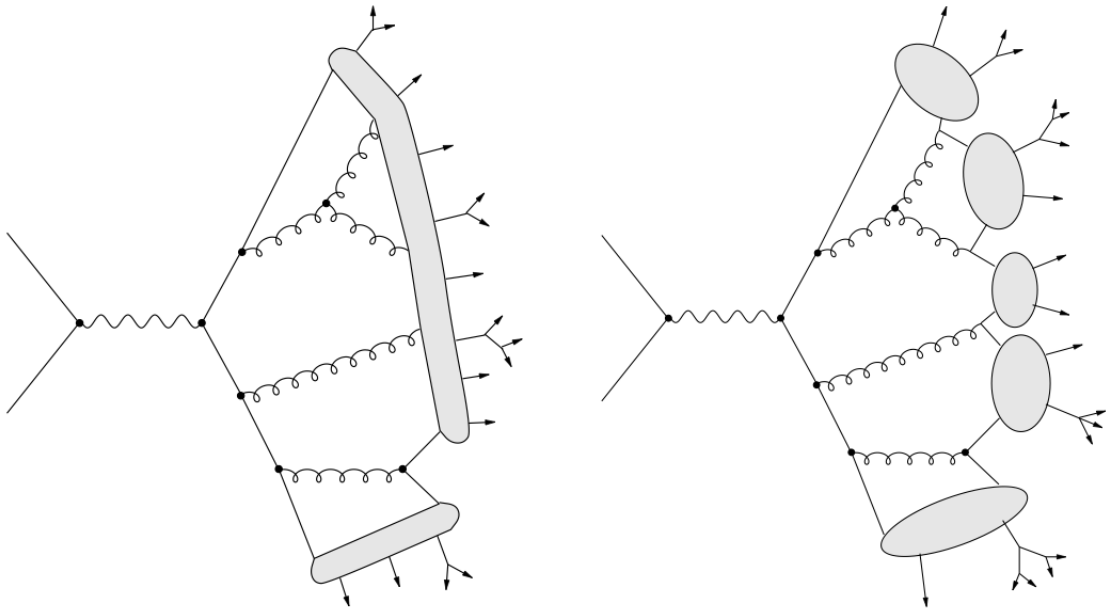


Figure 1.17: Parton shower with schematic string (left) and cluster (right) hadronization models.

- **String model:** This model is inspired by the string model of strong interaction. It was formulated prior to the quark and QCD models as well as modern lattice QCD calculations, which predict a linear rising potential between a quark q and an antiquark \bar{q} with increasing distance, caused - in a naive physical picture - by a colored flux tube or string due to the self interaction of gluons. If the energy stored in the string is large enough a second pair of quarks q' and \bar{q}' is produced, which form two color singlets $q\bar{q}'$ and $q'\bar{q}$. The breaking up of strings

stops, if only on-mass-shell hadrons remain. It should be noted, that the di-quark anti-di-quark pair-production allows also the production of baryons in this picture.

- **Cluster model:** In a first step, all gluons of the parton showers are split into light quark-antiquark or diquark-antidiquark pairs. In a second step, all quarks are combined with their nearest neighbors to form color singlet clusters. Basically, the mass of these clusters is used to determine in which hadrons they will finally decay.

1.6 Monte Carlo Generator

The purpose of Monte Carlo (MC) event generators is to simulate the theoretical models of physics processes, e.g.: the production of a Z boson in a proton-proton collision and its decay into two muons. Such a theoretical prediction is crucial to understand the measured data and to tune physics analyses, so both in the preparation stage of future experiments than during the data taking. The factorisation scheme, which was introduced in Section 1.5, is the basis of MC event generators, which rely at least partially on QCD. In case of LHC, MC generators must describe the structure of hadrons, the parton showers, the actual hard scattering process, and the hadronization. Various MC event generator programs use different approximations during the different steps and therefore the theoretical prediction relies at least partially on the choice of the underlying Event Generator. Some event generator programs, which have been used in this thesis, are outlined in the following:

- **PYTHIA:** PYTHIA is a general purpose event generator[36], which is commonly used in High Energy Physics, because of its easy handling and relatively large predictive power. It can simulate lepton-lepton, lepton-hadron and hadron-hadron interactions with a broad field of theoretical models, including supersymmetric models or models with leptoquarks. The hard scattering process is calculated in leading order approximation (LO). The higher order corrections are approximated with the parton shower approach. The hadronization process is based on the String-Model as introduced briefly in the previous Section.
- **HERWIG:** The Hadron Emission Reactions With Interfering Gluons (Herwig) program is also a general purpose event generator[37], which is quite similar to PYTHIA. The main differences are the modelling of the parton shower and the hadronization process. Herwig is based on the clusterization model.
- **MC@NLO:** This Monte Carlo event generator includes full next-to-leading-order calculations of rates of QCD processes during the hard scattering pro-

cess [38]. The output of the simulation is further processed by the Herwig Event generator, which adds higher order approximations of the parton shower and simulates the hadronization step. In the final step one compute the difference between exact next-to-leading order (NLO) formula and the previous step: this difference may be negative, so some negative weighted events may appear (actually up to 20% for $b\bar{b}$)

- RESBOS: this event generator[39] computes the fully differential cross section

$$\frac{d\sigma}{dp_T^B dy_B d(p^B)^2 d\Omega} \quad (1.24)$$

for processes $pp \rightarrow B \rightarrow l^+l^-$ with next-to-leading initial state QCD corrections. The label B indicates a boson, p_T its transverse momentum and y its rapidity. The calculation is based on the CSS resummation formalism[40] and contains the exact matrix element, including initial state soft gluon resummation effects, for the production of the specified boson. The hadronization is then modeled by the Herwig generator.

- POWHEG: POSitive Weight Hardest Emission Generator[41] is an event generator optimized for heavy quark production in hadronic collisions. It is accurate at the NLO, and it can be interfaced to Shower Monte Carlo programs like HERWIG and PYTHIA, in such a way that both the leading logarithmic accuracy of the shower and the NLO accuracy are maintained in the output. It is thus an alternative to the MC@NLO heavy flavour production program, with the advantage that one does not need to deal with negative event weights.

Chapter 2

The LHC and the CMS experiment

The Standard Model predictions have been confirmed so far by all the experimental tests, but still theoretical physicists have motivations to believe that the theory is not a fundamental theory. Standard Model is considered rather an effective field theory which predicts the particles and forces behaviour up to an energy scale $\Lambda \sim \text{TeV}$. The Higgs mechanism is mathematical consistent up to an energy scales above about 1 TeV, but needs experimental verifications. However, as already mentioned in the previous Chapter, there are alternative theories that invoke more symmetries such as supersymmetry or invoke new forces or constituents such as strongly-broken electroweak symmetry, technicolour, etc. An as yet unknown mechanism is also possible. Furthermore there are high hopes for discoveries that could pave the way towards a unified theory. These discoveries could take the form of supersymmetry or extra dimensions, the latter requiring modification of gravity at the TeV scale. Hence there are many compelling reasons to investigate the TeV energy scale (see Section 1.4 and mentioned references).

The beam energy (7 TeV) and the design luminosity $\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$) of the LHC have been chosen in order to study physics at the TeV energy scale. Hence a wide range of physics is potentially accessible with the seven-fold increase in energy and a hundred-fold increase in integrated luminosity with respect to the current hadron collider experiments at TeVatron[42]. These conditions also require a very careful design of the detectors.

The focus of this Chapter is to present a concise yet complete overview of the features of the Compact Muon Solenoid (CMS) experiment[43] which allow it to exploit the physics opportunities presented by the LHC[44]. The LHC machine characteristics are first discussed briefly.

2.1 The LHC Machine

The Large Hadron Collider (LHC) is a two ring superconducting hadron accelerator and collider installed in the existing 26.7 km tunnel that was constructed between 1984 and 1989 to host the CERN LEP machine. The LEP tunnel, sketched in Fig. 2.1 and 2.2, has eight straight sections and eight arcs and lies between 45 m and 170 m below the ground on a plane inclined at 1.4% sloping towards the Léman lake.

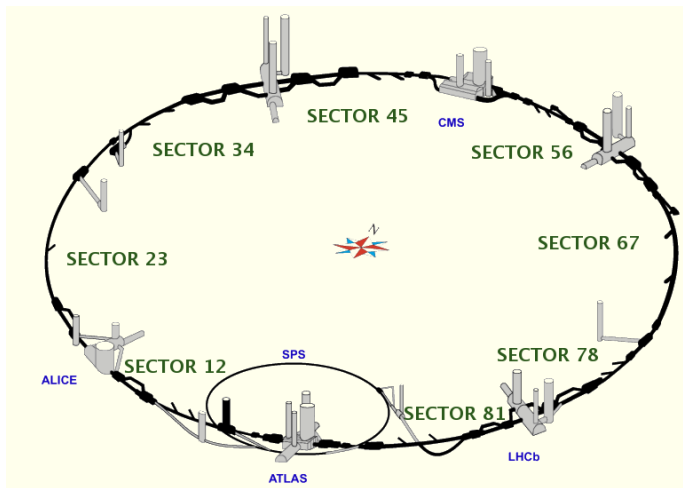


Figure 2.1: LHC sectors.

The LHC design[45] is linked with the latest results in the accelerator technology. The past SpS and the present TeVatron hadron colliders accelerate proton anti-proton beams in the same ring. LHC instead is a proton-proton collider and so it contains two rings with counter-rotating beams. Along the LHC ring there are four interaction points, where the caverns of the four experiment ALICE, ATLAS, CMS, LHCb, are installed

2.1.1 Performance goal

The centre of mass energies energy of 14 TeV is foreseen to be reached not before 2013. The first collision at the LHC has been delivered in autumn 2009 at a center of mass energy of 900 GeV and 2.136 TeV. From March 2010 LHC is operating at 7 TeV, the highest collision energy ever reached by a particle accelerator, and it is supposed to continue to work at this regime till end of 2011 and to obtain a physic reach comparable with the TeVatron[42]. The number of events per second generated in the LHC collisions is given by:

$$N_{event} = \mathcal{L}\sigma_{event} \quad (2.1)$$

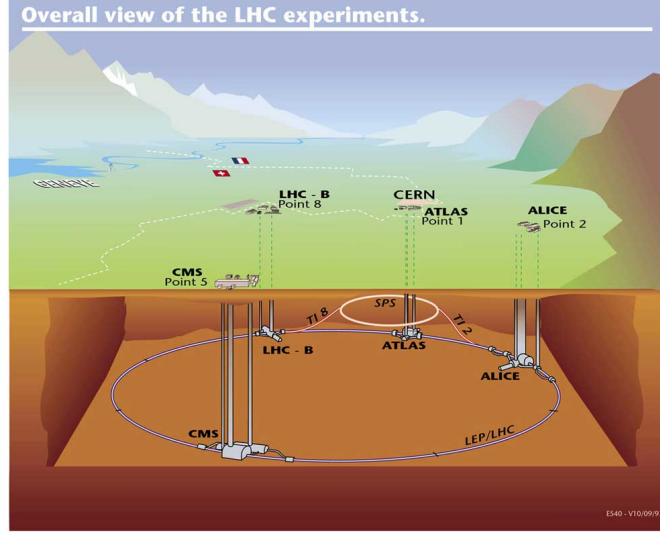


Figure 2.2: Overall view of the LHC experiments.

where σ_{event} is the cross section for the event type under study and \mathcal{L} the machine luminosity. The machine luminosity depends only on the beam parameters and can be written for a Gaussian beam distribution as:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} \quad (2.2)$$

where

- N_b is the number of particles per bunch,
- n_b the number of bunches per beam,
- f_{rev} the revolution frequency,
- γ_r the relativistic gamma factor,
- ϵ_n is the normalized transverse beam emittance, (with a design value of $3.75 \mu\text{m}$). The beam emittance is defined as the products of the particles positions distribution RMS, σ , times the RMS of the particle momenta distribution, σ' . The beam emittance is constant for all the beam life and the goal of the injection procedures is to introduce in the collider a beam with the lowest beam emittance.

- β^* , the beta function at the collision point, defined as the ratio σ/σ' . Contrary to ϵ , β^* can be reduced along the ring length, with a dedicated magnetic optics which squeezes the beams in the Interaction Point (IP). Note that a lower β^* means a higher spread in the particle momenta.
- F is the geometric luminosity reduction factor due to the crossing angle at the IP:

$$F = \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2\right)^{\frac{1}{2}} \quad (2.3)$$

where θ_c is the full crossing angle at the IP, σ_z the RMS bunch length, and σ^* the transverse RMS beam size at the IP.

The above expression assumes round beams, with $\sigma_z \ll \beta$, and with equal beam parameters for both beams. The exploration of rare events in the LHC collisions therefore requires both high beam energies and high beam intensities.

The LHC has two high luminosity experiments, ATLAS[46] and CMS[47], both aiming at a peak luminosity of $\mathcal{L} = 10^{34} \text{cm}^2 \text{s}^{-1}$ for proton operation. There are also two low luminosity experiments: LHCb[48] for B-physics, aiming at a peak luminosity of $\mathcal{L} = 10^{32} \text{cm}^2 \text{s}^{-1}$, and TOTEM[49] for the detection of protons from elastic scattering at small angles. In addition to the proton beams, the LHC will also be operated with heavy ion beams. The LHC has one dedicated ion experiment, ALICE[50], aiming at a peak luminosity of $\mathcal{L} = 10^{27} \text{cm}^2 \text{s}^{-1}$ for nominal lead-lead ion operation.

The high beam intensity required for a proton-proton collision luminosity of $\mathcal{L} = 10^{34} \text{cm}^2 \text{s}^{-1}$ excludes the use of anti-proton beams, and hence excludes the particle-anti-particle collider configuration of a common vacuum and magnet system for both circulating beams, as used for example in the Tevatron. To collide two counter-rotating proton beams opposite magnetic dipole fields in both rings are required. The LHC is therefore designed as a proton-proton collider with separate magnet fields and vacuum chambers in the main arcs and with common sections only at the insertion regions and where the experimental detectors are located. The two beams share an approximately 130 m long common beam pipe along the Interaction Regions (IRs). Together with a designed large number of bunches (2808 for each proton beam), and a nominal bunch spacing of 25 ns, the long common beam pipe implies 34 parasitic collision points at each experimental insertion region. Dedicated crossing angle orbit bumps separate the two LHC beams left and right from the IP in order to avoid collisions at these parasitic collision points. The LHC uses twin bore magnets that consist of two sets of coils and beam channels within the same mechanical structure and cryostat. The peak beam energy depends on the integrated dipole field around the storage ring, which implies a peak dipole field of 8.33 T for the 7 TeV in the LHC machine and the use of superconducting magnet technology.

2.1.2 Injection chain

The LHC will be supplied with protons from the injector chain Linac2 – Proton Synchrotron Booster (PSB) – Proton Synchrotron (PS) – Super Proton Synchrotron (SPS), as shown in Figure 2.3.

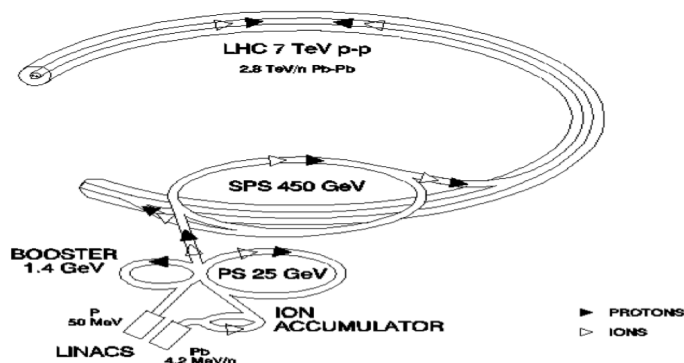


Figure 2.3: The LHC hadron injector complex.

These accelerators were existing at CERN from past experiments and were upgraded to meet the very stringent needs of the LHC: many high intensity proton bunches (design value: 2808 per LHC ring) with small transverse and well defined longitudinal emittances. The main challenges for the PS complex are (i) the unprecedented transverse beam brightness (intensity/emittance), and (ii) the production of a bunch train with the LHC spacing of 25 ns before the extraction from the PS (25 GeV). The transverse emittances of the LHC beam have to be maintained at their unusually small size throughout the injector chain. Small amounts of mis-steering and mismatch between the accelerators of the chain, virtually negligible for normal operation, become increasingly important and their effect has to be measurable, calling for high-resolution beam profile monitors.

The Equation for luminosity 2.2 summarizes the end-point beam requirements, but this implies many conditions that have to be satisfied, such as, the beam emittance must fit the small aperture of the LHC superconducting magnets; the beam intensity is limited by the synchrotron radiation that has to be absorbed by the cryogenic system; the beam-beam effect causes a spread in betatron tunes when the beams are colliding, and this has to be kept below a certain limit; and the space-charge limits in the injectors have to be respected. There are also conflicting requirements for the longitudinal emittance. An optimization procedure, taking into account these boundary conditions, has resulted in the LHC beam parameter set compiled in Table 2.1. During the first year of physics running, the LHC has operated at a much lower

intensity and luminosity level, see Section 2.1.5 The ultimate performance level in the table corresponds to the LHC beam-beam limit.

| | | | |
|------------------------------|---------------|----------------------|--------------------------------|
| Ultimate proton energy | E | 7 | TeV |
| Dipole field at 7 TeV | B | 8.33 | T |
| Ultimate luminosity | \mathcal{L} | 10^{34} | $\text{cm}^{-2} \text{s}^{-1}$ |
| Bunch spacing | | 25 | ns |
| # bunch | k_B | 2808 | |
| Ultimate intensity per bunch | N_p | $1.15 \cdot 10^{11}$ | |
| Beam current | I_b | 0.58 | A |
| Collisions | | | |
| β at IP | β^* | 0.55 | m |
| Beam radius RMS at IP | σ^* | 16.7 | μm |
| Luminosity mean life | τ_L | 15 | h |
| # collisions/crossing | n_c | ≈ 20 | |

Table 2.1: LHC ultimate proton beam and collision parameters.

As with the PS complex, the SPS is an old machine and is not optimised as an LHC injector. SPS is able at maximum to accelerate $\approx 4 \times 10^{13}$ protons per cycle. That limits the number of PS pulses per SPS cycle to a maximum of four. The total bunch length has to be below 4 ns to fit into the buckets of the SPS 200 MHz accelerating system, implying a longitudinal emittance of 0.35 eVs per PS bunch.

2.1.3 Luminosity lifetime

The luminosity in the LHC is not constant over a physics run, but decays due to the degradation of intensities and emittances of the circulating beams. The main cause of the luminosity decay during nominal LHC operation is the beam loss from collisions. The initial decay time of the bunch intensity, due to this effect, is:

$$\tau_{nuclear} = \frac{N_{tot;0}}{\mathcal{L}\sigma_{tot}k} \quad (2.4)$$

where $N_{tot;0}$ is the initial beam intensity, \mathcal{L} the initial luminosity, σ_{tot} the total cross section ($\sigma_{tot} = 10^{25} \text{cm}^{-2}$ at 14 TeV) and k the number of IPs. Assuming a peak luminosity of $\mathcal{L} = 10^{34} \text{cm}^{-2} \text{s}^{-1}$ and two high luminosity experiments, the above expression yields an initial decay time of $t = 44.85$ h. Equation 2.4 results in the following decay of the beam intensity and luminosity as functions of time:

$$N_{tot}(t) = \frac{N_{tot;0}}{1 + t/\tau_{nuclear}} \quad (2.5)$$

$$\mathcal{L}(t) = \frac{\mathcal{L}(0)}{(1 + t/\tau_{nuclear})^2}. \quad (2.6)$$

The time required to reach 1/e of the initial luminosity is given by:

$$t_{1/e} = (\sqrt{e} - 1)\tau, \quad (2.7)$$

yielding a luminosity decay time of $\tau_{nuclear;1/e} = 29$ h. Emittance blow-up is responsible for particles and eventually beam losses. Emittance blow-up can be caused by the scattering of particles on residual gas, the nonlinear force of the beam-beam interaction, Radio Frequency (RF) noise, and intra-beam scattering (IBS) scattering effects. Following the arguments set out in the Pink Book (1991 Design Study[51]), it is shown that LHC parameter adjustments can cancel the beam blow up due to the beam-beam interactions and RF noise. Approximating further the decay by an exponential process, the net luminosity lifetime can be estimated as:

$$\frac{1}{\tau_L} = \frac{1}{\tau_{IBS}} + \frac{1}{\tau_{resgas}} + \frac{1}{\tau_{nuclear;1/e}} \quad (2.8)$$

Assuming an IBS time constant of 80 hour and a rest gas time constant of 100 hour with the above nuclear decay time gives a net estimate of the luminosity lifetime of, $\tau_L = 14.9h$.

2.1.4 Integrated luminosity

Integrating the luminosity over one run yields,

$$\mathcal{L}_{int} = \mathcal{L}_0\tau_L[1 - e^{-T_{run}/\tau_L}], \quad (2.9)$$

where T_{run} is the total length of the luminosity run. The overall collider efficiency depends on the ratio of the length of the run to the average turnaround time. Assuming the machine can be operated for 200 days per year and assuming the luminosity lifetime of 15 hours obtained earlier, the optimum run time is 12 hours or 5.5 hours, for the average turnaround times of 7 hours and 1.15 hours, respectively. This leads to a maximum total integrated luminosity per year of 80 fb^{-1} to 120 fb^{-1} , depending on the average turnaround time of the machine.

2.1.5 LHC performance in 2010

LHC machine performed smoothly during the 7 TeV operation in 2010 increasing little by little the instantaneous luminosity up to $\mathcal{L} = 2 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$ reached in October 2010 (see Fig. 2.4). LHC operated proton-proton collision until the beginning of November 2010, when the heavy ion fill started. The proton-proton integrated luminosity delivered by LHC is almost equivalent to 50 pb^{-1} .

The filling scheme has evolved frequently and at the beginning of November 2010 was 248 bunches, with a spacing of 150 ns. Figure 2.5 shows the integrated luminosity produced by LHC in 2010 for the proton-proton operation.

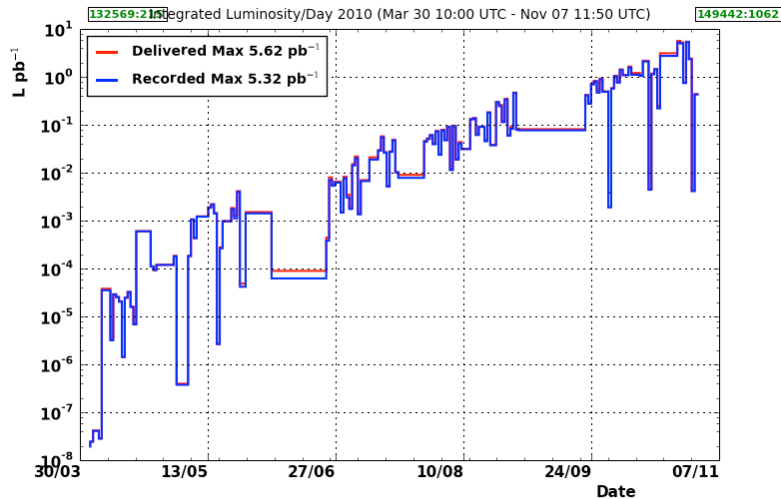


Figure 2.4: Maximum Instantaneous luminosity per day delivered to (red) CMS during stable beams at 7 TeV centre-of-mass energy.

2.2 The CMS experiment

The Compact Muon Solenoid (CMS) detector is a multi-purpose apparatus installed in one of the IP at the Large Hadron Collider (LHC) at CERN. CMS is installed about 100 metres underground close to the French village of Cessy, between Lake Geneva and the Jura mountains. Its installation was completed in 2008 and several test with cosmic rays has been carried on during the last three year before the exiting period started in 2009 with LHC collisions.

The designed LHC beam and collision conditions require a very careful design of the detectors. The total proton-proton cross section at $\sqrt{s} = 14$ TeV is expected to be about 100 mb. At design luminosity the general-purpose detectors will therefore observe an event rate of approximately 10^9 inelastic events/s. This leads to a great number of 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

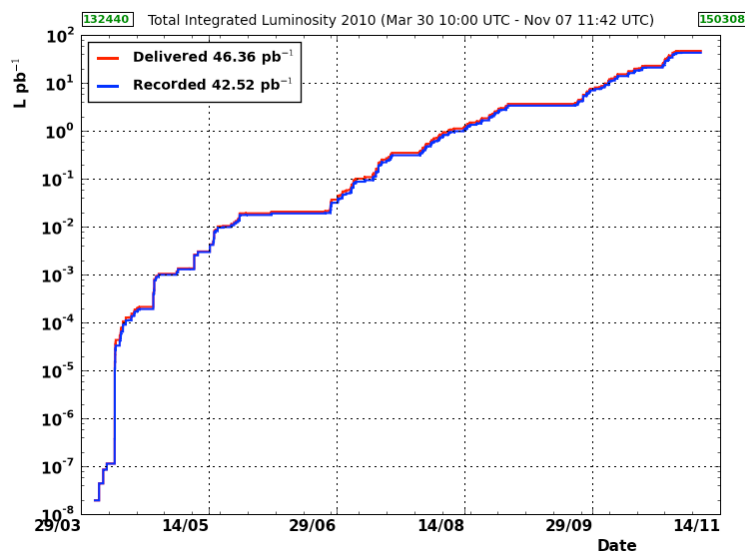


Figure 2.5: Integrated luminosity versus time delivered to (red), and recorded by CMS (blue) during stable beams at 7 TeV centre-of-mass energy in 2010.

the read-out and trigger systems.

At the design luminosity, a mean 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 (Pile-up). This problem clearly 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 with good time resolution, resulting in low occupancy. This requires 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, requiring 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:

- Good muon identification and momentum resolution over a wide range of momenta and angles, good di-muon mass resolution ($\approx 1\%$ at 100 GeV), and the ability to determine unambiguously the charge of muons with $p_T < 1$ TeV;
- Good charged-particle momentum resolution and reconstruction efficiency in the inner tracker. Efficient triggering and offline tagging of τ 's and b-jets,

requiring pixel detectors close to the interaction region;

- Good electromagnetic energy resolution, good di-photon and di-electron mass resolution ($\sim 1\%$ at 100 GeV), wide geometric coverage, π^0 identification and eventually rejection, and efficient photon and lepton isolation at high luminosities;
- Good missing-transverse-energy and jet energy resolution, requiring hadron calorimeters with a large hermetic geometric coverage and with fine lateral segmentation

The coordinate system adopted by CMS has the origin centered at the nominal collision point inside the experiment, the y axis pointing vertically upward, and the x axis pointing radially inward toward the center of the LHC. Thus, the z axis points along the beam direction toward the Jura mountains from LHC Point 5. The azimuthal angle ϕ is measured from the x axis in the $x - y$ plane and the radial coordinate in this plane is denoted by r . The polar angle θ is measured from the z axis. Pseudorapidity is defined as $\eta = -\ln \tan \theta/2$. Thus, the momentum and energy transverse to the beam direction, denoted by p_T and E_T , respectively, are computed from the x and y components. The imbalance of energy measured in the transverse plane is denoted by \cancel{E}_T .

At the core of the CMS detector, illustrated in Fig. 2.6 sits a high-magnetic field and 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 pseudorapidity coverage to high values ($|\eta| < 5$) assuring very good hermeticity. The overall dimensions of the CMS detector are a length of 21.6 m, a diameter of 14.6 m and a total weight of 12500 t.

2.2.1 Superconducting magnet

The superconducting magnet for CMS has been designed to reach a 4 T field in a free bore of 6 m diameter and 12.5 m length with a stored energy of 2.6 GJ at full current. The flux is returned through a 10000 t yoke comprising 5 wheels and 2 endcaps, composed of three disks each (Figure 2.7). The distinctive feature of the 220 t cold mass is the 4 layer winding made from a stabilised reinforced NbTi conductor. The ratio between stored energy and cold mass is high (11.6 kJ/kg), causing a large mechanical deformation (0.15%) during energising, well beyond the values of previous solenoidal detector magnets (see Fig. 2.8). In order to provide the necessary hoop strength, a large fraction of the CMS coil must have a structural function[52]. The

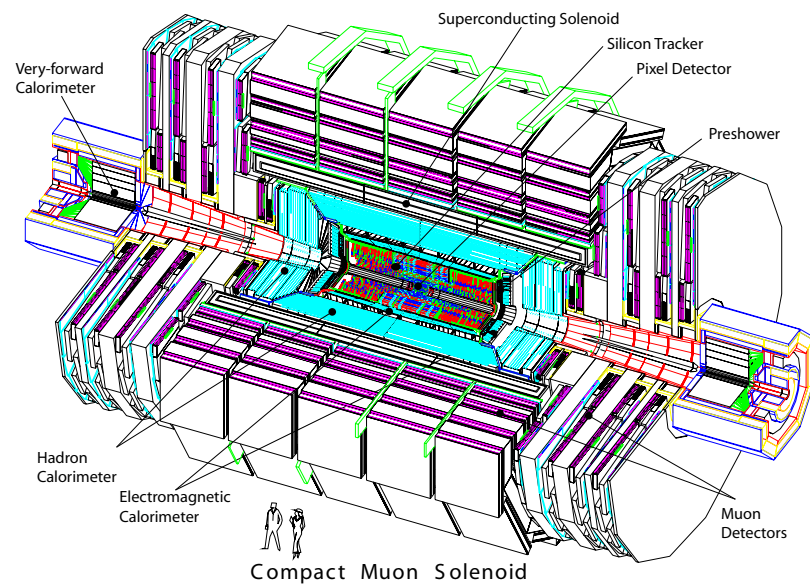


Figure 2.6: Perspective view of the CMS experiment.

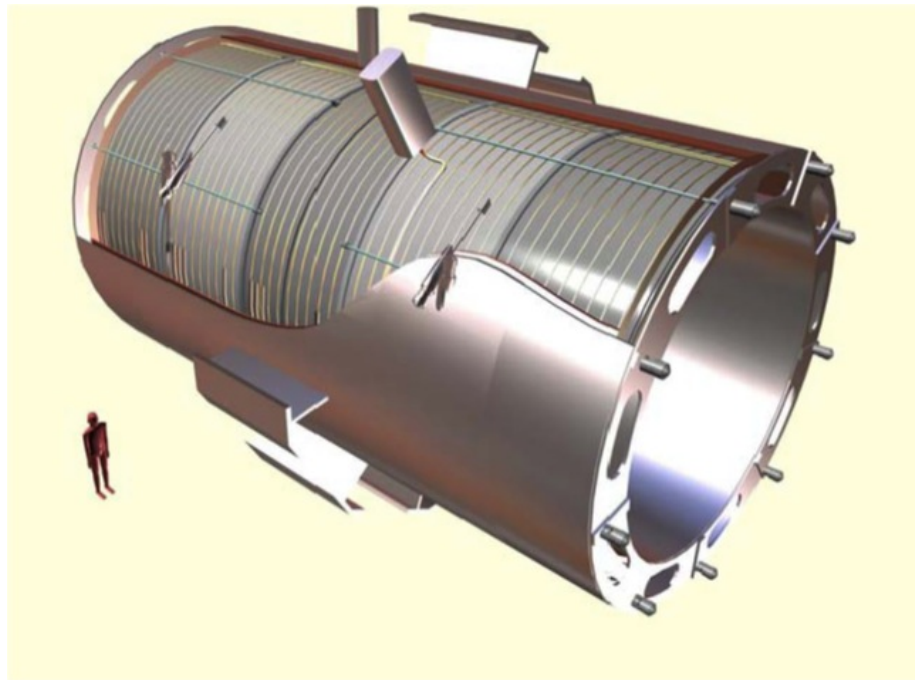


Figure 2.7: General artistic view of the 5 modules composing the cold mass inside the cryostat, with details of the supporting system (vertical, radial and longitudinal tie rods).

CMS Magnet has been fully and successfully tested and commissioned during autumn 2006.

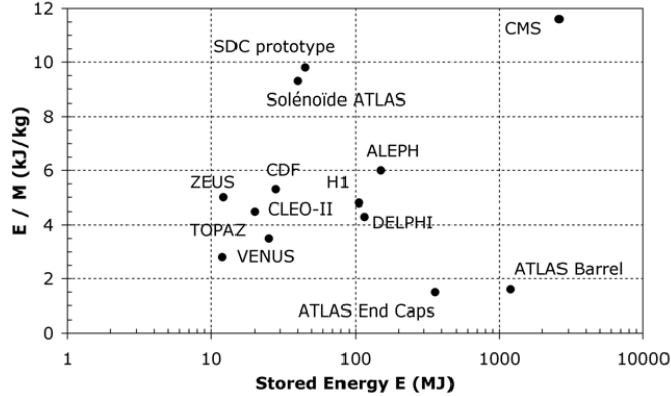


Figure 2.8: The energy over mass ratio E/M , for several detector magnets.

2.2.2 Inner tracking system

The inner tracking system of CMS is designed to provide a precise and efficient measurement of the trajectories of charged particles emerging from the LHC collisions. In addition a precise measurement of secondary vertexes and impact parameters is necessary for the efficient identification of heavy flavours decays which are produced in many of the interesting physics channels. Together with the electromagnetic calorimeter and the muon system the tracker has to identify electrons and muons. Tau leptons are searched in several discovery channels and need to be reconstructed in one-prong and three-prong decay topologies. In order to reduce the event rate from the LHC bunch crossing rate of 40 MHz to about 100 Hz which can be permanently stored, tracking information is heavily used in the high level trigger of CMS.

The tracking system surrounds the interaction point and has a length of 5.8 m and a diameter of 2.5 m. The CMS solenoid provides a homogeneous magnetic field of 4 T over the full volume of the tracker. At the LHC design luminosity of $10^{34} \text{cm}^{-2} \text{s}^{-1}$ there will be on average about 1000 particles from more than 20 overlapping proton-proton interactions traversing the tracker for each bunch crossing, i.e. every 25 ns. Therefore a detector technology featuring high granularity and fast response is required, such that the trajectories can be identified reliably and attributed to the correct bunch crossing. However, these features imply a high power

density of the on-detector electronics which in turn requires efficient cooling. This is in direct conflict with the aim of keeping to the minimum the amount of material in order to limit multiple scattering, bremsstrahlung, photon conversion and nuclear interactions. A compromise had to be found in this respect. The intense particle flux will also cause severe radiation damage to the tracking system. The main challenge in the design of the tracking system was to develop detector components able to operate in this harsh environment for an expected lifetime of about 10 years. These requirements on granularity, speed and radiation hardness lead to a tracker design entirely based on silicon detector technology. The CMS tracker is composed of a pixel detector with three barrel layers at radii between 4.4 cm and 10.2 cm and a silicon strip tracker with 10 barrel detection layers extending outwards to a radius of 1.1m. Each system is completed by endcaps which consist of 2 disks in the pixel detector and 3 plus 9 disks in the strip tracker on each side of the barrel, extending the acceptance of the tracker up to a pseudorapidity of $|\eta| < 2.5$. With about 200 m² of active silicon area the CMS tracker is the largest silicon tracker ever built[53] [54].

The operating conditions for a tracking system at the LHC are very challenging. As already mentioned, each LHC bunch crossing at design luminosity creates on average about 1000 particles hitting the tracker. This leads to a hit rate density of 1 MHz/mm² at a radius of 4 cm, falling to 60 kHz/mm² at a radius of 22 cm and 3 kHz/mm² at a radius of 115 cm. In order to keep the occupancy at or below 1% pixel detectors have to be used at radii below 10 cm. For a pixel size of 100×150 μm² in $r-\phi$ and z , respectively, which is driven by the desired impact parameter resolution, the occupancy is of the order 10⁻⁴ hits per pixel and LHC bunch crossing. At intermediate radii (20 cm < r < 55 cm) the reduced particle flux allows the use of silicon micro-strip detectors with a typical cell size of 10 cm×80 μm, leading to an occupancy of up to 2-3% hits per strip and LHC bunch crossing. In the outer region (55 cm < r < 110 cm) the strip pitch can be further increased. Given the large areas that have to be instrumented in this region, also the strip length has to be increased in order to limit the number of read-out channels. However, the strip capacitance scales with its length and therefore the electronics noise is a linear function of the strip length as well. In order to maintain a good signal to noise ratio of well above 10, CMS uses thicker silicon sensors for the outer tracker region with correspondingly higher signal (500 μm thickness as opposed to the 320 μm in the inner tracker). CMS is the first experiment using silicon detectors in this outer tracker region.

The radiation damage introduced by the high particle fluxes at the LHC interaction regions is a severe design constraint. Table 2.2 shows the expected fast hadron fluence (flux integrated over time) and radiation dose in the CMS barrel tracker for an integrated luminosity of 500 fb⁻¹ corresponding to about 10 years of LHC operation. Neutrons generated by hadronic interactions in the ECAL crystals make up a substantial contribution to the fast hadron fluence, which actually dominates in the outer tracker close to the ECAL surface.

| Radius (cm) | Fluence of fast hadrons (10^{14} cm^{-2}) | Dose (kGy) | Charged particle flux ($\text{cm}^{-2}\text{s}^{-1}$) |
|-------------|---|------------|---|
| 4 | 32 | 840 | 10^8 |
| 11 | 4.6 | 190 | |
| 22 | 1.6 | 70 | 6×10^6 |
| 75 | 0.3 | 7 | |
| 115 | 0.2 | 1.8 | 3×10^5 |

Table 2.2: Expected hadron fluence and radiation dose in different radial layers of the CMS tracker (barrel part) for an integrated luminosity of 500fb^{-1} (≈ 10 years).

A schematic drawing of the CMS tracker is shown in Figure 2.9. At radii of 4.4, 7.3 and 10.2 cm, three cylindrical layers of hybrid pixel detector modules surround the interaction point. They are complemented by two disks of pixel modules on each side.

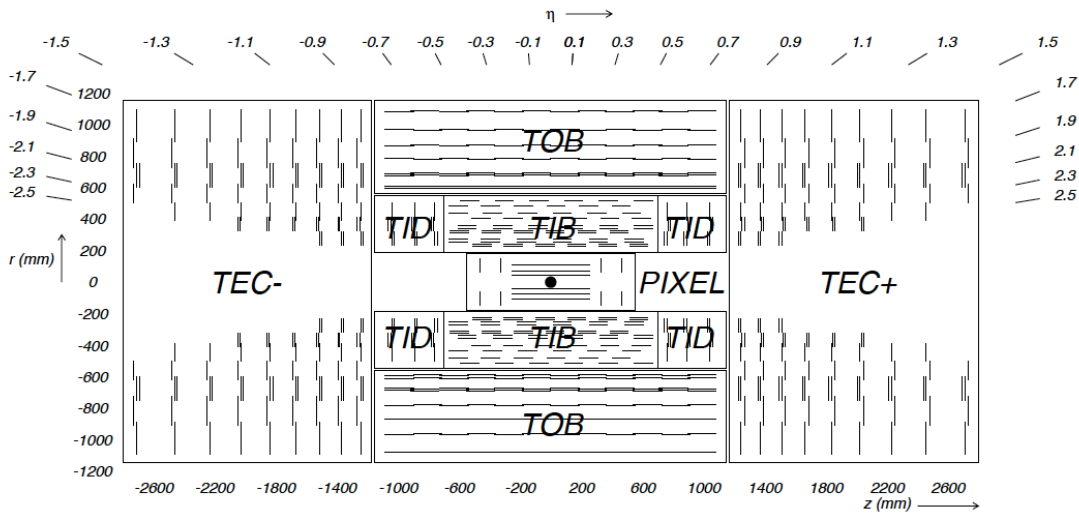


Figure 2.9: Schematic cross section through the CMS tracker. Each line represents a detector module. Double lines indicate back-to-back modules which deliver stereo hits.

The pixel system is the part of the tracking system that is closest to the interaction region. It contributes precise tracking points in $r-\phi$ and z and therefore is responsible for a small impact parameter resolution that is important for good secondary vertex reconstruction.

The pixel detector delivers three high precision space points on each charged

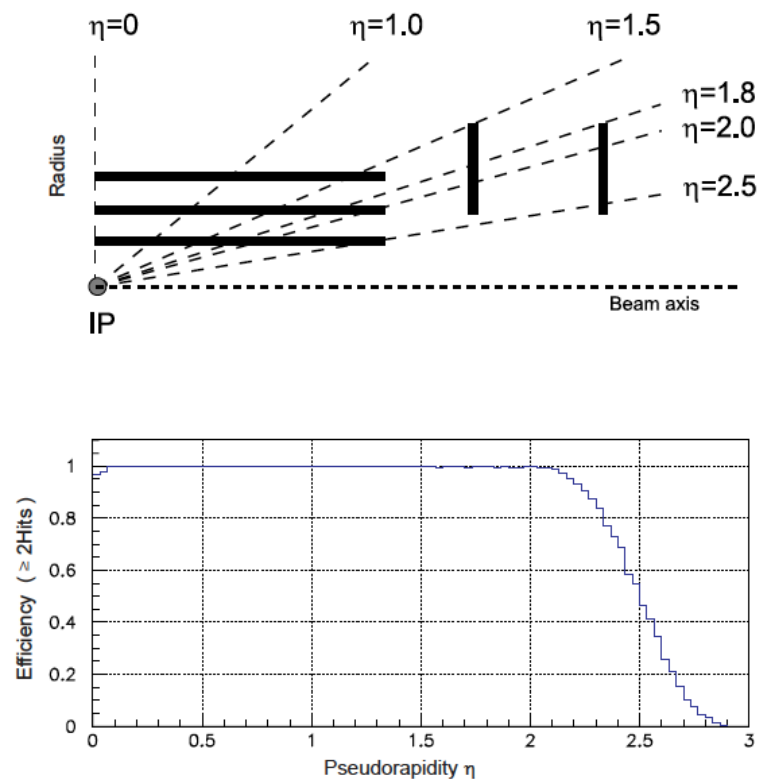


Figure 2.10: Geometrical layout of the pixel detector and hit coverage efficiency as a function of pseudorapidity.

particle trajectory, see Fig. 2.10. In total the pixel detector covers an area of about 1m^2 and has 66 million pixels.

With a pixel cell size of $100 \times 150 \mu\text{m}^2$ emphasis has been put on achieving similar track resolution in both $r - \phi$ and z directions. Through this a 3D vertex reconstruction in space is possible, which will be important for secondary vertexes with low track multiplicity. The pixel system has a zero-suppressed read out scheme with analog pulse height read-out. This improves the position resolution due to charge sharing and helps to separate signal and noise hits as well as to identify large hit clusters from overlapping tracks.

The pixel detector covers a pseudorapidity range $-2.5 < \eta < 2.5$, matching the acceptance of the central tracker. The pixel detector is essential for the reconstruction of secondary vertexes from b and tau decays, and forming seed tracks for the outer track reconstruction and high level triggering. It consists of three barrel layers (BPix) with two endcap disks (FPix). The 53 cm long BPix layers will be located at mean radii of 4.4, 7.3 and 10.2 cm. The FPix disks extending from ~ 6 to 15 cm in radius, will be placed on each side at $z = \pm 34.5$ cm and $z = \pm 46.5$ cm. BPix (FPix) contain 48 million (18 million) pixels covering a total area of 0.78 (0.28) m^2 . The arrangement of the 3 barrel layers and the forward pixel disks on each side gives 3 tracking points over almost the full η -range. Figure 2.10 shows the geometric arrangement and the hit coverage as a function of pseudorapidity η . In the high η region the 2 disk points are combined with the lowest possible radius point from the 4.4 cm barrel layer.

The vicinity to the interaction region also implies a very high track rate and particle fluences that require a radiation tolerant design. For the barrel layers the drift of the electrons to the collecting pixel implant is perpendicular to the 4 T magnetic field of CMS. The resulting Lorentz drift leads to charge spreading of the collected signal charge over more than one pixel. With the analog pulse height being read out a charge interpolation allows to achieve a spatial resolution in the range of 15-20 μm .

The radial region between 20 cm and 116 cm is occupied by the silicon strip tracker. It is composed of three different subsystems. The Tracker Inner Barrel and Disks (TIB/TID) extend in radius towards 55 cm and are composed of 4 barrel layers, supplemented by 3 disks at each end. TIB/TID delivers up to 4 $r - \phi$ measurements on a trajectory using 320 μm thick silicon micro-strip sensors with their strips parallel to the beam axis in the barrel and radial on the disks. The strip pitch is 80 μm on layers 1 and 2 and 120 μm on layers 3 and 4 in the TIB, leading to a single point resolution of 23 μm and 35 μm , respectively. In the TID the mean pitch varies between 100 μm and 141 μm . The TIB/TID is surrounded by the Tracker Outer Barrel (TOB). It has an outer radius of 116 cm and consists of 6 barrel layers of 500 μm thick micro-strip sensors with strip pitches of 183 μm on the first 4 layers and 122 μm on layers 5 and 6. It provides another 6 $r - \phi$ measurements with single point resolution of 53 μm and 35 μm , respectively. The TOB extends in z between

± 118 cm. Beyond this z range the Tracker EndCaps (TEC+ and TEC- where the sign indicates the location along the z axis) cover the region $124 \text{ cm} < |z| < 282 \text{ cm}$ and $22.5 \text{ cm} < |r| < 113.5 \text{ cm}$. Each TEC is composed of 9 disks, carrying up to 7 rings of silicon micro-strip detectors ($320 \mu\text{m}$ thick on the inner 4 rings, $500 \mu\text{m}$ thick on rings 5-7) with radial strips of $97 \mu\text{m}$ to $184 \mu\text{m}$ average pitch. Thus, they provide up to 9 ϕ measurements per trajectory. In addition, the modules in the first two layers and rings, respectively, of TIB, TID, and TOB as well as rings 1, 2, and 5 of the TECs carry a second micro-strip detector module which is mounted back-to-back with a stereo angle of 100 mrad in order to provide a measurement of the second coordinate (z in the barrel and r on the disks). The achieved single point resolution of this measurement is $230 \mu\text{m}$ and $530 \mu\text{m}$ in TIB and TOB, respectively, and varies with pitch in TID and TEC. This tracker layout ensures at least 9 measured hits in the silicon strip tracker in the full range of $|\eta| < 2.4$ with at least ~ 4 of them being two-dimensional measurements (Figure 2.11). The ultimate acceptance of the tracker ends at $|\eta| \approx 2.5$. The CMS silicon strip tracker has a total of 9.3 million strips and 198 m^2 of active silicon area.

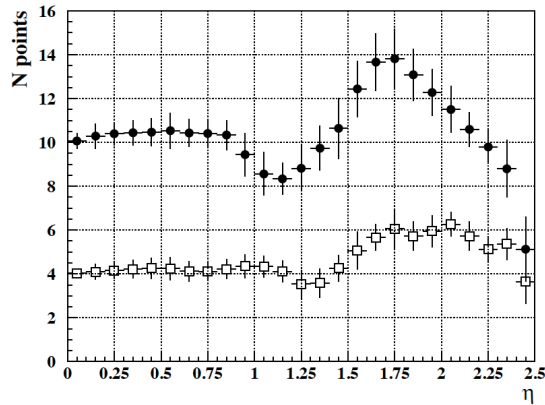


Figure 2.11: Number of measurement points in the strip tracker as a function of pseudorapidity η . Filled circles show the total number (back-to-back modules count as one) while open squares show the number of stereo layers.

