



University of Zagreb

FACULTY OF SCIENCE

Jelena Luetić

**Measurement of the cross section for  
associated production of a W boson and  
two b quarks with the CMS detector at the  
Large Hadron Collider**

DOCTORAL THESIS

Zagreb, 2015.





University of Zagreb

FACULTY OF SCIENCE

Jelena Luetić

**Measurement of the cross section for  
associated production of a W boson and  
two b quarks with the CMS detector at the  
Large Hadron Collider**

DOCTORAL THESIS

Supervisor:

Professor Vuko Brigljević, PhD

Zagreb, 2015





Sveučilište u Zagrebu

PRIRODOSLOVNO MATEMATIČKI FAKULTET

Jelena Luetić

**Mjerenje udarnog presjeka zajedničke  
produkциje  $W$  bozona i para b kvarkova  
CMS detektorom na Velikom hadronskom  
sudarivaču**

DOKTORSKI RAD

Mentor:

Prof. dr. sc. Vuko Brigljević

Zagreb, 2015.

*“Each time new experiments are observed to agree with the predictions, the theory survives and our confidence in it is increased; but if ever a new observation is found to disagree, we have to abandon or modify the theory.*

*At least that is what it is supposed to happen, but you can always question the competence of the person who carried out the observation.”*

Stephan Hawking

# Abstract

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

## *Acknowledgements*

The acknowledgments and the people to thank go here, don't forget to include your project advisor...

# Contents

<b>Abstract</b>	iii
<b>Acknowledgements</b>	iv
<b>Contents</b>	v
<b>List of Figures</b>	ix
<b>List of Tables</b>	xi
<b>1 Introduction</b>	1
<b>2 Theoretical overview and previous measurements</b>	3
2.1 Standard model overview . . . . .	4
2.1.1 Bottom quarks . . . . .	6
2.1.2 W boson . . . . .	7
2.2 W + b jets at hadron colliders . . . . .	8
2.2.1 Cross sections at hadron colliders . . . . .	9
2.2.2 Contributions to Wbb cross section . . . . .	13
2.2.2.1 Double parton scattering . . . . .	17
2.3 Previous measurements . . . . .	21
<b>3 Large Hadron Collider</b>	25
3.1 Physics goals for the LHC . . . . .	26
3.2 Design of the LHC . . . . .	27
3.3 LHC performance . . . . .	29
<b>4 Compact Muon Solenoid</b>	33
4.1 CMS coordinate system . . . . .	34
4.2 Solenoid magnet . . . . .	35
4.3 Inner tracker system . . . . .	36

4.3.1	Pixel Detector . . . . .	36
4.3.2	Strip detector . . . . .	38
4.4	Electromagnetic calorimeter . . . . .	39
4.5	Hadronic calorimeter . . . . .	41
4.6	Muon chambers . . . . .	42
4.7	Trigger . . . . .	42
4.8	Luminosity measurement . . . . .	45
<b>5</b>	<b>Physics objects definitions</b>	<b>47</b>
5.1	Electrons . . . . .	47
5.2	Muons . . . . .	50
5.3	Lepton isolation . . . . .	51
5.4	Jets . . . . .	52
5.4.1	Jet algorithms . . . . .	52
5.4.2	Jet corrections . . . . .	55
5.4.3	Jet identification . . . . .	57
5.4.4	Jets from b quarks . . . . .	58
5.5	Missing transverse energy . . . . .	60
<b>6</b>	<b>Event selection and background estimation</b>	<b>63</b>
6.1	Data and Monte Carlo samples . . . . .	64
6.2	Event selection . . . . .	66
6.3	Background estimation . . . . .	67
6.3.1	Top quark background . . . . .	67
6.3.2	Z+jets . . . . .	70
6.3.3	W+light jets and W+charm . . . . .	70
6.3.4	QCD . . . . .	70
6.3.5	Other backgrounds . . . . .	73
6.4	Monte-Carlo corrections . . . . .	74
6.4.1	Pileup . . . . .	74
6.4.2	Lepton efficiency measurement . . . . .	75
6.4.3	<i>b</i> -tagging scale factors . . . . .	76
<b>7</b>	<b>Cross section measurement</b>	<b>81</b>
7.1	Fit methodology . . . . .	81
7.2	Systematic uncertainties . . . . .	83
7.3	Acceptance and efficiency . . . . .	86
7.4	Results . . . . .	88
7.4.1	Effects of the systematics . . . . .	91
7.4.2	Tests of the fit stability . . . . .	91
7.4.3	Cross section measurement . . . . .	92
7.4.4	Theoretical calculations and comparison with the measurement . . . . .	92

<b>8 Conclusions</b>	<b>93</b>
<b>9 Prosireni sazetak - Mjerenje udarnog presjeka zajedničke produkcije W bozona i para b kvarkova</b>	<b>95</b>
9.1 Main Section 1 . . . . .	95
<b>A Lorentz angle measurement in Pixel detector</b>	<b>97</b>
A.1 Grazing angle method . . . . .	97
A.2 Minimum cluster size method (V-method) . . . . .	101
<b>B Acceptance and efficiency error calculation</b>	<b>103</b>
<b>Bibliography</b>	<b>105</b>



# List of Figures

2.1	List of Standard model elementary particles . . . . .	5
2.2	Strong force coupling constant . . . . .	9
2.3	Drawing of a proton-proton collision . . . . .	10
2.4	Parton distribution functions for different momentum transfers . . . . .	12
2.5	Proton-proton cross sections . . . . .	13
2.6	Leading order Wbb Feynmann diagram . . . . .	15
2.7	Scale dependence of Wbb cross section . . . . .	15
2.8	Wbb NLO scale dependence . . . . .	16
2.9	Wbb production within 5 flavor scheme . . . . .	17
2.10	Double parton scattering . . . . .	18
2.11	Results of $\sigma_{eff}$ measurements . . . . .	20
2.12	Atlas Wbb total cross section measurement . . . . .	22
2.13	Measured differential W+b-jets cross-sections as a function of leading b-jet $p_T$ . . . . .	23
2.14	CMS Wbb total cross section measurement . . . . .	24
3.1	Schematics of Large Hadron Collider . . . . .	26
3.2	Schematics of dipole magnets . . . . .	28
3.3	Luminosity delivered to the CMS experiment . . . . .	31
4.1	CMS detector . . . . .	34
4.2	CMS Pixel Detector . . . . .	37
4.3	CMS Pixel Detector psudorapidity range coverage and efficiency . . . . .	37
4.4	CMS Strip Detector . . . . .	39
4.5	CMS Electromagnetic Calorimeter . . . . .	40
4.6	CMS Hadronic Calorimeter . . . . .	41
4.7	CMS Muon Chambers . . . . .	43
4.8	Muon resolution measurements for tracker, muon chambers and combined .	43
4.9	A drawing of CMS Trigger System. . . . .	44
5.1	An example of configuration of IRC unsafe jet algorithm. . . . .	53
5.2	Clustering particles into jets with different algorithms. . . . .	54
5.3	Total jet energy correction as a function of pseudorapidity of two different jet $p_T$ values. . . . .	56

---

5.4	Total jet energy correction as a function of transverse momentum for four different $\eta$ values. . . . .	57
5.5	Combined secondary vertex discriminant value for multijet QCD sample (left) and tt enriched sample(right)[53] . . . . .	59
5.6	Combined secondary vertex misidentification probability for data and MC for medium working point.[53] . . . . .	60
6.1	Signal region detector level distributions. . . . .	68
6.2	Signal region detector level distributions. . . . .	69
6.3	Feynmann diagrams showing major backgrounds . . . . .	70
6.4	Top quark control region . . . . .	71
6.5	Distribution obtained using Wbb event selection before applying b-tagging criteria. . . . .	72
6.6	QCD diagram and illustration of QCD background determination . . . . .	73
6.7	Transverse mass distribution before and after QCD distribution determination. . . . .	74
6.8	Number of primary vertices before and after the pileup reweighing procedure. . . . .	75
6.9	Muon identification and isolation efficiencies using <i>tag and probe</i> method. . . . .	77
6.10	Muon trigger efficiency using <i>tag and probe</i> method. . . . .	77
6.11	B-tagging scale factors and misstag scale factors . . . . .	79
7.1	Comparison of the two W+bb samples used to obtain the number of signal events. . . . .	82
7.2	Shape of the transverse mass distribution for each systematic variation in both, signal region and TT control region. . . . .	87
7.3	Muon channel postfit distributions. . . . .	89
7.4	Electron channel postfit distributions. . . . .	90
A.1	Angle definitions for grazing angle method. . . . .	98
A.2	Depth at which electrons in silicon bulk were produced as a function of Lorentz drift. . . . .	99
A.3	The average drift of electrons as a function of the production depth. Slope of the linear fit result is the $\tan\theta_L$ . . . . .	99
A.4	Lorentz angle as a function of integrated luminosity for 2012. . . . .	100
A.5	An example of V-method fit. . . . .	101

# List of Tables

3.1	LHC performance in 2012 together with design performance[37] . . . . .	30
3.2	LHC performance highlights . . . . .	31
5.1	A summary of electron identification criteria. . . . .	49
5.2	A summary of muon identification criteria. . . . .	51
5.3	A summary of jet identification criteria. . . . .	58
6.1	Samples, generators and cross sections used for normalizations for signal and background simulation considered in this analysis. All samples are calculated to the NLO order except the W+jets which is NLO and $t\bar{t}$ which is normalized to the latest combined cross section measurement of ATLAS and CMS collaborations [67]. . . . .	65
7.1	Standard model cross section uncertainties used in the evaluation of MC normalization systematic effect. . . . .	86
7.2	Fiducial cuts used for cross section measurements. . . . .	87
7.3	Results of the $A \times \epsilon$ measurement for both, muon and electron channel together with the statistical uncertainty. . . . .	88
7.4	Yields obtained in the muon channel before and after the fitting procedure.	88
7.5	Yields obtained in the electron channel before and after the fitting procedure.	91
7.6	Systematics . . . . .	91
7.7	Systematics . . . . .	92
7.8	UPDATE NUMBERS! - Signal strengths obtained by fitting different distributions. Signal strengths are found to be consistent with each other. . . . .	92
A.1	Selection criteria for Lorentz angle measurement . . . . .	100



# Chapter 1

## Introduction



# Chapter 2

## Theoretical overview and previous measurements

Standard model of elementary particles is a theory which emerged in 1960s and 1970s, describing all of the known elementary particles and interactions except gravity. The final formulation of the Standard model incorporates several theories: quantum electrodynamics, Glashow-Weinberg-Salam theory of electroweak processes and quantum chromodynamics. First steps towards formulation of Standard model occurred in 1961. when Sheldon Glashow unified electromagnetic and weak interactions[1]. The difference in strength between weak and electromagnetic force was puzzling for the physicists at that time, and Glashow proposed that it can be accounted for if the weak force were mediated by massive bosons. However, he was not able to explain the origin of the mass for such mediators. The explanation came in 1967. when Steven Weinberg and Abdul Salam used Higgs mechanism in the electroweak theory[2, 3] which suggested the existence of additional particle called Higgs boson. After discovery of neutral currents which arise from the exchange of the neutral Z boson, electroweak theory became generally accepted. W and Z bosons were discovered in 1983 at CERN[4, 5], and their masses were in agreement with the Standard model prediction. Theory describing strong interactions got it's final form

in 1974. when it was shown that hadrons are consisting of quarks. The the final missing link in the Standard model, the Higgs boson, was discovered in 2012 at CERN[6, 7]. There are several unexplained phenomena which suggest the existence of physics beyond Standard model, but so far it's predictions were confirmed every time through numerous experimental tests.

In this chapter a brief overview of the Standard model particles and interactions will be shown with the emphasis on the W boson and b quarks which are the most relevant for this thesis. An introduction to cross section determination at hadron colliders is given. In the last part of the chapter historical overview of the development of W+b-jets theoretical calculations is described together with the existing experimental results.

## 2.1 Standard model overview

Elementary particle physics is described within the framework of the Standard model. We usually imagine particles as point like objects and some forces between them. These particles called are fermions, leptons or quarks of spin  $s = 1/2$ . There are three charged leptons, electron, muon and tau which properties are the same except for their mass. Each of the leptons has a corresponding neutrally charged neutrino with a very small mass. There are six different types of quarks with charge either  $Q = 2/3$  of  $Q = -1/3$ . They also carry one additional quantum number which is color charge. All objects observed in nature are colorless giving raise to the concept of quark confinement which will be explained later. Colorless composite objects are classifies into two categories. Bayons are fermions made out of three quarks, for example proton or neutron. The other category are mesons which are made of two quarks like pions. Quarks are divided into three generations where all properties are identical except for the mass of the particles.

From the point of view of the quantum field theory, Standard model is based on a gauge symmetry  $SU(3)_C \times SU(2)_L \times U(1)_Y$ . Strong interaction is described by  $SU(3)_C$ , while electroweak sector is described by  $SU(2)_L \times U(1)_Y$ . All interactions within the

Standard Model are mediated by an elementary particle which is a spin 1 boson. In the case of electromagnetic interaction, mediator is a massless photon, thus the range of electromagnetic interaction is infinite. For weak force mediators are three massive bosons  $W^\pm$  and  $Z$  and it's range is very small ( $10^{-16}$  m). These four bosons are the gauge bosons of  $SU(2)_L \times U(1)_Y$  group. The interaction between electroweak bosons is allowed in the Standard Model as long as charge conservation principle remains valid. Strong force is mediated by the exchange of 8 massless gauge bosons for  $SU(3)_C$  called gluons. Although gluons are massless, the range of the strong force is not infinite. Because of the effect of confinement, the range of the strong force is approximately the size of the lightest hadrons ( $10^{-13} cm$ ).

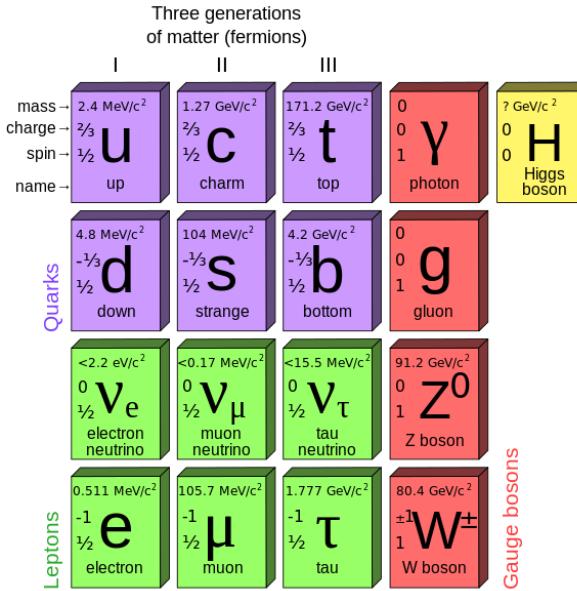


FIGURE 2.1: List of Standard model elementary particles.

The fact that weak gauge bosons are massive indicates that  $SU(2)_L \times U(1)_Y$  is not a good symmetry of the vacuum. In contrast with photon being massless,  $U(1)_{em}$  is a good symmetry of the vacuum which means that  $SU(2)_L \times U(1)_Y$  electroweak symmetry is somehow spontaneously broken to  $U(1)_{em}$  of electromagnetism. Spontaneous symmetry breaking is implemented through Higgs mechanism which gives masses to fermions,  $W^\pm$  and  $Z$  boson and leaves photon massless. Details of the mechanism can be found elsewhere [8] but the main point is that it also predicts a new scalar and electrically neutral particle

which is called Higgs boson. The search for Higgs boson lasted few decades before finally in 2012, a new particle was discovered with mass of 125 GeV [6, 7]. In subsequent years, properties of this new particle have been measured and at this point, all measurements agree with Standard Model predictions for Higgs boson.

### 2.1.1 Bottom quarks

Bottom quark was first predicted by Makoto Kobayashi and Toshihide Maskawa in 1974 when extending Cabibbo mixing angle to take into account CP violation observed in neutral K mesons [9]. The name "bottom" was introduced in 1975 by Haim Harari. The bottom quark was discovered in 1977 by the Fermilab E288 experiment team led by Leon M. Lederman through the observation of  $\Upsilon$  resonance [10]. Kobayashi and Maskawa won the 2008 Nobel Prize in Physics for their explanation of CP-violation.

At the LHC, the main production mechanism for b quarks is through strong interaction ( $g \rightarrow bb$ ) and top quark decay ( $t \rightarrow Wb$ ). Every b quark, after production, goes through the process of hadronization, forming one of the color neutral B mesons. B meson decays electromagnetically if produced in excited state to the ground state. Lowest state B mesons decay weakly, resulting in relatively long lifetime of  $\sim 1.5$  ps. Bottom quark can decay either to c quark or u quark. Both of these decays are suppressed by the CKM matrix 2.1. Long lifetime of B mesons makes it possible for them to traverse a substantial distance inside the detector. This fact is used in the creation of various b-tagging algorithms which are taking into account tracks originating from displaced vertices, discussed in Section 5.4.4.

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 0.974 & 0.225 & 0.003 \\ 0.225 & 0.973 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{pmatrix} \quad (2.1)$$

### 2.1.2 W boson

W boson is one of the massive mediators of weak interaction with a mass of  $m_W = 80.1$  GeV. The discovery of W and Z bosons in proton-antiproton collisions at UA1 and UA2 experiments was one of the major successes of the CERN experimental facility. Super Proton Synchrotron was the first accelerator powerful enough to produce W and Z bosons. Both collaborations reported their findings in 1983 [11, 12]. W boson at the LHC is primarily produced through quark-antiquark annihilation. In majority of the cases, W boson decays to quark-antiquark pair (66% of all W boson decays). Other decay channels include creation of a lepton and it's corresponding neutrino ( $\sim 10\%$  per lepton generation). This decay channel was the most important for W boson discovery and it's still essential for W boson detection at hadron colliders despite the large hadronic backgrounds because it includes easily identifiable isolated lepton and significant missing energy.

Detailed study of W boson produced in association with jets at hadron colliders started in 1980s motivated by the top quark searches. Additional jets come from radiation of additional quarks or gluons. However, because they carry color charge, quarks and gluons undergo the process of parton shower and hadronization forming jets in the detector. *Parton shower* is a process in which a high energy colored particle emits a low energy colored particle while *hadronization* is a process in which colored particles combine to form color neutral particles. Parton shower and hadronization cannot be computed analytically, instead have to be modeled using Monte Carlo simulations. As a result of these processes, the number of jets in the final state doesn't necessarily correspond to the number of incoming partons. This becomes relevant when trying to form an inclusive W+jets sample from exclusive (W + 1 jet, W + 2 jets...) samples and certain matching procedure has to be performed in order to avoid double counting. Matching procedure is described in detail in [13].

Many theoretical issues arise when trying to compute cross sections for W+jets processes. Divergences while calculating amplitudes come from emission of low energy particles or collinear jets. These problems are solved by introducing a cut-off called factorization scale. Other divergences come from integrating higher-order loops. Usually this type of divergence is than included into renormalized coupling constant. This procedure, however introduces a certain scale dependence into the result which will be further discussed in Section 2.2.1.

## 2.2 W + b jets at hadron colliders

First theoretical computations of W boson in association with b jets were published in 1993 [14], however only recently enough luminosity has been collected at hadron colliders to be able to make cross section measurements. This process was first interesting as a background to top quark searches and measurements where top quark decays to W boson and a b quark. In past few years, with the Higgs boson discovery, an important open question is whether this new particle also couples to fermions, and in particular to bottom quarks. Standard model Higgs boson branching ratio for decays into a bottom quark-antiquark pair (bb) is  $\approx 58\%$ . Study of this decay channel is therefore essential in determining the nature of the newly discovered boson. The measurement of the  $H \rightarrow bb$  decay will be the first direct test of whether the observed boson interacts as expected with the quark sector, as the coupling to the top quark has only been tested through loop effects. Determination of this coupling requires direct measurement of the corresponding Higgs boson decay. However, the large backgrounds for this measurement make it essential that all the contributing processes including W+b jets are well understood. The result was recently reported by the CMS experiment in the study of Higgs decays to bottom quarks [15, 16]. There are also beyond Standard model searches where contributions from W+b jets process is substantial including some Supersymmetry searches with lepton, b jets and missing energy in the final state.

### 2.2.1 Cross sections at hadron colliders

Determining cross sections for processes at hadron colliders is not an easy task. With proton being a composite object consisting of partons, it is necessary to include its internal structure as well as the diagrams for hard scattering of interest. This essentially means the soft and hard processes are occurring in the same event. Quarks and gluons within proton interact through strong force and are described using quantum chromodynamics. Two processes make it possible to perform calculations within the QCD, *asymptotic freedom* and *factorization theorem*. Since strong force coupling constant  $\alpha_s$  depends on the scale of the process, for high momentum transfers ( $Q >> \Lambda_{QCD} \approx 200\text{MeV}$ ) it becomes sufficiently small to make perturbative expansion in  $\alpha_s$  possible. This feature is called *asymptotic freedom* and it is used to determine the hard process cross section. Figure 2.2 shows the results of the  $\alpha_s$  measurements which is in complete agreement with the QCD predictions of asymptotic freedom.

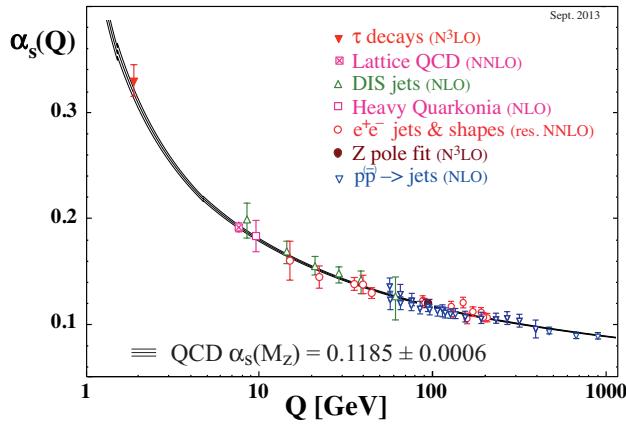


FIGURE 2.2: Summary of measurement of strong coupling constant  $\alpha_s$  [17]

*Factorization theorem* is introduced to separate the two contributions in the cross section calculation, the contribution from the hard process calculated using perturbative QCD and the contribution from the internal structure of the proton. This means that hard scattering between partons is independent from the proton internal structure. Factorization scale is introduced as a cut-off below which perturbative QCD calculation cannot be

performed. Scale dependence of strong coupling constant causes the hard and soft part of the process to happen at different time scales. The calculation the cross section for a

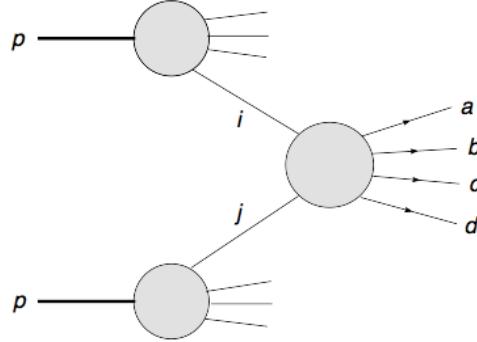


FIGURE 2.3: Drawing of a proton-proton collision.

process with two protons in the initial state and some interesting final state which we call X requires the following steps (as described in [18]):

1. Identify the leading order partonic processes that contribute to X
2. Calculate the corresponding hard scattering cross section
3. Determine the appropriate PDFs for initial state partons
4. Make a specific choices for factorization( $\mu_F$ ) and renormalization( $\mu_R$ ) scales
5. Perform integration over the fraction of momentum available for a given parton(x)

The cross section at hadron collides is thus a convolution of the hard scattering perturbative cross section and two incoming parton distribution functions.

$$\sigma_{AB} = \sum_{n=1}^{\infty} \alpha_s^n(\mu_R^2) \sum_{i,j} \int dx_1 dx_2 f_{i/A}(x_1, \mu_F^2) f_{j/B}(x_2, \mu_F^2) \sigma_{ij \rightarrow X}^{(n)}(x_1 x_2 s, \mu_R^2, \mu_F^2) \quad (2.2)$$

Equation 2.2 shows cross section perturbation series in  $\alpha_s$ ,  $n$  denotes the order of the series where  $n = 1$  is leading order,  $n = 2$  is next to leading order, etc. Hard process cross section between two partons  $\sigma_{ij \rightarrow X}^{(n)}$  is computed in the framework of perturbative

QCD and depends on  $s$  which is squared center of mass energy. Two functions denoted with  $f_{i/A}$  and  $f_{j/B}$  correspond to the probability density that parton  $i(j)$  with proton momentum fraction  $x_1(x_2)$  will be found inside a proton. and are called parton distribution functions (PDFs). These functions cannot be computed using perturbative QCD because momentum transfer values are small and the coupling constant becomes large. This phenomenon is called *confinement* and it requires different treatment for the quarks inside the proton. Internal structure of a proton is described using parton distribution functions(PDF) which are determined through deep inelastic scattering experiments. Sum over all combinations of partons has to be computed. Integral over available phase space for proton fraction momentum  $dx$  is usually carried out by simulations. Here  $\mu_F$  represents *factorization scale* and  $\mu_R$  is *renormalization scale* for running coupling constant. They are arbitrary cut-offs used to remove nonperturbative effects and be able to make perturbative calculations. If cross section is computed in full series,  $\mu_F$  and  $\mu_R$  should cancel out, and scale dependence should disappear. However, since fewer orders are used and some residual scale dependence is still present. This dependency can be used to estimate the contribution of the missing orders in the series.

Factorization scale controls soft and collinear emissions that can spoil the perturbative calculation. These emissions are than absorbed into the PDF for transverse momenta below  $\mu_F$ . Parton distribution functions for each of the partons inside a proton is shown in Using DGLAP equations, it is possible to evolve the PDFs for any momentum transfer value which is described in detail in [18] and making the factorization cut-off possible. PDF evolutions is shown figure 2.4 for one specific PDF function (MSTW) at momentum transfer values of  $Q^2 = 10 \text{ GeV}^2$  and  $Q^2 = 10^4 \text{ GeV}^2$  which corresponds to the typical momentum transfers for W boson production. At high momentum transfer values, sea quarks and gluons carry much larger portion of the proton momentum and  $b$  quark distributions become relevant. Renormalization scale is another cut-off used to control divergences from the integration of high momentum loops in parton cross sections. If the momenta is larger than  $\mu_R$ , the divergencies are absorbed in a redefined coupling constant  $\alpha_s$  and the cross section calculation becomes finite. This approach is common in renormalizable

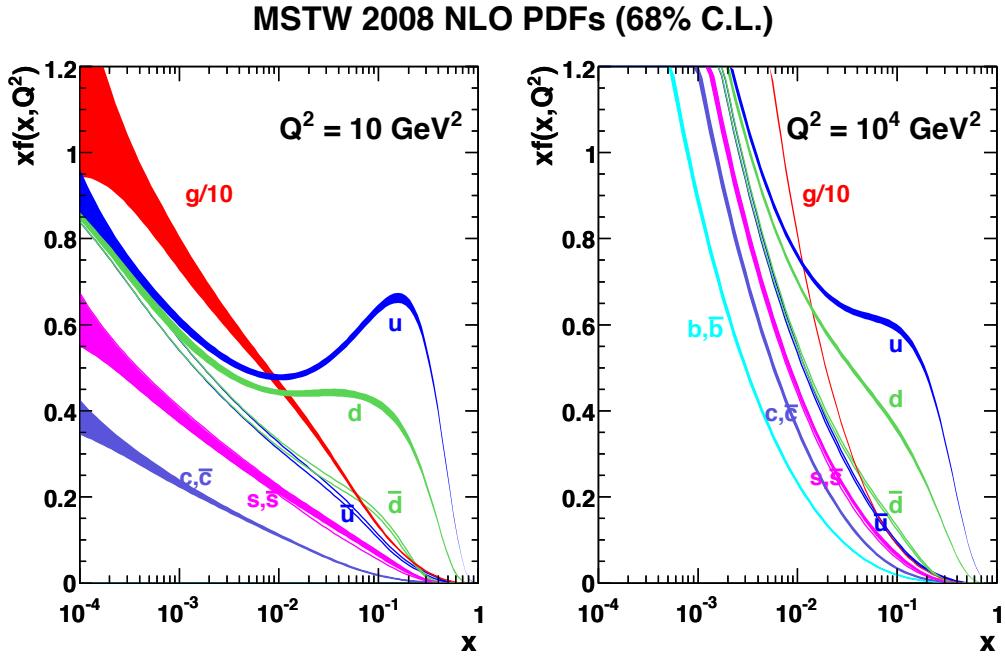


FIGURE 2.4: Parton distribution functions calculated by the MSTW group for  $Q^2 = 10 \text{ GeV}^2$  (left) and  $Q^2 = 10^4 \text{ GeV}^2$  (right) [19]

field theories. However, such result depends on the renormalization scale of the and the resulting dependency can be calculated using renormalization group equation (RGE) [20].

