

**A measurement of  $Wb\bar{b}$  production and a search for monophoton signals of dark matter  
using the CMS detector at the CERN LHC**

by

Thomas Mastrianni Perry

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

Doctor of Philosophy

(Physics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2016

Date of final oral examination: 3 August 2016

The dissertation is approved by the following members of the Final Oral Committee:

Wesley Smith (Advisor), Bjorn Wiik Professor, Physics

Sridhara Dasu, Professor, Physics

Matthew Herndon, Professor, Physics

Yang Bai, Professor, Physics

David C. Schwartz, Professor, Chemistry and Genetics

## ACKNOWLEDGMENTS

It is easy to work hard when your efforts are rewarded. I think this is one of the main appeals of working with your hands, that effort is correlated with progress and that tangible results come not just at the end, but along the way. This is why encouragement can be so much more effective than simply pointing out flaws. That even the illusion of progress can have its merits. But encouragement and progress are nothing without integrity, and the unrelenting integrity of Sascha Savin I take as inspiration and encouragement. He absolutely knows all of the short cuts, and systematically rejects them in his own life. For this he is a role model of mine, and I am truly grateful to have worked under him.

Maria Cepeda I will say less about. But she is my sister and one of the most genuine people I know. I consider myself extremely lucky not only to have worked under her, but to have her in my life. She is a beacon of light and sometimes just knowing that she's there is enough.

There are few people that I respect more as physicists than my advisor Wesley Smith and effective coadvisor Sridhara Dasu. In a very literal sense, I would not at all be where I am today without them, and as I played my small role in helping them deliver multiple superb trigger systems, I was able to see firsthand how they embody the lesson that nature can not be fooled. Even if your grant says otherwise.

I would like to thank my collaborators, particularly Bhawna Gomber and James Buchanan in the monophoton analysis and Andrea Schizzi and Jelena Luetic with  $Wb\bar{b}$ . And Aaron Levine is also someone I would like to mention in particular in a nonacademic sense. It has been five years now that we've been together and I can't imagine what this whole journey would have been like without him.

Of course I love and thank my family for their continual support. None of them have ever wavered from being completely there for me, even when I disappear for months into the wilds of Geneva.

And last here is Anita has been far more patient with me than I've given her reason to be.

# CONTENTS

	<b>Page</b>
<b>List of Figures</b>	<b>v</b>
<b>List of Tables</b>	<b>vi</b>
<b>Abstract</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The Standard Model . . . . .	2
1.1.1 <i>Standard Model Particles</i> . . . . .	2
1.1.2 <i>Electroweak symmetry breaking</i> . . . . .	3
1.1.3 <i>Cross sections and decay rates</i> . . . . .	6
1.1.4 <i>QCD and Proton Structure</i> . . . . .	9
1.2 Previous $Wb\bar{b}$ Measurements . . . . .	11
1.3 Dark Matter . . . . .	12
1.3.1 <i>Experimental motivations</i> . . . . .	12
1.3.2 <i>Simplified theoretical models</i> . . . . .	13
1.3.3 <i>Previous DM Searches</i> . . . . .	14
<b>2 Phenomenology of Processes</b>	<b>15</b>
2.1 The Standard Model process $pp \rightarrow Wb\bar{b} \rightarrow \ell\nu b\bar{b}$ . . . . .	15
2.1.1 $pp \rightarrow W$ . . . . .	16
2.1.2 $W \rightarrow \ell\nu$ . . . . .	16
2.1.3 $g \rightarrow b\bar{b}$ . . . . .	17
2.2 The Standard Model process $pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma$ . . . . .	18
2.2.1 $pp \rightarrow Z/\gamma$ . . . . .	19
2.2.2 $Z \rightarrow \nu\bar{\nu}$ . . . . .	19
2.2.3 $pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma$ . . . . .	19

2.3	Beyond the Standard Model: $pp \rightarrow \gamma + \text{DM}$ . . . . .	21
<b>3</b>	<b>The LHC and CMS</b>	<b>23</b>
3.1	The Large Hadron Collider . . . . .	23
3.1.1	<i>LHC pre-acceleration</i> . . . . .	24
3.1.2	<i>LHC acceleration</i> . . . . .	24
3.2	The Compact Muon Solenoid Detector . . . . .	27
3.2.1	<i>Geometry</i> . . . . .	28
3.2.2	<i>Magnet</i> . . . . .	30
3.2.3	<i>Tracking System</i> . . . . .	30
3.2.4	<i>Electronic Calorimeter</i> . . . . .	31
3.2.5	<i>Hadronic Calorimeter</i> . . . . .	33
3.2.6	<i>Muon System</i> . . . . .	35
3.2.7	<i>Trigger and Data Acquisition</i> . . . . .	37
3.2.7.1	Level-1 Trigger . . . . .	37
3.2.7.2	High Level Trigger . . . . .	41
<b>4</b>	<b>Simulation and Reconstruction</b>	<b>42</b>
4.1	Simulation of Events . . . . .	42
4.1.1	<i>Monte Carlo Event Generation</i> . . . . .	42
4.1.2	<i>Monte Carlo Generators</i> . . . . .	44
4.1.2.1	Matrix Element Calculators . . . . .	45
4.1.2.2	Showering and NLO Corrections . . . . .	46
4.1.3	<i>Detector Simulation</i> . . . . .	47
4.2	Reconstruction of Events . . . . .	47
4.2.1	<i>Track and Primary Vertex Reconstruction</i> . . . . .	48
4.2.2	<i>Electron ID and Reconstruction</i> . . . . .	49
4.2.3	<i>Muon ID and Reconstruction</i> . . . . .	50
4.2.4	<i>Lepton Isolation</i> . . . . .	51
4.2.5	<i>Photon ID and Reconstruction</i> . . . . .	51
4.2.6	<i>Missing Transverse Energy</i> . . . . .	52

4.2.7	<i>Jet ID and Secondary Vertices</i>	52
<b>5</b>	<b><i>Wb<math>\bar{b}</math></i> Cross Section Measurement</b>	<b>54</b>
5.1	Event Selection	54
5.2	Simulated samples	55
5.3	Background Estimation	57
5.3.1	<i>QCD</i>	57
5.3.2	<i>W+jets: light and charm component</i>	60
5.3.3	<i>Single top backgrounds</i>	61
5.3.4	<i>Z<math>\ell\bar{\ell}</math> backgrounds</i>	62
5.4	Analysis Strategy	64
5.5	Systematic uncertainties	68
5.6	Signal Extraction	69
5.7	Cross Section and Comparisons	74
<b>6</b>	<b>Monophoton Analysis</b>	<b>79</b>
6.1	Event Selection	79
6.2	Estimation of Background Contributions	82
6.2.1	<i>Reweighting</i>	83
6.2.2	<i>V<math>\gamma</math> Estimates</i>	84
6.2.3	<i>Electron Mis-ID</i>	84
6.2.4	<i>Non-collision Backgrounds</i>	85
6.2.5	<i>Minor SM Processes</i>	86
6.3	Results	86
6.3.1	<i>pp <math>\rightarrow</math> Z<math>\gamma \rightarrow \nu\bar{\nu}\gamma</math> Cross Section Measurement</i>	86
6.3.2	<i>Limits on Dark Matter</i>	89
<b>7</b>	<b>Conclusions and Future Prospects</b>	<b>92</b>
	<b>Bibliography</b>	<b>95</b>

# LIST OF FIGURES

1.1	One-loop corrections to vertices and propagator . . . . .	6
2.1	Feynman diagrams for $pp \rightarrow Wb\bar{b} \rightarrow \ell\nu b\bar{b}$ . . . . .	15
2.2	Feynman diagrams for $pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma$ . . . . .	18
2.3	Feynman diagrams for dark matter monophoton production . . . . .	21
3.1	The LCH accelerator complex at CERN . . . . .	25
3.2	The LHC dipole . . . . .	26
3.3	The CMS Detector . . . . .	29
3.4	The CMS Tracking System . . . . .	32
3.5	The CMS Electromagnetic Calorimeter . . . . .	34
3.6	The CMS Hadronic Calorimeter . . . . .	36
3.7	The CMS Muon System . . . . .	38
3.8	The 2012 Level-1 Trigger . . . . .	39
3.9	The 2015 Upgrade Calorimeter Trigger . . . . .	40
4.1	Example Parton Distribution Function from CTEQ6M . . . . .	44
5.1	QCD shape for $Wb\bar{b}$ analysis . . . . .	59
5.2	$Wjj$ control region for the $Wb\bar{b}$ measurement . . . . .	60
5.3	Diagram for single top $t$ -channel control region . . . . .	61
5.4	Single-top control region for the $Wb\bar{b}$ measurement . . . . .	62
5.5	$Zb\bar{b}$ control region for the $Wb\bar{b}$ analysis . . . . .	63
5.6	$Zjj$ control region for the $Wb\bar{b}$ measurement . . . . .	63
5.7	Three-step fitting procedure for $Wb\bar{b}$ measurement . . . . .	72
5.8	More variables in the $Wb\bar{b}$ signal region . . . . .	73
5.9	Cross section comparison for $Wb\bar{b}$ and generators . . . . .	78
6.1	Signal region distributions in the $pp \rightarrow \gamma + \text{invisible}$ analysis . . . . .	87

6.2	Exclusion plots in $m_\chi - m_M$ plane . . . . .	90
6.3	Expected lower limit for EFT cutoff parameter $\Lambda$ . . . . .	91

## LIST OF TABLES

1.1	Fermion couplings . . . . .	2
1.2	Illustrations of SM couplings . . . . .	5
1.3	Fundamental particle decay channels and rates . . . . .	9
5.1	$Wb\bar{b}$ Cutflow . . . . .	55
5.2	Systematic uncertainties in $Wb\bar{b}$ . . . . .	67
5.3	Measured breakdown of samples in $Wb\bar{b}$ . . . . .	71
5.4	Measured cross sections in the muon, electron, and combined lepton channels. . . . .	75
6.1	Monophoton Cutflow . . . . .	82
6.2	Estimated yields for $pp \rightarrow \gamma + \text{invisible}$ . . . . .	87
6.3	Systematic uncertainties in $pp \rightarrow \gamma + \text{invisible}$ . . . . .	88
6.4	Upper limits on DM cross sections for $m_M = 10$ TeV . . . . .	89

# ABSTRACT

This thesis presents the results of two analyses of data from proton-proton collisions collected by the CMS Collaboration at the CERN LHC during operation in the years 2012 and 2015. The 2012 data is analyzed in the context of the  $pp \rightarrow Wb\bar{b} \rightarrow \ell\nu b\bar{b}$  process where  $\ell = \mu$  or  $e$  at  $\sqrt{s} = 8$  TeV and compared with simulations within the context of the Standard Model (SM). A cross section of  $\sigma(pp \rightarrow Wb\bar{b} \rightarrow \ell\nu b\bar{b}) = 0.64 \pm 0.03(\text{stat}) \pm 0.10(\text{syst}) \pm 0.06(\text{theo}) \pm 0.02(\text{lumi})$  pb is measured, which is approximately one standard deviation higher than predictions. The data from 2015 is analyzed in the monophoton channel at  $\sqrt{s} = 13$  TeV both as a measurement of the cross section for  $pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma$  and as a search for dark matter (DM). The SM cross section is measured as  $\sigma(pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma) = 66.5 \pm 13.6(\text{stat}) \pm 14.3(\text{syst}) \pm 2.2(\text{lumi})$  fb which agrees with the NNLO SM prediction of  $65.5 \pm 3.3$  fb. Limits are therefore set on the mass of vector or axial-vector mediator models of DM production up to 600 GeV for particle DM mass up to 10 GeV. An effective field theory model is also considered and values of the suppression scale of the coupling between DM and photons are excluded up to 540 GeV.



# Chapter 1

## INTRODUCTION

One of the goals of modern society is to experimentally investigate particles and interactions, and to achieve this goal, the governments from nearly 100 different countries, states and territories are presently funding over 13,000 scientists, engineers and technicians to build, operate, maintain and analyze data from the Large Hadron Collider (LHC) at the European Center for Nuclear Research (CERN). This thesis presents analyses of data taken with the Compact Muon Solenoid (CMS) detector using proton-proton ( $pp$ ) collisions provided by the LHC during its operation in 2012 and 2015.

Using the 2012 data, a measurement of the production rate for the standard model (SM) process  $pp \rightarrow Wb\bar{b} \rightarrow \ell\nu b\bar{b}$  is performed. In this SM process the colliding protons ( $p$ ) create a  $W$  boson that mediates charged weak interaction, accompanied by a  $b$  quark-anti-quark pair. The  $W$  boson promptly decays to an electron or a muon ( $\ell$ ), accompanied by its neutrino ( $\nu_\ell$ ) partner that disappears without leaving a trace in the detector.

The other analysis is a Dark Matter (DM) search looking for a monophoton signature. Using the 2015 data, the monophoton signature is examined in the context of the SM process  $pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma$  and as a search for evidence of DM production. In the SM process, colliding protons create a  $Z$  boson which mediates neutral weak

interactions much like the photon ( $\gamma$ ), which mediates the electromagnetic interaction. The  $Z$  boson decays to a neutrino pair which disappears without a trace in the detector, producing a monophoton signature. Evidence for excessive production of such monophoton events constitutes evidence for DM, which is also expected to disappear without leaving a trace in the detector.

## 1.1 The Standard Model

### 1.1.1 Standard Model Particles

The SM Lagrangian uses the Yang-Mills construction on the group  $SU(3) \times SU(2) \times U(1)$ . Strong interactions are described by the symmetry  $SU(3)$  (color) which requires eight gauge bosons, known as gluons. Electroweak interactions are described by  $SU(2) \times U(1)$  are moderated by the massive  $W^\pm$  and  $Z$  bosons and the massless photon ( $\gamma$ ).

**Table 1.1:** *Fermions can exist in doublet and singlet configurations which are listed below along with the gauge fields the configuration interacts with, denoted by the symmetry the field respects.*

Fermion type	Interactions		
	$SU(3)$	$SU(2)$	$U(1)$
All quarks (doublet)	Yes	Yes	Yes
All leptons (doublet)	No	Yes	Yes
$u$ -type quarks (singlet)	Yes	No	Yes
$d$ -type quarks (singlet)	Yes	No	Yes
Charged leptons (singlet)	No	No	Yes

The matter and antimatter components are spin- $1/2$  fermions, which come in two families, known as leptons and quarks. Quarks and leptons come in three families, known as generations for the quarks and flavors for the leptons. The leptons are colorless and therefore do not couple to gluons, but the quarks do and additionally come in three color varieties for each generation. The singlet configurations contain only electrically charged fermions.

All quarks and leptons can exist in  $SU(2)$  doublet configurations, and the different  $I_3$  values further break quarks and leptons into types. For quarks, there are  $u$ -type  $(u, c, t) = (\text{up, charm, top})$  which have charge  $+2/3$ , and there are  $d$ -type  $(d, s, b) = (\text{down, strange, bottom})$  which have charge  $-1/3$ . Leptons are either charged  $(e, \mu, \tau) = (\text{electron, muon, tauon})$  or neutral  $(\nu_e, \nu_\mu, \nu_\tau) = (\text{electron-, mu-, tau-neutrino})$ , and all charged fermions are arranged such that mass increases with successive generations within a given type. In units where  $\hbar = c = 1$ , the top ( $m_t = 173 \text{ GeV}$ ) and bottom ( $m_b = 4 \text{ GeV}$ ) are the heaviest quarks of their respective types, with the bottom weighing three orders of magnitude greater than the lightest quark, up ( $m_u = 2 \text{ MeV}$ ). The mass separation for the charged leptons ( $m_\tau = 1.7 \text{ GeV}, m_e = 0.5 \text{ MeV}$ ) also spans multiple orders of magnitude, but while the observations of neutrino oscillations indicate that neutrinos have mass, only upper limits on the values they may have have been set on the order of MeV.

### 1.1.2 Electroweak symmetry breaking

Fermions and gauge bosons acquire mass through the Brout-Englert-Higgs mechanism in which a scalar field is introduced to the SM Lagrangian in the doublet configuration.

This generates mass for the Higgs boson ( $m_H$ ) as well as for the  $W^\pm$  and  $Z$  bosons, while leaving the photon massless,

$$M_Z = \frac{v}{2}\sqrt{g_1^2 + g_2^2}, \quad M_W = \cos \theta_W M_Z \quad , \quad (1.1)$$

where the Weinberg angle,  $\theta_W = g_1/g_2$ , is the ratio between the  $U(1)$  and  $SU(2)$  coupling strengths, and  $v \propto m_H$  is vacuum expectation value of the Higgs field. The  $\pm$  in  $W^\pm$  aligns with the electric charge of the  $W$  boson, and the  $Z$  and  $\gamma$  are both electrically neutral and orthogonal.

The vector boson couplings to the fermions are then expressed as three types of currents. Electromagnetic interactions involve couplings between the photon and charged particles, and any particle which interacts with the photon can also interact with the  $Z$  boson. The neutral current interactions have additional couplings to the lepton and quark doublets and charged current interactions are between the doublets and the  $W$  bosons. Couplings to each of the bosons are scaled by a coupling factor which decides the relative strengths of the interactions. The coupling factors are related to the strengths of the unbroken  $U(1)$  and  $SU(2)$  couplings as

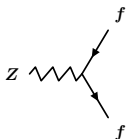
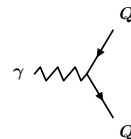
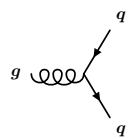
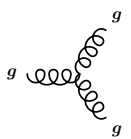
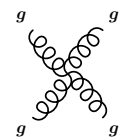
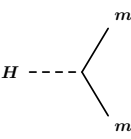
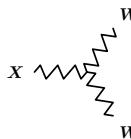
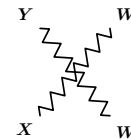
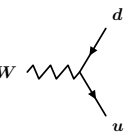
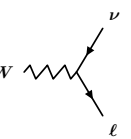
$$e = \frac{g_1 g_2}{\sqrt{g_1^2 + g_2^2}}, \quad g_Z = \frac{e}{\cos \theta_W \sin \theta_W}, \quad g_W = \frac{e}{\sqrt{2} \sin \theta_W} \quad (1.2)$$

In addition to generating mass terms for the bosons, the Higgs field gives rise to fermion mass via Yukawa couplings between the Higgs and fermion doublets. It is because the mass of the Higgs boson is tied to the masses of the gauge bosons and charged fermions that the discovery of the Higgs boson in 2012 was of such

importance. The result was announced just a few months after I moved to Geneva to do research at CERN, and it provided the first measurement of a parameter in the SM that had been previously unknown,  $m_H = 125$  GeV.

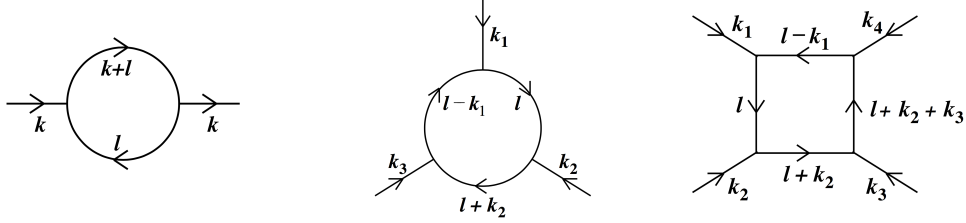
To describe the scattering of fermions and gauge bosons in the SM, diagrams with the appropriate initial and final state particles are constructed from the vertices illustrated in Table 1.2. External lines correspond to real particles which are observed in the final state and are on-shell. Internal lines correspond to virtual particles that not observed in the final state and therefore may or may not be on-shell. A diagram is tree-level or leading order (LO) if it uses the smallest number of vertices possible to depict an interaction having the correct initial and final states.