Figure 2.12 shows the material budget of the CMS tracker in units of radiation length. It increases from $0.4 X_0$ at $\eta \approx 0$ to about $1.8 X_0$ at $|\eta| \approx 1.4$, beyond which it falls to about $1 X_0$ at $|\eta| \approx 2.5$.

Figure 2.13 shows the expected resolution of transverse momentum, transverse impact parameter and longitudinal impact parameter, as a function of pseudorapidity [55] for single muons of transverse momenta of 1, 10 and 100 GeV. For high momentum tracks (100 GeV) the transverse momentum resolution is around 1-2% up

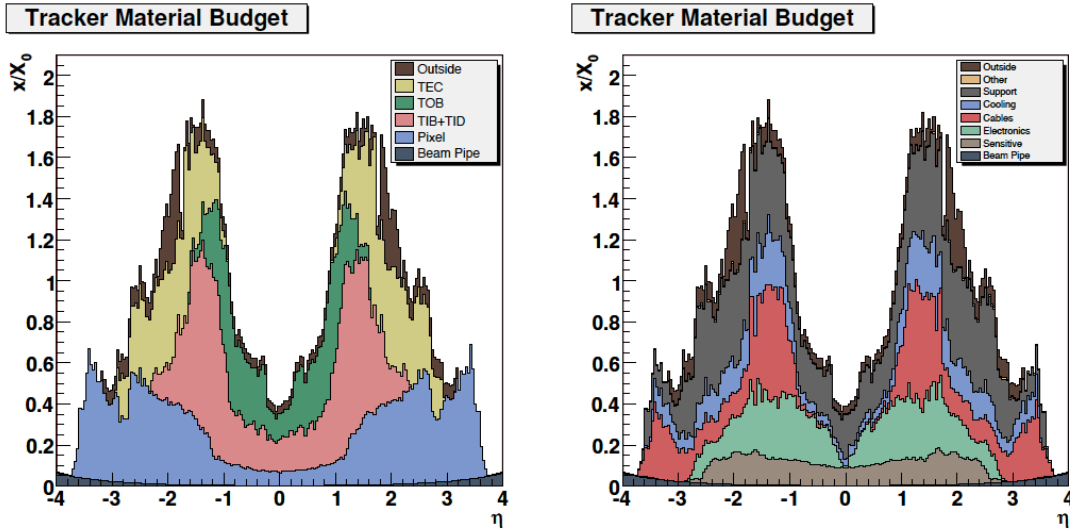


Figure 2.12: Material budget in units of radiation length as a function of pseudorapidity η for the different tracker sub-detectors (left panel) and broken down into the functional contributions (right panel). 30%.

to $|\eta| \approx 1.6$, beyond which it degrades due to the reduced lever arm. At a transverse momentum of 100 GeV multiple scattering in the tracker material accounts for 20 to 30% of the transverse momentum resolution while at lower momentum it is dominated by multiple scattering. The transverse impact parameter resolution reaches $10 \mu\text{m}$ for high p_T tracks, dominated by the resolution of the first pixel hit, while at lower momentum it is degraded by multiple scattering (similarly for the longitudinal impact parameter).

Figure 2.14 shows the expected track reconstruction efficiency of the CMS tracker for single muons and pions as a function of pseudorapidity. For muons, the efficiency is about 99 % over most of the acceptance. For $|\eta| \approx 0$ the efficiency decreases slightly due to gaps between the ladders of the pixel detector at $z \approx 0$. At high η the efficiency drop is mainly due to the reduced coverage by the pixel forward disks. For pions and hadrons in general the efficiency is lower because of interactions with the material in the tracker.

2.2.3 Electromagnetic calorimeter

The electromagnetic calorimeter of CMS (ECAL) is a hermetic homogeneous calorimeter made of 61200 lead tungstate (PbWO_4) crystals mounted in the central barrel part, closed by 7324 crystals in each of the two endcaps. A preshower detector is

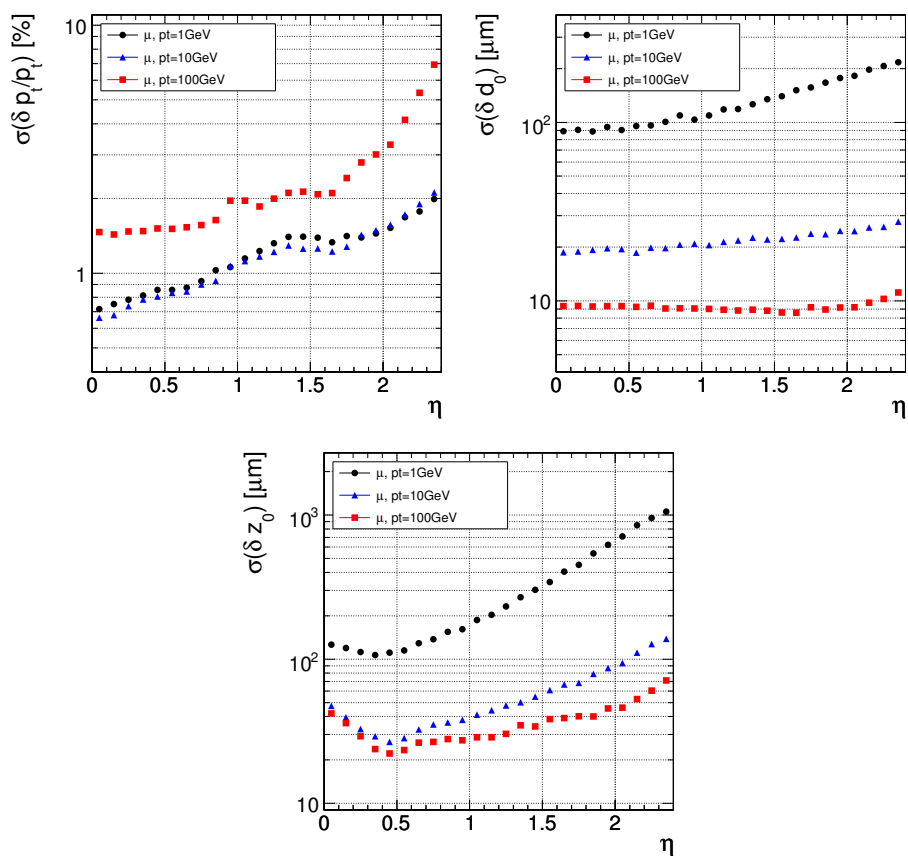


Figure 2.13: transverse momentum (left panel), transverse impact parameter (middle panel), and longitudinal impact parameter (right panel) resolution for single muons with transverse momenta of 1, 10 and 100 GeV.

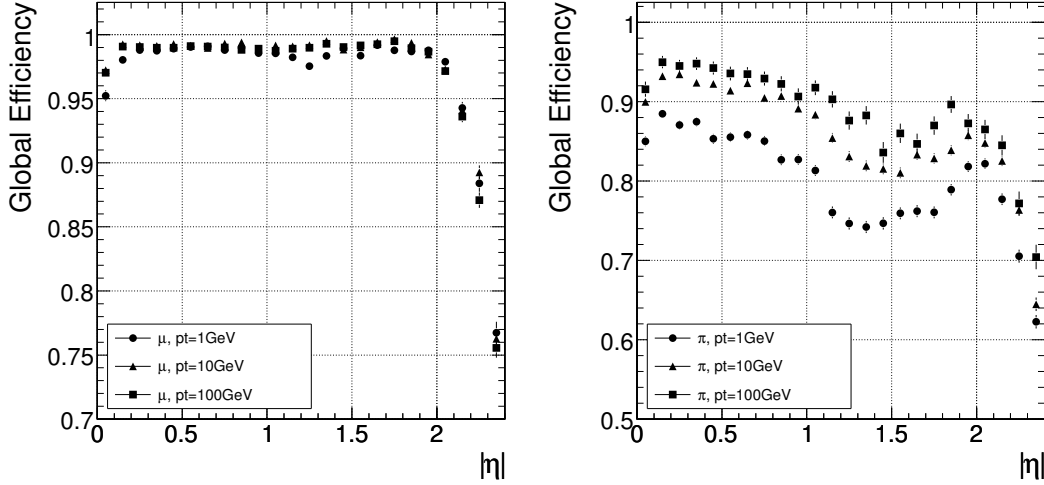


Figure 2.14: Global track reconstruction efficiency for muons (left panel) and pions (right panel) of transverse momenta of 1, 10 and 100 GeV.

placed in front of the endcap crystals. Avalanche photodiodes (APDs) are used as photodetectors in the barrel and vacuum phototriodes (VPTs) in the endcaps. The use of high density crystals has allowed the design of a calorimeter which is fast, has fine granularity and is radiation resistant, all important characteristics in the LHC environment. One of the driving criteria in the design was the capability to detect the decay to two photons of the postulated Higgs boson. This capability is enhanced by the good energy resolution provided by a homogeneous crystal calorimeter.

The characteristics of the PbWO_4 crystals make them an appropriate choice for operation at LHC. The high density (8.28 g/cm^3), short radiation length (0.89 cm) and small Molière radius (2.2 cm) result in a fine granularity and a compact calorimeter. The scintillation decay time of the crystals is of the same order of magnitude as the LHC bunch crossing time: about 80% of the light is emitted in 25 ns.

The barrel part of the ECAL (EB) covers the pseudorapidity range $|\eta| < 1.479$. The barrel granularity is 360-fold in ϕ and (2×85) -fold in η , resulting in a total of 61200 crystals. The crystals have a tapered shape, slightly varying with position in η . They are mounted in a quasi-projective geometry to avoid cracks aligned with particle trajectories, so that their axes make a small angle (3°) with respect to the vector from the nominal interaction vertex, in both the ϕ and η projections. The crystal cross section corresponds to approximately 0.0174×0.0174 in $\eta - \phi$ (in natural units) correspondent to $22 \times 22 \text{ mm}^2$ at the front face of crystal, and $26 \times 26 \text{ mm}^2$ at the rear face. The crystal length is 230 mm corresponding to $25.8 X_0$. The barrel crystal volume is 8.14 m^3 and the total weight is 67.4 t.

The centres of the front faces of the barrel crystals are at a radius 1.29 m. The crystals are contained in a thin-walled alveolar structure (submodule). 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.

The endcaps (EE) cover the rapidity range $1.479 < |\eta| < 3.0$. The longitudinal distance between the interaction point and the endcap envelope is 315.4 cm, taking account of the estimated shift toward the interaction point by 1.6 cm when the 4-T magnetic field is switched on. The endcap consists of identically shaped crystals grouped in mechanical units of 5×5 crystals (supercrystals, or SCs) consisting of a carbon-fibre alveola structure. Each endcap is divided into 2 halves, or “Dees”. Each Dee holds 3662 crystals. The crystals and SCs are arranged in a rectangular $x - y$ grid, with the crystals pointing at a focus 1300 mm beyond the interaction point, giving off-pointing angles ranging from 2 to 8 degrees. The crystals have a rear face cross section $30 \times 30 \text{ mm}^2$, a front face cross section $28.62 \times 28.62 \text{ mm}^2$ and a length of 220 mm ($24.7 X_0$). The endcaps crystal volume is 2.90 m^3 and the weight is 24.0 t. The layout of the calorimeter is shown in Fig. 2.15.

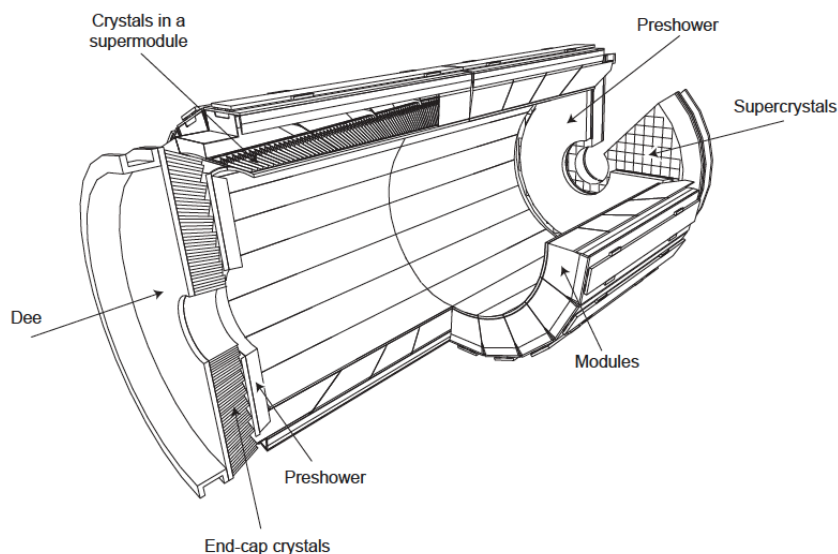


Figure 2.15: Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules, supermodules and endcaps, with the preshower in front.

The main goal of the CMS Preshower detector is to identify neutral pions in the endcaps within a fiducial region $1.653 < |\eta| < 2.6$. It also helps the identifi-

cation of electrons against minimum ionizing particles, and improves the position determination of electrons and photons with high granularity. The Preshower is a sampling calorimeter with two layers: lead radiators initiate electromagnetic showers from incoming photons/electrons while silicon strip sensors placed after each radiator measure the deposited energy and the transverse shower profiles. The total thickness of the Preshower is 20 cm. The material thickness of the Preshower traversed at $\eta = 1.653$ before reaching the first sensor plane is $2X_0$, followed by a further $1X_0$ before reaching the second plane. Thus about 95% of single incident photons start showering before the second sensor plane.

Calibration and monitoring

The number of scintillation photons emitted by the crystals and the amplification of the APD are both temperature dependent. Therefore a major task is to monitor the system's temperature and the verification that the required temperature stability of $(18 \pm 0.05)^\circ\text{C}$ of the crystal volume and the APDs is achieved.

Calibration is a severe technical challenge for the operation of the CMS ECAL. Many small effects which are negligible at low precision need to be treated with care as the level of precision of a few per mille is approached. ECAL calibration is naturally seen as composed of a global component, giving the absolute energy scale, and a channel-to-channel relative component, which is referred to as intercalibration. The essential issues are uniformity over the whole ECAL and stability, so that showers in different locations in the ECAL in data recorded at different times are accurately related to each other. Another feature to monitor is the evolution of the crystal transparency (a limited but rapid loss of optical transmission due to radiation), which is measured using laser pulses injected into the crystals via optical fibres.

Energy resolution

For energies below about 500 GeV, where shower leakage from the rear of the calorimeter starts to become significant, the energy resolution can be parametrized as in equation

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \quad (2.10)$$

where S is the stochastic term, N the noise term, and C the constant term. There are three basic contributions to the stochastic term:

- event-to-event fluctuations in the lateral shower containment,
- fluctuation of the number of photons contributes to 2.1%,
- fluctuations in the energy deposited in the preshower absorber (where present) with respect to what is measured in the preshower silicon detector.

The most important contributions to the constant term may be listed as follows:

- non-uniformity of the longitudinal light collection,
- intercalibration errors,
- leakage of energy from the back of the crystal

There are three contributions to the noise term:

- electronics noise,
- digitization noise,
- pileup noise.

A typical energy resolution found using test beams, summing the energy deposit in a matrix of 3×3 crystals was found to be:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right)^2 + (0.30\%)^2, \quad (2.11)$$

where E is in GeV.

2.2.4 Hadron calorimeter

The hadron calorimeters are particularly important for the measurement of hadron jets and missing transverse energy due to neutrinos or exotic particles. Figure 2.16 shows the longitudinal view of the CMS detector. The dashed lines correspond to fixed η values. The hadron calorimeter barrel and endcaps sit behind the tracker and the electromagnetic calorimeter as seen from the Interaction Point. The hadron calorimeter barrel is radially restricted between the outer extent of the electromagnetic calorimeter ($r=1.77$ m) and the inner extent of the magnet coil ($r = 2.95$ m). This constrains the total amount of material which can be inserted in to absorb the hadronic shower. Therefore, an outer hadron calorimeter or tail catcher is placed outside the solenoid complementing the barrel calorimeter. Beyond $|\eta| = 3$, the forward hadron calorimeters placed at 11.2 m from the interaction point extend the pseudorapidity coverage down to $|\eta| = 5.2$ using a Cherenkov-based, radiation-hard technology.

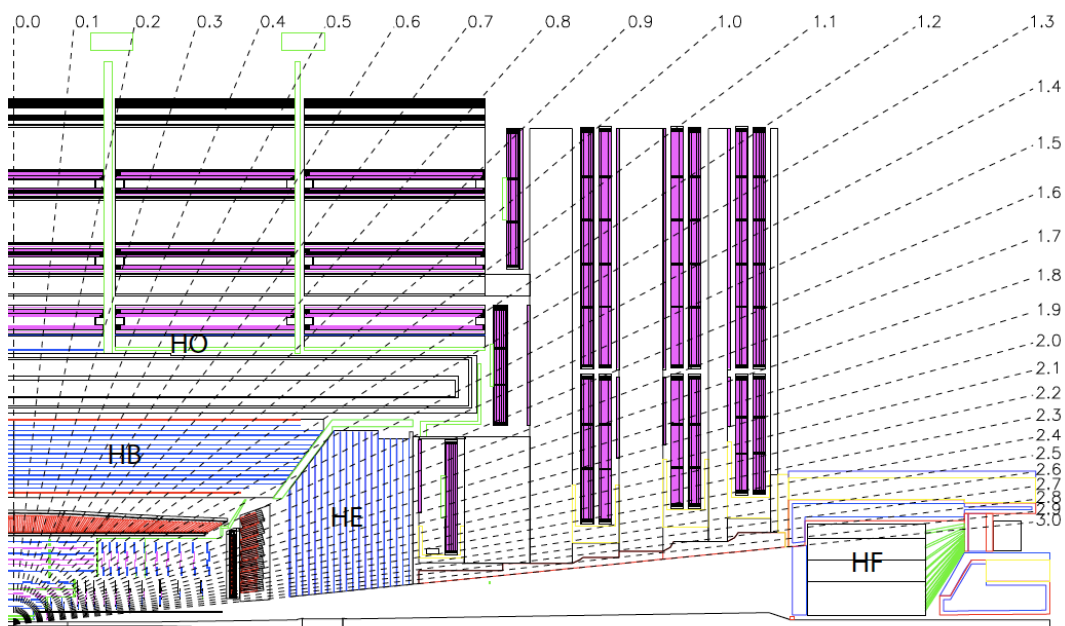


Figure 2.16: Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters.

Barrel design (HB)

The HB is a sampling calorimeter covering the pseudorapidity range $|\eta| < 1.3$. The HB consists of 36 identical azimuthal wedges which form the two half-barrels (HB+ and HB-). The wedges are constructed out of flat brass absorber plates aligned parallel to the beam axis. Each wedge is segmented into four azimuthal angle (ϕ) sectors. The plates are bolted together in a staggered geometry resulting in a configuration that contains no projective dead material for the full radial extent of a wedge (Figure 2.17). The innermost and outermost plates are made of stainless steel for structural strength. The plastic scintillator is divided into 16 η sectors, resulting in a segmentation $(\Delta\eta, \Delta\phi) = (0.087, 0.087)$. The absorber (Table 2.3) consists of a 40 mm thick front steel plate, followed by eight 50.5 mm thick brass plates, six 56.5 mm thick brass plates, and a 75 mm thick steel back plate. The total absorber thickness at 90° is 5.82 interaction lengths (λ_I). The HB effective thickness increases with polar angle (θ) as $1/\sin\theta$, resulting in $10.6 \lambda_I$ at $|\eta| = 1.3$. The electromagnetic crystal calorimeter in front of HB adds about $1.1 \lambda_I$ of material.

| layer | material | thickness |
|-------------|----------|-----------|
| front plate | steel | 40 mm |
| 1-8 | brass | 50.5 mm |
| 9-14 | brass | 56.5 mm |
| back plate | steel | 75 mm |

Table 2.3: Absorber thickness in the HB wedges.

The active medium uses the well known tile and wavelength shifting fibre concept to bring out the light. The CMS hadron calorimeter consists of about 70 000 tiles. In order to limit the number of individual elements to be handled, the tiles of a given ϕ layer are grouped into a single mechanical scintillator tray unit. The HB baseline active material is 3.7 mm thick Kuraray SCSN81 plastic scintillator, chosen for its long-term stability and moderate radiation hardness. The first layer of scintillator (layer 0) is located in front of the steel support plate. It is made of 9 mm thick Bicron BC408. The scintillators are summarized in Table 2.4. The purpose of layer zero is to sample hadronic showers developing in the inert material between the EB and HB. The larger thickness of layer 16 serves to correct for late developing showers leaking out the back of HB.

EndCap design(HE)

The hadron calorimeter endcaps (HE) [108] cover a substantial portion of the rapidity range, $1.3 < |\eta| < 3$ (13.2% of the solid angle), a region containing about 34% of the particles produced in the final state. The high luminosity of the LHC ($10^{34} \text{cm}^{-2} \text{s}^{-1}$)

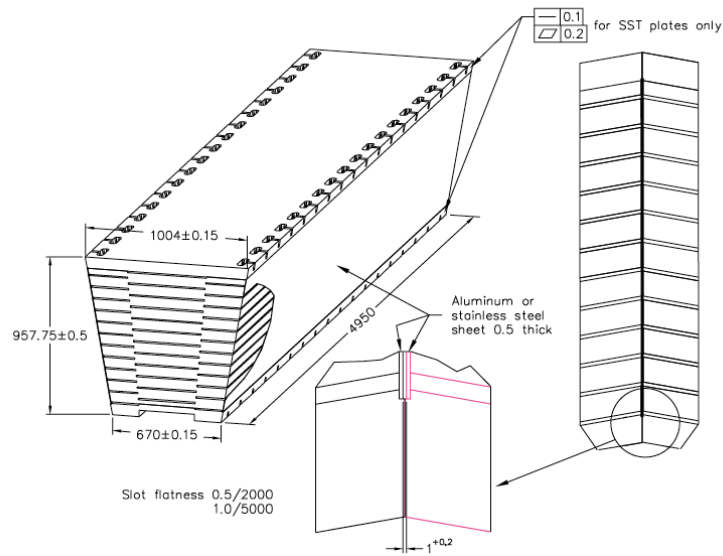


Figure 2.17: Isometric view of the HB wedges, showing the hermetic design of the scintillator sampling.

| layer | material | thickness |
|-------|----------------|-----------|
| 0 | Bicron BC408 | 9 mm |
| 1-15 | Kuraray SCSN81 | 3.7 mm |
| 16 | Kuraray SCSN81 | 9 mm |

Table 2.4: Scintillator in the HB wedges.

requires HE to handle high (MHz) counting rates and have high radiation tolerance (10 MRad after 10 years of operation at design luminosity) at $|\eta| < 3$. Since the calorimeter is inserted into the ends of a 4 T solenoidal magnet, the absorber must be made from a non-magnetic material. It must also have a maximum number of interaction lengths to contain hadronic showers, good mechanical properties and reasonable cost, leading to the choice of C26000 cartridge brass. The endcaps are attached to the muon endcap yoke as shown in Figure 2.18. Only a small part of the calorimeter structure can be used for the fixation to the magnet iron, because the majority of the space between HE and muon absorber is occupied with muon cathode strip chambers. A 10 t electromagnetic calorimeter (EE) with a 2 t preshower detector (ES) is attached at the front face of HE. The large weight involved (about 300 t) and a strict requirement to minimize non-instrumented materials along particle trajectories, has made the design of HE a challenge to engineers. An interface kinematic scheme was developed in order to provide precise positioning of the endcap detectors with respect to the adjacent muon station, and to minimize the influence of deformation under magnetic forces. The interface kinematic contains a sliding joint between the interface tube, and HE back-flange and the hinge connection between brackets and the iron disk (YE1). Structural materials used in the interface system are non-magnetic in order not to distort the axial magnetic field of up to 4 T. The brass plates are 79 mm thick with 9 mm gaps to accommodate the scintillators. The total length of the calorimeter, including electromagnetic crystals, is about 10 interaction lengths (λ_I). The scintillation light is collected by wavelength shifting (WLS) fibres. Trapezoidal-shaped scintillators, 3.7 mm thick SCSN81 for layers 1-17 and 9 mm thick Bicron BC408 for layer 0, have grooves in which the WLS fibres are inserted. The granularity of the calorimeters is $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 1.6$ and $\Delta\eta \times \Delta\phi \approx 0.17 \times 0.17$ for $|\eta| \leq 1.6$.

Outer calorimeter design (HO)

In the central pseudorapidity region, the combined stopping power of EB plus HB does not provide sufficient containment for hadron showers. To ensure adequate sampling depth for $|\eta| < 1.3$, the hadron calorimeter is extended outside the solenoid with a tail catcher called the HO or outer calorimeter. The HO utilises the solenoid coil as an additional absorber equal to $1.4/\sin\theta$ interaction lengths and is used to identify late starting showers and to measure the shower energy deposited after HB.

At $\eta = 0$, HB has the minimal absorber depth. Therefore, the central ring (ring 0) has two layers of HO scintillators on either side of a 19.5 cm thick piece of iron (the tail catcher iron) at radial distances of 3.82 m and 4.07 m, respectively. All other rings have a single HO layer at a radial distance of 4.07 m. The total depth of the calorimeter system is thus extended to a minimum of 11.8 λ_I except at the barrel-endcap boundary region. The HO is constrained by the geometry of the muon

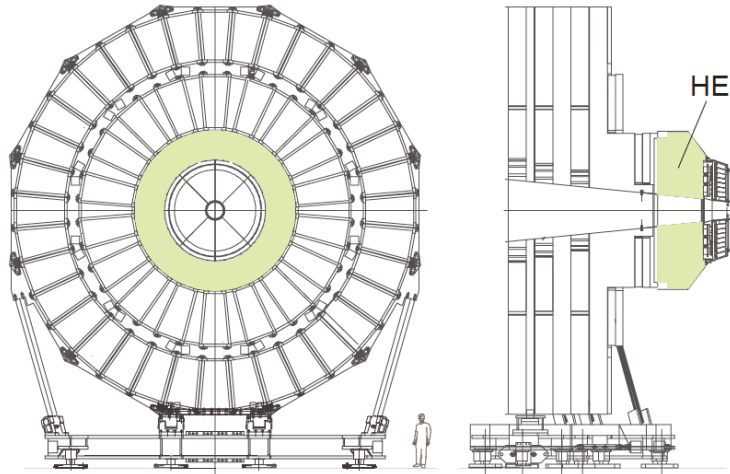


Figure 2.18: Hadron endcap (HE) calorimeter mounted on the endcap iron yoke.

system. Figure 2.19 shows the position of HO layers in the rings of the muon stations in the overall CMS setup.

Forward calorimeter design (HF)

The forward calorimeter will experience unprecedented particle fluxes. On average, 760 GeV per proton-proton interaction is deposited into the two forward calorimeters, compared to only 100 GeV for the rest of the detector. Moreover, this energy is not uniformly distributed but has a pronounced maximum at the highest rapidities. At $|\eta| = 5$ after an integrated luminosity of $5 \times 10^5 \text{ pb}^{-1}$ (≈ 10 years of LHC operation), the HF will experience a dose $\approx 10 \text{ MGy}$ (Gy is the universal unit measure of radiation absorption dose, and is defined as the absorption of one joule of ionizing radiation by one kilogram of matter). The charged hadron rates will also be extremely high. For the same integrated luminosity, inside the HF absorber at 125 cm from the beam-line, the rate will exceed 10^{11} per cm^2 . This hostile environment presents a considerable challenge to calorimetry, and the design of the HF calorimeter was first and foremost guided by the necessity to survive in these harsh conditions, preferably for at least a decade. Successful operation critically depends on the radiation hardness of the active material. This was the principal reason why quartz fibres (fused-silica core and polymer hard-cladding) were chosen as the active medium. The signal is generated when charged shower particles above the Cherenkov threshold ($E \lesssim 190 \text{ keV}$ for electrons) generate Cherenkov light, thereby rendering the calorimeter mostly sensitive to the electromagnetic component of showers. The

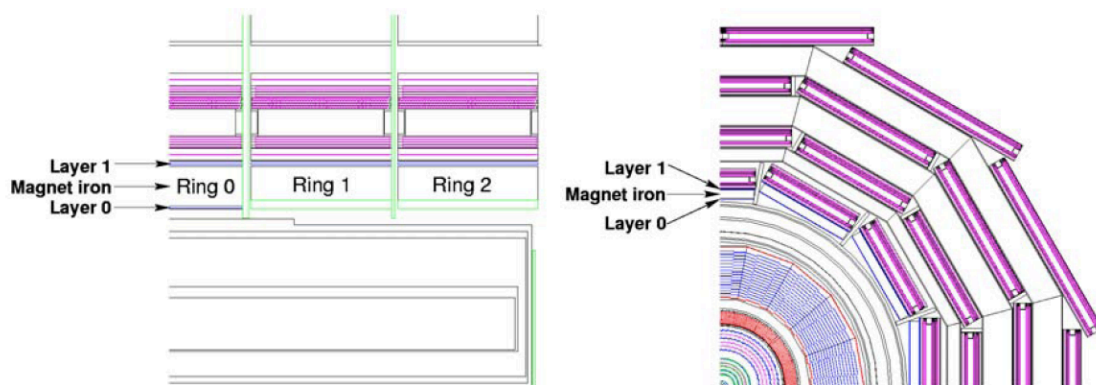


Figure 2.19: Longitudinal and transverse views of the CMS detector showing the position of HO layers.

forward calorimeter is essentially a cylindrical steel structure with an outer radius of 130.0 cm. The front face of the calorimeter is located at 11.2 m from the interaction point. The hole for the beam pipe is cylindrical, with radius 12.5 cm from the center of the beam line. This structure is azimuthally subdivided into 20° modular wedges. Thirty-six such wedges (18 on either side of the interaction point) make up the HF calorimeters.

HF luminosity monitor

The CMS luminosity measurement will be used to monitor the LHC's performance on a bunch-by-bunch basis in real time and to provide an overall normalization for physics analyses. The design goal for the real-time measurement is to determine the average luminosity with a 1% statistical accuracy with an update rate of 1 Hz. For offline analyses, the design goal is a systematic accuracy of 5%, although every reasonable effort will be made to produce a more accurate result. Both of these requirements must be met over a very large range of luminosities, extending from roughly $10^{28}\text{cm}^{-2}\text{s}^{-1}$ to $10^{34}\text{cm}^{-2}\text{s}^{-1}$, and possibly beyond. A number of techniques capable of providing suitable luminosity information in real time have been identified. One technique employs signals from the forward hadron calorimeter (HF). Two methods for extracting a real-time relative instantaneous luminosity with the HF have been studied. The first method is based on zero counting in which the average fraction of empty towers is used to infer the mean number of interactions per bunch crossing. The second method exploits the linear relationship between the average transverse energy per tower and the luminosity.

Castor

The CASTOR (CentauRO And Strange Object Research) detector is a quartz-tungsten sampling calorimeter[56], designed for the very forward rapidity region in heavy ion and proton-proton collisions at the LHC. Its physics motivation is to complement the nucleus-nucleus physics programme, developed essentially in the baryon-free mid-rapidity region, and also the diffractive and low- x physics in pp collisions. CASTOR is installed at 14.38 m from the interaction point, covering the pseudorapidity range $5.2 < |\eta| < 6.6$. Figure 2.20 shows the location of CASTOR in the CMS forward region. The calorimeter and its readout are designed in such a way as to permit the observation of the cascade development of the impinging particles as they traverse the calorimeter. The typical total and electromagnetic energies in the CASTOR acceptance range can be measured with a resolution better than $\approx 1\%$. The main advantages of quartz calorimeters are radiation hardness, fast response and compact detector dimensions, making them suitable for the experimental conditions encountered in the very forward region at the LHC. The CASTOR detector is a Cerenkov-based calorimeter, similar in concept to the HF. It is constructed from layers of tungsten (W) plates (density $\approx 18.5 \text{ g/cm}^3$) as absorber and fused silica quartz (Q) plates as active medium. For the electromagnetic (EM) section, the W plates have a thickness of 5.0 mm and the Q plates 2.0 mm. For the hadronic (HAD) section, the W and Q plates have thicknesses of 10.0 mm and 4.0 mm, respectively. The W/Q plates are inclined 45° with respect to the direction of the impinging particles, in order to maximize the Cerenkov light output in the quartz.

2.2.5 The muon system

Muon detection is a powerful tool for recognizing signatures of interesting processes over the very high background rate expected at the LHC with full luminosity. For example, the predicted decay of the Standard Model Higgs boson into ZZ or ZZ^* , which in turn decay into 4 leptons, has been called “gold plated” for the case in which all the leptons are muons. Besides the relative ease in detecting muons, the best 4 particle mass resolution can be achieved if all the leptons are muons because they are less affected than electrons by radiative losses in the tracker material. This example, and others from SUSY models, emphasize the discovery potential of muon final states and the necessity for wide angular coverage for muon detection. Therefore, as is implied by the experiment’s middle name, the detection of muons is of central importance to CMS: precise and robust muon measurement was a central theme from its earliest design stages. The muon system has 3 functions: muon identification, momentum measurement, and triggering. Good muon momentum resolution and trigger capability are enabled by the high field solenoidal magnet and its flux-return yoke. The latter also serves as a hadron absorber for the identification of muons.

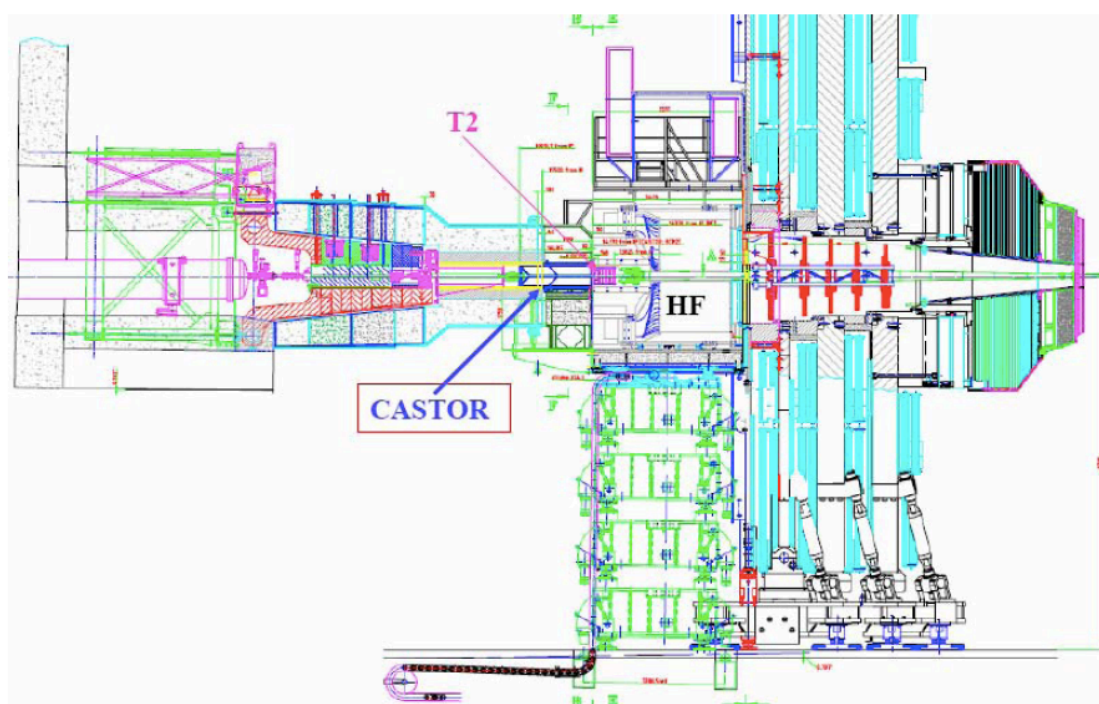


Figure 2.20: Location of CASTOR in the CMS forward region.

The amount of material thickness crossed by muons, as a function of pseudorapidity, is shown in Figure 2.21. The CMS muon system is designed to have the capability to reconstruct the momentum and charge of muons over the entire kinematic range of the momentum spectrum accessible in LHC collision event. CMS uses 3 types of gaseous particle detectors for muon identification[57]. Due to the shape of the solenoidal magnet, the muon system was naturally driven to have a cylindrical, barrel section and 2 planar endcap regions. Because the muon system consists of about 25000 m² of detection planes, the muon chambers had to be inexpensive, reliable, and robust.

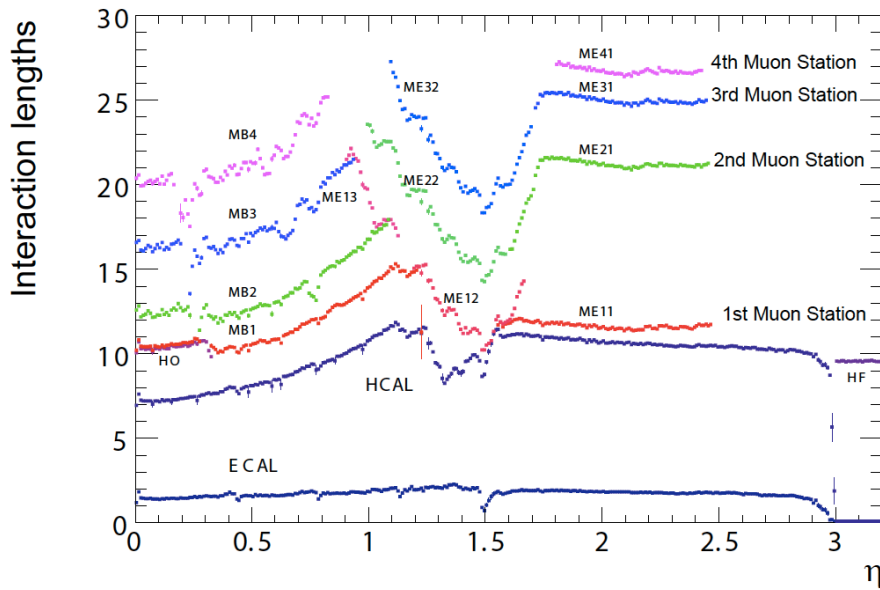


Figure 2.21: Material thickness in interaction lengths at various depths, as a function of pseudorapidity.

Drift tubes

In the barrel region, where we expect very muon contamination from neutron background, the muon rate is low, and the magnetic field is uniform and mostly contained in the steel yoke, drift chambers with standard rectangular drift cells are used. The barrel drift tube (DT) chambers cover the pseudorapidity region $|\eta| < 1.2$ and are organized into 4 stations forming concentric cylinders around the beam line: the 3 inner cylinders have 60 drift chambers each and the outer cylinder has 70. There are about 172000 sensitive wires interspersed among the layers of the flux return plates.

The wire length, around 2.4 m in the chambers measured in an $r - \phi$ projection, is constrained by the longitudinal segmentation of the iron barrel yoke. The transverse dimension of the drift cell, i.e., the maximum path and time of drift, was chosen to be 21 mm, corresponding to a drift time of 380 ns in a gas mixture of 85% Ar + 15% CO₂. This value is small enough to produce a negligible occupancy and to avoid the need for multi-hit electronics, yet the cell is large enough to limit the number of active channels to an affordable value. A tube was chosen as the basic drift unit to obtain protection against damage from a broken wire and to partially decouple contiguous cells from the electromagnetic debris accompanying the muon itself. The amount of iron in the return yoke was dictated by the decision to have a large and intense solenoidal magnetic field at the core of CMS. Two detector layers, one inside the yoke and the other outside, would be insufficient for reliable identification and measurement of a muon in CMS. Therefore, 2 additional layers are embedded within the yoke iron (Fig. 2.22). In each of the 12 sectors of the yoke there are 4 muon chambers per wheel, labeled MB1, MB2, MB3, and MB4. The yoke-iron supports that are between the chambers of a station generate 12 unavoidable dead zones in the ϕ coverage, although the supports are placed so as not to overlap in ϕ . A drift-tube (DT) chamber is made of 3 (or 2) superlayers (SL, see Fig. 2.23), each made of 4 layers of rectangular drift cells staggered by half a cell. The SL is the smallest independent unit of the design. The wires in the 2 outer SLs are parallel to the beam line and provide a track measurement in the magnetic bending plane ($r - \phi$). In the inner SL, the wires are orthogonal to the beam line and measure the z position along the beam. This third, z -measuring, SL is not present in the fourth station, which therefore measures only the ϕ coordinate. A muon coming from the interaction point first encounters a ϕ -measuring SL, passes through the honeycomb plate, then crosses the z -measuring SL and the second ϕ -measuring SL. In this scenario, there still exist limited regions of η in which the combined effect of the ϕ and z discontinuities limits to only 2 (out of 4), the number of stations crossed by a muon.

Cathode strip chambers

At the time of the LHC start-up, the CMS Endcap Muon system will consist of 468 cathode strip chambers (CSC) arranged in groups as follows: 72 ME1/1, 72 ME1/2, 72 ME1/3, 36 ME2/1, 72 ME2/2, 36 ME3/1, 72 ME3/2, and 36 ME4/1 (figure 2.24). The de-scoped 72 ME4/2 chambers will not be available during early years of CMS operation. The chambers are trapezoidal and cover either 10° or 20° in ϕ ; all chambers, except for the ME1/3 ring, overlap and provide contiguous ϕ -coverage. A muon in the pseudorapidity range $1.2 < |\eta| < 2.4$ crosses 3 or 4 CSCs. In the endcap-barrel overlap range, $0.9 < |\eta| < 1.2$, muons are detected by both the barrel drift tubes (DT) and endcap CSCs. In the baseline design, muons with $|\eta| < 2.1$ are also detected by resistive plate chambers (RPC); however, in the initial

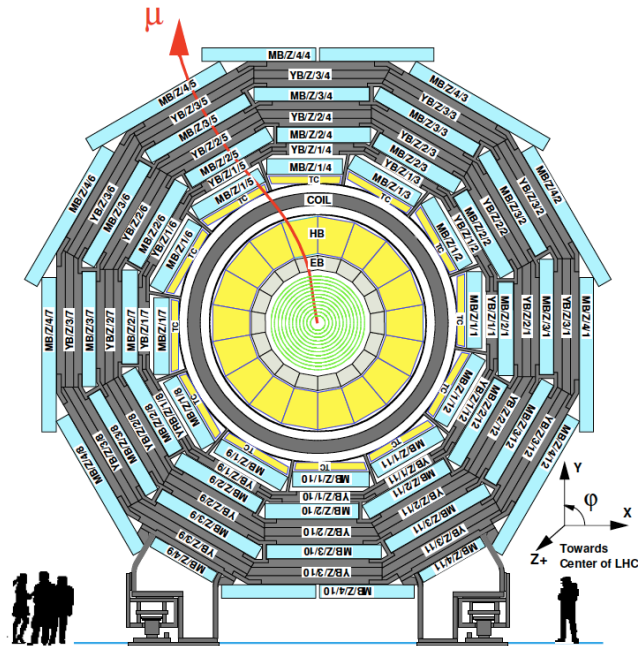


Figure 2.22: Layout of the CMS barrel muon DT chambers in one of the 5 wheels. The chambers in each wheel are identical with the exception of wheels -1 and +1 where the presence of cryogenic chimneys for the magnet shortens the chambers in 2 sectors.

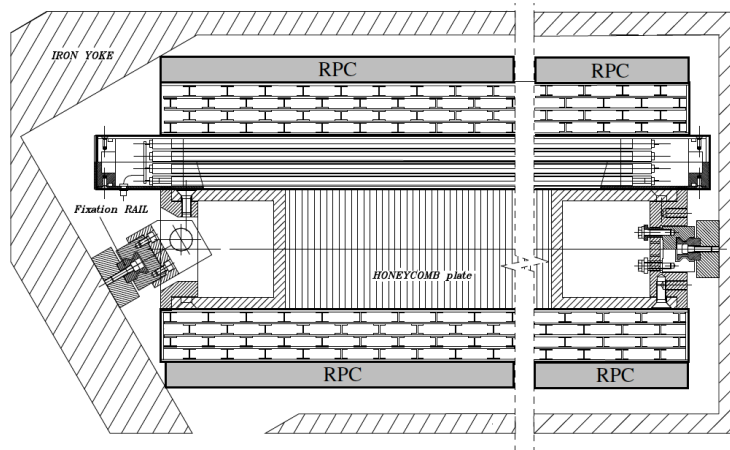


Figure 2.23: A DT chamber in position inside the iron yoke; the view is in the $(r-\phi)$ plane. One can see the 2 SLs with wires along the beam direction and the other perpendicular to it. In between is a honeycomb plate with supports attached to the iron yoke. Not shown are the RPCs, which are attached to the DT chambers via support plates glued to the bottom and/or top faces, depending on chamber type.

detector this coverage is reduced to $|\eta| < 1.6$. The CSCs are multiwire proportional chambers comprised of 6 anode wire planes interleaved among 7 cathode panels. Wires run azimuthally and define a track's radial coordinate. Strips are milled on cathode panels and run lengthwise at constant $\Delta\phi$ width. Following the original CSC idea[58], the muon coordinate along the wires (ϕ in the CMS coordinate system) is obtained by interpolating charges induced on strips. The largest chambers, ME2/2 and ME3/2, are about $3.4 \times 1.5 \text{ m}^2$ in size. The overall area covered by the sensitive planes of all chambers is about 5000 m^2 , the gas volume is greater than 50 m^3 , and the number of wires is about 2 million. There are about 9000 high-voltage channels in the system, about 220 000 cathode strip read-out channels with 12-bit signal digitisation, and about 180 000 anode wire read-out channels.

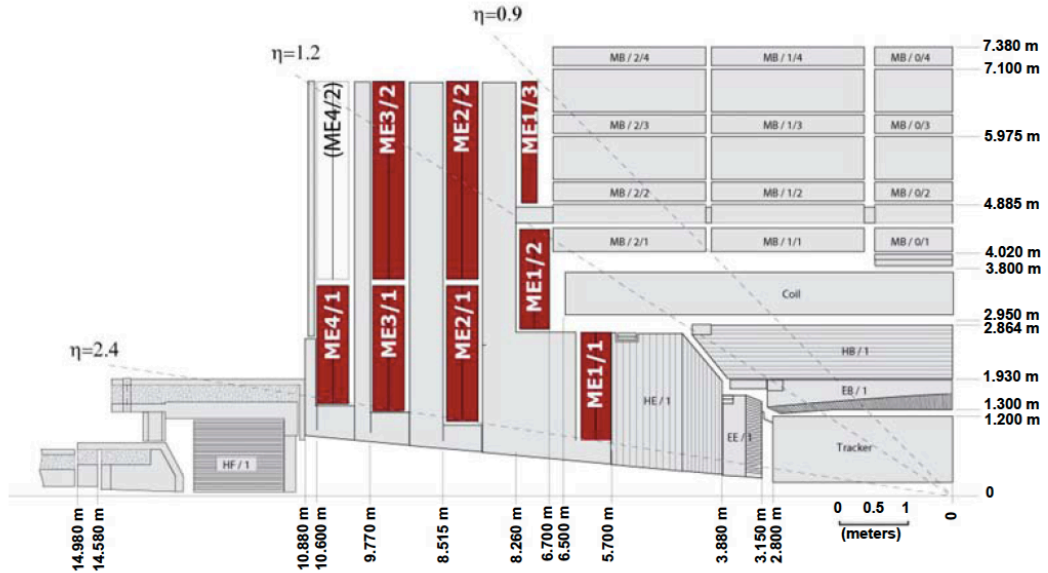


Figure 2.24: Quarter-view of the CMS detector. Cathode strip chambers of the Endcap Muon system are highlighted.