Usually factorization and renormalization scales are chosen to be identical and close to the scale of the process in question ( $\mu_F = \mu_R = \mu_0 \approx Q$ ). The choice of the scale in case of W boson production is usually around the mass of W boson. Taking into account specific kinematical properties of each event, a dynamic scale can be defined, for example  $\mu_0^2 = m_w^2 + p_{T,W}^2$ . In case of W boson and b jets production, adding b quark mass or transverse momentum to the scale is also a viable option.

Figure 2.5 shows some interesting Standard model cross sections in proton-proton and proton-antiproton collisions as a function of a center of mass energy. All cross sections have been computed to the NLO order using the above described procedure.

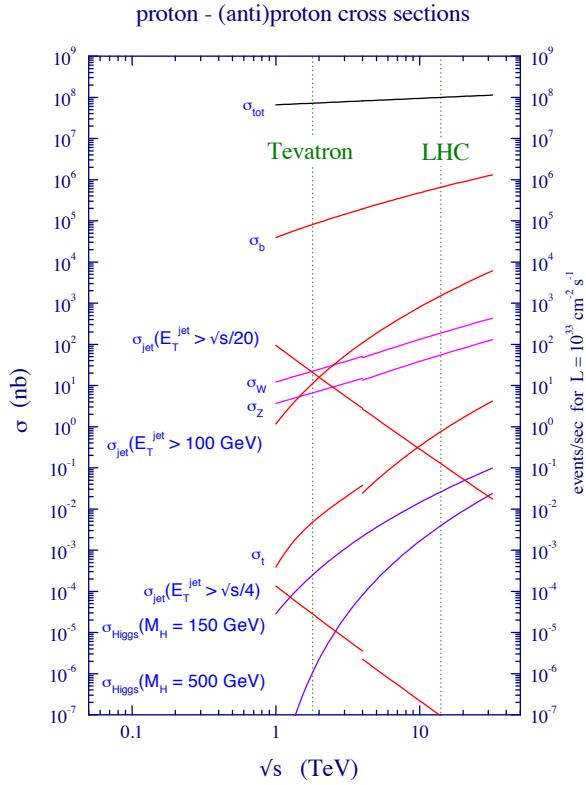


FIGURE 2.5: Standard model cross sections as a function of center of mass energy.[18]

### 2.2.2 Contributions to Wbb cross section

From theoretical point of view, calculations of W+b jets processes can be divided into two categories: only light quarks in the initial state shown in figure 2.6 (four flavour scheme - 4FS) and b quark in the initial state, usually called five flavor scheme (5FS) shown in Figure 2.9. Additional contribution to Wbb production at hadron colliders comes from double parton interactions where a W boson and a pair of b quarks is produced in different hard process inside the same collision as shown in Figure 2.10. This contribution will be discussed in Section 2.2.2.1.

The rationale behind using 4FS or 5FS is discussed in detail in [21]. Four flavor scheme approach assumes that bottom quarks are heavy and can only be created as pairs in collisions with high momentum transfer or as a decay product of t quark. Heavy quarks are not included in the initial state and their parton distribution function is set

to zero which means an effective theory is created where heavy quarks do not enter the computation of running coupling and the evolution of PDFs. If it happens that the scale of the process is much higher than the mass of the b quark, for example in the production of massive bosons, large logarithms of the type  $\log(Q^2/m_b^2)$  appear and can spoil the convergence of a fixed order perturbative expansion and introduce large scale dependence into the final result. In five flavor schemes calculations include b quark in the initial state allowing for some new and simpler processes to become available. These calculations allow resummation of possibly large logarithms of type  $\log(Q^2/m_b^2)$  into the b quark parton distributions function possibly transforming some higher order calculations into much simpler leading order calculations. The result in [21] shows that at the LHC 4-flavor calculations are well behaved and two schemes are in good agreement. The typical size of the possibly problematic logarithms in four flavor scheme at hadron colliders is not large enough to spoil convergence. On the other hand, five flavor scheme is less dependent on the scale of the process and show smaller uncertainties which is in general very good for predictions of inclusive observables.

First leading order calculations for associated production of a W boson and heavy quarks at hadron colliders were presented in 1993. Feynmann diagram for leading order  $W + 2 \text{ b jets}$  production is shown in Figure 2.6. Exact leading order matrix element has been computed and higher order corrections were estimated using Monte Carlo. Their results are summarized in the Figure 2.7 where the differential cross section for  $W+2 \text{ b jets}$  as a function of a leading b jet  $p_T$  is shown. Two scale choices have been studied, first one with  $\mu_0 = M_{bb}$  which is the invariant mass of the dijet system and is represented with solid line. Second choice is  $\mu_0 = m_W + p_T^W$  and is represented with the dotted line. Looking at the normalizations of two diagrams, the difference is clearly visible which indicates a strong total cross section scale dependence. However, the shape of the differential cross section shows the same behavior in both cases which means that the scale only affects total cross section.

Later development of theoretical calculations was strongly motivated by reducing the scale dependence of the result and it included adding additional partons to the final

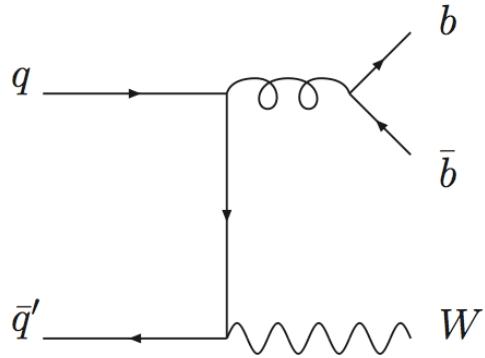


FIGURE 2.6: Leading order Wbb Feynmann diagram

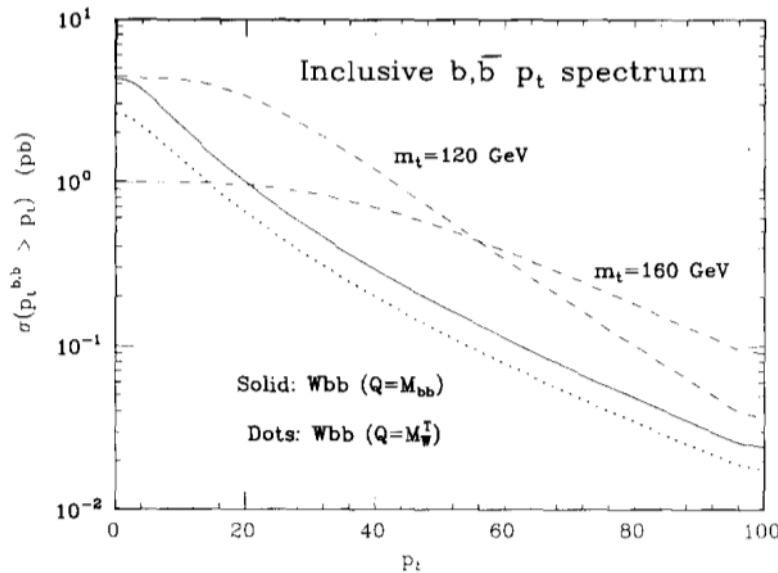


FIGURE 2.7: Scale dependence of Wbb cross section

state. This was a first step towards the full NLO calculation. The only thing missing was taking into account the loop effects. This approach made it possible to access some previously inaccessible kinematics, however at the expense of introducing additional scale dependence. The list of new final states is simple and it includes  $Wbbq$ ,  $Wbbq\bar{q}$ ,  $Wbb\bar{q}q'\bar{q}'\dots$ . For the measurements at the LHC in particular, calculations for new initial states  $qg$  and  $gg$  were of great importance. First results for  $W+2$  jets were published in [22]. Additional calculations were shown in [23] for up to six additional jets in the final state. Although these processes are suppressed by an additional  $\alpha_s$  factor, the gluon PDF inside a proton

is much larger than anti-quark, so this production mechanism is significant at the LHC energies.

First full NLO calculations were published in 2006 [24]. Events with b jet pair in the final state were selected, with momentum of the dijet system  $p_T > 15$  GeV and a pseudorapidity less than 2. The results were shown for two categories, inclusive and exclusive, depending on the treatment of extra jets. In the inclusive case events with additional jets were included, while in the exclusive case exactly two jets were required. Figure 2.8 shows the overall scale dependence of LO, NLO inclusive and NLO exclusive total cross-sections, when both renormalization scale and factorization scale are varied independently between  $\mu_0/2$  and  $4\mu_0$  (with  $\mu_0 = m_b + M_W/2$ ), including full bottom-quark mass effects. NLO cross sections have a reduced scale dependence over most of the range of scales shown, and the exclusive NLO cross-section is more stable than the inclusive one especially at low scales. The effect of the b quark mass has been shown to affect the total NLO cross section on the order of  $\approx 8\%$ . This is expected to be small when considering well separated jets.

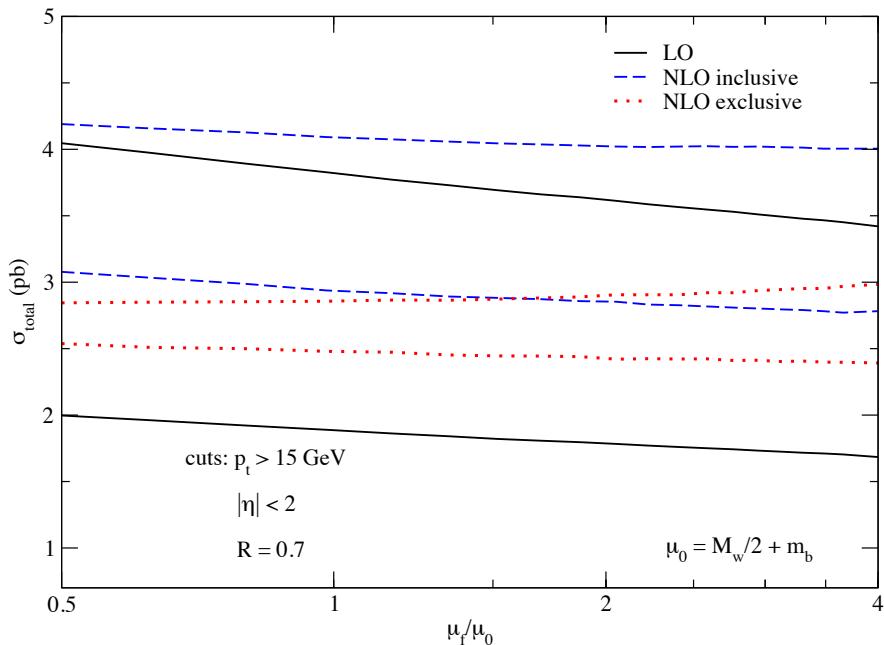


FIGURE 2.8: Wbb NLO scale dependence[24]

New results published in 2007 explored in particular NLO corrections for events with W boson and two jets where at least one is b-tagged. It was shown that for LHC the correction factor is  $\approx 1.9$ . This paper was interesting in particular for its study of soft and collinear topologies, where two b quarks merge into one. Additionally, b quark in the initial state was considered giving rise to the processes like  $bq \rightarrow Wbq'$  shown in figure 2.9. Parton distribution function for b quark needed to be determined perturbatively using DGLAP equations. Other approach is to consider a gluon in the initial state which then splits to  $b\bar{b}$ .

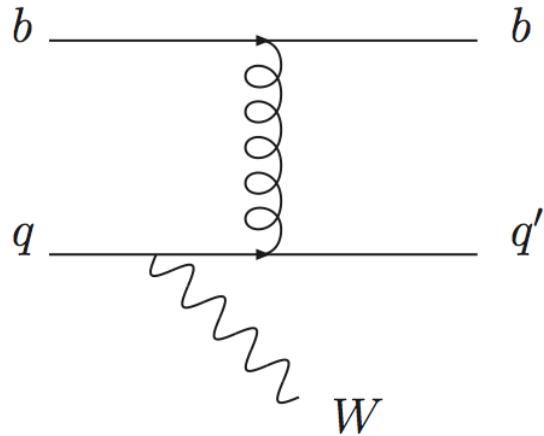


FIGURE 2.9: Wbb production within 5 flavor scheme

### 2.2.2.1 Double parton scattering

Multiple parton interactions happen due to composite nature of the proton. Usually inside a proton, only one parton has significant fraction of proton momentum  $x$  to produce a hard scattering. However, sometimes can happen that two such partons exist which results in two hard scatterings in the same collision. This phenomenon is called Double Parton Scattering (DPS) and is shown in Figure 2.10. In the framework of this thesis, two partons are responsible for creation of a W boson and other two for creation of pair of b jets.

Double parton scattering cannot be modeled in the framework of perturbative QCD, but it is approximated using simulations. The phenomenology of DPS starts from the

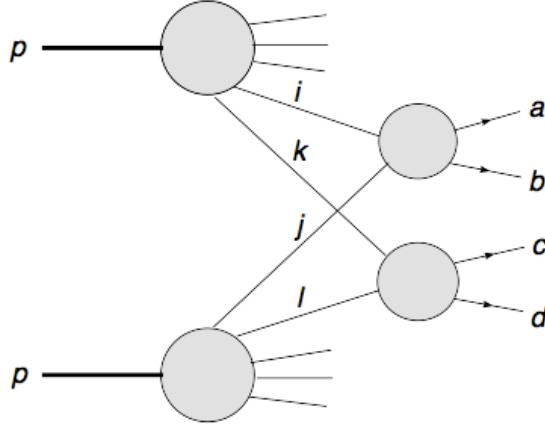


FIGURE 2.10: Double parton scattering

assumption that factorization between two hard processes is possible, as well as factorization between hard processes and proton kinematics. Cross sections for hard scatterings are computed separately of each pair of partons. However, instead of using regular parton distribution functions, a new set of distribution functions has been defined which are called Double Parton Distribution Functions (dPDFs). Factorized cross section for two hard processes A and B to happen in proton-proton scattering can be written as:

$$\sigma_{(A,B)}^{DPS} \sim \sum_{i,j,k,l} \int dx_1 dx_2 dx'_1 dx'_2 d^2 b \Gamma_{ij}(x_1, x_2, b; Q_1, Q_2) \sigma_{ik}^A(x_1, x'_1) \sigma_{jl}^B(x_2, x'_2) \Gamma_{kl}(x'_1, x'_2, b; Q_1, Q_2) \quad (2.3)$$

Parton level cross sections are denoted with  $\sigma_{ik}$ , for hard process between partons  $i$  and  $k$ , and  $\sigma_{jl}$  for hard process between partons  $j$  and  $l$ . These are the same as for single parton scattering and are known for most of the processes of interest today. Quantity  $\Gamma_{ij}(x_1, x_2, b; t_1, t_2)$  represents double parton distribution function which describes the probability of finding a parton  $i$  with momentum fraction  $x_1$  at scale  $Q_1$  inside a proton together with a parton  $j$  with momentum fraction  $x_2$  at scale  $Q_2$ . Another parameter in this distribution function is  $b$  which describes transverse distance between two partons. Scales  $Q_1$  and  $Q_2$  correspond to characteristic scales of hard processes  $A$  and  $B$ . For example in the framework of this thesis W boson production would correspond to process

$A$  and production of two b jets would correspond to process  $B$ . This study is described in detail in [25]. Usually, it is assumed that  $\Gamma_{ij}(x_1, x_2, b; t_1, t_2)$  can be decomposed into two components, longitudinal and transversal in the following way:

$$\Gamma_{ij}(x_1, x_2, b; t_1, t_2) = D_h^{ij}(x_1, x_2; t_1, t_2)F_j^i(b) \quad (2.4)$$

The interpretation of the function  $D_h^{ij}(x_1, x_2; t_1, t_2)$  within QCD is the probability of finding parton  $i$  with scale  $Q_1$  and parton  $j$  with scale  $Q_2$ . These functions cannot be determined using perturbative QCD, thus good modeling and to correctly take into account correlations between longitudinal momenta and transverse position is essential to making accurate cross section predictions.

More details on how to determine dPDFs can be found in [26], but in the simplest case  $D_h^{ij}(x_1, x_2; t_1, t_2)$  can be taken as a product of single parton distribution functions taking into account effects like  $x_1 + x_2 < 1$ . Since  $F_j^i(b)$  is the only part of  $\sigma_{(A,B)}^{DPS}$  that depends only on  $b$ , integration over  $b$  can be performed giving an effective cross section  $\sigma_{eff}$  which is related to the size of the proton and can be seen as an effective area of the interaction. This approach yields a simplified expression for the double parton scattering cross section:

$$\sigma_{(A,B)}^{DPS} \sim \frac{1}{\sigma_{eff}} \sigma_{(A)}^{SPS} \sigma_{(B)}^{SPS} \quad (2.5)$$

Here  $\sigma_{(A)}^{SPS}$  and  $\sigma_{(B)}^{SPS}$  are single parton scattering cross section which can be obtained using equation 2.2.

However, this factorized approach does not take into account some simple correlations like how finding a quark of some flavor affects the probability of finding another quark with the same flavor. While for some simple cases with low parton momentum fractions this factorized approach may give accurate results, for more complicated cases like calculating fiducial cross sections, acceptance cuts can spoil the equation. Thus, a simulation of the full kinematical effects is necessary.

First measurements of  $\sigma_{eff}$  have been performed by the AFS collaboration at the ISR (CERN) which was  $\sigma_{eff} \sim 5$  mb at 63GeV. Both CDF and D0 collaborations at Tevatron reported  $\sigma_{eff} \sim 15$  mb which is roughly 20% of the total  $p\bar{p}$  cross section at Tevatron energies. Their data also shows no sign of dependence on  $x$  in their measured  $\sigma_{eff}$  in the  $x$  ranges accessible. Later measurements performed by ATLAS and CMS collaborations are in reasonable agreement with previous results. All results are summarized in Figure 2.11.

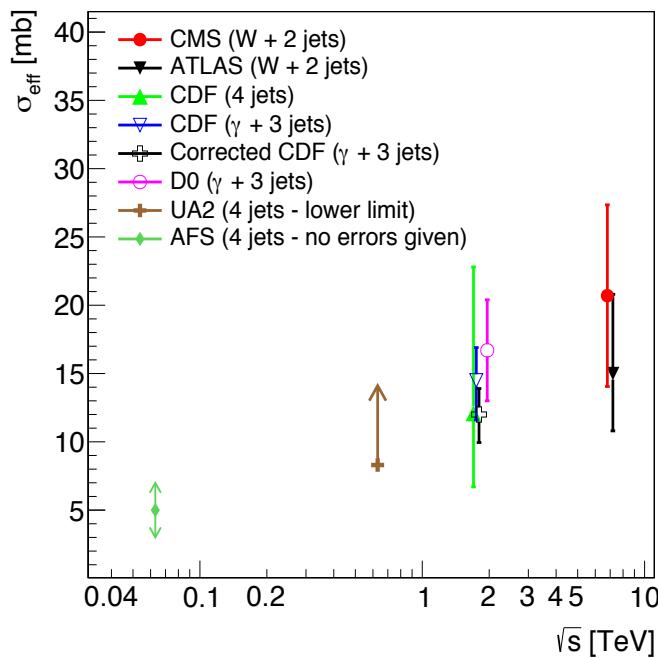


FIGURE 2.11: Center of mass energy dependence of  $\sigma_{eff}$  as reported from different collaborations. All these measurements use different approaches to estimate  $\sigma_{eff}$ . [27]

Double parton scattering measurement at CMS is performed by selecting the events with a W + 2-jet final state where one hard interaction produces a W boson and another produces a dijet [28]. The W + 2-jet process is attractive because the muonic decay of the W provides a clean tag and the large dijet production cross section increases the probability of observing DPS. Events containing a W + 2-jet final state originating from single parton scattering (SPS) constitute an irreducible background. Results were obtained by performing a template fit to two uncorrelated variables: the relative  $p_T$  balance between

two jets ( $\Delta p_T$ ) and the angle between W boson and a dijet system. Obtained results again show that contribution of DPS to total cross section is  $\sim 20\%$  which is in good agreement with previous Tevatron results. The DPS contribution in the case of W + 2 b jets is estimated to be  $\sim 15\%$ . [29]

## 2.3 Previous measurements

Previous measurements of a W boson produced in association with b quarks have been performed on different experiments. However, the final states and phase space used in these measurements were different, which means that the results cannot be directly compared, but they can be compared with theoretical predictions. This process was measured for the first time at Tevatron with D0 and CDF experiments at  $\sqrt{s} = 1.96$  TeV. The CDF collaboration published its result in 2009 and the cross-section measured is that of “jets from b-quarks produced with a W boson” [30]. The event selection is based on reconstructing a leptonically decaying W boson, and one or two jets where at least one has to be b-tagged. Events with jets from light quarks are vetoed with a cut on the secondary vertex mass. Contribution of other background events containing a b quark in final state (e.g. events with top quark) is estimated using Monte Carlo simulations. The measured cross section is 2.8 standard deviations higher than corresponding theoretical prediction.

D0 collaboration published their result in 2012. with somewhat different phase space definition[31]. The difference with respect to the CDF measurement consists in the inclusion of the events with 3 jets and reduced pseudorapidity range in which the measurement was performed. The measurement technique is similar to that of CDF, although b-tagging algorithms were slightly different. The measured cross section was in good agreement with the Standard model prediction.

First measurements at the LHC were published by the ATLAS collaboration based on 36/pb of integrated luminosity at  $\sqrt{s} = 7$  TeV. One year later they improved their measurement using  $4.6/\text{fb}^{-1}$  [32]. Selected events contain one reconstructed electron or

muon, significant amount of missing transverse energy and one or two jets where exactly one is b-tagged. The phase space is divided in two regions, depending on the number of jets. Events with exactly 2 b jets and events with more than 2 jets are vetoed in order to suppress background events from top quark decay. The results are shown in Figure 2.12. The cross section measurement in the one jet region shows an excess corresponding to 1.5 standard deviations. In the two jet region, the measured cross section is in good agreement with theoretical predictions. A differential cross section measurement as a function of leading b jet transverse momentum has been performed for the first time and shown in figure 2.13. The cross section measurement in the one jet region is again higher than NLO predictions but within theoretical and experimental uncertainties. The cross section measured for the events with two jets is in good agreement with the theoretical prediction.

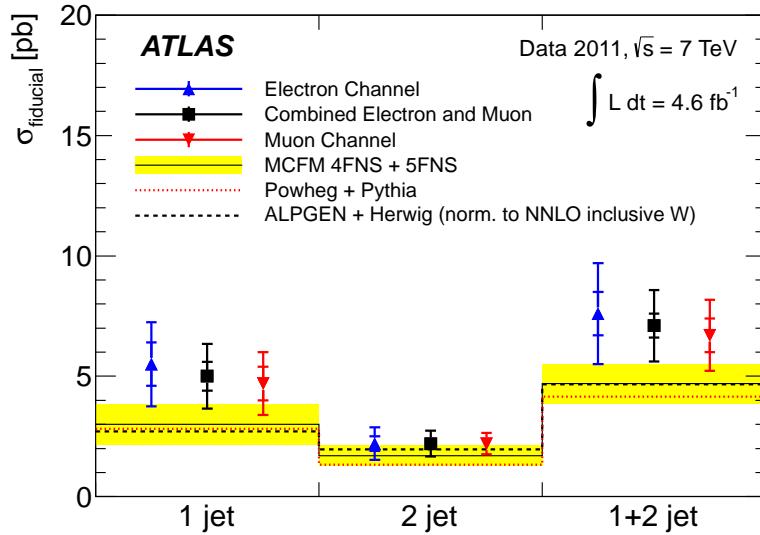


FIGURE 2.12: Measured fiducial cross-sections in the electron, muon, and combined electron and muon channels. The cross-sections are given in the 1-jet, 2-jet, and 1+2-jet fiducial regions. [32]

CMS collaboration published its results corresponding to data collected during 2011 at 7 TeV corresponding to  $5\text{fb}^{-1}$  of data. Selected events contained a muon and missing transverse energy in the final state, together with two b-tagged jets. All additional lepton

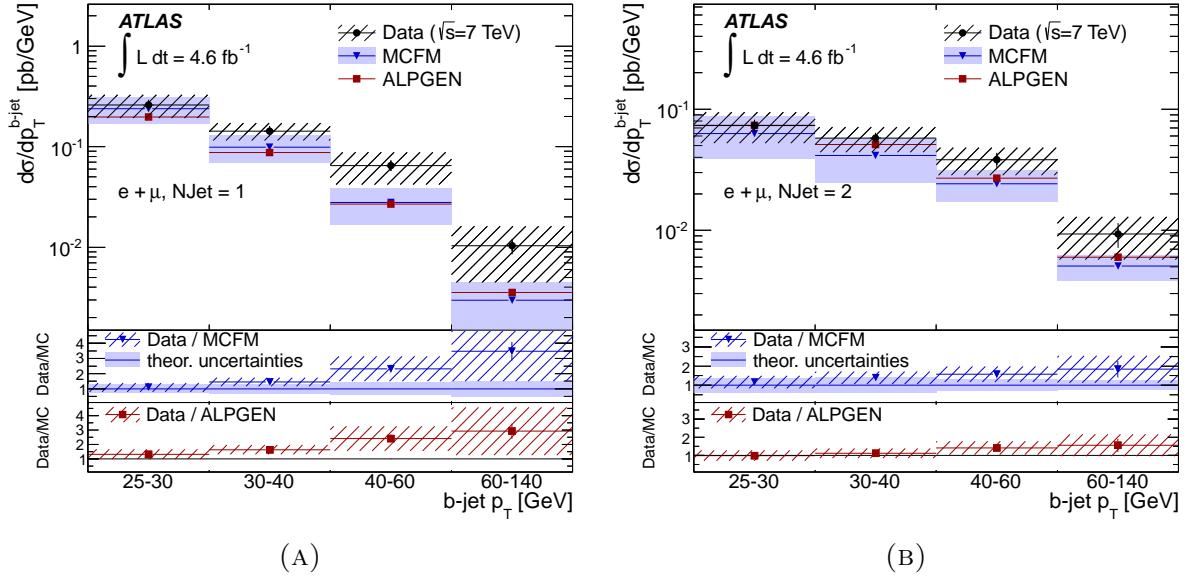


FIGURE 2.13: Measured differential  $W+b\text{-jets}$  cross-sections as a function of leading b-jet  $p_T$  in the 1-jet (2.13a) and 2-jet (2.13b) fiducial regions, obtained by combining the muon and electron channel results. [32]

and jet activity was vetoed to reduce the background contributions. Figure 2.14 shows the leading jet transverse momentum distribution used for signal extraction. The measured cross section is in excellent agreement with the Standard model prediction. [29]

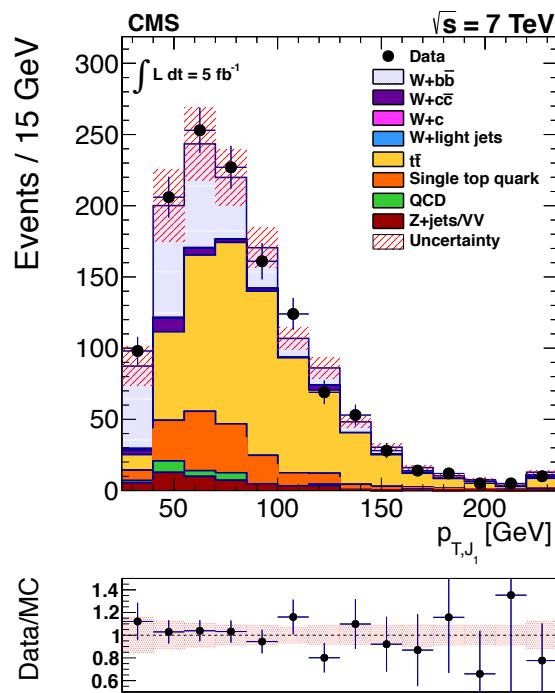


FIGURE 2.14: Leading jet transverse momentum distribution used for signal extraction in  $W+bb$  total cross section measurement with the CMS experiment. [29]

# Chapter 3

## Large Hadron Collider

CERN is the largest particle physics laboratory in the world, located near the city of Genava, on the French-Swiss border. It was founded in 1953 by 12 countries and today it has 21 member states. Its main function is to provide particle accelerators and infrastructure for high energy physics experiments. Major physics results at CERN include the discovery of neutral currents, discovery of W and Z bosons, creation of antihydrogen atom and direct observation of CP violation among others.