**Table 1.2:** The couplings for the SM particles interactions.

				
$f$ is any fermion	$Q$ is electrically charged	$q$ is any quark	$g$ is any gluon	
				
$m$ is any massive particle	$X$ is $Z$ or $\gamma$	$X$ and $Y$ are EW bosons such that charge is conserved	$u$ is an up-type quark and $d$ is a down-type quark	$\ell$ is a lepton and $\nu$ is the corresponding neutrino

Renormalization is necessary to account for corrections to the tree-level propagators and vertices which arise from contributions of virtual particles connecting to form

**Figure 1.1:** Renormalization takes place via a modification of the coupling parameters  $g_1, g_2, g_3$  to account for the contributions that loops of virtual particles make on the propagators and vertices. Below are diagrams keeping track of the flow of momentum for one-loop corrections to the propagator and three and four point vertices.



closed internal loops. The lowest order diagrams with one internal loop, are illustrated in Figure 1.1. These are termed next-to-LO (NLO), and some present calculations can be performed at next-to-NLO (NNLO). Renormalization is accomplished by introducing an energy scale, and assuming that the couplings are small compared this scale. The coupling constants are therefore functions of energy and are quoted at a particular renormalization scale,  $\mu_R$ .

### 1.1.3 Cross sections and decay rates

Quantum mechanics dictates that only predictions of probability are possible, and the final probability of observing a particular interaction is dependent on many variables, including the energies, types and angular momenta of the incoming and outgoing particles as well as the masses of the propagators and the orientation and efficiency of the detector.

A quantity typically measured is therefore the interaction cross section,  $\sigma$ , and for the scattering of two incoming particles going to  $n'$  particles,  $2 \rightarrow n'$ , in the CM

frame, the differential is

$$d\sigma = \frac{1}{4|\mathbf{k}_1|_{\text{CM}}} |\mathcal{T}|^2 d\text{LIPS}_{n'}(k_1 + k_2) \quad (1.3)$$

where the scattering matrix element,  $\mathcal{T}$ , is defined as

$$\langle f|i\rangle = (2\pi)^4 \delta^4 \left( \sum k_{\text{in}} - \sum k_{\text{out}} \right) i\mathcal{T} \quad (1.4)$$

and the Lorentz-invariant measure of the phase space for the  $n'$  outgoing particles is

$$d\text{LIPS}_{n'}(k) = (2\pi)^4 \delta^4 \left( k - \sum_{j=1}^{n'} k'_j \right) \prod_{j=1}^{n'} dq'_j \quad (1.5)$$

with the Lorentz-invariant differential  $dq$ .

The cross section is used to calculate the rate at which a process occurs, but is not the only relevant factor in determining the overall production rate. The production rate of a given final state is also dependent on the incoming rate of possible interactions and is known as luminosity,  $\mathcal{L}$ . Luminosity has the units of inverse area per unit time and the total number of events produced is therefore proportional to  $\int \mathcal{L} dt$ . In any real detector, final state particles are collected only within a finite solid angle and the number of particles scattered into a given solid angle,  $\Omega$ , is given by

$$\frac{dN}{d\Omega} = \mathcal{L} \frac{d\sigma}{d\Omega} \quad (1.6)$$

It is also possible for particles to decay as  $1 \rightarrow n'$ . Massive particles decay to lighter ones in both the fermion and boson sectors, with all massive bosons able

to spontaneously decay via the diagrams in Figure 1.2. Of the charged fermions, only the first generation is stable for each type, and neutrinos are not known to spontaneously decay, but oscillate between flavors while propagating in free space. Like the differential cross section, the differential decay rate is a function of the scattering amplitude and has integration measure  $d\text{LIPS}$ ,

$$d\Gamma = \frac{1}{2E} |\mathcal{T}|^2 d\text{LIPS}_{n'}(k) \quad . \quad (1.7)$$

The differential decay rate is inversely proportional to the energy of the particle,  $E = \sqrt{m^2 + p^2}$ . This means that comparatively heavy particles will decay faster than comparatively light ones and that energetic particles will appear to live longer for a stationary observer due to relativistic time dilation effects. The total decay rate of a given particle is found by summing the decay rates from each of the contributing processes, and the primary decay channels and rates for the fundamental particles are given in Table 1.3.

At CMS, the heaviest quarks and the heaviest lepton decay before reaching the detector volume, but by tracing the trajectories of the decay products backwards, the location of the decay can be determined. The heaviest fundamental particles which can be observed to decay in a location that is spatially separated from the initial  $pp$  collision are  $b$  quarks. Their heavy mass means that they couple strongly with the Higgs boson which still has many properties that are under investigation, making  $b$  quarks an interesting object of study. The  $W$ ,  $Z$  and Higgs bosons are each so massive that they decay before reaching the innermost layers of the detector and are often identified by their decay products pointing back to the same common vertex as



**Table 1.3:** Below are listed the decay channels and rates for unstable fundamental particles. At CMS, with the detection apparatus located a finite distance away from the interaction vertex, particles such as the  $W$ ,  $Z$  and Higgs bosons, as well as the  $t$  and  $\tau$ , decay too quickly to be resolved as having decayed in a different location than the initial  $pp$  collision. Particles such as the  $b$  and  $c$ , can be observed as having travelled a finite distance away from the initial  $pp$  collision before decaying. The  $s$  quark can decay on time scales of  $10^{-23} - 10^{-10}$  s, depending on the hadron it is bound in and the mode of decay, and the muon typically decays after having passed through the entire detector.

Particle	Primary decay mode	Total rest-frame $d\Gamma$	Vertex Resolved
$W$	$W \rightarrow \ell\nu$	2.1 GeV	No
$Z$	$Z \rightarrow f\bar{f}$	2.5 GeV	No
$H$	$H \rightarrow b\bar{b}$	6.1 MeV	No
$\tau$	$\tau \rightarrow W\nu_\tau$	2.2 keV	No
$\mu$	$\mu \rightarrow W\nu_\mu$	300 peV	Yes
$t$	$t \rightarrow W^+b$	1.4 GeV	No
$b$	$b \rightarrow W^-c$	500 eV	Yes
$c$	$c \rightarrow W^+s$	2 meV	Yes
$s$	$s \rightarrow W^-u$	$\mu\text{eV-MeV}$	Yes

the initial  $pp$  collision.

### 1.1.4 QCD and Proton Structure

Feynman diagrams are used to describe interactions between fundamental particles, but at the LHC, collisions take place between protons, which are composite.

One feature of the  $SU(3)$  symmetry of the strong force is that gluons carry one unit of color and one unit of anticolor while the quarks carry one unit of color charge. This is what allows gluons to interact with each other as well as with quarks. That

quark confinement is necessitated by the  $SU(3)$  structure has not been conclusively determined, but observationally, a free gluon or quark has never been observed. Instead, quarks appear as bound in colorless (singlet) combinations called hadrons which are further classified as mesons ( $q\bar{q}$ ) or as baryons ( $qqq$  or  $\bar{q}\bar{q}\bar{q}$ ), and are held together by gluons. Evidently, the binding energy of the quarks has a form such that after a distance of roughly  $10^{-15}$  meters, the energy stored in the gluon field is greater than the energy needed to create a quark-antiquark pair, bringing the pair into existence. This process of energetic quarks creating particles as they separate is called hadronization and is an important effect at the LHC.

Protons are a type of baryon and at low energy, may combine with a single electron to form a neutral hydrogen atom. At higher energies, the internal structure of the proton becomes more evident, and it contains three valence quarks,  $uud$ , which are constantly exchanging gluons. When probed at high enough energy, or equivalently, at short enough length scales, these gluons can also each split into a  $q\bar{q}$  pair which typically reannihilate with each other. With gluons inside the proton splitting into quarks and coupling with other gluons, this forms a ‘sea’ of quarks and gluons, and as protons are accelerated to energies of GeV or TeV as is the case at the LHC, the fraction of the momentum of the proton attributed to the gluons becomes higher than that attributed to the valence quarks.

A proton-proton collider was therefore a sensible choice for the LHC. The physics goals of the project are to measure quantities associated with a wide range SM processes and to continue the search for evidence of new physics. Quarks interact with all of the SM gauge bosons as well as with the Higgs boson and the proton

contains the lightest quarks of each type in addition to the gluons and sea. Colliding proton beams thus allow for the interactions between many different initial particle configurations to be explored, and with the exception of the neutrinos which interact only via the weak exchange of the  $Z$  boson and escape the detectors, all other fundamental SM particles have been directly observed at CERN.

## 1.2 Previous $Wb\bar{b}$ Measurements

The production of  $W$  bosons in association with  $b$  jets has been studied at a center-of-mass energy of 7 TeV using data samples with up to  $5 \text{ fb}^{-1}$  of integrated luminosity, by the ATLAS and CMS experiments [1, 2], as well as by the Collider Detector at Fermilab (CDF) and D0 Collaborations using  $p\bar{p}$  collisions provided by the Tevatron [3, 4] at Fermilab at  $\sqrt{s} = 1.96 \text{ TeV}$ . Neither of the experiments at the Tevatron directly examined a final state with more than one  $b$ -tagged jet, but both published results on the process  $p\bar{p} \rightarrow Wbj \rightarrow \ell\nu bj$  where  $j$  signifies any jets. Here, and throughout the remainder of this thesis, leptons,  $\ell$ , are taken to mean only electrons or muons. The CDF Collaboration observed a cross section that was a factor of two higher than the best NLO prediction available at the time, and the D0 Collaboration observed a cross section that was 20% – 40% higher than various NLO predictions. The ATLAS Collaboration also examined a final state having only one  $b$ -tagged jet in the process  $pp \rightarrow Wb \rightarrow \ell\nu b$ . They measured cross sections for final states with exactly one jet and exactly two jets, and their observed cross section in the one-jet channel was 70% higher than the NLO prediction. The agreement improved in the

two-jet channel, where the observed cross section was 30% higher than the NLO calculation predicts.

The first measurement of a final state having two identified  $b$  jets and the products of a  $W$  decay was made by the CMS Collaboration at  $\sqrt{s} = 7$  TeV in the process  $pp \rightarrow Wb\bar{b} \rightarrow \mu\nu b\bar{b}$  where exclusively two jets were allowed, both of which  $b$ -tagged. In this measurement, the cross section was found to be within 4% of the NNLO prediction, The analysis presented in this thesis extends the  $pp \rightarrow Wb\bar{b} \rightarrow \mu\nu b\bar{b}$  analysis to  $\sqrt{s} = 8$  TeV and additionally incorporates  $W$  decays  $W \rightarrow e\nu$ , with the goal of providing a measurement of this process at higher energy than ever before and comparisons with the latest simulations and predictions.

## 1.3 Dark Matter

### 1.3.1 Experimental motivations

Albert Einstein's theory of general relativity, GR, has many experimental predictions which run counter to human intuition. GR predicts that massive objects warp a four-dimensional spacetime and thus feel mutual attraction. This has as a consequence, the prediction that even massless objects such as photons will experience a net deviation in their path near a massive object as a result of gravity and this effect was famously verified by Arthur Eddington through the observation of stars around the sun during a full solar eclipse. More recently, the direct detection of gravitational waves by the LIGO Collaboration also aligns with the GR predictions of distorted spacetime around colliding black holes. The time distortion effects due to the varying strengths

of Earth's gravitational field on the surface and at the GPS satellites, provide precise tests of the quantitative predictions of GR. However, though these tests and others provide evidence that GR is an accurate theory of gravity, some basic predictions related to gravitational interactions do not agree with observations, motivating the concept for DM.

The first observational evidence for DM came from an analysis of the speeds of galaxies in the Coma cluster by Fritz Zwicky. The magnitude of the angular velocities of the galaxies was too great to be explained by the visible matter alone and DM is now believed to outweigh visible matter in a ratio of 5 : 1 throughout the universe and 10 : 1 throughout the Milky Way galaxy.

### 1.3.2 Simplified theoretical models

The defining features of DM are that it is massive and appears to interact on large scales only via the gravitational force. On the galactic and supergalactic scales, DM is distributed along similar structures as is visible matter, and it surrounds visible matter in extended halos.

Because DM has not yet been observed, the models of DM being considered in this thesis are simplified and based on minimal assumptions, the first being that DM is even capable of interacting with hadrons and is thus possible to produce at the LHC. While the visible sector of particles is diverse, the models used in this analysis consist of a single DM particle,  $\chi$ , which is assumed to be a fermion and may be different from  $\bar{\chi}$ .

One way DM could couple to the SM is via the addition of a  $U(1)$  symmetry that

gives rise to a vector gauge mediator,  $M$ . If some quarks are also charged under  $U(1)$ , then DM may be produced in the  $s$  channel as  $f\bar{f} \rightarrow M \rightarrow \chi\bar{\chi}$ . If  $M$  conserves parity in  $f\bar{f} \rightarrow M$ , it is said to have a vector coupling, and if it violates parity, it is termed axial-vector. In these models,  $M$  is assumed not to couple to leptons, but an effective field theory (EFT) model is also considered in this analysis which estimates a direct interaction between DM and photons. This coupling is mediated by a vertex  $\gamma\gamma\chi\bar{\chi}$ . and allows for DM production via the channel  $pp \rightarrow \gamma \rightarrow \gamma\chi\bar{\chi}$ .

### 1.3.3 Previous DM Searches

A search for particle DM in the monophoton channel [5] used  $19.6 \text{ fb}^{-1}$  of  $pp$  data collected in 2012 by the CMS Collaboration at  $\sqrt{s} = 8 \text{ TeV}$ , but no evidence was found and a new upper limit for the DM cross section was set. A similar search was performed by the ATLAS experiment [6] using  $3.2 \text{ fb}^{-1}$  of  $pp$  collision data at  $\sqrt{s} = 13 \text{ TeV}$  collected in 2015, but again no evidence for new physics was found. This imposed limits on the possible masses DM particles and the mediator mass as  $m_\chi > 150 \text{ GeV}$  or  $m_M > 710 \text{ GeV}$ . Similar searches have also been performed looking for particle DM produced in conjunction with ISR radiation of particles other than photons, in particular jets [7] and leptons [8]. Both of these searches explore a range of masses for  $\chi$  and  $M$ . In the EFT interpretation, both the monojet and monolepton searches give similar results, concluding that for  $m_\chi < 100 \text{ GeV}$  the suppression scale must obey  $\Lambda > 1 \text{ TeV}$ .

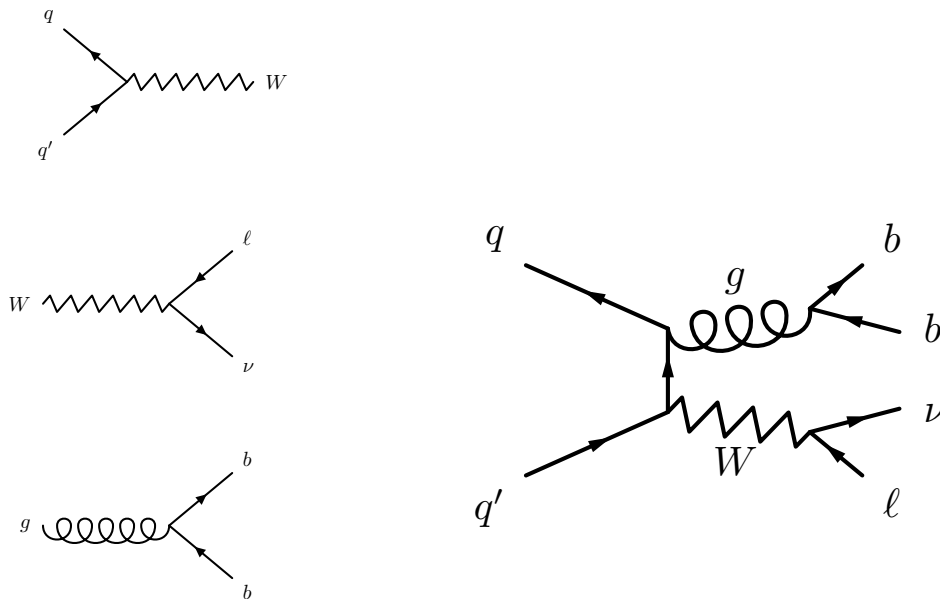
# Chapter 2

## PHENOMENOLOGY OF PROCESSES

### 2.1 The Standard Model process $pp \rightarrow Wb\bar{b} \rightarrow \ell\nu b\bar{b}$

The Feynman diagram depicting the full SM process  $pp \rightarrow Wb\bar{b} \rightarrow \ell\nu b\bar{b}$  along with the component vertices making it up are illustrated in Figure 2.1.

**Figure 2.1:** The Feynman diagram for the process  $pp \rightarrow Wb\bar{b} \rightarrow \ell\nu b\bar{b}$  is illustrated below, and is composed from the individual vertices illustrated on the left, each of which is described in Section 2.1.



### 2.1.1 $pp \rightarrow W$

The  $W$  boson couples to all charged fermions and can be created during the collision of a quark-antiquark pair with a relative charge difference of  $e$ . In the proton are quarks and the most prevalent valence quark is the  $u$ . Therefore in a  $pp$  collision, the channel by which most  $W$  bosons are produced is via the annihilation of a valence  $u$  quark from one proton with a  $\bar{d}$  from the sea of the other,  $u\bar{d} \rightarrow W^+$ . Quarks of higher generation can also be found inside the sea as the result of gluons splitting into  $q\bar{q}$  pairs, but all interactions are modified by a coefficient in the CKM matrix and higher generation mixing is thus suppressed. In this thesis, all modes of  $pp \rightarrow W^\pm$  production are considered.

### 2.1.2 $W \rightarrow \ell\nu$

Just as the  $W$  boson can be created by the collision  $q\bar{q}' \rightarrow W$ , it can also decay as  $W \rightarrow q\bar{q}'$ . This is known as hadronic  $W$  decay and can be a useful analysis channel for experimentalists, especially for decay products with energies approaching the TeV scale. Leptonic  $W$  decay,  $W \rightarrow \ell\nu$ , is also an important channel for experimentalists and is the one considered in this analysis. Because leptons constitute a negligible fraction of the sea, the detection of leptons at high energy after a  $pp$  collision is often a good indicator of the decay of a massive gauge boson,  $W \rightarrow \ell\nu$  or  $Z \rightarrow \ell\bar{\ell}$ .

The  $W$  boson is much heavier than any of the leptons and therefore decays with roughly equal probability to any of  $e\nu_e, \mu\nu_\mu, \tau\nu_\tau$ . From Table 1.3, taus created at CMS subsequently decay before reaching the detector, so for this analysis, the decay channel of the  $W$  investigated is  $W \rightarrow \ell\nu$  where  $\ell \in e, \mu$ .



To reconstruct muons from the decay products, the transverse mass,  $m_T$  variable is often used. It is defined by  $m_T^2 = m^2 + p_x^2 + p_y^2$  where  $p_i$  is the component of the momentum along the  $i$  axis, and in the case of a massive particle decaying to two massless particles, can be rewritten as

$$m_T^2 = 2p_{T,1}p_{T,2}(1 - \cos \phi) \quad (2.1)$$

where  $\phi$  is the angle between the particles and  $p_{T,j}$  is the component of the momentum of the particles in the transverse plane.

In the decay of a  $W$  boson, a neutrino is produced, but can not be detected. The CMS detector is designed to capture the energy from all of the other particles produced in a collision, so neutrinos are accounted for as  $E_T^{\text{miss}}$ . The variable  $E_T^{\text{miss}}$  is the transverse component of the negative vector sum of all of the energy identified as having come from a particular interaction vertex, known as the primary vertex. Therefore the transverse mass of the  $W$  boson is

$$\left(m_T^W\right)^2 = 2p_T^\ell E_T^{\text{miss}}(1 - \cos \phi) \quad (2.2)$$

where  $\phi$  is the angle between the lepton and  $E_T^{\text{miss}}$ .