Because the muon detector elements cover the full pseudorapidity interval $|\eta| < 2.4$ with no acceptance gaps, muon identification is ensured over the range corresponding to $10^\circ < \theta < 170^\circ$. Offline reconstruction efficiency of simulated single-muon samples (Figure 2.25) is typically 95-99% except in the regions around $|\eta| = 0.25$ and 0.8 (the regions between 2 DT wheels) and $|\eta| = 1.2$ (the transition region between the DT and CSC systems), where the efficiency drops. A negligible punch-through fraction reaches the system due to the amount of material in front of the muon system, which exceeds 16 interaction lengths. Due to multiple-scattering in the detector material before the first muon station, the offline muon momentum resolution of the

“standalone muon” system (i.e. the track muon reconstructed only with the muon system information) is about 9% for small values of η and p for transverse momenta up to 200 GeV. At 1 TeV the standalone momentum resolution varies between 15% and 40%, depending on $|\eta|$. A global momentum fit using also the inner tracker improves the momentum resolution by an order of magnitude at low momenta. Below a momentum value of about 100 GeV, the muon momentum resolution given by the tracker information (“tracker muon” track) is better than the standalone muon system, while above 100 GeV the situation is the opposite. At high momenta (1 TeV) both detector parts together yield a momentum resolution of about 5%. Note that the muon system and the inner tracker provide independent muon momentum measurements; this redundancy enhances fault finding and permits cross-checking between the systems. A crucial characteristic of the DT and CSC subsystems is that they can each trigger on the p_T of muons with good efficiency and high background rejection, independent of the rest of the detector. The Level-1 trigger p_T resolution is about 15% in the barrel and 25% in the endcap.

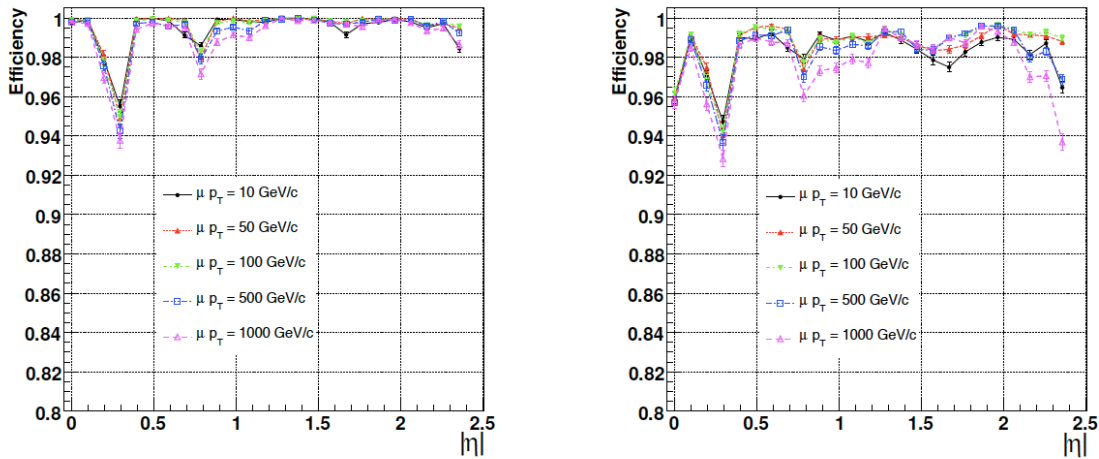


Figure 2.25: Muon reconstruction efficiency as a function of pseudorapidity for selected values of p_T . Left panel: standalone reconstruction (using only hits from the muon system with a vertex constraint). Right panel: global reconstruction (using hits from both the muon system and the tracker).

Resistive plate chambers

Because of the uncertainty in the eventual background rates and in the ability of the muon system to measure the correct beam-crossing time when the LHC reaches full

luminosity, a complementary, dedicated trigger system consisting of resistive plate chambers (RPC) was added in both the barrel and endcap regions. Resistive Plate Chambers (RPC) are gaseous parallel-plate detectors that combine adequate spatial resolution with a time resolution comparable to that of scintillators[59]. An RPC is capable of tagging the time of an ionising event in a much shorter time than the 25 ns between 2 consecutive LHC Bunch Crossings (BX). Therefore, a fast dedicated muon trigger device based on RPCs can identify unambiguously the relevant BX to which a muon track is associated even in the presence of the high rate and background expected at the LHC. Signals from such devices directly provide the time and position of a muon hit with the required accuracy. A trigger based on RPCs has to provide the BX assignment to candidate tracks and provide a fast estimate of the transverse momenta with high efficiency in an environment where rates may reach 10^3 Hz/cm².

The RPCs provide a fast, independent, and highly-segmented trigger with a sharp p_T threshold over a large portion of the rapidity range ($|\eta| < 1.6$) of the muon system. The RPCs are double-gap chambers (see Fig. 2.26), with bakelite strips, operated in avalanche mode to ensure good operation at high rates. They produce a fast response, with good time resolution but coarser position resolution than the DTs or CSCs. They also help to resolve ambiguities in attempting to make tracks from multiple hits in a chamber. A total of 6 layers of RPCs are embedded in the barrel muon system, 2 in each of the first 2 stations, and 1 in each of the last 2 stations. The redundancy in the first 2 stations allows the trigger algorithm to work even for low p_T tracks that may stop before reaching the outer 2 stations. In the endcap region, there is a plane of RPCs in each of the first 3 stations in order for the trigger to use the coincidences between stations to reduce background, to improve the time resolution for bunch crossing identification, and to achieve a good p_T resolution. Finally, a sophisticated alignment system measures the positions of the muon detectors with respect to each other and to the inner tracker, in order to optimize the muon momentum resolution.

2.2.6 Trigger

The LHC provides proton-proton and heavy-ion collisions at high interaction rates. For protons the beam crossing interval is 25 ns, corresponding to a crossing frequency of 40 MHz, but the the maximum storage rate allowed is around 100 Hz. Depending on luminosity, several collisions occur at each crossing of the proton bunches. Since it is impossible to store and process the large amount of data associated with the resulting high number of events, a drastic rate reduction has to be achieved to preserve the events of physics interests. This task is performed by the trigger system, which is the start of the physics event selection process. The rate is reduced in two steps called Level-1 (L1) Trigger and High-Level Trigger (HLT), respectively. The Level-1 Trigger consists of custom-designed, largely programmable electronics, whereas the

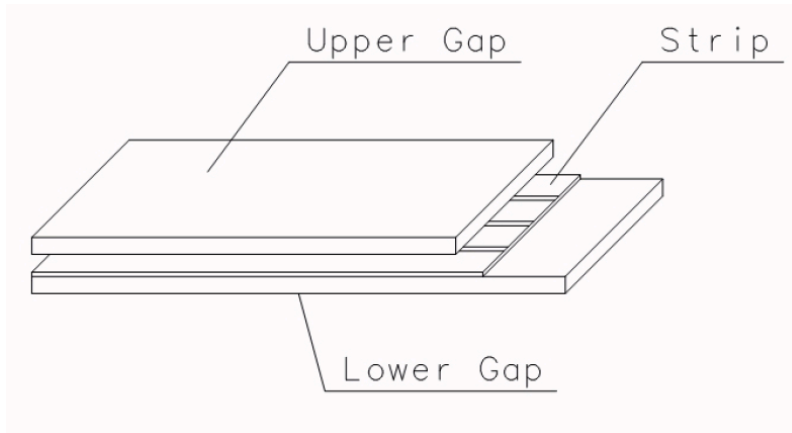


Figure 2.26: Layout of a double-gap RPC.

HLT is a software system implemented in a filter farm of about one thousand commercial processors. The rate reduction capability is designed to be at least a factor of 10^6 for the combined L1 Trigger and HLT. The design output rate limit of the L1 Trigger is 100 kHz, which translates in practice to a calculated maximal output rate of 30 kHz, assuming an approximate safety factor of three. The L1 Trigger uses coarsely segmented data from the calorimeters and the muon system, while holding the high-resolution data in pipe-lined memories in the front-end electronics. The HLT has access to the complete read-out data and can therefore perform complex calculations similar to those made in the the analysis off-line software if required for specially interesting events. HLT algorithms will evolve with time and experience.

At the bottom end, the Local Triggers, also called Trigger Primitive Generators (TPG), are based on energy deposits in calorimeter trigger towers and track segments or hit patterns in muon chambers, respectively. Regional Triggers combine their information and use pattern logic to determine ranked and sorted trigger objects such as electron or muon candidates in limited spatial regions. The rank is determined as a function of energy or momentum and quality, which reflects the level of confidence attributed to the L1 parameter measurements, based on detailed knowledge of the detectors and trigger electronics and on the amount of information available. The Global Calorimeter and Global Muon Triggers determine the highest-rank calorimeter and muon objects across the entire experiment and transfer them to the Global Trigger, the top entity of the Level-1 hierarchy. The latter takes the decision to reject an event or to accept it for further evaluation by the HLT. The architecture of the L1 Trigger is depicted in Figure 2.27. The L1 Trigger has to analyze every bunch crossing. The allowed L1 Trigger latency, between a given bunch crossing and the distribution of the trigger decision to the detector front-end electronics, is $3.2 \mu\text{s}$.

The processing must therefore be pipelined in order to enable a quasi-deadtime-free operation.

The event reconstruction and selection in the HLT[60] take place in steps which correspond roughly to what would have been distinct trigger systems, the Level-2 and Level-3 trigger systems. It is thus convenient to use the terminology, and to refer to a "Level-2 trigger" or a "Level-3 step" to describe the selection algorithms and criteria of the HLT. As mentioned the CMS HLT architecture does not include a sharp division between these trigger steps, other than the order in which they are applied. Typically, a Level-2 trigger, which has the maximum rate of events input to it, uses only information from the calorimeter and muon detectors. In contrast, "Level-3" refers to selection that includes the reconstruction of full tracks in the tracker.

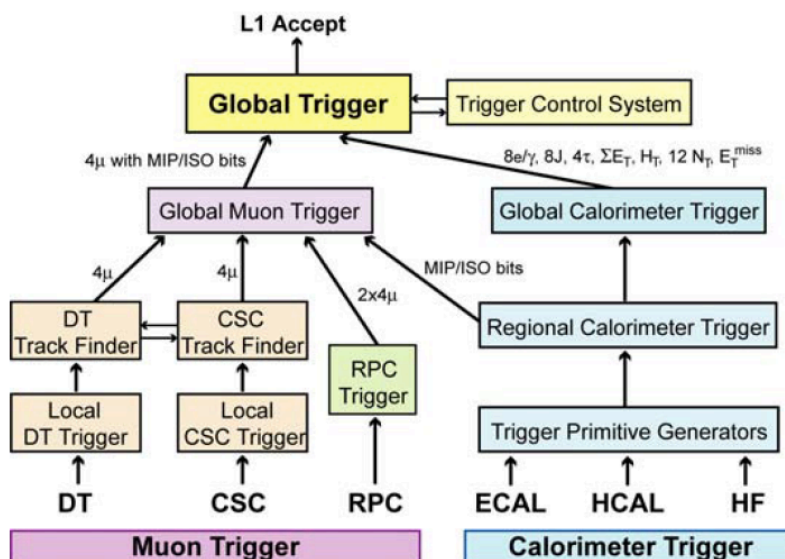


Figure 2.27: Architecture of the Level-1 Trigger.

Muon trigger

All three muon systems - the DT, the CSC and the RPC - take part in the L1 trigger. The barrel DT chambers provide local trigger information in the form of track segments in the ϕ -projection and hit patterns in the η -projection. The endcap CSCs deliver 3-dimensional track segments. All chamber types also identify the bunch crossing from which an event originated. The Regional Muon Trigger consists

of the DT and CSC Track Finders, which join segments to complete tracks and assign physical parameters to them. In addition, the RPC trigger chambers, which have excellent timing resolution, deliver their own track candidates based on regional hit patterns. The Global Muon Trigger then combines the information from the three sub-detectors, achieving an improved momentum resolution and efficiency compared to the stand-alone systems. The initial rapidity coverage of the muon trigger is $|\eta| \leq 2.1$ at the start-up of LHC. The design coverage is $|\eta| \leq 2.4$. Then the Global Muon Trigger improves the trigger efficiency, reduce trigger rates and suppress background by making use of the complementarity and redundancy of the three muon systems. The muon track reconstruction algorithm used by the HLT is seeded by the up to four muon candidates found by the Level-1 Global Muon Trigger, including those candidates that did not necessarily lead to a Level-1 trigger accept by the Global Trigger. The algorithm uses the reconstructed hits built from the digitized signals in the muon system, and constructs tracks according to the Kalman Filter technique[61]. The resulting trajectories are used to validate the Level-1 decision as well as to refine the muon measurement in this Level-2 muon selection. The main feature of the Level-3 muon selection is to add silicon tracker hits to the muon trajectory, thus greatly improving the muon momentum measurement and stiffening the trigger threshold. Isolation criteria can be applied to the muon candidates to provide additional rejection: at Level-2 using the calorimetric energy sum in a cone around the muon, and at Level-3 using the number of pixel tracks in a region around the projected muon trajectory. This suppresses muons from b, c, π , and K decays.

The expected overall efficiency for muons to pass the Level-1 through Level-3 single muon trigger criteria cumulatively as a function of the generated is shown in Figure 2.28. Muons were generated at in the intervals $5 < p_T < 100$ GeV/c and $|\eta| < 2.1$ without any pileup. The average combined Level-1 through Level-3 efficiency without any requirements on the reconstructed p_T is 97%, but is lower in some particular regions because of gaps in the geometrical coverage of the chambers. If we also look at the the p_T we find that the efficiency at Level-3 for relatively high p_T muons is around 95%.

The trigger performance on data were studied with $Z \rightarrow \mu^+\mu^-$ events in this thesis work and the simulated expectation were compared with 7 TeV collision data in Section 6.4.10 of the final chapter of the thesis.

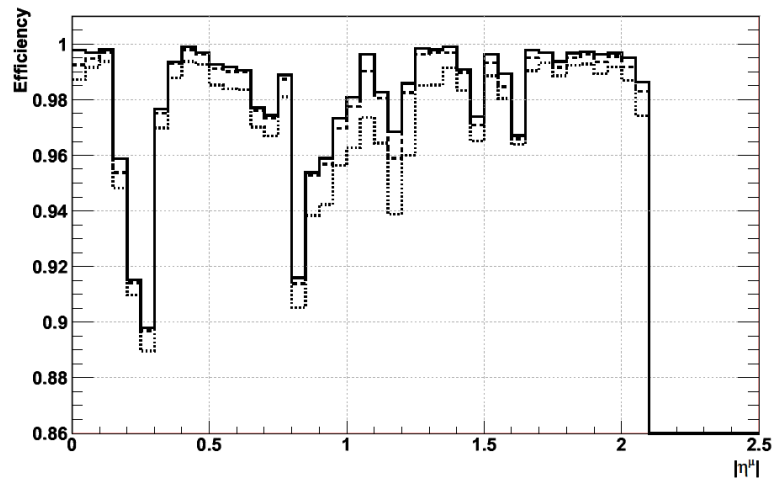


Figure 2.28: Efficiency for single muons to pass the Level-1 (solid), Level-2 (dashed), and Level-3 (dotted) triggers as a function of the generated muon pseudo-rapidity. No thresholds on p_T are applied. Note that the vertical scale starts y-axis. The dips at $|\eta|$ 0.3 and 0.8 are due to gaps in the muon chamber coverage.

Part II

CMS data processing workflow: non event data

Chapter 3

Computing challenges in CMS: data processing work-flow

A key element to the success of the experiment is the adequate design, implementation and smooth operation of the data processing work-flows from the detector to the end user analysis.

This Chapter describes the technical details of the data flow from the detector to the final analysis. It briefly explains first the CMS computing model structure, then it explains the data acquisition system and the various online and offline computing systems, and describes the software and the work-flows used in the data taking chain. In this Chapter particular focus will be given to the CMS condition data work-flow in which I personally contributed to, while the Chapter 5 and 6 will be focused on the reconstruction and the physics analysis algorithms.

The month-long data taking exercise known as the Cosmic Run At Four Tesla (CRAFT) in 2008-2009 and the first months of 7 TeV operation in 2010 represented major tests for these work-flows[62].

3.1 CMS computing model

A major challenge of the experiments is the distribution of the great amount of data produced in order to make them available for analysis. That is not achievable with a standard computing farm connected to the online data taking structure, but one needs to require and develop distributed systems.

The CMS offline computing system must support the storage, transfer and manipulation of the recorded data for the lifetime of the experiment. The system accepts real-time detector information from the data acquisition system at the experimental site; ensures safe treatment of the raw data; performs pattern recognition, event filtering, and data reduction; supports the physics analysis activities of the collaboration.

The system also supports production and distribution of simulated data, and access to conditions and calibration information and other non-event data. The users of the system, and the physical computer centres it comprises, are distributed worldwide, interconnected by high-speed international networks. Unlike previous generations of experiments, the majority of data storage and processing resources available to CMS lie outside the host laboratory. A fully distributed computing model has therefore been designed from the outset. The system is based upon Grid middleware, with the common Grid services at centres defined and managed through the world wide LHC Computing Grid (WLCG) project[63], a collaboration between LHC experiments, computing centres, and middleware providers.

The distributed computing centres available to CMS[64] around the world are configured in a tiered architecture, that behaves as a single coherent system. The computing centre at CERN hosts the Tier-0 of the distributed computing system of CMS. The Tier-0 hosts the initial processing of data coming from the detector and corresponds to about 20% of all computing resources available to CMS. The Tier-1 level takes care of subsequent processing and re-processing work-flows and has approximately 40% of the CMS computing resources available, while the Tier-2 level hosts Monte Carlo (MC) simulation and analysis and uses the remaining $\sim 40\%$ of all CMS computing resources. All streams defined by the online system and the HLT are written in a binary data format, referred to as streamer files. A transfer system copies the streamer files from the online systems at the detector site to the main CERN computing centre to be converted to a ROOT-based event data format[65], split into primary datasets and stored on tape. A first reconstruction is performed and its output is stored in separate datasets. The event content of the detector measurements is called the RAW data-tier and the output of the reconstruction pass is called the RECO data-tier.

3.2 Online system

The CMS data acquisition system is designed to collect and analyse the detector information at the LHC bunch-crossing frequency of 40 MHz. The rate of events to be recorded for offline processing and analysis is about a few hundred Hz. The first-level trigger (L1) is designed to reduce the incoming data rate to a maximum of 100 kHz, by processing fast trigger information coming from the calorimeters and the muon chambers, and selecting events with interesting signatures. The data acquisition system must sustain a maximum input rate of 100 kHz, or a data flow of about 100 GB/s, coming from approximately 650 data sources from the different detector components. The data acquisition system then reduces this rate by a factor of 1000 using a high-level trigger (HLT), a software filtering system running on a large processor farm. The architecture of the system is described in detail in Reference [66].

The CMS HLT algorithms are executed in a farm comprising about 1000 computing nodes, the event filter farm, executing the HLT reconstruction and selection algorithm sequence on individual events in parallel. The products of the HLT execution (e.g. reconstructed physics objects like tracks, muons, jet etc.) can be added to the event before it is sent to storage, thus facilitating later debugging and analysis of the HLT performance. The HLT reconstruction uses the same framework as the offline reconstruction. The HLT configuration (menu) is delivered to the individual processes by the run control system. HLT configurations are managed by a configuration system designed around a relational database abstraction of the individual components (reconstruction modules, filters, etc.) and their parameters. An HLT menu consists of a set of trigger paths, each consisting of a sequence of reconstruction and selection modules. Each path is normally designed to select a specific physics signature (e.g. inclusive muon events). Events accepted by the HLT are delivered to the storage manager system (SM) via the same switched network used for event building (see next Section 3.3).

Routing of individual event data to files in the SM is driven by the definition of output streams, which group events selected by specific HLT paths. Several different streams are normally defined to group together events according to their offline usage (e.g. primary “physics” stream, “express” stream, calibration streams, etc.). The same path can feed multiple streams and hence, in general, individual streams can overlap. Within a stream, sets of paths selecting similar signatures (e.g. “inclusive muons”, etc.) can be further grouped into primary datasets (PDs). A PD is defined as a subset of the stream consisting of the events satisfying a certain group of paths selected by that stream. The PD definition is subsequently used by the Tier-0 repackaging step to split the contents of a stream into its component PDs. Overlaps between streams affect the transfer bandwidth to the Tier-0 while overlaps between PDs primarily affect the disk and tape space consumption of the recorded data. Both the stream and PD definition are intimately connected with the HLT menu, and hence the three are handled as a unit. The same configuration management system is used to maintain and distribute them, and a single “identification key” (a unique string labelling the particular configuration used online) is used by the HLT, the SM, and the Tier-0 to retrieve the relevant portion of the configuration from the database.

3.3 Framework and Event Data Model

The central concept of the CMS data model is the Event. The Event provides access to the recorded data from a single triggered bunch crossing, and to new data derived from it. This may include raw digitised data, reconstructed products, or high-level analysis objects, for real or simulated collisions. The Event also contains information describing the origin of the raw data, and the provenance of all derived data products.

The inclusion of provenance information allows users to unambiguously identify how each event contributing to a final analysis was produced; it includes a record of the software configuration and conditions/calibration setup used to produce each new data product. Events are physically stored as persistent ROOT files[65]. The Event is used by a variety of physics modules, which may read data from it, or add new data, with provenance information automatically included. Each module performs a well-defined function relating to the selection, reconstruction or analysis of the Event. Several module types exist, each with a specialised interface. These include: event data producers, which add new data products into the event; filters used in online triggering as well as offline selection; analysers, producing summary information from an event collection; and input and output modules for both disk storage and data acquisition.

Modules are separated from the computing environment, execute independently from one another, and communicate only through the Event; this allows modules to be developed and verified independently. A complete CMS application is constructed by specifying to the Framework (the basic structure underlying the CMS data handling and analysis software) one or more ordered sequences of modules through which each Event must flow, along with the configuration for each module. The Framework configures the modules, schedules their execution, and provides access to global services and utilities (Figure 3.1).

The same framework is used for the HLT and the offline.

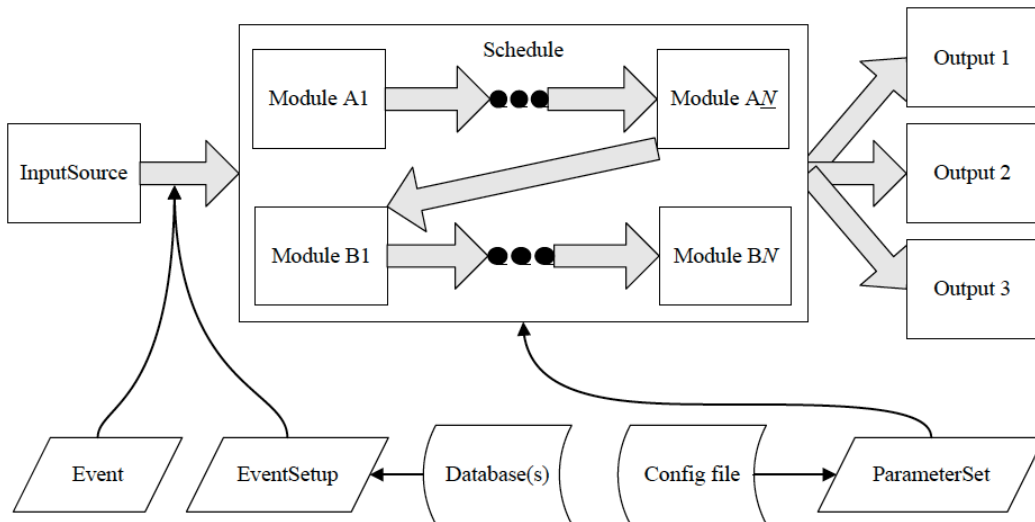


Figure 3.1: Modules within the CMS Application Framework.

3.4 Data Formats and processing

In order to achieve the required level of data reduction, whilst maintaining flexibility, CMS makes use of several event formats with differing levels of detail and precision. Other specialised event formats are used for heavy-ion data. The process of data reduction and analysis takes place in several steps, typically carried out at different computer centres as illustrated in Figure 3.2.

In addition to event data recorded from the detector, a variety of non-event data is also required in order to interpret and reconstruct events, see Chapter 4 for a detailed discussion of non event data, while this section will be devoted to event data description.

RAW format

RAW events contain the full recorded information from the detector, plus a record of the trigger decision and other meta-data. RAW data is accepted into the offline system at the HLT output rate (nominally 300 Hz for pp collisions). An extension

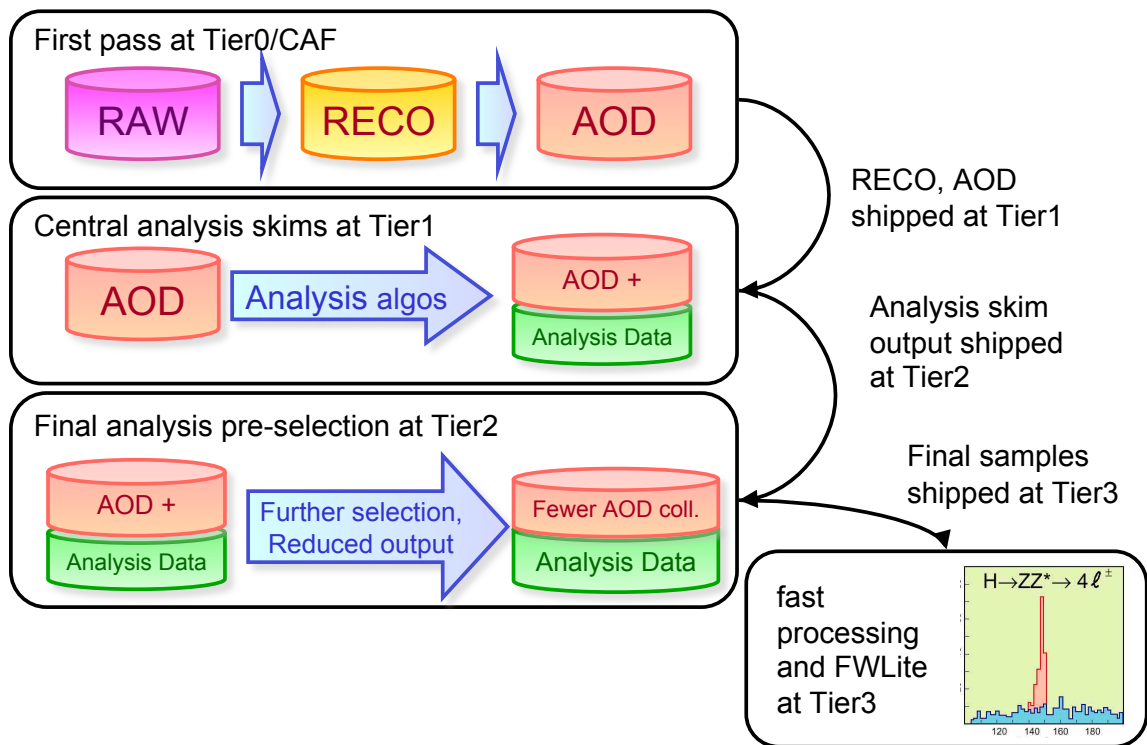


Figure 3.2: CMS data reduction work-flow.

of the RAW data format is used to store the output of CMS Monte Carlo simulation tools. The RAW data is permanently archived in safe storage, and is designed to occupy around 1.5 MB/event (2 MB/event for simulated data, due to additional Monte Carlo truth information). The RAW data will be classified by the online system into several distinct primary datasets, based upon the trigger signature. Event classification at the earliest possible stage has several advantages, including the possibility of assigning priorities for data reconstruction and transfer in the case of backlog, and balancing of data placement at centres outside CERN. CMS will also define one or more flexible “express streams” used for prompt calibration and rapid access to interesting or anomalous events.

RECO format

Reconstructed (RECO) data is produced by applying several levels of pattern recognition and reduction algorithms to the RAW data. These algorithms include: detector-specific filtering and correction of the digitised data; cluster- and track-finding; primary and secondary vertex reconstruction; and particle ID, using a variety of algorithms operating on cross-detector information. Reconstruction is the most CPU-intensive activity in the CMS data processing chain. The resulting RECO events contain high-level physics objects, plus a full record of the reconstructed hits and clusters used to produce them. Sufficient information is retained to allow subsequent application of new calibrations or algorithms without recourse to RAW data, though basic improvements in pattern recognition or event formats will probably require re-production of the RECO data at least once per year. RECO events are foreseen to occupy around 0.5 MB/event.

AOD format

AOD (Analysis Object Data) is the compact analysis format, designed to allow a wide range of physics analyses whilst occupying sufficiently small storage so that very large event samples may be held at many centres. AOD events contain the parameters of high-level physics objects, plus sufficient additional information to allow kinematic refitting. This format will require around 100 kB/event, small enough to allow a complete copy of the experimental data in AOD format to be held at computing centres outside CERN. AOD data is produced by filtering of RECO data, either in bulk production, or in a skimming process which may also filter a primary dataset into several analysis datasets.

3.5 Data quality monitor

Data quality monitoring (DQM) is critically important for the detector and operation efficiency, and for the reliable certification of the recorded data for physics analyses. The CMS-wide DQM system comprises:

- tools for the creation, filling, transport and archival of histogram and scalar (i.e. single numbers) monitor elements, with standardised algorithms for performing automated quality and validity tests on value distributions;
- monitoring systems, live online for the detector, the trigger, and the data acquisition hardware status and data throughput, for the offline reconstruction and for validating calibration results, software releases and simulated data;
- visualisation of the monitoring results;
- certification of datasets for physics analyses;
- organisation and operation of the activities, including shifts and tutorials.

The online DQM system consists of a number of consumer applications, labelled as DQM in Figure 3.3, usually one per subsystem, which receive event data through a storage manager event server and fill histograms at an event rate of 10-15 Hz. In addition, a small number of histograms is filled in the HLT filter units, which process events at up to 100 kHz. These histograms are shipped out to DQM consumer applications periodically.

The offline DQM system accumulates monitoring data from several work-flows in CMS, namely Tier-0 prompt reconstruction, re-reconstruction at the Tier-1s and the validation of the alignment and calibration results, the software releases, and all the simulated data. CMS has standardised the monitoring of the event data processing into a two-step work-flow:

1. The histogram monitor elements are created and filled with CMS event data information. The histograms are stored along with the processed events into the normal output event data files. When the CMS data processing systems merge output files, the histograms are automatically summed to form the first partial result.
2. At the end of the data processing the histograms are extracted from the job output data files and summed together across entire runs to yield full event statistics. The final histograms are then used to calculate efficiencies and are checked for quality, by making comparisons with reference distributions. The histograms, certification results, and quality test results are saved into a ROOT

file, which is then uploaded to a central DQM GUI web server. In the web server, the files are merged and backed up to tape; recent data are kept cached on disk for several months.

3.6 Analysis model and tool

The CMS analysis model is data-location driven, i.e. the user analysis runs where data are located. The related work-flow is mainly characterised by the following steps: interactive code development using small data samples; job preparation and configuration to run over higher statistics (hence to access the whole dataset or a significant part of it); and interactive analysis of the obtained results. With the increasing complexity of the computing infrastructure, the implementation of such a work-flow became more and more difficult for the end user. In order to provide the physicists an efficient access to the distributed data while hiding the underlying

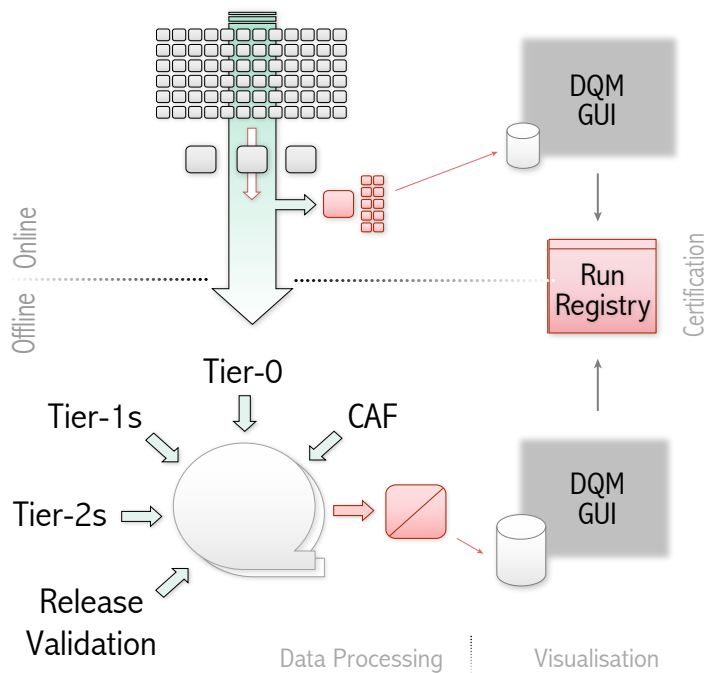


Figure 3.3: Sketch of the DQM system, consisting of branches for online and offline monitoring.

complexity, CMS developed and deployed a dedicated tool named CMS Remote Analysis Builder (CRAB)[77].

During the thesis all the steps of the analysis work-flow described in this Chapter and illustrated in Figure 3.2 were followed to get the results on the physics channel which is the object of my work (see Chapter 6):

- Data were analyzed at Tier-2 center using CRAB. In the specific case of this thesis the data input is the physics stream known as “muon dataset”, because it is produced from the muon trigger paths, and the Tier-2 center exploited is the Legnaro Italian center;
- a reduced output with only few AOD collections were saved locally at the university of Naples Tier-3;
- fast processing on this reduced outputs is performed to produce the physics plots and results. This last and fast step has been repeated as many times as needed with different selection cuts.

The AOD reduced collections step is made possible through the use of the CMS Physics Analysis Toolkit (PAT[67]), a high-level analysis layer enabling the development of common analysis efforts. In the analysis described in the last Chapter, we reconstructed composite particle (in the specific case di-muon) candidates using one of the modules provided by PAT and illustrated graphically in Figure 3.4. So only events containing these di-muon collection were saved, and only the information needed in the following steps of the analysis were saved together with the collections.

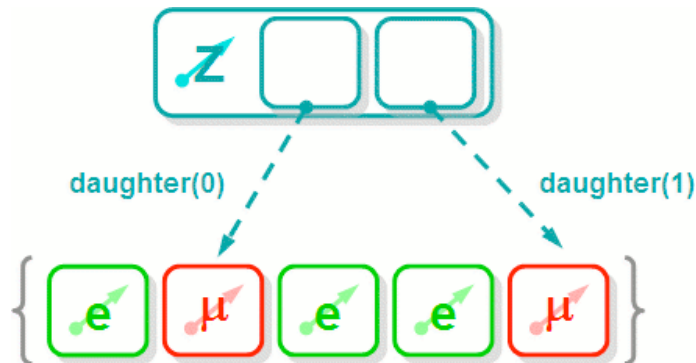


Figure 3.4: Illustration of the reconstruction of composite particle candidates in PAT.

Chapter 4

Non Event Data in CMS

Most of the first year of my phd program and later was spent at CERN working in the CMS database team to work and develop tool to populate and monitor the activity of the CMS database containing the calibration constants (see Sub-section for a description of this type of data). This Chapter aims to describe the physics and computing motivation to such a work, the tool built and deployed and the achievement done for the CMS community.

In the CMS experiment, the non event data needed to set up the detector, or being produced by it, and needed to calibrate the physical responses of the detector itself are stored in ORACLE databases (a commercial relational database system[68]). The large amount of data to be stored, the number of clients involved and the performance requirements make the database system an essential service for the experiment to run. This Chapter describes the CMS condition database architecture, the data-flow and PopCon, the tool built in order to populate the offline databases. Finally, the first experience obtained during the 2008 and 2009 cosmic data and 2010 7 TeV collision data taking are presented.

The large amount of data needed to set up the detector (tens of GBytes) and produced by it (few TBs per year) makes the database system a service which is essential for CMS data-taking operation. The CMS collaboration decided to converge towards an unique database technology and a set of central software tools supported by the experiment for all data taking.

The current layout of the database model and data-flow was developed after close interaction with the CMS subsystems.

The two most important requirements identified by CMS are:

- CMS has to be able to operate without network connection between LHC Point 5 (P5) and the outside world (CERN network included). Therefore CMS must own an independent database structure based at P5.
- The offline condition data work-flow has to fit a multi-tier distributed structure

as used for event data.

The Database Project group, with the help of CERN Information Technology (IT) group and in collaboration with all the CMS sub-projects, designed a system based on 3 different databases, all based on the ORACLE technology.

- Online Master Database System (**OMDS**) is the online master database located at P5 on the CMS online network. It stores the configuration of the detector and the non event data (condition data) produced by the sub-systems like slow control, electronics, data acquisition and trigger data. It is a purely relational database.
- Offline Reconstruction Condition database for ONline use (**ORCON**), on the online network, stores all the offline condition data required online by the High Level Trigger (HLT) and offline for the event data reconstruction. It also contains conditions needed offline for data quality indication and for more detailed offline analysis. ORCON serves only as an intermediate storage of the latest offline condition data. The entire history of off line condition data is stored in ORCOFF. The data contained in it are written using the POOL-ORA[69] technology and are retrieved by the HLT programs as C++ objects.
- Offline Reconstruction Condition database for OFFline use (**ORCOFF**) is the master offline database located at the CERN Tier-0 site and it contains a copy of ORCON made through ORACLE streaming. ORCOFF contains the entire history of all CMS condition data and serves prompt reconstruction as well as the condition deployment service to Tier-1/Tier-2 sides as input source. Data contained in it are retrieved by the reconstruction algorithms as C++ objects.

4.0.1 Non-Event Data Description

For each sub-detector, the non-event data to be stored in the CMS databases can be classified in different groups, according to their needs for meta-data (i.e., data to describe the data):

- **Construction data** During the construction of the detector, data are gathered from both the production process and the produced items. Some of the construction data also belongs to the data types described below, and therefore were moved to the common data storage at the end of construction. The different CMS sub-detectors agreed to keep their construction data available for the lifetime of the experiment in order to be able to trace back production errors.

- **Equipment management data** Detector items should be tracked in order to log their history of placements and repairs. The classification of CERN as INB (Installation Nucleaire de Base[70]) requires, in addition, to keep a continuous trace of the location of irradiated items. Equipment management data contain, therefore, the location history of all items being installed at the experiment, in particular detector parts as well as off detector electronics. Hence, the required meta-data must be time validity information. This data are stored in OMDS.
- **Configuration data** The data needed to bring the detector into any running mode are classified as configuration data. They comprise voltage settings of power supplies as well as programmable parameters for front-end electronics. Configuration data require a version and time validity information. This data are stored in OMDS.
- **Condition data** The data describing the state of any detector subsystem are defined as condition data. These conditions are measured online and are stored in OMDS. They include data quality indicators such as bad channel lists and settings of the detectors needed offline such as calorimeter channel threshold (pedestals). Condition data in OMDS are used in the online system for post mortem analysis of detector errors. Condition data needed for HLT and offline reconstruction are uploaded in ORCON, and must be described by a version and the time validity information corresponding to the set of data for which they are measured.
- **Calibration data** The data describing the calibration and the alignment of the individual components of the different sub-detectors are labeled as calibration data. These quantities (such as drift velocities, alignments constants, etc.) are evaluated by running dedicated algorithms offline. Since they are needed by HLT and for offline reconstruction, they appear only in the offline databases (ORCON and ORCOFF). Calibrations must match the corresponding raw data coming from the collision events revealed by the detector. Calibration data can be grouped by the version and the time range in which they are valid.

4.1 The Database Architecture

Different data usage and access between online and offline determines the overall database architecture for the CMS experiment. In the online network, data are mainly written into the database, so that the time for a database transaction to be committed is critical, while, in the offline network, data are mainly read from the databases. Moreover, the online data are being written at random times, while the offline data must be synchronized with the events. Since online data are used

for error tracking, different data items must be accessible in order to be compared between each other; on the other hand, offline data must be grouped before they are read, so that they can be decoded according to predefined rules.

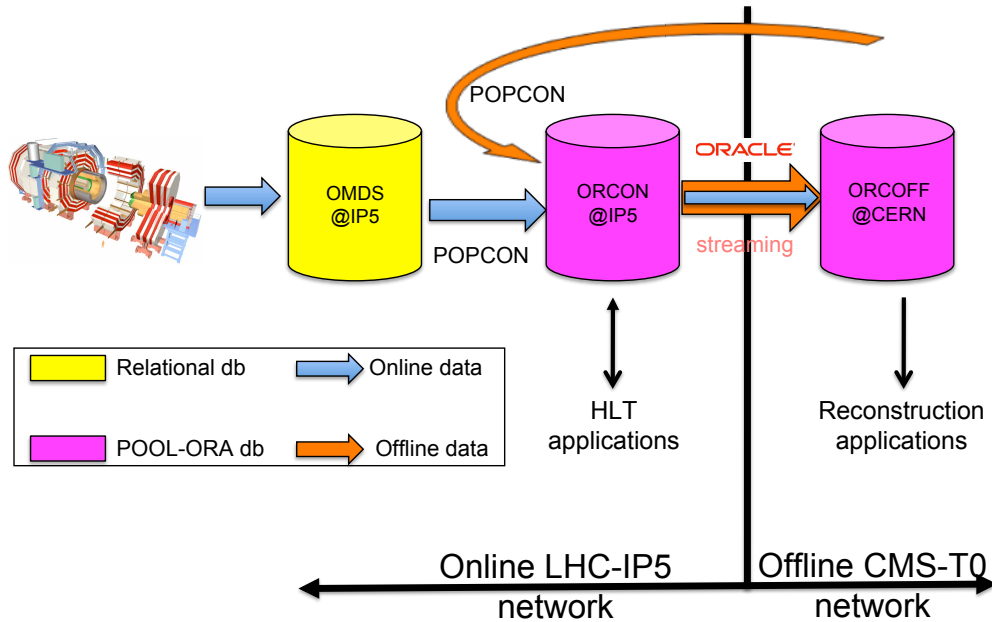


Figure 4.1: Condition databases architecture.

The general non-event data flow can be described as follows (see Figure 4.1): every sub-project calculates and measures in advance all the parameters needed to setup its hardware devices, mainly related to the detector, data acquisition and trigger. Hence, configuration data are prepared using the equipment management information, for both hardware and software. Different hardware setups can be stored at the same time in the configuration database, but only one will be loaded before the run starts. During data taking, the detector produces many kind of conditions, to be stored in OMDs, from the slow control[71] and from other tools like the data acquisition tools, Run Control and data quality monitoring. Part of OMDs data, needed by the HLT and offline reconstruction, will be transferred to ORCON. A software application named PopCon (Populator of Condition Objects) operates the online to offline condition data transfer, and encapsulates the data stored in relational databases. PopCon adds to non-event data the version and the time range in which they are valid, so that they can be retrieved by both the HLT and the offline software.

In addition to condition data transferred from OMDs to ORCON, calibration and alignment data determined offline are also written to ORCON, using again PopCon. Finally, data are transferred to the offline using ORACLE streaming. For massively

parallel read-access, the databases are interfaced with a cache system referred to as Frontier, which in case of ORCOFF is the mechanism used to distribute conditions data to the Tier-1 and Tier-2 centres outside CERN. Caching servers are used to cache requested objects to avoid repeated access to the same data, significantly improving the performance and greatly reducing the load on the central database servers. Further details can be found in Reference [72].

A better understanding on how the detector works will require a new version of calibrations. When it will be available, it will be uploaded into ORCON using PopCon, and then streamed offline to ORCOFF.

4.1.1 The Online Master Database

In the CMS experiment, the non event data needed to set up the detector, or being produced by the detector itself, is stored in OMDS. The online database must allow for accessing individual, ideally self explaining data items: hence a pure ORACLE access and manipulation structure has been chosen for OMDS.

The data size is expected to become very large (several TBs), and, since condition data will constantly flow into the database, the time needed to store these data in OMDS is a critical issue. To fulfill these requirements, each sub-detector has designed its own database schema, reflecting as far as possible the detector structure.

The total amount of data stored in OMDS is about 1.5 TB in 100 days of data taking. This rate was extrapolated from the 2008 and 2009 cosmic runs: in particular, in the account on OMDS for non-event data coming from the electromagnetic calorimeter (ECAL), about 5 GB of data per day are stored.

4.1.2 Offline database

As shown in Figure 4.1 and already said in the previous section, the CMS database infrastructure envisages two offline databases intended for condition data: ORCON and ORCOFF. ORCON possess identical “schemas” as ORCOFF, the latest being optimized for the offline usage.

Together with the production databases, CMS users can also use a “development” and an “integration” database, intended for tests, and accessible from the offline network:

The data access (both insertion and retrieval) is controlled by an C++ application. In the offline databases, only a subset of configuration data and condition data, as well as all calibration data, must be stored. All these data need a tag, labeling their version, and an interval of validity for describing their time information. The interval of validity (IOV[73]) is the contiguous (in time) set of events for which non-event data are to be used in reconstruction. According to the use-case, the IOV will be defined in terms either of Global Positional System (GPS) time (mainly for condition

data) or “run-number” and “lumi-sections”¹ range (usually for calibrations) Whilst the IOV for some electronic related conditions (e.g. pedestals and noises) is identical to the time interval in which these data were used in the online operations, some calibration data may possess an IOV different from the time range in which they were defined. For this reason, the IOV assignment for a given set of condition data is carried out at the offline level. Each calibration object (in the database language known as “payload”) stored in ORCOFF, is indexed by its IOV and a tag, a label describing the calibration version, while the data themselves do not contain any time validity information.

The matching with the raw data from the collision events is indeed possible via these meta-data: the reconstruction algorithms for the analysis of a given run query the offline condition data corresponding to the same run grouped through a set of tags, called *global tag*[73].

The policy established by the Database Project for the CMS community is to write any condition/calibration data in ORCON; the data are then copied to ORCOFF using the ORACLE streaming tool.

The size of condition data stored in ORCON and ORCOFF, where only a subset of condition data will be uploaded, is decreased by a factor of 20 with respect to OMDS. This is a great success of the entire architecture.

In Section 4.1.3 the PopCon framework is described, while in Section 4.2.3 more information about the online-to-offline (O2O) transfer operated by PopCon is given.

4.1.3 PopCon

PopCon[3] [2] is a mini-framework within the CMS software CMSSW[74] that transfers the condition objects from a user-defined data source to the offline databases.