Current accelerator complex is a chain of smaller accelerators with increasingly higher energies of which the largest one is Large Hadron Collider (LHC) (Figure 3.1). LHC is a proton-proton collider which is also able to deliver lead-lead and proton-lead collisions. Protons are obtained by taking hydrogen atoms and stripping them of the orbiting electrons, and are accelerated by a small linear accelerator Linac2 to 50 MeV and injected to PS Booster. After reaching 1.4 GeV, protons are injected to Proton Synchrotron and accelerated to 25 GeV. Next accelerators in chain are Super Proton Synchrotron (SPS) with energy of 450 GeV, and Large Hadron Collider with designed beam energy of 7 TeV.

In this chapter, we will briefly go through the motivation for the LHC design, building blocks of the accelerator together with its performance during the past few years.

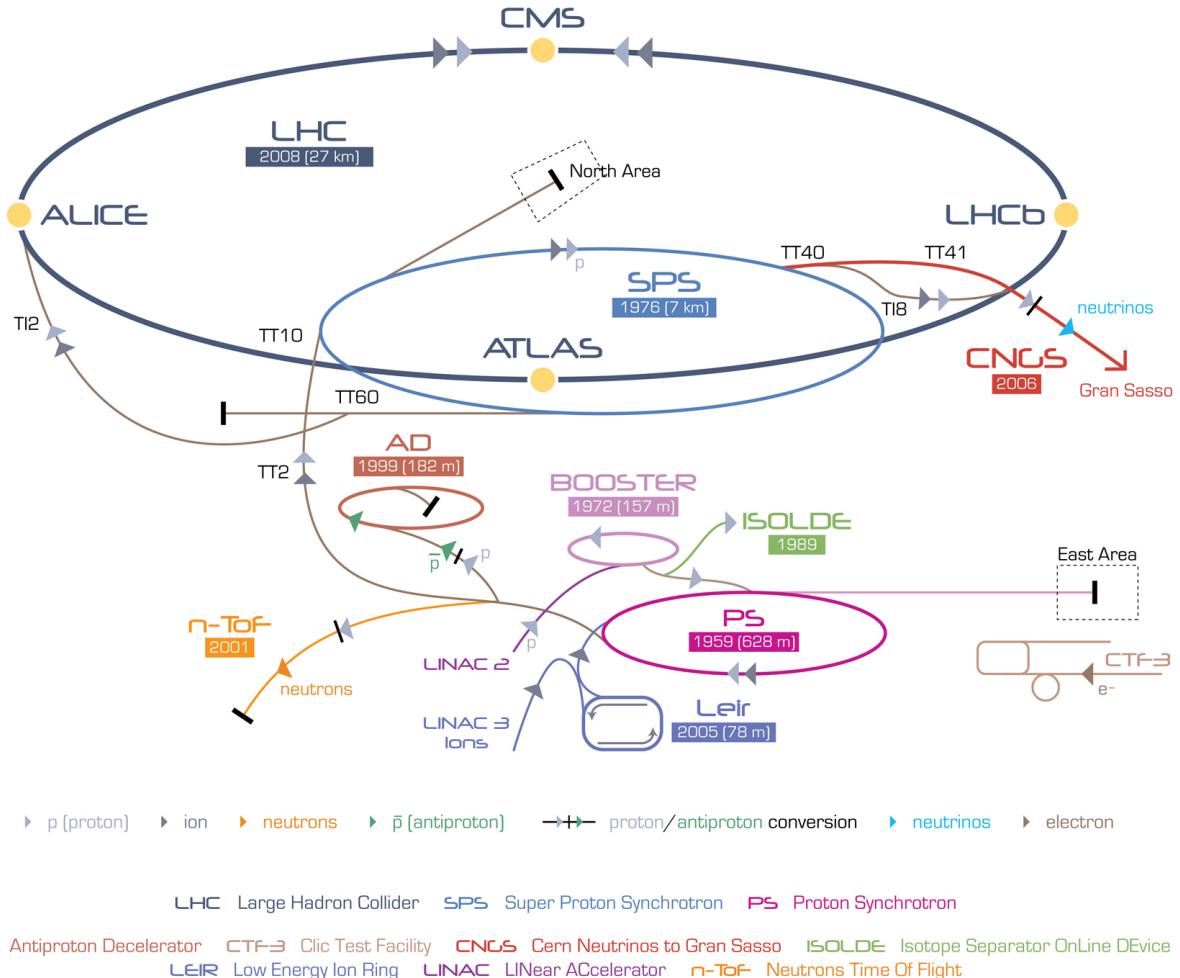


FIGURE 3.1: Schematics of Large Hadron Collider

### 3.1 Physics goals for the LHC

The Standard model of elementary particles describes nicely all known particles and interactions, however there are still some unanswered questions. One of the major open questions was the existence of Higgs boson which was recently answered with the discovery of a new boson at 125 GeV. In order to be able to claim such a discovery, all known Standard model processes have to be well measured and the behavior of the experimental device has to be well understood. These requirements lead to many precision measurements which determined precisely cross sections, couplings, masses and other parameters within the Standard model. Any deviation from predicted values can be an evidence for the existence of physics beyond Standard model. One of the questions that remain open

is the unification of fundamental forces. One attempt to achieve this goal is the theory of supersymmetry which predicts that each particle has its heavier supersymmetric partner and at high energies could unify strong and electroweak forces. If the theory of supersymmetry is correct, lightest supersymmetric particles would be stable and could be detected at the LHC. Such particle would also be a great candidate for the dark matter considering it would interact only weakly and as such would fit nicely into the present dark matter theories. On the other hand, the problem of matter-antimatter asymmetry could be addressed trying to discover why is the world built only of matter. Other theories that involve extra dimensions, bound states of quarks and leptons and other exotic models can be tested as well. Within LHCs heavy ion program, lead-lead and lead-proton collisions were performed in which a state called quark-gluon plasma is produced that resembles the conditions in the early universe.

During the past few years, various models for new physics have been extensively tested, and new exclusion limits have been set. After three years of data taking at 7 and 8 TeV, and a shutdown period of two years, LHC is now almost ready to deliver collisions at record energies of 13 TeV which could hopefully show signs of new physics.

## 3.2 Design of the LHC

The LHC is located inside a 27 km tunnel, which lies between 45 m and 170 m below the ground surface and previously housed LEP accelerator. Beams circulating inside the LHC, collide at four interaction points. At each of these points, a detector has been built to record the products of particle collisions. This thesis was done using data collected with the CMS (Compact Muon Solenoid) detector [33]. Another detector with the same purpose but different design is the ATLAS (A Toroidal LHC Apparatus) detector located at the opposite side of the LHC ring [34]. These two are so called multiple purpose particle detectors, which cover a wide range of physics topics, from searches for Higgs boson and supersymmetry to Standard model precision measurements. ALICE (A Large

Ion Collider Experiment) is designed to study quark-gluon plasma by measuring mainly lead-lead collisions and in addition lead-proton and proton-proton collisions [35]. LHCb (LHC Beauty) is aimed towards B physics, by studying decays of B mesons [36]. Two other experiments TOTEM and LHCf are placed away from the interaction point to measure the collision products along the beam direction.

The LHC is made out of nearly 9600 different magnets, including dipoles, quadrupoles, sextupoles, octupoles, etc. The largest portion of the accelerator is made out of 1232 dipole magnets which are made using superconducting niobium-titanium (NbTi) cables that undergo a phase transition to a superconductive state at 9.2K. In order to achieve superconductivity and withstand very high currents (11850 A), cables have to be cooled with superfluid helium to less than 2K. They create large magnetic fields, which can reach 8.2T which bend the proton beams around the ring. Other higher order magnets are used to focus and correct the beam.

Two proton beams are counter-circulating inside a single cryogenic structure which requires opposite magnetic field direction for each of the beams in order to be steered along the same circumference. One of the LHC dipole magnets is shown schematically in Figure 3.2 together with the drawing of the magnetic field inside the dipole.

Beams in the LHC are injected in series of bunches separated by a vacuum gaps, with each bunch having more than  $10^{11}$  protons. Bunches are arranged in trains of 72 bunches with 25 ns spacing between them and 12 empty bunches between trains. Acceleration is provided by the radio frequency superconducting cavities (RF). It takes approximately 20 minutes for the beams to be accelerated from the injection at 450 GeV to the full beam energy. Moreover, RF chambers provide a small corrections of the order of  $\sim 7$  keV per turn to the beam due to the energy loss from synchrotron radiation. After the acceleration, beams are tuned at the interaction points to achieve intersection. Peak collision rate of 40 MHz is achieved when collisions happen at every bunch crossing. Beams are squeezed to a transverse size of  $\sim 17 \mu\text{m}$  at the interaction point in order to maximize the probability of collision.

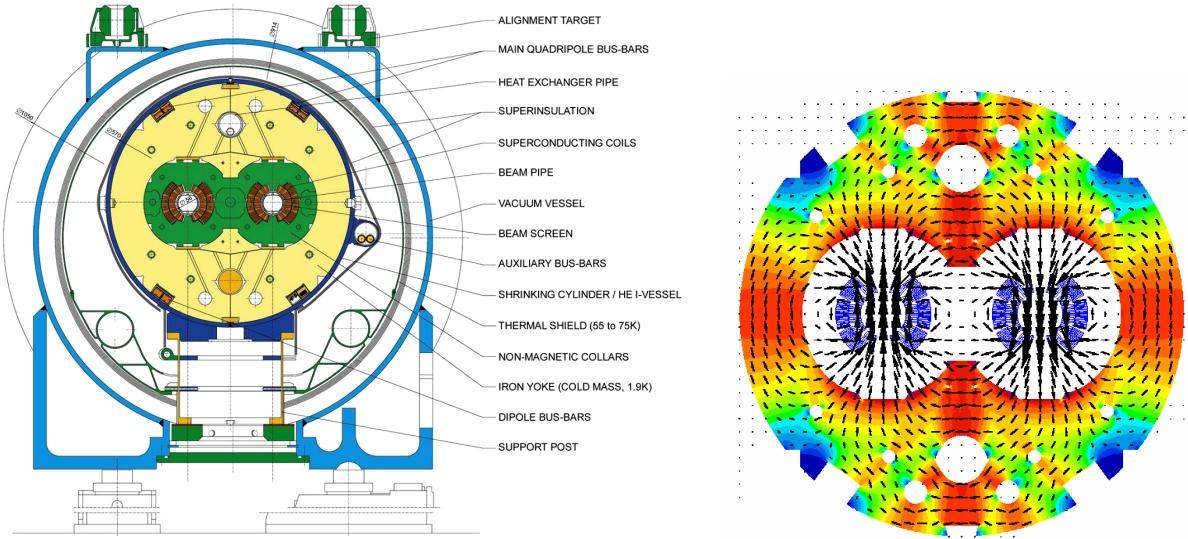


FIGURE 3.2: Schematics of Dipole magnets [37, 38]

### 3.3 LHC performance

Since the start of the LHC in 2009, there were three years of machine operation, which yielded many physics results among which the discovery of Higgs boson should be highlighted. First year of operation, when LHC center of mass energy was 900 GeV, was devoted to commissioning and understanding machine characteristics with the emphasis on safety and tests of the machine protection systems. In 2011 new energy of 7 TeV, breaking the record of 1.96 TeV previously set by Tevatron. This number was increased once again in 2012 with center of mass energy going to 8 TeV.

High bunch intensity with 50 ns bunch spacing was used in order to get a good instantaneous luminosity performance. This came at a cost of high number of collisions in one bunch crossing (pile-up) which was around 12 collisions during 2011, and in some cases this number went as high as 20 interactions. With the increase of instantaneous luminosity in 2012, number of pile-up interactions was on the average around 30. Besides proton-proton collisions, LHC successfully delivered lead-lead ion runs in 2010 and 2011, primarily for the ALICE experiment, but also for CMS and ATLAS. On top of that, in

the beginning of 2013 there was a successful proton-lead run performed for the first time. LHC design together with the 2012 operations parameters are shown in Table 3.1.

TABLE 3.1: LHC performance in 2012 together with design performance[37]

Parameter	Design value	Value in 2012
Beam energy [TeV]	7	4
Bunch spacing [ns]	25	50
Number of bunches	2808	1374
Protons per bunch	$1.15 \times 10^{11}$	$1.6-1.7 \times 10^{11}$
Peak luminosity [ $\text{cm}^{-2}\text{s}^{-1}$ ]	$1 \times 10^{34}$	$7.7 \times 10^{33}$
Max. number of events per bunch crossing	19	$\approx 40$
Stored beam energy [MJ]	362	$\approx 140$

Luminosity ( $L$ ) indicates the number of collisions per unit of time over the interaction cross section  $\sigma$ :

$$L = \frac{1}{\sigma} \frac{dN}{dt} \quad (3.1)$$

Luminosity for collider experiments is connected to beam parameters:

$$L = \frac{n \cdot N^2 f}{A_{eff}} \quad (3.2)$$

where  $n$  is a number of bunches with  $N$  protons inside, that are colliding at the revolution frequency  $f$  and effective beam area  $A_{eff}$ . The amount of data collected in a certain period of time is called total integrated luminosity and is defined as:

$$\mathcal{L} = \int L dt \quad (3.3)$$

During the Run 1 data taking period, the LHC delivered around  $24\text{fb}^{-1}$  of data (figure 3.3) at the energy of 8 TeV with highest instantaneous luminosity of  $8 \cdot 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ . Some of the LHC performance highlights are listed in table 3.2.

Following a two year shutdown, LHC is anticipating operations at even higher energies of 13 TeV and later 14 TeV. The long term plan includes even higher peak luminosities, installation of the new injector complex and later the beginning of HL-LHC era. The

TABLE 3.2: LHC performance highlights

Max. luminosity delivered in one fill	$237 \text{ pb}^{-1}$
Max. luminosity delivered in 7 days	$1.35 \text{ fb}^{-1}$
Longest time in stable beams (2012)	22.8 hours
Longest time in stable beams over 7 days	91.8 hours (55%)

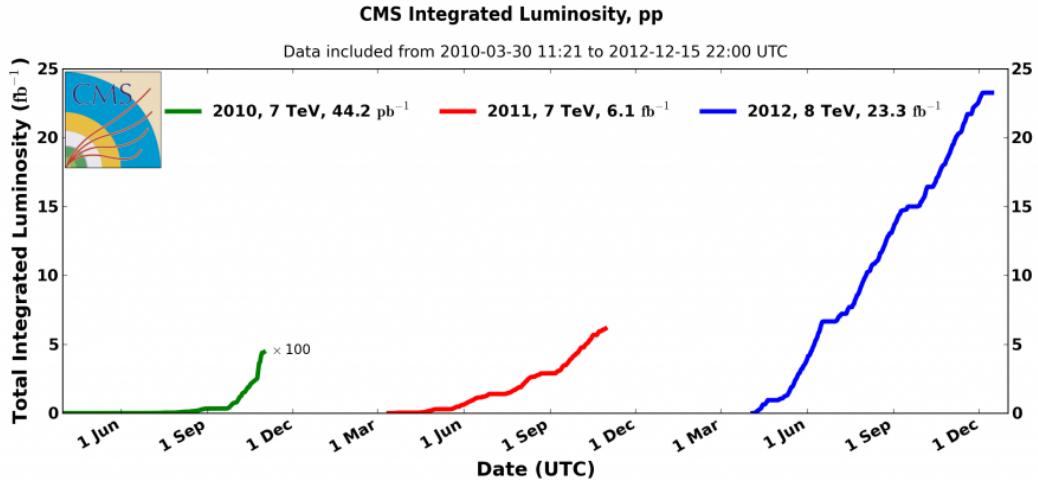


FIGURE 3.3: Luminosity delivered to the CMS experiment

timeline will, of course, be highly affected by the performance and results of the next run.



# Chapter 4

## Compact Muon Solenoid

Compact Muon Solenoid (CMS) is a general purpose detector designed to cover a wide range of physics topics at the LHC with a layered design approach and coverage of a large portion or the solid angle around the interaction point. Tracking system and calorimeters are placed inside a large soleonid in order to improve the resolution of the momentum and energy measurements. Detectors outside the solenoid are aimed primarily to detect muons. A drawing of the CMS detector is shown in Figure 4.1.

The motivation for the CMS design with respect to its purpose in the LHC program is a very good muon identification and good momentum resolution over wide range of phase space and unambiguous determination of muon charge. Very good inner tracking system allows for detection of charged particles and high efficiency offline b quark and  $\tau$  tagging. Other important requirements, specially for Higgs searches, is diphoton mass resolution, and photon and electron identification and isolation at high energies. CMS detector with its design meets all these requirements as is shown in following sections of this chapter.

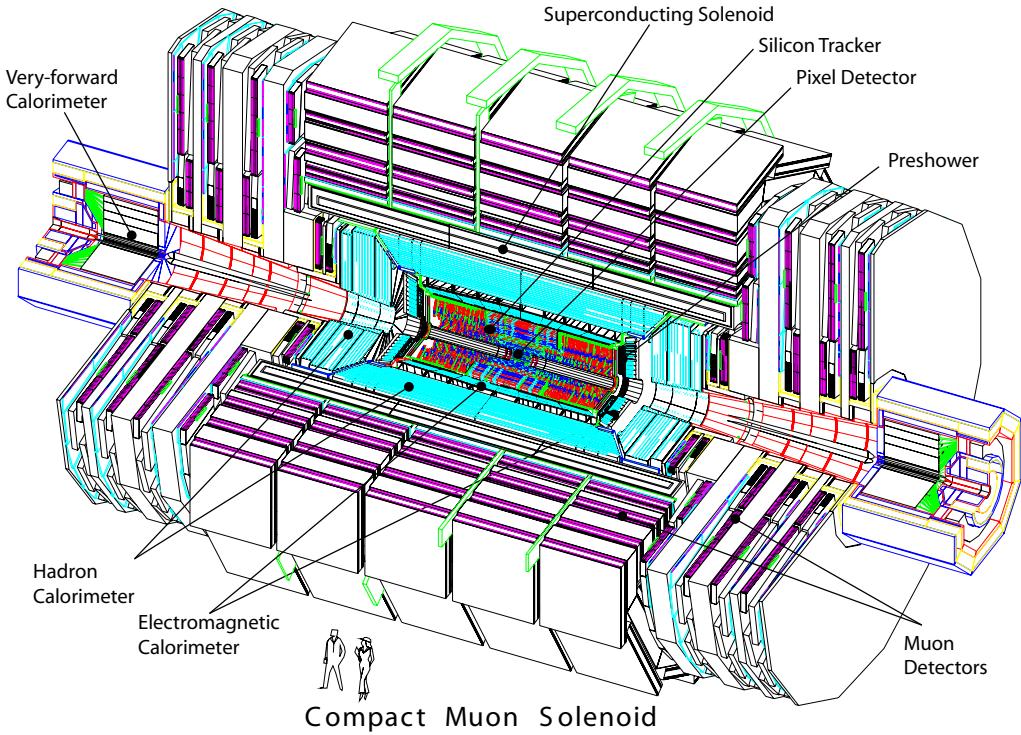


FIGURE 4.1: A drawing of the CMS detector. [33]

## 4.1 CMS coordinate system

CMS uses a right-handed coordinate system with the origin in the nominal interaction point.  $z$ -axis is pointing along the beam line.  $x$ -axis is pointing towards the center of the LHC ring while  $y$ -axis points upwards. Two angles are used to describe position inside the detector, azimuthal angle  $\phi$  and polar angle  $\theta$ .  $\phi$  angle lies in  $x - y$  plane with a range  $[-\pi, \pi]$  and is defined as  $\phi = \arctan(y/x)$ . The other angle  $\theta$  is usually not used in high-energy physics because differences in  $\theta$  are not Lorentz invariant. The variable that is Lorentz invariant is rapidity:

$$y = \frac{1}{2} \ln \left[ \frac{E + p_z}{E - p_z} \right] \quad (4.1)$$

In high energy experiments in the relativistic limit where  $E \gg m$ , a quantity called pseudorapidity is a good approximation of rapidity:

$$\eta = -\ln \left[ \tan \frac{\theta}{2} \right] \quad (4.2)$$

The Lorentz invariance of pseudorapidity means that a measurement of  $\Delta\eta$  between particles is not dependent on specifying a reference frame, such as the rest frame of a particle or the laboratory frame. The term "forward" direction refers to regions of the detector that are close to the beam axis, at high  $|\eta|$ . When the distinction between "forward" and "backward" is relevant, the former refers to the positive  $z$ -direction and the latter to the negative  $z$ -direction.

In proton-proton collisions, colliding objects are partons and gluons. Given the energy-momentum conservation and the fact that the proton momentum in the plane perpendicular to the beam axis is negligible, the momenta of the final state particles have to be balanced in the  $x - y$  plane. This is why transverse momentum is often used in various analyses and is computed as  $p_T = \sqrt{p_x^2 + p_y^2}$ .

## 4.2 Solenoid magnet

The CMS solenoid magnet has the length of 12.9 m, an inner diameter of 5.9 m and provides a magnetic field of 3.8 T. The solenoid is large enough to contain inner tracking system and calorimeters which reduces the material budget before the energy measurement in the calorimeters. The strong magnetic field increases the curvature of the trajectories of the highly energetic particles created in the collision thus improving the momentum resolution.

Solenoid magnet is built of superconducting materials with the operational temperature of 4.6 K. It is composed of four layers of superconducting material inserted in

aluminum. Muon detectors outside the solenoid operate in 2 T magnetic field enhanced by the 10 000 t iron yoke.

## 4.3 Inner tracker system

The role of inner tracking system is to provide a precise measurement of the trajectories of charged particles with  $p_T > 1$  GeV and the pseudorapidity  $|\eta| < 2.5$ . Additionally, it allows for precise secondary vertex positions reconstruction and impact parameter determination. The size of CMS inner tracker is 5.8 m in length with a diameter of 2.5 m. Large magnetic field of 3.8 T is provided by the surrounding solenoid and is homogeneous across the entire inner tracking system. With the design LHC luminosity, expected occupancy of inner tracking system amounts to more than 1000 particles from 20 primary interactions in each bunch crossing. This requires high granularity detectors with fast responses and low dead time. In addition the amount of material in the detector has to be kept at minimum and the radiation hardness must be taken into account which lead to the solution of building an all-silicon detector with high granularity. CMS inner tracking system has two separate parts, Pixel detector and Strip detector, both described below.

### 4.3.1 Pixel Detector

Pixel detector is the innermost part of the CMS, closest to the interaction point. The central part, called barrel pixel, consists of three layers located at radii of 4.4 cm, 7.3 cm and 11 cm. On each side of the barrel pixel, there are two discs at  $z = 34.5$  cm and 46.5 cm. The detector covers pseudorapidity range  $-2.5 < \eta < 2.5$  which is illustrated in figure 4.3. Its purpose is to provide precise three dimensional space points for charged particle tracking and vertex position determination.

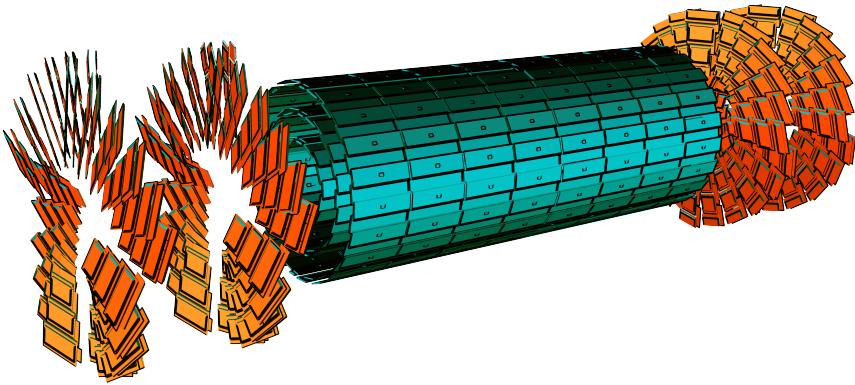
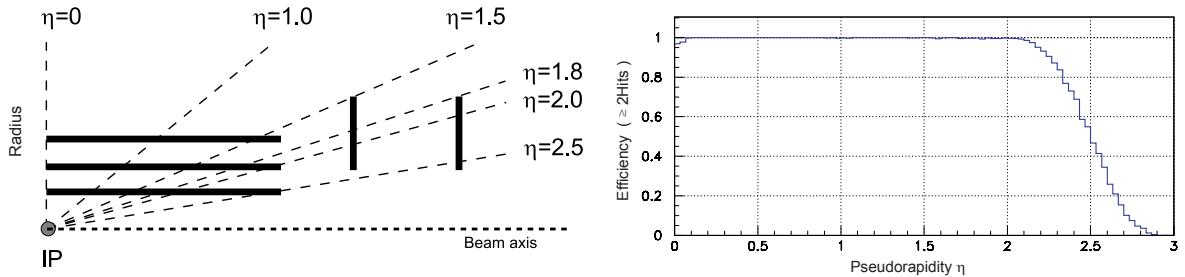


FIGURE 4.2: A drawing of the CMS pixel detector. [33]

FIGURE 4.3: A sketch of CMS Pixel Detector pseudorapidity range coverage(*left*) and hit efficiency as a function of pseudorapidity(*right*). [33]

Pixel detector is fully modular, consisting of rectangular modules in barrel part and quasi-triangular modules in the discs. Modules are arranged in a way to ensure measurements in at least three layers for each of the trajectories passing through the detector. Pixel size of  $100 \times 150 \mu\text{m}^2$  results in similar resolution in  $z$  and  $r - \phi$  directions. In the barrel part the resolution of  $15 - 20 \mu\text{m}$  is achieved due to charge sharing. Electrons inside the silicon shift under Lorentz force which is used in the reconstruction to determine the correct hit position. Detailed measurement of the Lorentz angle is described in the Appendix A. Pixel detector consists of 66 million pixels in total with 48 million being

in the barrel pixel and 18 million in the forward. The closeness to the interaction point implies high track occupancy and necessity for radiation resistant materials.

Readout of the pixel detector goes through read-out chips (ROC) to which each pixel is bump bonded and read out individually. There are around 16000 ROCs in the detector. Each ROC consists of  $52 \times 80$  pixels. Only pixels with signal above certain threshold are read out which can be tuned manually for each pixel. The average noise level in the detector is around 170 electrons at  $T = -10^\circ\text{C}$ . The information for each event is stored in a temporary buffer awaiting the signal from the Level-1 trigger in order to be read-out. Data is read out serially, with packets containing all hits corresponding to a single trigger. Each pixel hit uses six values, five to encode pixel position, and sixth value is the analog signal charge. ROC header is added at the beginning of each ROC sequence in order to make ROC hit-association possible. Signal is than digitized and sent to central data acquisition for further processing. Various other systems are installed in order to monitor and adjust the temperature, humidity, voltages, etc.

With the design LHC luminosity, there are more than 1000 particles hitting the detector in every bunch-crossing. Very small pixel size results in the occupancy for each pixel of the order  $10^{-4}$ . The Pixel detector has been operational for several years and shows very little drop in performance due to irradiation. The plan is to keep the present detector during the Run 2, until 2017, and than replace it with new, four-layer pixel detector which is currently being built.

### 4.3.2 Strip detector

Silicon pixel tracker is built in layers around Pixel detector where track particle flux is lower and lower granularity detector can be used instead. Detector is built of strips in which a passing charged particle induces current. Current is than transferred to silicon detectors connected to the wires. The barrel section of the strip detector consists of four layers in the inner part (TIB) and 6 layer in the outer part (TOB). In the forward regions

there are three tracker inner discs (TID) on each side of the barrel and 9 layers in the tracker endcap (TEC).

Some strips are built in double layers tilted against each other by an angle of 100 mrad to precisely measure the position of both  $r\phi$  and  $rz$  directions. The pitch size between strips varies from 80  $\mu m$  in the TIB to 184  $\mu m$  in TOB and TEC. With the increasing distance from the interaction point, both strip pitch and strip length increase and sensor thickness becomes larger which affects the resolution.