### 2.1.3 $g \rightarrow b\bar{b}$

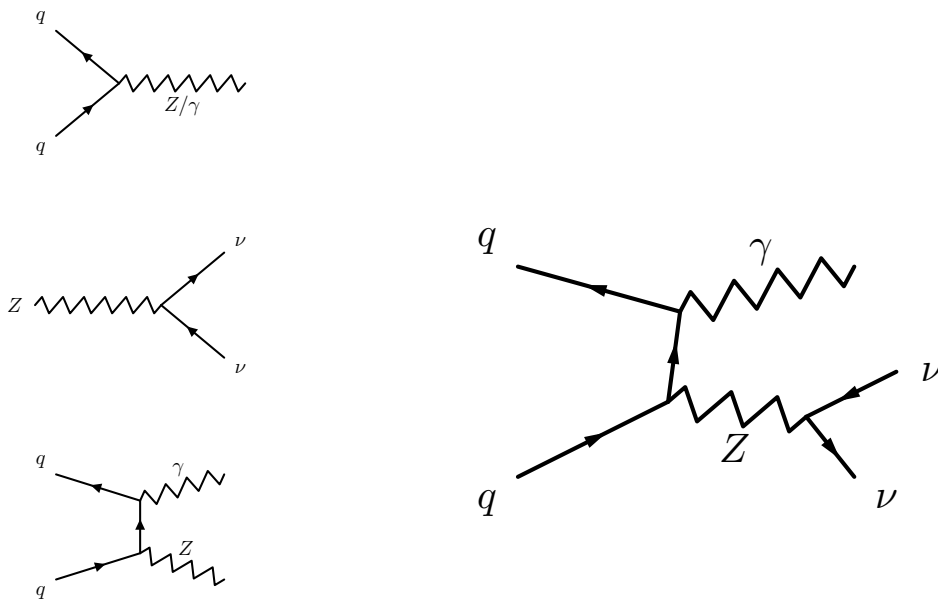
Because quarks couple strongly to gluons and  $q\bar{q}' \rightarrow W$  has been shown to be an important production channel in  $pp$  collisions, it is possible for one of the initial state quarks to radiate a gluon. This is called initial state radiation, ISR, and if the gluon

is produced with enough energy, it is capable of splitting to a quark-antiquark pair. In particular, a  $g \rightarrow b\bar{b}$  vertex can be added to either of the incoming quarks to form  $pp \rightarrow Wb\bar{b} \rightarrow \ell\nu b\bar{b}$ .

## 2.2 The Standard Model process $pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma$

The Feynman diagram depicting the full SM process  $pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma$  along with the component vertices making it up are illustrated in Figure 2.2.

**Figure 2.2:** The Feynman diagram for the process  $pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma$  is illustrated below, and is composed from the diagrams illustrated on the left, each of which is described in Section 2.2.



### 2.2.1 $pp \rightarrow Z/\gamma$

Similar to the  $W$  boson, the  $Z$  boson and the photon can also each be produced via the collision of quarks in the process  $q\bar{q} \rightarrow Z/\gamma$ . Unlike interactions with the  $W$  boson, interactions with  $Z/\gamma$  conserve parity invariance and do not transport charge. Any interaction which can happen as mediated by a photon can also happen with the exchange of a  $Z$  boson, but for collisions at  $\sqrt{s} < M_Z = 90$  GeV, the  $Z$  can not be made on-shell. In this low energy regime  $\gamma$  exchange dominates, but in 2015, the LHC ran at  $\sqrt{s} = 13$  TeV and the relative mass difference between the  $Z$  and the  $\gamma$  played a negligible role in their relative rates of production.

### 2.2.2 $Z \rightarrow \nu\bar{\nu}$

The only particle which the  $Z$  boson can couple to but the photon can not is the neutrino. At  $\sqrt{s} = 13$  TeV, the mass differences between the five lightest flavor of quark and the six leptons are negligible and the  $Z$  boson can decay into any kinematically allowed pairs,  $Z \rightarrow f\bar{f}$ . Including the three color possibilities for each quark, these are  $3 \times 5 + 6 = 21$  final states, each of which has roughly the same branching fraction. Therefore only approximately  $2/7$  of  $Z$  decays happen in the leptonic channel  $Z \rightarrow \ell\bar{\ell}$ . and of these decays, approximately  $2/3$  happen as  $Z \rightarrow \nu\bar{\nu}$ .

### 2.2.3 $pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma$

With the photon coupling only to electrically charged particles, the only place where a vertex containing a photon could be attached to either of the upper two left diagrams

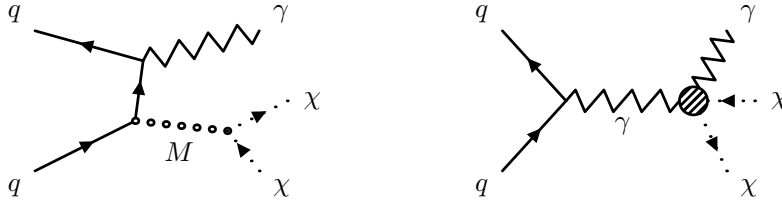
in Figure 2.2 is on one of the quarks. Photons are massless and therefore stable and so are a final state observable. Like the gluon in  $pp \rightarrow Wb\bar{b} \rightarrow \ell\nu b\bar{b}$ , the photon is an example of ISR. In the CM frame of the colliding  $q\bar{q}$ , which is approximately the lab rest frame for colliding beams of equal energy as is the case at the LHC, conservation of momentum dictates that the  $Z$  boson and photon should have equal and opposite momenta.

Unlike any of the other fermions, neutrinos are electrically neutral and therefore only interact via the weak force. So while the cross sections for most fermion-fermion interactions involve contributions from the comparatively stronger electromagnetic and strong forces, the neutrino cross section contains contributions from only the  $W$  and  $Z$  bosons at tree-level and is much smaller than that of the charged fermions. This makes the detection of neutrinos very difficult in general, and impossible to do with present technology given the extreme backgrounds present in a collider setting.

In the case where the ISR photon is recoiling against a  $Z$  boson which decays to neutrinos, no direct detection of the  $Z$  boson or of its decay products is possible, leaving only the photon visible in the final state. This is called the monophoton signature, where a photon is observed recoiling against apparently nothing, and while the monophoton signature is predicted to be observed as a result of SM process as in  $pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma$ , if the observed monophoton cross section is measured to be higher than predicted, it could also be an indicator of physics beyond the SM (BSM). Specifically, the monophoton signature used in searches for dark matter.

## 2.3 Beyond the Standard Model: $pp \rightarrow \gamma + \text{DM}$

**Figure 2.3:** Feynman diagrams for the DM process  $pp \rightarrow \gamma + \text{DM}$  using simplified models are illustrated below. On the left is the  $U(1)$  gauge model in which DM production is mediated by  $M$  which can be either vector or axial-vector,  $pp \rightarrow \gamma M \rightarrow \gamma \chi \bar{\chi}$ . On the right is the diagram for DM production using an EFT model of the  $\gamma\gamma\chi\bar{\chi}$  coupling, for the total process  $pp \rightarrow \gamma \rightarrow \gamma \chi \bar{\chi}$ .



The existence of particle dark matter is well motivated, and the simplified model theories of DM used in this thesis allow for interactions which can result in the monophoton signature. One of the classes of models considered is a  $U(1)$  gauge theory in which  $\chi\bar{\chi}$  is produced via a vector or axial-vector mediator  $M$  which couples to quarks. The tree-level process in this model which leaves a monophoton signature is illustrated on the left of Figure 2.3, and the relevant parameters governing the cross section of this interaction are the masses of the two particles,  $m_\chi$  and  $m_M$ , and the strengths of the couplings between  $M$  and quarks,  $g_{Mq}$ , and between  $M$  and DM,  $g_{M\chi}$ . An EFT describing the four-point interaction vertex  $\gamma\gamma\chi\bar{\chi}$  is also considered and illustrated on the right side of Figure 2.3. In this theory, the coupling is a function of two parameters,  $k_1$  and  $k_2$ , and is moderated by a mass scale,  $\Lambda$ . The other parameter in the EFT is  $m_\chi$ , and by measuring the cross section for  $pp \rightarrow \gamma + \text{DM}$  in comparison

with the SM prediction, estimations or limits can be set on the parameters used in either of these two models.

# Chapter 3

## THE LHC AND CMS

The measurements presented in this thesis are performed on data of proton-proton collisions collected by the Compact Muon Solenoid (CMS) detector and provided by the Large Hadron Collider (LHC) at the European Center for Nuclear Research (CERN). The LHC was designed to probe physics at the scale of TeV and is capable of operating at multiple energy scales. As measured in the CM frame of protons colliding inside CMS, the LHC operated at  $\sqrt{s} = 8$  TeV in 2012 and  $\sqrt{s} = 13$  TeV in 2015. The measurement of the  $pp \rightarrow Wb\bar{b}$  cross section is performed using  $19.8 \text{ fb}^{-1}$  of integrated luminosity collected at  $\sqrt{s} = 8$  TeV and the measurement of the  $pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma$  cross section and the extensions to set limits on DM models use  $2.3 \text{ fb}^{-1}$  of data collected at  $\sqrt{s} = 13$  TeV.

### 3.1 The Large Hadron Collider

The LHC is a single-ring, double-bore particle accelerator and collider located on the border of France and Switzerland outside Geneva. It was built using the existing 26.7 km of tunnels from the Large Electron Positron collider and hosts four primary experiments, located at four interaction points where beams of hadrons are made to cross. Of the four experiments, two (CMS and ATLAS) are built for studying SM processes and searching for new physics in general, one (ALICE) is designed to

investigate quark-gluon plasma resulting from the high energy collisions of heavy ions such as lead, and one (LHCb) was built for the study of b-mesons and CP violation.

### **3.1.1 LHC pre-acceleration**

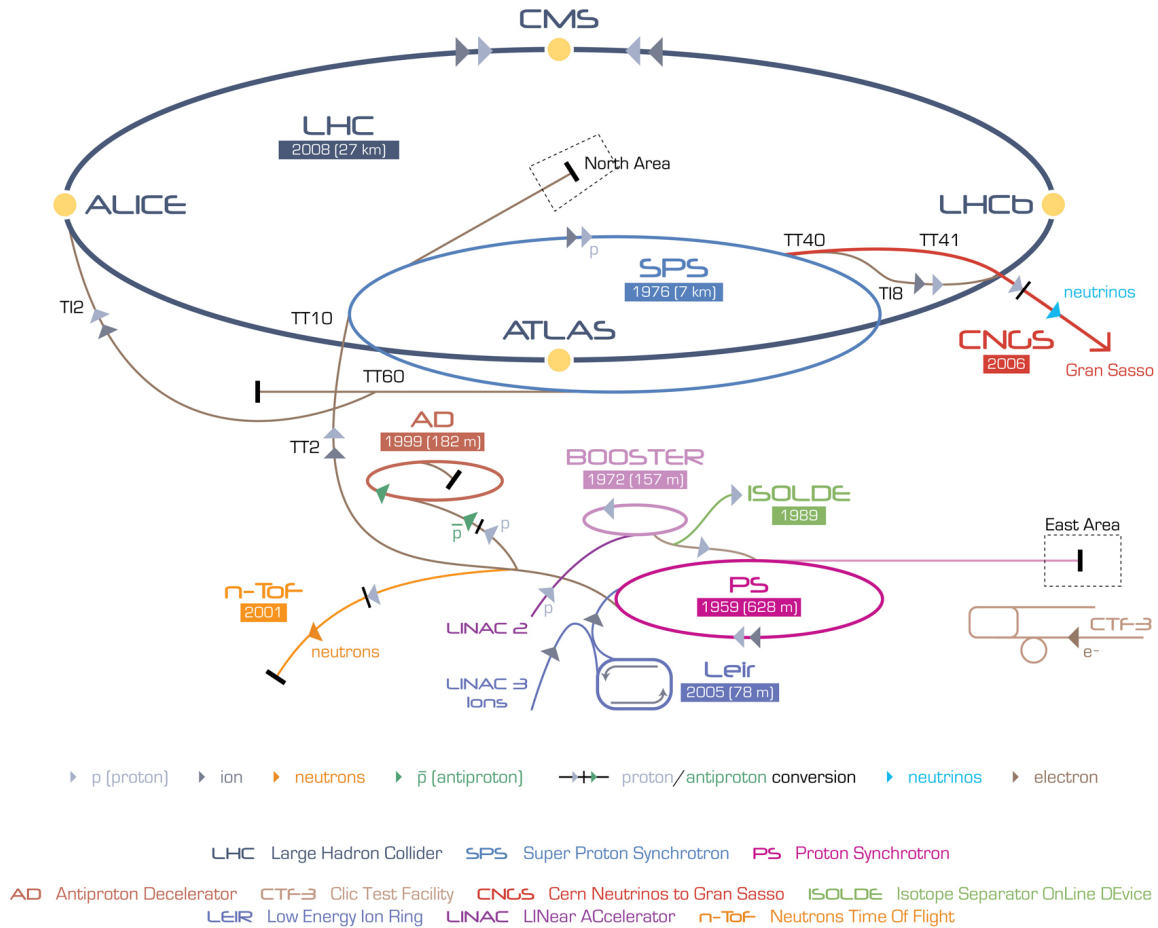
To accelerate protons to their collision energy, a multi-stage procedure is used and the major components of the accelerator infrastructure are illustrated in Figure 3.1. First, protons are separated from electrons in neutral hydrogen gas before entering the linear accelerator (LINAC2) which brings them up to an energy of 50 MeV using a series of oscillating electric potentials. In this process, rather than having a continuous stream of accelerating protons, the protons are grouped into bunches, and the beam retains this structure of distinct groups of protons separated by gaps throughout the acceleration procedure. After the LINAC2, protons enter the Proton Synchrotron Booster (BOOSTER) where they are accelerated to 1.4 GeV and prepared for injection into the Proton Synchrotron (PS). Inside the PS, bunches are accelerated to 26 GeV before being injected into the Super Proton Synchrotron (SPS) where they are further accelerated to 450 GeV. After the SPS, bunches of protons are sent into the LHC.

### **3.1.2 LHC acceleration**

The work of accelerating and containing the protons which form the beam of the LHC is done by superconducting magnets. They are cooled to a temperature of 1.9 K using liquid helium and are housed in the LHC dipole apparatus diagrammed in Figure 3.2. The dipole contains two beam pipes which are each surrounded by superconducting coils of Niobium Titanium (NbTi) which carry oscillating currents when in operation.



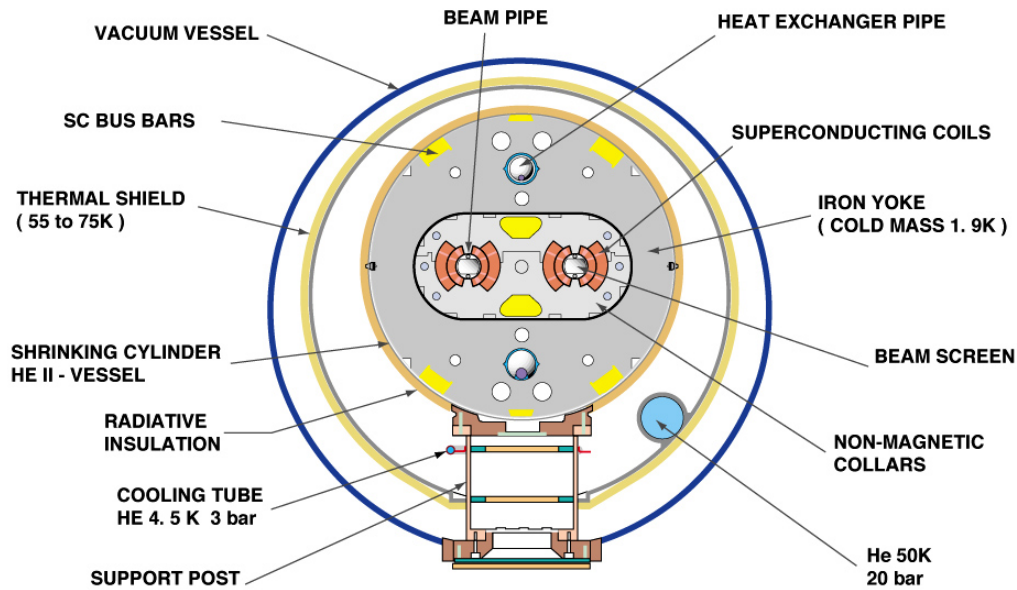
**Figure 3.1:** Before protons are released into the LHC for final acceleration and collision, they pass through the LINAC2, BOOSTER, PS, and SPS, undergoing acceleration at each stage.



The LHC contains RF cavities operating at 400 MHz which circulate proton bunches in opposing directions in the two beam pipes with a spacing of 25 ns between bunches. The magnets are capable of reaching a strength of over 8 T, a constraint imposed by the desired energy scale of the accelerator and the radius of the existing LEP tunnels in which the LHC was built. At individual beam energies of 4 and 6.5 TeV in 2012

**Figure 3.2:** Below is a cross section of the LHC dipole apparatus. It contains two beam pipes, each surrounded by superconducting magnetic coils which are held in place by an iron yolk. The system is cooled to a temperature of 1.9 K and is thermally isolated as well as protected from radiation.

### CROSS SECTION OF LHC DIPOLE



CERN AC\_HE107A\_V02/02/98

and 2015 respectively, the rest mass of the proton constitutes a negligible fraction of the proton momentum and the minimum magnetic field required by the magnets is 3 and 5 Tesla respectively. In practice, due to the noncircular geometry of the LHC, this field is closer to 8 T.

The rate at which a particular collision process occurs at the LHC is proportional to the cross section of that interaction and the luminosity of the colliding beams

as given in Equation 1.6. Assuming a Gaussian beam distribution, the machine parameters determine  $\mathcal{L}$  as

$$\mathcal{L} = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi \epsilon_n \beta^*} \mathcal{F}(\theta) \quad (3.1)$$

where  $N_b$  is the number of particles per bunch,  $n_b$  is the number of bunches per beam,  $f_{\text{rev}}$  is the revolution frequency of the bunches,  $\gamma_r = E_p/m_p$  is the relativistic gamma factor for protons at energy  $E_p$ ,  $\epsilon_n$  is the normalized emittance which characterizes bunch width,  $\beta^*$  is a measure for the betatron oscillation envelope, and  $\mathcal{F}(\theta)$  is a relativistic geometrical correction factor which is a function of the angle at which the beams cross. In addition to pushing the energy frontier, the LHC also has a significantly greater  $\mathcal{L}$  than previous hadron colliders. [Reference to Tevatron]

## 3.2 The Compact Muon Solenoid Detector

The CMS detector was built at Interaction Point 5 on the LHC ring to collect particle collision data exploiting the full physics reach of the LHC. The analysis of these data includes the discovery of the Higgs boson [9, 10] and high precision measurements of SM processes, as well as searches for physics beyond the standard model. To be able to perform such precision measurements, CMS was designed with four main subdetectors that work in concert and with a superconducting solenoid. The tracking and most of the calorimetric detectors are inside the solenoid while the muon detectors are outside. When running, the solenoid produces a 3.8 T uniform magnetic field in its interior, and has a uniform 2 T field over the bulk of the detector external to the

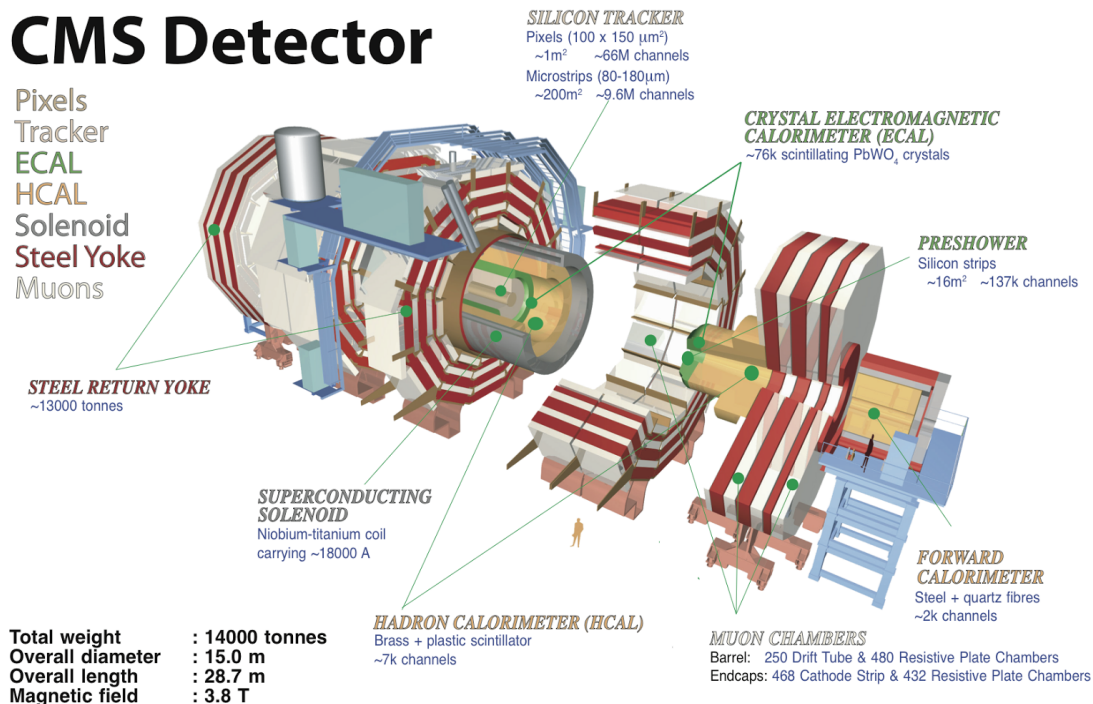
solenoid.