PopCon is integrated in the infrastructure of the CMS software Framework[75]. It is possible to use different data sources such as databases, ROOT files, ASCII files, and so on. A C++ object type (built in type, structure, class, template class) which contains the non event data, must be defined into the CMS software framework. For each condition object class a PopCon application is created.

The core framework consists of three classes (two of them are C++ templates), as can be seen in Figure 4.2:

- PopCon
- PopConSourceHandler
- PopConAnalyzer

¹A CMS “run” is an interval of time in which the CMS detectors is active with the same online configuration. One “lumi-section” corresponds to 23 s and is the smallest part of a CMS “run”. A run usually lasts hundreds of lumisections.

Once the C++ condition object is embedded into CMSSW, the “detector user” provides the code which handles the data source and specifies the destination for the data, writing a derived class of PopConSourceHandler, where all the online (source handling) code goes. The user instantiates the objects, provides the IOV information for such objects and configures the database output module. The PopCon configuration file associates the tag name defined according to some specific rules, to the condition object. Once the object is built, the PopCon application writes the data to the specified account in the offline database. Sub-detector code does not access the target output database: it only passes the objects to the output module.

The analyzer object holds the source handling object. It also serves to implement some additional functionality such as:

- Locking mechanism.
- Transfer logging.
- Payload verification (IOV sequence).
- Application state management.
- Database output service.

The writer in PopCon iterates over the container of user objects and stores it in the user-configured data destination.

Any transaction towards ORCON is logged by PopCon, and the process information is sent to a database account as well. A monitoring tool for this information was developed, in order to check the correctness of the various transactions, and to keep trace of every upload for condition data, see Section 4.3.

4.2 First Experience in Operating the Population on the Condition Databases

In the 2008 and 2009 global runs (with and without the magnetic field) and in the 2010 7 TeV collision runs, the great majority of the condition data was transferred offline using a PopCon application. Great effort was devoted by the CMS database project team to the integration of all the software and to the infrastructural chain used to upload the calibration constants into the CMS condition databases. Many tools were provided to help the sub-detector responsible people to populate their database accounts. A central procedure, based on an automatic up-loader into ORCON on a dedicated machine in the online network, was successfully deployed during 2008, and became the recommended way to populate ORCON since 2009 data taking and later on.

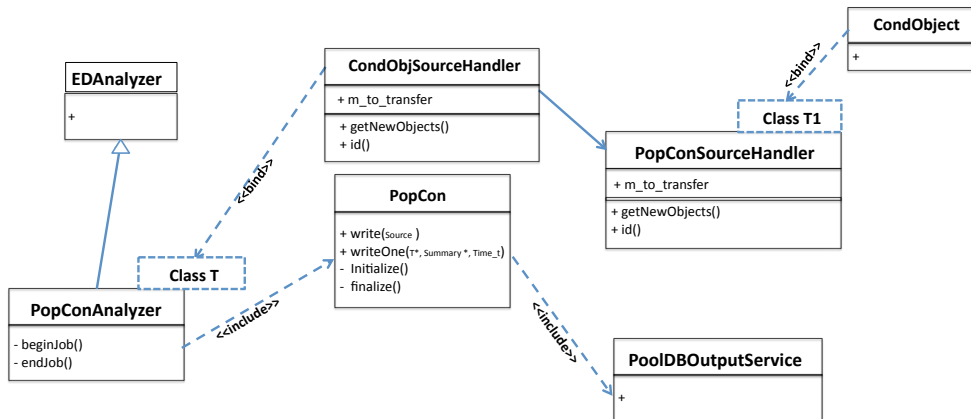


Figure 4.2: Schema of the classes for the PopCon package.

4.2.1 Alignment and calibration work-flow

The basic offline work-flow for alignment and calibration in CRAFT and 2010 collision data period is illustrated in Figure 4.3. Event information relevant for alignment and calibration are streamed from the CMS detector site via the standard physics event stream, and via a special calibration stream and streams with special event content, labeled “AlCaRaw” dedicated to particular calibration procedures. Events from these streams pass the conversion to the ROOT based event data format at the Tier-0 and in the case of the physics stream enter the prompt reconstruction process. The reconstructed data are then skimmed to create a series of “AlCaReco” datasets that are transferred to the CAF to be used as input to alignment and calibration algorithms. The AlCaReco datasets are designed to contain only the minimal amount of information required by the associated alignment and calibration work-flows. The skims producing them performs both event selection, starting from a selection based on HLT bits, and reduction of event content. The alignment and calibration work-flows uses the AlCaReco datasets to generate alignment and calibration constants that are validated and uploaded to the conditions database. If new constants are generated a re-reconstruction at the Tier-1 sites is performed, generating new AlCaReco datasets that are used in turn as input to the next series of improvements on alignment and calibration constants.

4.2.2 Condition Objects Written Using PopCon in 2009

As stated before, each piece of condition data (pedestals, Lorentz angles, drift time, etc.) corresponds to a C++ object in the CMS software. Each object is associated with a PopCon application which writes the payload into ORCON. Table 4.1 lists all the CMS condition objects used in 2009, grouped according to the subsystem they belong to. For each object the type, the approximate data size in ORCON and the upload frequency are also reported.

4.2.3 Central Population of the Condition Databases

A central procedure was set up in 2008, and used since then, for populating the CMS condition databases: it exploits a central account, explicitly devoted to the deployment of tools and services for the condition databases, in the CMS online network. On that account, a set of automatic jobs was centrally set up for any single sub-detector user, in order to both populate ORCON and monitor any transactions to it.

Two possibilities are given to users:

1. to run automatically the application that reads from any online source, assigns tag and interval of validity, and uploads the constants into ORCON (mainly

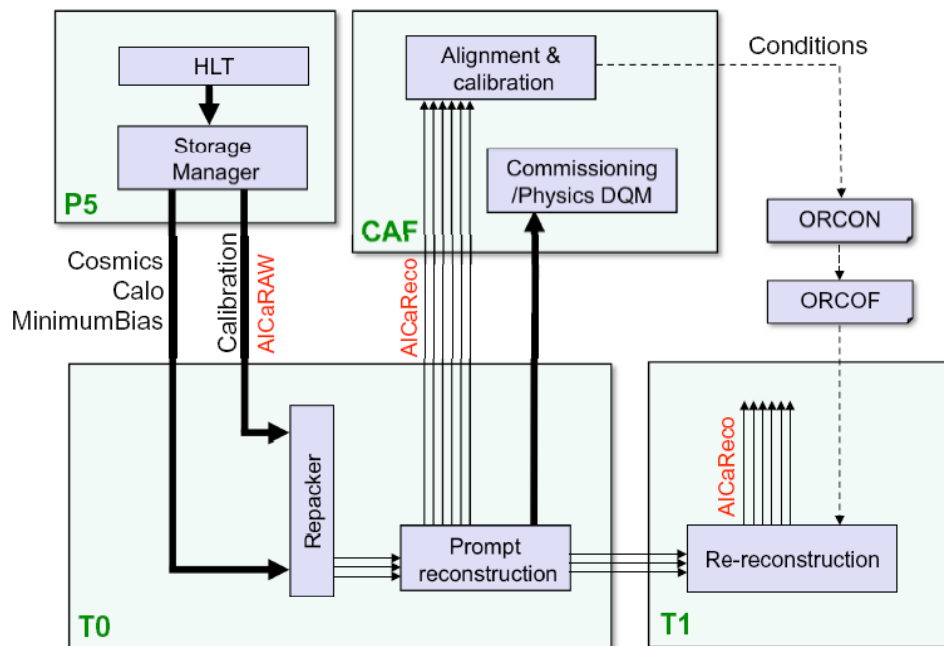


Figure 4.3: Offline work-flow for alignment and calibration used in CRAFT and 2010 collision data period.

Table 4.1: 2009 CMS condition objects list: “conf.” stays for configuration, “calib.” for calibration and “cond.” for condition.

| Subsystem | Name | Type | Data size | Frequency |
|-----------|---------------------------|----------------|-----------|-------------------------|
| Pixel | SiPixelFedCablingMap | online conf. | 1K | once (before the run) |
| | SiPixelLorentzAngle | offline calib. | 1MB | each run (if different) |
| | SiPixelCalibConfiguration | online calib. | 5KB | each calibration run |
| Tracker | SiStripFedCabling | online config. | 1K | once |
| | SiStripBadStrip | online cond. | 1MB | each run (if different) |
| | SiStripThreshold | offline calib. | 1MB | each run (if different) |
| | SiStripPedestals | offline calib. | 1MB | each run (if different) |
| | SiStripNoise | offline calib. | 1MB | each run (if different) |
| Ecal | EcalPedestals | online calib. | 2MB | daily |
| | EcalLaserAPDPNRatios | online calib. | 2MB | hourly |
| Hcal | HcalElectronicsMap | online conf. | 1MB | once (before the run) |
| | HcalGains | offline calib. | 1MB | each run |
| | HcalPedestals | offline calib. | 1MB | each run |
| | HcalPedestalsWidths | offline calib. | 1MB | each run |
| | HcalQIEData | online calib. | 1MB | each run |
| CSC | CSCChamberMap | online conf. | 10KB | monthly |
| | CSCCrateMap | online conf. | 10KB | monthly |
| | CSCDDUMap | online conf. | 10KB | monthly |
| | CSCChamberIndex | online conf. | 10KB | monthly |
| | CSCGains | offline calib. | 2MB | each run |
| | CSCNoiseMatrix | offline calib. | 2MB | each run |
| | CSCPedestals | offline calib. | 2MB | each run |
| DT | DtReadOut | online config. | 10MB | once |
| | DtCCBConfig | online config. | 100KB | once (before the run) |
| | DtT0 | offline calib. | 10MB | rare |
| | DtTTrig | offline calib. | 1MB | at trigger change |
| | DtMTime | offline calib. | 1MB | daily |
| RPC | RPCMap | online config. | 10MB | once |
| | L1RPCCConfig | online config. | 10MB | once |
| | RPCCCond | online cond. | 10MB | daily |
| DAQ | RunSummary | run cond. | 10KB | run start/end |

for condition data). The time interval of the automatic jobs is negotiated with the users;

2. to use the account as a drop-box: users copy the calibrations in the light format into a dedicated folder for each sub-detector, and then these data are automatically exported in ORCON (mainly for offline calibration data).

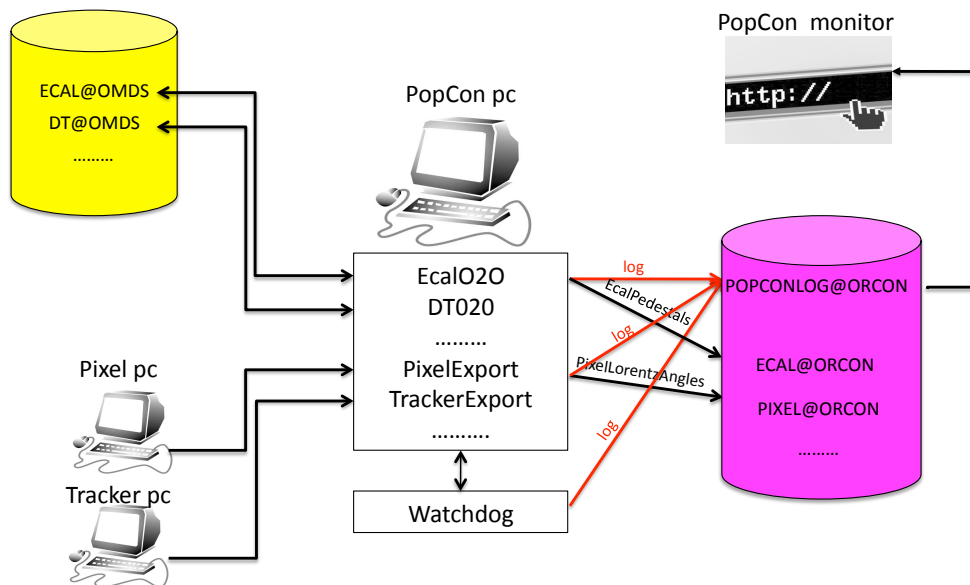


Figure 4.4: Schematic illustration of the central system used to populate ORCON, and of the web monitoring system.

Figure 4.4 shows a sketch of the central system used to populate the condition database, and to ensure the synchronization between online and offline. Each sub-detector responsible person may transfer the payload onto the central PopCon machine, that then automatically manages the exportation into the ORCON database (using a specific set of sub-detector export scripts). Other automatic scripts check to see if new conditions have appeared in the online table, and, if so, perform the data

transfer from OMDS to ORCON. The PopCon applications transfer each payload into the corresponding account, and create some log information which are subsequently stored in the PopConLog account on ORCON itself.

Each automatic job is associated with a monitor tool that checks its status (“watchdog” in the Figure 4.4). The job monitoring information are also logged into the PopConLog account on ORCON.

4.3 PopCon Web Based Monitoring

A dedicated web based application, *PopCon monitoring*[76] was set up on a CMS web server in order to look at all the logged information, hence monitoring the activity on the condition databases. The architecture of PopCon monitoring is illustrated in Figure 4.5.

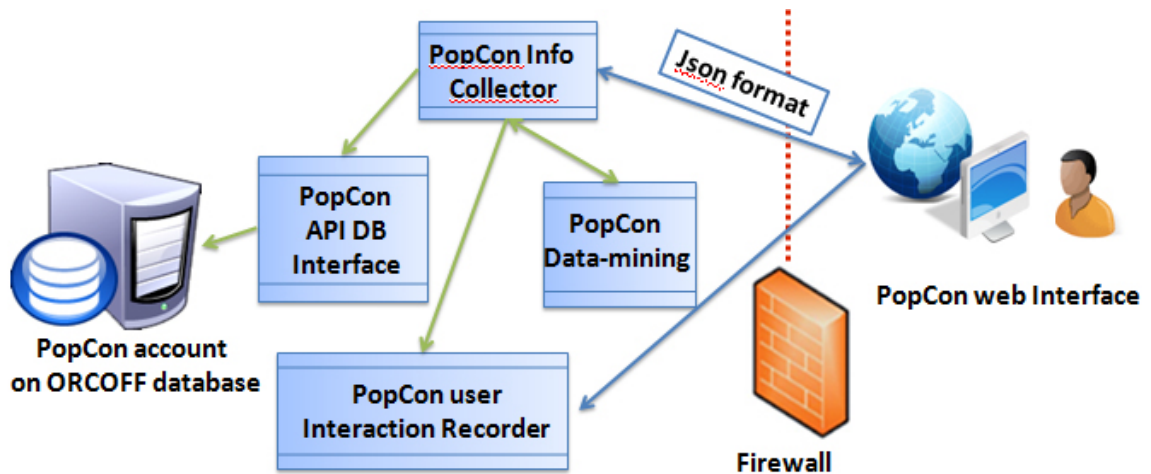


Figure 4.5: PopCon monitoring Architecture

The PopCon web interface represents information about database transactions in different types: both charts and tables. The main page shows ORCOFF activity during the day, and a summary of the the transactions status (success or error), the payload, iov and tag transferred. Fig. 4.6 shows how it appears during a recent day with LHC 7 TeV collision.

A user can easily add or remove columns by clicking the check-box and columns can also be sorted. Information could be grouped according to different filters.

The PopCon web interface represents information about database transactions in different types: both charts and tables. A user can easily add or remove columns by

| EXEC TIME | IOV TAG | Payload container | Payload Name | DESTINATION DB | Exec Mess | PAYLOAD INDEX | USER TEXT |
|----------------------|----------------------|-------------------|----------------|---|-----------|---------------|--|
| 23-05-10 08:03:41 | SiStripDetVOff_3X_v2 | SiStripDetVOff | SiStripDetVOff | oracle://CMSDEVR_LB /CMS_COND_STRIP | OK | 491 | PopCon v3.0; USER=popcondev;HOSTNAME=srv-C2C05-15; PWD=/ntfshome0/popcondev/SiStripJob/CMS_SW_3_5_4_DCS /src/CalibTracker/SiStripDCS/test;first payload Since 5474311925540143424,last payload Since 5474312195971083072;@@@ Number of payloads transferred 2. time #LV off #HV off 2010-05-23T04:06:09.797000 84 15034 2010-05-23T04:07:12.645000 95 95 |
| 23-05-10 08:03:40 | SiStripDetVOff_3X_v2 | SiStripDetVOff | SiStripDetVOff | oracle://CMSDEVR_LB /CMS_COND_STRIP | OK | 490 | PopCon v3.0; USER=popcondev;HOSTNAME=srv-C2C05-15; PWD=/ntfshome0/popcondev/SiStripJob/CMS_SW_3_5_4_DCS /src/CalibTracker/SiStripDCS/test;first payload Since 5474311925540143424,last payload Since 5474312195971083072;@@@ Number of payloads transferred 2. time #LV off #HV off 2010-05-23T04:06:09.797000 84 15034 2010-05-23T04:07:12.645000 95 95 |
| 23-05-10 08:01:58 | DTHVStatus_v1 | DTHVStatus | DTHVStatus | oracle://cms_orcon_prod /CMS_COND_31X_DT | OK | 1376 | PopCon v3.0; USER=popcondev;HOSTNAME=srv-C2C05-15; PWD=/ntfshome0/popcondev/DTJob/CMS_SW_3_5_4_DCS /src;first payload Since 5474300268202103080,last payload Since 5474341298025175768; |

Figure 4.6: Screen-shot of part of ORCOFF transactions activity during a typical high-activity day of CMS 2010 data taking.

clicking the check-box and columns can also be sorted. Information could be grouped according to different filters.

In addition, two other web pages, very useful for database transaction monitoring and error detection, are produced:

1. an activity summary, in which the number of ORCON transactions, the sub-system involved, the IOV and tag can be displayed, according to the users' requests. An example is shown in Figure 4.7.
2. the logs of all the central scripts, as produced by the watchdog tools. Looking at those logs, the correct behaviour of the central uploader machine is controlled, so that an alarm system, based on that information, can recognize if some exports were not successful and, eventually, inform the end-user of the error occurred. If that is the case the users can recognize the error and request a new transaction, thus preventing the online to offline calibration synchronization to fail, and the assignment of wrong condition to the reconstructed data and the HLT. A screen-shot of the page is shown in Figure 4.8.

Figure 4.7 reports all the transactions towards the condition database accounts occurring in a two months of 7 TeV collision data taking in 2010. As the summary plot points out, almost all sub-detectors used PopCon to upload calibration constants to the condition databases. An average of more than one hundred PopCon applications per day were run during the test runs in 2008-2009 and 2010 data taking, hence more than one hundred connections per day to the condition databases took place.

During the whole year 2009 commissioning exercises, the total amount of condition data written in the production database was approximately 1 TB. The same rates and volumes has been experienced the first two months of LHC collision data.

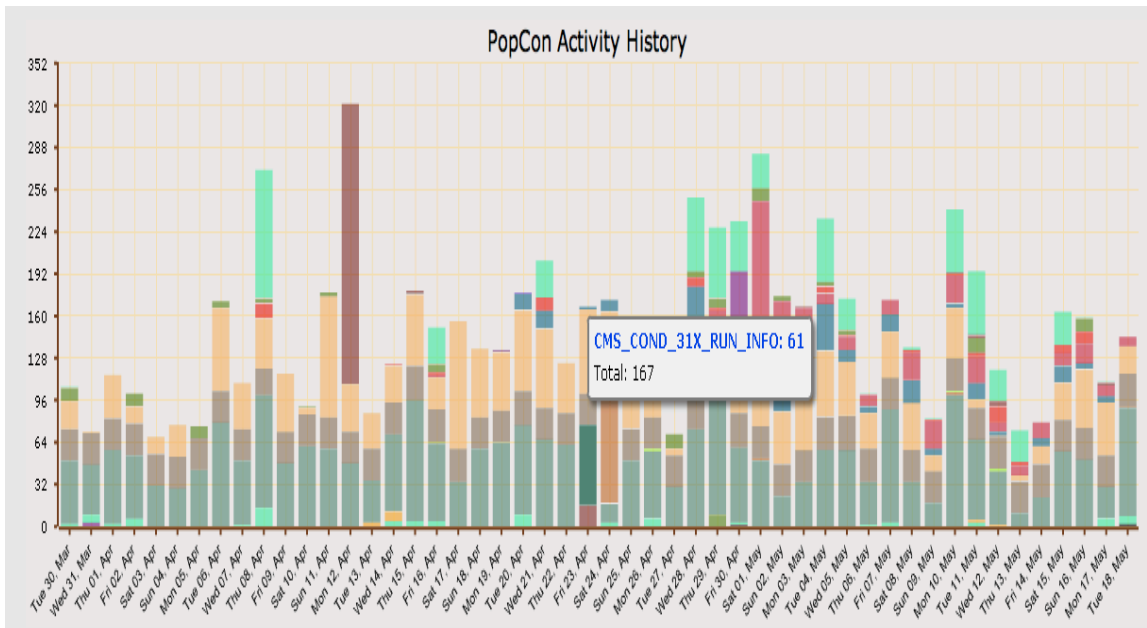


Figure 4.7: PopCon activity between March and beginning of May 2010.

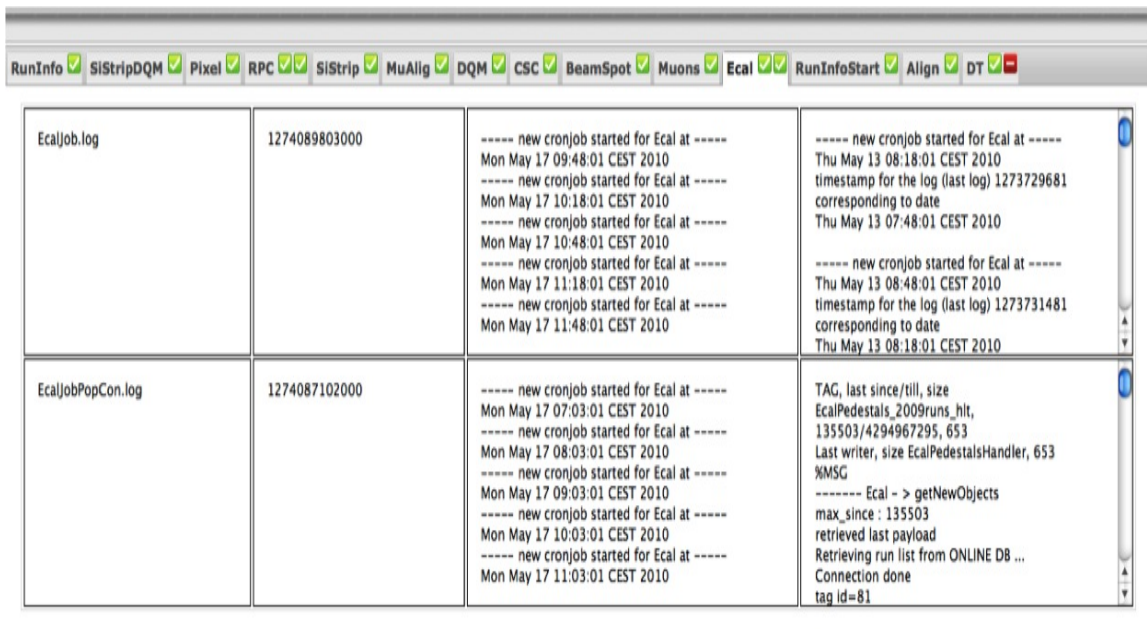


Figure 4.8: Screen-shot of the web page produced in a typical high-activity 2010 collision day by the monitoring system that checks the watchdog tools for the automatic population of ORCON. Different colours helps to identify, quickly, the seriousness of the problem.

Moreover, no network problems, neither for the online-offline streaming, nor for Frontier were detected. All the conditions and calibrations were properly evaluated in 2008-2009 and have been evaluated in the 2010 7 TeV collision runs and reprocessing, leading to several global tags used for the reconstruction and the analysis of the collision data by the whole CMS community.

Part III

$Z \rightarrow \mu^+ \mu^-$ cross-section measurement
and luminosity monitor

Chapter 5

Muon and $Z \rightarrow \mu^+ \mu^-$ reconstruction in CMS

Many of the rare physics processes searched for in hadron collider events are characterized by leptons in the final state. Leptons provide a striking signature in a large background of jets, allowing for the identification of rare signal events. This is especially true for muons and indeed the CMS experiment also in its name focus the attention on physics reach with muons. Thus it is crucial for CMS to be able to detect muons with high efficiency, low fake-rate from backgrounds, and excellent momentum resolution.

5.1 Muon identification

Muon reconstruction in CMS and the response of various sub-detectors to muons have been studied in great detail using muons from cosmic rays[78]. After a year of 7 TeV data taking we can report for the first times on the performance of the CMS muon reconstruction[79] and trigger in proton-proton collision data. A muon event display of 7 TeV collision is shown in Fig. 5.1. The CMS reconstruction of tracks uses the Kalman Filter technique[61]. For the muons particle, tracks are first reconstructed independently in the silicon tracker (tracker track) and in the muon spectrometer (standalone-muon track). Based on these, two reconstruction approaches are used:

1. Global muon reconstruction (outside-in): starting from a standalone muon in the muon system, a matching tracker track is found and a global-muon track is fitted combining hits from the tracker track and standalone-muon track. At large transverse momenta ($p_T \gtrsim 200$ GeV/c), the global-muon fit can improve the momentum resolution compared to the tracker-only fit.
2. Tracker Muon reconstruction (inside-out): in this approach, all tracker tracks

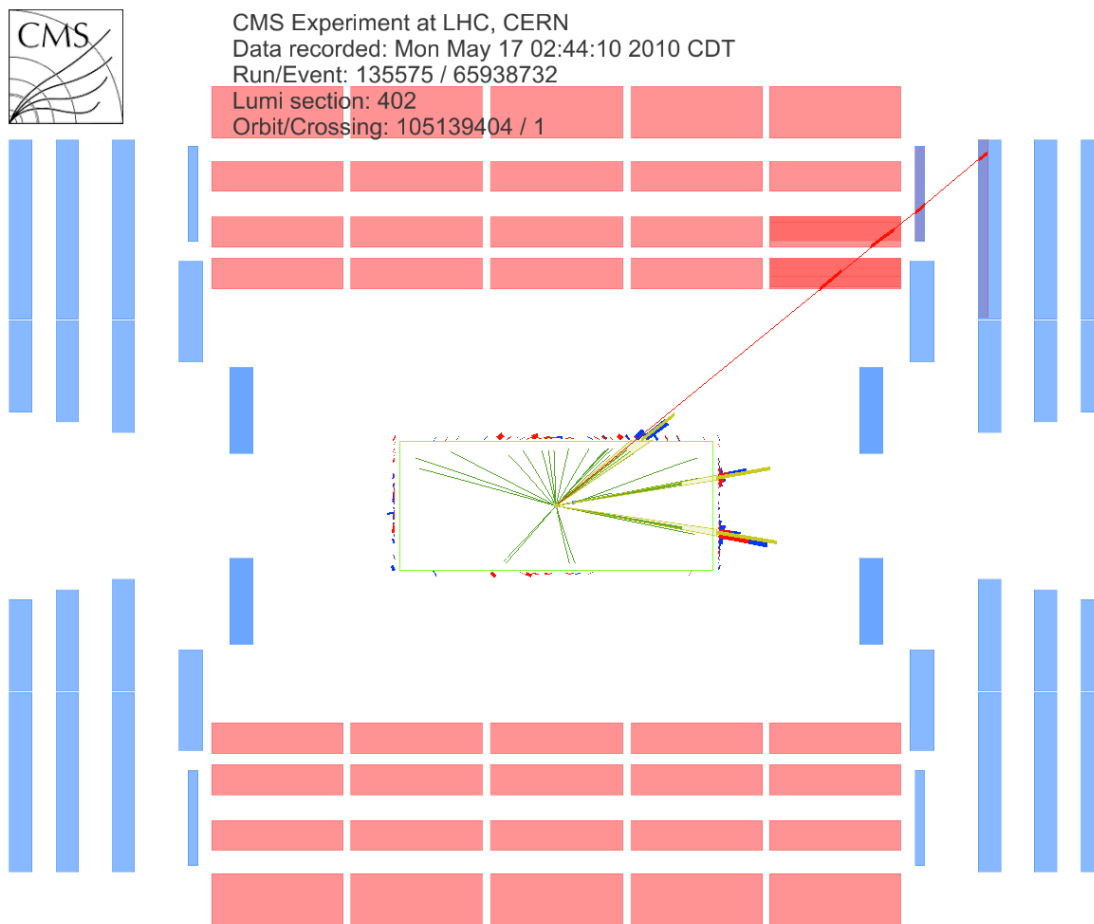


Figure 5.1: A display of a collision event with a muon crossing the DT and CSC chambers in the overlap region between barrel and endcap.

with $p_T > 0.5$ GeV/c and $p > 2.5$ GeV/c are considered as possible muon candidates and are extrapolated to the muon system, taking into account the expected energy loss and the uncertainty due to multiple scattering. If at least one muon segment (i.e. a short track stub made of DT or CSC hits) matches the extrapolated track in position, the corresponding tracker track qualifies as a tracker-muon track. At low momentum (roughly $p < 5$ GeV/c) this approach is more efficient than the global muon reconstruction, since it requires only a single muon segment in the muon system, while global muon reconstruction typically becomes efficient with two or more segments.

Most of the muons from collisions (with sufficiently high momentum) are reconstructed either as a Global Muon or a Tracker Muon, or very often as both. However, if both approaches fail and only a standalone-muon track is found, this leads to a third category of muon candidates:

3. Standalone-muon track only: this occurs only for about 1% of muons from collisions, thanks to the high tracker-track efficiency in CMS. In addition the acceptance of this type of muon track for cosmic-ray muons is a factor 10^2 to 10^3 larger.

Particles detected as muons are produced in pp collision from different sources which lead to different experimental signatures:

- Muons from heavy flavour: here most of muon chamber hits associated to the muon candidate were produced by a true muon. The muon's parent particle can be a beauty or charmed meson, a tau lepton or any other particle (like W or Z). These muons are usually prompt and with relatively high momentum.
- Muons from light flavour: In this category most of muon chamber hits associated to the muon candidate were produced by a true muon. This muon originated from light hadron decays (π and K) or, less frequently, from a calorimeter shower or a product of a nuclear interaction in the detector.
- Hadron punch-through: Here most of muon chamber hits of the muon candidate were produced by a particle other than a muon. The so-called "punch-through" is the source of the most of these candidates ($\sim 88\%$ for Global Muons) although "sail-through" (i.e. particles that does not undergo before the muon system) is present as well.
- Duplicate: If one simulated particle gives rise to more than one reconstructed muon candidate, the one with the largest number of matched hits is flagged according to one of the other categories. Any others are labeled as "duplicate". These are duplicate candidates created by instrumental effects or slight imperfections in the pattern recognition algorithm of the reconstruction software.

The standard CMS reconstruction provides additional information for each muon, useful for muon quality selection and identification (ID) in physics analyses. Track quality information includes the number of hits associated to the tracker track, standalone-muon track and global-muon track, the χ^2 of the various fits, the number of “good” hits (those having small residual with respect to the track), and the number of layers where a hit would be expected but no hit was found, the number of chambers in different stations with “matching” segments.

Plots of some of these variable on muons from selected $Z \rightarrow \mu^+\mu^-$ events are shown for data and simulation in Figure 5.2. A dataset of 2.9 pb^{-1} of 7 TeV collision data is used, the same which will be the input for the $Z \rightarrow \mu^+\mu^-$ cross section measurement reported in the next Chapter.

Other muon ID information includes matching information for segments in the different stations of the muon system matched to the tracker track extrapolation, with residuals in position and direction and their corresponding uncertainties. Compatibility variables (having a value between 0 and 1) are provided based on muon segment information, giving the most weight to muon segments in the outer layers of the muon system (the “muon segment compatibility”) and based on the matched energy deposits in the calorimeters (the “calorimeter compatibility”)

Summarizing, there are specific handles helping to select high quality muons. For the purpose of the $Z \rightarrow \mu^+\mu^-$ and $W \rightarrow \mu\nu$ analysis we aim to select prompt high p_T ($> 20 \text{ GeV}/c$) muons. The current data sample do not contain so far ultra-high momentum muons, thus we so not address here specific issues with very high energy muons. In addition the triggering muon has to be in the well examined range $|\eta| < 2.1$ (2.4) of the HLT trigger which, effectively, also suppresses the muon background and serves as a muon ID preselection. A list of cuts to improve the muon ID selection contains the following requirements:

- A muon has to be identified both as a tracker (TRK) and a global muon (GLB). This is effective against decays-in-flight, punch-through and accidental matching (with noisy or background tracks or segments).
- The number of hits in the tracker track part of the muon has to be larger than 10. Generally tracks with small number of hits give bad p_T estimate. In addition decays in flight give rise in many cases to lower hit occupancy in the tracks.
- There should be at least one pixel hit in the tracker track part of the muon. The innermost part of the tracker is an important handle to discard non-prompt muons. By requiring just a minimal number of hits we introduce negligible reconstruction inefficiency Such inefficiency can be measured with Z events directly.

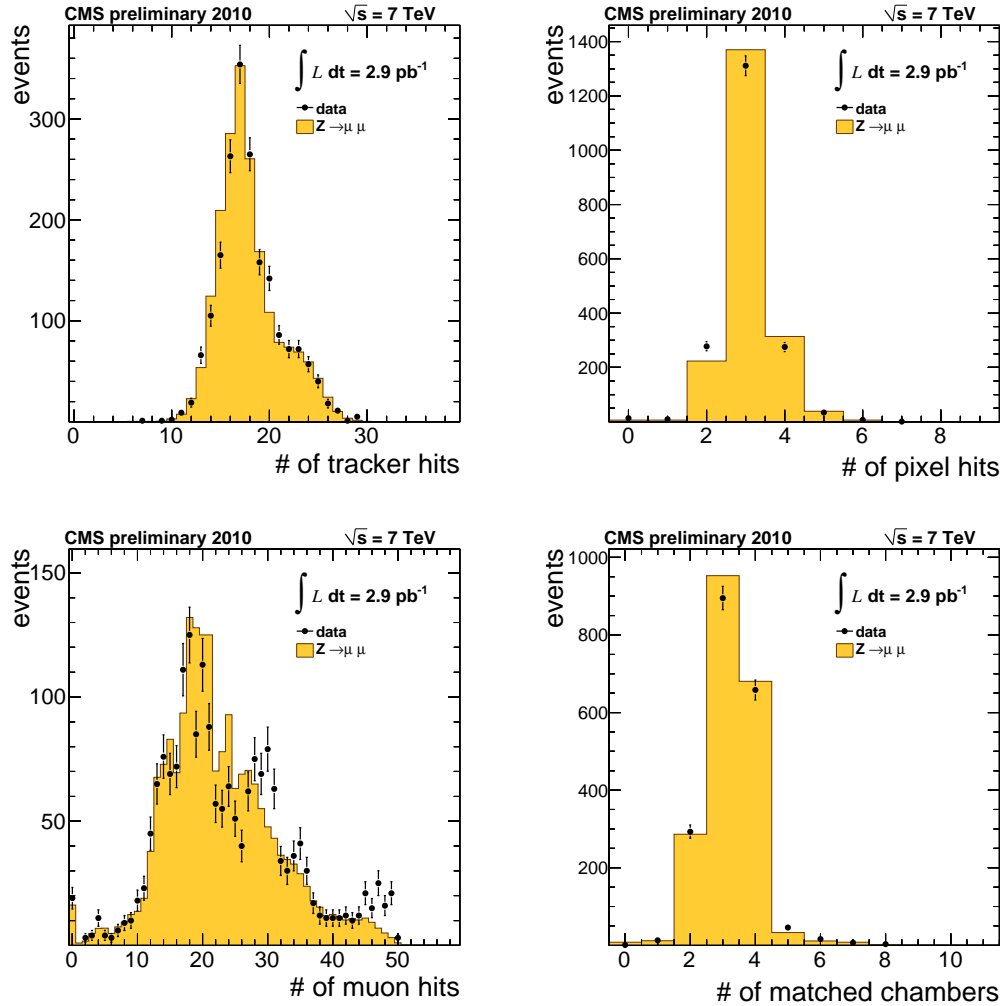


Figure 5.2: Muon ID selection variables on selected $Z \rightarrow \mu^+ \mu^-$ events in data and simulation. Top left: number of tracker hits (pixel plus strip) associated to the muon track; top right: number of pixel hits associated to the muon track; bottom left: number of muon hits in the muon chambers associated to the track; bottom right: number of muon chambers with matches.

- In addition, the muon track has to have at least two chambers in different stations with “matching” (consistent with the propagated to the muon chambers tracker track) segments. This is also to comply with a similar looser requirement in the trigger.
- Very bad fits are rejected by requiring reasonable global muon fit quality: $\chi^2/\text{NDF} < 10$ (NDF is the number of degrees of freedom of the fit). If there is a decay in flight inside the tracking volume, the trajectory could contain a sizeable “kink”, resulting in a poorer χ^2 of the fit used to determine the trajectory.
- The global muon has to contain at least one “valid” muon hit. By this requirement we make sure that the global muon is not a “bad” match between the information from the muon system and the tracker. This could happen in particular for non-prompt muons.
- The impact parameter d_{xy} , defined as the distance of closest approach of the muon track with respect to the beam spot has to be compatible with the interaction point hypothesis (muon from the interaction point). A loose, yet a powerful against cosmic background, cut is set at $d_{xy} < 2$ mm.

Another useful physical variable to distinguish signal leptons in an hadronic environment is the isolation variables, i.e. the amount of particle activity in the neighbourhood of the lepton: signal leptons, as the ones resulting from the decay of W and Z , tend to be very isolated, while most of the background leptons, are contained inside a jet or high particle density remnants of the partonic interactions. Muon reconstruction assigns to every muon three isolation variables I_{trk} , I_{Ecal} , I_{Hcal} , where I_{trk} , I_{Ecal} , and I_{Hcal} are respectively the sum of track p_T , of the transverse energy E_T in the electromagnetic calorimeter, and of the transverse energy in the hadron calorimeter, all measured in a cone of a given size, $\Delta R = 0.3$, with $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, around the muon track direction. The track of the muon itself is excluded from the computation of I_{trk} , and the muon calorimeter deposits are excluded from the computation of I_{Ecal} and I_{Hcal} by removing calorimeter towers in a “veto cone” around the extrapolated path of the muon.

Frequently, muons originating from background processes tend to have lower p_T than those from the signal. In many physics analyses, therefore, the separation between signal and background muons is improved by normalizing the isolation energy to the p_T of the muon, giving the relative isolation variable $I_{rel} = \frac{I}{p_T}$. The most powerful to discriminate background muon from signal muons is usually the relative combine isolation variable, I_{rel}^{comb} , defined as $I_{rel}^{comb} = \frac{I_{trk} + I_{Ecal} + I_{Hcal}}{p_T}$. The isolation properties depend on the event selection and on the type of muons considered, and will thus be different from analysis to analysis.

Plots of the three isolation variables I_{trk} , I_{Ecal} , I_{Hcal} and I_{rel}^{comb} isolation on muons from selected $Z \rightarrow \mu^+\mu^-$ events are shown for data and simulation in Figure 5.3 on a dataset of 2.9 pb^{-1} of 7 TeV collision data.

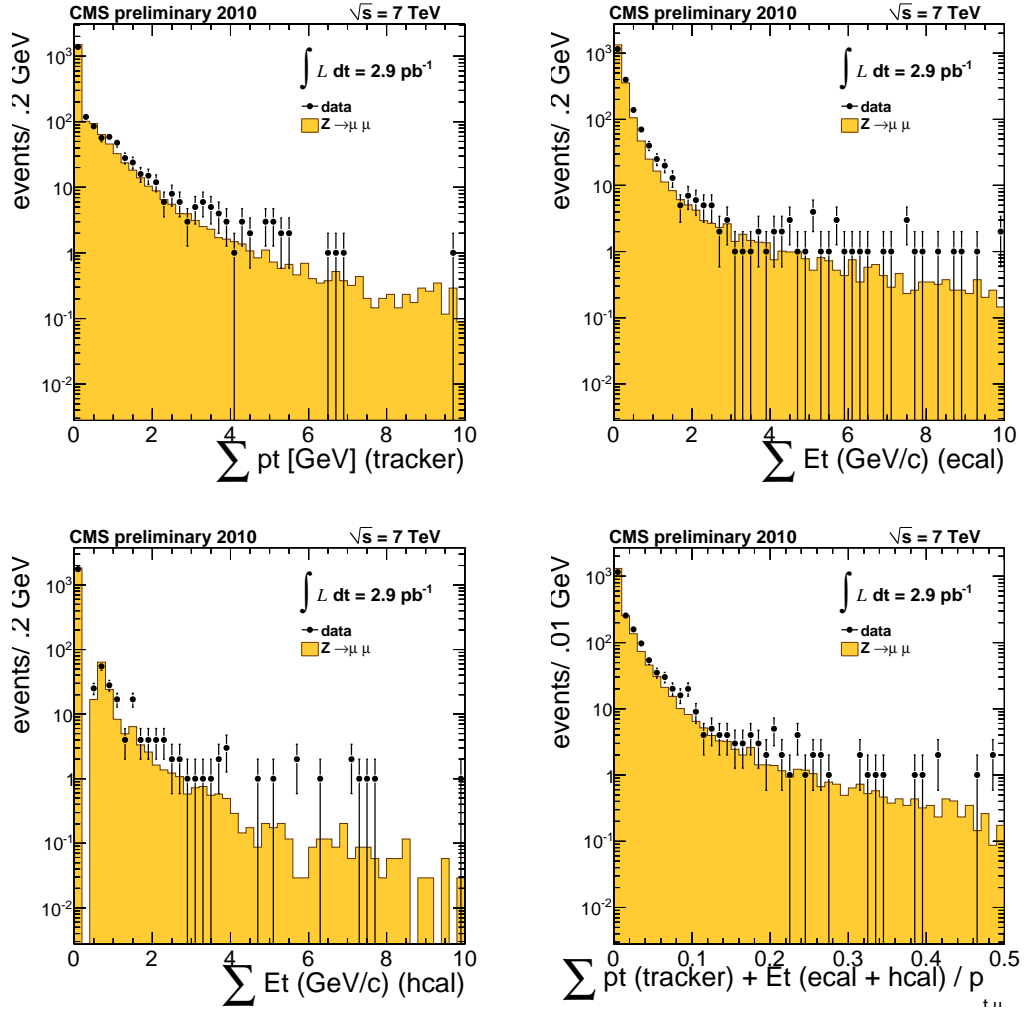


Figure 5.3: Muon isolation variables on selected $Z \rightarrow \mu^+\mu^-$ events in data and simulation. Top left: tracker isolation; top right: ecal isolation; bottom left: hcal isolation; bottom right: relative combined isolation.

5.2 Z Bosons Production at LHC

The mass and the width of the Z boson are measured by the LEP experiments to highest precision[80]. The world average values are 91.1875 ± 0.0021 GeV and 2.4952 ± 0.0023 GeV respectively. The branching ratios of Z decay channels are given in Table 5.1.

| Decay mode | Probability (%) |
|-----------------|-------------------|
| $e^+ e^-$ | 3.363 ± 0.004 |
| $\mu^+ \mu^-$ | 3.366 ± 0.007 |
| $\tau^+ \tau^-$ | 3.370 ± 0.008 |
| Invisible | 20.00 ± 0.06 |
| Hadrons | 69.91 ± 0.06 |

Table 5.1: Branching fractions of the Z boson in the main allowed channels.

Z bosons production at LHC proceeds mainly via the Drell-Yan process. The dominating processes at the LHC are $q\bar{q} \rightarrow Z/\gamma^* \rightarrow \mu^+ \mu^-$ (65%) and $qg \rightarrow q\gamma^*/Z \rightarrow q\mu^+ \mu^-$ (35%) which are illustrated in Figure 5.4. The dominant higher order correction of the first process is the scattering of a quark with a gluon, which contributes roughly 1/3 of the overall cross section of this process.

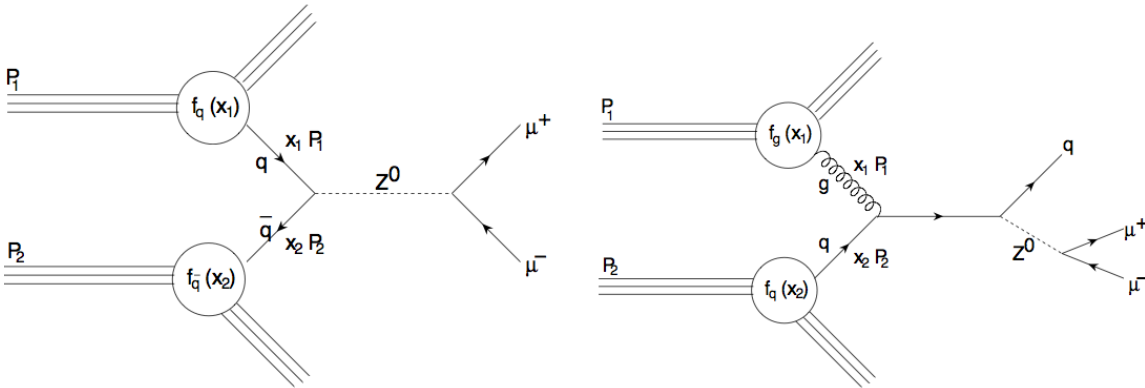


Figure 5.4: (Left) Leading order diagram of the Drell-Yan process. (Right) One of the next to leading order diagram of the Drell-Yan process.

The fraction of the proton's four-momentum carried by the initial partons, indicated as x -value, can be estimated with a simple calculation. The total energy in the center of mass frame of two colliding protons with four momenta P_1 and P_2 is determined by

$$S = (P_1 + P_2)^2 \approx 2P_1P_2,$$

when neglecting the rest mass of the protons. The square of the total energy in the center of mass frame of the two partons with the momenta p_1 and p_2 is then given by

$$s = (p_1 + p_2)^2 = (x_1P_1 + x_2P_2)^2 \approx 2x_1x_2P_1P_2 = 2x_1x_2S,$$

where x_i denotes the fraction of the proton's four-momentum relative to the i^{th} partons. This energy \sqrt{s} must be greater than m_Z and therefore:

$$x_1x_2 > \frac{m_Z^2}{S}.$$

For a typical value $x \approx x_1 \approx x_2$ (i.e. assuming for a rough estimate that the two interacting parton carry the same energy fraction) it follows

$$x \approx \frac{m_Z}{\sqrt{S}} \approx 0.01,$$

for the collision energy at the Large Hadron Collider of 7 TeV. Comparing this small value with the structure functions of the proton shown in Figure 1.15 it becomes obvious that gluons are the dominating partons, raising the question why the process $qg \rightarrow qZ/\gamma^*$ is not dominating. This can be explained by two facts: at these high energies, the assumption $x_1 \approx x_2$ may not be valid; moreover, the annihilation of $q\bar{q}$ leads to a resonant production of the Z boson which is enhanced by many orders of magnitude. A next-to-next-to-leading order calculation of the total cross section of the process $pp \rightarrow Z + X \rightarrow \mu^+\mu^- + X$ at a center of mass of 7, 10 and 14 TeV using MSTW08NNLO[81] PDF are listed in the Table 5.2.

| | 7 TeV | 10 TeV | 14 TeV |
|--------------------------------|----------------|-----------------|-----------------|
| $\sigma(Z \rightarrow l^+l^-)$ | 0.963 nb | 1.456 nb | 2.037 nb |
| $60 < M_{ll} < 120$ | | | |
| PDF uncertainty | +3.59 % -3.26% | +3.38 % -3.21 % | +3.28 % -3.11 % |

Table 5.2: Next-to-next-to leading order prediction for $Z \rightarrow l^+l^-$ cross section for different energies of LHC.