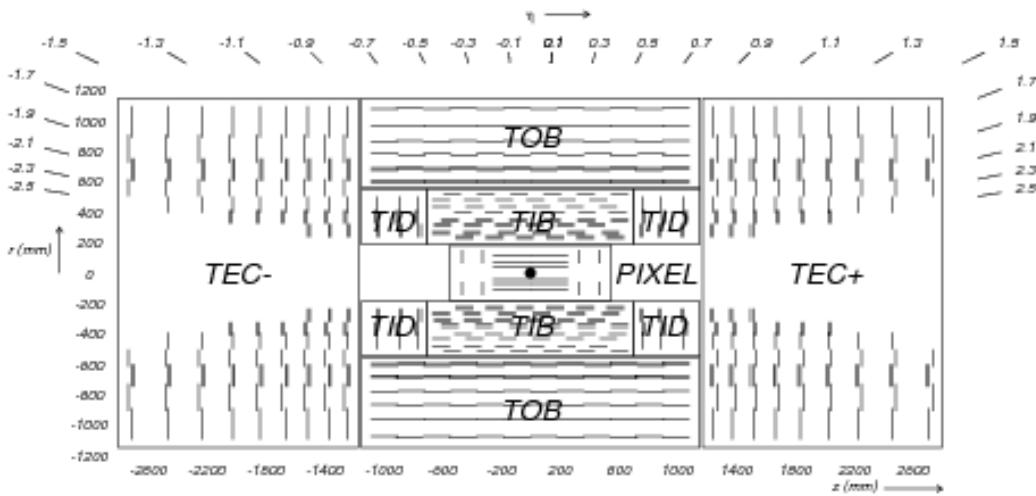


FIGURE 4.4: A drawing of the CMS strip detector. [33]

## 4.4 Electromagnetic calorimeter

The role of the electromagnetic calorimeter is precise measurement of electron and photon energies. It is built from lead tungstate ( $PbWO_4$ ), a material with very high density (8.28 g/cm<sup>3</sup>) and a small Moliere radius (0.89 cm) which is a scale of transverse dimension of the fully contained electromagnetic showers. The scintillation light emitted within a single bunch crossing of 25 ns is about 80% of the total light which means the response time of the detector is very small and is a large advantage of this material. The calorimeters is built of 61 200 crystals in the barrel region and 14 670 crystals in the endcaps. Each

crystal has a size of  $22 \times 22 \text{ mm}^2$  in the front,  $26 \times 26 \text{ mm}^2$  at the back side and length on 23 cm in the barrel region. In the endcaps, the size of the crystals varies from  $28.62 \times 28.62$  in the front to  $30 \times 30 \text{ mm}^2$  in the back with a length of 22 cm. The whole systems covers the range  $|\eta| < 3$ .

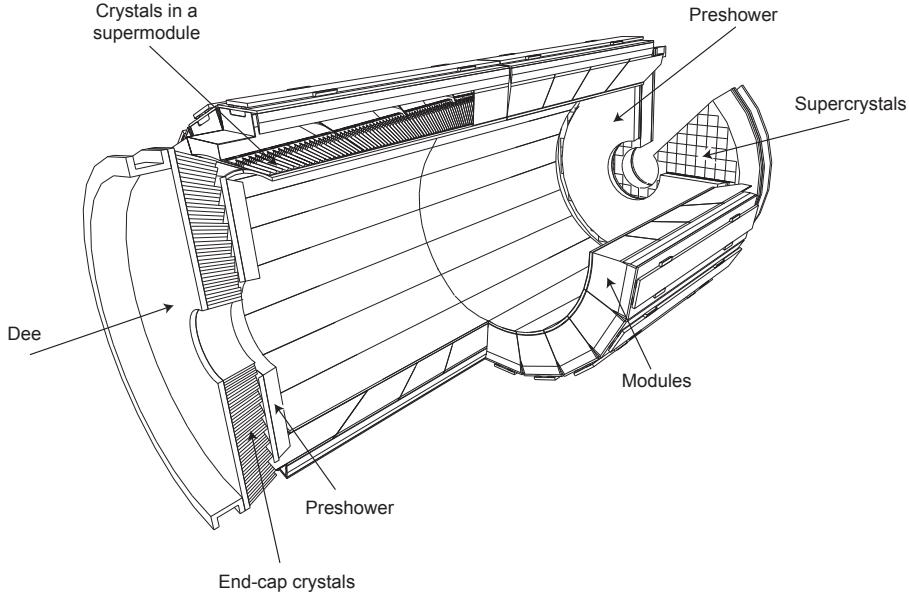


FIGURE 4.5: A drawing of the CMS electromagnetic calorimeter. [33]

Operation temperature of the detector is  $18^\circ\text{C}$  at which  $\sim 4.5$  photoelectrons are collected per MeV of deposited energy. The blue-green scintillation light is measured by the avalanche photodiodes in the barrel and vacuum phototriodes in the endcaps.

The ECAL energy resolution is affected by three uncorrelated sources and can be described by the relation 4.3. Parameters  $a$ ,  $b$  and  $c$  are determined from the test beam. The stochastic term  $a$  is very low for the lead tungstate crystals ( $a = 2.83 \pm 0.3\%$ ) which means that showers are mostly contained within the crystals. The noise term  $b$  is determined from the electronics and amounts to  $b = 124 \text{ MeV}$ . The last term  $c$  is the constant term which limits the ECAL accuracy at high energies.

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{b}{E}\right)^2 + c^2 \quad (4.3)$$

## 4.5 Hadronic calorimeter

Hadronic calorimeter is used to measure energies of hadron particles such as pions, kaons, protons, neutrons etc. Barrel and endcap hadronic calorimeters cover the pseudorapidity range to  $|\eta| = 3$ . Since the absorber in the transverse direction thickness is only 5.82 interaction lengths, additional layer was placed outside the solenoid. Hadronic calorimeter is a sampling calorimeter which consists of layers of brass and plastic scintillator layers. Showers are produced mostly in brass and are detected in the scintillator and reemitted in the narrow wavelength range in which photodetectors operate. In the endcap region, steel and quartz are used because of their higher radiation hardness. There is an additional part of the detector placed 11.2 meters from the interaction point on both sides called forward HCAL which extends the coverage to  $|\eta| = 5.2$ . Large HCAL coverage and good energy measurement are very important for jet reconstruction as well as for missing transverse energy determination.

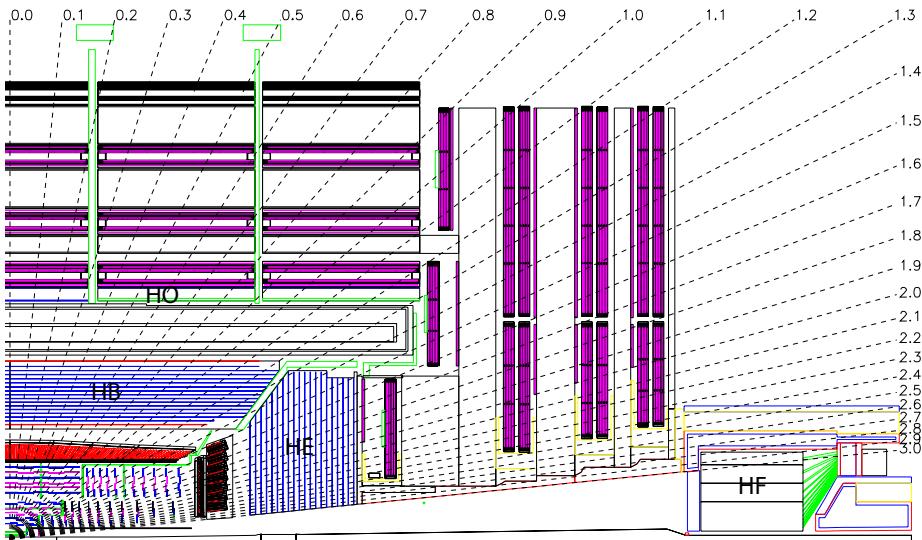


FIGURE 4.6: A drawing of the CMS Hadronic Calorimeter. [33]

## 4.6 Muon chambers

Muons are the only particles that can pass the calorimeters and the solenoid. Their charge and momentum is measured also in the outer part of the detector by the muon chambers. There are three different types of the gaseous detectors used in the muon system, Drift tubes (DT), Resistive Plate Chambers (RPC) and Cathode Strip Chambers (CSC). Drift tubes are used in the barrel region where muon rate is relatively low and covers pseudorapidity range of  $|\eta| < 1.2$ . The signal in Drift tubes is generated when a particle ionizes the gas inside the tube and the charge is collected by wires which are at high voltages. Cathode Strip Chambers are used in the endcap region where muon rate is much higher and magnetic field is not uniform. These are multi-wire proportional chambers with anodes that collect charge from the gas ionization. Resistive Plate Chambers are placed both in barrel and endcap region. These detectors are designed as two parallel plates which create a uniform electric field in the gas between them. The electrodes on the plates are highly resistive so when charged particle passes, it causes an electron avalanche which passes through the plates and is collected by the external metallic strips. Their time resolution is of the order  $\sim 1$  ns which makes RPCs a good choice for triggering although their spatial resolution is not so good.

Large magnetic field enables even for high  $p_T$  muons to be measured with a reasonable cell size in the muon chambers. The limiting factor for good resolution of low  $p_T$  muons is multiple scattering, and for high  $p_T$  muons the chamber resolution. The momentum resolution as a function of muon  $p_T$  is shown in Figure 4.8 for both muon chambers and inner tracking as well as the combined result.

## 4.7 Trigger

The design rate of the proton collisions at the LHC is 40 MHz, although during Run 1 data taking period, the rate was 20 MHz which corresponds to 50 ns bunch spacing. Data

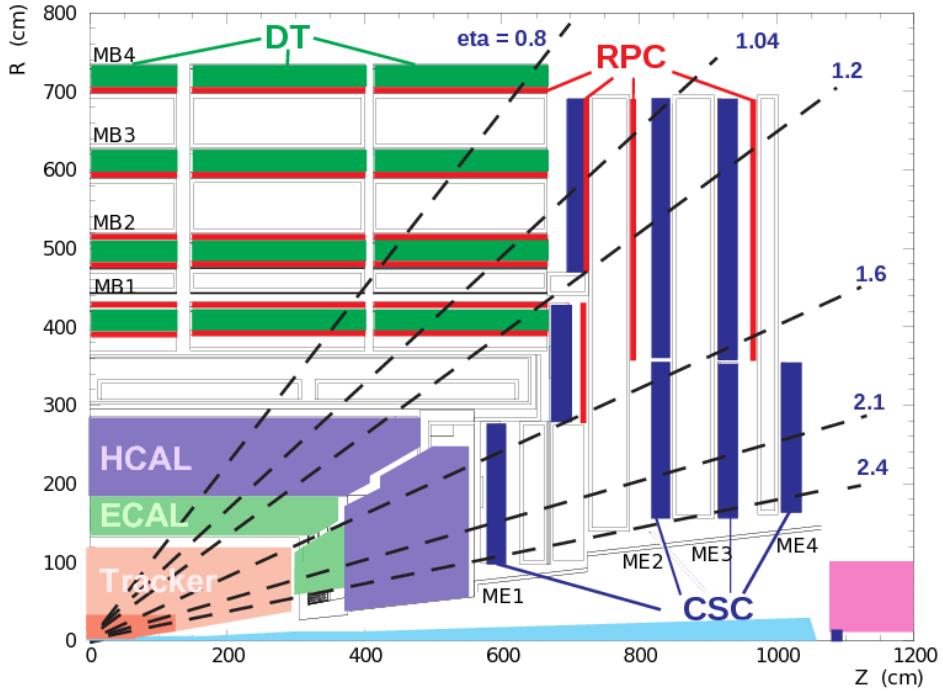


FIGURE 4.7: A drawing of the CMS muon chambers which consist of three different types of the detectors: Drift tubes, Cathode Strip Chambers and Resistive Plate chambers. [33]

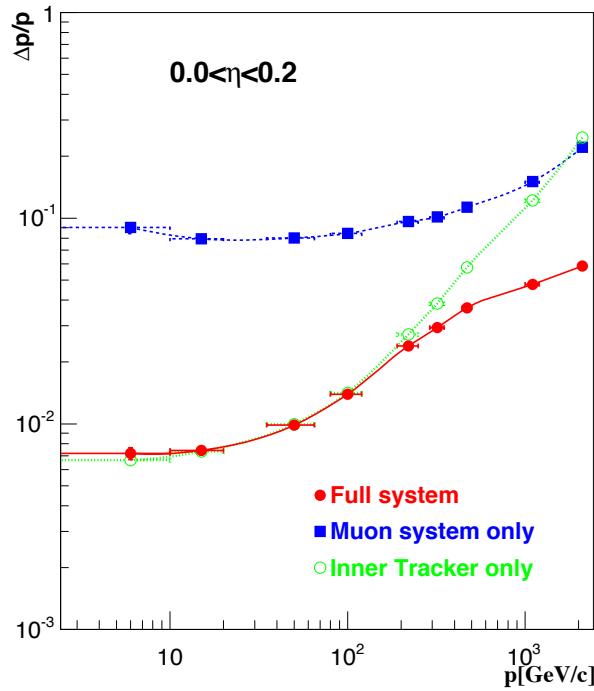


FIGURE 4.8: Muon resolution measurements for tracker, muon chambers and combined [33]

from each of the bunch crossings is called an event. Since there are huge amounts of data coming from the subdetectors, it is necessary to apply some conditions which can reduce the rate to about 100 events per second. This is done by two level triggering system, first one called Level 1(L1) trigger and second one called High Level Trigger(HLT). L1 trigger uses a special custom made electronics designed to reduce the output rate from 40 MHz to 100 kHz. Events which pass some loose criteria are than passed to the HLT. The L1 trigger uses the information from calorimeters and muon chambers to take the decision whether the event should be accepted or rejected usually searching for the presence of muons, jets above certain  $p_T$ , or looking at the total amount of  $E_T$  and  $E_T^{miss}$ . The time needed to send the signals to the electronics, run the L1 selection and send the information back to the subdetectors in  $3.2 \mu s$ . If L1 triggers accept the event, it is stored in the readout buffers where partial reconstruction takes place and the event is than processed by the HLT. This is a software farm which reduces the number of events to about 100 per second. The schematic of the trigger system is shown in figure 4.9.

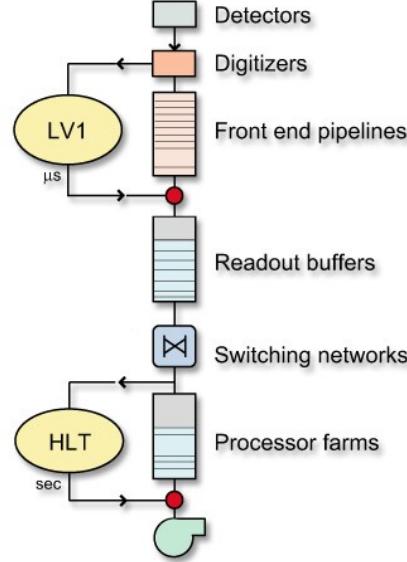


FIGURE 4.9: A schematic of CMS Trigger System. [33]

## 4.8 Luminosity measurement

Accurate and precise luminosity measurement is essential for physics analyses. CMS measurement is based either on the activity in the forward hadronic calorimeter (HF) or on the number of clusters in the pixel detector. The forward calorimeter is placed in the high pseudorapidity region, covering range  $3 < |\eta| < 5$ . The measurement relies on the fraction of nonempty calorimeter towers to estimate the number of interactions. The uncertainty to this measurement comes from the nonlinear response of the HF to luminosity and the occurrence of the afterglow, the effect when the energy deposits created in a given bunch crossing create a signal in subsequent bunch crossings.

Luminosity measurement with the pixel detector is described in detail in [39, 40]. This method relies on the effective cross section measurement for pixel clusters determined with Van der Meer scan. The value of this cross section is then used to determine integrated luminosity for each lumi-section (23.3 seconds). This approach is not suitable for online luminosity measurements as pixel bias voltage is not turned on until the stable beams are declared and the data is not available if the central data acquisition system is busy. On the other hand, in the offline analysis this is not a problem as the data recorded without the pixel detector is discarded anyway. The total luminosity delivered in 2012 was around  $24 \text{ fb}^{-1}$  while luminosity recorded by the CMS was around  $22 \text{ fb}^{-1}$ . Ideally these two numbers would be the same, but due to the downtime of data acquisition or some of the subsystems, some of the delivered luminosity is not recorded.



# Chapter 5

## Physics objects definitions

CMS detector is designed to efficiently identify and reconstruct interesting physics objects. The reconstruction procedure which takes as input the signals from all subdetectors and combines them to get physics objects is called *particle flow*. [41] This algorithm classifies all the objects into one of the following categories: charged hadrons, neutral hadrons, photons, electrons and muons. These are built from reconstructed tracks, energy deposits in the calorimeters and signals in the muon chambers creating a global event description. Additionally, a set of cuts is imposed on both input signals and reconstructed object in order to minimize the misidentification, e.g wrongly identifying electron as a jet.

Following sections show electron and muon reconstruction procedures and identification criteria. Additionally, jet reconstruction procedure is described together with necessary jet corrections and b-tagging algorithms. After the reconstruction of all other objects, missing transverse energy is computed.

### 5.1 Electrons

Electrons in CMS are detected as a tracks in tracking system and an energy deposit in the electromagnetic calorimeter. Two different algorithms are used for electron reconstruction,

*tracker driven* seeding which is more suitable for low  $p_T$  electrons and electrons inside jets and *ECAL driven* seeding optimized for high  $p_T$  isolated electrons. Both approaches take electromagnetic crystals with deposited energy and join them into *clusters*. Electron passing through the detector bends due to magnetic field and interacts with the detector material emitting *bremsstrahlung* photons. ECAL energy deposits from these photons are spread in  $\phi$  direction in very narrow  $\eta$  range and combined with the existing cluster forming a *supercluster*. Trajectories are reconstructed using modeling of electron energy loss in detector material and fitted with a Gaussian Sum Filter(GSF)[42].

Matching ECAL superclusters and reconstructed tracks is where the two approaches differ. Tracker driven seeding uses track from the tracking system and tries to match it with the supercluster in the ECAL while ECAL driven seeding starts from the superclusters. Each electron candidate has to pass various quality cuts in order to maximize the probability of identifying the electron coming from the hard interaction, and reject electrons from jets or conversion. These selection cuts can be divided into three categories: identification, isolation and conversion cuts. Details on electron reconstruction and performance can be found in [43].

## Electron identification

Electron identification procedure first focuses on good matching between reconstructed track and supercluster, by imposing cuts on angular distance  $\Delta\eta$  and  $\Delta\phi$  between the two. These variables are computed as absolute  $\eta$  and  $\phi$  distance between the supercluster and electron track extrapolated to the ECAL surface. Additionally, a cut is imposed on  $\sigma_{inj\eta}$  which describes a shower shape spread in  $\eta$  direction. This variable is particularly discriminating against clusters coming from electrons and energy deposits from photons and fakes. Shower shape is defined as:

$$\sigma_{inj\eta} = \sqrt{\frac{\sum_i^{5\times5} w_i (\eta_i - \eta_{seed})^2 \times \Delta\eta_{xtal}^2}{\sum_i^{5\times5}}} \quad (5.1)$$

where  $i$  runs over all crystals in  $5 \times 5$  block around supercluster seed,  $\eta_i - \eta_{seed}$  is the distance in number of crystals in  $\eta$  direction between  $i$ -th crystal in supercluster and seed crystal and  $\Delta\eta_{xtal}^2$  is the average width of a single crystal. Each crystal is given a weight defined as  $w_i = \max(0, 4.7 + \ln(E_i/E_{5\times 5}))$ , where  $E_i$  is a single crystal energy, and  $E_{5\times 5}$  is the sum of energy deposits inside a  $5\times 5$  crystal block. Additional cut on the ratio between the energy deposits in the hadronic and electromagnetic calorimeter for electrons is used to discard the events with significant hadron activity.

Electrons coming from photon conversions are rejected by requiring a hit in every layer of the inner tracking system. Additionally, for each electron track a fit is performed trying to combine it with another electron track under the hypothesis that both electrons originate from converted photon. Electron is selected only if this probability is sufficiently small. Electron compatibility with the primary vertex is estimated by looking at the impact parameters in both  $xy$  and  $z$  planes. Due to the gap in the electromagnetic calorimeter in  $1.4442 < |\eta| < 1.566$ , all electrons which have a supercluster position reconstructed in this range are rejected. A full list of identification criteria is summarized in Table 5.1.

TABLE 5.1: A summary of electron identification criteria.

Variable	Barrel	Endcap
$\Delta\eta <$	0.004	0.005
$\Delta\phi <$	0.03	0.02
$\sigma_{inj} <$	0.01	0.03
$H/E <$	0.12	0.10
$d_{xy} <$	0.02 cm	0.02 cm
$d_z <$	0.1 cm	0.1 cm
$(1/E - 1/p) <$	0.05	0.05
Missing hits	0	1
Vertex Fit Probability	$10^{-6}$	$10^{-6}$

## 5.2 Muons

Muons in CMS are reconstructed by combining a reconstructed track inside the tracker (*tracker track*) and track in muon chambers (*standalone muon track*). Individual track segments in the muon chambers are fitted using Kalman fitter technique [44] in order to obtain a standalone muon track. Similar as with electrons, two approaches are used for combining track from tracker and standalone track.

*Global muon reconstruction* approach uses a standalone muon track in the muon chambers and tries to find a matching tracker track by combining parameters of two tracks by projecting it to the common surface. This *outside-in* approach uses Kalman fitter technique to combine these two objects in an object called *global muon*. Muon momentum is then determined from this global muon track using all available systems which shows improved precision in comparison to other approaches.

The second approach for muon reconstruction is *tracker muon reconstruction* which starts from tracks inside the tracker with  $p_T > 0.5$  GeV/c and total momentum  $p > 2.5$  GeV/c as potential muon candidates. Extrapolation is then performed to the muon chambers taking into account the magnetic field, Coulomb scattering in the material and other energy losses. *Tracker moun* is found if at least one muon segment matches the extrapolated track. The efficiency of the *Tracker muon* reconstruction is higher for low energy muons than the efficiency for the global muons, because only a single muon segment in the muon chambers is required. For high energy muons where more there are more segments inside muon chambers, *global muon* algorithm is designed to have high efficiency. Detailed muon reconstruction procedure is shown in [45].

### Muon identification

In this analysis *particle flow* muon identification selection is applied to the *global muons*. Selection is applied in order to minimize misidentification of charged hadrons as muons, maximize the efficiency of muon identification inside jets and ensure good momentum

measurement. Muons used in the analysis have  $|\eta| < 2.1$  and transverse momentum  $p_T > 30$  GeV with more than 5 hits in the inner tracker system and at least one hit in pixel detector. At least one good muon chamber hit in the *global muon* track fit is required to have  $\chi^2/ndof < 10$ , at least two segments in two different muon stations should be matched to a track in order to suppress muons from in-flight decays. Cosmic muons are rejected by applying cuts on the impact parameter with respect to the primary vertex of  $|d_{xy}| < 0.2$  cm and  $|d_z| < 0.5$  cm. Muon identification criteria are summarized in Table 5.2.

TABLE 5.2: A summary of muon identification criteria.

Variable	Requirement
number of pixel hits >	0
number of inner tracker hits >	5
$\chi^2/ndof <$	10
number of muon hits >	0
chambers with matched segments >	1
$d_{xy} <$	0.2 cm
$d_z <$	0.5 cm

### 5.3 Lepton isolation

Leptons from W decays are in general expected to be well isolated. The degree of isolation is calculated using *particle flow* approach by summing the transverse momenta contributions of particles around the lepton inside a specific cone. All charged particles are considered as well as photons and neutral hadrons with  $p_T > 0.5$  GeV. The cone used for determination of energy deposits is defined as  $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$  around the lepton axis and isolation measure is defined as:

$$I_{PF}^{rel} = \frac{\sum p_T^{charged} + \max(0, \sum E_T^\gamma + \sum E_T^{neutral} - 0.5 \sum E_T^{PU})}{p_T^l} \quad (5.2)$$

where  $p_T^{charged}$  is the sum of the momenta of charged hadrons and  $E_T^\gamma$  and  $E_T^{neutral}$  are the sums of photon and neutral hadron momenta.  $E_T^{PU}$  is the sum of the pile-up transverse

energies from neutral particles and is calculated as a sum of track transverse momenta not coming from the primary vertex inside the isolation cone. This is divided by the factor of 0.5 which corresponds approximately to the ratio of neutral to charged hadron production in the hadronization process of pile-up interactions. Selected muons are required to pass isolation cut  $I_{rel}^{PF} < 0.12$ . On the other hand, if there is a requirement that there be no leptons in the event, the isolation cut is  $I_{rel}^{rel} < 0.2$ . Electron isolation is computed in the same way with cut of  $I_{rel}^{PF} < 0.1$  for selected electrons and  $I_{rel}^{PF} < 0.15$  for vetoed additional electrons.

## 5.4 Jets

In high energy physics, jet is a collimated group of hadrons which emerges as a result of quark or gluon fragmentation and hadronization process. Hadrons reconstructed in a particle detector need to be combined in order to form a jet and give information about the initial parton. A set of rules has to be created which define how to group particles and how to assign momentum to the jet. Usually this is done by summing the four-momentum of each particle in a jet.

### 5.4.1 Jet algorithms

Jet algorithms take into account the distance between particles and define rules to determine which particle belongs to what jet. Same jet algorithms should be applicable to both, experimental data and theoretical calculation. Other important properties of jet algorithms are *infrared safety* and *collinear safety* which means if an event is modified by addition of soft emission or collinear splitting, the final number of hard jets will remain unchanged. These two properties together are called *IRC safety*. IRC unsafe jet algorithms may break the cancellation of divergences by yielding one set of jets for tree-level splitting while loop diagrams lead to another, as shown in figure 5.1, giving infinite cross-sections in

the final calculations. Jet definitions, jet relation to partons and an overview of different jet algorithms are summarized in [46].

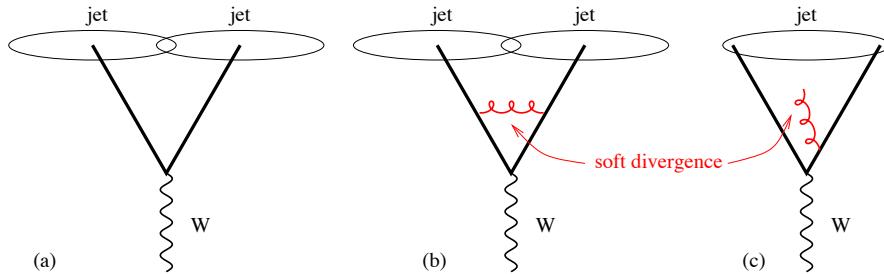


FIGURE 5.1: Configuration showing IC unsafety with W boson and two partons. Adding a soft gluon causes two jets to be reconstructed as one. [46]

There are two types of jet algorithms which are most commonly used: *cone algorithms* and *sequential recombination algorithms*. In the case of *cone algorithms*, a jet is defined as a set of particles inside a stable cone around their center of mass. Most popular cone algorithm is *iterative cones* (IC) where a seed particle is chosen and momenta of all particles around that initial particle inside a cone of radius  $R$  are summed. After adding each new particle to the sum, the direction of the new sum is taken as a seed direction, and the procedure repeats until the direction of the resulting cone is stable. Particles inside the cone are then removed from the list of available particles and the procedure repeats. This approach is not IRC safe given that nearly collinear splitting of the hardest particle in the event can be reconstructed as two jets. In that case, a less energetic, particle, pointing in another direction, can become the hardest particle in the event, yielding different set of jets. Cone algorithms can be IRC safe using a *seedless cone* (SC) algorithm where all stable cone solutions are identified at once. However this approach is very time consuming even for small number of particles and thus very impractical to use. In the *sequential recombination algorithms* at hadron colliders two longitudinally invariant distances are introduced:  $d_{ij}$  which is the distance between each pair of particles and  $d_{iB}$  which is the particle-beam distance. These distances are defined as:

$$d_{ij} = \min(k_{T,i}^{2p}, k_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \quad (5.3)$$

$$d_{iB} = k_{T,i}^{2p} \quad (5.4)$$

where  $\Delta R_{ij}$  denotes the distance in the  $\eta - \phi$  plane and is computed as

$$\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2.$$

$k_T$  is transverse momentum of the particle,  $R$  is an angular cut-off similar to the one in *cone algorithms*, and  $p$  defines which particles are clustered first and is described below. Both  $R$  and  $p$  are free parameters of the algorithm. The algorithm is applied using the following approach: first two distances  $d_{ij}$  and  $d_{iB}$  are computed and minimal values are found. If  $d_{ij}$  is smaller, that two particles are combined, treated as a new particle and the distance with next particle in the list is computed. In case of  $d_{iB}$  being smaller,  $i$  is declared to be the final jet and removed from the list of particles. The procedure continues until there are no more particles in the list.