The innermost of the subdetectors is the tracker which uses silicon pixel and strip detectors to record the tracks of charged particles passing through it. The tracks are used in conjunction with the 3.8 T magnetic field to measure the momentum of these particles and this information is used for identifying the  $pp$  interaction vertex as well as locating secondary vertices from the decay of heavy flavor quarks such as the b or c. Outside the tracker is the electromagnetic calorimeter (ECAL), which is designed to have good energy resolution in recording the electromagnetic interactions of charged particles such as electrons or photons over a wide range of angles. The hadronic calorimeter (HCAL) is outside the ECAL and is designed to absorb energy which comes in the form of neutral hadrons and provide good resolution in missing transverse energy,  $E_T^{\text{miss}}$ . Outside the calorimeters is the solenoid and steel return yolk, and the outermost layers of the detector are dedicated to the efficient detection of muons. The overall length of CMS is 21.6 m, with a radius of 7.3 m and a total weight of 12500 tons.

### 3.2.1 Geometry

The coordinate system used by CMS is one in which the z-axis is aligned with the beam pipe, the y-axis is pointing upward vertically and the x-axis points radially inward toward the center of the LHC ring. The detector itself is mostly cylindrically symmetric about the beam pipe so cylindrical coordinates are also used. In this system,  $r$  is the radial distance as measured from the beam pipe, the azimuthal angle,  $\phi$ , is measured up from the x-axis in the x-y plane, and the polar angle,  $\theta$ , is measured

**Figure 3.3:** The CMS detector consists primarily of a tracker and electromagnetic and hadronic calorimeters which are mostly located inside a 3.8 T field provided by a superconducting solenoid, as well as a muon detection system located outside the solenoid.



down from the z-axis. The angle  $\theta$  is commonly replaced by pseudorapidity,

$$\eta = -\ln(\tan \theta/2) \quad (3.2)$$

since the distribution of particles is roughly constant as a function of  $\eta$ . For the calorimeters, "barrel" refers to the region of  $|\eta| < 1.4442$ , and "endcap" to the region  $3.0 > |\eta| > 1.566$ . Instrumentation cables are run through the gap between the barrel

and endcap, so this area has detecting components. The HCAL forward region (HF) covers  $3.0 < |\eta| < 5$  and the tracker extends to  $|\eta| < 2.5$ .

### 3.2.2 Magnet

To precisely measure the momentum of a charged particle, it is necessary to measure radius of curvature of that particle as it moves through a magnetic field. The momentum resolution varies as

$$\frac{\delta p}{p} \sim \frac{1}{L^2 B} \quad (3.3)$$

where  $L$  is the length of the track of the particle through a magnetic field of strength  $B$ . For particles at high energy, this requires a very strong magnetic field which is achieved by the superconducting solenoid in CMS. The solenoid operates at 3.8 T with a bore of 3 m in radius and 12.5 m in length and is constructed from four layers of NbTi superconductor. The steel yoke which provides physical support for the CMS structure and serves as an absorber for the muon system is fully saturated by the fringe magnetic field from the solenoid.

### 3.2.3 Tracking System

The inner tracking system of CMS is designed to provide precise and efficient measurements of the trajectories of charged particles produced during collisions, as well as a precise reconstruction of secondary vertices. The tracker has a length of 5.8 m and a radius of 1.25 m in a cylindrical structure surrounding the interaction point, as

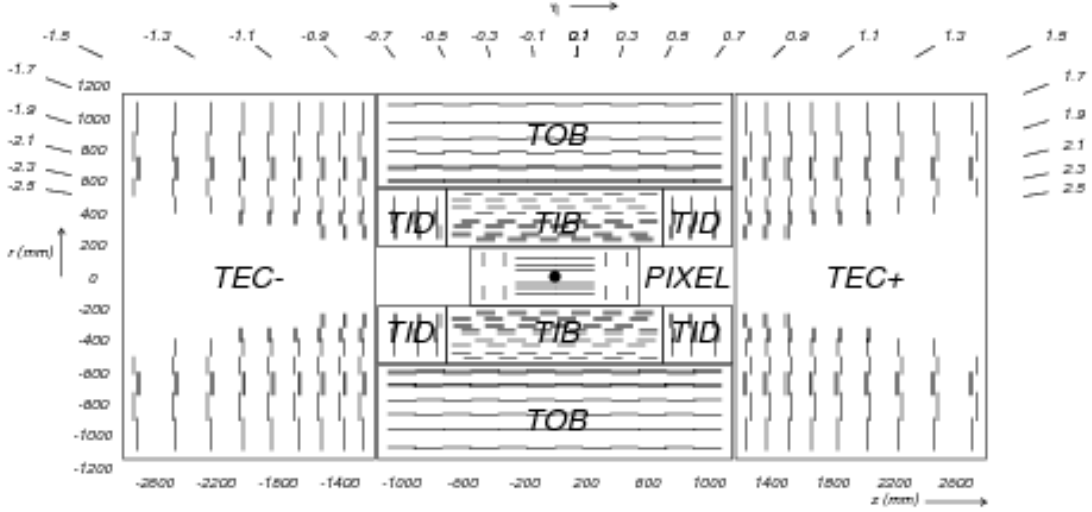
illustrated in Figure 3.4.

At the core of the tracker and closest to the beam line are three concentric cylindrical layers of hybrid pixel detector modules which are complemented by two discs of silicon pixel modules on each end and extend to a distance of 10 cm from the beam line. In total, the pixel component of the tracker covers an area of about 1 m<sup>2</sup> with 66 million pixels. External to the pixel detector are the tracker inner barrel and discs (TIB/TID) which are made from silicon strips and extend out to a distance of 55 cm. There are four layers of strips in the TIB, with 3 discs at each end. The tracker outer barrel (TOB) is composed of 6 layers of micro-strip sensors and extends in  $z$  between  $\pm 118$  cm and to a radius of 116 cm. At the end of the  $z$  range for the TOB are the tracker end caps (TEC) which cover the ranges  $124 < |z| < 282$  cm and  $22.5 \text{ cm} < r < 113.5$  cm. Each TEC is composed of 9 discs, each carrying up to 7 rings of silicon micro-strip detectors. In total, the tracker contains 9.3 million strips which cover an area of 198 m<sup>2</sup> and extends to an acceptance of  $|\eta| < 2.5$ . For tracks with momentum on the order of 100 GeV, the momentum resolution is around 1-2% up to  $|\eta| < 1.6$  and degrades to around 10% with increasing  $\eta$ .

### 3.2.4 Electronic Calorimeter

The electronic calorimeter (ECAL) is a homogeneous calorimeter made from nearly 76000 crystals of lead tungstate (PbWO<sub>4</sub>) mounted in the barrel and endcap sections with a preshower detector located in front of the endcaps, arranged as shown in Figure 3.5 In the barrel, avalanche photodiodes are used as photodetectors, and in the endcap vacuum phototriodes are used. The material PbWO<sub>4</sub> was chosen for its

**Figure 3.4:** Below is a schematic of the CMS tracking system where each line represents a detector module. The system is made from silicon pixels and silicon microstrips distributed into four sections, TIB, TID, TOB, TEC.



properties of being dense, optically transparent and radiation hard. The radiation length inside the ECAL is typically less than 1 cm with a Moliere radius of 2.2 cm and about 80% of the light is emitted from a crystal within the first 25 ns. Since the length of a given crystal is on the order of 20 cm, most photons and electrons deposit all of their energy within the ECAL, and do not reach the HCAL.

The use of  $\text{PbWO}_4$  crystals allows for excellent position and timing resolution with the energy resolution given by

$$\left(\frac{\delta_E}{E}\right) = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right) + (0.30\%)^2. \quad (3.4)$$

In this expression, the first term comes from the statistical error in the measurement



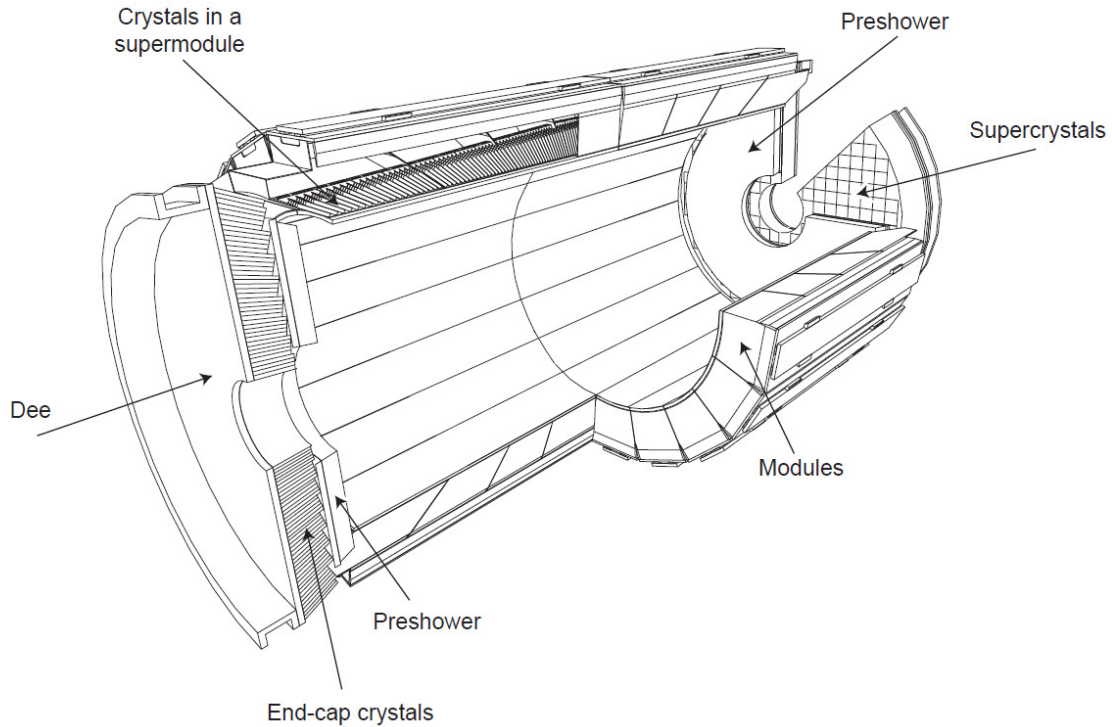
which arises from the stochastic nature of electromagnetic shower evolution and the second term represents the error in the measurement which results from noise in the electronics or energy deposits from additional soft interactions. The ECAL provides stable and accurate measurements of energies over a range from 1 GeV to 1 TeV, with the upper limit set by the energy at which electromagnetic showers penetrate through the ECAL into the HCAL. With time, after undergoing a heavy bombardment of high energy radiation, the  $\text{PbWO}_4$  crystals physically deteriorate and develop nonuniform light transmission properties. This is monitored and corrected for using a laser calibration system that probes for changes in crystal transparency.

### 3.2.5 Hadronic Calorimeter

Situated mostly between the ECAL and the superconducting solenoid is the hadronic calorimeter (HCAL) which plays a crucial role in the measurement of hadron jets and particles such as neutrinos which escape the detector and result in apparent missing transverse energy. The HCAL is designed to contain the energy of neutral particles which pass through the ECAL and is therefore made from dense materials such as steel and brass interleaved with scintillating material. Because the HCAL is designed to fit between these two components, it takes the shape of a hollow cylinder of inner radius 1.77 m and outer radius 2.95 m and one half of the HCAL is illustrated in Figure 3.6.

The barrel of the HCAL (HB) extends to  $|\eta| < 1.3$  and is constructed from brass absorber plate wedges aligned parallel to the beam axis and mounted in an overlapping configuration, with a smaller amount of steel used in the inner and

**Figure 3.5:** Below is a diagram of the ECAL, which sits between the tracker and HCAL in CMS. It is made from  $\text{PbWO}_4$  crystals throughout the volume with avalanche photodiodes in the barrel and vacuum phototriodes in the endcaps.



outermost wedges for structural stability. The endcap of the HCAL (HE) extends this coverage to  $|\eta| < 3.0$  and is complemented by the forward hadron calorimeter (HF) which is made from the comparatively radiation-hard steel plates embedded with quartz fibers. Inside the barrel region there is an additional layer of the HCAL, the outer calorimeter (HO), which is located just outside the solenoid and uses it as an absorber for energetic showers which start late in the HB.

In the HB, HO and HE, light from particle showers inside scintillators and collected by quartz fibers is used as an estimate of the total energy of the shower. In the HF, this estimate is made using the Cherenkov radiation from particles with energy above 190 keV collected by the quartz fibers. For the two cases, the energy resolution takes the same functional form

$$\left(\frac{\delta_E}{E}\right) = \left(\frac{A}{\sqrt{E}}\right)^2 + (B)^2 \quad (3.5)$$

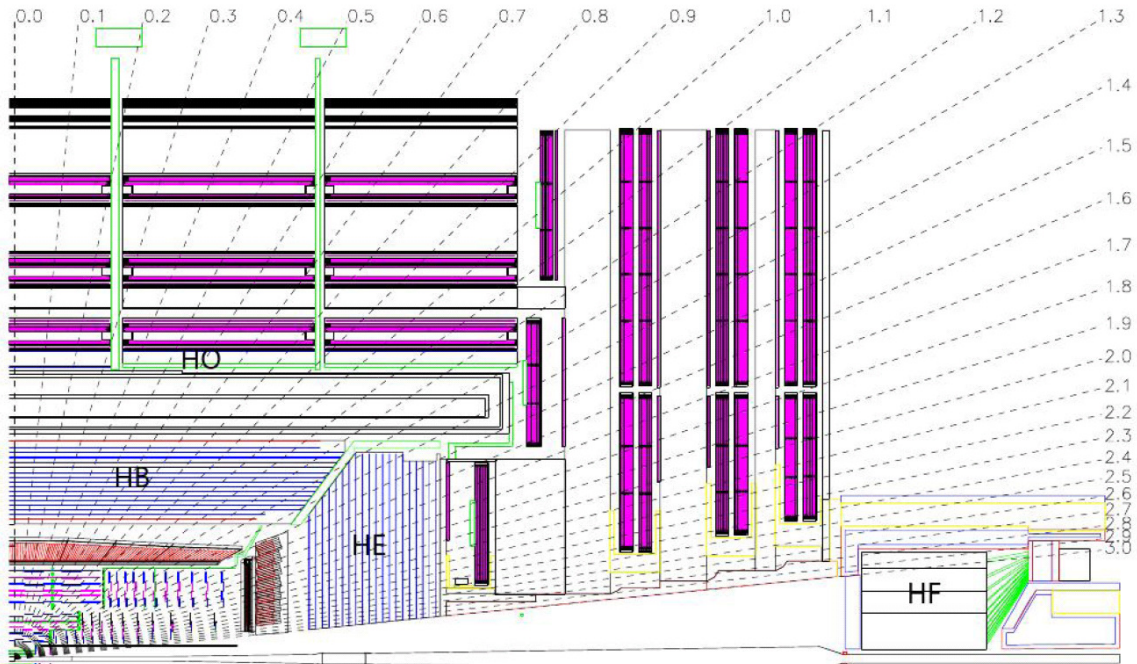
where  $A$  is 90% (172%) in the HB/HO/HE (HF) and relates to the stochastic uncertainty of shower evolution and  $B$  is 4.5% (9.0%) and comes from uncertainties in calibration.

### 3.2.6 Muon System

Muons play a central role in the physics program outlined by CMS and the muon detection system is positioned as the outermost layer of the detector. Unlike the other charged leptons, muons typically pass through the ECAL and HCAL and deposit only a fraction of their energy, so a dedicated muon system is necessary in order to determine the momentum of these particles. The muon system is composed of three different kinds of gaseous detectors, drift tubes (DTs), resistive plate chambers (RPCs) and cathode strip chambers (CSCs) and their layout is illustrated in Figure 3.7.

The barrel region of the muon system is covered by DTs in the range  $|\eta| < 1.2$  and the endcaps are covered by CSCs in the range  $0.9 < |\eta| < 2.4$ . The RPCs are located in the range  $|\eta| < 1.6$  and provide fast, independent and highly segmented

**Figure 3.6:** A schematic layout of the HCAL, which complements the ECAL in providing a measurement of the total energy produced in a collision. The HCAL is made from brass and steel plates, embedded with quartz fibers.



transverse momentum measurements of muons.

The DT system is composed of 4 stations which form concentric cylinders about the beam line and contain 172000 sensitive wires. As charged particles enter the DTs, they ionize the Ar/CO<sub>2</sub> gas mixture, knocking off electrons which then are attracted to the positively charged wires.

The CSCs are less sensitive to uneven magnetic fields and high particle rates so are therefore used in the endcaps. They are made from crossed arrays of positively

charged wires and negatively charged strips in gas and are composed of six layers, giving them precise timing as well as positional information. As an upgrade between the 2012 and 2015 data taking periods, a fourth layer of CSCs was added to the CMS detector, adding to the three which were present in 2012.

The RPCs are built from two sheets held at opposite charges and separated by a gas volume. As muons move through the chamber, electrons are ionized from the gas and attracted to small metallic strips which they reach after a small but well known time delay. The timing resolution of RPCs is on the order of 1 ns.

### **3.2.7 Trigger and Data Acquisition**

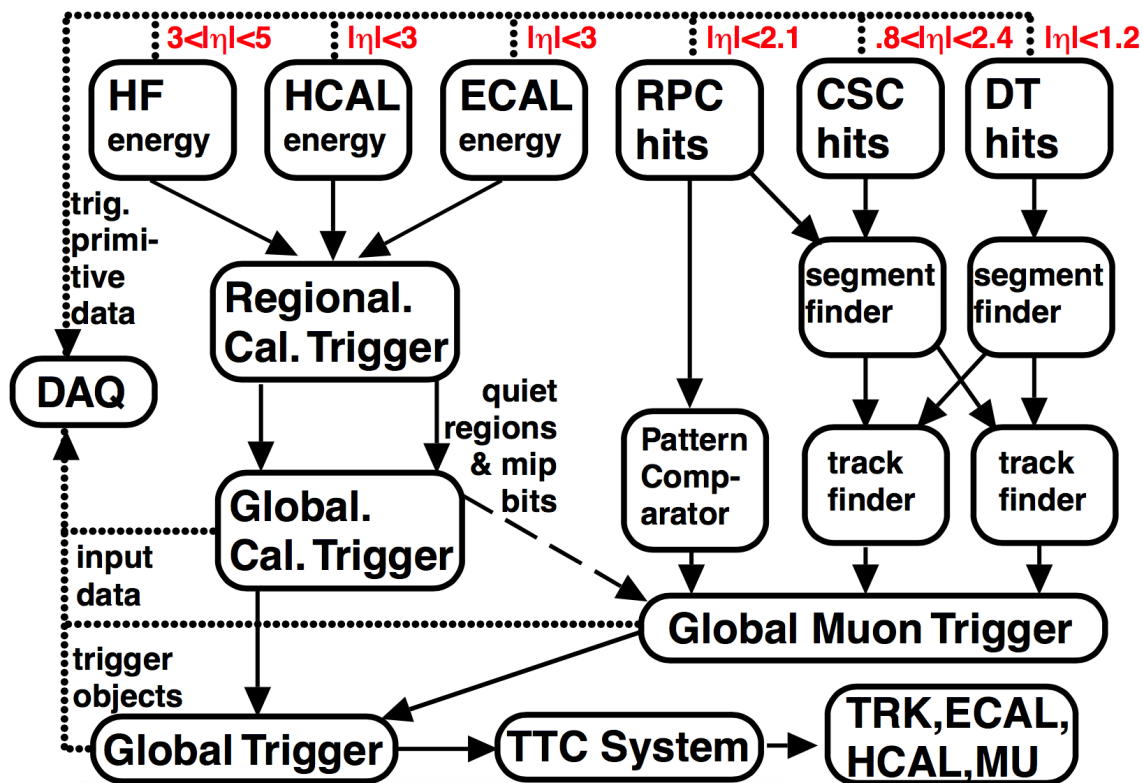
In 2012, the LHC delivered proton-proton collisions with a spacing of 50 ns between collisions, and in 2015 this spacing was reduced to 25 ns. With these rates of collisions, it is not possible to permanently save the complete set of information about energy deposits in the detector, so the triggering system is of crucial importance. The purpose of the trigger system is to reduce the input rate of data flow down to a level at which it can be recorded, specifically choosing to keep those events which are most likely to involve the physics processes being studied. The trigger system is further divided into two parts.