The systematic theoretical uncertainty of the cross section arises from the uncertainties of the PDF function used for the computation.

5.3 Reconstruction of Z boson decays in CMS

For each collision event the $Z \rightarrow \mu^+\mu^-$ candidates are built from all combinations of muons pairs (di-muons) with relatively high momentum ($p_T > 15$ GeV/c) detected in the angular acceptance of CMS $|\eta| < 2.4$. The tool used is the PAT toolkit, already mentioned in section 3.6. The opposite charge is not required at the first step, hence the muon charge mis-assignment can be studied. Muon candidates are classified according to the type of tracks they are associated to. They can be reconstructed in the inner tracker (tracker track muon, i.e. a tracker track with muon mass hypothesis), in the muon system (standalone muon), in the full CMS detector (global muon) as explained in the first part of this Chapter. Figure 5.5 shows a schematic image of the three kinds of reconstructed muons.

In addition also tracker muon candidates (which is a more stringent requirement than a tracker track, because some hits associated in the muon chambers are needed to form a tracker muon track) are considered to make a comparison with the tracker track and study the efficiency of the standalone-tracker muon matching to form a global muon.

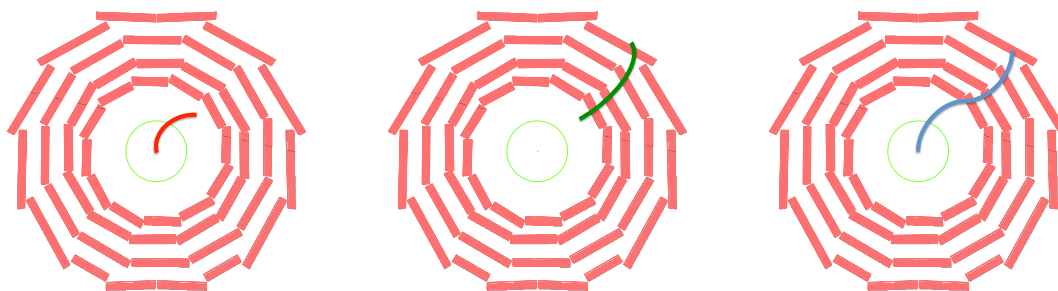


Figure 5.5: Muon candidates classification, from left to right: tracker track, standalone track and global muon.

Therefore, three categories of $Z \rightarrow \mu^+\mu^-$ candidates are built, according to the type of input muons used in the reconstruction:

- two global muons, called ‘golden’ candidates;
- global muon plus a standalone muon
- a global muon plus a tracker track muon.

The property of these various categories and other similar ones used as control check sample in the analysis will be deeply discussed in the next chapter.

In the following Figure 5.6 some of the kinematic variable for the selected $Z \rightarrow \mu^+\mu^-$ events are shown for data and simulation. Golden candidates are prented (with

the selection described in the next Chapter). A dataset of 2.9 pb^{-1} of 7 TeV collision data is used. From the plots one can see the excellent agreement between data and simulation. In addition we reported the number of selected $Z \rightarrow \mu^+ \mu^-$ events, showing the performance of the di-muon candidate reconstruction algorithm. When two candidates are selected one of the two is fake (being less than 10% the probability that a ZZ decaying in muons has been produced in the collision with the analyzed statistics).

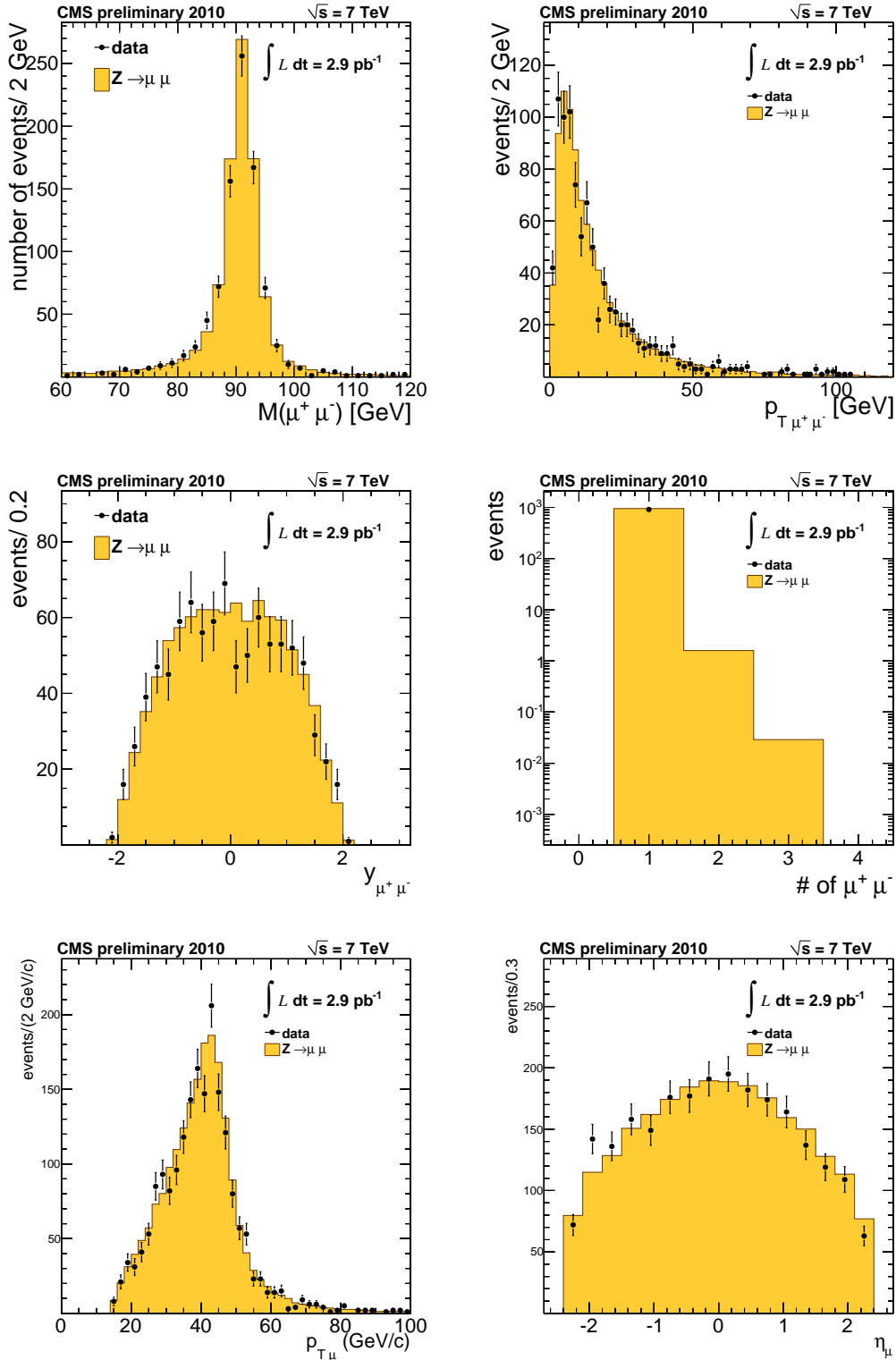


Figure 5.6: Golden $Z \rightarrow \mu^+ \mu^-$ candidates variables distribution on 3 pb^{-1} 7 TeV collision data and simulation. Top left: di-muon mass; top right: di-muon p_T ; center left: di-muon y ; center right: number of di-muon candidates per event; bottom left: p_T of muons from selected $Z \rightarrow \mu^+ \mu^-$ candidates; bottom right: η of muons from selected $Z \rightarrow \mu^+ \mu^-$ candidates.

Chapter 6

Z analysis

The inclusive Z production: $pp \rightarrow Z + X \rightarrow l^+l^-$ ($l = e, \mu$) has relatively large cross section and clear signature at LHC. Therefore is one of the first processes studied from the very early phase of data-taking. It is worth observing that the Z measurements at LHC will be less precise than those at leptonic machines such as LEP II. Nevertheless, the Z channel is considered a “Standard Candle” for many LHC measurements. Indeed they are essential for the calibration of the detector and to assess its performance. Events with Z (+ jets) also constitute the main source of background for many searches beyond the Standard Model. Therefore a precise understanding of these processes is critical before any discovery claim. Their inclusive study represents a first step in the detailed understanding of reference physics processes at the LHC: Z boson transverse momentum spectra, Z production in association with jet activity, beyond-leading-order effects on the total rate and parton density functions. In addition Z and W production rate, once we assure the consistency with the theoretical prediction (dominated by the parton density function systematic uncertainty), can be used to have an offline luminosity estimate. In this thesis a study of the inclusive cross section of the process $pp \rightarrow Z + X \rightarrow \mu^+\mu^-X$ with the first 2.9 pb^{-1} of LHC 7 TeV collision in the CMS experiment is presented extensively, with a first look at higher statistics of 35 pb^{-1} , which correspond to the entire 2010 proton-proton collision dataset. The method to extract the cross section is based on a simultaneous fit of the yield of $Z \rightarrow \mu^+\mu^-$ events, the average reconstruction muon efficiencies in the tracker and in the muon detector, the trigger efficiency, as well as the efficiency of the cut applied to select isolated muons without any estimate of those efficiencies from MC. The extracted Z yield has to be just corrected for the geometrical acceptance and for the integrated luminosity in order to measure the inclusive cross section.

With the very first few hundred of nb^{-1} of collision data not entirely all aspects of this data driven approach were possible, and so a simple “cut and count” approach has been preferred. Section 6.1 contains all the details of the first measurement.

In addition an online “ W/Z counting” has been developed for the machine luminosity check purpose. The physics performance of such tool will be describe in detail in Section 6.5

6.1 $Z \rightarrow \mu^+ \mu^-$ cross section determination

Any physics channel cross section can be determined from an event counting as

$$\sigma = \frac{N_{sig} - N_{bkg}}{\epsilon \times A \times \mathcal{L}}, \quad (6.1)$$

where N_{sig} and N_{bkg} are the number of signal and background events passing the selection, ϵ is the efficiency used in the selection, A is the geometrical acceptance, i.e. the fraction of generated events with the selection kinematic cuts (i.e. p_T , $|\eta|$), and \mathcal{L} is the machine integrated luminosity.

The evolution of the $Z \rightarrow \mu^+ \mu^-$ cross section extraction strategy from the very first day of data taking up to higher integrated luminosity is graphically represented in Figure 6.1.

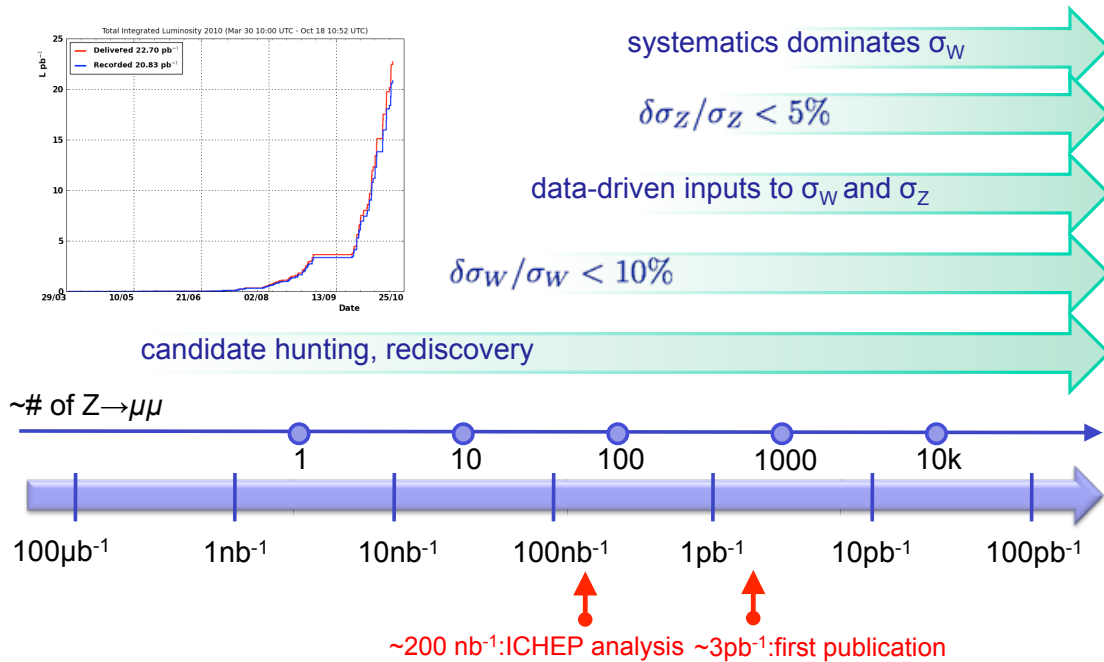


Figure 6.1: Scenario of the $Z \rightarrow \mu^+ \mu^-$ cross section extraction versus the integrated luminosity.

When the first era of candidates “hunting” was over, two different methods came one after the other to measure the $Z \rightarrow \mu^+\mu^-$ cross section with the first 7 TeV collision data. I personally gave a great contribution to both measurements:

1. The first method is MC driven and has been carried out in the first months of 7 TeV data taking, to achieve a results with the very first few hundred nb^{-1} of data. This method relies on MC expectation to estimate the efficiency values. MC predictions are corrected for remaining differences with respect to data via efficiency correction factors.
2. The second method is fully data driven and aims to extract the signal yield and the efficiency terms directly for data using a simultaneous fit to different di-muon categories. This method has been studied in details on simulated data before the collision era and showed to be fully valid with an amount of data corresponding to few inverse pb of integrated luminosity.

In the following we first describe briefly the first period of 7 TeV collision with the first candidates “hunting” activity, then the first method used to perform the first $Z \rightarrow \mu^+\mu^-$ cross section measurement with at a center of mass energy of 7 TeV showed at ICHEP2010 conference[82] and then we fully concentrate on the second method witch is the core of my thesis. Both measurement represent my original contribution to the experiment.

6.2 W/Z rediscovery phase

The first two months immediately after the start of the 7 TeV operation, April and May 2010, were characterized by the “rediscovery” of the W and Z bosons, as also shown in Figure 6.1. This period coincides approximately with the production of the first 10 nb^{-1} of integrated luminosity, when one expects few tens of $W \rightarrow \mu\nu$ and few $Z \rightarrow \mu^+\mu^-$ candidates. In this period CMS operated only with the minimum bias trigger. The events containing the first W and Z bosons were searched eagerly between the great amount of data resulting from this unique stream. The rediscovery of the electro-weak bosons was considered a first step to assess the performance and correct operation of the CMS experiment. I contributed very actively in this hunting period, which resulted in the observation of the $W \rightarrow \mu\nu$ and $Z \rightarrow \mu^+\mu^-$ first events displayed in Figure 6.2.

6.3 MC driven first era analysis

The first $Z \rightarrow \mu^+\mu^-$ measurement at 7 TeV center of mass energy with the CMS experiment has been done using the first 198 nb^{-1} of collision events. The muon

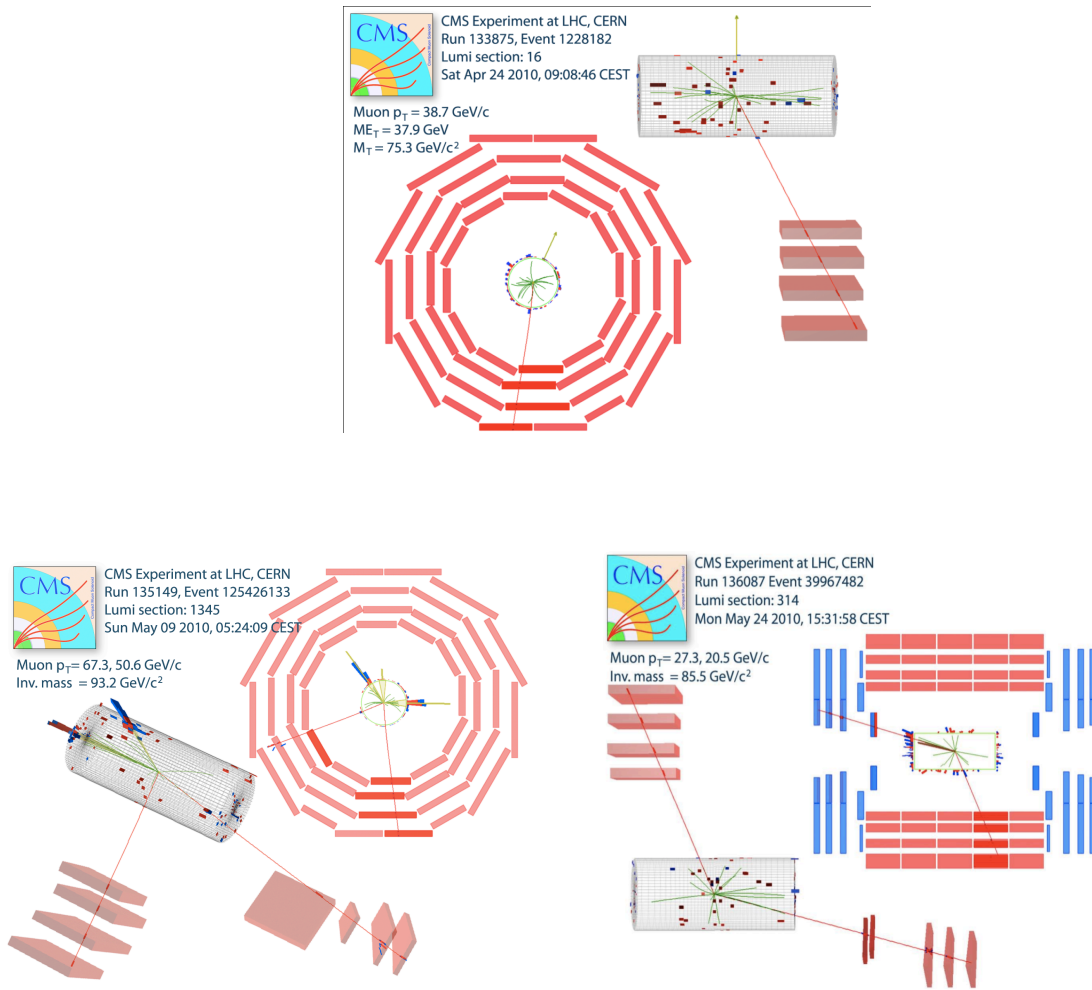


Figure 6.2: Event display with the first $W \rightarrow \mu\nu$ and $Z \rightarrow \mu^+\mu^-$ candidates detected at CMS. Top: $W \rightarrow \mu\nu$ candidate; bottom left: first $Z \rightarrow \mu^+\mu^-$ candidate; bottom right: second $Z \rightarrow \mu^+\mu^-$ candidate.

trigger stream was active after the first 10 nb^{-1} of integrated luminosity, hence the muon primary dataset was used for the analysis.

6.3.1 Monte Carlo data samples

Several high statistic Monte Carlo simulated samples are used to evaluate signal and background efficiencies and to validate the analysis techniques deployed. Samples of simulated signal $Z \rightarrow \mu^+\mu^-$ and background $W \rightarrow \mu\nu$ and $Z \rightarrow \tau^+\tau^-$ are produced with POWHEG interfaced with PYTHIA parton-shower generator (see Section 1.6 for a brief description and references on the event generator used at LHC). QCD events with a muon in the final state and $t\bar{t}$ events are studied with PYTHIA. Generated events are processed through the full GEANT4[83] detector simulation, trigger emulation and event reconstruction chain of the CMS experiment. The analysed samples are reported in Table 6.1. In the QCD inclusive muons sample decays-in-flight are included, but no punch through. Signal samples simulated with PYTHIA are used as cross-check with respect to POWHEG samples.

| Generator | Process | Kinematic cuts (in GeV, $c = 1$) | σ (pb) | Events | PDF set |
|-----------|------------------------------|--|---------------|--------|---------|
| POWHEG | $Z \rightarrow \mu^+\mu^-$ | $m_{\mu\mu} > 20$ | 1686 | > 1M | CTEQ66 |
| POWHEG | $W \rightarrow \mu^+\nu$ | no cuts | 6152 | ~700k | CTEQ66 |
| POWHEG | $W \rightarrow \mu^-\nu$ | no cuts | 4179 | ~700k | CTEQ66 |
| POWHEG | $Z \rightarrow \tau^+\tau^-$ | $m_{\tau+\tau^-} > 20$ | 1686 | > 1M | CTEQ66 |
| PYTHIA | $t\bar{t}$ | no cuts | 94.3 | 500k | CTEQ6L |
| PYTHIA | Inclusive μ QCD | $\hat{p}_T > 20, p_T^\mu > 15, \eta_\mu < 2.5$ | 109853 | 6M | CTEQ6L |
| PYTHIA | $Z \rightarrow \mu^+\mu^-$ | $m_{\mu\mu} > 20$ | 1300 | 2M | CTEQ6L |

Table 6.1: Summary of analyzed Monte Carlo samples for the various signal and background processes.

6.3.2 Event selection

$Z \rightarrow \mu^+\mu^-$ are characterized by the presence of two high- p_T isolated muons. The expected background to this process is very low.

The analysis described in this section, due to the low statistics, is based on a cut-and-count strategy, where the $Z \rightarrow \mu^+\mu^-$ candidates are selected using a robust and high purity selection.

With $\int \mathcal{L} dt = 198 \text{ nb}^{-1}$ from MC studies we expect of the order of few tenths of candidate events containing two global muons (i.e. the category with the highest purity defined later in this Chapter (Section 6.4.1), while the other lower purity categories (with only one global muon and the other muon failing at least one selection

cut and defined as well in Section 6.4.1) lack of sufficient statistics to perform precise quantitative estimates.

The amount of background under the signal peak and the relevant efficiencies are estimated from MC, with reasonable systematic errors, and cross-checked as much as possible with the available data.

We select events which satisfy the single non-isolated muon trigger HLT_Mu9. For each event, we consider all the possible di-muon pairs made by opposite-charge muons and with invariant mass $60 < m_{\mu\mu} < 120$ GeV/c². The muons in the pair must satisfy, in addition, the following acceptance cuts:

- both muons must have $p_T > 20$ GeV/c;
- at least one of the muons must be within $|\eta| < 2.1$ for triggering purposes the remaining muon must be within $|\eta| < 2.4$ (fiducial region of the muon system).

Both muons must be identified as *Global Muons* and must have more than 10 total (pixel+strips) hits in the Tracker detector (referred in the following as “loose” muon selection). In addition to the above loose quality cuts, at least one of the muons must pass all the quality cuts described in Section 5.1:

- the muon must be also identified as *Tracker Muon*;
- > 0 pixel hits in the Tracker detector;
- ≥ 2 chambers in different stations of the muon system with matching segments;
- χ^2/NDF of the global fit < 10 ;
- ≥ 1 valid muon hit;
- for cosmics rejection we require the impact parameter d_{xy} w.r.t. beam spot < 2 mm.

The muon passing all the quality cuts must also match to a Level-3 (L3) trigger object firing the HLT_Mu9 trigger path (trigger and muon quality selections are referred in the following as “tight” muon selection). For the trigger matching we require the L3 muon to be within a cone of aperture $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} = 0.2$ around the offline muon, and their relative p_T difference $\Delta p_T/p_T$ to be < 1 , which is a very loose requirement given the L3 muon resolution in this p_T range. The choice of requiring trigger match and muon quality selection is motivated by the correlation of trigger and muon identification selection, which, applied to the same muon, make easier the evaluation of related systematic uncertainties.

Both muon candidates must be isolated, according to the default tracker isolation variable provided by the CMS muon reconstruction, I_{trk} , defined as the sum of the transverse momenta of all tracks within a cone of radius $\Delta R = 0.3$:

$$I_{trk} = \sum_{\Delta R < 0.3} p_T . \quad (6.2)$$

Only tracks with $\Delta z < 0.2$ w.r.t. the muon track are considered for the sum. Tracks within a cone of $\Delta R = 0.01$ are vetoed in order to avoid counting the muon track. We require for each muon $I_{trk} < 3$ GeV/c.

6.3.3 Results on data and comparison with MC

We have analyzed a data sample of 198 nb^{-1} and 77 events pass the $Z \rightarrow \mu^+ \mu^-$ event selection. The di-muon invariant mass, p_T , and rapidity distributions of the selected Z candidates are shown in Figures 6.3, 6.4 respectively. They are compared to the expected MC distribution for the nominal luminosity. NLO generator cross section (POWHEG with CTEQ66 PDF) are considered for the MC samples generated with POWHEG. We have applied a global scaling factor to the expectation from the PYTHIA $t\bar{t}$ sample given by the ratio of NLO MCFM[87] to LO PYTHIA cross sections ($162 \text{ pb}/94.3 \text{ pb} = 1.718$).

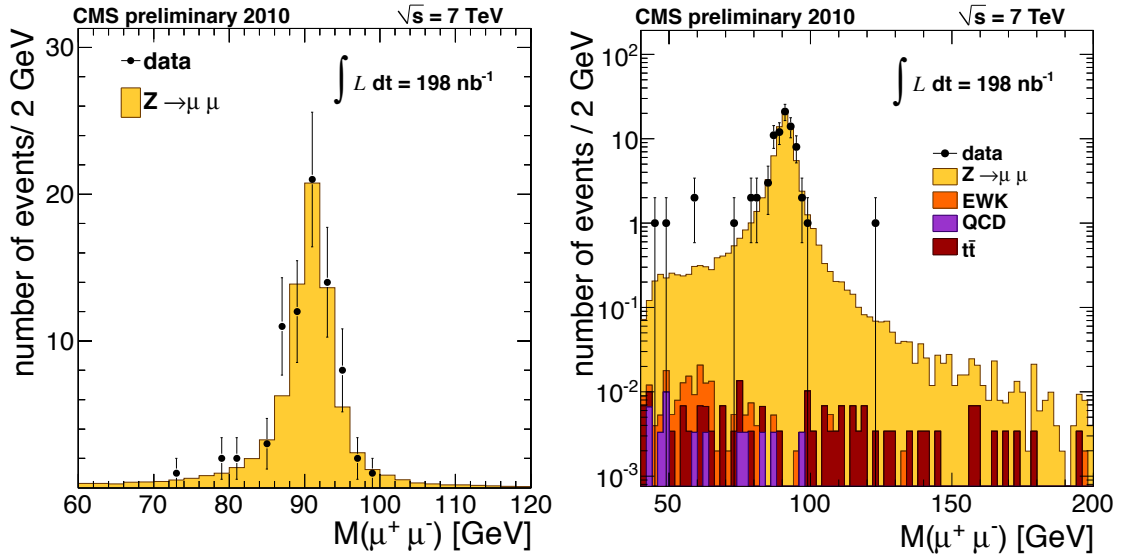


Figure 6.3: Invariant mass distribution of the selected $Z \rightarrow \mu^+ \mu^-$ candidates in data superimposed to the MC expectation. (a): linear scale; (b): log scale.

From MC we expect 78.2 signal events in the $[60, 120]$ GeV/ c^2 mass range and very low background: 0.16 expected events, dominated by $t\bar{t}$ and $Z \rightarrow \tau^+ \tau^-$ events,

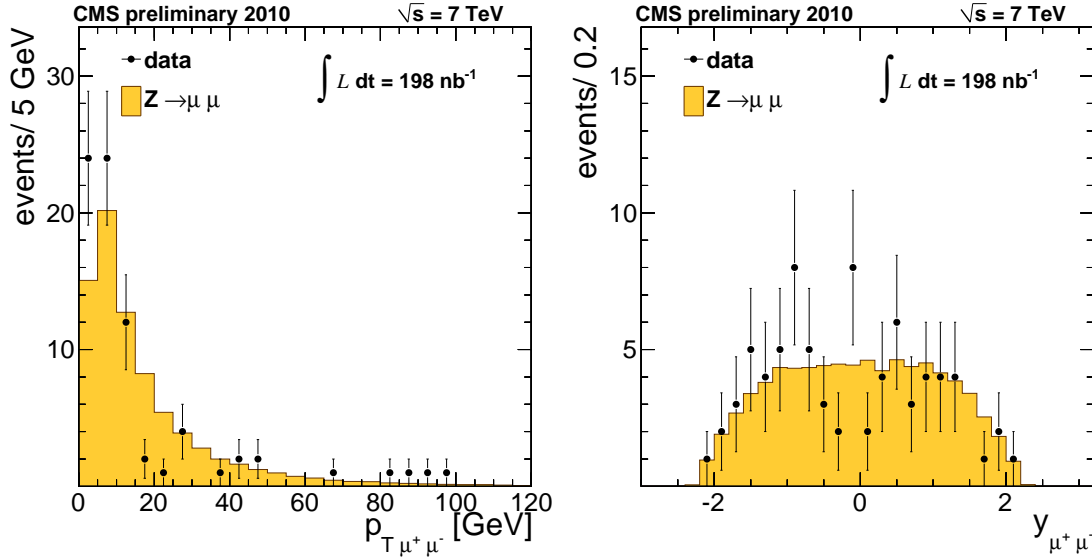


Figure 6.4: (left) Transverse momentum distribution of the selected $Z \rightarrow \mu^+ \mu^-$ candidates in data superimposed to the MC expectation. (right) Rapidity distribution of the selected $Z \rightarrow \mu^+ \mu^-$ candidates in data superimposed to the MC expectation.

0.10 and 0.10 expected events respectively. In Table 6.2 we report the performance of the selection evaluated on the signal and background MC samples and the expected yields for the nominal luminosity.

Due to the low statistics of possible control samples from data we may only estimate the expected background with the lowest possible uncertainty from Monte Carlo. Considering the small expected background size ($\approx 0.3\%$ of the signal), the systematic uncertainty due to MC prediction will not affect the cross section measurement.

6.3.4 Muon efficiencies

We use two muon selections: one “tight” and one “loose” in $Z \rightarrow \mu^+ \mu^-$ selection, and we require that both muon legs must satisfy the “loose” selection and that at least one of the muon legs satisfies the “tight” selection. We can write the efficiency to select a “loose” muon as:

$$\epsilon_{\text{rec}} = \epsilon_{\text{trk}} \times \epsilon_{\text{sa}} \times \epsilon_{\text{Nhits}} \quad (6.3)$$

where ϵ_{trk} is the reconstruction efficiency of a muon in the tracker, ϵ_{sa} is the reconstruction efficiency in the muon detector, and ϵ_{Nhits} is the probability for a reconstructed muon to pass the cut on the number of tracker hits. All those efficiencies are taken from MC. The efficiency to reconstruct a “tight” muon is:

$$\epsilon_{\text{tight}} = \epsilon_{\text{rec}} \times \epsilon'_{\text{ID}} \times \epsilon'_{\text{trig}} \quad (6.4)$$

Table 6.2: MC cross sections, cross sections times acceptance and efficiency of the full selection, and expected yields for the nominal luminosity. All the numbers have been evaluated on the reference MC samples (POWHEG). The reported errors are only the statistical uncertainties due to the MC statistics.

| | σ (pb) | $\sigma \times A \times \epsilon$ (pb) | Expected yields |
|-------------------------------|--------------------|--|-------------------|
| $Z \rightarrow \mu^+ \mu^-$ | 1686 | 395.0 ± 2.0 | 78.21 ± 0.40 |
| $W \rightarrow \mu \nu$ | 10331 | 0.03 ± 0.03 | 0.005 ± 0.005 |
| QCD | $296.9 \cdot 10^6$ | 0.10 ± 0.03 | 0.020 ± 0.006 |
| $t\bar{t}$ | 162 | 0.48 ± 0.07 | 0.095 ± 0.014 |
| $Z \rightarrow \tau^+ \tau^-$ | 1686 | 0.23 ± 0.05 | 0.046 ± 0.010 |

where ϵ'_{ID} is the probability that a muon passing the “loose” selection also passes all muon identification cuts¹, and ϵ'_{trig} is the probability that a muon passing the “loose” selection plus the muon identification cuts is also matched to a trigger muon. The efficiency ϵ'_{ID} for a Z candidate to pass the muon quality requirements is estimated from MC and the observed data/MC discrepancies are considered as source of systematics. We find in MC that the difference of ϵ'_{trig} w.r.t. the efficiency ϵ_{trig} , measured on all ‘loose’ muons (not necessarily passing the muon identification cuts) is about 0.6%, due to the (small) correlation of muon trigger performance with muon quality. The trigger efficiency ϵ'_{trig} is estimated from MC after applying the proper data/MC correction factors. The MC trigger efficiency is estimated by counting the numbers $N_{\mu\mu}^{2\text{HLT}}$ of $Z \rightarrow \mu^+ \mu^-$ candidates having both legs matched to trigger muons and $N_{\mu\mu}^{1\text{HLT}}$ of $Z \rightarrow \mu^+ \mu^-$ candidates having only one leg trigger-matched:

$$\frac{N_{\mu\mu}^{2\text{HLT}}}{N_{\mu\mu}^{2\text{HLT}} + N_{\mu\mu}^{1\text{HLT}}} = \frac{\epsilon_{\text{trig}}}{2 - \epsilon_{\text{trig}}} \quad (6.5)$$

We consider for this estimate the $Z \rightarrow \mu^+ \mu^-$ candidates having both legs which pass the tight muon selection. Thus, the estimate of the trigger efficiency takes correctly into account correlations with the muon selection cuts. We find $N_{\mu\mu}^{2\text{HLT}} = 56$ and $N_{\mu\mu}^{1\text{HLT}} = 10$, which gives, using Clopper-Pearson binomial confidence interval[85]:

$$\epsilon_{\text{trig}} = 0.916_{-0.034}^{+0.026}. \quad (6.6)$$

This number agrees with the MC estimate within the statistical uncertainty.

The isolation efficiency is determined from data by counting the number of di-muon candidates $N_{\mu\mu}^{\text{non iso}}$ which fail the isolation cut on one of the legs (but pass all the other cuts). We find 4 candidates in this category and set: $\epsilon_{\text{iso}} = 0.980_{-0.018}^{+0.010}$.

¹Note that the cuts on number of tracker hits is already applied in the loose selection.

The tracker efficiency is cross-checked with data by counting the di-muon candidates where one of the legs is a standalone muon but fails the global muon reconstruction. We find no candidate in this category and set $\epsilon_{\text{trk}} = 1.00_{-0.01}^{+0.00}$. The standalone muon efficiency is cross-checked in data by counting the number of di-muon candidates where one of the legs is a tracker track but fails the global muon reconstruction. We find 2 candidates in this category and set: $\epsilon_{\text{sa}} = 0.987_{-0.016}^{+0.008}$

In Table 6.3 we report the muon efficiencies estimated from MC and the results from cross-checks in data (largely limited by the statistics), where possible, using the Z categories that will be used in the full data driven analysis with more data, and described in the next section.

Table 6.3: Muon efficiencies for $Z \rightarrow \mu^+\mu^-$ analysis. In the first column we report the MC efficiencies estimated from the $Z \rightarrow \mu^+\mu^-$ reference MC sample. In the second column we report the results obtained by counting di-muon candidates in different categories.

| | MC efficiency | Data estimates (di-muon counting) |
|---------------------------|---------------------|-----------------------------------|
| ϵ_{trig} | 0.927 ± 0.001 | – |
| ϵ'_{trig} | 0.932 ± 0.001 | $0.916_{-0.034}^{+0.026}$ |
| ϵ_{trk} | 0.9992 ± 0.0001 | $1.00_{-0.01}^{+0.00}$ |
| ϵ_{sa} | 0.9894 ± 0.0005 | $0.987_{-0.016}^{+0.008}$ |
| ϵ'_{ID} | 0.9758 ± 0.0006 | – |
| ϵ_{Nhits} | 0.9979 ± 0.0001 | – |
| ϵ_{iso} | 0.9837 ± 0.0007 | $0.980_{-0.018}^{+0.010}$ |

6.3.5 $Z \rightarrow \mu^+\mu^-$ cross section determination with the first 198 nb^{-1} of 7 TeV collision data

The inclusive $Z \rightarrow \mu^+\mu^-$ cross section is determined from the $N_{\mu\mu}$ yield in a data sample corresponding to an integrated luminosity \mathcal{L}_{int} , by using:

$$\sigma(pp \rightarrow Z + X \rightarrow \mu^+\mu^- + X) \times \mathcal{A} = \frac{N_{\mu\mu}}{\epsilon_{\text{rec}}^2 \epsilon_{\text{iso}}^2 [1 - (1 - \epsilon'_{\text{ID}} \epsilon'_{\text{trig}})^2] \mathcal{L}_{\text{int}}}, \quad (6.7)$$

where \mathcal{A} is the kinematic cut acceptance, and the efficiency terms have been introduced above. The cross section is determined in a kinematic region defined by the p_T and η cuts on the muons and by the $m_{\mu\mu}$ invariant mass cut.

Using the corrected efficiencies in Table 6.3 and $\mathcal{L}_{\text{int}} = 198 \pm 22 \text{ nb}^{-1}$ we obtain, quoting only statistical uncertainty:

$$\sigma(pp \rightarrow Z + X \rightarrow \mu^+\mu^- + X) \times \mathcal{A} = 0.418 \pm 0.048(\text{stat.}) \text{ nb} \quad (6.8)$$

Using the acceptance given the used kinematic cuts ($60 < m_{\mu\mu} < 120\text{GeV}$, $p_{T,\mu} > 20$, $|\eta^1| < 2.1$, $|\eta^2| < 2.4$) estimated at the generator level from the POWHEG $Z \rightarrow \mu^+\mu^-$ reference MC sample:

$$\mathcal{A}(\text{POWHEG}) = 0.476 \pm 0.002 \quad (6.9)$$

we can extrapolate the cross section beyond the kinematics cuts, and we can determine:

$$\sigma(pp \rightarrow Z + X \rightarrow \mu^+\mu^- + X) [60 < m_{\mu\mu} < 120\text{GeV}/c^2] = 0.88 \pm 0.10(\text{stat.}) \text{ nb} . \quad (6.10)$$

Only the statistical error is quoted; systematic errors are described in Section 6.3.6.

6.3.6 Systematic uncertainties

The largest uncertainty for the cross section measurement comes from the luminosity measurement, currently estimated to be 11% using the HF measurements[84]. This uncertainty should decrease in the future. We quote it separately from the other systematic uncertainties. The first group of sources of systematics is related with uncertainties in the experimental response predicted by CMS simulations. Particularly important are estimates of muon reconstruction, identification, trigger and isolation efficiencies. These sources are quantified as the accuracy of the tests done to get the efficiencies ratio between MC and data[11].

Sub-dominant systematic uncertainties come from the muon energy/momentum scale and resolution. Studies of high-energy cosmic ray events, alignment discrepancies between inner tracker extrapolations and muon chamber positions, low-mass di-muon resonances extrapolation and estimated uncertainties on tracker alignment show that scale shifts above 1% for muons with $p_T \sim 40 \text{ GeV}/c$ can be excluded.

Theoretical uncertainties in the $Z \rightarrow \mu^+\mu^-$ cross section measurement enter in the determination of the acceptance of the detector and selection cuts. The Monte Carlo estimates are based on simulations that use a NLO generator (POWHEG) as input. Events are re-weighted according to different PDF set assumptions (CTEQ6.6, MSTW08NLO, NNPDF2.0). The observed variations in the acceptance are $\leq 2\%$. Remaining theoretical uncertainties due the treatment of initial-state radiation, final-state QED radiation, missing electroweak effects of renormalization/factorization scale assumptions amount to $\sim 1.5\%$. Given the very low statistics we can neglect any systematics associated to the very low and irreducible background subtraction.

Table 6.4 shows the the systematic uncertainties in the $Z \rightarrow \mu^+\mu^-$ cross section measurement. The main uncertainties are statistical in nature, and will decrease as more data is collected and analyzed, and the more we exploit data driven technique to determine the efficiencies.

Table 6.4: Summary of systematic uncertainties for the $Z \rightarrow \mu^+\mu^-$ channel using the first 198 nb⁻¹ of 7 TeV collision data.

| Source | % |
|---------------------------------------|------|
| Muon reconstruction/identification | 2.5 |
| Trigger efficiency | 0.7 |
| Isolation efficiency | 1.0 |
| Muon momentum scale/resolution | 0.5 |
| PDF uncertainty in acceptance | 2.0 |
| Other theoretical uncertainties | 1.6 |
| TOTAL(without luminosity uncertainty) | 3.8 |
| Luminosity | 11.0 |

Collecting all the pieces it is possible to measure for the first time a cross sections for $Z \rightarrow \mu^+\mu^-$ production at 7 TeV, using 198 nb⁻¹:

$$\sigma(pp \rightarrow Z(\gamma^*)+X \rightarrow \mu^+\mu^-+X) = 0.881+0.104(\text{stat.})+0.042(\text{syst.})\pm 0.097(\text{lumi.})nb$$

The measurement has been presented in July 2010 at the ICHEP conference in Paris, France and is contained in a CMS public document [11].

The reported $Z \rightarrow \mu^+\mu^-$ cross sections are limited to the di-lepton invariant mass range [60, 120] GeV/c², and corrected for the kinematic acceptance. The NNLO prediction for Z production is 0.97 ± 0.04 nb, in good agreement with our measurements. Figure 6.5 shows the EWK CMS measurements together with measurements at lower-energy hadron colliders as shown at ICHEP 2010 conference.

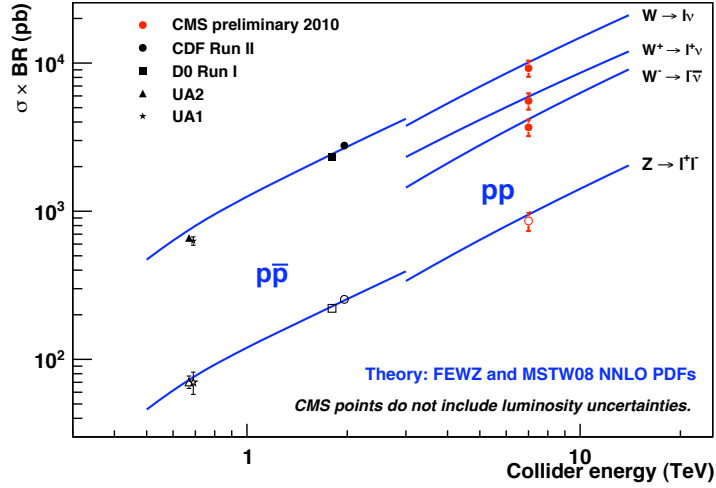


Figure 6.5: Measurements of inclusive cross sections from CMS and experiments at lower-energy colliders. The solid symbols represent $\sigma(W \rightarrow l\nu) \times BF(W \rightarrow l\nu)$ and the hollow symbols, $\sigma(Z(\gamma^*) \rightarrow l^+l^-) \times BF(Z \rightarrow l^+l^-)$.

6.4 Data driven analysis

Finally it presented a data driven study to determine of the inclusive cross section of the process $pp \rightarrow ZX \rightarrow \mu^+\mu^-X$. The method to extract the cross section is based on a simultaneous fit of the yield of $Z \rightarrow \mu^+\mu^-$ events, the average reconstruction muon efficiencies in the tracker and in the muon detector, the trigger efficiency, as well as the efficiency of the cut applied to select isolated muons. The extracted Z yield has to be just corrected for the geometrical acceptance and for the integrated luminosity in order to measure the cross section. The results are based on a collision data samples of 2.9 pb^{-1} .

6.4.1 Method description and trigger requirements

The number of produced $Z \rightarrow \mu^+\mu^-$ events in a collected data sample can be determined from the number of observed events with two reconstructed isolated global muons having an invariant mass within a range centered at the Z mass peak, corrected by the efficiency of reconstructing the two muons, the trigger selection efficiency, and the efficiency of the isolation cut.

We want to determine both the yield of produced Z events, corrected by the efficiency effects, and the involved efficiency terms from data.

We consider, as muon candidates, global muons, stand alone muons and tracks.

We build $Z \rightarrow \mu^+\mu^-$ candidates as pairs of muon candidates, and we define the following categories of events with at least one reconstructed $Z \rightarrow \mu^+\mu^-$ candidates:

- $Z_{\mu\mu}$: a pair of isolated global muons. This category can be further split into two independent samples:
 - $Z_{\mu\mu}^{2\text{HLT}}$: a pair of isolated global muons, both matched to an HLT trigger object;
 - $Z_{\mu\mu}^{1\text{HLT}}$: a pair of isolated global muons only one matched to a trigger object;
- $Z_{\mu s}$: a pair of one isolated global muon and one isolated stand-alone muon; the global muon must be matched to a trigger object;
- $Z_{\mu t}$: a pair of one isolated global muon and one isolated tracker track; the global muon must be matched to a trigger object;
- $Z_{\mu\mu}^{\text{non iso}}$: a pair of global muons, where at least one is non isolated; at least one muon must be matched to trigger primitives.

The five categories are explicitly forced to be mutually exclusive in our event selection: if one event falls in the first category it is excluded from the second; if it does not fall in the first category and falls in the second, it is excluded from the third, and so on, in order to have non-overlapping, hence statistically independent, event samples. In case of multiple di-muon candidates for an event falling in one of the categories, all the possible combinations are considered.

We introduce the differential event yields for signal plus background with the following Probability Density Functions (PDF) for each of the four categories:

$$\frac{dN_{\mu\mu}}{dm} = f_{\mu\mu}(m) = N_{\mu\mu} f_{peak}(m), \quad (6.11)$$

$$\frac{dN_{\mu\mu}^{2\text{HLT}}}{dm} = f_{\mu\mu}(m)^{2\text{HLT}} = N_{\mu\mu}^{2\text{HLT}} f_{peak}(m), \quad (6.12)$$

$$\frac{dN_{\mu\mu}^{1\text{HLT}}}{dm} = f_{\mu\mu}(m)^{1\text{HLT}} = N_{\mu\mu}^{1\text{HLT}} f_{peak}(m), \quad (6.13)$$

$$\frac{dN_{\mu s}}{dm} = f_{\mu s}(m) = N_{\mu s} f_{peak}^s(m) + b_{\mu s}(m), \quad (6.14)$$

$$\frac{dN_{\mu t}}{dm} = f_{\mu t}(m) = N_{\mu t} f_{peak}(m) + b_{\mu t}(m), \quad (6.15)$$

$$\frac{dN_{\mu\mu}^{\text{non iso}}}{dm} = f_{\mu\mu}^{\text{non iso}}(m) = N_{\mu\mu}^{\text{non iso}} f_{peak}(m) + b_{\mu\mu}^{\text{non iso}}(m). \quad (6.16)$$

In the above equations, the total signal yield in the different categories is factorized in the terms $N_{\mu\mu} = N_{\mu\mu}^{2\text{HLT}} + N_{\mu\mu}^{1\text{HLT}}$, $N_{\mu s}$, $N_{\mu t}$ and $N_{\mu\mu}^{\text{non iso}}$, so that the functions $f_{peak}(m)$ and $f_{peak}^s(m)$ are normalized to the unity. We have assumed, according to our Monte Carlo estimates, that the background in the samples with two isolated global muons is negligible: we expect $\approx 0.1\%$ of background from non- Z processes, and 0.030% from combinatorial background in Z events producing fake di-muon combinations.