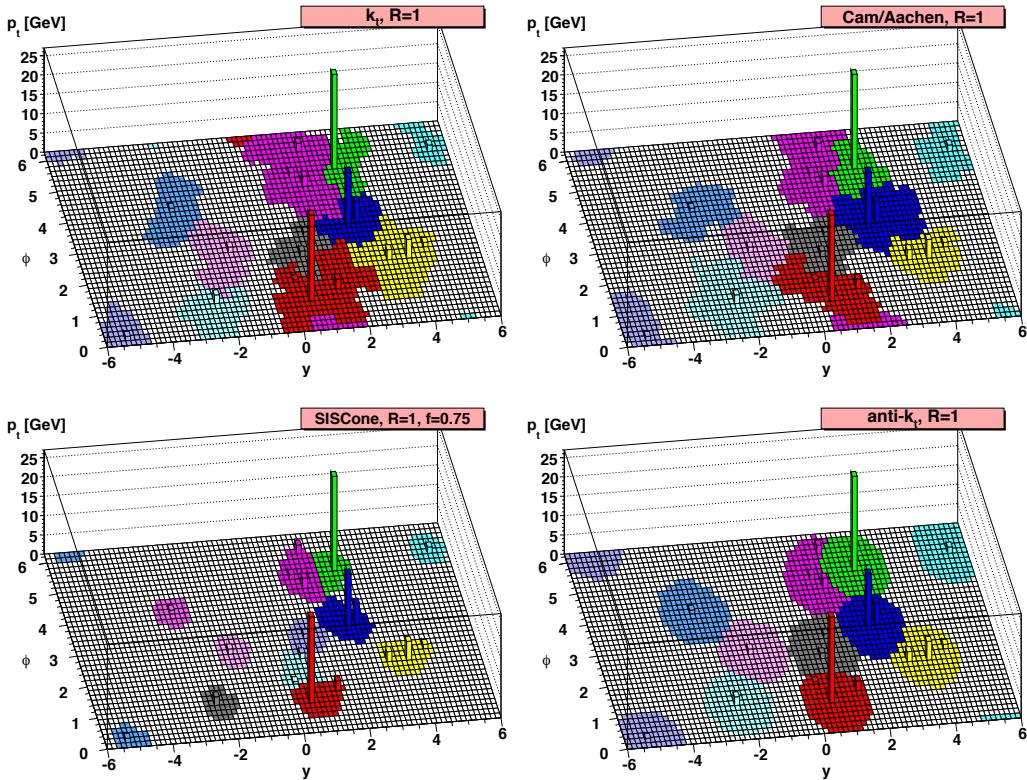


FIGURE 5.2: Clustering same set of reconstructed particles into jets using different jet algorithms. [46]

Parameter  $p$  defines which particles are clustered first thus defining the type of algorithm. The  $k_T$  algorithm uses  $p = 1$ , clustering soft particles first. This results in irregularly shaped jets, as shown in figure 5.2, which are sensitive to radiation in the event and difficult to calibrate. The *Cambridge-Aachen* algorithm (CA) uses  $p = 0$  thus relying only on angular distribution of the input particles. This approach is particularly useful for jet substructure analysis and is less sensitive to radiation. The algorithm used in this analysis is *anti*- $k_T$  algorithm where  $p = -1$  clusterizing the hardest particles first [47]. *Anti*- $k_T$  is an IRC safe algorithm with jets that are circular in shape because they are not affected by the softer components of the jet.

### 5.4.2 Jet corrections

Measured jet energy at detector level in general doesn't correspond to the energy of the originating particle. Jet calibration procedure is introduced to compensate for the nonlinear response of the calorimeters. This is done using a factorized approach where corrections on each level of correction are determined separately as described in [48]. Final corrected jet momentum is obtained from measured  $p^{raw}$  according to:

$$p^{corr} = C_{offset}(p_T^{raw}, \eta) \times C_{rel}(\eta) \times C_{abs}(p'_T) \times C_{res}(p''_T, \eta) \times p^{raw} \quad (5.5)$$

where offset correction  $C_{offset}$  and calibration factors  $C_{rel}$  and  $C_{abs}$  are applied to both data and simulation,  $C_{res}$  is applied only to data. Corrections are applied sequentially, in a fixed order such that  $p'_T = C_{offset} \times C_{rel}(\eta) \times p^{raw}$ , and  $p''_T = C_{rel}(\eta) \times C_{abs}(p'_T)$ . Correction factors used in this analysis can be found in [49]. Each level of corrections is summarized below:

- Offset correction  $C_{offset}$  compensates for energy contributions arising from pile-up events or instrument noise. The offset is determined in dependence of pseudorapidity, jet area  $p_T density$  which is described in [50].

- Relative correction  $C_{rel}$  is aimed to flattening the jet scale in pseudorapidity. The correction is determined from simulation, adjusting the jet scale in all  $\eta$  regions to one of the jets in  $|\eta| < 1.3$  without changing the absolute scale.
- Absolute correction  $C_{abs}$  flattens the jet scale in  $p_T$ . This correction is also determined from QCD multijet events as inverse of average response at fixed  $p_T^{gen}$ .
- Residual correction  $C_{res}$  is applied only to data in order to account for possible residual differences in data and simulation agreement after applying absolute and relative corrections. These corrections are derived using events with momentum balance in transverse plane, like dijet events or  $Z/\gamma + \text{jet}$  events.

The total jet correction for a fixed jet  $p_T$  as a function of pseudorapidity is shown in figure 5.3.

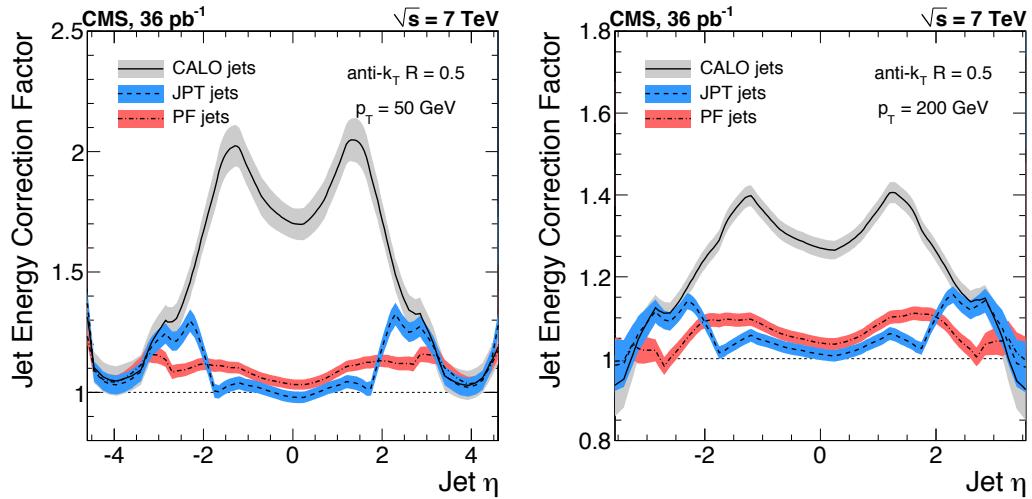


FIGURE 5.3: Total jet energy correction as a function of pseudorapidity of two different jet  $p_T$  values. Corrections are shown for all three types of jets, calo, JPT and PF jets. Bands indicate corresponding uncertainty.

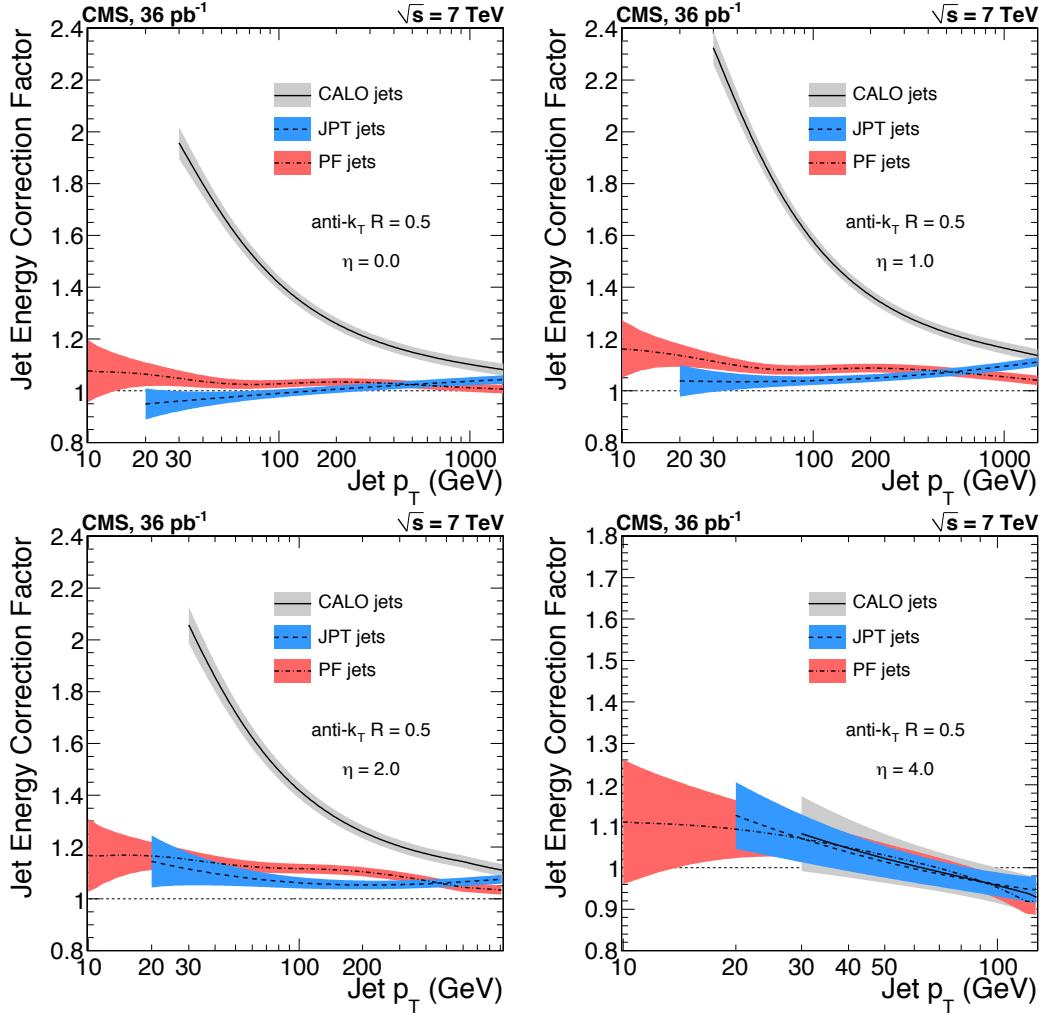


FIGURE 5.4: Total jet energy correction as a function of transverse momentum for four different  $\eta$  values. Corrections are shown for all three types of jets, calo, JPT and PF jets. Bands indicate corresponding uncertainty.

### 5.4.3 Jet identification

This analysis uses  $anti - k_T$  algorithm with cone size  $R = 0.5$ . Jet algorithm implementation is done in the *Fast-jet* package [51]. Depending on which signals is the algorithm applied to, there are different kinds of jets: calo jets (using calorimeter deposits), jet-plus-track jets (calorimeter deposits complemented with tracker information) and most widely used *particle flow* jets (PF). These jets are clustered from PF particles identified with PF algorithm, thus using the information not only from HCAL, but also from tracking

system and ECAL which show much better resolution. Only neutral fraction of the jet is measured only with HCAL which makes about 15% of the total jet composition. PF jets show excellent performance and are the default jets for most CMS analysis. Pile-up information is also taken into account by removing charged hadrons originating from pile-up vertices from the list of particles available for the jet clusterization. This procedure is called *charged hadron subtraction*. Some additional cuts to the jet composition are applied in order to endure good jet identification.

TABLE 5.3: A summary of jet identification criteria.

Variable	Requirement
Neutral hadron fraction	< 0.99
Neutral EM fraction	< 0.99
Number of Constituents	> 1
<hr/>	
Charged hadron fraction	> 0
Charged multiplicity	> 0
Charged EM fraction	< 0.99

#### 5.4.4 Jets from b quarks

Unique properties of the bottom quark can be used to identify hadronic jets originating from b quarks which are usually referred to as b-jets. Long lifetime of B hadrons is a consequence of weak force decay which results in their displacement by few micrometers at the LHC energies. B hadron decays show a large number of tracks with hard  $p_T$  spectrum and soft leptons emerging from semi-leptonic decays. The process of b-jet identification is called *b – tagging*. This process takes one or more variables and produces a single discriminant value for each jet. This value shows how much the observed jet looks like a b-jet. There are several *b-tagging* algorithms in use at CMS which are described in detail in [52] and the following were used in 2012 data analysis:

- *Track counting*(TC) - The discriminant value is the impact parameter significance which is calculated as impact parameter value divided by the respective impact

parameter uncertainty. Impact parameter values are sorted in the decreasing order. Depending on whether second or third value is chosen, the algorithm is denoted as high efficiency or high purity.

- *Jet Probability*(JP) - This algorithm combines information from several tracks inside a jet by computing a likelihood that all tracks originated from the primary vertex.
- *Combined secondary vertex*(CSV) - This is the most efficient *b-tagging* algorithm currently used at CMS. Both secondary vertex and track related information are combined to build a CSV discriminant value. It shows high efficiency even when no good secondary vertex can be reconstructed. Some of the variables used in CSV algorithm are flight distance, vertex mass, impact parameter significance, track multiplicity at the vertex and track multiplicity in a jet. The distribution of CSV discriminant value is shown in figure 5.5.

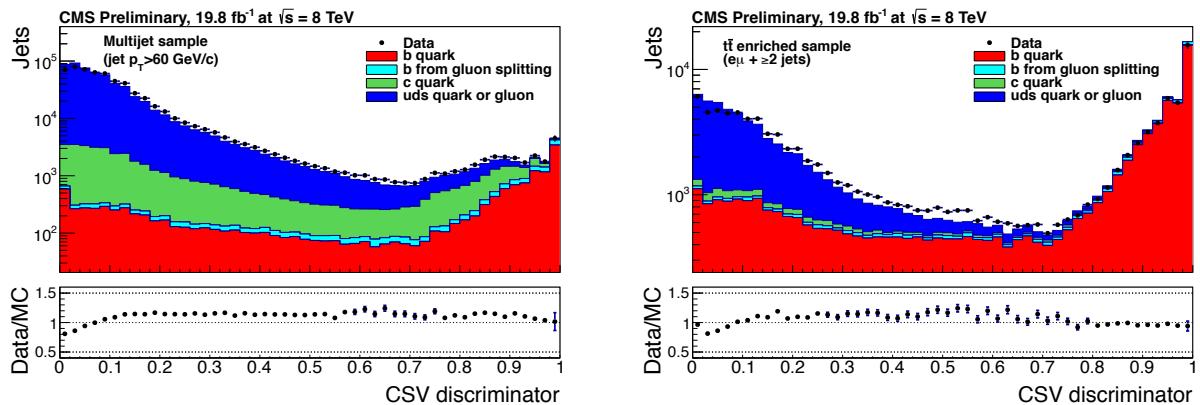


FIGURE 5.5: Combined secondary vertex discriminant value for multijet QCD sample (left) and  $t\bar{t}$  enriched sample(right)[53]

For each non-*b*-jet there is a chance that it would be identified as *b*-jet. Based on this misidentification rate, three operating points have been defined for the discriminant value: loose, medium and tight. For an average jet of 80 GeV, these values correspond to misidentification rates of 10%, 1% and 0.1% respectively. Misidentification probabilities as a function of *b*-jet efficiency for several algorithms are shown in figure 5.6.

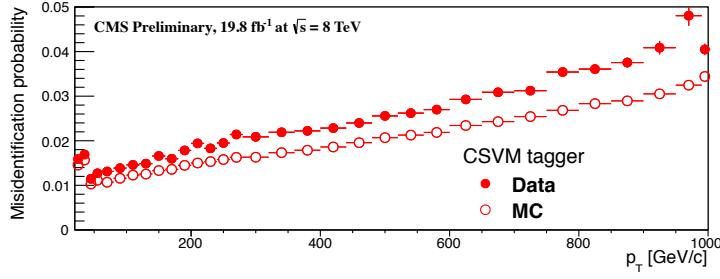


FIGURE 5.6: Combined secondary vertex misidentification probability for data and MC for medium working point.[53]

## 5.5 Missing transverse energy

Missing transverse momentum is the imbalance in the vectorial sum of transverse momenta of all measured particles. Momentum conservation delegates that the imbalance arises from weakly interacting neutral particles such as neutrinos. Missing transverse energy is the magnitude of the missing transverse momentum and is calculated as:

$$E_T^{miss} = \left| - \sum_i \vec{p}_i \right| \quad (5.6)$$

where  $i$  goes over all visible particles. Measurement of the missing transverse energy relies on the good measurement of all other particles in the event and as such is very sensitive to detector inefficiencies, particle missmeasurements, limited acceptance of the detector, cosmic-ray particles all of which can cause artificial missing energy. There are several approaches to determine  $E_T^{miss}$ . In this analysis, particle flow technique is used which tries to identify each particle in the event by combining the information from all subdetectors and gives the best missing energy resolution.[54, 55] Several corrections are applied to the  $E_T^{miss}$  which correct for the possible bias in the missing energy measurement:

- Type-I correction: propagates jet energy corrections described in Section 5.4.2 to missing energy. This correction replaces the missing energy calculated by summing transverse momenta of particles in a jet by transverse momentum of a jet to which JEC were applied.

- xy-shift correction: aimed at correcting the observed missing energy  $\phi$  modulation.

True missing energy distribution is expected not to depend not to depend of  $\phi$  because of the rotational symmetry of collisions around the beam axis. The possible cause for such modulation include unisotropic detector response, detector misalignment, the displacement of the beam spot. The amplitude of the modulation is observed to increase with the number of pile-up interactions so this correction can be seen as mitigation for the pile-up effects.



# Chapter 6

## Event selection and background estimation

This chapter describes the event selection criteria for identification of the  $pp \rightarrow W + bb$  process necessary for the cross section measurement. The selection is focused on leptonic decay modes of W boson, identified by the presence of isolated muon or electron and missing energy, and two b-tagged jets. Reconstruction and identification of these object has been described in the previous chapters. All major background contributions will be determined using simulation and subtracted in order to get the final yields used for cross section measurement. This chapter is organized as follows. Samples used in the analysis together with a short description on simulation procedures is described in Section 6.1. Signal selection criteria is listed in Section 6.2 and background sources are identified in Section 6.3. In order to obtain good agreement between data and simulation, certain corrections need to be applied to take into account trigger and reconstruction efficiencies, pileup reweighing and b-tagging inefficiencies. Applied corrections are listed in Section 6.4.

## 6.1 Data and Monte Carlo samples

Data samples used in this analysis are collected with CMS experiment in proton proton collisions during 2012. After performing necessary data-quality checks,  $19.8\text{fb}^{-1}$  of data was marked as good quality for physics analysis. Selected events are required to pass one of the following triggers:

- Isolated muon with  $p_T > 24 \text{ GeV}$  (HLT\_IsoMu24\_eta2p1)
- Electron with  $p_T > 27 \text{ GeV}$  which passes some additional identification criteria as described in 5.1.(HLT\_Ele27\_WP80)

Simulated samples for signal and background processes were obtained using Monte Carlo simulation, as a part of the official 2012 CMS production campaign. Several event generators were used to produce samples used in the analysis:

- **Pythia** [56, 57] is a multi-purpose generator which can also simulate parton shower. Pythia is only able to calculate tree-level  $1 \rightarrow 2$  and  $2 \rightarrow 2$  processes while higher orders are approximated with parton shower algorithm. Parton showering in all samples uses so called Z2 tune for modeling the underlying event [58, 59]. Diboson samples were generated with Pythia event generator while W+jets, Z+jets and TT samples were produced with Madgraph and showered with Pythia.
- **Madgraph** [60] calculates matrix elements on tree level to arbitrary order (up to 10). Radiation of hard gluons in initial and final state radiation is taken into account on the matrix element calculation level. A minimum  $p_T$  threshold has to be defined in order to avoid soft gluon emissions which causes total cross section to be strongly scale dependent. Thus, cross section is normalized to predictions from other software, such as MCFM[61] for Standard model processes.
- **aMC@NLO** [62] unifies the tree-level and next-to-leading order automated computation tools within the MadGraph family.

- **Powheg** [63] is a package optimized for heavy quark production in hadronic collisions. The hard process is calculated at the NLO order, but for fragmentation and hadronization other software is used (e.g. Pythia). Single top events were produced using this generator and showered with Pythia.
- **Tauola** [64] is a package for simulation of  $\tau$  decays.

Detector response is simulated using GEANT4 simulation package [65]. Data was analyzed using ROOT framework [66]. List of the used samples together with the corresponding luminosities is listed in the Table 6.1.

TABLE 6.1: Samples, generators and cross sections used for normalizations for signal and background simulation considered in this analysis. All samples are calculated to the NLO order except the W+jets which is NLO and  $t\bar{t}$  which is normalized to the latest combined cross section measurement of ATLAS and CMS collaborations [67].

Sample	Generator	$\sigma(pb)(NLO)$
W( $\rightarrow l\nu$ )+jets	Madgraph + Pythia	37509 (NNLO)
W + 1 jet	Madgraph + Pythia	—
W + 2 jets	Madgraph + Pythia	—
W + 3 jets	Madgraph + Pythia	—
W + 4 jets	Madgraph + Pythia	—
W + bb	Madgraph + Pythia	377.4
Z + jets	Madgraph + Pythia	3531.9
$t(\bar{t})$ semileptonic	Madgraph + Pythia	107.7
$t\bar{t}$ full leptonic	Madgraph + Pythia	25.8
single $t$ - t-channel	Powheg + Pythia	56.4
single $t$ - s-channel	Powheg + Pythia	3.97
single $t$ - tW-channel	Powheg + Pythia	11.1
single $\bar{t}$ - t-channel	Powheg + Pythia	30.7
single $\bar{t}$ - s-channel	Powheg + Pythia	1.76
single $\bar{t}$ - tW-channel	Powheg + Pythia	11.1
WZ	Pythia	33.6
WW	Pythia	56.0

Signal events are simulated in exclusive W+1,2,3,4 jets samples using Madgraph event generator and showered with Pythia. These samples are generated using the five-flavour scheme with massless b-quark in the initial state. Samples are then divided into three subsamples labeled as W+b(b), W+c(c) and W+light. W+b(b) subsample is selected if

there is a generated B hadron starting with pdgId= $\pm 5$  in one of the jets. W+c(c) samples is selected if c quark is generated in one of the jets (pdgId=  $\pm 4$ ). Events are classified sequentially with W+b(b) events selected first. All other events are labeled as W+light jets (W+udcsg). Additionally, measurement has been preformed using the shape from four-flavour sample listed in the Table 6.1 normalized to five-flavour cross section which has much larger number of generated signal events.

## 6.2 Event selection

Signal events are characterized by presence of a W boson and two jets which have been tagged as coming from b quarks. Candidates for a W boson are identified as isolated muons or electron and significant missing energy. Jets are identified as particle flow objects clustered with anti- $k_T$  algorithm with a cone size of 0.5. Combined secondary vertex (CSV) algorithm is then used to identify jets arising from fragmentation and hadronization of b-quarks which is explained in 5.4.4. Signal events are selected requiring the following:

- One muon or electron with  $p_T > 30$  GeV, within  $|\eta| < 2.1$ . which passes the trigger requirement and tight ID criteria described in 5.2, and has  $I_{rel}^{PF} < 0.12(0.10)$  in case of muons(electrons).
- Exactly two jets with jet  $p_T > 25$  GeV, within  $|\eta| < 2.4$  passing loose ID criteria from 5.4.3 with distance between lepton and jet  $\Delta R > 0.5$ . Events with more such jets are rejected.
- Events containing jets with  $p_T > 25$  in high pseudorapidity range  $2.4 < |\eta| < 5$  are rejected.
- Events containing additional lepton with  $p_T > 10$  GeV, within  $|\eta| < 2.1$ , tight ID and isolation are rejected.
- Both selected jets are required to pass tight CSV discriminator cut of 0.898.

After applying the described criteria, in the final distribution only around 20% of the selected events come from the Wbb events. Therefore it is essential to understand the contributions from all major backgrounds which is described in the next section. Several detector level distributions are shown in figure 6.1 for the muon channel and in figure 6.2 for the electron channel.

## 6.3 Background estimation

After applying all selection cuts described in the previous section, major backgrounds that remain are top quark, Z+jets, W+jets, diboson and QCD background. Each of the background contributions is described in detail below.

### 6.3.1 Top quark background

Production of  $t\bar{t}$  pairs and single top represent a challenging background at the LHC because of their relatively large production cross sections.  $t\bar{t}$  events are largely reduced by requiring additional jet veto. Single top events are more difficult to reject relative to signal using just topological cuts, but production cross-section is smaller resulting in a smaller contribution in the final distributions. Nevertheless, the contribution from the top quark backgrounds in the final distribution is around 60%. With very large contribution of the  $t\bar{t}$  background, a test of the normalization was performed. A separate control region was created which was defined to be same as signal region, but requiring additional jet activity. This results in a  $t\bar{t}$  enriched sample where it is visible that the shape of the distribution is well described in the simulation, but the difference in the overall normalization after applying all scale factors is of the order of 12%. The disagreement is taken into account while performing the global fit for number of signal events extraction which will be described further in the next section.

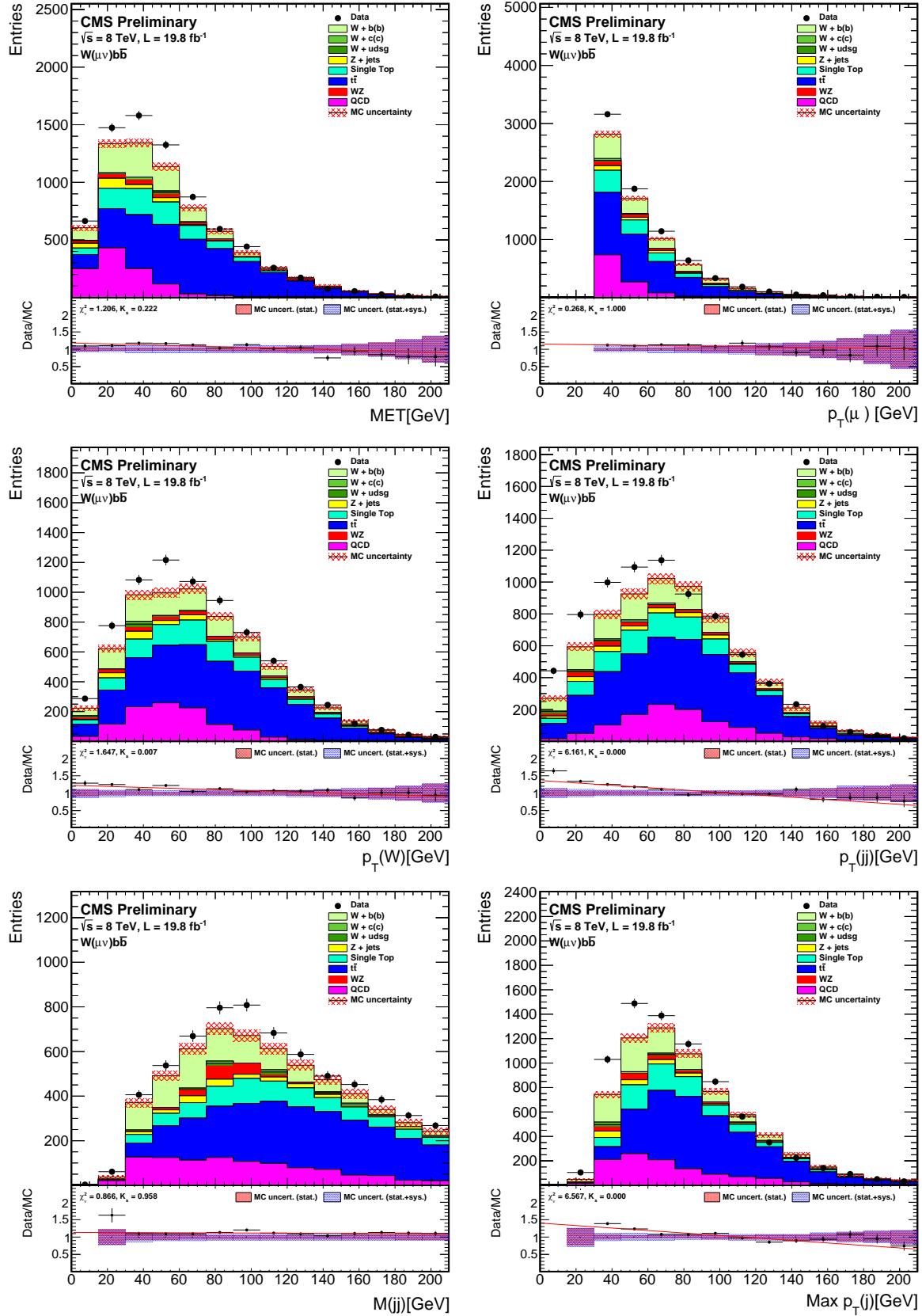


FIGURE 6.1: Signal region detector level distributions. Good shape agreement between data and simulation is observed, however simulation normalization is smaller than expected from data.