#### **3.2.7.1 Level-1 Trigger**

The Level-1 trigger (L1T) is the first layer of processing, and this processing is done using custom hardware and software to reduce the rate of data flow down to 100 kHz. A schematic of the L1T as used in 2012 is given in Figure 3.8. showing how the L1T



**Figure 3.8:** Below is a diagram illustrating the flow of information through the 2012 Level-1 Trigger. The L1T is composed of the RCT the GMT, and the GT. The RCT and GMT both feed in to the GT which makes the final trigger decisions and passes this information off to the DAQ for storage and to the TTC system for feedback.

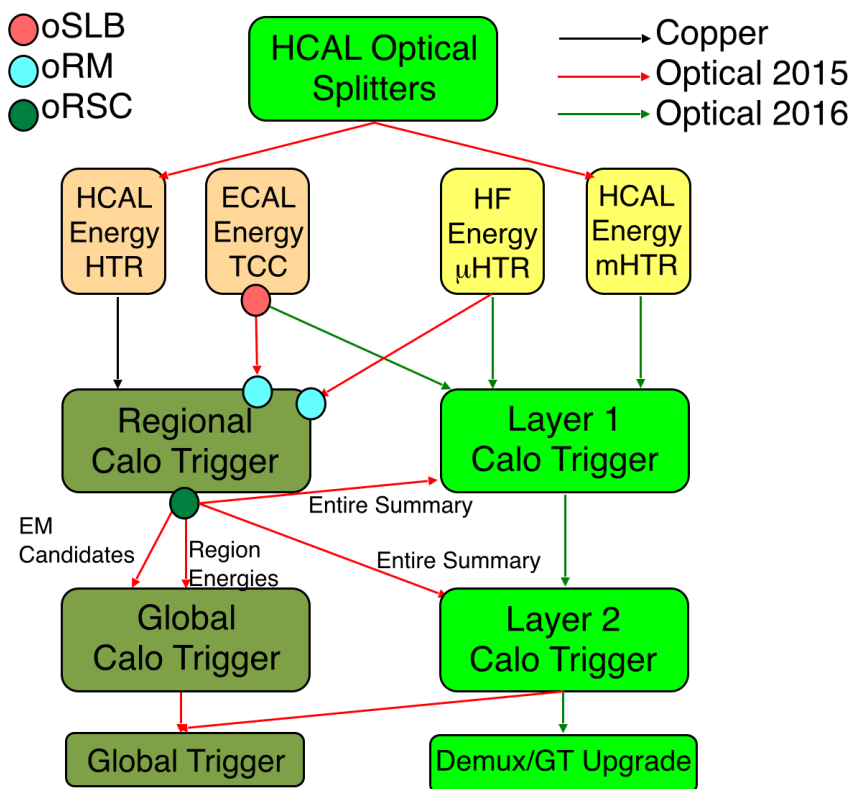


energy. Objects are considered isolated if there is a relatively low fraction of energy in nearby towers. This information is sent to the Global Trigger (GT) along with information about regions in the detector without energy deposits to be used in coordination with the Global Muon Trigger (GMT) for determining muon isolation.

During the 2015 data-taking period, the RCT was run in parallel with an upgraded calorimeter trigger (UCT), as illustrated in Figure 3.9. The UCT fulfilled the same

functionality as the RCT, but incorporated improvements in both the hardware and software to accommodate for the higher energies and rates of collisions that had been anticipated for the 2015 data-taking. These improvements included switching the copper cables which connected the RCT to the calorimeters with optical fibers, as well as the upgrading to modern FPGAs which can transmit data at 10 GB/s. This allowed for more sophisticated identification and pileup subtraction algorithms.

**Figure 3.9:** Below is a diagram illustrating the flow of information through the UCT in 2015. The RCT and the UCT were both run in parallel in 2015.



The GMT uses information about the geometry and total energy of hits in the RPC, CSC and DT systems collectively to find candidate muons. Because of the



overlap between the muon subsystems, information on reconstructed tracks is shared between the CSCs and the DTs and the CSCs also take as input information from the RPCs to disambiguate multiple hits in the same chamber.

All of this information is used, along with the information about quiet regions in the calorimeters sent from the RCT, to determine candidate isolated and non-isolated muons. The RCT and GMT both feed in to the GT which makes the final trigger decisions on which events to investigate further and passes this information off to the Data Acquisition (DAQ) system for validation and storage as well as to the Trigger Timing and Control (TTC) system for real time feedback.

### 3.2.7.2 High Level Trigger

The High Level Trigger (HLT) uses custom software and algorithms on commercial hardware to further process the output from the GT and reduce the rate of events being processed to 100 Hz. Because the HLT has access to TPs from all of the detector subsystems on events which have already passed the first selection stage of the L1T, algorithms designed to mimic the final object reconstruction described in Section 4.2 are used. These algorithms are optimized to compute quantities which can be computed quickly, and as events are rejected during the chain of HLT processing, the processing on the events is stopped so resources can be reallocated quickly. In the  $pp \rightarrow Wb\bar{b}$  analysis, the triggers used are based on finding an isolated muon (electron) with  $p_T > 24$  (27) GeV. In the  $pp \rightarrow \gamma + \text{invisible}$  analyses, localized clusters of energy deposits are required to have 90% of the total energy deposited in the ECAL.

# Chapter 4

## SIMULATION AND RECONSTRUCTION

### 4.1 Simulation of Events

Vital to the analysis of the data gathered using the CMS detector are accompanying predictions to be compared against. Good predictions can be used not only for direct comparison against data as in the case of a cross section measurement, but can also be used in the design of future detectors and experiments, or for optimizations of parameters in blinded analyses. Predictions are made by simulating  $pp$  collisions and the subsequent decays and interactions that take place inside the CMS detector volume using Monte Carlo (MC) techniques. The first step in producing these simulations is the generation of the collision event itself, and the second is in simulating the interaction of the collision products with the detector.

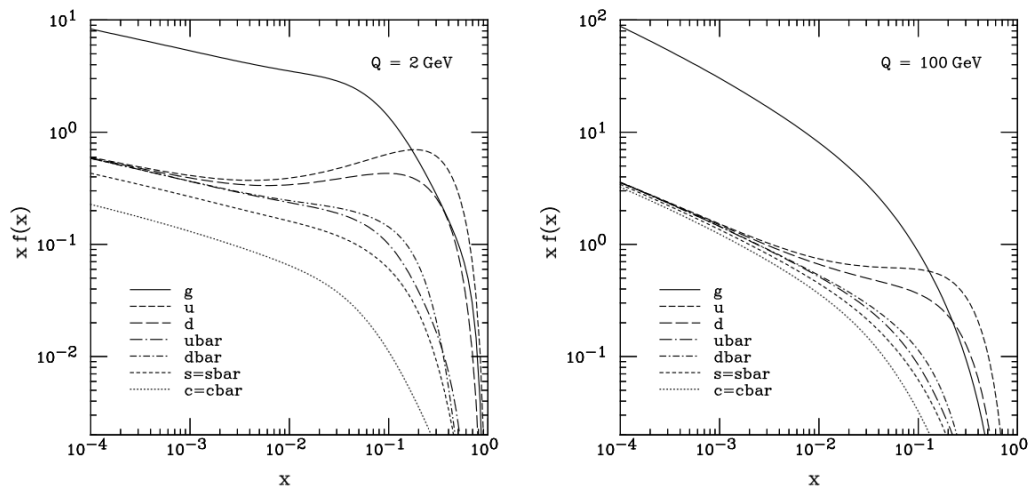
#### 4.1.1 Monte Carlo Event Generation

There are two complimentary methods used in producing a simulation of the collision event, the direct calculation of the scattering amplitude (also known as the matrix element, ME), and the showering of particles as they decay, hadronize and radiate. Particles produced in the in the collision before showering and hadronization are referred to as prompt.

As discussed in Section 1.1.4, protons are composite objects which form as the result of strong interactions between bound quarks. Protons are therefore modeled using parton distribution functions (PDFs) which describe the probability of finding a given constituent particle, or parton, to have a given fraction of the momentum of the proton. Parton densities vary as a function of the energy used to probe the proton and the quantity of energy absorbed by the proton is known as the momentum transfer,  $Q$ . The PDF is thus actually a set of density functions, one for each parton taken into account and for each momentum transfer, and examples of the CTEQ6M parton distribution are shown in Figure 4.1. Weighted with probabilities from the PDF set and summed over, MEs are calculated explicitly from Feynman diagrams which have initial state particles found in the PDF and the desired final state. If only the first four lightest quarks are included in the PDF, the PDF is considered four-flavor (4F), and if the  $b$  quark is also considered directly in the PDF, then it is five-flavor (5F).

Radiation is also important to correctly model. At any point in the collision any colored particle can radiate a gluon, and any charged particle can radiate a photon. Gluon radiation from initial state partons is always present in  $pp$  collisions at the LHC and results in jets, columnated showers of particles which are the products of quark hadronization and gluon splitting. The effects of radiation and parton showering are simulated using a Markov process in which vertices iteratively are added to partons with probabilities based on the coupling strengths, energies of the participants and the generation of random numbers. For the constituents of the proton which did not participate in the hard interaction, quarks must be created from the vacuum

**Figure 4.1:** Below are proton PDFs shown at two different values of momentum transfer,  $Q$ . The horizontal axis shows the momentum fraction carried by the parton and the vertical axis shows the parton density.



to enforce confinement. This produces low energy, soft, radiation and is known as the underlying event. The underlying event must be simulated along with the hadronization effects for all colored particles.

## 4.1.2 Monte Carlo Generators

Each of the MC generators used in this thesis are designed to serve one of two purposes. They either calculate MEs for hard scattering processes or they compute corrections such as hadronization or NLO effects.

#### 4.1.2.1 Matrix Element Calculators

For a given  $2 \rightarrow n'$  scattering process, the differential cross section is a function of the Lorentz-invariant phase space as in Equation 1.3. To calculate the cross section within a finite phase space,  $d\sigma$  is integrated and MADGRAPH numerically does this at LO through the sampling of random numbers. The phase space can be interpreted as a multidimensional hypercube spanning all degrees of freedom for all final state particles and Equation 1.3 is integrated over using the VEGAS package to calculate a weight,  $dw$ , for each point sampled in the phase space. The average of the weights converges towards  $\int dw$ .

To produce events with the frequency predicted by the theory being modeled, MADEVENT uses the Von Neumann method to unweight events. For each event, a random number,  $g$ , is generated between 0 and 1 and compared to the ratio  $dw/dw_{\max}$  where  $dw_{\max}$  is the largest event weight sampled. If  $dw/dw_{\max} > g$  then the event is kept and is otherwise rejected. Accepted events generated with MADGRAPH/MADEVENT in this way have the same frequency and follow the same kinematic distributions as predicted from the input Equation 1.3.

A Monte Carlo for FeMtobarn (MCFM) and Fully Exclusive W,Z Production (FEWZ) also use VEGAS to perform phase space integrations, but MCFM does so at NLO and FEWZ at NNLO. With a much larger phase space to integrate over than MADGRAPH, MCFM performs optimizations on the integration and produces weighted events. With weighted events, an estimation of a cross section to NLO accuracy can be made, but is made only at the level of partons interacting, and in particular does not include corrections due to hadronization effects. Estimations made by MCFM thus

need to be corrected for hadronization effects before being compared with observed cross sections.

#### 4.1.2.2 Showering and NLO Corrections

Radiation and hadronization are modeled with PYTHIA, which uses the Lund string model in which quarks are confined to the ends of strings and gluons are represented as kinks on that string. As quarks separate, the string breaks and creates a  $q\bar{q}$  pair, thus building confinement directly into the model. Jets can thus either appear in the final state as a result of the ME calculation or as radiation and to avoid double counting, the  $k_t$ -MLM matching scheme is used. The underlying event is modeled in PYTHIA as a set of  $2 \rightarrow 2$  processes which are correlated with each other via the color connections present in the proton, and the set of parameters used by PYTHIA to perform these calculations is referred to as the tune. The most recent PYTHIA Z2\* tune is derived from the Z1 tune [11], which uses the CTEQ5L parton distribution set, whereas Z2\* adopts CTEQ6L [12].

Corrections due to NLO effects are handled by the Positive Weight Hardest Emission Generator (POWHEG). Powheg replaces the hardest jet in the event with the NLO prediction. A similar technique is used by aMC@NLO, which was modified to interface directly with MADGRAPH to produce fully consistent NLO predictions for use on the 2015 LHC dataset. The combined generator is called MADGRAPH\_aMC@NLO.

### 4.1.3 Detector Simulation

After events have been produced, they are passed to Geant4 for simulation of the passage of particles through the physical mass of CMS. The Geant4 toolkit includes a full model of CMS, including all of the subdetectors as well as the inert material from the support structure and readout electronics. The magnetic field is emulated using data from measurements on the real field and Geant4 uses all of this information to register hits in the simulated detector as a consequence of the interaction between particles produced in the simulated event and the simulated material of CMS. Additionally, hits are added to the simulated detector taking into account the rates of background noise, and the final output is emulated data which is stored in the same way as would be data as taken from the real detector.

## 4.2 Reconstruction of Events

Real data collected from the detector and simulated data output from Geant4 consist of time-correlated energy deposits in the various subdetectors of CMS. As a result of the coordinated designs of the subdetectors, the final-state particles which arise from  $pp$  collisions at the LHC can be individually identified and reconstructed using the combined information from the entirety of CMS. The associated global event description from this particle-flow (PF) reconstruction provides excellent performance for the identification of electrons and muons, as well as for vertex identification and the evaluation of  $E_T^{\text{miss}}$ . As a result of the PF reconstruction of an event, particles are identified and placed into the mutually exclusive categories: charged and neutral

hadrons, photons, electrons and muons.

### 4.2.1 Track and Primary Vertex Reconstruction

The subdetector closest to the interaction vertex is the tracker, which records precise information about the trajectories of charged particles as they pass through it. Combined with the magnetic field, this allows for the measurement of the momenta of these particles as well as a means of identifying the the location of the primary interaction.

Tracks are identified via an iterative process. The first tracks to be reconstructed are those which pass strict seeding criteria, designed to have a moderate efficiency, but negligibly small fake rate. Then the detector hits associated with these tracks are masked and the remaining hits are used to form track seeds with slightly relaxed criteria. This operation is repeated, with every iteration imposing more complex and time-consuming seeding, filtering and track fitting algorithms.

Because bunches of protons instead of single protons are made to cross in the LHC, multiple collisions can take place during the same bunch crossing. The vertex with the highest scalar sum transverse momentum,  $p_T$ , of tracks and passing further quality selections based on the goodness of fit for the tracks and the number of tracks associated with a given vertex is chosen as the primary vertex (PV). The other vertices are referred to as pileup vertices and the associated collision products as pileup.

In  $pp \rightarrow Wb\bar{b} \rightarrow \ell\nu b\bar{b}$  events, the two  $b$  quarks and the lepton from the  $W$  decay all leave energy deposits in the tracker, thus making the choice of PV unambiguous.



However, in the  $pp \rightarrow \gamma + \text{invisible}$  events, the only visible final state object is a photon, and photons do not leave hits in the tracker. This makes the identification of the PV in the monophoton analysis difficult and motivates the using of variables that are less sensitive to correct PV identification.

## 4.2.2 Electron ID and Reconstruction

Electrons are reconstructed using tracker hits and ECAL deposits. The seed of an electron candidate is selected as an energy deposit in the ECAL with  $E_T > 4$  GeV having nearby deposits in the tracker. As electrons move in magnetic fields, they emit bremsstrahlung radiation tangential to their flight path and this radiation both appears in the detector, and alters the course of the electron.

The effects of this radiation are taken into account via the Gaussian Sum Filter (GSF) track fitting algorithm. This algorithm uses weighted sums of Gaussian functions to describe electron energy loss and thus allows for non-Gaussian corrections to the fitting of tracks. In the CMS detector, bremsstrahlung from electrons results in the emission of photons in an extended strip in the  $\phi$  direction and electron superclusters (SCs) are made by including the energy deposits from these photons in the ECAL as part of the candidate electron object.

Further requirements during the reconstruction of the electron improve the purity of selection. The SC and the GSF track are required to be separated by no more than  $|\eta| < 0.02$  and  $|\phi| < 0.15$  and the fraction of energy deposited in the HCAL directly behind the SC, and the SC is required to be no more than 15%.

### 4.2.3 Muon ID and Reconstruction

Muon identification is performed using two reconstruction and filtering methods to produce ‘tracker muons’ and ‘standalone muons’ which are combined to form ‘global muons.’ Tracker muons are identified starting with a track,  $p_T > 0.5$  GeV and  $p > 2.5$  GeV, which is then extrapolated to the muon system. If the distance between the the extrapolated track and the nearest hit in one of the muon chambers is less than 3 cm, a tracker muon is identified. Tracker muons are also identified if the pull between the extrapolated track and the matched station hit is less than four, where pull is defined as the distance between the track and the station hit divided by the uncertainties on both measured quantities. Tracker muons are built from the inside of the detector towards the outside, and standalone muons are built in the other direction. Only hits in the muon stations are used to reconstruct standalone muons, with the additional constraint that the path reconstructed from the hits points back toward the interaction region. Thus, the tracker muon algorithm is well-suited for the identification of low- $p_T$  muons by having low thresholds and requiring only one track and one station hit, while the standalone muon algorithm is aimed at high- $p_T$  muons which have the energy to penetrate multiple layers of muon stations to form tracks which can be traced back to the interaction. Global muons are required to pass the criteria for both standalone muons and tracker muons, and, starting with the standalone muons, the global muon trajectory is refit using information from both the muon stations and the tracker, yielding an improved energy resolution than either one.

#### 4.2.4 Lepton Isolation

After having been identified and reconstructed, leptons are also assessed in terms of their isolation. Isolation is an important variable for differentiating prompt leptons from leptons resulting from secondary decays or hadronization effects and is defined as

$$I = \frac{\sum p_T^{\text{charged}} + \max(0, \sum p_T^\gamma + \sum E_T^{\text{neutral}} - 0.5 \cdot p_T^{\text{PU}})}{p_T^\ell}, \quad (4.1)$$

with the sum running over the PF candidates (hadrons, electrons, photons) in a cone of size  $\Delta R < 0.4$  (0.3) around the muon (electron) direction. The last term in the numerator,  $p_T^{\text{PU}}$ , is a correction for pileup effects and is based on the scalar sum of transverse momenta of charged particles not associated with the primary vertex that are within the isolation cone.

#### 4.2.5 Photon ID and Reconstruction

Photons are reconstructed using the same ECAL clustering algorithms as are used for electrons. This allows for the simultaneous reconstruction of photons that have and have not split to  $e\bar{e}$  pairs. The size of the SC is determined dynamically and the center is determined to be the barycenter of the distribution, with weights assigned using the logarithm of the fractional energy deposits of the ECAL crystals clustered in the SC. The angular width of the cluster is  $\sigma_{\eta\eta}$  and photons tend to have narrow deposits.