The signal yield in the four categories can be further rewritten in terms of the number of produced $Z \rightarrow \mu^+\mu^-$ events, $N_{Z \rightarrow \mu^+\mu^-}$, and the average efficiencies for muon reconstruction in the tracker (ϵ_{trk}), in the muon detector as a stand-alone track (ϵ_{sa}), the average efficiency of the isolation cut (ϵ_{iso}) and the average HLT efficiency (ϵ_{HLT}):

$$N_{\mu\mu}^{2\text{HLT}} = N_{Z \rightarrow \mu^+\mu^-} \epsilon_{\text{HLT}}^2 \epsilon_{\text{iso}}^2 \epsilon_{\text{trk}}^2 \epsilon_{\text{sa}}^2, \quad (6.17)$$

$$N_{\mu\mu}^{1\text{HLT}} = 2N_{Z \rightarrow \mu^+\mu^-} \epsilon_{\text{HLT}} (1 - \epsilon_{\text{HLT}}) \epsilon_{\text{iso}}^2 \epsilon_{\text{trk}}^2 \epsilon_{\text{sa}}^2, \quad (6.18)$$

$$N_{\mu s} = 2N_{Z \rightarrow \mu^+\mu^-} \epsilon_{\text{HLT}} \epsilon_{\text{iso}}^2 \epsilon_{\text{trk}} (1 - \epsilon_{\text{trk}}) \epsilon_{\text{sa}}^2, \quad (6.19)$$

$$N_{\mu t} = 2N_{Z \rightarrow \mu^+\mu^-} \epsilon_{\text{HLT}} \epsilon_{\text{iso}}^2 \epsilon_{\text{trk}}^2 \epsilon_{\text{sa}} (1 - \epsilon_{\text{sa}}), \quad (6.20)$$

$$N_{\mu\mu}^{\text{non iso}} = N_{Z \rightarrow \mu^+\mu^-} (1 - (1 - \epsilon_{\text{HLT}})^2) (1 - \epsilon_{\text{iso}}^2) \epsilon_{\text{trk}}^2 \epsilon_{\text{sa}}^2. \quad (6.21)$$

This factorization is done neglecting the correlations between the number of entries in the various categories. This assumption will be justified and discussed in more details in Section 6.4.2. We can assume that the peak distribution is identical in the categories $Z_{\mu\mu}$, $Z_{\mu t}$ and $Z_{\mu\mu}^{\text{non iso}}$ because the muon momentum resolution in CMS is determined by the tracker measurement for muon with $p_T \leq 200$ GeV, as for the muon from Z decay. We can also neglect the background in the $Z_{\mu\mu}$ category (of the order of few per mille) and we take as distribution for $f_{peak}(m)$ the spectrum of the di-muon invariant mass in the $Z_{\mu\mu}$ category. We have re-binned the distribution in order to match the bin width in the $Z_{\mu t}$ and $Z_{\mu\mu}^{\text{non iso}}$ categories. In Fig. 6.6 and 6.7 we show the level of agreement of the invariant mass distribution for $Z_{\mu\mu}$ candidates selected in signal events with the distributions for $Z_{\mu t}$ and $Z_{\mu\mu}^{\text{non iso}}$ candidates.

We have also assumed that the isolation efficiency is identical for global muons, tracks and stand-alone muons. For the latter, in particular, the worse direction resolution could produce a slightly different isolation efficiency. We measured on the Monte Carlo signal sample that the difference in isolation efficiency is very small, and compatible with zero within errors:

$$\epsilon_{\text{iso}}^{\text{s.a.}} - \epsilon_{\text{iso}}^{\text{glob.}} = 0.007 \pm 0.057\% .$$

In order to determine from data a model for the PDF of the peak function for the $Z_{\mu s}$ category, $f_{peak}^s(m)$, we consider the $Z_{\mu\mu}$ candidates, and for one of the muons

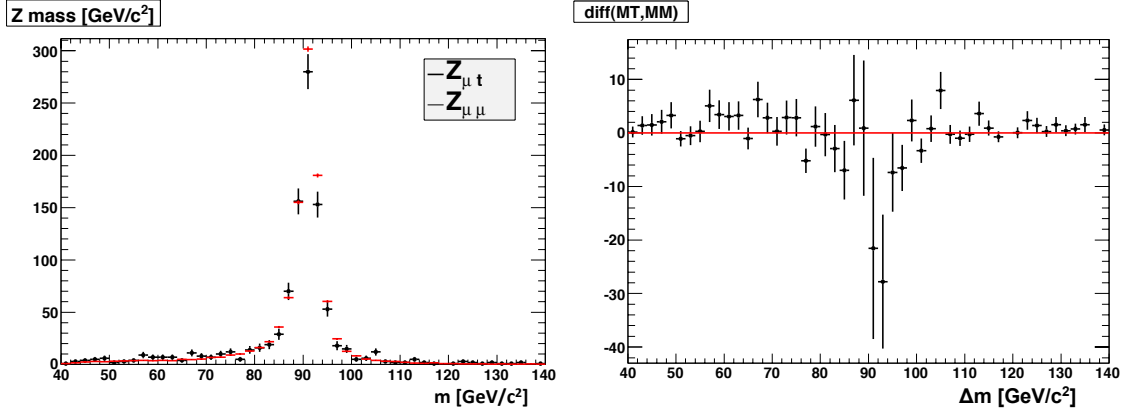


Figure 6.6: Left: invariant mass distribution for selected $Z_{\mu\mu}$ (red points) and $Z_{\mu t}$ (black points) candidates in signal events. The $Z_{\mu\mu}$ distribution is normalized in order to have the same number of events as the $Z_{\mu t}$ sample. Right: difference between the $Z_{\mu\mu}$ and $Z_{\mu t}$ distributions.

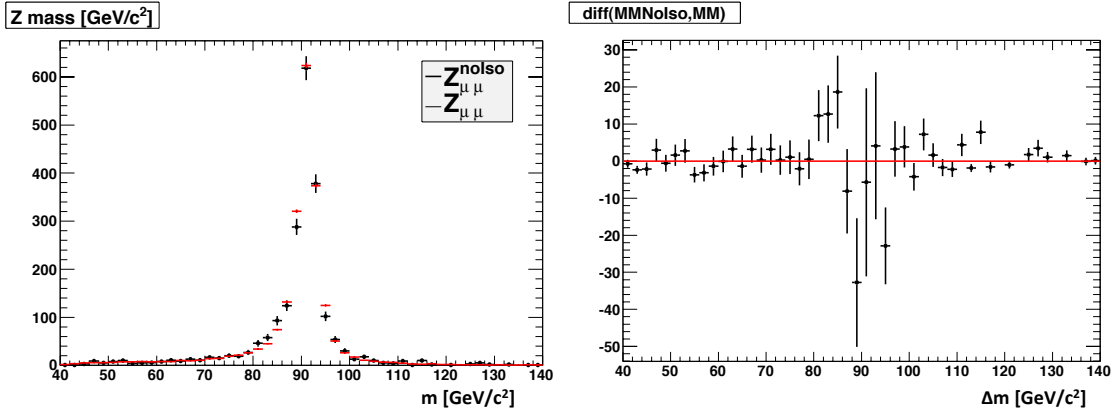


Figure 6.7: Left: invariant mass distribution for selected $Z_{\mu\mu}$ (red points) and $Z_{\mu\mu}^{\text{non iso}}$ (black points) candidates in signal events. The $Z_{\mu\mu}$ distribution is normalized in order to have the same number of events as the $Z_{\mu\mu}^{\text{non iso}}$ sample. Right: difference between the $Z_{\mu\mu}$ and $Z_{\mu\mu}^{\text{non iso}}$ distributions.

we take the momentum measured from the muon detector track fit only, in order to mimic a stand-alone muon. We avoid to put the same event twice in the histogram, by choosing alternatively the first (second) muon for even (odd) events respectively. This makes the signal shape description entirely data-driven. Figure 6.8 compares the invariant mass distribution of the selected $Z_{\mu s}$ candidates with the shape obtained from $Z_{\mu\mu}$ candidates.

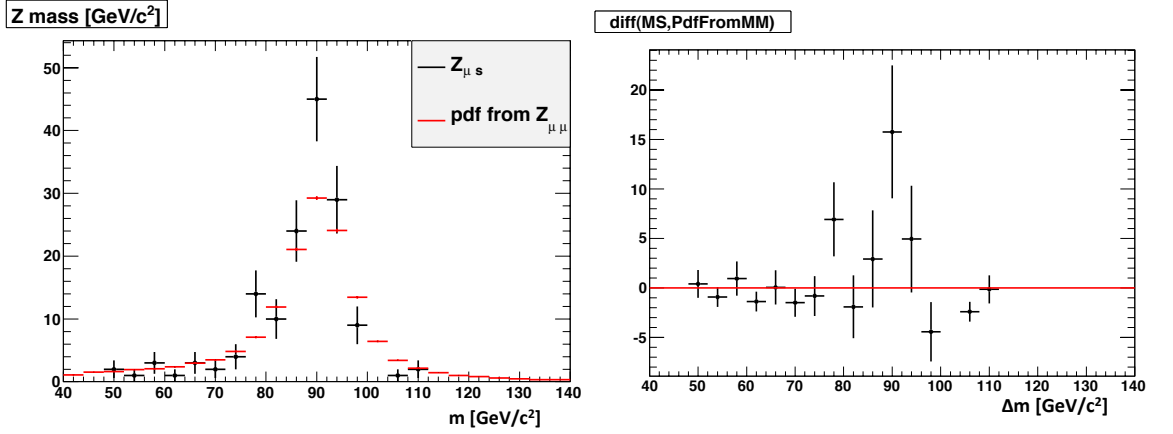


Figure 6.8: Left: invariant mass distribution for selected $Z_{\mu s}$ candidates in signal events (black points) superimposed to the pdf (red points) determined from $Z_{\mu\mu}$ candidates by using, for one of the muons in the pair, the momentum of the associated stand-alone muon. Right: difference between the $Z_{\mu\mu}$ and $Z_{\mu s}$ distributions.

Background functions are modeled as products of exponential terms with polynomials of different order for the three samples for which the background is not neglected:

$$b_{\mu t}(m) = N_{\mu t}^b (1 + a_1 m + a_2 m^2) e^{-\alpha m} \quad (6.22)$$

$$b_{\mu\mu}^{\text{non iso}}(m) = N_{\mu\mu}^{\text{non iso}} (1 + b_1 m + b_2 m^2) e^{-\beta m} \quad (6.23)$$

$$b_{\mu s}(m) = N_{\mu s}^b (1 + c_1 m + c_2 m^2) e^{-\gamma m} \quad (6.24)$$

We divide the invariant di-muon mass spectrum of the different categories into bins whose size depends on the statistics of each of the five categories. We define, for each category, minus two times the negative logarithmic of the following Poissonian likelihood ratio [86]:

$$\chi_\lambda^2 = -2 \ln \lambda(m) = -2 \ln \frac{\text{Poiss}(n_i, \nu_i)}{\text{Poiss}(n_i, n_i)} = \sum_{i=1, \dots, n_{\text{bins}}} \nu_i - n_i + n_i \log \frac{n_i}{\nu_i}, \quad (6.25)$$

where ν_i is the expected number of events in the i^{th} -bin of the di-muon invariant mass histograms and n_i is the measured number of events in that bin, and $\text{Poiss}(n, \nu)$ is

a Poissonian distribution. In order to perform the simultaneous fit, we minimize the sum R of the five χ_λ^2 , i.e.:

$$R = \sum_{j=1,\dots,5} \chi_{\lambda,j}^2 = 2 \sum_{j=1,\dots,5} \sum_{i=1,\dots,n_{\text{bins}}^{(j)}} \nu_i^{(j)} - n_i^{(j)} + n_i^{(j)} \log \frac{n_i^{(j)}}{\nu_i^{(j)}}, \quad (6.26)$$

where the index j runs on the five categories, and the index i on the bins of the j^{th} category. A benefit of this statistic technique is that it allows a goodness-of-fit test, as for sufficiently large ν_i the minimum of R follows a χ^2 distribution. Using a single bin (counting only) and a Gaussian approximation for the two categories with the largest statistics, $Z_{\mu\mu}^{2\text{HLT}}$ and $Z_{\mu\mu}^{1\text{HLT}}$, the Z yield and the four efficiencies are obtained by minimizing:

$$R = \frac{(N_{\mu\mu}^{2\text{HLT}} - N_{Z \rightarrow \mu^+ \mu^-} \epsilon_{\text{HLT}}^2 \epsilon_{\text{iso}}^2 \epsilon_{\text{trk}}^2 \epsilon_{\text{sa}}^2)^2}{N_{\mu\mu}^{2\text{HLT}}} + \frac{(N_{\mu\mu}^{1\text{HLT}} - 2N_{Z \rightarrow \mu^+ \mu^-} \epsilon_{\text{HLT}} (1 - \epsilon_{\text{HLT}}) \epsilon_{\text{iso}}^2 \epsilon_{\text{trk}}^2 \epsilon_{\text{sa}}^2)^2}{N_{\mu\mu}^{1\text{HLT}}} + \chi_{\lambda,\mu s}^2 + \chi_{\lambda,\mu t}^2 + \chi_{\lambda,\mu\mu}^{\text{non iso } 2},$$

where we count the events $Z_{\mu\mu}^{1\text{HLT}}$ and $Z_{\mu\mu}^{2\text{HLT}}$ golden categories and $\chi_{\lambda,\mu s}^2$, $\chi_{\lambda,\mu t}^2$ and $\chi_{\lambda,\mu\mu}^{\text{non iso } 2}$ are the Poisson likelihood ratio of the di-muon mass binned histograms for the three categories $Z_{\mu s}$, $Z_{\mu t}$, and $Z_{\mu\mu}^{\text{non iso}}$. The likelihood ratio fit is equivalent to an ordinary χ^2 for enough large statistics and we verified that, with the current data sample, it gives the same central value and statistical uncertainty. We perform the fit in the range $60 < m < 120$ GeV/ c^2 .

6.4.2 Correlation studies

In the following sub-sections we consider the possible effect of:

- kinematic correlation between the two muons
- correlation of muon detector reconstruction and HLT efficiency distributions as functions of η and p_T
- correlation of tracker reconstruction and isolation efficiency distributions

Efficiency correlation between the two muons

In the above fit model we assumed that we can factorize the efficiency terms for the two muons (that are products of reconstruction, isolation and trigger efficiencies).

We now justify this assumptions, that consists in neglecting the correlation of the efficiency terms and provide a method to estimate the uncertainty caused by this assumption.

The differential Z yield, as a function of the two muon three-momenta, can be written as:

$$\frac{d^3n^0}{d^3p_1d^3p_2} = N^0 f_0(\vec{p}_1, \vec{p}_2) \times \epsilon_1(\vec{p}_1)\epsilon_2(\vec{p}_2) , \quad (6.27)$$

where N^0 is the total number of produced events, \vec{p}_1 and \vec{p}_2 are the two muons three-momenta, $f_0(\vec{p}_1, \vec{p}_2)$ is the probability density function that takes into account the process matrix element, phase space, and detector resolutions. We introduce the efficiency terms for the two muons as: $\epsilon_1(\vec{p}_1)$, $\epsilon_2(\vec{p}_2)$, whose interpretation varies for the different samples we consider. In the case of the sample reconstructed as a pair of global muons, for instance, the two functions ϵ_1 and ϵ_2 coincide, and are equal to the product $\epsilon_{\text{trk}}(\vec{p})\epsilon_{\text{sa}}(\vec{p})\epsilon_{\text{iso}}(\vec{p})\epsilon_{\text{HLT}}(\vec{p})$.

The differential Z yield as a function of the muon pair invariant mass is:

$$\frac{dn}{dm} = N_0 \int d^3p_1d^3p_2 f_0(\vec{p}_1, \vec{p}_2) \delta(m_{12}(\vec{p}_1, \vec{p}_2) - m) \epsilon_1(\vec{p}_1)\epsilon_2(\vec{p}_2) , \quad (6.28)$$

or sum of similar terms with different $\epsilon_1(\vec{p}_1)$ and $\epsilon_2(\vec{p}_2)$. The efficiency terms for the other categories are reported in Table 6.5 for completeness. In the above equation m_{12} is the di-muon invariant mass, which can be written as $m_{12} = (\vec{p}_1, \vec{p}_2) = 2p_1p_2(1 - \cos\theta_{12})$ neglecting the muon mass.

$$\frac{dn^0}{dm} = N_0 \int d^3p_1d^3p_2 f_0(\vec{p}_1, \vec{p}_2) \delta(m_{12}(\vec{p}_1, \vec{p}_2) - m) , \quad (6.29)$$

In the fit model above we made the approximation that the efficiency terms can be factorized as average terms $\bar{\epsilon}_1$, $\bar{\epsilon}_2$:

$$\frac{dn}{dm} \simeq \frac{dn'}{dm} = N_0 \bar{\epsilon}_1 \bar{\epsilon}_2 \int d^3p_1d^3p_2 f_0(\vec{p}_1, \vec{p}_2) \delta(m_{12}(\vec{p}_1, \vec{p}_2) - m) , . \quad (6.30)$$

where:

$$\bar{\epsilon}_1 = \langle \epsilon_1(\vec{p}_1) \rangle = \int d^3p_1d^3p_2 f_0(\vec{p}_1, \vec{p}_2) \epsilon_1(\vec{p}_1) , \quad (6.31)$$

$$\bar{\epsilon}_2 = \langle \epsilon_2(\vec{p}_2) \rangle = \int d^3p_1d^3p_2 f_0(\vec{p}_1, \vec{p}_2) \epsilon_2(\vec{p}_2) . \quad (6.32)$$

The difference between the approximated and exact expressions, due to the normalization of $f(\vec{p}_1, \vec{p}_2)$:

$$\int d^3p_1d^3p_2 f_0(\vec{p}_1, \vec{p}_2) = 1 \quad (6.33)$$

| Category | $\epsilon_1(\vec{p}_1)$ | $\epsilon_2(\vec{p}_2)$ |
|-------------------------------|---|---|
| $Z_{\mu\mu}^{2\text{HLT}}$ | $\epsilon_{\text{trk}}(\vec{p}_1)\epsilon_{\text{sa}}(\vec{p}_1)\epsilon_{\text{iso}}(\vec{p}_1)\epsilon_{\text{HLT}}(\vec{p}_1)$ | $\epsilon_{\text{trk}}(\vec{p}_2)\epsilon_{\text{sa}}(\vec{p}_2)\epsilon_{\text{iso}}(\vec{p}_2)\epsilon_{\text{HLT}}(\vec{p}_2)$ |
| $Z_{\mu\mu}^{1\text{HLT}}$ | $\epsilon_{\text{trk}}(\vec{p}_1)\epsilon_{\text{sa}}(\vec{p}_1)\epsilon_{\text{iso}}(\vec{p}_1)\epsilon_{\text{HLT}}(\vec{p}_1)$ $\epsilon_{\text{trk}}(\vec{p}_1)\epsilon_{\text{sa}}(\vec{p}_1)\epsilon_{\text{iso}}(\vec{p}_1)(1 - \epsilon_{\text{HLT}}(\vec{p}_1))$ | $\epsilon_{\text{trk}}(\vec{p}_2)\epsilon_{\text{sa}}(\vec{p}_2)\epsilon_{\text{iso}}(\vec{p}_2)(1 - \epsilon_{\text{HLT}}(\vec{p}_2))$ $\epsilon_{\text{trk}}(\vec{p}_2)\epsilon_{\text{sa}}(\vec{p}_2)\epsilon_{\text{iso}}(\vec{p}_2)\epsilon_{\text{HLT}}(\vec{p}_2)$ |
| $Z_{\mu s}$ | $\epsilon_{\text{trk}}(\vec{p}_1)\epsilon_{\text{sa}}(\vec{p}_1)\epsilon_{\text{iso}}(\vec{p}_1)\epsilon_{\text{HLT}}(\vec{p}_1)$ $(1 - \epsilon_{\text{trk}}(\vec{p}_1))\epsilon_{\text{sa}}(\vec{p}_1)\epsilon_{\text{iso}}(\vec{p}_1)$ | $(1 - \epsilon_{\text{trk}}(\vec{p}_2))\epsilon_{\text{sa}}(\vec{p}_2)\epsilon_{\text{iso}}(\vec{p}_2)$ $\epsilon_{\text{trk}}(\vec{p}_2)\epsilon_{\text{sa}}(\vec{p}_2)\epsilon_{\text{iso}}(\vec{p}_2)\epsilon_{\text{HLT}}(\vec{p}_2)$ |
| $Z_{\mu t}$ | $\epsilon_{\text{trk}}(\vec{p}_1)\epsilon_{\text{sa}}(\vec{p}_1)\epsilon_{\text{iso}}(\vec{p}_1)\epsilon_{\text{HLT}}(\vec{p}_1)$ $\epsilon_{\text{trk}}(\vec{p}_1)(1 - \epsilon_{\text{sa}}(\vec{p}_1))\epsilon_{\text{iso}}(\vec{p}_1)$ | $\epsilon_{\text{trk}}(\vec{p}_2)(1 - \epsilon_{\text{sa}}(\vec{p}_2))\epsilon_{\text{iso}}(\vec{p}_2)$ $\epsilon_{\text{trk}}(\vec{p}_2)\epsilon_{\text{sa}}(\vec{p}_2)\epsilon_{\text{iso}}(\vec{p}_2)\epsilon_{\text{HLT}}(\vec{p}_2)$ |
| $Z_{\mu\mu}^{\text{non iso}}$ | $\epsilon_{\text{trk}}(\vec{p}_1)\epsilon_{\text{sa}}(\vec{p}_1)\epsilon_{\text{iso}}(\vec{p}_1)\epsilon_{\text{HLT}}(\vec{p}_1)$ $\epsilon_{\text{trk}}(\vec{p}_1)\epsilon_{\text{sa}}(\vec{p}_1)(1 - \epsilon_{\text{iso}}(\vec{p}_1))\epsilon_{\text{HLT}}(\vec{p}_1)$ $\epsilon_{\text{trk}}(\vec{p}_1)\epsilon_{\text{sa}}(\vec{p}_1)(1 - \epsilon_{\text{iso}}(\vec{p}_1))\epsilon_{\text{HLT}}(\vec{p}_1)$ $\epsilon_{\text{trk}}(\vec{p}_1)\epsilon_{\text{sa}}(\vec{p}_1)\epsilon_{\text{iso}}(\vec{p}_1)\epsilon_{\text{HLT}}(\vec{p}_1)$ $\epsilon_{\text{trk}}(\vec{p}_1)\epsilon_{\text{sa}}(\vec{p}_1)(1 - \epsilon_{\text{iso}}(\vec{p}_1))\epsilon_{\text{HLT}}(\vec{p}_1)$ $\epsilon_{\text{trk}}(\vec{p}_1)\epsilon_{\text{sa}}(\vec{p}_1)(1 - \epsilon_{\text{iso}}(\vec{p}_1))\epsilon_{\text{HLT}}(\vec{p}_1)$ $\epsilon_{\text{trk}}(\vec{p}_1)\epsilon_{\text{sa}}(\vec{p}_1)\epsilon_{\text{iso}}(\vec{p}_1)(1 - \epsilon_{\text{HLT}}(\vec{p}_1))$ $\epsilon_{\text{trk}}(\vec{p}_1)\epsilon_{\text{sa}}(\vec{p}_1)(1 - \epsilon_{\text{iso}}(\vec{p}_1))(1 - \epsilon_{\text{HLT}}(\vec{p}_1))$ $\epsilon_{\text{trk}}(\vec{p}_1)\epsilon_{\text{sa}}(\vec{p}_1)(1 - \epsilon_{\text{iso}}(\vec{p}_1))(1 - \epsilon_{\text{HLT}}(\vec{p}_1))$ | $\epsilon_{\text{trk}}(\vec{p}_2)\epsilon_{\text{sa}}(\vec{p}_2)(1 - \epsilon_{\text{iso}}(\vec{p}_2))\epsilon_{\text{HLT}}(\vec{p}_2)$ $\epsilon_{\text{trk}}(\vec{p}_2)\epsilon_{\text{sa}}(\vec{p}_2)\epsilon_{\text{iso}}(\vec{p}_2)\epsilon_{\text{HLT}}(\vec{p}_2)$ $\epsilon_{\text{trk}}(\vec{p}_2)\epsilon_{\text{sa}}(\vec{p}_2)(1 - \epsilon_{\text{iso}}(\vec{p}_2))\epsilon_{\text{HLT}}(\vec{p}_2)$ $\epsilon_{\text{trk}}(\vec{p}_2)\epsilon_{\text{sa}}(\vec{p}_2)(1 - \epsilon_{\text{iso}}(\vec{p}_2))\epsilon_{\text{HLT}}(\vec{p}_2)$ $\epsilon_{\text{trk}}(\vec{p}_2)\epsilon_{\text{sa}}(\vec{p}_2)\epsilon_{\text{iso}}(\vec{p}_2)(1 - \epsilon_{\text{HLT}}(\vec{p}_2))$ $\epsilon_{\text{trk}}(\vec{p}_2)\epsilon_{\text{sa}}(\vec{p}_2)(1 - \epsilon_{\text{iso}}(\vec{p}_2))(1 - \epsilon_{\text{HLT}}(\vec{p}_2))$ $\epsilon_{\text{trk}}(\vec{p}_2)\epsilon_{\text{sa}}(\vec{p}_2)(1 - \epsilon_{\text{iso}}(\vec{p}_2))\epsilon_{\text{HLT}}(\vec{p}_2)$ $\epsilon_{\text{trk}}(\vec{p}_2)\epsilon_{\text{sa}}(\vec{p}_2)\epsilon_{\text{iso}}(\vec{p}_2)\epsilon_{\text{HLT}}(\vec{p}_2)$ $\epsilon_{\text{trk}}(\vec{p}_2)\epsilon_{\text{sa}}(\vec{p}_2)(1 - \epsilon_{\text{iso}}(\vec{p}_2))\epsilon_{\text{HLT}}(\vec{p}_2)$ |

Table 6.5: List of the efficiency terms to be used in Equation (6.28) for the different r reconstructed Z categories.

is:

$$\frac{dn}{dm} - \frac{dn'}{dm} = N_0 \int d^3p_1 d^3p_2 f_0(\vec{p}_1, \vec{p}_2) \delta(m_{12}(\vec{p}_1, \vec{p}_2) - m) (\epsilon_1(\vec{p}_1)\epsilon_2(\vec{p}_2) - \bar{\epsilon}_1\bar{\epsilon}_2) . \quad (6.34)$$

Integrating over the mass m , in order to extract the cross section, in a range $[m_1, m_2]$, one has:

$$N = \int_{m_1}^{m_2} dm \frac{dn}{dm} , \quad (6.35)$$

$$N' = \int_{m_1}^{m_2} dm \frac{dn'}{dm} , \quad (6.36)$$

hence:

$$\begin{aligned} N - N' &= N_0 \int d^3p_1 d^3p_2 f_0(\vec{p}_1, \vec{p}_2) (\epsilon_1(\vec{p}_1)\epsilon_2(\vec{p}_2) - \bar{\epsilon}_1\bar{\epsilon}_2) \\ &= N_0 \langle \epsilon_1(\vec{p}_1)\epsilon_2(\vec{p}_2) - \bar{\epsilon}_1\bar{\epsilon}_2 \rangle \\ &= N_0 \langle (\epsilon_1(\vec{p}_1) - \bar{\epsilon}_1)(\epsilon_2(\vec{p}_2) - \bar{\epsilon}_2) \rangle \\ &= N_0 \text{cov}(\epsilon_1(\vec{p}_1), \epsilon_2(\vec{p}_2)) \equiv N_0 \text{cov}_{12} , \end{aligned}$$

or equivalently:

$$\frac{N - N'}{N_0} = \frac{\Delta N}{N_0} = \text{cov}(\epsilon_1(\vec{p}_1), \epsilon_2(\vec{p}_2)) . \quad (6.37)$$

So, the assumption we made is equivalent to neglect the correlation term between the two muon efficiencies, cov_{12} .

Note that the above term is quadratic in the statistical dispersion of the efficiencies in the p_t and η range considered. So, taking two efficiencies for the table 6.5, if we assume that $\epsilon_k(\vec{p}_k) - \bar{\epsilon}_k$ ($k = 1, 2$) is at most δ , the relative systematic error introduced by the approximation will be smaller than δ^2 . This would give a first way to estimate an upper limit to this systematic effect just looking at the efficiency excursion in the efficiency tables obtained with other method such as the ‘‘Tag and Probe’’ (see Section 6.4.10 for an explanation of this method): a 10% dispersion would give a 1% effect.

A more precise way to estimate this effect could be done using the T&P efficiency tables ($\epsilon_k^{tp}(\vec{p})$). We could estimate the needed terms as discrete averages over the signal sample:

$$\bar{\epsilon}_k^{tp} = \frac{1}{N_{\text{obs}}^{\mu, k}} \sum_{i=1, \dots, n}^{N_{\text{obs}}^{\mu, k}} \epsilon_k^{tp}(\vec{p}_i), \quad k = 1, 2, \quad (6.38)$$

and:

$$\text{cov}_{12}^{tp} = \frac{1}{N_{\text{obs}}^Z} \sum_{i=1, \dots, n}^{N_{\text{obs}}^Z} (\epsilon_1^{tp}(p_{1(i)}) - \bar{\epsilon}_1^{tp})(\epsilon_2^{tp}(p_{2(i)}) - \bar{\epsilon}_2^{tp}) \quad (6.39)$$

$$= \frac{1}{N_{\text{obs}}^Z} \sum_{i=1, \dots, n}^{N_{\text{obs}}^Z} \epsilon_1^{tp}(p_{1(i)}) \epsilon_2^{tp}(p_{2(i)}) - \bar{\epsilon}_1^{tp} \bar{\epsilon}_2^{tp}. \quad (6.40)$$

Above, N_{obs}^Z is the number of observed Z events and is the number of observed muons in Z events for the two category $k = 1, 2$ or for the unique category, in case of Z reconstructed from a pair of global muons.

Studies based on MC samples demonstrated that the correlation effect can be neglected at the 0.01% level.

Correlation between trigger efficiency and reconstruction efficiency

In this section we discuss the correlation between trigger and reconstruction efficiency. This correlation cannot be neglected a-priori and brings to the definition of an effective average trigger efficiency. The trigger efficiency we are considering is indeed the product of the efficiencies L1, L2 and L3 trigger event selection paths.

We rewrite Equation (6.31) as follows:

$$\bar{\epsilon}_i = \langle \epsilon_i(\vec{p}_i) \rangle = \int d^3 p_1 d^3 p_2 f_0(\vec{p}_1, \vec{p}_2) \epsilon_i(\vec{p}_i), \quad i = 1, 2, \quad (6.41)$$

where $\epsilon_i(\vec{p}_i)$, $i = 1, 2$, is one of the terms listed in Table 6.5. We define for simplicity, omitting the subscript i :

$$\bar{\epsilon} = \langle \epsilon(\vec{p}) \rangle = \int d^3p f_0(\vec{p}) \epsilon(\vec{p}) , \quad (6.42)$$

where, for $i = 1$, $f_0(\vec{p}_1) = \int d^3p_2 f_0(\vec{p}_1, \vec{p}_2)$, and similarly for $i = 2$, $f_0(\vec{p}_2) = \int d^3p_1 f_0(\vec{p}_1, \vec{p}_2)$.

In the case of the sample reconstructed as a pair of global muons, for instance, the average:

$$\bar{\epsilon} = \langle \epsilon_{\text{trk}}(\vec{p}) \epsilon_{\text{sa}}(\vec{p}) \epsilon_{\text{iso}}(\vec{p}) \epsilon_{\text{HLT}}(\vec{p}) \rangle \quad (6.43)$$

does not coincide with the product of the averages: $\langle \epsilon_{\text{trk}}(\vec{p}) \rangle \langle \epsilon_{\text{sa}}(\vec{p}) \rangle \langle \epsilon_{\text{iso}}(\vec{p}) \rangle \langle \epsilon_{\text{HLT}}(\vec{p}) \rangle$, and again we can assume factorization wherever correlation terms can be neglected. It is reasonable to assume that $\epsilon_{\text{iso}}(\vec{p})$ and $\epsilon_{\text{trk}}(\vec{p})$ are uncorrelated w.r.t. the other terms while it could be not the case for $\epsilon_{\text{sa}}(\vec{p})$ and $\epsilon_{\text{HLT}}(\vec{p})$, which can be correlated, since single muon trigger is very related to the geometry of the muon detector.

When we express the differential Z yields of the different categories, we have, for each of the two muons, efficiency terms that contain $\epsilon_{\text{sa}}(\vec{p})$ and $\epsilon_{\text{HLT}}(\vec{p})$ either as products $\epsilon_{\text{sa}}(\vec{p}) \cdot \epsilon_{\text{HLT}}(\vec{p})$, or as single terms containing just $\epsilon_{\text{sa}}(\vec{p})$. We never find single terms in $\epsilon_{\text{HLT}}(\vec{p})$. Thus, when we compute the average terms, we are still allowed to use factorization in Equations (6.17), (6.18), (6.19), (6.20) and (6.4.4), but we have to re-define ϵ_{HLT} as:

$$\epsilon_{\text{HLT}} = \frac{\langle \epsilon_{\text{sa}}(\vec{p}) \cdot \epsilon_{\text{HLT}}(\vec{p}) \rangle}{\langle \epsilon_{\text{sa}}(\vec{p}) \rangle} , \quad (6.44)$$

and so we interpret ϵ_{HLT} as the efficiency to trigger a muon which has been correctly reconstructed in the muon system.

Correlation between tracking efficiency and isolation efficiency

A correlation between tracking efficiency and isolation efficiency may occur in case of very bad tracker noise or large event pile-up situation, in which a simultaneous loss of tracker efficiency and isolation power generated by excess of noise in some detector regions could be present.

If we don't neglect this correlation, a similar treatment as it was discussed in Section 6.4.2 can be done. In a similar way, we can re-define an "effective" isolation efficiency, similarly to Eq. (6.44):

$$\epsilon_{\text{iso}} = \frac{\langle \epsilon_{\text{trk}}(\vec{p}) \cdot \epsilon_{\text{iso}}(\vec{p}) \rangle}{\langle \epsilon_{\text{trk}}(\vec{p}) \rangle} , \quad (6.45)$$

interpreting ϵ_{iso} as the efficiency of the isolation cut on muon fully reconstructed in the tracking system. All efficiency terms in the definition of Z categories from

Equations (6.17), (6.18), (6.20) and (6.4.4) remain unchanged, but correlation must be taken into account in the efficiency term of the $Z_{\mu s}$ category, in Equation (6.19). The term, including correlation, is:

$$\begin{aligned} & \langle \epsilon_{\text{trk}}(\vec{p}_1) \epsilon_{\text{iso}}(\vec{p}_1) (1 - \epsilon_{\text{trk}}(\vec{p}_2)) \epsilon_{\text{iso}}(\vec{p}_2) \rangle + \\ & \langle (1 - \epsilon_{\text{trk}}(\vec{p}_1)) \epsilon_{\text{iso}}(\vec{p}_1) \epsilon_{\text{trk}}(\vec{p}_2) \epsilon_{\text{iso}}(\vec{p}_2) \rangle = \\ & 2 \langle \epsilon_{\text{trk}} \epsilon_{\text{iso}} \rangle (\langle \epsilon_{\text{iso}} \rangle - \langle \epsilon_{\text{trk}} \epsilon_{\text{iso}} \rangle) . \end{aligned}$$

Replacing in the above term:

$$\begin{aligned} \langle \epsilon_{\text{trk}} \epsilon_{\text{iso}} \rangle &= \epsilon_{\text{trk}} \epsilon_{\text{iso}} \\ \langle \epsilon_{\text{iso}} \rangle &= \epsilon_{\text{iso}} - \frac{\text{COV}_{\text{trk iso}}}{\epsilon_{\text{trk}}} = \epsilon_{\text{iso}} \left(1 - \frac{\text{COV}_{\text{trk iso}}}{\epsilon_{\text{trk}} \epsilon_{\text{iso}}} \right) . \end{aligned}$$

We can correct Equations (6.19) including a possible correlation term:

$$N_{\mu s} = 2N_{Z \rightarrow \mu^+ \mu^-} \epsilon_{\text{HLT}}^2 \epsilon_{\text{iso}}^2 \epsilon_{\text{trk}}^2 \epsilon_{\text{sa}}^2 \left((1 - \epsilon_{\text{trk}}) - \frac{\text{COV}_{\text{trk iso}}}{\epsilon_{\text{trk}} \epsilon_{\text{iso}}} \right) . \quad (6.46)$$

This correction would only affect the $Z_{\mu s}$ category that is used to determine the tracker efficiency which we expect that, under normal detector operation, would be very close to one. So, we expect this category to be the one with the smallest statistics. A deviation of the number of $Z_{\mu s}$ events would result in a corresponding variation on the tracker inefficiency $(1 - \epsilon_{\text{trk}})$, that would result in a much smaller relative variation of ϵ_{trk} , being ϵ_{trk} close to the unity.

If this covariance term would turn out to be significantly different from zero, as alternative, we can drop the isolation request to the stand-alone muon, and this will allow to fully absorb the covariance term into the redefinition of ϵ_{iso} , as done in Section 6.4.2 for ϵ_{HLT} .

Tracker background and event pile-up don't have such a serious impact, especially at low luminosity, to impair dramatically the tracker performance, as in the case of the current measurement described in this work. Anyway, in order to estimate correctly the correlation term under those pessimistic conditions, a realistic estimate would need either a proper simulation of those detector and run conditions, or control samples from real data taken under those conditions. It's important to note that the same effect may also affect other currently used methods to estimate detector efficiencies, such as the Tag and Probe method (see Section 6.4.10).

6.4.3 Data samples

The 2010 7 TeV collision data sample used for this analysis (Run2010A) amounts to an integrated luminosity of 2.9 pb^{-1} (see Section 6.5 for details on how the luminosity

measurement is done in CMS). We have used the following Monte Carlo samples generated with the standard CMS generator, simulation and reconstruction chain, assuming a center of mass energy of $\sqrt{s} = 7$ TeV with the same detector condition and reconstruction parameter used for the data taking:

- $pp \rightarrow Z/\gamma^* X \rightarrow \mu^+ \mu^- X$
- $pp \rightarrow W^\pm X \rightarrow \mu^\pm \nu_\mu X$
- QCD jets events containing at least one muon with $p_t > 15 \text{ GeV}/c$
- $t\bar{t}$.
- WZ and WW

The number of events for each MC sample, the product of the NLO cross section (σ) times the generator kinematic filter efficiency (ϵ_{filter} , i.e. the fraction of generated event witch fall effectively in the kinetic region accessible by the detectors), and the equivalent integrated luminosity ($\int \mathcal{L} dt$) are shown in Table 6.6.

| Sample | $\sigma \times \epsilon_{\text{filter}}$ (pb) | $\int \mathcal{L} dt$ (pb^{-1}) |
|--|---|--|
| $pp \rightarrow ZX \rightarrow \mu^+ \mu^- X (m > 20)$ | 1686 | 100 |
| $pp \rightarrow W^+ X \rightarrow \mu^+ \nu_\mu X$ | 6152 | 100 |
| $pp \rightarrow W^- X \rightarrow \mu^- \nu_\mu X$ | 4179 | 100 |
| QCD jets, muon $p_t > 15 \text{ GeV}/c$ | 79688 | 60 |
| $t\bar{t}$ | 162 | 100 |
| $pp \rightarrow ZX \rightarrow \tau^+ \tau^- X$ | 1686 | 100 |
| WZ | 43 | 537 |
| WW | 18 | 100 |
| ZZ | 5.9 | 100 |

Table 6.6: Analyzed MC data samples.

The analyzed samples are first processed through a pre-selection phase denominated “skimming” where HLT requirements are applied. At skimming level, it is also required the presence of at least two reconstructed muons or one muon plus one tracker track, with $p_T > 15 \text{ GeV}/c$. After that, in order to further reduce the data samples, a secondary skim (sub-skim) was run and the a very light output was has been produced.

6.4.4 Event selection

Events are required to satisfy the single non-isolated muon trigger with a (L3 trigger step) p_T cut of 9 GeV (7 GeV at L1 trigger step) .

Muons used for Z reconstruction are checked for matching with an HLT object (see Sect. 6.4.10 for the HLT matching requirement used in the different samples of Z candidates and for additional muon trigger efficiencies estimation results with $Z_{\mu\mu}$).

We require that both muon candidates, either global or stand-alone muons, or tracker tracks, must satisfy:

$$p_T > 20 \text{ GeV}/c, \text{ and } |\eta| < 2.1 .$$

We perform the fit in the di-muon mass range $60 < m_{\mu^+\mu^-} < 120 \text{ GeV}/c^2$.

We define as isolation variable the sum of the transverse momenta of all tracks within a cone of radius $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} = 0.3$ plus the sum of the calorimetric deposit in Ecal and Hcal, divided by the muon momentum itself.

$$I_{rel}^{comb} = \frac{\sum_{\Delta R < 0.3} I_{trk} + I_{Ecal} + I_{HCal}}{p_T} . \quad (6.47)$$

We require:

$$I_{rel}^{comb} < 0.15 .$$

Quality cuts to the tracks are applied in order to select a very pure sample. The efficiencies of these additional selection cuts are included together with the reconstruction efficiency terms. We require than a tracker track have:

- at least 11 tracker hits (pixel + silicon tracker layers)
- at least 1 pixel hit

The efficiency ϵ_{trk} , which appears in the equations 6.17-, is then re-defined as the efficiency to reconstruct a track and the two additional efficiencies $\epsilon_{trk} \rightarrow \epsilon_{trk} \times \epsilon_{\#TrackerHits>10} \times \epsilon_{\#PixelHits>0}$. Events reconstructed as global muon pairs in which one of the two muon does not satisfy these two additional requirements fall in the $Z_{\mu s}$ category. In addition we require that the standalone muon track should have:

- at least 1 muon hit
- at least 2 segments matched to the muon track.

As a consequence ϵ_{sa} is re-defined as the efficiency to reconstruct a standalone track and the two additional efficiencies $\epsilon_{sa} \rightarrow \epsilon_{sa} \times \epsilon_{\#MuonHits>0} \times \epsilon_{\#Matches>1}$. Events reconstructed as global muon pairs in which one of the two muon does not satisfying these two additional requirements fail in the $Z_{\mu t}$ category.

Figures 6.9 to 6.13 show the invariant mass distributions of the selected Z candidates for each category considered in the analysis for data and MC scaled to the data luminosity. Table 6.7 reports the number of selected candidates for data and MC signal and background in the $[60 - 120] \text{ GeV}/c^2$ mass range. The MC expected numbers of entries in each categories are scaled to the given data luminosity.

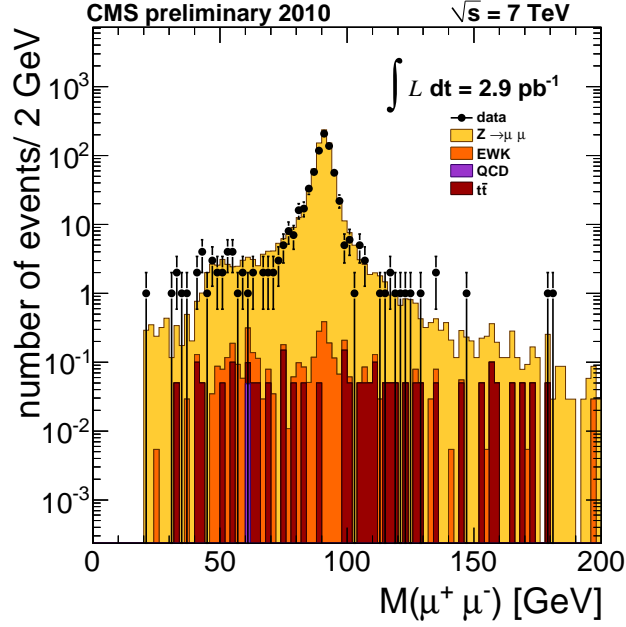


Figure 6.9: Invariant mass distribution of $Z_{\mu\mu}^{2HLT}$ candidates for data and MC signal and background events for a luminosity of 2.9 pb^{-1} .

| sample | $Z_{\mu\mu}$ | $Z_{\mu s}$ | $Z_{\mu t}$ | $Z_{\mu\mu}^{\text{non iso}}$ |
|-------------------------------------|-----------------|-------------------|-----------------|-------------------------------|
| $Z \rightarrow \mu^+ \mu^-$ | 950 ± 6 | 12.5 ± 0.6 | 50.2 ± 1.2 | 33.3 ± 1.0 |
| $W^\pm \rightarrow \mu^\pm \nu_\mu$ | 0.03 ± 0.03 | 0.23 ± 0.08 | 2.0 ± 0.2 | 0.55 ± 0.12 |
| $t\bar{t}$ | 1.3 ± 0.2 | 0.10 ± 0.06 | 1.5 ± 0.3 | 1.9 ± 0.3 |
| QCD | 0.05 ± 0.05 | 0.5 ± 0.2 | 0.7 ± 0.2 | 29.6 ± 1.2 |
| $Z \rightarrow \tau^+ \tau^-$ | 0.52 ± 0.12 | 0.03 ± 0.03 | 2.6 ± 0.3 | No events |
| WZ | 0.72 ± 0.06 | 0.011 ± 0.007 | 0.10 ± 0.02 | 0.06 ± 0.02 |
| WW | 0.32 ± 0.09 | No events | 0.38 ± 0.10 | No events |
| ZZ | 0.55 ± 0.13 | No events | 0.09 ± 0.05 | 0.06 ± 0.04 |
| data | 913 | 21 | 75 | 66 |

Table 6.7: Number of candidates expected (MC) and selected (data) in each category with an invariant mass in the range $[60-120] \text{ GeV}/c^2$ for 2.9 pb^{-1} in data and MC. Here $Z_{\mu\mu} = Z_{\mu\mu}^{1HLT} + Z_{\mu\mu}^{2HLT}$. For MC the separate contributions from signal and background processes are shown.