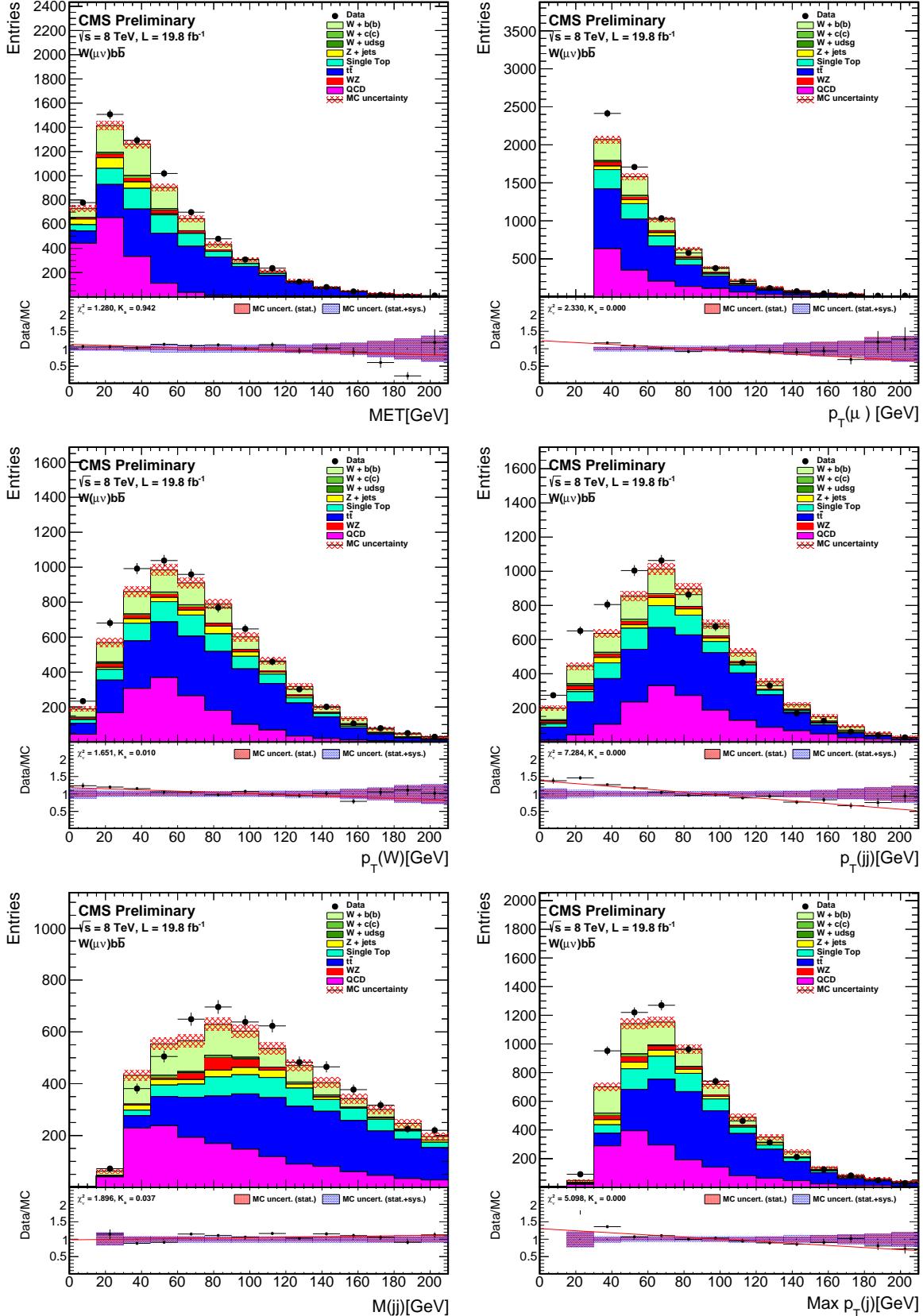


FIGURE 6.2: Signal region detector level distributions. Good shape agreement between data and simulation is observed, however simulation normalization is smaller than expected from data.

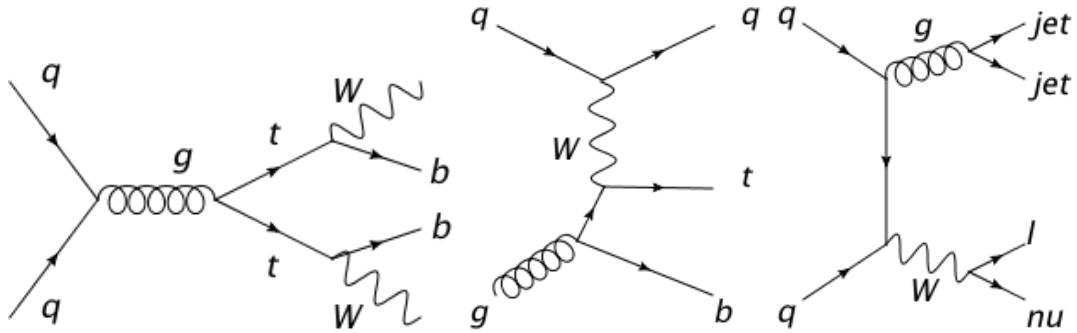


FIGURE 6.3: Feynmann diagrams showing major backgrounds.

### 6.3.2 $Z+jets$

The contribution from the events where  $Z$  boson is produced in association with two  $b$  jets is largely suppressed by requiring only one lepton in the event. However, it can happen that one of the leptons from  $Z$  decay escapes the detection or is missidentified which possibly causes significant missing energy. Such events are than passing all selection criteria and have to be taken into account in the final cross section measurement.

### 6.3.3 $W+light\;jets$ and $W+charm$

$W+jets$  is the major background before applying the  $b$ -tagging criteria which can be visible in figure 6.5. Both shape and normalization agree well between data and Monte Carlo for several distributions shown. Very tight  $b$ -tag selection reduces both  $W+light\;jets$  and  $W+charm$  to almost negligible levels.

### 6.3.4 QCD

QCD background arises from QCD multijet events containing soft leptons. An example of such event is shown in the left part of figure 6.6. This is one of the more challenging backgrounds as it is difficult to simulate significant amount of such events without restrictions. Therefore, the contribution of QCD events in the signal region is determined from data. The illustration of the method is shown in the right part of figure 6.6. Two uncorrelated

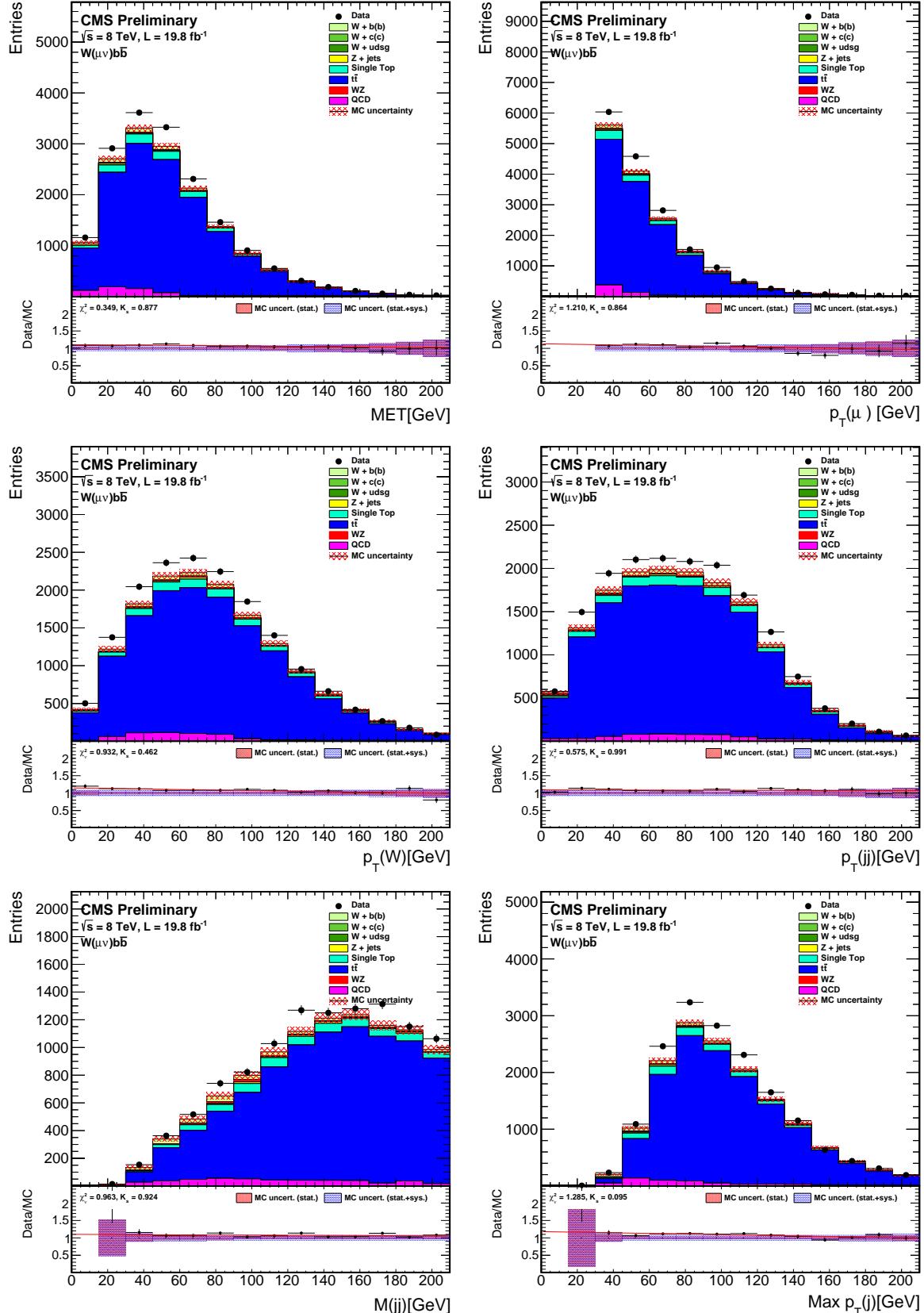


FIGURE 6.4: Top quark control region. Good shape agreement between data and simulation is observed, however simulation normalization is smaller than expected from data.

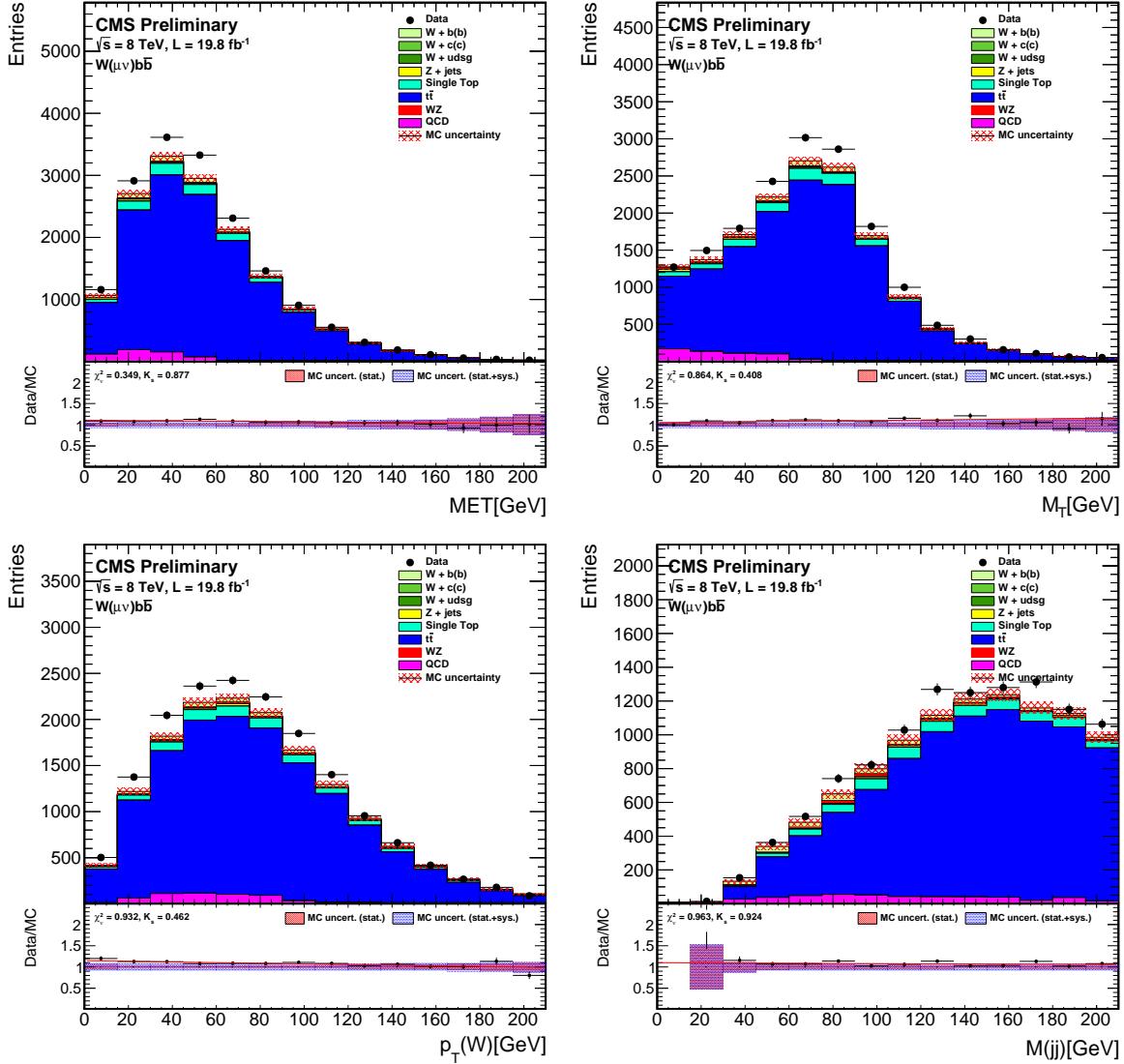


FIGURE 6.5: PLACEHOLDER - Distribution obtained using Wbb event selection before applying b-tagging criteria.

variables are chosen, in this case transverse mass and lepton isolation. Transverse mass distribution shows Jakobian peak for the events containing a W boson, thus it is a natural choice for discrimination from the non W boson final states. In the case of QCD multijet, transverse mass values concentrate in the low region of Signal region is shown in the region A. Control sample dominated by QCD events is created by inverting the lepton isolation cut to  $I_{rel}^{PF} > 0.2(0.15)$  for muons(electrons) which is shown as region C. The rest of the selection criteria in the control sample is the same as in signal region. It is assumed that the QCD distribution has the same shape in regions A and C. The obtained sample is

relatively clean, the shape of the distribution is determined by subtracting the simulation events that pass the selection. Normalization of the distribution is determined from the  $M_T$  region below 30 GeV. Fake rate is determined from regions B and D by subtracting simulation number of simulated events that pass the cuts from the number of data events. The final normalization is then expressed as:

$$QCD^A = \frac{N_{data}^B - N_{MC}^B}{N_{data}^D - N_{MC}^D} \times QCD_{data}^C \quad (6.1)$$

where  $N_{data}^B$  and  $N_{data}^D$  are the number of data events in data in regions B and D respectively, and  $N_{MC}^C$  and  $N_{MC}^D$  are the number of MC events in regions B and D respectively. The signal region before and after the QCD contribution determination is shown in figure 6.7.

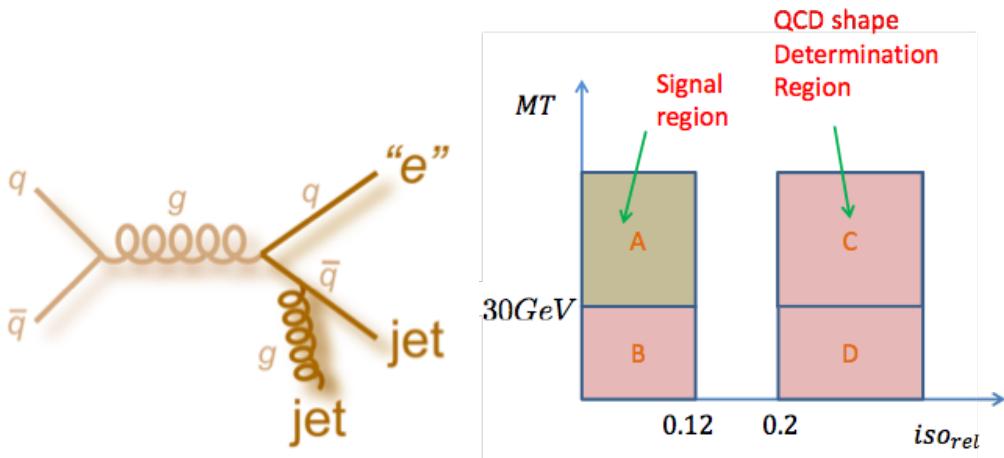


FIGURE 6.6: An example of QCD event which looks like signal event (left) and illustration of ABCD method used for QCD background determination (right)

### 6.3.5 Other backgrounds

Other backgrounds include processes with final states that match the final state of the signal. One of such signals is WZ where W decays leptonically and Z decays in a pair of b quarks. Another example is the production of Higgs boson in association with W boson

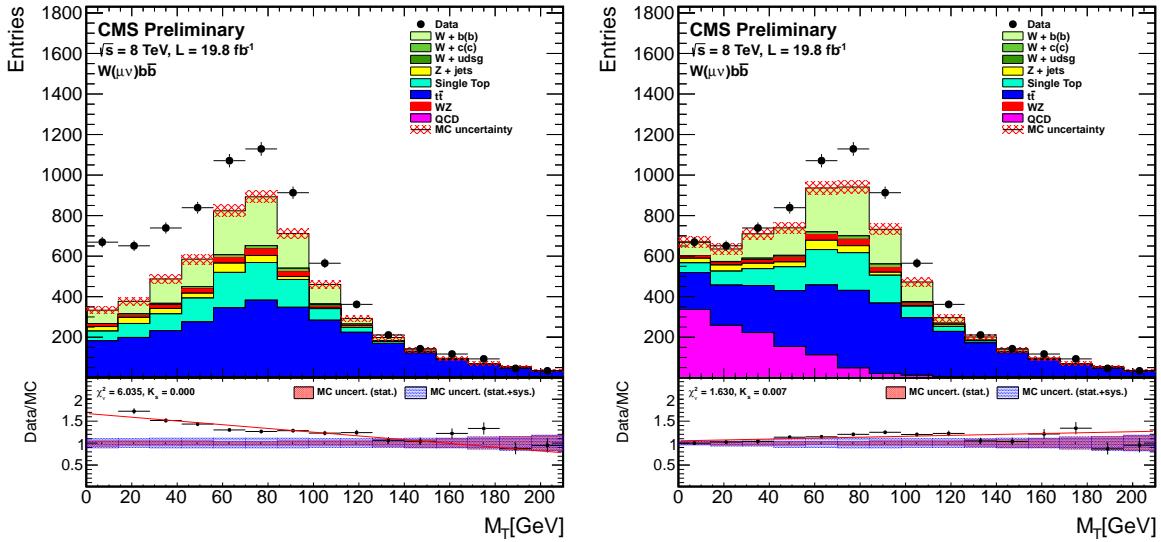


FIGURE 6.7: Transverse mass distribution before(*left*) and after QCD background determination(*right*).

where Higgs also decays to a pair of quarks. Such backgrounds are called irreducible backgrounds.

## 6.4 Monte-Carlo corrections

### 6.4.1 Pileup

In proton-proton collisions at high beam intensities, there is a high probability that multiple interactions could happen. These additional interactions are usually referred to as pileup interactions. The average number of additional interactions during 2012 was 21 with some events going up to 70. With these conditions it is important to be able to recognize the signature from such interactions. Usually pileup originates from low- $p_T$  QCD jets. The identification of pileup jets as well as their removal is described in detail in [68].

Simulated events have different distribution for number of pileup interaction with respect to data. This occurs because when generating simulated events, it was difficult to

predict the exact pileup distribution in data. Therefore, simulated events to be reweighed to match the distribution in data. The data pile up distribution in the collision period was estimated assuming total proton-proton cross section of 68 mb. For each simulated event, a weight  $w_{PU}$  is derived based on the number of pileup events provided by the generator. Figure 6.8 shows number of pile-up events before and after the reweighting procedure for signal events in the muon channel. The agreement between data and simulation has improved after the procedure which is also visible in the ratio plot.

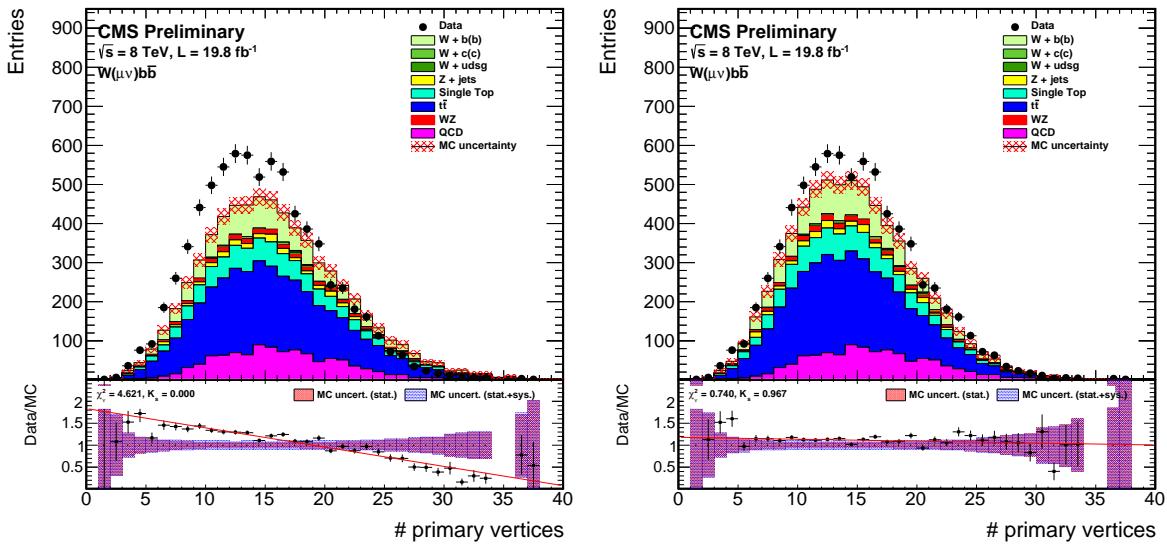


FIGURE 6.8: Number of primary vertices before (*left*) and after (*right*) the pileup reweighting procedure.

#### 6.4.2 Lepton efficiency measurement

Events used for cross section measurement are required to pass certain triggers in order to be selected. However, trigger selection is not 100% efficient and the selection efficiency has to be measured. The following steps in the analysis like reconstruction and isolation have some inefficiencies as well. Efficiency estimation from simulation shows large systematic errors due to inaccuracy in signal modeling and detector response. This was the main motivation for development of the fully data-driven efficiency estimation called *Tag and probe*. Using well-known mass resonances, such as Z boson mass resonance, a

selection criteria is applied to the decay products. *Tag* lepton is the one passing very tight selection cuts with low missidentification probability. The efficiency for certain cuts is than measured by counting *probe* leptons which are leptons passing this looser cut divided by number of all leptons. Probes are selected in such way that the invariant mass of the two leptons falls into the Z mass resonance. The following relation is used for the measurement:

$$\epsilon = \frac{N_{pass}}{N_{pass} + N_{fail}} \quad (6.2)$$

Final selection contains a number of events where Z boson was not actually produced and which have to be subtracted. Both signal and background contributions are parametrized their relative contributions are estimated using Maximum likelihood fit. For signal events a convolution of Z generator shape with a Gaussian is used to take into account the detector effects, while for background parametrization, a combination of exponential function and polinomial was used.

The efficiency was measured as a function of pseudorapidity and transverse momentum of a passing probe. Trigger, identification and isolation criteria were used in electron and muon channels separately. Both data and Monte Carlo efficiencies were measured and their ratio was used as a scale factor for each event in order to match simulated lepton efficiencies to measured data. Muon identification and isolation efficiency measurement for data and MC is shown in figure 6.9 while trigger efficiency measurement is shown in figure 6.10.

### 6.4.3 *b*-tagging scale factors

CMS simulations describe very well the detector performance, however, it can be difficult to accurately model all parameters used in b-tagging algorithms. Procedure used to identify b-jets is described in 5.4.4 and it depends on track reconstruction efficiency, tracking resolution and other tracking related parameters. Efficiency and missidentification probability are functions of transverse momentum and pseudorapidity of a jet. Therefore, it is

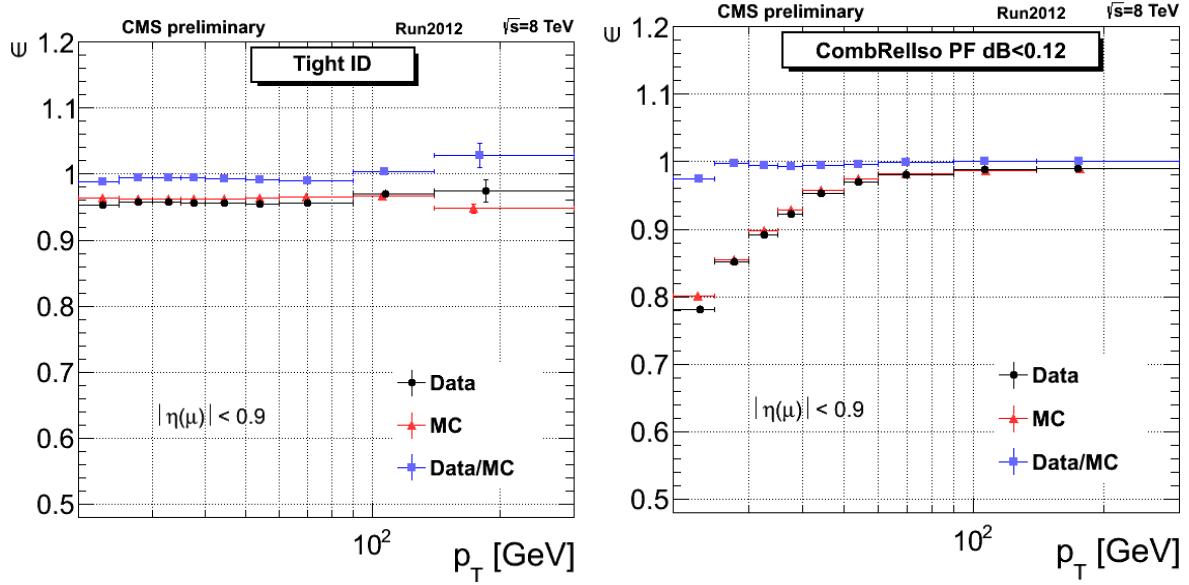


FIGURE 6.9: Muon identification and isolation efficiencies as a function of a probe  $p_T$  using *tag and probe* method. The measurement shown here is for barrel part of the detector ( $\eta < 0.9$ ), but similar measurements were performed for the region between barrel and endcap ( $0.9 < \eta < 1.2$ ) and endcap ( $1.2 < \eta < 2.1$ ) separately.