In an ideal tracker, photons would not interact at all and objects that leave signatures similar to those of photons could be rejected through the rejection of tracks.

However, some photons do convert to  $e\bar{e}$  pairs inside the tracker volume which leave tracks, so the rejection of tracks is not a perfect way to distinguish between photons and electrons.

For a photon object to be considered isolated, scalar sums of the transverse momenta of PF charged hadrons, neutral hadrons, and photons within a cone of  $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.3$  around the candidate photon must individually fall below bounds defined for 80% signal efficiency. Only the PF candidates that do not overlap with the electromagnetic shower of the candidate photon are used in the isolation sums.

#### 4.2.6 Missing Transverse Energy

The transverse mass of the  $W$  boson is defined in Equation 4.1, and is a natural discriminator against non- $W$  final states such as QCD multijet events, that have a lepton candidate and  $E_T^{\text{miss}}$ , but a relatively low value of  $m_T$ . In calculating  $m_T$ , the  $E_T^{\text{miss}}$  is corrected for noise in the ECAL and HCAL [13]. and corrections to limit the effects of pileup are also included [14].

#### 4.2.7 Jet ID and Secondary Vertices

The reconstruction of jets is accomplished using the anti- $k_t$  clustering algorithm on particles identified in the PF. The anti- $k_t$  algorithm is both infrared and collinear safe, meaning that it is stable against soft (low energy) radiation getting clustered into individual jets, and also stable against hard (high energy) jets splitting collinearly and affecting the shape of the jet.

Jets are corrected in simulation and in data to remove energy believed to come from elsewhere than the PV, thus removing the luminosity dependence of the jet. Jets are also corrected to have a response that is independent of  $\eta$  by studying dijet events and calibrating the jets to anti-align. To make the jet response independent of the  $p_T$  of the jet, an absolute correction is applied, and in data, one further correction on the relative energy scale is applied. After all of these corrections are applied, simulated jets are observed to have sharper energy resolution than is observed, so jets in MC are smeared in energy.

Bottom quarks have a relatively long lifetime and are the heaviest fundamental particle that has been seen to decay inside the volume of the CMS detector. A  $b$  quark produced in a  $pp$  collision at CMS therefore has enough time to hadronize into a jet before decaying, and such jets are called  $b$ -jets. The identification, or tagging, of  $b$ -jets is focused around the vertex associated with the  $b$ -hadron which, since it is not the PV but is still a vertex associated with the event, is called a secondary vertex, SV. The tagging of  $b$ -jets is accomplished using a multivariate analysis technique in which information about the number and energy of tracks associated with the SV and their corresponding alignment with the SV, as well as the presence and energies of soft leptons is combined into a single discriminator value.

# Chapter 5

## $Wb\bar{b}$ CROSS SECTION MEASUREMENT

The first of the two processes examined in this thesis is the SM production of  $pp \rightarrow Wb\bar{b} \rightarrow \ell\nu b\bar{b}$  at  $\sqrt{s} = 8$  TeV, described in Section 2.1.

### 5.1 Event Selection

Two decay channels of the  $W$  boson are considered,  $W \rightarrow \mu\nu_\mu$  and  $W \rightarrow e\nu_e$ , and events are selected using single-muon (single-electron) triggers with a loosely isolated muon (electron) with transverse momentum  $p_T > 24$  (27) GeV and pseudo-rapidity  $|\eta| < 2.1$  (2.5). Individual particles emerging from each collision are reconstructed with the particle-flow (PF) technique described in Section 4.2.

Both the muon and electron candidates are required to have  $p_T > 30$  GeV,  $|\eta| < 2.1$  and to originate from the PV of the event, as defined in Section 4.2.1. These leptons must also pass tight ID and isolation requirements, with  $I < 0.12$  (0.10) for selected muons (electrons) as defined in Equation 4.1. Additionally, to reject some of the background events contributed by processes such as  $pp \rightarrow t\bar{t}$  and  $pp \rightarrow Zj$ , events with more than one lepton  $p_T > 10$  GeV and  $|\eta| < 2.4$  are rejected. Jets are constructed from PF candidates using the anti- $k_t$  clustering algorithm [15] with a distance parameter of 0.5, and jets are  $b$ -tagged using the multivariate discriminator at an operating point with a efficiency of 40% and a misidentification probability

**Table 5.1:** Listed below are the raw number of events in data passing the selection listed in the first column as well as all selections in higher rows. The second column is for the  $W \rightarrow \mu\nu$  decay channel, and the third column is for  $W \rightarrow e\nu$ .

Selection	Events passing selection	
	$W \rightarrow \mu\nu$	$W \rightarrow e\nu$
$p_T^{\ell_1} > 30 \text{ GeV}$ and $ \eta^{\ell_1}  < 2.1$	67249140	58079638
$p_T^{\ell_2} \not> 10 \text{ GeV}$ or $ \eta^{\ell_2}  \not< 2.1$	58297430	52911675
$p_T^{j_1, j_2} > 25 \text{ GeV}$ and $ \eta^{j_1, j_2}  < 2.4$	5251248	5880076
$p_T^{j_3} \not> 25 \text{ GeV}$ or $ \eta^{j_3}  \not< 4.7$	3212465	3618448
$b$ -tag $j_1, j_2$	7432	7357

of 0.1% for light jets and 1% for charm jets described in Section 4.2.7. Exactly two  $b$ -tagged jets with  $p_T > 25 \text{ GeV}$  and  $|\eta| < 2.4$  are required to be present in selected events, to remove contamination in the signal region from charm and light flavor jets. To reduce the contribution from  $pp \rightarrow t\bar{t}$  events, events with a third jet with  $p_T > 25 \text{ GeV}$  and  $|\eta| < 4.7$  are rejected. The effects of the various cuts with regard to the total number of events passing selections on the data is illustrated in Table 5.1.

## 5.2 Simulated samples

After all selection requirements detailed in Section 5.4 are applied, the contributing processes to the overall yield are the associated production of a massive vector boson and jets ( $V + \text{jets}$ ), as well as diboson ( $WW, WZ, ZZ$ ),  $t\bar{t}$ , single top quark,  $\gamma + \text{jets}$ , and QCD multijet production. The corresponding contributions are estimated from simulation, except for the QCD background, which is estimated from data as described

in Section 5.3.1.

Simulated samples of  $V + \text{jets}$ ,  $\gamma + \text{jets}$  and  $t\bar{t} + \text{jets}$  are generated at tree-level with MADGRAPH 5.1 [16, 17] using the CTEQ6L [12] Parton Distribution Function (PDF) set. These samples are interfaced with PYTHIA 6.4 [18] for hadronization using the Z2\* tune for the underlying event. The  $k_t$ -MLM [19, 20] matching scheme is used. For the signal distributions, the  $Wb\bar{b}$  component of an inclusive  $W + \text{jets}$  sample is used, with the shapes of the distributions taken from a dedicated high-statistics generated sample of exclusive  $Wb\bar{b}$ . The shape of the  $Wb\bar{b}$  signal distribution is obtained by separating the  $W + \text{jets}$  simulated sample into three subsamples labeled as  $Wb\bar{b}$ ,  $Wc\bar{c}$ , and  $Wusdcg$ . These three categories are described below. The separation is done at the particle list level. If an event contains a  $b$  jet, from the matrix element or parton shower, it falls into the  $Wb\bar{b}$  category. A  $b$  quark at generator level requires the presence of a  $b$  hadron within a cone of radius  $R = 0.4$  with respect to the jet axis. The jets are constructed using generator-level information using all stable particles in the event excluding neutrinos. Jets with a distance smaller than  $R = 0.5$  with respect to a lepton are removed from the event. If an event contains no  $b$  jets, but an even, non-zero, number of charm jets, again from the matrix element or parton shower, it falls into the  $Wc\bar{c}$  category. The remaining events fall into the  $Wusdcg$  category. The energy of the selected leptons at the generator level is corrected for final-state radiation by summing up the four-momenta of all the photons generated within a cone of radius  $R = 0.1$  around the lepton. Generated leptons which do not originate in simulation from the primary interaction vertex are not considered for selection.

Single top quark event samples are generated at NLO with POWHEG 2.0 [21, 22,



23, 24] using the CTEQ6M PDF set. Hadronization is performed using PYTHIA 6.4 with the Z2\* tune. Diboson samples are generated and hadronized with PYTHIA 6.4 at LO using the CTEQ6L PDF set and the Z2\* tune.

The cross sections for the  $V + \text{jets}$  and  $\gamma + \text{jets}$  processes are normalized using the predictions for inclusive W, Z and photon production from FEWZ 3.1 [25, 26] evaluated at NNLO. Single top quark and diboson production cross sections are normalized to the NLO cross section predictions from MCFM 7.0 [27, 28] using the MSTW08 NLO PDF set [29]. The  $t\bar{t}$  cross section used is  $241.5 \pm 8.5$  pb and was determined from data collected by the ATLAS and CMS experiments [30, 31, 32] at the LHC at  $\sqrt{s} = 8\text{TeV}$ .

Events induced by additional simultaneous  $pp$  interactions (pileup) as described in Section 4.2.1 are simulated using events generated with PYTHIA 6. During the 2012 data taking, the average pileup rate was 21 interactions per bunch crossing and the simulated number of pileup interactions have been reweighted to match this distribution in the data.

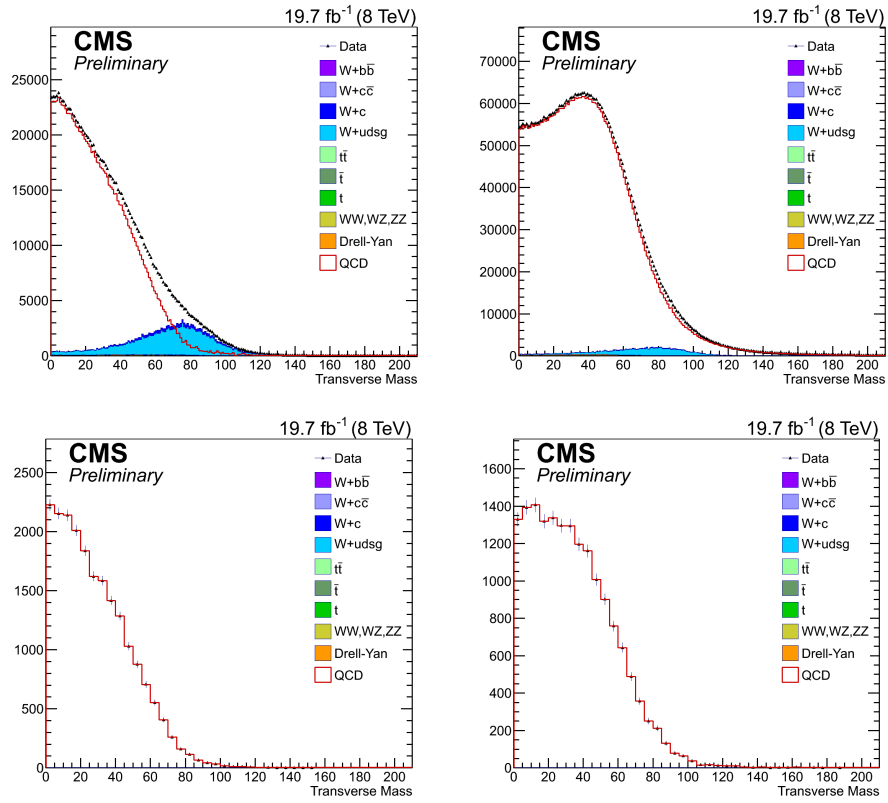
## 5.3 Background Estimation

### 5.3.1 QCD

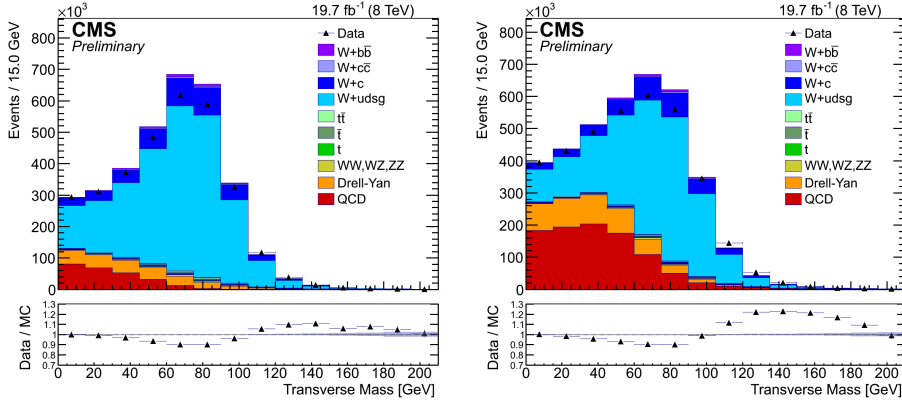
The QCD multijet sample is derived using a data-driven method. The shapes of the distributions for QCD multijet events are taken as the difference between the data sample and the sum of the other simulated backgrounds in a region of phase space enriched in multijets as illustrated in Figure 5.1. This region is found using the same

selection requirements as those in the signal region, but requiring the muon (electron) to be antiisolated:  $I > 0.20$  (0.15). In the fiducial regions used in this analysis, no correlation is observed between  $I$  and  $m_T$ , validating the use of an inverted isolation requirement to obtain the QCD background shape. This is not the case for the  $\Delta R$  distance between the two  $b$ -tagged jets,  $\Delta R(b, \bar{b})$ , or the lepton  $p_T$ . The shape of the QCD distribution for these variables is therefore taken from an  $m_T < 30$  GeV sideband and validated against QCD multijet simulation. The normalization of the QCD background in these variables is set to the final normalization resulting from the fit to the  $m_T$  variable which was derived using the inverted isolation requirement. To obtain an initial estimate of the number of QCD multijet events passing signal region selections, the shape derived in the antiisolated region, is put in the signal region and scaled by  $(d_{20} - m_{20})/q_{20}$  where  $d_{20}$  is the yield in data in the range  $0 < m_T < 20$ ,  $m_{20}$  is the combined yield from the simulated samples in this range, and  $q_{20}$  is the corresponding unnormalized yield of QCD multijet events. This has the effect of normalizing the QCD sample such that the combination of the QCD and the simulated backgrounds has the same total yield as data in the range  $0 < m_T < 20$ . If  $d_{20} < m_{20}$ , the QCD contribution is taken to be negligible. The relative uncertainty in the yield of QCD multijet events is estimated to be  $\pm 50\%$ , taking into account both the fit result and the extrapolation from  $0 < m_T < 20$  to the high- $m_T$  range. This relative uncertainty also covers shape mismodelings of the multijet contribution in the final sample.

**Figure 5.1:** The shape for the QCD is found by inverting the lepton isolation and subtracting MC from the data. Shown above is the data, MC background and extrapolated QCD shape (difference between data and MC backgrounds) for both the muon and electron channels in the  $W + jj$  and  $W + b\bar{b}$  phase spaces. The requirement of two well-identified  $b$  tags essentially eliminates all MC backgrounds in the  $W + b\bar{b}$  region, leaving the QCD shape the same as that of the data.



**Figure 5.2:** Selecting for a tight ID muon with  $p_T > 30$  GeV and exactly two central jets passing loose ID, we recover the  $m_T$  distributions in the muon channel (left) and the electron channel (right). The shaded band indicates the statistical uncertainty.

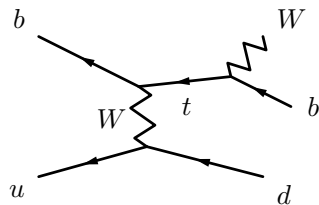


### 5.3.2 $W$ +jets: light and charm component

$W$ +light jets is the dominant background in the  $Wjj$  phase space, which is found using identical selections as are used in the signal region with the exception of the  $b$  tag requirement. In the  $W + jj$  phase space, no requirements on  $b$  tags are made. This control region therefore serves as a cross check on the reconstructed objects observed in the signal region before the added complication of  $b$ -tagging has been introduced. In Figure 5.2 is shown the  $m_T$  between the identified lepton and  $E_T^{\text{miss}}$  in both decay channels. Agreement between simulation and data is on the order of 10% – 20%.

The process  $pp \rightarrow Wb$  with a single  $b$  quark produced is CKM suppressed by two generations if light quarks are interacting. The process  $pp \rightarrow Wc$  is only suppressed by one generation but is still found to contribute negligible rate in the signal region

**Figure 5.3:** Below is the diagram for the process attempting to be isolated in the single top  $t$ -channel control region.

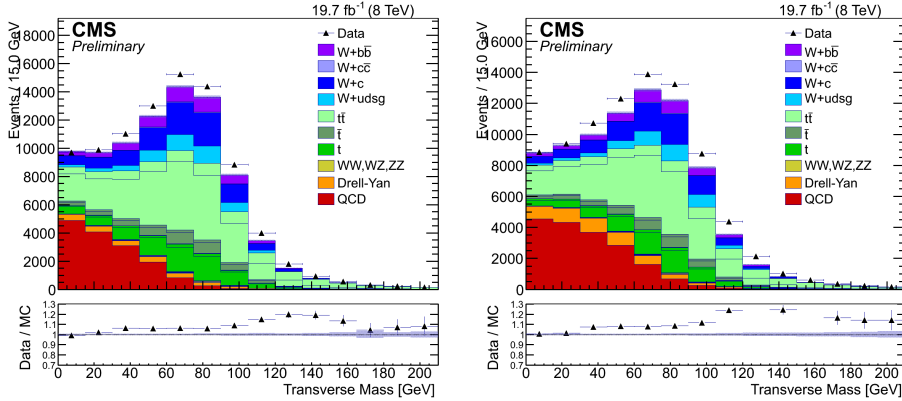


due to the requirement of a second  $b$  tag along with a veto on a third jet. A second jet can be added to  $ppwc$  via ISR or FSR, and light quarks require less energy to produce than heavy quarks, making them more probable. Therefore, the dominant contribution to  $ppwc$  in the signal region comes from the mistag of a  $c$  jet for a  $b$  jet, along with the misidentification of an ISR/FSR light jet. However, the contribution to events in the signal region from high energy  $g \rightarrow c\bar{c}$  is not negligible and moreover these events have kinematics closely related to that of the signal.

### 5.3.3 Single top backgrounds

The single top  $t$ -channel control region is defined by the signal selection requirements without the third and forward jet vetoes, and with the leading jet required to be central ( $|\eta| < 2.4$ ) and tightly  $b$ -tagged, while the subleading jet has no  $b$  requirement and must fall within  $2.4 < |\eta| < 5.0$ . The reasons for these selections are based on examination of Figure 5.3. The  $b$  quark is a product of the  $t$  decay and is recoiling against a  $W$ , making it a high- $p_T$  jet. The other jet comes from an initial state quark that exchanges a  $W$  and recoils against the  $t$ . Because it is the product of an initial

**Figure 5.4:** The single top control region is defined by one  $b$ -tagged central jet and one forward jet. Shown below are the  $m_T$  distributions in the single top control region in the muon channel (left) and electron channel (right). The shaded band indicates the statistical uncertainty.

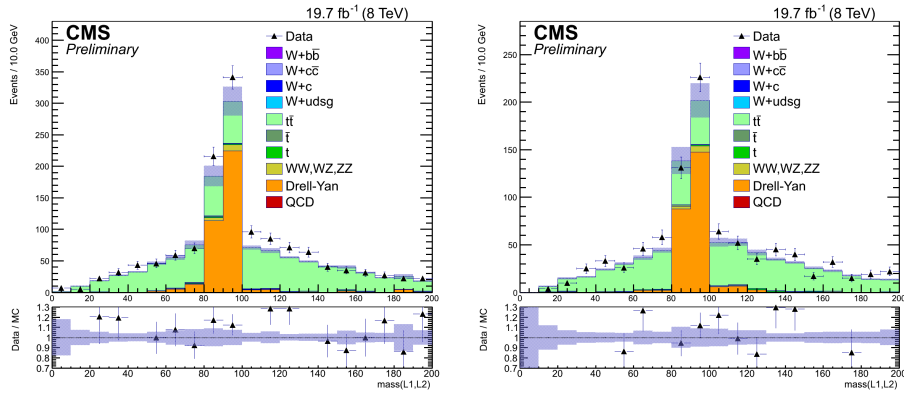


state quark which is most likely to be light, the final state jet is also most likely to be light. The fact that it is recoiling against the much more massive  $t$  means that it is likely to be forward by conservation of momentum. As can be seen in Figure 5.4, there are many backgrounds contaminating the purity of this phase space, but agreement between data and simulation is on the order of 5-10% up to 100 GeV.