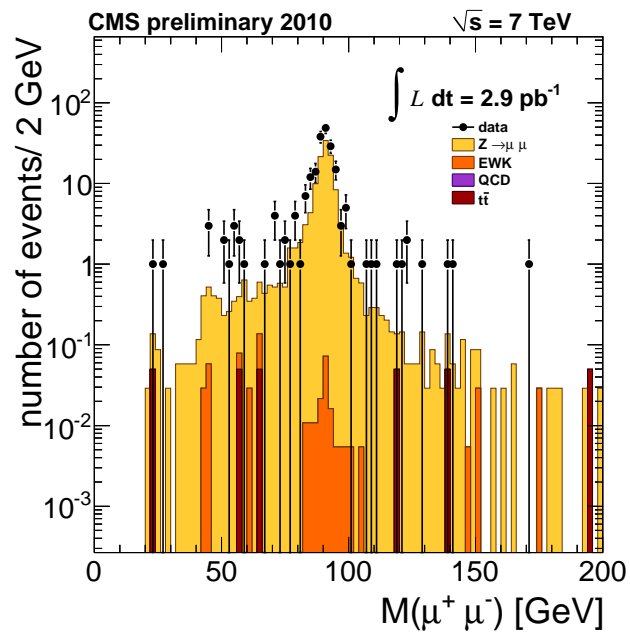


Figure 6.10: Invariant mass distribution of $Z_{\mu\mu}^{\text{HLT}}$ candidates for data and MC signal and background events for a luminosity of 2.9 pb^{-1} . The discrepancy one sees is due to the different ϵ_{HLT} in data and simulation.

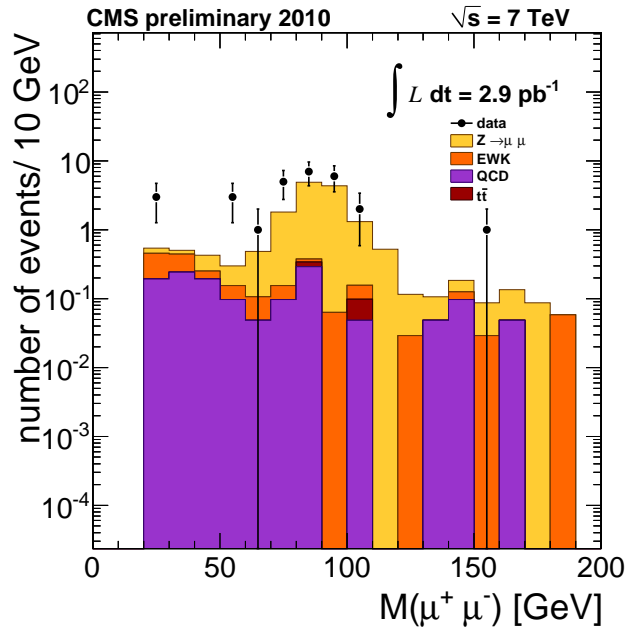


Figure 6.11: Invariant mass distribution of $Z_{\mu s}$ candidates for data and MC signal and background events for a luminosity of 2.9 pb^{-1} .

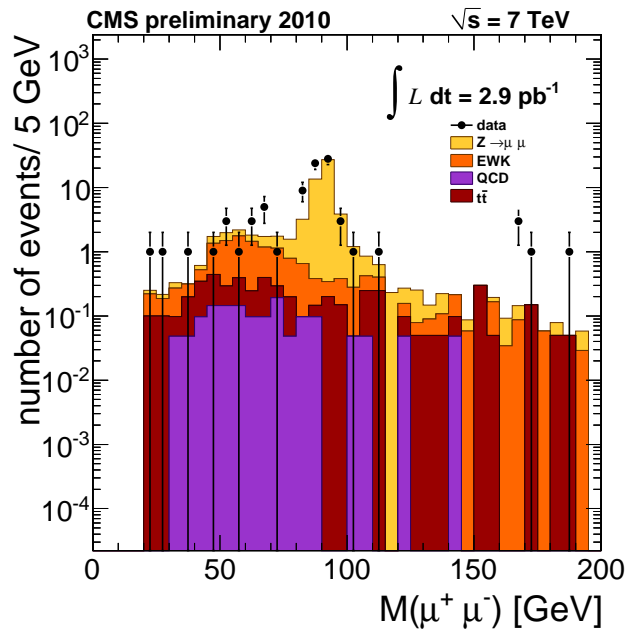


Figure 6.12: Invariant mass distribution of $Z_{\mu t}$ candidates for data and MC signal and background events for a luminosity of 2.9 pb^{-1} .

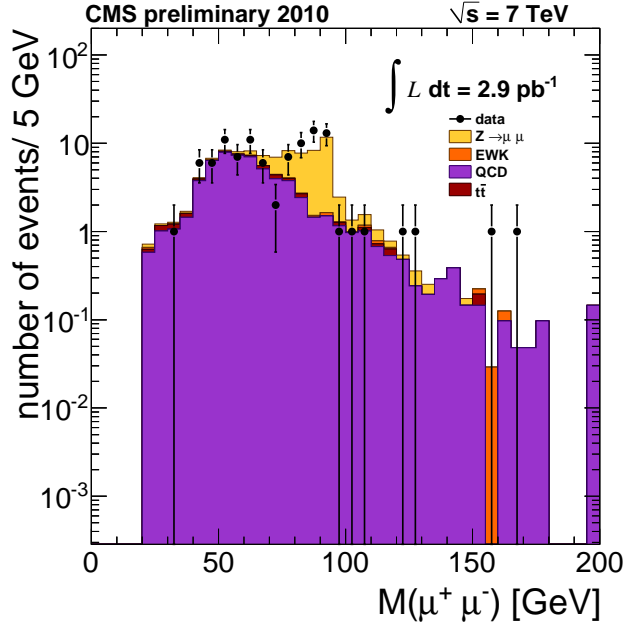


Figure 6.13: Invariant mass distribution of $Z_{\mu\mu}^{\text{non iso}}$ candidates for data and MC signal and background events for a luminosity of 2.9 pb^{-1} .

6.4.5 Fit results

We have performed the fit on the 2.9 pb^{-1} 7 TeV collision data. Figures 6.14-6.16 show the fit result superimposed to the histograms for the $Z_{\mu s}$, $Z_{\mu t}$, and $Z_{\mu\mu}^{\text{non iso}}$ samples. Table 6.9 reports the yield and efficiencies determined from the fit. They are compared to the results obtained by a fit performed using the the MC signal plus background only scaled to 2.9 pb^{-1} . In addition, the MC truth values of the average efficiencies, obtained from a sample of $Z \rightarrow \mu^+ \mu^-$ MC signal events, are also reported in the Table. The resulting χ^2 and correspondent p -value (i.e. the probability of obtaining a value greater that the obtained one, evaluated from the χ^2 distribution) assures the goodness of the fit procedure.

From the fit results and the comparison with the MC-truth, one can see that all the efficiencies found in data agree quite well with the expectation (even tough slightly lower) with the exception of the trigger efficiency which is significant lower in data. That has already be seen from the plots in Fig. 6.10 and 6.9, which shows that the events in the $Z_{\mu\mu}^{\text{1HLT}}$ are visible more than the expectation and the opposite for the $Z_{\mu\mu}^{\text{2HLT}}$ category. This extra trigger inefficiency in data is however well know in the CMS community and some more plots, reasons and discussion will be given in the following Section 6.4.10.

Given the low statistics, the polynomial degree of the background shape is trun-

cated to the first order, and the $Z_{\mu s}$ histogram, which contains only about 20 entries, is supposed to be background free (Table 6.7 shows that the background expected fraction is about 10%, not possible to subtract with such a low number of entries).

The correlation coefficients of the fit parameters is reported for completeness in Table 6.8. The first line of the matrix correlation table reports the correlations between the Z yield and the other fit parameter. No correlations greater than 13% are present.

An additional test is doing evaluating the Poissonian likelihood ratio variation from the minimum versus the yield value (after fixing all the efficiency term parameter in the fit). This variation is compared to an ideal Gaussian parabolic shape as a function of the Z yield, where the standard deviation of the Gaussian is fixed to one standard deviation error around the best fit minimum. The comparison in Fig. 6.17 demonstrates that we can safely consider the Gaussian approximation adequate to evaluate the error of the fitted $Z \rightarrow \mu^+ \mu^-$ yield.

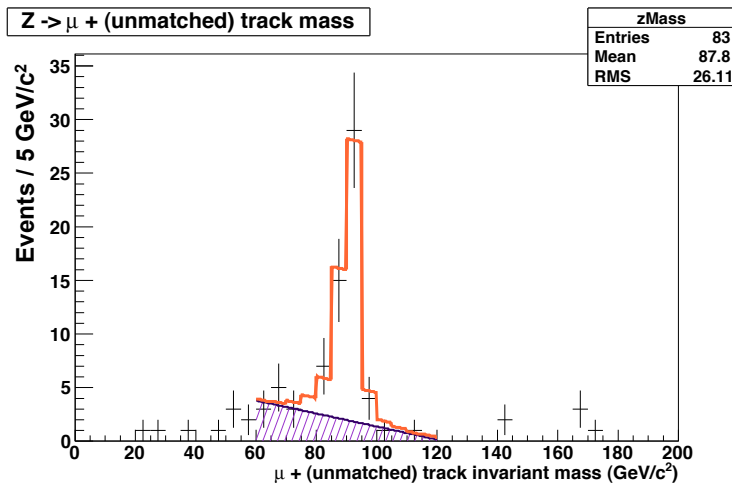


Figure 6.14: Fit curve superimposed to the invariant mass histogram of $Z_{\mu t}$ candidates for 2.9 pb^{-1} of LHC 7 TeV collision data.

6.4.6 Kinematic acceptance

Once the $Z \rightarrow \mu^+ \mu^-$ yield has been determined, we need to evaluate the kinematic acceptance at generator level in order to determine the cross section in an enlarged kinematic region.

The acceptance of the applied kinematic selection can be evaluated with Monte Carlo, and is somewhat sensitive to the generator adopted.

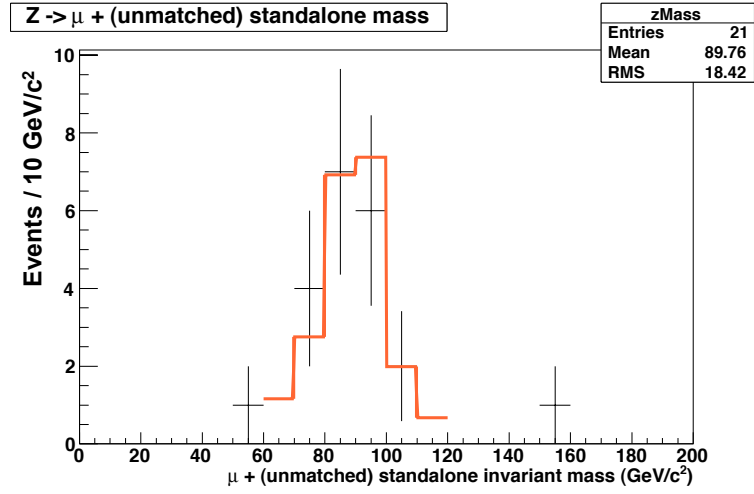


Figure 6.15: Fit curve superimposed to the invariant mass histogram of $Z_{\mu s}$ candidates for a sample corresponding to an integrated luminosity of 2.9 pb^{-1} of LHC 7 TeV collision data.

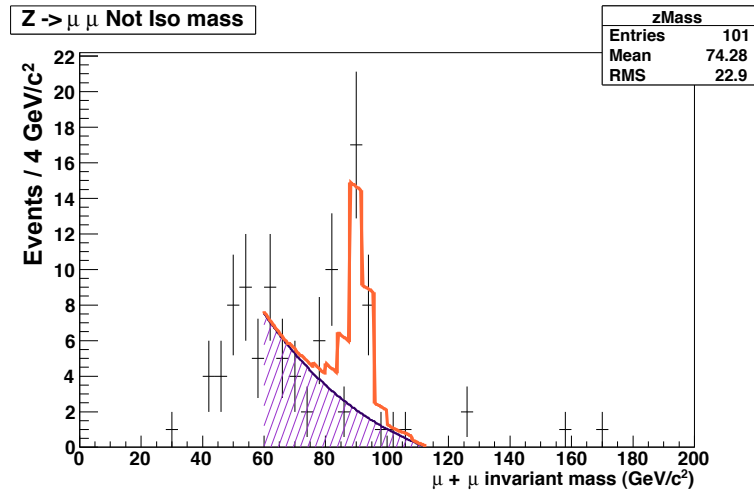


Figure 6.16: Fit curve superimposed to the invariant mass histogram of $Z_{\mu\mu}^{\text{non iso}}$ candidates for a sample corresponding to an integrated luminosity of 2.9 pb^{-1} of LHC 7 TeV collision data.

| Param. | Description | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-------------------------|---|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Yield | $Z \rightarrow \mu^+ \mu^-$ yield | 1.000 | -0.049 | -0.126 | -0.122 | -0.081 | -0.064 | -0.055 | 0.036 | 0.052 | -0.055 |
| ϵ_{trk} | tracking efficiency | -0.049 | 1.000 | 0.029 | 0.017 | 0.032 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| ϵ_{sa} | standalone efficiency | -0.126 | 0.029 | 1.000 | 0.0024 | 0.045 | 0.000 | 0.177 | -0.114 | 0.000 | 0.000 |
| ϵ_{iso} | isolation efficiency | -0.122 | 0.017 | 0.024 | 1.000 | 0.000 | -0.270 | 0.000 | 0.000 | -0.220 | 0.231 |
| ϵ_{HLT} | trigger efficiency | -0.081 | 0.032 | 0.045 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| α | background exponential slope | -0.064 | 0.000 | 0.000 | 0.270 | 0.000 | 1.000 | 0.000 | 0.000 | -0.957 | 0.932 |
| A_0 | $Z_{\mu t}$ bkg polyn. 1 st degree term | -0.055 | 0.000 | 0.177 | 0.000 | 0.000 | 0.000 | 1.000 | -0.978 | 0.000 | 0.000 |
| A_1 | $Z_{\mu t}$ bkg polyn. 2 nd degree term | 0.036 | 0.000 | -0.114 | 0.000 | 0.000 | 0.000 | -0.978 | 1.000 | 0.000 | 0.000 |
| B_0 | $Z_{\mu\mu}^{\text{non iso}}$ bkg polyn. 1 st degree term | -0.052 | 0.000 | 0.000 | -0.220 | 0.000 | -0.957 | 0.000 | 0.000 | 1.000 | -0.995 |
| B_1 | $Z_{\mu\mu}^{\text{non iso}}$ bkg polyn. 2 nd degree term | -0.055 | 0.000 | 0.000 | 0.231 | 0.000 | 0.932 | 0.000 | 0.000 | -0.995 | 1.000 |

Table 6.8: Correlation coefficients of the fit parameters from the simultaneous fit minimization

| $\int L dt = 2.9 \text{pb}^{-1}$ | fit results on data | simulation (2.9pb^{-1}) | MC-truth efficiencies |
|----------------------------------|---------------------|-------------------------------------|-----------------------|
| ϵ_{HLT} | 0.883 ± 0.008 | 0.930 ± 0.007 | 0.9319 ± 0.0014 |
| ϵ_{iso} | 0.985 ± 0.004 | 0.990 ± 0.004 | 0.9914 ± 0.0004 |
| ϵ_{sa} | 0.964 ± 0.004 | 0.971 ± 0.004 | 0.9724 ± 0.0006 |
| ϵ_{trk} | 0.994 ± 0.005 | 0.994 ± 0.004 | 0.9927 ± 0.0007 |
| $N_{Z \rightarrow \mu^+ \mu^-}$ | 1050 ± 35 | 1107 ± 37 | |
| χ^2/ndof | 1.074 | 1.034 | |
| p - value | 0.373 | 0.402 | |

Table 6.9: Comparison between fit parameters results with the fit model described in this chapter performed in data with simulation signal and background scaled to the data luminosity. χ^2/NDOF and p -value are also requested. MC-truth values of the average efficiencies are also shown for comparison.

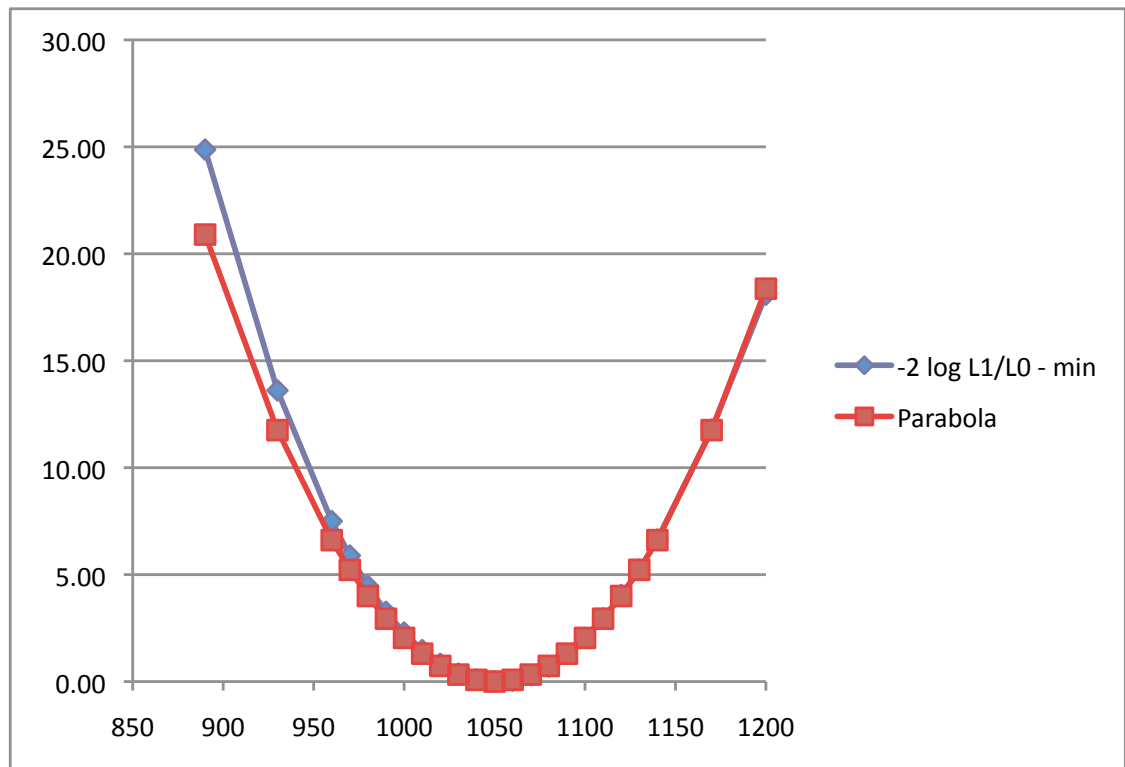


Figure 6.17: Poissonian likelihood ratio variation from the minimum versus the yield value compared with a Gaussian parabola.

Aiming to determine the cross section in the same mass region used in the fit, but without any p_T and η cut the acceptance we need to correct is:

$$\epsilon^{\text{kin}} = \frac{N(60 < m_Z < 120, p_{T,\mu} > 20, |\eta_\mu| < 2.1)}{N(60 < m_Z < 120)}, \quad (6.48)$$

where the numerator of the formula is the number of events considering muon after the final state radiation (because these are the muons we detect), while in the denominator the mass of the di-muon system is evaluated before the final state radiation. That is done in order to compare the final cross section number with a theoretical prediction which includes also the final state radiation in the calculation.

Running on the MC $Z \rightarrow \mu^+\mu^-$ sample used for the analysis I found for the geometric acceptance the value is:

$$\epsilon^{\text{kin}} = 0.3977 \pm 0.0017. \quad (6.49)$$

In order to use the Monte Carlo estimate, we need to verify that the acceptance estimated on generator particle and on reconstructed muons is identical and not affected too much from resolution effects. The good agreement has been verified in data with the statistics got so far and the relative systematics quoted as 0.2%.

6.4.7 Systematic Uncertainties

We can account for sources of systematics to the $Z \rightarrow \mu^+\mu^-$ cross section measurement the followings:

- The first source of systematic uncertainty is the LHC machine luminosity[84] estimate which amount to 11% and will be quoted separately in the final number.
- A second source is the theoretical uncertainty which affect the acceptance, due to the uncertainties on parton density function. On top of higher order QCD corrections, an attempt has been carried on to estimate the effect of Electroweak diagrams not fully implemented in our baseline MC: final state radiation and virtual and non-virtual corrections. A complete list of results can be found in Reference [7].
- A subtle source of systematics due to the online data taking, hence the L1 trigger, is the loss of muon events for trigger pre-firing, i.e. wrong assignment to the muon(s) bunch crossing number. This effect is due to wrong timing in the DT, CSC and RPC system, and is especially affecting the muon system overlap region ($0.9 < |\eta| < 1.2$) in which one should take care not only on the timing of the DT, CSC or RPC only, both also on the synchronization of

the three subsystems. Anyway the correction to be applied for those timing problem has been estimated for the current data to be at maximum $1.0 \pm 0.5\%$ for the $Z \rightarrow \mu^+\mu^-$ channel. This estimate has been obtained searching events in which a standalone-standalone pair of muons, peaking at the Z mass value is found: muon pre-firing indeed role out the tracker system from the muon measurements, resulting in the impossibility to built a L3 muon online and a global muon in the offline reconstruction. We just remark that it is a global scale factor to the cross section (to be enhanced by 1%, because we loss these 1% of events) and the 0.5% uncertainty is taken as systematic uncertainty.

- The remaining sources of systematics are then the one introduced with the fit procedure.

We build the probability functions for the signal in a fully data driven way. The same is not true for the background shape, built as polynomial of first degree times an exponential. We tried to change the background from first to second degree polynomial functions, then to fix to zero the slope of the exponential and then vary the binning. The fit yield value is very stable. So we took half the difference between the maximum and minimum fitted yield, and we assigned it as systematic error. We estimate a 0.7% uncertainty due to this effect.

The $Z_{\mu\mu}^{2\text{HLT}}$ and $Z_{\mu\mu}^{1\text{HLT}}$ histograms are supposed to be background free, but we know from MC that we have an irreducible contamination of less than 0.5% with the given selection. We added a flat background contribution to the two histograms and we saw that the fit yield output changes by an amount of one third of the background over signal ratio. So we can quote as 0.2% as a conservative estimate of systematics error due to neglecting the background in the $Z_{\mu\mu}^{2\text{HLT}}$ and $Z_{\mu\mu}^{1\text{HLT}}$ histograms.

Adding these two numbers in quadrature we can quote of 1% as a conservative estimate for the systematics induced by the fit strategy.

Table 6.10 below reports all the contribution to the systematics uncertainties to the the data driven $Z \rightarrow \mu^+\mu^-$ cross section measurement.

6.4.8 Cross section determination

At this point all the ingredients to finally evaluate the $Z \rightarrow \mu^+\mu^-$ cross section with the first 2.9 pb^{-1} of 7 TeV collision data are available. We start from the well known formula for the extraction of the cross section already introduced in the Equation 6.1.

As can be seen in Table 6.7 the expected contamination of irreducible background, f_{bkg} , in the $Z_{\mu\mu}^{2\text{HLT}}$ and $Z_{\mu\mu}^{1\text{HLT}}$ categories is $f_{bkg} = 0.37\%$ (3.5 ± 0.3 events on a total

Table 6.10: Table of systematic uncertainties for the simultaneous fit $Z \rightarrow \mu^+\mu^-$ cross section measurement.

| Source | % |
|---------------------------------------|------|
| Trigger firing | 0.5 |
| Muon momentum scale/resolution | 0.2 |
| Fit Background shape and subtraction | 1.0 |
| PDF uncertainty in acceptance | 1.2 |
| Other theoretical uncertainties | 1.6 |
| TOTAL(without luminosity uncertainty) | 2.3 |
| Luminosity | 11.0 |

of 950 expected signal events²).

The simultaneous fit strategy gives directly the Yield corrected by the efficiencies to select the $Z \rightarrow \mu^+\mu^-$ candidates, so we can evaluate the cross section as:

$$\sigma = \frac{Y(1 - f_{bkg})}{A \int \mathcal{L} dt} \quad (6.50)$$

A 1% correction due to the loss of muon events due to trigger pre-firing in the first era of CMS data taking is also applied (as explained in Section 6.4.7). The following cross sections for Z production is then measured:

$$\sigma(\mathbf{pp} \rightarrow \mathbf{ZX}) \times \mathbf{BF}(Z \rightarrow \mu^+\mu^-) = \mathbf{0.924} \pm \mathbf{0.031}(\mathbf{stat.}) \pm \mathbf{0.022}(\mathbf{syst.}) \pm \mathbf{0.101}(\mathbf{lumi.})\mathbf{nb}, \quad (6.51)$$

The reported Z cross section is limited to the invariant mass range $60 < m_{\mu^+\mu^-} < 120\text{GeV}/c^2$, and is corrected for the kinematic acceptance. The NNLO prediction for Z production is 0.97 ± 0.04 nb, so we find very good agreement with the Standard Model prediction.

²From the Table 6.7 also one can see that the QCD contribution to the background is very little w.r.t. other EWK channels and $t\bar{t}$. So we don't need to care for any possible increase of the QCD contribution w.r.t. to simulation where QCD (especially $b\bar{b}$) decay and punch-throughs in the muon system are not correctly simulated. This scale factor to apply to the QCD background has been anyway estimated to be not greater than 1.5. As we subtract the background in the other loose categories where the QCD is the main source of background, we don't need to consider this additional increment also for these histograms.

6.4.9 Correction factors for $W \rightarrow \mu\nu$ analysis

The $W \rightarrow \mu\nu$ selection requires the presence of an high p_T muon accompanied by a significant amount of missing transverse energy in the event.

For the first $W \rightarrow \mu\nu$ data driven measurement at CMS one needs to get the correction factor DATA/Monte Carlo (ρ_{eff}) for the efficiencies to reconstruct, trigger and identify a muon. That is done to correct the efficiencies taken directly from the simulation in a way similar to the one used in the first era of the $Z \rightarrow \mu^+\mu^-$ analysis, described in Section 6.3 . The correction value for the $W \rightarrow \mu\nu$ has been obtained by mean of the simultaneous fit to the $Z \rightarrow \mu^+\mu^-$ sample described in the thesis comparing the results obtained in data and simulation.

The muon selection in the $W \rightarrow \mu\nu$ analysis is equivalent to the one used for the $Z \rightarrow \mu^+\mu^-$ except for two additional requirements on the global muon track: the χ^2 of the fit to be less than 10 and the requirement for the global muon to be also a tracker muon. We will denote as ϵ_{sel} the efficiency for the two additional cuts.

With the $Z \rightarrow \mu^+\mu^-$ selected events we can measured ϵ_{sel} as a ratio of di-muons from global-global pairs in the mass range with and without those cuts applied to both muons (having applied first all the other cuts).

The measurements of ϵ_{HLT} , ϵ_{iso} , ϵ_{trk} , ϵ_{sa} on data and simulation has been already shown in table 6.9. Now we show the resulting correction factors and the numerical values of ϵ_{sel} . The total correction factor ρ_{eff} is the product of all the corrections. A summary of the results is given in Table 6.11.

| Efficiency | Data | Simulation | Data/Simulation(ρ_{eff}) |
|-------------------------|-------------------|---------------------|--|
| ϵ_{HLT} | 0.883 ± 0.008 | 0.9319 ± 0.0014 | 0.947 ± 0.009 |
| ϵ_{iso} | 0.985 ± 0.004 | 0.9914 ± 0.0004 | 0.994 ± 0.004 |
| ϵ_{sa} | 0.964 ± 0.004 | 0.9724 ± 0.0006 | 0.992 ± 0.005 |
| ϵ_{trk} | 0.994 ± 0.005 | 0.9927 ± 0.0007 | 0.998 ± 0.003 |
| ϵ_{sel} | 0.997 ± 0.003 | 0.9967 ± 0.0005 | 1.0 ± 0.003 |
| Net(W) | 0.828 ± 0.011 | 0.8874 ± 0.0013 | 0.933 ± 0.012 |

Table 6.11: Final efficiency factors used in the $W \rightarrow \mu\nu$ analysis. There were obtained applying the simultaneous fit technique on a clean $Z \rightarrow \mu^+\mu^-$ sample.

As can be seen in the table, most of the correction values are very close to unity. Only the trigger efficiency is about 5% lower in data than in simulation, a fact that has been confirmed many times in CMS and will be discussed in the next Section.

Additional studies performed with the $Z \rightarrow \mu^+\mu^-$ selection and analysis strategy essential for the $W \rightarrow \mu\nu$ analysis one are the following:

| Subset | Data/Simulation(Net ρ_{eff}) |
|-------------------------------------|---|
| positive muons | 0.935 ± 0.018 |
| negative muons | 0.931 ± 0.019 |
| barrel ($ \eta < 0.9$) | 0.955 ± 0.024 |
| transition ($0.9 < \eta < 1.2$) | 0.89 ± 0.04 |
| endcap ($1.2 < \eta < 2.1$) | 0.92 ± 0.03 |

Table 6.12: Correction factors for subsets of muons

- We evaluated the correction factors for positive and negative muons separately, as shown in Table 6.12. The difference amounts to 0.4% and is compatible with zero. Hence the same correction can be applied for the $W^+ \rightarrow \mu^+\nu$ and $W^- \rightarrow \mu^-\nu$ and in the ratio W^+/W^- .
- The table also lists the correction values for the barrel, endcap and transition regions of the muon system. The correction factors are fairly uniform with perhaps a somewhat lower value in the transition, or overlap region between the DTs and the CSCs.
- Another important result found is the global correlation between the product of the efficiencies for the muon selection and the $Z \rightarrow \mu^+\mu^-$ yield. This number is used for the error propagation in the W over Z ratio, which is one of the most important EWK measurement. This value has been obtained from the simultaneous fit described above reparametrizing one of the efficiency as $\epsilon_{\text{tot}} = \epsilon_{\text{HLT}}\epsilon_{\text{trk}}\epsilon_{\text{sa}}\epsilon_{\text{iso}}$, and introducing this parameter in the fit minimization. The total correlation found is

$$\text{correlation}(\text{Yield}, \epsilon_{\text{tot}}) = -0.236 \quad (6.52)$$

6.4.10 Trigger efficiency estimate using $Z \rightarrow \mu^+ \mu^-$

Di-muon resonance such as J/Ψ , $\Upsilon(1s)$, Z are used to study the detector reconstruction, isolation and trigger efficiencies using the so called ‘‘Tag and Probe’’ method. This method, which has been successfully used in some form or another by past and present experiments, relies upon $Z \rightarrow \mu^+ \mu^-$ decays to provide an unbiased, high-purity, muon sample with which to measure the efficiency of a particular selection cut, including the trigger. With the intention to study the trigger efficiency, a single muon trigger sample is used, from which a subset of di-muon events are selected. One of the muons, the ‘‘tag’’, is required to pass stringent muon identification criteria in order to have as low background as possible, while the other muon, the ‘‘probe’’, is only required to pass a set of identification criteria depending on the efficiency under study. The invariant mass of the tag and probe muon candidates is required to be within a window around m_Z . The tight criteria imposed on the tag coupled to the invariant mass requirement is sufficient to ensure high muon purity.

This method has been applied to study the muon trigger efficiency, to study the single muon trigger selection path named HLT_Mu9. Both tag and prob muons are required to pass stringent identification and isolation cuts (see Section 6.3.2), in order to provide the efficiency to be used for the $W \rightarrow \mu\nu$ analysis. Hence the probe muon is required to pass these tight identification and isolation cuts and the trigger firing is investigated.

We find for the average efficiency for $p_T > 20$ GeV/c and $|\eta| < 2.1$ in the first 2.9 pb^{-1} of collision data the value reported in Table 6.13 and in the plots of Fig. 6.18 and 6.19. For comparison we report also the efficiency we expect from the simulation, using the MC $Z \rightarrow \mu^+ \mu^-$ simulated sample used in the thesis. As already remarked the correction we found to be applied to pass from simulated MC to data efficiency is of the order of 95%, to be precise 0.947 ± 0.008 for these tight selected muons.

The CMS community is investigating the reason of this 5% of extra inefficiency in data. As can be seen by the numbers and plots it is clear that the region much affected by this trigger loss is the barrel-endcap overlap region ($0.9 < |\eta| < 1.2$), due to synchronization problem between DT, CSC and RPC detectors. It has been found that all these subsystem need to better calibrate their timing constants. That calibration can be done when more collision data will be available and so it is expected that the L1 trigger efficiency will increase in the future collision. See Section 6.7 for an update of the measurement with about ten times data, in which a net improvement in the trigger performance is already visible.

In addition a second loss of efficiency has been found to be introduced in the L2 and L3 step, especially in the endcaps, due to a probability some per cent higher than expected to assign a wrong charge to the muon in the L2-L3 steps. For that reason new L3 algorithms will be soon deployed online.

Table 6.13: Table of muon single muon trigger efficiency versus η obtained in data using the tag and prob method on $Z \rightarrow \mu^+\mu^-$ selected event.

| | overall ($ \eta < 2.1$) | barrel ($ \eta < 0.9$) | transition region ($0.9 < \eta < 1.2$) | endcaps ($1.2 < \eta < 2.1$) |
|--------------------------------|-------------------------------|------------------------------|---|-------------------------------------|
| $Z \rightarrow \mu^+\mu^-$ MC | 0.9333 ± 0.0010 | 0.9607 ± 0.0012 | 0.871 ± 0.003 | 0.920 ± 0.002 |
| data (2.9 pb^{-1}) | 0.884 ± 0.007 | 0.934 ± 0.009 | 0.75 ± 0.03 | 0.871 ± 0.014 |
| scale factor (data/MC) | 0.947 ± 0.008 | 0.972 ± 0.010 | 0.86 ± 0.04 | 0.947 ± 0.017 |

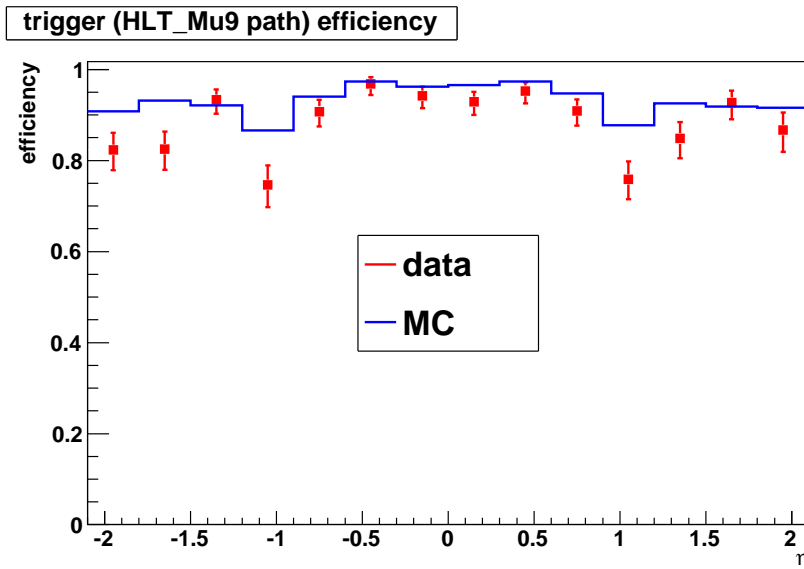


Figure 6.18: Muon single muon trigger efficiency versus muon η , obtained using tag and prob method on $Z \rightarrow \mu^+\mu^-$ selected events on 2.9 pb^{-1} of LHC 7 TeV collision data.

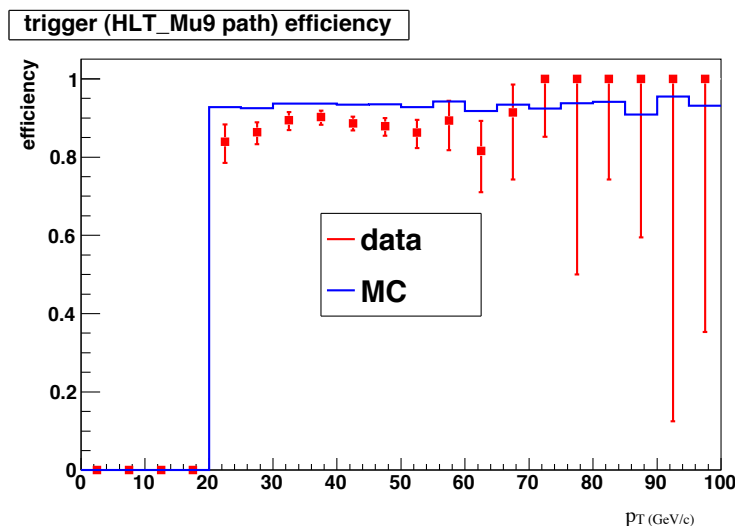


Figure 6.19: Muon single muon trigger efficiency versus muon p_T obtained using tag and prob method on $Z \rightarrow \mu^+\mu^-$ selected events on 2.9 pb^{-1} of LHC 7 TeV collision data.

6.5 Luminosity measurement using W and Z

The machine luminosity monitor and measurement is one of the most important thing to assess in the first years of data taking of a new accelerator and through-tout its whole life.

The integrated luminosity in CMS is based on signals from the Forward Hadronic Calorimeter, HF (see Section 2.2.4). Two methods for extracting a real-time relative instantaneous luminosity are used. The “zero counting” method in which the average fraction of empty towers is used to infer the mean number of interactions per bunch crossing. The second method exploits the linear relationship between the average transverse energy per tower and the luminosity. The choice of the online algorithm chosen to report the luminosity summary might get adjusted based on the data integrity/reliability, namely the statistical uncertainty or systematics uncertainty due to the Forward Calorimeter or the background sensitivity. The different algorithm agree to within 5% so far. The final normalization of the luminosity is based on Van der Meer scans, which determine the size of the colliding beams and thus the luminosity with minimal reliance on simulation. This technique was first pioneered

by Van Der Meer at the ISR. See reference [4], article to which I contributed, for a review of the luminosity measurement issue at LHC, and Reference [84] for a details on the CMS measurement systematic uncertainties.

The resulting measurement is the most reliable one and the recommended to be used for data analysis and is stored run by run (to be more precise lumisection by lumisections, where the lumisections is the smallest part of a CMS run lasting 23 s) in the CMS Condition Database and is accessible by the whole CMS community.

As already said the current precision of this method, resulting in a systematics for all cross section measurement is 11%. The uncertainty associated to this measurement is supposed to lower with the passing of the time, when all the LHC fill beam information will be calibrated and measured always better, but is eventually limited by the uncertainty on the knowledge of all the QCD processes contributing to the very forward physics.

Several offline method have been used in the first era of the LHC: above all the primary vertexes counting. The method counts the primary vertexes falling in each lumisections and knowing the expectation for the theory manage to obtain a luminosity value for the lumisections. As the online HF method also the vertex counting is dominated finally by the QCD physics uncertainties.

With high integrated luminosity we can exploit EWK channels, once the Standard Model cross section value will be definitely confirmed, to use them as luminosity monitor and eventually to measure the luminosity with low systematics. Muon decays are for sure more feasible for such a use, because in general CMS reaches better muon reconstruction and identification than electrons. As already seen in the previous Section already with about 3 pb^{-1} we have a measurement of the $Z \rightarrow \mu^+\mu^-$ cross section in which the systematical error is comparable with the statistical error, and it is negligible for the $W \rightarrow \mu\nu$ channel, with has about ten times greater cross section.

Already in the future year of the machine, when the instantaneous luminosity will reach the value of $L = 10^{33} \text{ s}^{-1} \text{ cm}^{-2}$ we will collect in less then 10 hours (hence the time interval a CMS run typically lasts) an integrated luminosity of the order of the pb^{-1} , and when the luminosity will reach the design value of $L = 10^{34} \text{ s}^{-1} \text{ cm}^{-2}$ W and Z bosons decaying into muons will be produced with the rate of 10 and 1 Hz respectively (see also Table 1.4).

Given that, we found very useful to develop a tool, inserted in the offline DQM framework (see Section 3.5), that counts the number of selected $W \rightarrow \mu\nu$ and $Z \rightarrow \mu^+\mu^-$ events and publishes them on a web page.

The software program is run centrally with the complete DQM offline chain producing the histogram containing the $Z \rightarrow \mu^+\mu^-$ mass and $W \rightarrow \mu\nu$ transverse mass distribution. A subsequent script developed also by myself access the web repository where this histograms are saved, counts the entries and applies the data driven method described in the previous Section and determine the integrated luminosity (assuming the Standard Model cross section value).

This tool has been put in place already with the first runs and shows to agree well with the online measurements. We wait a larger instantaneous luminosity to better exploit its utility with the resulting small uncertainties.

A snapshot of the $W \rightarrow \mu\nu$ and $Z \rightarrow \mu^+\mu^-$ counting web page is shown in Figure 6.20.

Luminosity expectation counting W and Z into muons (using dqm minimum bias(may27rerec run2010/Mu prompt-reco), assuming B/S = 10% for W and 1% for Z; disclaimer:all numbers

| run | LumiFromWmn(nW*0.24) 1 / nb | LumiFromZmm(nZ*3.0) 1 / nb |
|--------|-----------------------------|----------------------------|
| 143827 | 140.16 | 144.0 |
| 143831 | 12.48 | 6.0 |
| 143833 | 125.04 | 147.0 |
| 143835 | 30.48 | 24.0 |
| 143953 | 106.32 | 123.0 |
| 143954 | 31.92 | 33.0 |
| 143955 | 14.16 | 6.0 |
| 143956 | 7.68 | 3.0 |
| 143957 | 29.04 | 18.0 |
| 143959 | 17.28 | 33.0 |
| 143960 | 13.68 | 24.0 |
| 143961 | 52.32 | 54.0 |
| 143962 | 52.56 | 51.0 |
| 143977 | 0.0 | 0.0 |
| 144010 | 19.44 | 27.0 |
| 144011 | 78.24 | 78.0 |
| 144083 | 0.0 | 0.0 |
| 144086 | 64.56 | 63.0 |
| 144089 | 247.92 | 294.0 |
| 144112 | 280.8 | 285.0 |
| 144114 | 9.84 | 12.0 |
| tot | 3225.36+/-27.8 | 3312.0+/-99.7 |

Figure 6.20: Snapshot of the web page produced with the offline luminosity measurement tool using W and Z bosons decaying into muons. The numbers are in nb^{-1} .

6.6 Overview of all CMS EWK results in 2010

In this section we report a list of all the results regarding the W and Z boson measurements performed by CMS in 2010[1]. The $Z \rightarrow \mu^+\mu^-$ cross section has been measured with the method reported in the this thesis. Many inputs for the $W \rightarrow \mu\nu$ cross section measurements and W over Z ratio has been obtained using the same method: the muon reconstruction, isolation and trigger efficiency, separately for all muons, for positive and negative muons, the correlation with the $Z \rightarrow \mu^+\mu^-$ yield respectively. The measurements on the muon channel are compared with the electron ones (which are less precise at this point of the CMS operation).

With the dataset corresponding to 2.9 pb^{-1} the following cross sections for inclusive W production has been obtained:

$$\begin{aligned}\sigma(pp \rightarrow WX) \times BF(W \rightarrow e\nu) &= 9.801 \pm 0.112(\text{stat.}) \pm 0.495(\text{syst.}) \pm 1.078(\text{lumi.})\text{nb}, \\ \sigma(pp \rightarrow WX) \times BF(W \rightarrow \mu\nu) &= 9.969 \pm 0.090(\text{stat.}) \pm 0.309(\text{syst.}) \pm 1.097(\text{lumi.})\text{nb}, \\ \sigma(pp \rightarrow WX) \times BF(W \rightarrow l\nu) &= 9.903 \pm 0.070(\text{stat.}) \pm 0.289(\text{syst.}) \pm 1.089(\text{lumi.})\text{nb}.\end{aligned}$$

The NNLO prediction is $10.44 \pm 0.52 \text{ nb}$. The results for charge-specific W production are:

$$\begin{aligned}\sigma(pp \rightarrow W^+X) \times BF(W^+ \rightarrow e^+\nu^-) &= 5.802 \pm 0.079(\text{stat.}) \pm 0.306(\text{syst.}) \pm 0.638(\text{lumi.})\text{nb}, \\ \sigma(pp \rightarrow W^+X) \times BF(W^+ \rightarrow \mu^+\nu^-) &= 5.868 \pm 0.068(\text{stat.}) \pm 0.188(\text{syst.}) \pm 0.646(\text{lumi.})\text{nb}, \\ \sigma(pp \rightarrow W^+X) \times BF(W^+ \rightarrow l^+\nu^-) &= 5.840 \pm 0.051(\text{stat.}) \pm 0.182(\text{syst.}) \pm 0.642(\text{lumi.})\text{nb};\end{aligned}$$

and:

$$\begin{aligned}\sigma(pp \rightarrow W^-X) \times BF(W^- \rightarrow e^-\nu^+) &= 3.945 \pm 0.067(\text{stat.}) \pm 0.224(\text{syst.}) \pm 0.434(\text{lumi.})\text{nb}, \\ \sigma(pp \rightarrow W^-X) \times BF(W^- \rightarrow \mu^-\nu^+) &= 4.101 \pm 0.059(\text{stat.}) \pm 0.140(\text{syst.}) \pm 0.451(\text{lumi.})\text{nb}, \\ \sigma(pp \rightarrow W^-X) \times BF(W^- \rightarrow l^-\nu^+) &= 4.034 \pm 0.044(\text{stat.}) \pm 0.137(\text{syst.}) \pm 0.444(\text{lumi.})\text{nb}.\end{aligned}$$

The NNLO predictions for these cross sections are $6.15 \pm 0.29 \text{ nb}$ for W^+ and $4.29 \pm 0.23 \text{ nb}$ for W^- .

The following cross sections for Z production have been measured:

$$\begin{aligned}\sigma(pp \rightarrow ZX) \times BF(Z \rightarrow e^+e^-) &= 0.960 \pm 0.037(\text{stat.}) \pm 0.059(\text{syst.}) \pm 0.105(\text{lumi.})\text{nb}, \\ \sigma(pp \rightarrow ZX) \times BF(Z \rightarrow \mu^+\mu^-) &= 0.924 \pm 0.031(\text{stat.}) \pm 0.022(\text{syst.}) \pm 0.102(\text{lumi.})\text{nb}, \\ \sigma(pp \rightarrow ZX) \times BF(Z \rightarrow l^+l^-) &= 0.931 \pm 0.026(\text{stat.}) \pm 0.023(\text{syst.}) \pm 0.103(\text{lumi.})\text{nb}.\end{aligned}$$

The reported Z cross sections pertain to the invariant mass range $60 < m_{l+l^-} < 120 \text{ GeV}$, and are corrected for the kinematic acceptance but not for γ^* exchange. The NNLO prediction for Z production is $0.97 \pm 0.04 \text{ nb}$.

Figure 6.21 shows the comparison shows the EWK CMS measurements together with measurements at lower-energy hadron colliders with about 2.9 pb^{-1} of 7 TeV proton-proton collision data, while Figure 6.22 reports the detailed comparison for the Z cross section measurements with the theory.