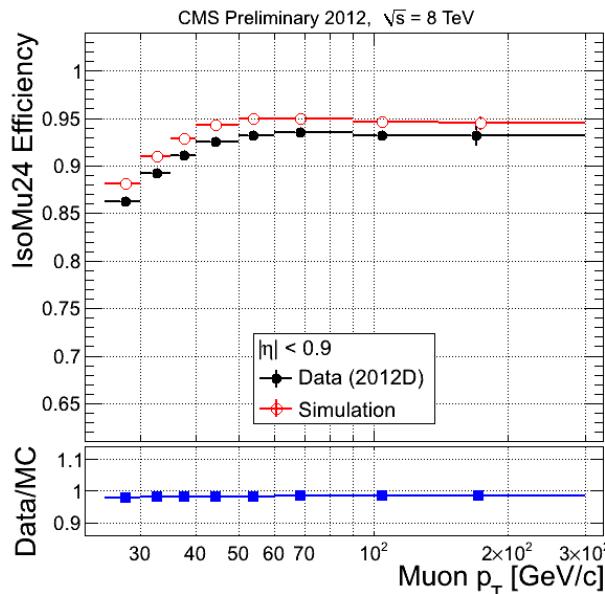


FIGURE 6.10: Muon trigger efficiency for HLT IsoMu24 using *tag and probe* method of barrel part of the detector ( $\eta < 0.9$ ).

very important to determine the b-tagging efficiency from data. The obtained corrections are applied to simulated events as scale factors which is ratio between efficiency measured in collisions  $\epsilon_b^{data}$  and efficiency from simulated events  $\epsilon_b^{MC}$ :

$$SF_b = \frac{\epsilon_B^{data}}{\epsilon_b^{MC}} \quad (6.3)$$

Scale factor determination has to be performed using b-jet enriched sample such as  $t\bar{t}$  or multijets events with jet containing a muon within a  $\Delta R < 0.4$  cone from the jet axis. The choice of the muon within jet relies on the fact that in B hadron decays, the semileptonic branching ratio is much higher than that for other hadrons ( $\sim 20\%$  when including  $c\rightarrow$  decays) and such jets are much more likely to arise from B hadron decay. With very high muon detection efficiency at CMS, it is relatively easy to obtain a clean sample with jets containing nonisolated muons. Additional kinematic criteria is applied to the selected muons to determine the efficiency in different  $p_T$  ranges of the jets. Other efficiency measurement is performed using  $t\bar{t}$  enriched sample by cutting on number of selected jets and isolated leptons in the event and approximating that the  $t$  quark decays to  $W + b$  exclusively. By combining the results from both measurements, scale factors were obtained as a function of jet  $p_T$  together with statistical and systematic error for each  $p_T$  bin. The same strategy is used to obtain missidentification rates by using the inverted cut on b-tag discriminator. The behavior of both scale factors is approximated by an analytical parametrization as a function of jet  $p_T$ . The usage of the scale factors depends on the number of b-tagged jets in the event. In this analysis two b-tagged jets are required and weight for each event is derived as:

$$w(2|2) = SF_{b||light}(1st \text{ jet}) \times SF_{b||light}(\text{second jet}) \quad (6.4)$$

where  $w(2|2)$  is event weight with 2 jets where both jets are b-tagged. The choice between  $SF_b$  and  $SF_{light}$  depends on the flavor of the jet in the simulation. A jet is considered a b jet if there is a B hadron present among the jet constituents within a cone of 0.4 from the jet axis.

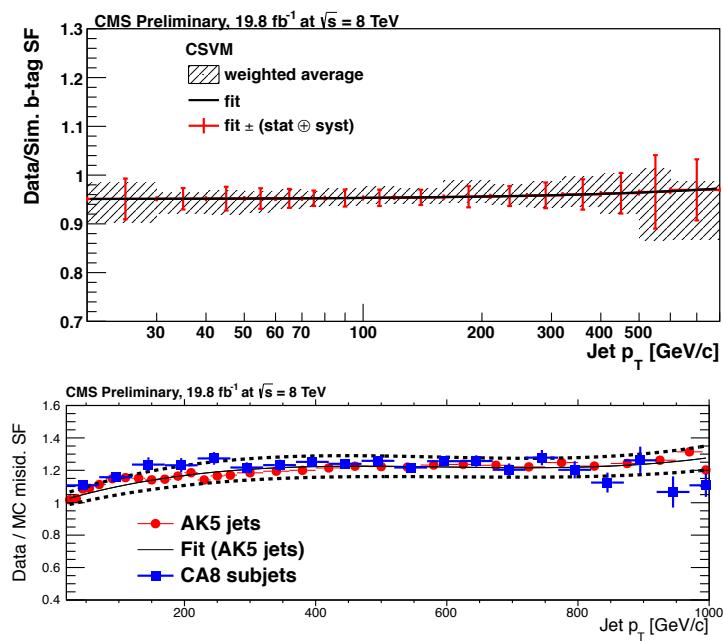


FIGURE 6.11: B-tagging scale factors and mistag scale factors



# Chapter 7

## Cross section measurement

Cross section for  $\text{pp} \rightarrow \text{W} + \text{bb}$  process is measured separately in electron and muon channel using this simple relation:

$$\sigma = \frac{N_{sig}}{A \times \epsilon \cdot \mathcal{L}} \quad (7.1)$$

where  $N_{sig}$  is the number of signal events after applying selection criteria described in 6.2,  $A \times \epsilon$  is the detector acceptance and efficiency (section 7.3) and  $\mathcal{L}$  is the total integrated luminosity for the period over which the data was collected as described in 4.8.

In the following sections the fitting procedure for the extraction of signal events is described taking into account the systematic effects. After the subtraction of background contribution, final  $N_{sig}$  for cross section calculation was obtained. The procedure for calculation of  $A \times \epsilon$  is described in 7.3. Finally, cross section are computed and compared to theoretical predictions.

### 7.1 Fit methodology

After applying selection criteria described in section 6.2, transverse mass distributions are obtained for both electron and muon channels. Modeling signal region was done using

four-flavor scheme and five-flavor scheme separately. These samples are described in detail in 6.1. The comparison between the two samples is given in figure 7.1.

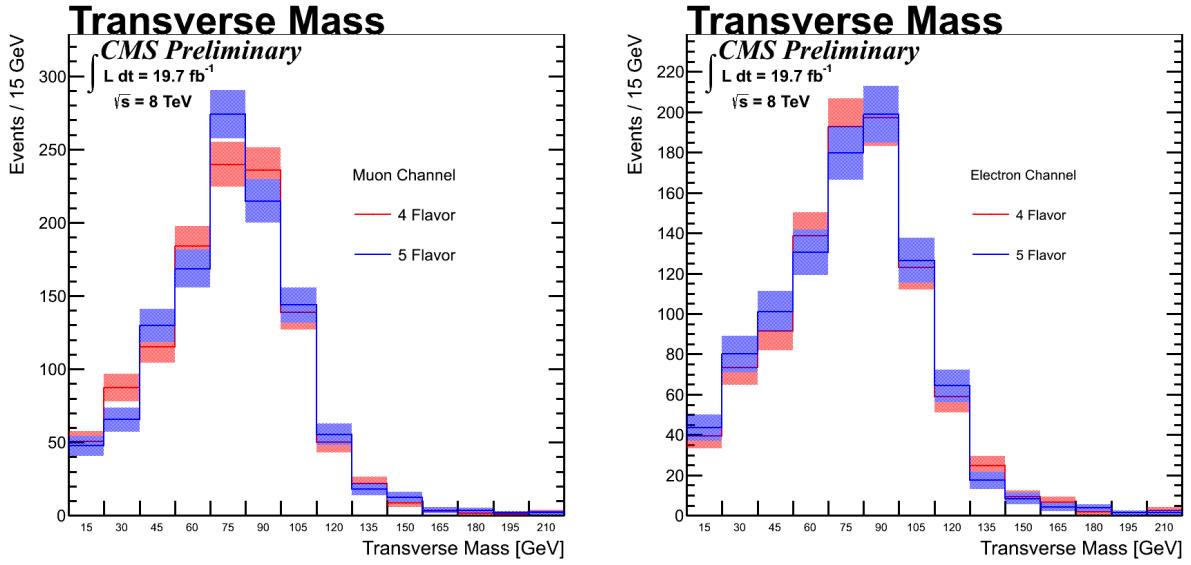


FIGURE 7.1: Shape comparison of the two samples used to obtain the number of signal events. Left figure shows muon channel while the right figure shows the electron channel.

Final signal yields are extracted using a binned maximum likelihood fit. The details of the fitting procedure are described in [69]. Simultaneous fit was performed in Wbb region and in TT multijet control region. The result of the fit is represented as *signal strength*  $\mu$  which is the ratio between the cross section under test and the theoretical cross section, in this case Wbb cross-section. The predictions for both, signal and background yields, depend on various uncertainties which are included in the fit as nuisance parameters. Likelihood function is constructed as:

$$\mathcal{L}(\text{data}|\mu, \theta) = \text{Poisson}(\text{data}|\mu \cdot s(\theta) + b(\theta)) \cdot p(\tilde{\theta}|\theta) \quad (7.2)$$

In this expression "data" represents the actual measurements,  $\theta$  is a set of nuisance parameters describing the uncertainties while  $s(\theta)$  and  $b(\theta)$  describe signal and background yields respectively, which depend on the nuisance parameters. Probability density functions(*pdf*)  $p(\theta)$  with some  $\tilde{\theta}$  as the best estimate of the parameter set describing each of the nuisances are used to characterize the nuisances. Different choices of *pdf* in general include

flat distribution, Gaussian, log-normal and gamma distribution. In this analysis, systematic uncertainties are treated in two different ways. In cases where systematic uncertainty doesn't change the shape of the fitted distribution, log-normal *pdf* is used because it is suitable for the description of the positively defined variables like cross-section, luminosity of cut efficiency. The width of the log-normal is defined by the parameter  $\kappa$ :

$$\rho(\theta) = \frac{1}{\sqrt{2\pi} \ln(\kappa)} \exp\left(\frac{(\ln(\theta/\tilde{\theta}))^2}{2(\ln\kappa)^2}\right) \frac{1}{\theta}. \quad (7.3)$$

The value of the  $\kappa$  implies by how much an observable can be larger or smaller, both deviations having a chance of 16%. Other way of treating systematic uncertainties is by producing two additional input shapes for each process affected by the uncertainty, by shifting up and down that parameter by one standard deviation. When building the likelihood, each shape uncertainty is associated to a nuisance parameter taken from a unit Gaussian distribution, which is used to interpolate or extrapolate using the specified histograms. After the construction of the likelihood function, standard maximum likelihood procedure is performed yielding the values of  $\theta$  and  $\mu$  which make the measured data most probable. Next section lists the major sources of the systematic uncertainties together with the strategy used for determination of the corresponding nuisance *pdf*.

## 7.2 Systematic uncertainties

Systematic uncertainties on the expected signal and background yields and shapes affect the final distribution used to obtain  $N_{sig}$ . For a given systematic variation, a new set of signal and background templates was created which may differ both in shape and normalization from the original template. These shape variations are included in the final fit. Several sources of systematic variations have been considered.

## Luminosity uncertainty

Luminosity measurement is performed using cluster counting in the Pixel detector described in section 4.8. An uncertainty of 2.2% for luminosity measurement during 2012 datataking is reported by the CMS luminosity group [39].

## Jet energy scale uncertainty

The source of jet energy scale uncertainty arises from different jet energy corrections applied to unify the detector response in energy and pseudorapidity as described in section 5.4.2. The uncertainty for each level of corrections is estimated separately and added in quadrature to get the final uncertainty [48]. The jet energy scale for each jet is varied within one standard deviation of the applied jet energy corrections. and the efficiency of the analysis selection is recomputed to assess the systematic variation on the normalization and shape of the signal and all background components.

## Jet energy resolution

Jet energy resolution in simulation is smeared in order to take into account differences between data and Monte Carlo. The uncertainty on the applied smearing factors is used to produce modified signal and background templates. These modified templates are then used in the final fit.

## Jet b-tagging efficiencies

Jet b-tagging efficiencies are determined from the  $t\bar{t}$  Monte Carlo events for the jets selected in the analysis as described in section 6.4.3. These efficiencies are applied as weight factors for each event and depend on  $p_T$ ,  $\eta$  and flavour of the selected jets. To estimate the effect of the uncertainties of the efficiency determination, each scale factor is shifted up and down by the corresponding uncertainty, and event weights are recalculated.

The procedure is done separately for b jet efficiencies and for light jet mistag rate. The uncertainty for jets from c quark is taken to be twice as the one from b jets since there is no proper measurement for this uncertainty. The scale factor variation is found to be of the order of few percent.

### Lepton scale factors

Muon and electron trigger, reconstruction, and identification efficiencies are determined in data using the standard tag-and-probe technique with Z bosons as described in 6.4.2. The effects of the corresponding uncertainties were determined by varying the corresponding scale factors by one standard deviation from the tag-and-probe technique.

### Lepton energy scale

The lepton energy scale measurement uncertainty corresponds to 1% for muons in the whole detector and electrons in the barrel region. Systematic uncertainty od 2.5% is associated to the electrons in the endcap region. The systematic effect is evaluated by varying lepton energy scale for each lepton within one standard deviation and creating corresponding transverse mass distributions used in the final fit.

### Unclustered missing energy

The uncertainty on missing energy measurement from unclustered energy, e.g. jets with  $p_T > 10 \text{ GeV}$  and  $|\eta| < 4.7$  is estimated. The energy scale of such jets is varied by 10%, which is propagated into the calculation of missing energy. New transverse mass distributions are created and taken into account in the final fit.

### MC samples normalizations

The finite size of the signal and background MC samples are included in the normalization uncertainties. Normalizations for each of the Monte-Carlo samples are also allowed to vary within the uncertainties of measured Standard model cross-sections. Cross section uncertainties are summarized in the table 7.1

Process	Cross section uncertainty
W+c(c)	8.1%
W+udsg	13.2%
Z+jets	7.9%
Single Top	5.4%
T <bar>T</bar>	7.4%
VV	8.1%

TABLE 7.1: Standard model cross section uncertainties used in the evaluation of MC normalization systematic effect.

Sum of the Monte Carlo distributions for each source of systematic uncertainty included in the final fit are shown in the figure 7.2. The largest systematic variations comes from the b-tagging uncertainties. A more detailed study of the final signal strength dependence on the systematic variations is shown in section 7.4.1.

## 7.3 Acceptance and efficiency

Due to the limitations of the detector, not all produced signal events will be detected. Some final state particles will end up outside the functional part of the detector. The fraction of the phase space covered with functional detector for signal final state particles is called *acceptance*. Usually this part of the phase space is called *fiducial region* and applied cuts are summarized in table 7.2. However, fraction of the events that fall into the fiducial volume will not be detected due to trigger and reconstruction inefficiency or selection cuts imposed by trigger or the analysis. Usually, acceptance and efficiency are

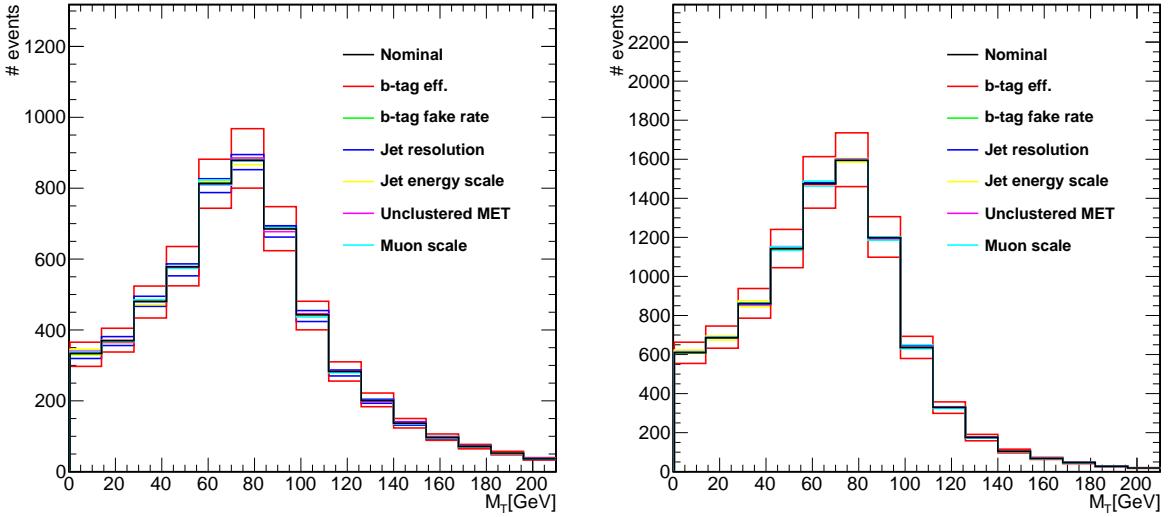


FIGURE 7.2: Shape of the transverse mass distribution in the muon channel for each systematic variation in both, signal region (left) and TT control region (right).

Variable	Cut
Lepton $p_T$	$> 30 \text{ GeV}$
Lepton $ \eta $	$< 2.1$
Jet $p_T$	$> 25$
Jet $ \eta $	$< 2.4$
$\Delta R(\text{jet}, \text{B hadron})$	$< 0.5$

TABLE 7.2: Fiducial cuts used for cross section measurements.

estimated as a single quantity which is a product of these two numbers, defined as:

$$A \times \epsilon = \frac{\text{number of selected Wbb events}}{\text{number of generated Wbb events in the fiducial volume}} \quad (7.4)$$

This ratio is computed using simulated Wbb sample for each of the channels separately. Number of selected events is obtained by applying the selection cuts described in 6.2. Number of generated hits is obtained by applying generator-level cuts summarized in the table 7.2. With this ratio being derived from simulation, it is necessary to correct it for the difference between data and Monte-Carlo. These corrections include pile-up  $w^{PU}$ , lepton trigger, reconstruction and identification scale factors  $w^{lep}$ , and b-tagging scale factors

$w^{b\text{-tag}}$  all described in 6.4. With all the corrections,  $A \times \epsilon$  for each channel becomes:

$$A \times \epsilon = \frac{\sum^{\text{sel}} w^{\text{lep}} w^{\text{PU}} w^{b\text{-tag}}}{N_{\text{fiducial}}^{\text{gen}}} \quad (7.5)$$

Obtained results are summarized in the table 7.3 for each channel.

Channel	$A \times \epsilon$
Muon channel	$9.14 \pm 0.01 \%$
Electron channel	$7.72 \pm 0.01 \%$

TABLE 7.3: Results of the  $A \times \epsilon$  measurement for both, muon and electron channel together with the statistical uncertainty.

## 7.4 Results

Transverse mass distributions in signal region and  $t\bar{t}$  region were fitted simultaneously in order to obtain the signal strength for Wbb. Yields before and after the fit are shown in Table 7.4 for muon channel and 7.5 for the electron channel. Contributions from different sources of systematic variations are summarized in table 7.6 for the muon channel and table 7.7 for the electron channel. Table also shows relative contribution to the signal strength uncertainty for each source of uncertainty together with the change in total systematic uncertainty when removing specific source of uncertainty.

TABLE 7.4: Yields obtained in the muon channel before and after the fitting procedure.

Sample	Prefit yields	Fitted yields	Prefit yields( $M_T > 45\text{GeV}$ )	Fitted yields( $M_T > 45\text{GeV}$ )
W+bb	$1129.7 \pm 25.5$	$1629.1 \pm 80.5$	$871.7 \pm 29.5$	$1254.5 \pm 75.0$
W+cc	$68.9 \pm 7.4$	$81.9 \pm 9.8$	$52.8 \pm 7.3$	$62.6 \pm 9.1$
W+udscg	$47.1 \pm 10.8$	$61.4 \pm 15.6$	$35.8 \pm 6.0$	$46.8 \pm 14.3$
Z+jets	$208.7 \pm 22.4$	$217.0 \pm 18.8$	$121.8 \pm 11.0$	$126.6 \pm 14.4$
Single Top	$913.1 \pm 16.7$	$986.4 \pm 33.4$	$700.2 \pm 26.5$	$748.1 \pm 30.4$
$T\bar{T}$	$3093.2 \pm 13.0$	$3488.0 \pm 36.2$	$2499.6 \pm 50.0$	$2713.9 \pm 32.8$
VV	$145.2 \pm 3.1$	$151.8 \pm 6.1$	$108.4 \pm 10.4$	$113.1 \pm 5.7$
QCD	$1110.5 \pm 28.2$	$908.4 \pm 33.6$	$293.9 \pm 17.1$	$240.4 \pm 11.6$
Sum	$6716.3 \pm 50.7$	$7524.1 \pm 103.8$	$4684.2 \pm 68.4$	$5306.0 \pm 91.0$
Data		$7481.0$		$5372.0$

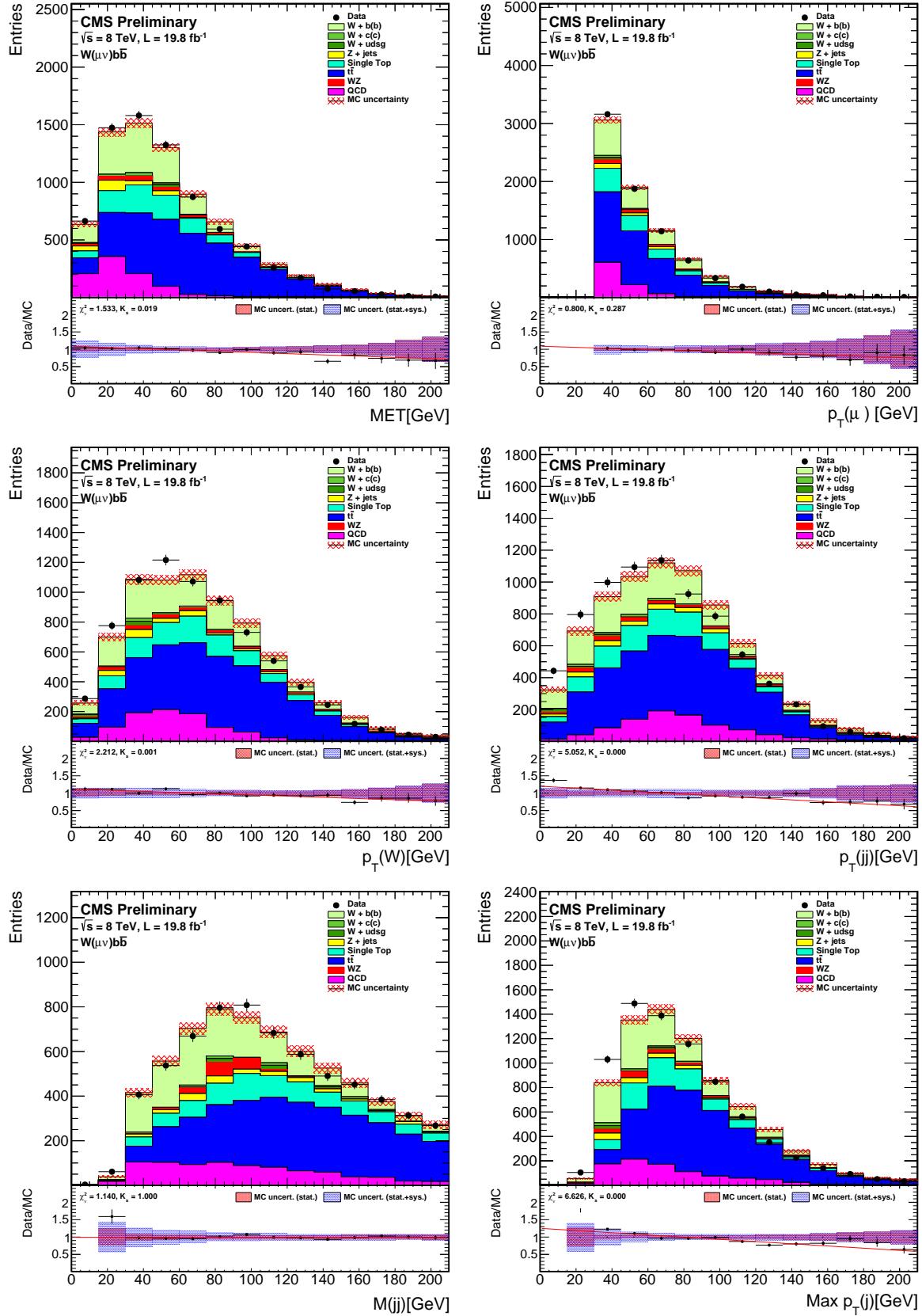


FIGURE 7.3: Muon channel postfit distributions.

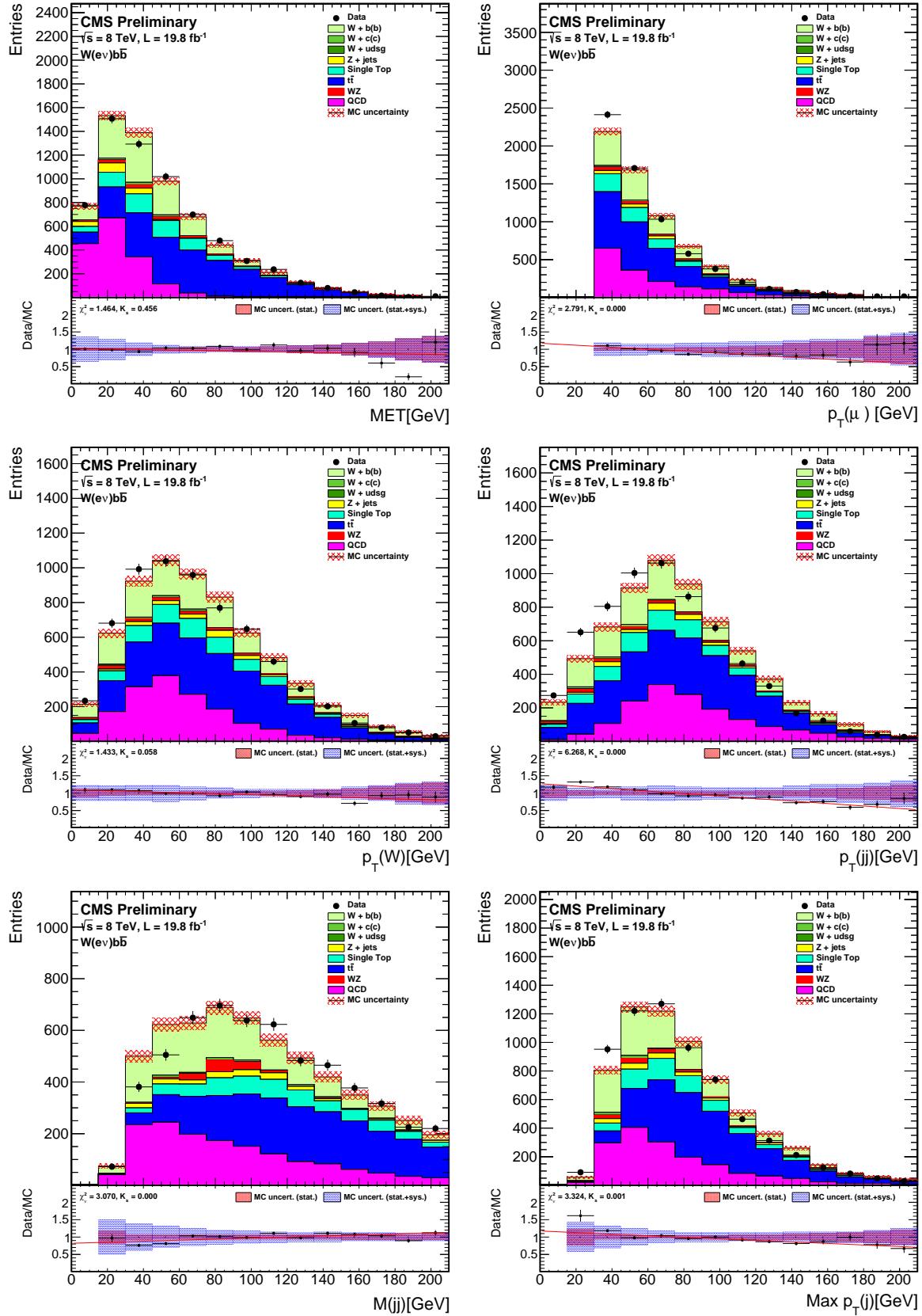


FIGURE 7.4: Electron channel postfit distributions.