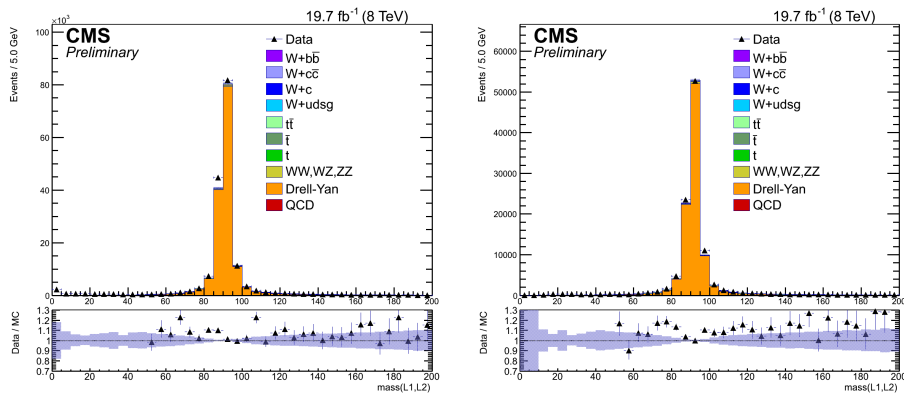
### 5.3.4 $Z\ell\bar{\ell}$ backgrounds

The Drell-Yan background is validated in a control region where the  $Wb\bar{b}$  selection requirements are applied, but the lepton veto is inverted, requiring two isolated, same-flavor leptons which are presumed to be the decay products of the  $Z$  boson. This is referred to as the  $Zb\bar{b}$  region and distributions of the mass of the dilepton pair is shown in Figure 5.5. Contamination from  $t\bar{t}$  is evident. A cleaner Drell-Yan phase

**Figure 5.5:** Below are distributions of the invariant mass of the dilepton pair in the  $Zb\bar{b}$  phase space. The left plot shows the dimuon decay channel and the right plot shows the dielectron decay channel.



**Figure 5.6:** Below are distributions of the invariant mass of the dilepton pair in the  $Zjj$  phase space. The left plot shows the dimuon decay channel and the right plot shows the dielectron decay channel.



space is found by requiring exactly two jets but placing no  $b$  tag requirement and is referred to as  $Zjj$ . Figure 5.6 shows the same distributions as Figure 5.5 in this phase space.

## 5.4 Analysis Strategy

The  $Wb\bar{b}$  yield is measured using a likelihood fit to the  $m_T$  distribution in the signal region. With the exception of QCD processes, the initial shapes and normalizations of all contributions in the fit are taken from simulation. Consequently, it is important to verify that the simulation describes the data. The dominant background in the signal region arises from the  $t\bar{t}$  process. The data and simulation are thus compared in two  $t\bar{t}$ -dominated control regions: one characterized by a pair of opposite flavor leptons, and the other by the presence of three or more jets. The simulation is reweighted to describe the data in the control regions and then is used to predict the  $m_T$  distributions in the signal region.

The signal region requires a muon (electron) with  $p_T > 30$  GeV,  $|\eta| < 2.1$ , and satisfying  $I < 0.12$  (0.10). Exactly two  $b$ -tagged jets with  $p_T > 25$  GeV and  $|\eta| < 2.4$  are also required. Events with additional leptons with  $p_T > 10$  GeV and  $|\eta| < 2.4$  or a third jet with  $p_T > 25$  GeV and  $|\eta| < 4.7$  are rejected. The  $t\bar{t}$ -multijet control region is obtained using the same selection criteria as for the signal region, but requiring at least three jets in the event with  $p_T > 25$  GeV and  $|\eta| < 2.4$  instead of vetoing events which have more than two jets. These are to capture events in which one of the  $W$  bosons from the  $t$  decay decays leptonically and the other decays hadronically, producing jets. The  $t\bar{t}$ -multilepton control region uses similar selection criteria as the signal region, but changing the lepton requirement from vetoing events which contain a second lepton, to requiring two isolated leptons of different flavor, both with  $p_T > 30$  GeV and  $|\eta| < 2.1$ . This selection is to capture  $t\bar{t}$  events where both  $W$  bosons decay leptonically, while avoiding same-flavor dileptons which are characteristic of  $Z$  boson



decay. In the  $t\bar{t}$ -multilepton region, the  $m_T$  variable is calculated with respect to the electron in the electron channel and the muon in the muon channel.

The normalizations and shapes of the simulated backgrounds are allowed to vary in the fit within the uncertainties listed in Table 5.2. The uncorrelated normalization uncertainties are uncertainties on the cross section of the given sample.

Two major parameters in the simulations significantly affect the normalization of the simulated distributions: the  $b$ -tagging efficiency and the jet energy scale (JES). Both control and the signal regions show similar sensitivity to the  $b$ -tagging efficiency, and its adjustment affects all the regions in a correlated manner. Because the  $t\bar{t}$  production has more than two jets in the final state, the rejection of events with a third jet makes it sensitive to JES. The effect on the leading jets is moderate but JES variations lead to significant migration of jets into and out of the veto region. The  $t\bar{t}$ -multijet control region, since it has no veto on a third jet, is less sensitive to JES variations than the  $t\bar{t}$ -multilepton control region. The same is true for the  $Wb\bar{b}$  process, which at LO has only two jets in the final state from the ISR gluon splitting as illustrated in Figure 2.1 and therefore is not affected by the jet veto requirement.

The fit procedure consists of three consecutive steps where the simulated distributions in two control regions and the signal region are fit to data using the  $m_T$  variable, which is chosen because it has a well known shape for  $W + \text{jets}$  production that allows for reliable signal extraction. First, the fit is performed in the  $t\bar{t}$ -multijet region. In this control region, the uncertainty is dominated by the uncertainty in the  $b$ -tagging efficiency so the fit results in a correction of the  $b$ -tagging efficiency that is measured separately in the muon and electron channels and then combined. The

simulation is corrected by 14% using this result and the corrected simulated samples are fit to the data in the  $t\bar{t}$ -multilepton region. The result of the second step is used to adjust JES by 3.4% and as a result of this procedure, the simulation is expected to better describe the  $t\bar{t}$  contribution. The final step is to extract the number of  $Wb\bar{b}$  events from the fit to  $m_T$  in the signal region.

**Table 5.2:** Breakdown of the main sources of systematic uncertainty in the  $Wb\bar{b}$  signal region. The column labeled “Variation” indicates the bounds on the normalization change of a given sample due to a variation of the uncertainty by one standard deviation. The last column indicates the contribution of the given systematic to the overall uncertainty in the measured cross section. The uncertainty labeled “b-tag eff rescaling” is the uncertainty associated with the rescaling of the b-tagging efficiency. UES refers to the scale of energy deposits not clustered into jets, and MES and EES refer to the muon and electron energy scales. The uncertainty labeled as “Id/Iso/Trg” is the uncertainty associated with the efficiency of the lepton identification, isolation, and trigger. The uncertainties in the integrated luminosity [33] and on the acceptance due to PDF uncertainties and scale choices are not included in the fit, and are treated separately.

		Uncertainty	Variation	Effect on the measured cross section
Normalization	Uncorrelated	$t\bar{t}$	3.5%	3.8%
		Single top	5.4%	2.5%
		$Wusdcg$	13.2%	< 2%
		$Wc\bar{c}$	13.2%	< 2%
		Diboson	8.1%	< 2%
		Drell–Yan	7.9%	< 2%
		$\gamma$ +jets	10.0%	< 2%
		QCD	50%	2-3%
Shape	Correlated	b-tag eff rescaling	8.4%	9.2%
		JES rescaling	0-6%	3.8%
		MES	0-3%	< 2%
		UES	0-3%	< 2%
		EES	0-3%	< 2%
		Id/Iso/Trg	0-4%	< 2%
Luminosity			2.6%	
Scales ( $\mu_R, \mu_F$ )			10%	
PDF choice			1%	

## 5.5 Systematic uncertainties

The main sources of the systematic uncertainties are listed in Table 5.2. The size of the variation is shown for each uncertainty source, together with its effect on the measured cross section and is included as a nuisance parameter in the fit. Some of the uncertainties affect only the normalization in the respective contributions. These include the uncertainties on the theoretical cross section of a given sample, which are uncorrelated between samples and are included as log-normal constraints on the rate. For any sample, the effect on the final normalization from uncertainties which only effect normalization is

$$N_f = N_i \prod a_j^{x_j} \quad (5.1)$$

where  $a$  is the input bound given in the “Variation” column of Table 5.2 and  $x$  is the parameter being fit for the  $j$  nuisances. The uncertainty due to the b-tagging efficiency and the uncertainty due to the JES are allowed in principle to have shape dependencies in this analysis, but only affect the normalizations of the samples in the  $m_T$  variable in practice.

The uncertainties which affect both the normalization and the shape of the  $m_T$  distributions are those listed in the table under “Shape” and are incorporated into the fit via binned templates. These templates are obtained by varying the source of the given uncertainty and reprocessing the simulated sample. Taking  $h_0$  as the unshifted template and  $h_j^\pm$  as the up and down templates obtained through reprocessing after

shifting nuisance  $j$ , the final fit shape  $h$  is given by

$$h = h_0 + \sum_j [A(x)h_j^- + B(x)h_0 + C(x)h_j^+] \quad (5.2)$$

$$A(x) = x(x+1)/2, \quad B(x) = -x^2, \quad C(x) = x(x-1)/2 \quad (5.3)$$

The 50% uncertainty in the QCD background is taken as a conservative estimate and ultimately has a 2-3% contribution to the uncertainty on the measured cross section. The b-tagging efficiency and JES rescaling uncertainties are taken from their respective fits. The renormalization and factorization scale uncertainties are estimated by simultaneously changing the renormalization and factorization scales,  $\mu_R$  and  $\mu_F$ , up and down by a factor of two to account for the scale choice of the couplings and the energy scale at which jets are modeled as coming from the ME or from the shower. The PDF uncertainties are estimated from the change in acceptance found by varying the PDF set following the LHAPDF/PDF4LHC prescription [34, 35, 36, 37] considering PDF sets from CTEQ, MSTW, NNPDF, and HERA Collaborations.

## 5.6 Signal Extraction

The fit in the  $t\bar{t}$ -multijet region is used to obtain b-tag rescaling factors separately for the muon and electron channels in order to better describe the b-tagging efficiency in the simulation as described in Section 5.4. The results of the fit are presented the two plots at the top of Fig. 5.7. The central values of the b-tag rescaling factors,  $1.12 \pm 0.08$  (muon channel) and  $1.16 \pm 0.08$  (electron channel), are combined to  $1.14 \pm 0.08$ . The

simulation is reweighted accordingly for the next fit and the uncertainty on this fit sets the one standard deviation bound on the b-tag efficiency rescale factor in subsequent fits.

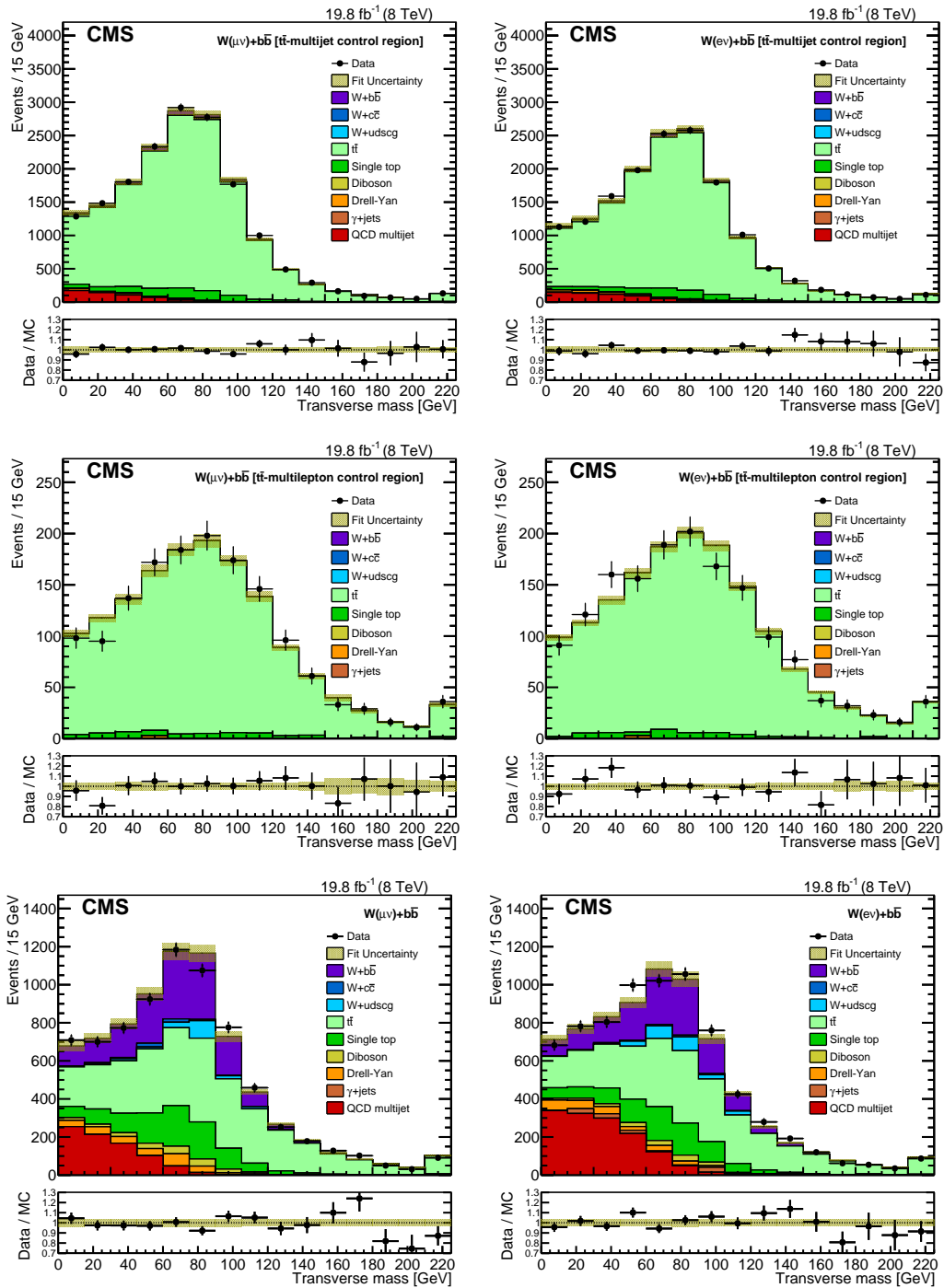
A fit to the  $t\bar{t}$ -multilepton region adjusts the JES, as described in Section 5.4. As a result, the simulated  $m_T$  distributions change normalization. The best fit suggests changing the normalization by approximately 3.4% from its central value which corresponds to 1.3 standard deviations in JES. The middle plots in Fig. 5.7 show the results of the fits in the  $t\bar{t}$ -multilepton event enhanced control region for the muon (left) and the electron (right) channels. The JES is therefore shifted by 1.3 standard deviations in the simulation with the uncertainty taken from the fit. Thus the simulation is tuned to describe the  $t\bar{t}$  control regions and is used to extract the signal yield in the signal region.

The results of the fit in the  $Wb\bar{b}$  signal region are shown in the bottom of Fig. 5.7. All background contributions are allowed to vary in the fit within their uncertainties, while the  $Wb\bar{b}$  normalization remains a free parameter of the fit. The correlations across all simulated samples are taken into account as shown in Table 5.2. The composition of the event sample in the signal region is summarized in Table 5.3. Events coming from the production of a Higgs boson in association with a vector boson constitute a negligible fraction of the overall event yield and are not considered. Distributions for variables other than those being directly fit are also produced by applying the results from the three fits to the simulated samples. Distributions of  $\Delta R(b\bar{b})$  and  $p_T^\ell$  combining both lepton flavors are presented in Fig. 5.8.

**Table 5.3:** Initial and final yields obtained in the  $Wb\bar{b}$  signal region. The uncertainties in the signal strength represent the total uncertainty of the fit.

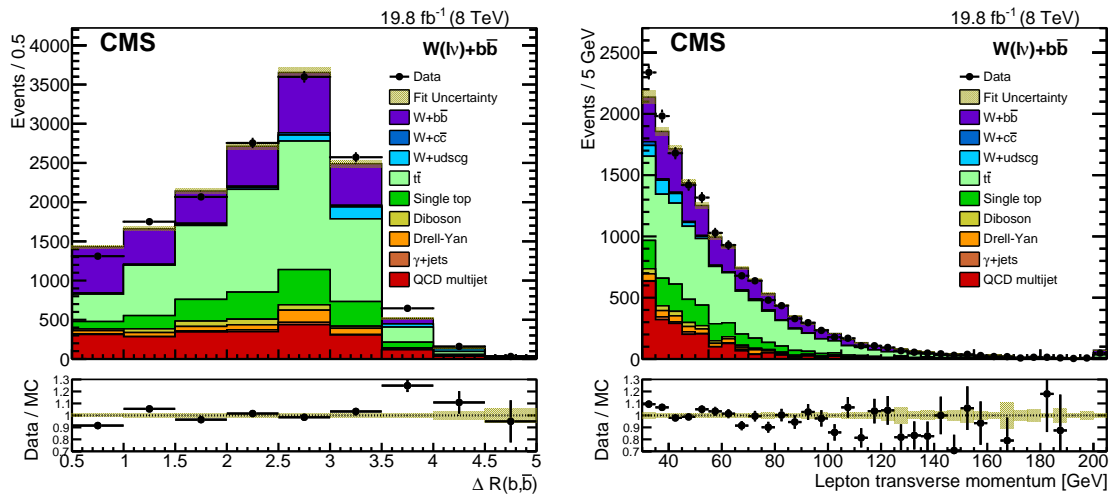
	Muon		Electron	
	Initial	Fitted	Initial	Fitted
Data	7432		7357	
$Wb\bar{b}$	1323	1712	1121	1456
$Wc\bar{c}$	60	61	36	37
$Wusdcg$	182	179	220	217
$t\bar{t}$	3049	3296	2640	2864
Single top	958	1008	820	865
Drell-Yan	261	265	220	224
Diboson	175	181	139	144
$\gamma$ +jets	-	-	98	105
QCD	1109	803	1654	1373
Total MC	7116	7505	6948	7284
Signal strength	$1.21 \pm 0.19$		$1.37 \pm 0.23$	
Combined	$1.26 \pm 0.17$			

**Figure 5.7:** The transverse mass distributions (upper) in the  $t\bar{t}$ -multijet phase space after fitting to obtain the  $b$ -tag efficiency rescale factors, (middle) in the  $t\bar{t}$ -multilepton event enhanced control region after fitting to find the appropriate JES, and (lower) in the  $Wb\bar{b}$  signal region after fitting simultaneously muon and electron decay channels. The lepton channels are shown separately with the muon sample on the left and the electron sample on the right. The last bin contains overflow events. The shaded area represents the total uncertainty in the simulation after the fit.





**Figure 5.8:** Distributions of  $\Delta R(b\bar{b})$  and  $p_T^\ell$  after applying the results from the fits to the simulation. The QCD background shape is taken from an  $m_T < 30$  GeV sideband and the muon and electron channels have been combined in these distributions. The last bin contains overflow events and the shaded area represents the total uncertainty in the simulation after the fit.