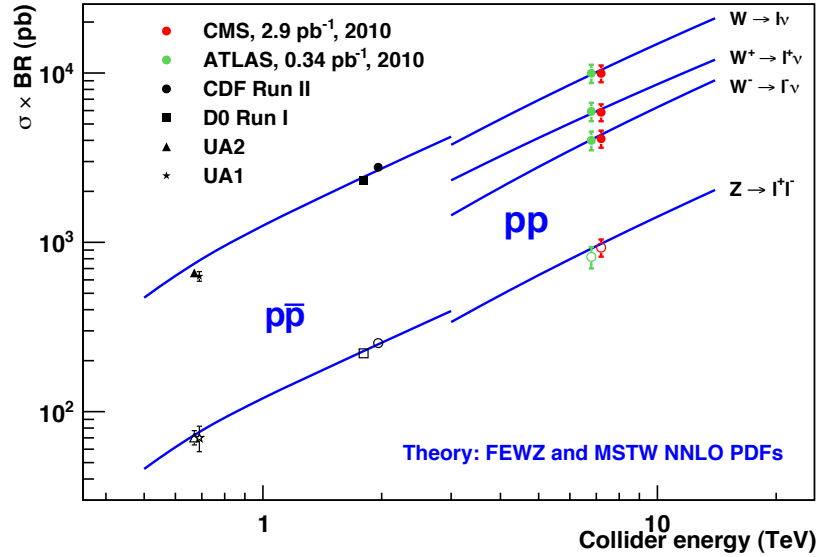


Figure 6.21: Measurements of inclusive cross sections from CMS and experiments at lower-energy colliders. The plain red symbols represent $\sigma(W \rightarrow l\nu) \times BF(W \rightarrow l\nu)$ and the empty red symbols $\sigma(Z(\gamma^*) \rightarrow l^+l^-) \times BF(Z \rightarrow l^+l^-)$.

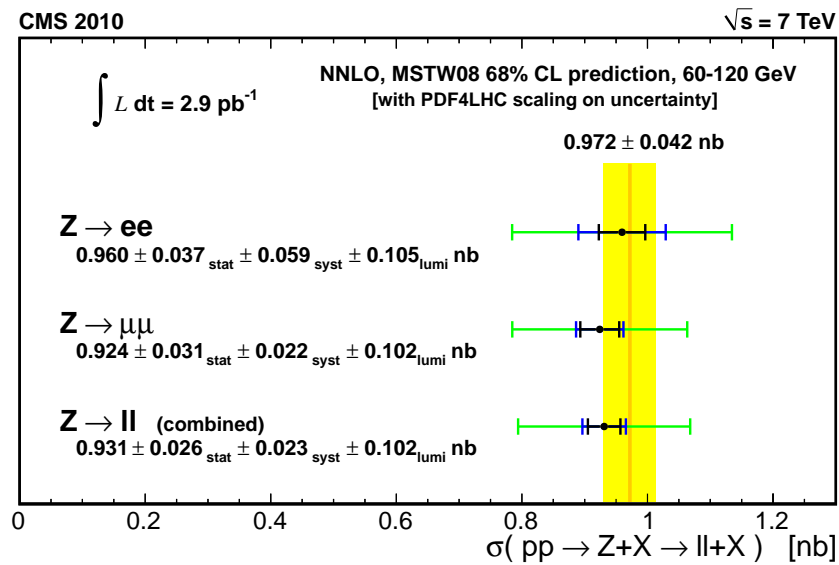


Figure 6.22: Comparison with the theory for the CMS inclusive Z cross section measurements.

Combination of electron and muon results are evaluated by calculating an average value weighted by the statistical uncertainties (i.e., by $1/\sigma_{stat}^2$). The systematic uncertainties are taken into account numerically, by shifting central values of yields, acceptances, and efficiencies by the assigned systematic uncertainty. The luminosity uncertainty cancels exactly in these ratios. When uncertainties are correlated, the shifts are also correlated.

Figure 6.23 reports the ratio between the CMS EWK inclusive measurements with the theory, while Figure 6.24 reports the comparison for the Z cross section measurements with the ATLAS measurements[88]. The CMS measurement shown at ICHEP 2010 and in the paper [1] are compared with the one reported by the ATLAS collaboration.

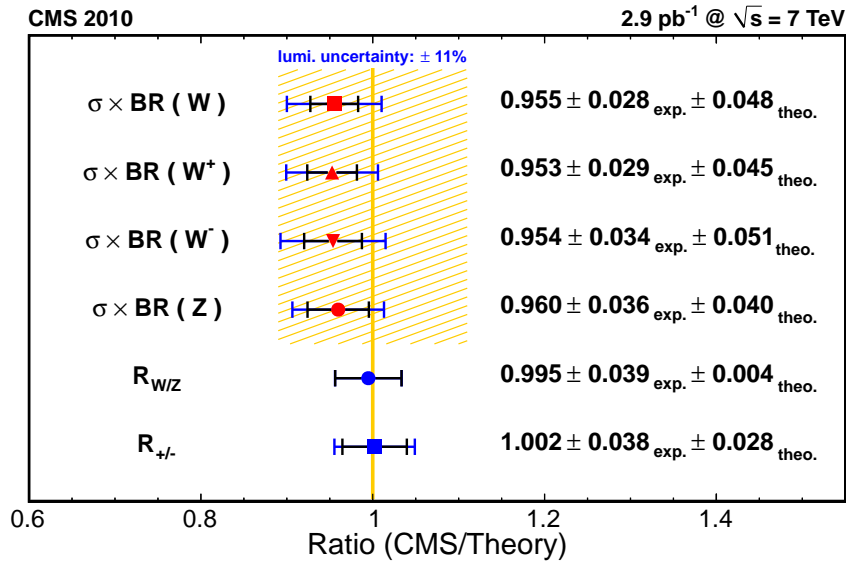


Figure 6.23: Comparison with the theory for the the CMS inclusive Z and W cross section measurements.

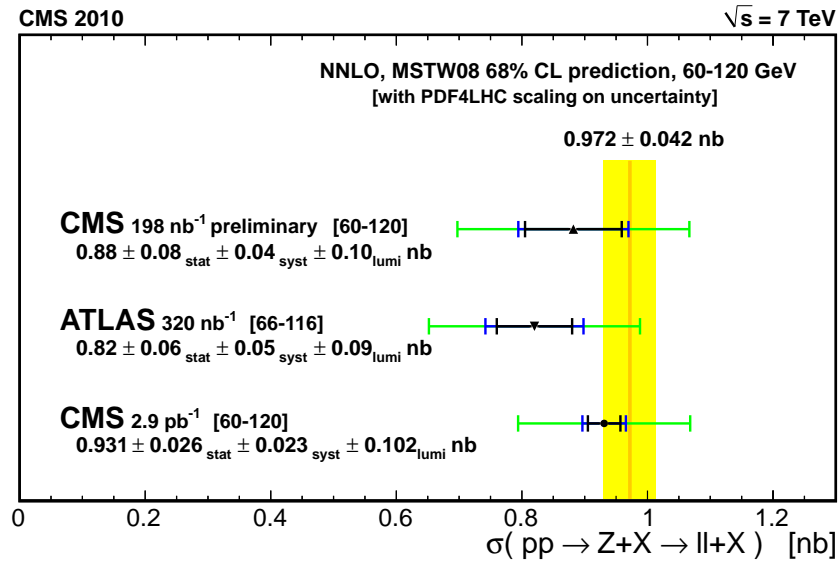


Figure 6.24: Comparison of the CMS and ATLAS measurements reported in 2010 7 TeV collision data.

6.7 Update of $Z \rightarrow \mu^+\mu^-$ results with the entire 2010 proton-proton dataset

All the results presented in this last Section should be considered preliminary, due to the very short time occurring from the data availability and the delivery dead-line of this thesis.

LHC provided proton-proton collision to CMS and the other experiments till the beginning of November 2010 when the heavy ion program started. The entire 2010 proton collision dataset has been analyzed and preliminary results has been obtained. The same selection and the same systematics errors has been assumed. The amount of collision data analyzed accounts to 35 pb⁻¹.

The selection applied to measure the $Z \rightarrow \mu^+\mu^-$ cross section with the full 2010 amount of data is the same used for the measurement with 2.9 pb⁻¹, except with two modifications needed for the muon trigger and isolation selection:

- With the progressive increase of the instantaneous luminosity up to $2 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$ the single muon trigger rate based on a 9 GeV cut was not anymore sustainable. Hence the threshold for the highest unrescaled single muon trigger has been moved from 9 to 11 and finally to 15 GeV. As we use an offline cut to 20 GeV for selecting the muon to reconstruct the $Z \rightarrow \mu^+\mu^-$ candidates, in principle we should not be sensible to the different behaviours of these three

different triggers. Of course a detailed study on that is needed, which is neglected here for lack of time. In addition, as already said in Section 6.4.10 new development on L1 and L3 muon trigger has been introduced in the last month of CMS proton-proton operation and trigger performance are expected to improve. In Table 6.14 the results for average trigger muon efficiency ($|\eta| < 2.1$, $p_T > 20$ GeV/c) with the $Z \rightarrow \mu^+ \mu^-$ selection are reported dividing the data-taking period according to the high threshold unprescaled muon trigger available online. The data-taking is divided in RunA and RunB according to the software version used to take and reconstruct the data (note that RunA coincides approximately with the first 2.9 pb^{-1} used for the previous measurement). The trigger efficiency is obtained with the $Z \rightarrow \mu^+ \mu^-$ “tag and probe”. From the Table one can see that indeed the trigger muon efficiencies raised in RunB, i.e. in the last part of data-taking. As also expected, the performance of the high threshold muon trigger in RunB are compatible, after applying an offline cut of 20 GeV/c to the muon p_T .

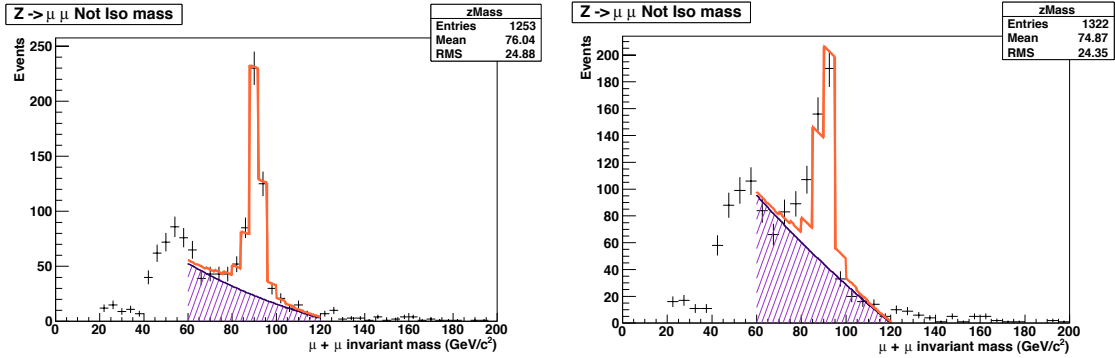
- The isolation cut used for the previous results is the relative combined isolation cut. With more data a disagreement is observed between the template shape taken for the $Z \rightarrow \mu^+ \mu^-$ “golden” categories and the $N_{\mu\mu}^{\text{non iso}}$ one. The reason for that has been found on the photon from final state radiation, which may give raise to high deposit in the ECAL around the muon. The $N_{\mu\mu}^{\text{non iso}}$ sample, when combined isolation is applied, results to be enriched by these type of events, in which the di-muon mass is lower that the peak value of 91 GeV. That feature is not present with tracker only isolation. Figure 6.25 below reports graphically the fit with the data driven template from $Z \rightarrow \mu^+ \mu^-$ to the tracker and combined $N_{\mu\mu}^{\text{non iso}}$ sample with the full 2010 data. The isolation cut applied are $I_{rel}^{comb} = \frac{\sum_{\Delta R < 0.3} I_{trk} + I_{Ecal} + I_{HCal}}{p_t} < 0.15$ and $I_{trk} < 3.0$ GeV/c respectively. The disagreement is cured taking the template for the fit from the simulation, but we prefer not to consider that possibility for baseline selection of the $Z \rightarrow \mu^+ \mu^-$ cross section in order to minimize any systematic error source introduced from the simulation.

We use the same selection used for the 2.9 pb^{-1} results except for the trigger and the tracker only isolation cut, which permits to use the fully data driven strategy for the simultaneous determination of yield and efficiencies. Figures 6.26 below reports the mass plots in linear and logarithmic scale for $Z \rightarrow \mu^+ \mu^-$ “golden” sample obtained adding the $Z_{\mu\mu}^{\text{HLT1}}$ and $Z_{\mu\mu}^{\text{HLT2}}$ categories. A can be seen the shape obtained in data matches well with the simulation.

We have performed the fit on the 35 pb^{-1} 7 TeV collision data. Figures 6.27-6.29 show the fit result superimposed to the histograms for the $Z_{\mu s}$, $Z_{\mu t}$, and $Z_{\mu\mu}^{\text{non iso}}$ samples. Table 6.15 reports the yield and efficiencies determined from the fit. In addition,

Table 6.14: Table of average single muon trigger efficiency versus integrated luminosity. The highest unscaled muon trigger is reported in each period.

| period | integrated luminosity (pb^{-1}) | trigger | ϵ_{HLT} | data/simulation |
|--------|--|----------|-------------------------|-------------------|
| RunA | 3.07 | HLT_Mu9 | 0.876 ± 0.008 | 0.942 ± 0.009 |
| RunB | 4.41 | HLT_Mu9 | 0.919 ± 0.005 | 0.966 ± 0.006 |
| RunB | 9.47 | HLT_Mu11 | 0.925 ± 0.003 | 0.972 ± 0.004 |
| RunB | 17.91 | HLT_Mu15 | 0.926 ± 0.002 | 0.973 ± 0.003 |

**Figure 6.25:** fit with the data driven template from $Z \rightarrow \mu^+\mu^-$ to non isolated sample with the full 2010 data. Left: tracker only isolation case. Right: relative combined isolation case. Clear disagreement below the mass peak is visible for the combined isolation plot.

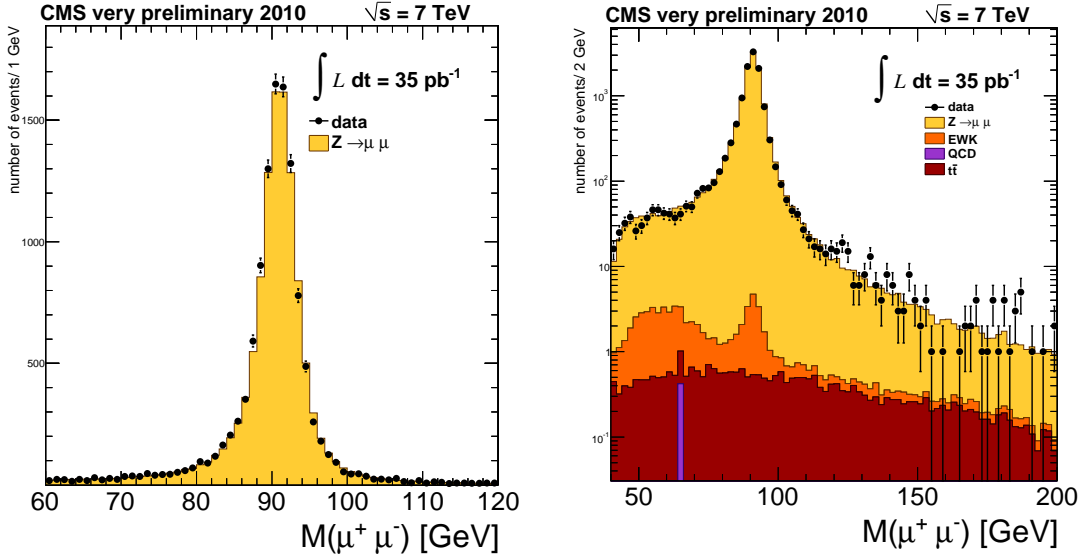


Figure 6.26: Invariant mass distribution of $Z \rightarrow \mu^+\mu^-$ golden candidates for data and MC signal and background events for a luminosity of 35 pb^{-1} . Left: linear scale. Right: logarithmic scale.

the MC truth values of the average efficiencies and the correction data/simulation factors are reported, obtained from a sample of $Z \rightarrow \mu^+\mu^-$ MC signal events, are also reported in the Table. The resulting fit χ^2 and correspondent p -value (i.e. the probability of obtaining a value greater than the obtained one, evaluated from the χ^2 distribution) assures the goodness of the fit procedure.

From the fit results and the comparison with the MC-truth, one can see that all the efficiencies found in data agree quite well with the expectation (even though slightly lower) with the exception of the trigger efficiency which is still lower in data but less than with the 2.9 pb^{-1} result reported in Table 6.15.

With the given statistics, a second degree polynomial function is taken for modelling the background shape.

Assuming the same systematics sources as the past studies, the following cross sections for Z production is measured with 2010 data:

$$\sigma(\text{pp} \rightarrow \text{ZX}) \times \text{BF}(Z \rightarrow \mu^+\mu^-) = 0.954 \pm 0.009(\text{stat.}) \pm 0.022(\text{syst.}) \pm 0.105(\text{lumi.})\text{nb}, \quad (6.53)$$

The value reported is in agreement both with the NNLO prediction for Z production and the previous measurements. The uncertainty on the measurement is now dominated by the systematics error, rather than the statistical. Given that the systematic error has been reduced already to the minimum with the use of data-driven methods, that means that we can think to perform the ultimate measurements of

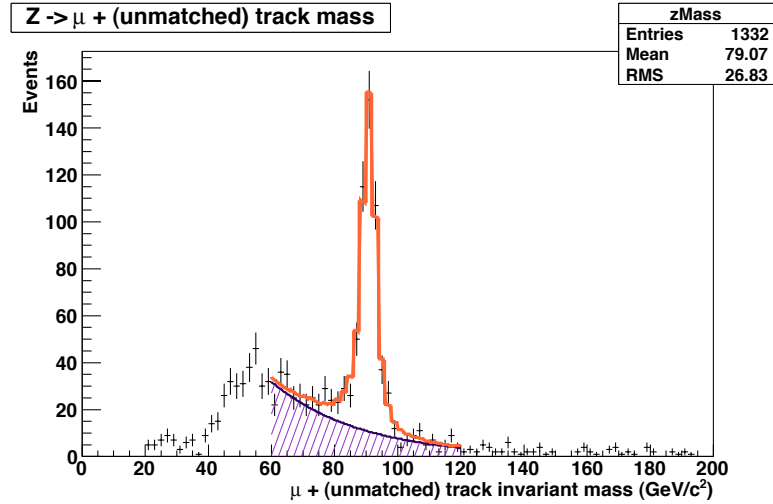


Figure 6.27: Fit curve superimposed to the invariant mass histogram of $Z_{\mu t}$ candidates for 35 pb^{-1} of LHC 7 TeV collision data.

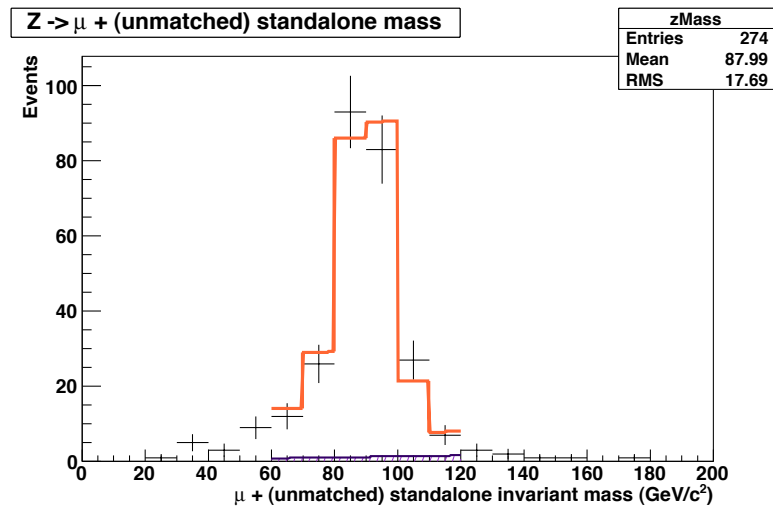


Figure 6.28: Fit curve superimposed to the invariant mass histogram of $Z_{\mu s}$ candidates for a sample corresponding to an integrated luminosity of 35 pb^{-1} of LHC 7 TeV collision data.

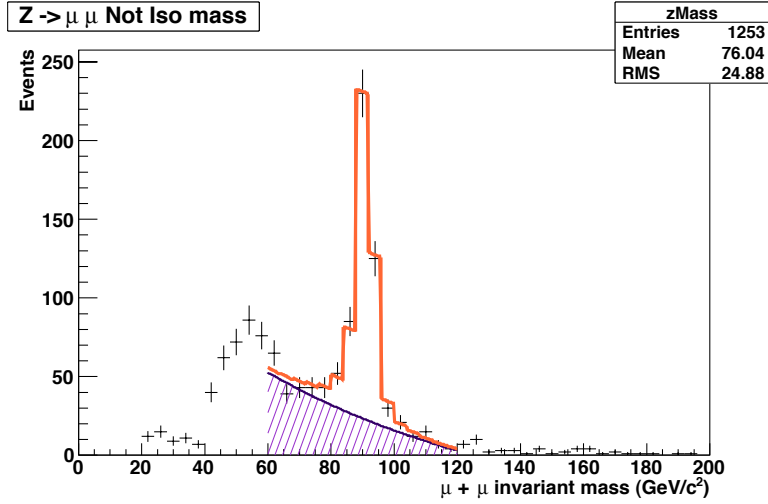


Figure 6.29: Fit curve superimposed to the invariant mass histogram of $Z^{\text{non iso}}_{\mu\mu}$ candidates for a sample corresponding to an integrated luminosity of 35 pb^{-1} of LHC 7 TeV collision data.

| $\int Ldt = 35\text{pb}^{-1}$ | fit results on data | data/simulation |
|---------------------------------|---------------------|---------------------|
| ϵ_{HLT} | 0.9201 ± 0.0019 | 0.9670 ± 0.0020 |
| ϵ_{iso} | 0.9815 ± 0.0011 | 0.9962 ± 0.0012 |
| ϵ_{sa} | 0.9767 ± 0.0013 | 0.9972 ± 0.0014 |
| ϵ_{trk} | 0.9890 ± 0.0009 | 0.9949 ± 0.0010 |
| $N_{Z \rightarrow \mu^+ \mu^-}$ | 13152 ± 119 | |
| χ^2/ndof | 1.12 | |
| p - value | 0.44 | |

Table 6.15: Comparison between fit parameters performed in data with simulation signal and background scaled to the data luminosity, using the full 2010 data, χ^2/NDOF and p -value are also requested. Data/simulation ratio are also shown for comparison.

the $Z \rightarrow \mu^+\mu^-$ cross section already with the 2010 data.

Conclusions

The CMS experiment collected about 35 pb^{-1} proton-proton collision data in 2010, most of which in the last weeks of proton operation in October and November 2010, the first year of 7 TeV collision for the LHC.

The measurement of the cross-section for the process $pp \rightarrow ZX \rightarrow \mu^+\mu^-X$ has been amongst the earliest conducted with the CMS detector in 2010. For this measurement it is crucial to understand the response of the CMS detector, especially the Muon Spectrometer, which is responsible for the identification and the precise measurement of relatively highly energetic muons.

Several methods have been developed and discussed in the context of the total cross-section measurement of $pp \rightarrow ZX \rightarrow \mu^+\mu^-X$, depending on the integrated luminosity available: from the first months “cut and count” analysis to a fully data-driven technique. A data-driven measurement strategy of the $Z \rightarrow \mu^+\mu^-$ cross section during its start-up phase has been described in detail. This strategy has been tuned for an integrated luminosity of few pb^{-1} and indeed has been applied for the first time on a dataset of 2.9 pb^{-1} of 7 TeV collision data but can easily be adjusted for later periods of the experiment. The method described, which permitted to CMS collaboration to publish the $Z \rightarrow \mu^+\mu^-$ measurement for the first time [1], relies on a fully data driven approach: the $Z \rightarrow \mu^+\mu^-$ inclusive cross-section is determined with a simultaneous fit of the yield of $Z \rightarrow \mu^+\mu^-$ events, the average reconstruction muon efficiencies in the tracker and in the muon detector, the trigger efficiency, as well as the efficiency of the cut applied to select isolated muons. The extracted Z yield has to be just corrected for the geometrical acceptance and for the integrated luminosity in order to measure the cross section. The measurements obtained with the first 2.9 pb^{-1} is

$$\sigma(\text{pp} \rightarrow \text{ZX}) \times \text{BF}(Z \rightarrow \mu^+\mu^-) = \mathbf{0.924} \pm \mathbf{0.031}(\text{stat.}) \pm \mathbf{0.022}(\text{syst.}) \pm \mathbf{0.101}(\text{lumi.})\text{nb}, \quad (6.54)$$

referred to the invariant mass range $60 < m_{\mu^+\mu^-} < 120 \text{ GeV}/c^2$. The value measured is in agreement with the Standard Model prediction.

The systematic uncertainties has been reduced to the minimum. It arises mainly on theoretical uncertainties of the process contributing to the Z boson production.

It is a major goal of the presented approach to minimize the dependence on Monte Carlo simulations.

The systematic uncertainty dominates over the statistical error with a luminosity greater than 10 pb^{-1} , as presented in the very last section of the thesis, where an update of the $Z \rightarrow \mu^+\mu^-$ cross-section measurement using 35 pb^{-1} has been presented as a preliminary result. This amount of data correspond to the full 2010 collision dataset. We measure for the $Z \rightarrow \mu^+\mu^-$ cross-section with the full 2010 data the preliminary value:

$$\sigma(\mathbf{pp} \rightarrow \mathbf{ZX}) \times \mathbf{BF}(Z \rightarrow \mu^+\mu^-) = \mathbf{0.954} \pm \mathbf{0.009}(\text{stat.}) \pm \mathbf{0.022}(\text{syst.}) \pm \mathbf{0.105}(\text{lumi.})\mathbf{nb}, \quad (6.55)$$

in agreement with the previous measurement and the NNLO prediction.

The method presented can be used for all the life of the experiment as a machine luminosity indicator and measurement. Once we believe in the Standard Model prediction, as demonstrated to be true in the work, we can revert the cross section formula and determine the integrated luminosity run by run, when the LHC instantaneous luminosity will reach the design value.

Bibliography

- [1] CMS Collaboration, "Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s}=7$ TeV". CMS-EWK-10-002, CERN-PH-EP-2010-050, Dec 2010. Temporary entry e-Print: arXiv:1012.2466 [hep-ex]
- [2] M.De Gruttola et al., *Persistent storage of non-event data in the CMS databases*, JINST 5:P04003,2010.
- [3] M.De Gruttola et al. First experience in operating the population of the condition databases of the CMS experiment, *International Conference on Computing in High-Energy Physics Conference (CHEP 2009), Distributed Processing and Analysis track*, 23–27 March 2009, Prague, Czech Republic.
- [4] S. de Capua, (CERN) , F. Ferro, (INFN, Genoa) , M. De Gruttola, (INFN, Naples) , M. Villa, (INFN, Bologna) . 2008. 12pp.“Luminosity measurements at LHC”, prepared for 5th Italian Workshop on LHC Physics (In Italian), Perugia, Italy, 30 Jan - 2 Feb 2008.Published in Nuovo Cim.123B:423-434,2008.
- [5] M.De Gruttola et al., "Determination of the $Z \rightarrow \mu^+\mu^-$ inclusive cross section with a simultaneous fit on Z Yield and muon reconstruction efficiency", CMS AN-2008/062.
- [6] M.De Gruttola et al, "Determination of the $Z \rightarrow \mu^+\mu^-$ inclusive cross section with a simultaneous fit on Z Yield, muon reconstruction efficiency and High Level Trigger efficiency", CMS AN-2009/005.
- [7] M.De Gruttola et al., “Updated Measurements of Inclusive W and Z Cross Sections at 7 TeV”, CMS AN-2010-264.
- [8] M.De Gruttola et al., "Determination of the $Z \rightarrow \mu^+\mu^-$ inclusive cross section with a simultaneous fit of Z yield, muon reconstruction efficiencies and trigger efficiency with the first 2.9 pb-1 of 7 TeV collision data", CMS AN-2010-345.
- [9] CMS Collaboration, "Towards a measurement of the inclusive $W \rightarrow \mu\nu$ e $Z \rightarrow \mu^+\mu^-$ cross section in pp collision at $\sqrt{s}=14$ TeV",CMS PAS-EWK-07-002.

-
- [10] CMS Collaboration, "Measurement of the W and Z cross section with muons (10 TeV update)" - CMS PAS-EWK-09-001.
- [11] CMS Collaboration, "Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{7}$ TeV", CMS PAS EWK-2010-002 (2010).
- [12] Gell-Mann, Murray (1995). The Quark and the Jaguar. Owl Books. ISBN 978-0805072532.
- [13] S. L. Glashow, Nucl. Phys. 22, 579 (1961).
- [14] S. Weinberg, Phys. Rev. Lett. 19, 1264 (1967). January 2, 2009 14:52 WSPC - Proceeding.s
- [15] A. Salam In Elementary Particle Theory, ed. N. Svartholm (Almqvist and Wiksells, Stockholm 1969), 367-377.
- [16] P. Langacker, "Introduction to the Standard Model and Electroweak Physics," arXiv:0901.0241 [hep-ph].
- [17] J. Goldstone, A. Salam and S. Weinberg, Phys. Rev. 127, 965 (1962).
- [18] F J Hasert et al. 1973a Phys. Lett. 46B 121.
- [19] G. Arnison et al., Phys. Lett. B166, p. 484 (1986).
- [20] R. Ansari et al., Phys. Lett. B186, p. 440 (1987).
- [21] Guido Altarelli, New Physics and the LHC., (Rome III U. & INFN, Rome3 & CERN) . RM3-TH-08-9, CERN-PH-TH-2008-085, May 2008. 35pp.
- [22] The LEP Electroweak Working Group, <http://lepewwg.web.cern.ch/LEPEWWG/>
- [23] T. Aaltonen et al. (CDF and D0 Collaborations), Phys. Rev. Lett. 104, 061802 (2010).
- [24] S.P. Martin, "A Supersymmetry Primer", hep-ph/9709356 (1997).
- [25] For a review and a list of references, see, for example, J. Hewett and M. Spiropulu, Ann.Rev.Nucl.Part.Sci. 52, 397 (2002), hep-ph/0205196.
- [26] CMS Collaboration, "Selection of Top-Like Events in the Dilepton and Lepton-plus-Jets Channels in Early 7 TeV Data", CMS-PAS-TOP-10-004.
- [27] CMS Collaboration, CMS Physics Technical Design Report, Vol II: Physics Performance, CERN/LHCC/2006-021 (2006), J. Phys. G Nucl. Part. Phys. 34 (2006) 995.

-
- [28] ATLAS Collaboration, Expected performance of the ATLAS experiment: detector, trigger, physics, CERN Report, CERN-OPEN-2008-020, arXiv:0901.0512.
- [29] CMS Collaboration, "Search for Quark Compositeness with the Dijet Centrality Ratio in pp Collisions at $\sqrt{s}=7$ TeV." CMS-EXO-10-002, CERN-PH-EP-2010-038, Oct 2010. e-Print: arXiv:1010.4439 [hep-ex]
- [30] DØ Collaboration, V.M. Abazov, et al., Fermilab-PUB-09-326-E, Phys. Rev. Lett., arXiv:0906.4819, submitted for publication.
- [31] CMS Collaboration, "Performance of Methods for Data-Driven Background Estimation in SUSY Searches", CMS-PAS-SUS-10-001.
- [32] ATLAS Collaboration, Expected performance of the ATLAS experiment: detector, trigger, physics, CERN Report, CERN-OPEN-2008-020, arXiv:0901.0512.
- [33] Georgi, H.; Glashow, S.L. (1974). "Unity of All Elementary Particle Forces". Physical Review Letters 32: 438-441. doi:10.1103/PhysRevLett.32.438.
- [34] P. Nadolsky, D. R. Stump, and C. P. Yuan. Semi-inclusive hadron production at hera: The effect of QCD gluon resummation. Phys. Rev., D61:014003, 2000.
- [35] G. Altarelli, G. Parisi, and R. Petronzio. Transverse momentum in drell-yan processes. Phys. Lett., B76:351, 1978.
- [36] T. Sjostrand, S. Mrenna, and P. Skands. Pythia 6.4 physics and manual. JHEP, 05:026, 2006.
- [37] G. Marchesini et al. Herwig: A monte carlo event generator for simulating hadron emission reactions with interfering gluons. version 5.1 - april 1991. Comput. Phys. Commun., 67:465508, 1992.
- [38] S. Frixione and B. R. Webber. The mc@nlo event generator. 2002.
- [39] G. A. Ladinsky and C. P. Yuan. The nonperturbative regime in qcd resummation for gauge boson production at hadron colliders. Phys. Rev., D50:4239, 1994.
- [40] F. Landry, R. Brock, G. Ladinsky, and C. P. Yuan. New fits for the non-perturbative parameters in the css resummation formalism. Phys. Rev., D63:013004, 2001.
- [41] P.Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. JHEP 0411:040,2004.
- [42] Fermilab web site: <http://www.fnal.gov/>.

-
- [43] *CMS Physics. Technical Design Report (vol. 1: Detector Performance and Software)*, CMS collaboration, CERN-LHCC-2006-01.
- [44] Lyndon Evans and Philip Bryant (editors), *LHC Machine*, 2008 JINST 3 S08001.
- [45] O.S. Brüning et al. (editors), The LHC design report v.1: the LHC Main Ring, CERN-2004-003-V-1, <http://cdsweb.cern.ch/record/782076>; The LHC design report v.2 : the LHC Infrastructure and General Services, CERN-2004-003-V-2, <http://cdsweb.cern.ch/record/815187>; M. Benedikt et al. (eds.), The LHC design report v.3 : the LHC Injector Chain, CERN-2004-003-V-3, <http://cdsweb.cern.ch/record/823808>.
- [46] ATLAS Collaboration, ATLAS: technical proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN, CERN-LHCC-94-43, <http://cdsweb.cern.ch/record/290968>.
- [47] CMS collaboration, CMS technical proposal, CERN-LHCC-94-38, <http://cdsweb.cern.ch/record/290969>.
- [48] LHCb collaboration, LHCb technical proposal, CERN-LHCC-98-004, <http://cdsweb.cern.ch/record/622031>.
- [49] TOTEM collaboration, W. Kienzle et al, TOTEM, Total cross section, elastic scattering and diffractive dissociation at the LHC: Technical Proposal, CERN-LHCC-99-007, <http://cdsweb.cern.ch/record/385483>.
- [50] ALICE collaboration, ALICE: Technical proposal for a Large Ion collider Experiment at the CERN LHC, CERN-LHCC-95-71, <http://cdsweb.cern.ch/record/293391>.
- [51] THE LHC STUDY GROUP, Design study of the large hadron collider (LHC): a multiparticle collider in the LEP tunnel, CERN-91-03, <http://cdsweb.cern.ch/record/220493>.
- [52] P. Fabbriatore et al., The construction of the modules composing the CMS superconducting coil, *IEEE Trans. Appl. Supercond.* 14 (2004) 552.
- [53] CMS collaboration, The CMS tracker system project: technical design report, CERN-LHCC-98-006, <http://cdsweb.cern.ch/record/368412>.
- [54] CMS collaboration, The CMS tracker: addendum to the technical design report, CERN-LHCC-2000-016, <http://cdsweb.cern.ch/record/490194>.

-
- [55] CMS collaboration, CMS Physics Technical Design Report Volume 1: Detector Performance and Software, CERN-LHCC-2006-001, <http://cdsweb.cern.ch/record/922757>; CMS collaboration, CMS Physics Technical Design Report Volume 2: Physics Performance, J. Phys. G 34 (2006) 995, CERN-LHCC-2006-021, <http://cdsweb.cern.ch/record/942733>.
- [56] CMS-CASTOR group, CASTOR Engineering Design Report: A Calorimeter for CMS at Very Forward Rapidity (CMS, 2007).
- [57] CMS collaboration, The CMS muon project, technical design report, CERN-LHCC-97-032, <http://cdsweb.cern.ch/record/343814>.
- [58] G. Charpak and F. Sauli, High-accuracy, two-dimensional read-out in multiwire proportional chambers, Nucl. Instrum. Meth. 113 (1973) 381.
- [59] R. Santonico and R. Cardarelli, Development of resistive plate counters, Nucl. Instrum. Meth. 187 (1981) 377.
- [60] CMS COLLABORATION group, W. Adam et al., The CMS high level trigger, Eur. Phys. J. C 46 (2005) 605 [hep-ex/0512077].
- [61] P. Billoir, R. Fr auwirth and M. Regler, Nucl Instr. Meth. A241 (1985) 115.
- [62] CMS data processing workflows during an extended cosmic ray run, CMS Collaboration 2010 JINST 5 T03006, doi: 10.1088/1748-0221/5/03/T03006.
- [63] World-wide LHC Computing Grid (WLCG), <http://lcg.web.cern.ch/LCG/public/default.htm>.
- [64] CMS collaboration, CMS: The computing project. Technical design report, CERN-LHCC-2005-023.
- [65] R. Brun and F. Rademakers, ROOT: an object oriented data analysis framework, Nucl. Instrum. Meth. A 389 (1997) 81; see also <http://root.cern.ch>.
- [66] P. Sphicas ed., CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger, CERN-LHCC-2002-026.
- [67] PAT: The CMS Physics Analysis Toolkit, W Adam et al 2010 J. Phys.: Conf. Ser. 219 032017 doi: 10.1088/1742-6596/219/3/032017.
- [68] Kevin Loney, Oracle Database 11g The Complete Reference, Oracle, Release 1 (11.1), 2008.

-
- [69] Z.Xie et al., POOL Persistency Framework for the LHC: New Developments and CMS Applications *Proc. of "Frontier Science 2005: New Frontiers in Sub nuclear Physics*, 12–17 September, 2005 Milan, Italy”.
- [70] C. Janot, "L'assurance de la qualité pour une Installation Nucléaire de Base" - CERN-ST-2000-001.
- [71] M. Gonzalez-Berges, The Joint Controls Project Framework, *Proceeding of the International Conference on Computing in High Energy Physics (CHEP 2003)*, 24–28 March 2003, La Jolla, California.
- [72] B.Blumenfeld et al., CMS conditions data access using FroNTier, *2008 J. Phys.: Conf. Ser. 119 072007*.
- [73] G.Govi et al., CMS Offline Conditions Framework and Services, *International Conference on Computing in High-Energy Physics Conference (CHEP 2009), Distributed Processing and Analysis track*, 23–27 March 2009, Prague, Czech Republic.
- [74] C.D. Jones et al., Analysis environments for CMS, *J. Phys.: Conf. 2008 Ser. 119 032027*.
- [75] C.D. Jones et al., The New CMS Event Data Model and Framework, *Proceedings of the International Conference on Computing in High-Energy Physics (CHEP 2006)*, 13–17 February 2006, T.I.F.R. Mumbai, India.
- [76] Ignas Butenas et al., PopCon monitoring: web application for detailed real-time database transaction monitoring, *Proceedings of the 15th International Conference on Distributed Multimedia Systems*, 10–12 September 2009, San Francisco Bay, USA.
- [77] D. Spiga et al., The CMS Remote Analysis Builder (CRAB), *Lect. Notes Comput. Sci. 4873 (2007) 580*.
- [78] CMS Collaboration, “Performance of CMS muon reconstruction in cosmic-ray events”, JINST 5 (2010) T03022. doi:10.1088/1748-0221/5/03/T03022.
- [79] CMS Collaboration, “Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV”, CMS PAS MUO-10-002 (2010).
- [80] the LEP Electroweak Working Group, Precision electroweak measurements on the z resonance. *Phys. Rept.*, 427:257, 2006.
- [81] A. D. Martin et al. *Eur. Phys. J. C64:653-680*, 2009.

- [82] ICHEP 2010 conference web site: <http://www.ichep2010.fr/>.
- [83] GEANT4 Collaboration, S. Agostinelli et al., GEANT4: A simulation toolkit, Nucl. Instrum. Meth. A506 (2003) 250303.
- [84] CMS Collaboration, "Measurement of CMS luminosity", CMS PAS EWK-2010-004 (2010).
- [85] Clopper, C.; Pearson, E. S. (1934). "The use of confidence or fiducial limits illustrated in the case of the binomial". Biometrika 26: 404-413.
- [86] S. Baker and R. Cousins, Clarification of the Use of Chi-Square and Likelihood Functions in Fits to Histograms, NIM 221:437 (1984).
- [87] J. Campbell and R.K. Ellis, Monte Carlo for FeMtobarn processes, 1449 <http://mcfm.fnal.gov/>.
- [88] ATLAS Collaboration, "Measurement of the $W \rightarrow l\nu$ and $Z/\gamma^* \rightarrow ll$ production cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector." CERN-PH-EP-2010-037, Oct 2010. 38pp.e-Print:arXiv:1010.2130[hep-ex].

Curriculum Vitae (Nov. 2010)

MICHELE DE GRUTTOLA

Address

Via Cerreto 15A Ariano Irpino (AV) 83031 Italy

Phone: +39-320-4572190

Email: michele.de.gruttola@cern.ch

Homepage: <http://people.na.infn.it/~degruttola/>

Personal Details

Gender: Male

Date of birth: 24th of April, 1983

Place of birth: Ariano Irpino, Italy

Present Citizenship: Italian

Education

07/2002 Scientific lyceum diploma with grade 100/100 - Liceo Parzanese Ariano Irpino

10/2005 Bachelor degree MAGNA CUM LAUDE in Physics at the University of Naples “Federico II”, with the Thesis *Study on CP violation in the decay $\Phi \rightarrow K_S K_L \rightarrow \pi^+ \pi^- \pi^+ \pi^-$ at the experiment DAΦNE, with the KLOE detector*: Supervisor: Prof M.Napolitano.

01/2007–12/2007 Technical Student at CERN working in the CMS database project group; Supervisor V.Innocente.

10/2007 Master degree MAGNA CUM LAUDE in Physics at the University of Naples “Federico II”, with the Thesis *Study of the Z decay in two muons with the CMS detector at LHC-CERN*; Supervisor prof. C.Sciacca and L.Lista.

09/2009–08/2010 CERN associate on behalf of INFN

end 2010 Expected completion of PhD Program

Awards

2006 INFN grant for best graduated master students in Particle Physics.

2007 selected for the Technical Studentship program at CERN

2009 selected for the CERN associate program on behalf of INFN

2009 awarded as “illustrious citizen” from my city major, for the work done at CERN which “made the city proud” of me.

Language Knowledge

Italian native
English near native
French near native

Courses and schools attended

| | |
|----------------------------------|-----------------------------|
| Oracle-Sql | CERN, Apr 2008 |
| Python Hands-On Introduction | CERN, Oct 2008 |
| CERN 2009 EUROPEAN SCHOOL OF HEP | Bautzen (Germany), Jun 2009 |

Responsibilities

- 2008** Co-responsible of the PopCon software application for transferring condition data from the various CMS sub-detectors to the offline condition database
- 2010** CERN associate on behalf of the INFN with the mandate to play a leading role in the CMS Electro-weak Physics group first analysis and paper with 7 TeV data.

Personal presentation at Conferences

- Feb 2008** *Luminosity measurement at LHC with known physics channels*, at the “V Italian Workshop on p-p Physics at LHC”, Perugia, Italy, 30/01-02/02 2008
- Mar 2009** *First experience in operating the population of the condition database for the CMS experiment*, Computing in Nuclear and High Energy Physics (CHEP09), Prague, Czech Republic, 21-27/03 2009
- Apr 2010** *Inclusive W and Z production at LHC startup with the CMS experiment*, IFAE2010 Incontri di Fisica delle Alte Energie "Sapienza" Università di Roma - Dip. di Fisica Roma, 7-9 Aprile 2010
- May 2010** *Time-critical Database Conditions Data-handling for the CMS Experiment*, 17TH REAL TIME CONFERENCE 24-28 May 2010 Lisbon, Portugal
- Jul 2010** *Recent Results with High-Pt Leptons from CMS*, Blois2010: “22nd Rencontres de Blois on Particle Physics and Cosmology”; First Results from the LHC”, 15-20 Jul 2010, Blois (France)".
- Sep 2010** *Osservazione dei bosoni W e Z con l'esperimento CMS in collisioni pp ad energia nel centro di massa di 7 TeV* at SIF 2010: XCVI CONGRESSO NAZIONALE SOCIETÀ ITALIANA DI FISICA, 20-24 Sep 2010, University of Bologna, Bologna (Italy)

List of contributions to CMS experiment internal and public note

- CMS AN-2007/031** *Towards a measurement of the inclusive $W \rightarrow \mu\nu$ and $Z \rightarrow \mu\mu$ cross section in pp collision at $\sqrt{s}=14$ TeV*, M.De Gruttola et al., CMS AN-2007/031 Dec. 2007

- CMS PAS-EWK-07/002** *Towards a measurement of the inclusive $W \rightarrow \mu\nu$ and $Z \rightarrow \mu\mu$ cross section in pp collision at $\sqrt{s}=14$ TeV*, CMS PAS-EWK-07/002, Dec. 2007
- CMS AN-2008/062** *Determination of the $pp \rightarrow ZX \rightarrow \mu\mu X$ inclusive cross section with a simultaneous fit on Z Yield and muon reconstruction efficiency*, M.De Gruttola, S.Di Guida, F.Fabozzi, L.Lista, P.Noli, CMS AN-2008/062, July 2008
- CMS AN-2009/005** *Determination of the $pp \rightarrow ZX \rightarrow \mu\mu X$ inclusive cross section with a simultaneous fit on Z Yield, muon reconstruction efficiency and High Level Trigger efficiency*, M.De Gruttola, A.de Cosa, S.Di Guida, F.Fabozzi, L.Lista, P.Noli, CMS AN-2009/005 Jan. 2009
- CMS PAS-EWK-09/001** *Measurement of the W and Z cross section with muons (10 TeV update)* - CMS PAS-EWK-09-001, Mar. 2009
- CMS NOTE-2010/001** *Persistent storage of non-event data in the CMS databases*, CMS NOTE-2010/001, M.De Gruttola et al., Jan. 2010
- CMS AN-2010/116** *Measurement of the W and Z inclusive production cross sections at $\sqrt{s}=7$ TeV with the CMS experiment at the LHC*, M. De Gruttola et al., CMS AN-2010/116, Jul. 2010
- CMS PAS-EWK-10/002** *Measurement of the W and Z inclusive production cross sections at $\sqrt{s}=7$ TeV with the CMS experiment at the LHC*, CMS PAS-EWK-10-002, Jul. 2010
- CMS AN-2010/216** *Updated Measurements of Inclusive W and Z Cross Sections at 7 TeV*, M.De Gruttola et al., CMS AN-2010-216, Oct. 2010
- CMS AN-2010/345** *Determination of the $pp \rightarrow ZX \rightarrow \mu\mu X$ inclusive cross section with a simultaneous fit of Z yield, muon reconstruction efficiencies and trigger efficiency with the first 2.9 pb^{-1} of 7 TeV collision data*, M.De Gruttola, A.de Cosa, F.Fabozzi, L.Lista, P.Noli, , CMS AN-2010/345, Nov. 2010