TABLE 7.5: Yields obtained in the electron channel before and after the fitting procedure.

Sample	Prefit yields	Fitted yields	Prefit yields( $M_T > 45\text{GeV}$ )	Fitted yields( $M_T > 45\text{GeV}$ )
W+bb	$940.6 \pm 23.3$	$1516.6 \pm 74.6$	$715.8 \pm 26.8$	$1160.5 \pm 68.6$
W+cc	$69.5 \pm 8.4$	$62.3 \pm 8.4$	$61.0 \pm 7.8$	$54.0 \pm 7.6$
W+udscg	$15.4 \pm 5.2$	$13.7 \pm 2.2$	$12.5 \pm 3.5$	$11.0 \pm 2.0$
Z+jets	$201.1 \pm 21.4$	$184.4 \pm 10.7$	$88.7 \pm 9.4$	$94.5 \pm 7.1$
Single Top	$719.6 \pm 14.8$	$672.8 \pm 20.1$	$547.6 \pm 23.4$	$509.9 \pm 18.6$
$t\bar{t}$	$2496.8 \pm 11.6$	$2369.9 \pm 30.3$	$2056.3 \pm 45.3$	$1879.6 \pm 27.6$
VV	$110.2 \pm 2.7$	$106.9 \pm 3.9$	$85.4 \pm 9.2$	$84.1 \pm 3.7$
QCD	$1506.7 \pm 29.1$	$1622.5 \pm 45.5$	$727.0 \pm 27.0$	$778.1 \pm 27.3$
Sum	$6060.0 \pm 48.1$	$6549.0 \pm 95.7$	$4294.3 \pm 65.5$	$4571.7 \pm 81.7$
Data		6530.0		4639.0

TABLE 7.6: Information about each source of systematic uncertainty together with the information whether just normalization or shape is included in the final fit. The table also shows the uncertainty on signal and background yields and the relative contribution to the signal strength uncertainty. Due to correlations, the total systematic uncertainty is smaller than the quadrature sum of individual uncertainties. The last column shows the decrease in total systematic uncertainty when removing specific source of uncertainty.

Source	Type	Event yield uncertainty range (%)	Individual contribution to $\mu$ uncertainty (%)	Effect of removal on $\mu$ uncertainty (%)
b-tag efficiency	shape	6.72	0.99	3.54
Lepton ID/Iso/Trig	shape	0.57	0.28	< 0.01
Jet resolution	shape	0.03	< 0.01	0.10
Jet energy scale	shape	0.03	3.41	0.53
Unclustered MET	shape	0.00	< 0.01	0.06
Muon energy scale	shape	1.10	0.15	0.02
Luminosity	norm.	2.60	0.68	0.06
Monte Carlo statistics	norm.	0.75	3.62	3.58

#### 7.4.1 Effects of the systematics

#### 7.4.2 Tests of the fit stability

Additional tests were performed in order to verify consistency of the obtained signal strength. This was done by fitting different combinations of distributions in both signal region and  $t\bar{t}$  control region. Additional distributions include missing energy in the signal region and  $t\bar{t}$  control region shown in figures 6.1 and 6.4 and invariant mass of third and fourth jet in  $t\bar{t}$  control region. All distributions show good agreement in shapes between data and Monte Carlo. Fitting procedure is performed in the muon channel as previously

TABLE 7.7: Same as table 7.6 for the electron channel.

Source	Type	Event yield uncertainty range (%)	Individual contribution to $\mu$ uncertainty (%)	Effect of removal on $\mu$ uncertainty (%)
b-tag efficiency	shape	6.80	1.01	2.13
lepton ID/iso/trigg	shape	0.57	0.35	0.05
Jet resolution	shape	0.03	0.02	0.24
Jet energy scale	shape	0.05	0.30	3.71
Unclustered MET	shape	0.00	< 0.01	1.15
Muon scale	shape	1.09	0.15	0.24
Luminosity	norm.	2.60	0.30	11.11
Monte Carlo statistics	norm.	0.75	1.16	11.96

described. Obtained signal strengths are summarized in Table 7.8 and are found to be consistent with the transverse mass fit.

Fitted distribution ( $Wbb/t\bar{t}$ )	Signal Strength	Yield Ratio
$M_T/M_T$	$1.34 \pm 0.15$	1.55
$M_T/E_T^{miss}$	$1.31 \pm 0.14$	1.52
$M_T/M(j_3j_4)$	$1.35 \pm 0.16$	1.53
$E_T^{miss}/M_T$	$1.43 \pm 0.21$	1.64
$E_T^{miss}/E_T^{miss}$	$1.33 \pm 0.17$	1.53
$E_T^{miss}/M(j_3j_4)$	$1.38 \pm 0.21$	1.55

TABLE 7.8: UPDATE NUMBERS! - Signal strengths obtained by fitting different distributions. Signal strengths are found to be consistent with each other.

### 7.4.3 Cross section measurement

### 7.4.4 Theoretical calculations and comparison with the measurement

# Chapter 8

## Conclusions



# **Chapter 9**

**Prosireni sazetak - Mjerenje udarnog  
presjeka zajedničke produkcije  $W$   
bozona i para b kvarkova**

## **9.1 Main Section 1**



# Appendix A

## Lorentz angle measurement in Pixel detector

### A.1 Grazing angle method

Lorentz angle is measured by using grazing angle method described in detail in [70]. From the individual signals in the detector, using reconstruction algorithms, tracks of muon candidates are obtained. From these reconstructed track it is possible to extract the entry point ( $x_{reco}, y_{reco}$ ) to each layer of the detector. Distance between reconstructed entry point and the actual hit in the detector is then defined as ( $\Delta x, \Delta y$ ):

$$\Delta x = x_{center} - x_{reco} \quad (\text{A.1})$$

$$\Delta y = y_{center} - y_{reco} \quad (\text{A.2})$$

where  $(x_{center}, y_{center})$  is the position of each individual pixel center in the observed cluster. Drift of the electrons can be determined using three impact angles defined in the following way:

$$\tan\alpha = \frac{p_z}{p_x} \quad (\text{A.3})$$

$$\tan\beta = \frac{p_z}{p_y} \quad (\text{A.4})$$

$$\tan\gamma = \frac{p_x}{p_y} \quad (\text{A.5})$$

where  $p_x, p_y$  and  $p_z$  are momentum components in local coordinate system which are calculated from reconstructed track parameters (Fig. A.1).

Drift of the electrons depends on the depth at which electrons are created. Depth of the

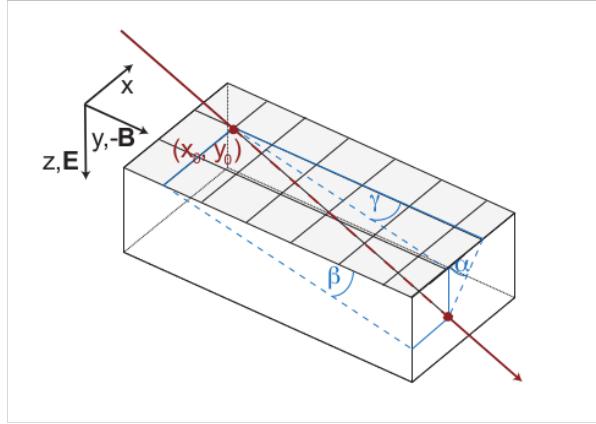


FIGURE A.1: Angle definitions for grazing angle method.

electron production  $z$  and drift due to magnetic field  $d$  are defined:

$$z = \Delta y \tan\beta \quad (\text{A.6})$$

$$d = \Delta x - \Delta y \tan\gamma \quad (\text{A.7})$$

This procedure is repeated for each pixel over many tracks in order to obtain charge drift distance vs depth. The Lorentz angle is the slope of this distribution. Without a magnetic field, the direction of the clusters largest extension is parallel to the track projection on the (x, y) plane. The average drift distance of an electron created at a certain depth is obtained from Fig. A.2. A linear fit is performed over the total depth

of the detector excluding the first and last  $50 \mu\text{m}$  where the charge drift is systematically displaced by the finite size of the pixel cell (Fig:A.3).

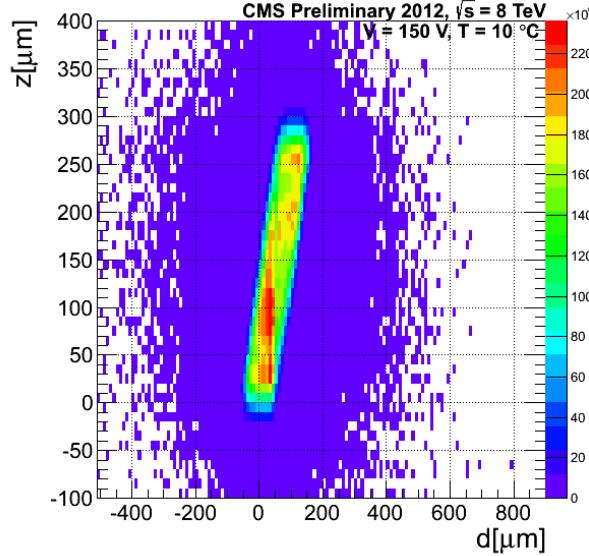


FIGURE A.2: Depth at which electrons in silicon bulk were produced as a function of Lorentz drift.

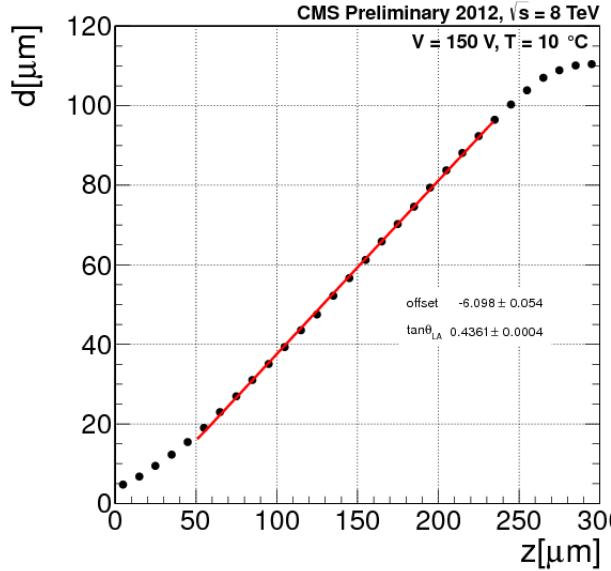


FIGURE A.3: The average drift of electrons as a function of the production depth. Slope of the linear fit result is the  $\tan\theta_L$ .

In order to obtain a good measurement, it is important to use clean tracks. Therefore, it required to have a well reconstructed muon tracks with  $p_T > 3\text{GeV}$  and  $\chi^2/\text{ndof} < 2$

which are required to have shallow impact angle with respect to local  $y$  direction with cluster size of at least 4 pixels in this direction. Summary of the selection criteria can be found in table A.1.

TABLE A.1: Selection criteria for Lorentz angle measurement

Cluster size in $y$	$> 3$
Track $p_t$	$> 3\text{GeV}/c$
$\chi^2/\text{ndof}$	$< 2$
Hit residuals	$< 50\mu m$
Cluster charge	$< 120000e$

Figure A.4 shows how Lorentz angle changes with integrated luminosity. Results are shown for  $23\text{fb}^{-1}$  of delivered luminosity in 2012. Increase in Lorentz angle measured with grazing angle method has been observed in all layers, with largest effect ( 6%) visible in layer 1 over this period of data taking.

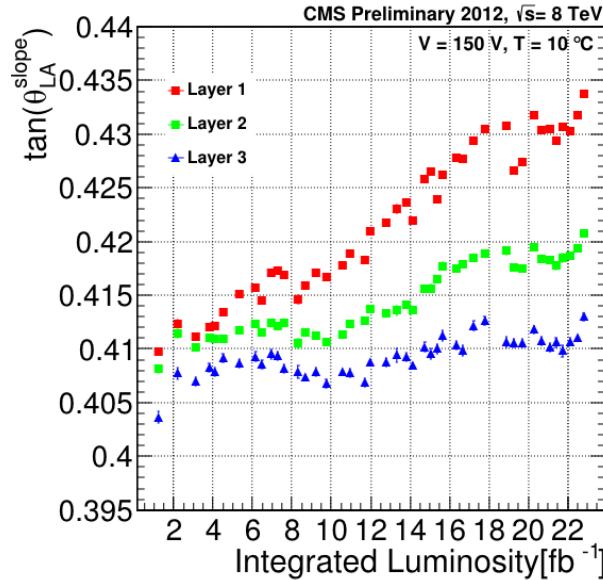


FIGURE A.4: Lorentz angle as a function of integrated luminosity for 2012.

## A.2 Minimum cluster size method (V-method)

The pixel cluster size in the drift direction depends on the incident angle and is minimal when incident angle is equal to the Lorentz angle. Thus, measuring the average cluster size in drift direction as a function of incident angle and obtaining a minimum of that distribution is an alternative and direct method of measuring the Lorentz angle. The method is usually referred to as V-method due to a shape of distribution which in the simple case can be approximated with formula

$$p_1 * \text{abs}(\tan(\theta) - p_0) + p_2 \quad (\text{A.8})$$

where  $p_0$ ,  $p_1$  and  $p_2$  are parameters obtained from the fit and  $p_0 = \tan(\theta_{LA})$ .

The method was successfully applied to cosmic muon tracks during CMS commissioning period in 2008 and again in 2015. The fit result is shown in figure A.5.

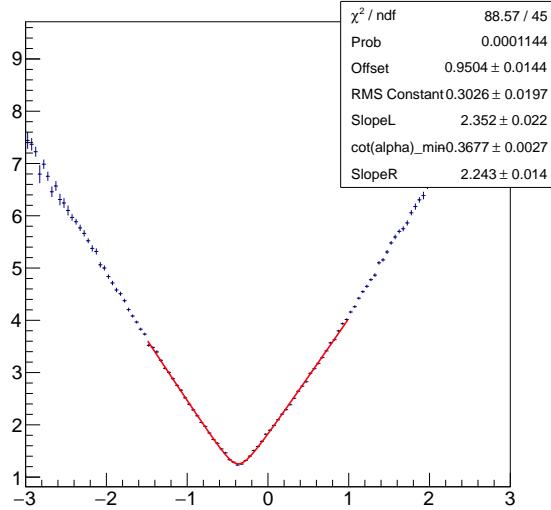


FIGURE A.5: An example of V-method fit.

Application to collision data is more challenging. Coordinates of a track passing through the detector, its incoming angle, and its  $p_T$  are correlated and therefore incoming angles from collision tracks have limited range. With standard running conditions the

value of Lorentz angle is at the edge of that range where tracks with very low  $p_T$  ( $<0.5$  GeV) dominate. Because of that average cluster size as a function of incoming angle cannot be described by a simple model like above mentioned for cosmic data. While results for collision data obtained with V-method are in general agreement with the default calculation, the uncertainty of the method at present is too big to be used as a viable alternative.

## **Appendix B**

### **Acceptance and efficiency error calculation**

Write your Appendix content here.



# Bibliography

- [1] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579 – 588, 1961. ISSN 0029-5582. doi: [http://dx.doi.org/10.1016/0029-5582\(61\)90469-2](http://dx.doi.org/10.1016/0029-5582(61)90469-2). URL <http://www.sciencedirect.com/science/article/pii/0029558261904692>.
- [2] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967. doi: 10.1103/PhysRevLett.19.1264. URL <http://link.aps.org/doi/10.1103/PhysRevLett.19.1264>.
- [3] Abdus Salam. Weak and Electromagnetic Interactions. *Conf.Proc.*, C680519:367–377, 1968.
- [4] G. Arnison et al. Experimental observation of isolated large transverse energy electrons with associated missing energy at  $s=540$  gev. *Physics Letters B*, 122(1):103 – 116, 1983. ISSN 0370-2693. doi: [http://dx.doi.org/10.1016/0370-2693\(83\)91177-2](http://dx.doi.org/10.1016/0370-2693(83)91177-2). URL <http://www.sciencedirect.com/science/article/pii/0370269383911772>.
- [5] G. Arnison et al. Experimental observation of lepton pairs of invariant mass around 95 gev/c<sup>2</sup> at the cern sps collider. *Physics Letters B*, 126(5):398–410, 1983. doi: 10.1016/0370-2693(83)90188-0. URL <http://www.scopus.com/inward/record.url?eid=2-s2.0-23044516930&partnerID=40&md5=8be07e2a211517aafdf5135170ca9591>.

- [6] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys.Lett.*, B716:1–29, 2012. doi: 10.1016/j.physletb.2012.08.020.
- [7] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys.Lett.*, B716:30–61, 2012. doi: 10.1016/j.physletb.2012.08.021.
- [8] D. Griffiths. *Introduction to Elementary Particles*. John Wiley & Sons, New York, USA, 1987.
- [9] Makoto Kobayashi and Toshihide Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog.Theor.Phys.*, 49:652–657, 1973. doi: 10.1143/PTP.49.652.
- [10] S. W. Herb, D. C. Hom, L. M. Lederman, J. C. Sens, H. D. Snyder, J. K. Yoh, J. A. Appel, B. C. Brown, C. N. Brown, W. R. Innes, K. Ueno, T. Yamanouchi, A. S. Ito, H. Jöstlein, D. M. Kaplan, and R. D. Kephart. Observation of a dimuon resonance at 9.5 gev in 400-gev proton-nucleus collisions. *Phys. Rev. Lett.*, 39:252–255, Aug 1977. doi: 10.1103/PhysRevLett.39.252. URL <http://link.aps.org/doi/10.1103/PhysRevLett.39.252>.
- [11] G. Arnison et al. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at  $s^{**}(1/2) = 540\text{-GeV}$ . *Phys.Lett.*, B122:103–116, 1983. doi: 10.1016/0370-2693(83)91177-2.
- [12] M. Banner et al. Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider. *Phys.Lett.*, B122:476–485, 1983. doi: 10.1016/0370-2693(83)91605-2.
- [13] John M. Campbell. Overview of the theory of  $W/Z + \text{jets}$  and heavy flavor. 2008.
- [14] Michelangelo L. Mangano. Production of  $W$  plus heavy quark pairs in hadronic collisions. *Nucl.Phys.*, B405:536–554, 1993. doi: 10.1016/0550-3213(93)90558-7.

- [15] Serguei Chatrchyan et al. Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks. *Phys.Rev.*, D89(1):012003, 2014. doi: 10.1103/PhysRevD.89.012003.
- [16] Serguei Chatrchyan et al. Evidence for the direct decay of the 125 GeV Higgs boson to fermions. *Nature Phys.*, 10:557–560, 2014. doi: 10.1038/nphys3005.
- [17] K.A. Olive et al. Review of Particle Physics. *Chin.Phys.*, C38:090001, 2014. doi: 10.1088/1674-1137/38/9/090001.
- [18] John M. Campbell, J.W. Huston, and W.J. Stirling. Hard Interactions of Quarks and Gluons: A Primer for LHC Physics. *Rept.Prog.Phys.*, 70:89, 2007. doi: 10.1088/0034-4885/70/1/R02.
- [19] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur.Phys.J.*, C63:189–285, 2009. doi: 10.1140/epjc/s10052-009-1072-5.
- [20] Michael Edward Peskin and Daniel V. Schroeder. *An introduction to quantum field theory*. Advanced book program. Westview Press Reading (Mass.), Boulder (Colo.), 1995. ISBN 0-201-50934-2. URL <http://opac.inria.fr/record=b1131978>.
- [21] Fabio Maltoni, Giovanni Ridolfi, and Maria Ubiali. b-initiated processes at the LHC: a reappraisal. *JHEP*, 1207:022, 2012. doi: 10.1007/JHEP04(2013)095, 10.1007/JHEP07(2012)022.
- [22] R. Keith Ellis and Sinisa Veseli. Strong radiative corrections to W b anti-b production in p anti-p collisions. *Phys.Rev.*, D60:011501, 1999. doi: 10.1103/PhysRevD.60.011501.
- [23] Michelangelo L. Mangano, Mauro Moretti, and Roberto Pittau. Multijet matrix elements and shower evolution in hadronic collisions:  $Wb\bar{b} + n$  jets as a case study. *Nucl.Phys.*, B632:343–362, 2002. doi: 10.1016/S0550-3213(02)00249-3.

- [24] F. Febres Cordero, L. Reina, and D. Wackerlo. NLO QCD corrections to W boson production with a massive b-quark jet pair at the Tevatron p anti-p collider. *Phys.Rev.*, D74:034007, 2006. doi: 10.1103/PhysRevD.74.034007.
- [25] Seth Quackenbush, Edmond L. Berger, C.B. Jackson, and Gabe Shaughnessy. LHC Sensitivity to  $Wb\bar{b}$  Production via Double Parton Scattering. 2011.
- [26] Jonathan R. Gaunt and W. James Stirling. Double Parton Distributions Incorporating Perturbative QCD Evolution and Momentum and Quark Number Sum Rules. *JHEP*, 1003:005, 2010. doi: 10.1007/JHEP03(2010)005.
- [27] Sunil Bansal, Paolo Bartalini, Boris Blok, Diego Ciangottini, Markus Diehl, et al. Progress in Double Parton Scattering Studies. 2014.
- [28] Serguei Chatrchyan et al. Study of double parton scattering using W + 2-jet events in proton-proton collisions at  $\sqrt{s} = 7$  TeV. *JHEP*, 1403:032, 2014. doi: 10.1007/JHEP03(2014)032.
- [29] Serguei Chatrchyan et al. Measurement of the production cross section for a W boson and two b jets in pp collisions at  $\sqrt{s}=7$  TeV. *Phys.Lett.*, B735:204–225, 2014. doi: 10.1016/j.physletb.2014.06.041.
- [30] T. Aaltonen et al. First Measurement of the b-jet Cross Section in Events with a W Boson in p anti-p Collisions at  $s^{**}(1/2) = 1.96$ -TeV. *Phys.Rev.Lett.*, 104:131801, 2010. doi: 10.1103/PhysRevLett.104.131801.
- [31] V.M. Abazov et al. Measurement of the  $p\bar{p} \rightarrow W + b + X$  production cross section at  $\sqrt{s} = 1.96$  TeV. *Phys.Lett.*, B718:1314–1320, 2013. doi: 10.1016/j.physletb.2012.12.044.
- [32] Georges Aad et al. Measurement of the cross-section for W boson production in association with b-jets in pp collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector. *JHEP*, 1306:084, 2013. doi: 10.1007/JHEP06(2013)084.

- [33] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008. doi: 10.1088/1748-0221/3/08/S08004.
- [34] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008. doi: 10.1088/1748-0221/3/08/S08003.
- [35] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008. doi: 10.1088/1748-0221/3/08/S08002.
- [36] Jr. Alves, A. Augusto et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008. doi: 10.1088/1748-0221/3/08/S08005.
- [37] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008. doi: 10.1088/1748-0221/3/08/S08001.
- [38] Oliver Sim Brüning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC Design Report*. CERN, Geneva, 2004.
- [39] Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update. Technical Report CMS-PAS-SMP-12-008, CERN, Geneva, 2012. URL <https://cds.cern.ch/record/1434360>.
- [40] CMS Collaboration. CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update. Technical Report CMS-PAS-LUM-13-001, 2013.
- [41] CMS Collaboration. Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.
- [42] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov. RESEARCH NOTE FROM COLLABORATION: Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC. *Journal of Physics G Nuclear Physics*, 31:9, September 2005. doi: 10.1088/0954-3899/31/9/N01.
- [43] CMS Collaboration. Electron reconstruction and identification at  $\sqrt{s} = 7$  TeV. CMS Physics Analysis Summary CMS-PAS-EGM-10-004, 2010.

- [44] R. Frühwirth. Application of kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2–3):444 – 450, 1987. ISSN 0168-9002. doi: [http://dx.doi.org/10.1016/0168-9002\(87\)90887-4](http://dx.doi.org/10.1016/0168-9002(87)90887-4). URL <http://www.sciencedirect.com/science/article/pii/0168900287908874>.
- [45] The CMS Collaboration. Performance of CMS muon reconstruction in pp collision events at  $\sqrt{s} = 7$  TeV. *Journal of Instrumentation*, 7:2P, October 2012. doi: 10.1088/1748-0221/7/10/P10002.
- [46] Gavin P. Salam. Towards Jetography. *Eur.Phys.J.*, C67:637–686, 2010. doi: 10.1140/epjc/s10052-010-1314-6.
- [47] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 0804:063, 2008. doi: 10.1088/1126-6708/2008/04/063.
- [48] Serguei Chatrchyan et al. Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS. *JINST*, 6:P11002, 2011. doi: 10.1088/1748-0221/6/11/P11002.
- [49] CMS Collaboration. 8 TeV Jet Energy Corrections and Uncertainties based on 19.8  $\text{fb}^{-1}$  of data in CMS. Technical report, Oct 2013. URL <http://cds.cern.ch/record/1627305>.
- [50] CMS Collaboration. Pileup subtraction using jet areas. *Physics Letters B*, 659(1–2):119 – 126, 2008. ISSN 0370-2693. doi: <http://dx.doi.org/10.1016/j.physletb.2007.09.077>. URL <http://www.sciencedirect.com/science/article/pii/S0370269307011094>.
- [51] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur.Phys.J.*, C72:1896, 2012. doi: 10.1140/epjc/s10052-012-1896-2.
- [52] Serguei Chatrchyan et al. Identification of b-quark jets with the CMS experiment. *JINST*, 8:P04013, 2013. doi: 10.1088/1748-0221/8/04/P04013.

- [53] CMS Collaboration. Performance of b tagging at  $\sqrt{s}=8$  TeV in multijet, ttbar and boosted topology events. CMS Physics Analysis Summary CMS-PAS-BTV-13-001, 2013.
- [54] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, 2009. Geneva, Apr 2009. URL <http://cds.cern.ch/record/1194487>.
- [55] Serguei Chatrchyan et al. Missing transverse energy performance of the CMS detector. *JINST*, 6:P09001, 2011. doi: 10.1088/1748-0221/6/09/P09001.
- [56] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 0605:026, 2006. doi: 10.1088/1126-6708/2006/05/026.
- [57] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput.Phys.Commun.*, 178:852–867, 2008. doi: 10.1016/j.cpc.2008.01.036.
- [58] Rick Field. Early LHC Underlying Event Data - Findings and Surprises. 2010.
- [59] Serguei Chatrchyan et al. Jet and underlying event properties as a function of charged-particle multiplicity in proton–proton collisions at  $\sqrt{s} = 7$  TeV. *Eur.Phys.J.*, C73(12):2674, 2013. doi: 10.1140/epjc/s10052-013-2674-5.
- [60] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5 : Going Beyond. *JHEP*, 1106:128, 2011. doi: 10.1007/JHEP06(2011)128.
- [61] John M. Campbell and R.K. Ellis. MCFM for the Tevatron and the LHC. *Nucl.Phys.Proc.Suppl.*, 205-206:10–15, 2010. doi: 10.1016/j.nuclphysbps.2010.08.011.
- [62] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 1407:079, 2014. doi: 10.1007/JHEP07(2014)079.

- [63] Carlo Oleari. The POWHEG-BOX. *Nucl.Phys.Proc.Suppl.*, 205-206:36–41, 2010. doi: 10.1016/j.nuclphysbps.2010.08.016.
- [64] S. Jadach, Z. Was, R. Decker, and Johann H. Kuhn. The tau decay library TAUOLA: Version 2.4. *Comput.Phys.Commun.*, 76:361–380, 1993. doi: 10.1016/0010-4655(93)90061-G.
- [65] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl.Instrum.Meth.*, A506: 250–303, 2003. doi: 10.1016/S0168-9002(03)01368-8.
- [66] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl.Instrum.Meth.*, A389:81–86, 1997. doi: 10.1016/S0168-9002(97)00048-X.
- [67] CMS Collaboration. Combination of ATLAS and CMS top quark pair cross section measurements in the emu final state using proton-proton collisions at 8 TeV. CMS Physics Analysis Summary CMS-PAS-TOP-14-016, 2014.
- [68] CMS Collaboration. Pileup Jet Identification. CMS Physics Analysis Summary CMS-PAS-JME-13-005, 2013.
- [69] Procedure for the LHC Higgs boson search combination in summer 2011. Technical Report ATL-PHYS-PUB-2011-011, CERN, Geneva, Aug 2011. URL <https://cds.cern.ch/record/1375842>.
- [70] B Henrich and R Kaufmann. Lorentz-angle in irradiated silicon. *Nucl. Instrum. Methods Phys. Res., A*, 477(1-3):304–307, 2002.