## 5.7 Cross Section and Comparisons

The cross section for the process  $\sigma(pp \rightarrow Wb\bar{b} \rightarrow \ell\nu b\bar{b})$ , is derived from the signal strength measurement as obtained from the fit. The cross section is written as

$$\sigma(pp \rightarrow Wb\bar{b} \rightarrow \ell\nu b\bar{b}) = \frac{N_{\text{signal}}^{\text{Data}}}{A \cdot \epsilon \cdot \mathcal{L}} = \frac{N_{\text{signal}}^{\text{Data}}}{(N_{\text{signal}}^{\text{MC}}/N_{\text{generated}}^{\text{MC}}) \cdot \mathcal{L}} = \alpha\sigma_{\text{gen}}$$

where  $N_{\text{signal}}^{\text{Data}}$  is the number of observed signal events,  $N_{\text{signal}}^{\text{MC}}$  is the number of expected signal events from simulation,  $N_{\text{generated}}^{\text{MC}}$  is the number of generated events in the fiducial region,  $A$ ,  $\epsilon$  are the acceptance and efficiency correction factors,  $\alpha$  is the measured signal strength in the given lepton channel, and  $\sigma_{\text{gen}}$  is the simulated fiducial cross section of the signal sample.

In this analysis, the fiducial cross section was calculated in the following manner: MADGRAPH is used to compute the  $Wb\bar{b}$  cross section with fiducial cuts applied. Then a k-factor for inclusive W production is applied, obtained from the ratio of the inclusive W cross sections calculated with FEWZ at NNLO using the 5F CTEQ 6M PDF set with the cross section calculated using MADGRAPH. The product  $A \cdot \epsilon$  is 13 (11)% in the muon (electron) channels and results from the combined effect of the efficiency from lepton identification requirements (80%), and  $b$  tag efficiency (40% per jet). The uncertainty on this product is 10% as listed in the bottom row of Table 5.2, which was calculated by varying the PDF set using the LHAPDF/PDF4LHC [38, 35, 36, 37] prescription considering PDF sets from CTEQ, MSTW, NNPDF, and HERA as well as varying the choice of scales  $\mu_{\text{F}}$ ,  $\mu_{\text{R}}$  simultaneously up and down by a factor of two. Varying the PDFs and the choice of scale is a way to estimate the

dependence of the measured cross section on the choices of these parameters.

The  $Wb\bar{b}$  cross section is measured within a fiducial volume, which is defined by requiring leptons with  $p_T > 30\text{GeV}$  and  $|\eta| < 2.1$  and exactly two  $b$ -tagged jets of  $p_T > 25\text{GeV}$  and  $|\eta| < 2.4$ . The measured cross sections are presented in Table 5.4. The combination of the muon and electron measurements is done using a simultaneous fit to both channels, taking into account correlations between different sources of uncertainties.

**Table 5.4:** Measured cross sections in the muon, electron, and combined lepton channels.

Channel	$\sigma(pp \rightarrow Wb\bar{b} \rightarrow \ell\nu b\bar{b})$ pb
Combined	$0.64 \pm 0.03(\text{stat}) \pm 0.10(\text{syst}) \pm 0.06(\text{theo}) \pm 0.02(\text{lumi})$
Muon	$0.62 \pm 0.04(\text{stat}) \pm 0.11(\text{syst}) \pm 0.06(\text{theo}) \pm 0.02(\text{lumi})$
Electron	$0.70 \pm 0.05(\text{stat}) \pm 0.15(\text{syst}) \pm 0.07(\text{theo}) \pm 0.02(\text{lumi})$

The measured cross sections are compared to theoretical predictions from MCFM [27, 28] with the MSTW 2008 PDF set, as well as from MADGRAPH 5 interfaced with PYTHIA6 in the four- and five-flavor schemes and MADGRAPH 5 with PYTHIA8 [39] in the four-flavor scheme. In the five-flavor scheme, the PDF set CTEQ 6L was used and PYTHIA6 was run using TuneZ2\*. The two four-flavor samples were produced using a NNLO PDF set interfaced with PYTHIA (version 6 in one sample, version 8 in the other) in the CUETP8M1 tune.

Comparisons between the results of calculations performed under different assumptions provide important feedback on the functioning and validity of the techniques employed. Differences in predictions arising from the modelling of  $b$  quarks as massive or massless are possible, as are variations in predictions arising from the use of

different showering packages (PYTHIA6 vs. PYTHIA8) or matrix element generators (MADGRAPH vs. MCFM). In the phase space explored here, these predictions are all very close in their central value and agree with each other well within their respective uncertainties.

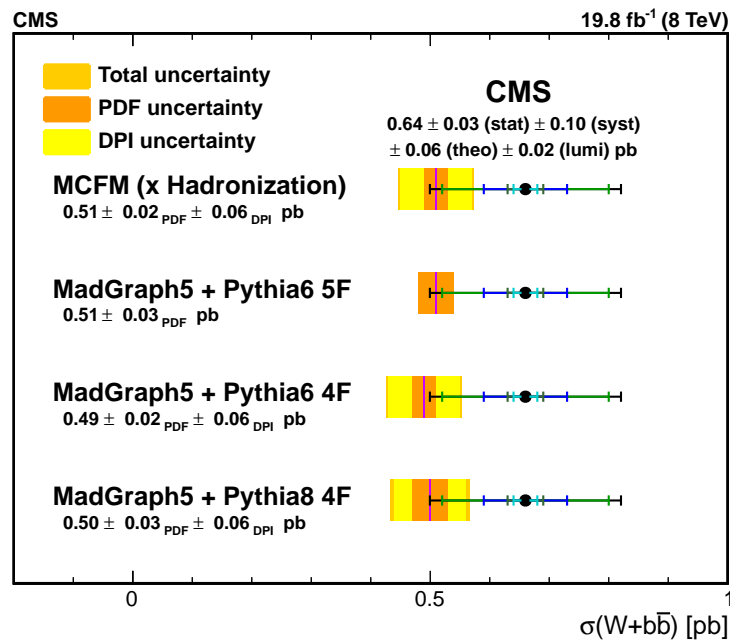
The MCFM cross section calculation is performed at the level of parton jets and thus requires a hadronization correction. The multiplicative hadronization correction factor  $0.81 \pm 0.07$  is calculated using the MADGRAPH + PYTHIA6 sample and agrees well with a similar factor calculated in the 7TeV Z+b analysis calculated as  $0.84 \pm 0.03$  [40]. This factor is determined by taking the ratio of the predicted cross sections in a given sample after having undergone a simulation of the hadronization process, and using jets identified at the particle list level. The correction factor is obtained for jets computed excluding neutrinos from the particle list, as particles such jets are closer in kinematics to particle jets at the detector level. The uncertainty reflects both the statistics of the MADGRAPH + PYTHIA6 sample as well as a comparison with the MADGRAPH + PYTHIA8 sample.

The MCFM and four-flavor MADGRAPH predictions do not account for  $Wb\bar{b}$  production where the  $b\bar{b}$  system comes from multiple parton scattering. CMS simulations of MADGRAPH + PYTHIA events that include double parton interactions (DPIs) reproduce the  $W + \text{jets}$  data [41], therefore a MADGRAPH + PYTHIA8 sample of a  $W$  boson produced in association with a  $b\bar{b}$  pair coming from DPS was generated to study the effect on the fiducial cross section. Using this dedicated sample, an additive correction  $\sigma_{\text{DPS}}$  is estimated to be  $0.06 \pm 0.06$  pb, where the uncertainty is conservatively assigned to be 100% of the value.

The uncertainty in the theoretical cross sections arising from the choice of PDF is also accounted for, using the LHAPDF/PDF4LHC [38, 35, 36, 37] prescription in which PDF sets from CTEQ, MSTW, NNPDF, and HERA are considered. Uncertainties in the theoretical cross section due to the choice of scale are also estimated by varying the scales  $\mu_F$ ,  $\mu_R$  simultaneously up and down by a factor of two.

The resulting cross section predictions in the fiducial phase space at the hadron level and including the estimated hadronization and DPS corrections when needed are compared in Fig. 5.9 with the measured value. Within one standard deviation the predictions agree with the measured cross section. The results also agree within one standard deviation with previously published  $Wb\bar{b}$  measurements at 7 TeV, where data are found to be well described by the same predictions.

**Figure 5.9:** Comparison between the measured  $pp \rightarrow Wb\bar{b} \rightarrow \ell\nu b\bar{b}$  cross section and various QCD predictions. The orange band indicates the uncertainty in the given sample associated with PDF choice and the yellow band represents the uncertainty associated with DPI. The labels 4F and 5F refer to the four- and five-flavor PDF schemes, and in the case of the MADGRAPH + PYTHIA 6 (5F) sample, the effects of DPI are already included in the generated samples so the DPI correction is not needed. The measured cross section is also shown with associated uncertainties.



# Chapter 6

## MONOPHOTON ANALYSIS

The second analysis presented in this thesis is of the monophoton final state described in Section 2.2. The data are analyzed in the context of the SM process  $pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma$  in which the  $E_T^{\text{miss}}$  is interpreted as coming from the invisible decay of the  $Z$  boson,  $Z \rightarrow \nu\bar{\nu}$ . The data are also analyzed as a dark matter search under the interpretation that the  $E_T^{\text{miss}}$  arises from the annihilation of incoming particles into DM and the  $\gamma$  is initial state radiation recoiling against the process. Under this interpretation, limits are set on the cross section of DM as a function of the mediator mass for vector and axial-vector models. This analysis is performed using protons colliding at  $\sqrt{s} = 13$  TeV provided by the LHC and detected by the CMS detector.

### 6.1 Event Selection

The sample of data analyzed was collected using a trigger that requires at least one photon HLT SC candidate with  $p_T^\gamma > 165\text{GeV}$  as described in Section 3.2.7.2. Because photons interact electromagnetically, they are expected to deposit all of their energy in the ECAL, while jets typically have a neutral component and deposit some energy in the HCAL as well. To increase photon efficiency, the trigger therefore also requires at least 90% of the energy deposited in the calorimeters to be deposited in the ECAL. This trigger is 98% efficient at selecting photons which pass the other

analysis selections. Events passing the trigger are further required to have at least one PF photon with  $p_T^\gamma > 175\text{GeV}$  in the barrel fiducial region ( $|\eta| < 1.44$ ).

To distinguish photons from electrons, which leave a similar signature of energy deposits in the ECAL and HCAL, candidate photons are required to not have any associated track seeds in the pixel detector. To distinguish photons from jets, selections based on calorimetric information and isolation are applied. The fraction of energy deposited in the ECAL compared to the total deposit in the calorimeters is tightened relative to the trigger to be 95% and the shower shape variable describing the spread of the energy deposits in the  $\eta$  direction,  $\sigma_{i\eta i\eta}$ , is required to be  $\sigma_{i\eta i\eta} < 0.0102$ . Additionally, the photon is required to pass the isolation requirements described in Section 4.2.5.

Because photon objects are not reconstructed from tracks, there is an ambiguity in identifying the collision vertex that the photon originates from in the presence of pileup collisions. Association of a vertex to the photon candidate impacts the photon in two ways. First, the photon momentum direction is defined by the straight line which connects the ECAL cluster position and the identified vertex. Additionally, the isolation sum uses only the PF charged hadrons having tracks associated to the vertex. While the first effect is minor and is not relevant for this analysis, the second will cause photon candidates that are actually not isolated to appear isolated, if the vertex is misassigned. In practice, photon momentum is always computed with respect to the PV, but for the charged hadron isolation sum, all vertices are considered, and the maximum value of the isolation sum is used as a conservative estimate of the true isolation sum.



To reduce the contribution of backgrounds arising from occurrences in the CMS detector which did not originate from collisions, the energy pulse which seeded the photon cluster described in Section 4.2.5 is required to be within  $\pm 3\text{ns}$  of the time expected for particles from a collision, and the cluster must not be so narrow that it is consistent with a cluster formed by a single crystal. To reduce contamination from beam halo, the ECAL crystals not associated with the photon candidate are examined for evidence of the passage of a minimum-ionizing particle (MIP) roughly parallel to the beam axis (beam halo tag). If at least 4.9 GeV of energy is found deposited along this trajectory, the event is rejected. This value was determined by through optimizing the signal to background photon efficiencies, with a 95% identification for prompt photons and a 20% misidentification rate for deposits originating in beam halo events.

The candidate events are required to have  $E_T^{\text{miss}} > 170$  GeV. The azimuthal opening angle between the candidate photon and  $E_T^{\text{miss}}$  is required to be greater than 2 radians to ensure that the main source of  $E_T^{\text{miss}}$  is not photon energy mismeasurement.

Because jet energy mismeasurement can also give rise to  $E_T^{\text{miss}}$ , events are rejected if the minimum azimuthal opening angle between  $E_T^{\text{miss}}$  and up to four leading jets ( $\min\Delta\phi(E_T^{\text{miss}}, j)$ ) is less than 0.5 radians. As was the case in the  $Wb\bar{b}$  analysis, jets are reconstructed using the PF algorithm, but in this analysis the jet clustering cone size is  $\Delta R < 0.4$  radians as opposed to 0.5 radians.

Finally, events are also vetoed if they contain a charged lepton (an electron or a muon) with  $p_T > 10$  GeV that is separated from the photon by  $\Delta R > 0.5$  radians.

The effects of the various cuts with regard to the total number of events passing

**Table 6.1:** Listed below are the raw number of events in data passing the selection listed in the first column as well as all selections in higher rows.

Selection	Events passing selection
Photon kinematics/trigger/ID	109384
$E_T^{\text{miss}} > 170 \text{ GeV}$	304
$\Delta\phi(E_T^{\text{miss}}, \gamma) > 2 \text{ radians}$	183
Veto charged lepton	149
$\Delta\phi(E_T^{\text{miss}}, j) > 0.5 \text{ radians}$	77

selections on the data is illustrated in Table 6.1. After applying all of the selection criteria, 77 candidate events are found in data.

## 6.2 Estimation of Background Contributions

The dominant SM processes contributing to this phase space of the candidate events are the associated productions of a  $Z$  or  $W$  boson with a high-energy photon. If the  $Z$  boson decays into a neutrino-antineutrino pair, the final state exhibits a high- $E_T$  photon and large missing transverse energy. Similarly, if the  $W$  boson decays into a lepton-neutrino pair and the lepton is outside of the detector acceptance or fails reconstruction, the event appears to be  $\gamma + E_T^{\text{miss}}$ . Together, these two processes account for approximately 75% of the events as estimated using MC simulations. Hard-scattering events are generated with MADGRAPH5\_AMC@NLO version 2.2 [42] at LO in QCD, with NNPDF3.0 as the PDF. Parton shower and hadronization is performed by PYTHIA 8.2 [43]. Generated particles are processed through the full GEANT-based simulation of the CMS detector [44, 45] and event reconstruction used

for data. Minimum-bias simulations are overlaid to model pileup interactions.

### 6.2.1 Reweighting

To account for differences arising from imperfect modeling of the data in the simulation, a total correction factor  $\rho = 0.99 \pm 0.06$  is applied to all MC-based estimates. This is the product of individual correction factors which are each taken as the ratio of the efficiency measured in data and in simulation. The efficiency for photon identification is measured and provided centrally using  $Z \rightarrow e\bar{e}$  events as  $0.99 \pm 0.016$  for photon identification measured using  $Z \rightarrow e\bar{e}$  events. The photon seed trigger efficiency is also measured using  $Z \rightarrow e\bar{e}$  events, and is found to be  $1.00 \pm 0.0246$  using jet triggers as a reference. The efficiencies for worst isolation, beam halo tag and lepton veto were measured using events in data which were triggered as having at least one muon and on MC using a combination of Drell-Yan,  $t\bar{t}$  and  $VV$  samples requiring  $Z\gamma \rightarrow \mu\bar{\mu}\gamma$  events to be identified by the dimuon pair with mass in the range  $61 < m_{\mu\mu} < 121$  GeV. The measured relative efficiencies are  $1.00 \pm 0.05$ .

Generated samples are weighted on an event by event basis with a product of two factors. The first factor matches the distribution of the generator-level photon  $p_T$  to that calculated at NNLO in QCD using the DYRES [46] calculator and the second factor, taken from Refs. [47, 48], further corrects this distribution to account for electroweak NLO effects.

### 6.2.2 $V\gamma$ Estimates

After accounting for the event selection efficiency difference between data and MC, respectively  $42.1 \pm 6.3$  and  $10.7 \pm 1.5$  events are estimated from  $Z\gamma \rightarrow \nu\bar{\nu}\gamma$  and  $W\gamma \rightarrow \ell\nu\gamma$ . Four sources of systematic uncertainty on  $Z\gamma$  and  $W\gamma$  estimates are considered: PDF and scale uncertainties are found using the the LHAPDF/PDF4LHC recommendations of varying the scales up and down by a factor of 2 and using Pdf sets from CTEQ, MSTW and NNPDF to be 5.37% and 8.9% respectively. Electroweak correction uncertainties are estimated conservatively as the quoted uncertainties [47, 48] which are 11% for  $Z\gamma$  and 7% for  $W\gamma$ . Scale factors are estimated to have an uncertainty of 6% which mostly arises from the statistical limitations of the data samples used. The systematic uncertainty due to  $\text{jet}/E_T^{\text{miss}}/\gamma$  energy scale and pileup, is estimated at 6.2% by shifting the energy of the respective PF object and observing the relative change in the number of events passing selections. As a crosscheck, the total contribution from  $Z\gamma \rightarrow \nu\bar{\nu}\gamma$  is estimated in data using a sample of  $Z\gamma \rightarrow \ell\bar{\ell}\gamma$  candidates, where the leptons from the decay of the Z boson are removed and considered as  $E_T^{\text{miss}}$ . This provides an estimate of  $64.6 \pm 17.6$ , where the uncertainty is dominated by the size of the sample.

### 6.2.3 Electron Mis-ID

The most important SM background comes from events where electrons are misidentified as photons, mainly in the  $W \rightarrow e\nu$  process. Seeding efficiency in the pixel detector for electron tracks is  $\epsilon = 0.982 \pm 0.004$  for electrons with  $p_T > 100\text{GeV}$ . This efficiency is measured in data using the tag-and-probe method [49] on  $Z \rightarrow e\bar{e}$  events,

and is verified with MC simulation. Electrons from  $W$  boson decay that are not seeded appear as isolated photons accompanied with large  $E_T^{\text{miss}}$  from the escaping neutrino. This class of events is modeled by an electron proxy event sample selected in data using criteria that are identical to those described in Sec. 6.1, except the photon candidate is required to have a pixel seed. The number of electron proxy events is then scaled by  $(1 - \epsilon)/\epsilon$  where  $\epsilon$  is the efficiency to yield an estimated contribution of  $7.4 \pm 1.2$  from electron misidentification events. The dominant uncertainty in the estimate is the statistical uncertainty in the tag-and-probe fit, and is assessed by generating a large ensemble of toy dielectron mass distributions on which the fit procedure is repeated. The standard deviation of the number of  $Z \rightarrow e\bar{e}$  events obtained from the fits is then propagated to the uncertainty in the efficiency.

### 6.2.4 Non-collision Backgrounds

Non-collision backgrounds, from things such as detector noise, cosmic rays, and beam halo, are estimated from the time distribution of the cluster seeds since each process exhibits a distinctive time distribution when the cluster is in the ECAL barrel. Templates for anomalous signals, cosmic ray muons, and beam halo events are obtained by inverting the shower shape and beam halo tag requirements, and are fitted to the timing distribution of the candidate sample. The only nonnegligible residual contribution to the candidate sample is found by fitting the template shapes to data and arises from the beam halo, with an estimated  $5.9 \pm 4.7$  events.

### 6.2.5 Minor SM Processes

The SM processes  $W \rightarrow l\nu\gamma$ ,  $Z \rightarrow \ell\bar{\ell}\gamma$ ,  $W(l\nu)$  and  $\gamma + jets$  are generated with MADGRAPH\_aMC@NLO at LO [50] with up to 2 jets and then processed with PYTHIA 6.426 generator [51] for showering and hadronization, with the NNPDF3.0 LO PDF. The total background expectation from these processes is  $3.05 \pm 0.67$  events, where the uncertainty includes the statistical and systematic uncertainty due to scale factor and  $\text{jet}/E_T^{\text{miss}}/\gamma$  energy scale.

## 6.3 Results

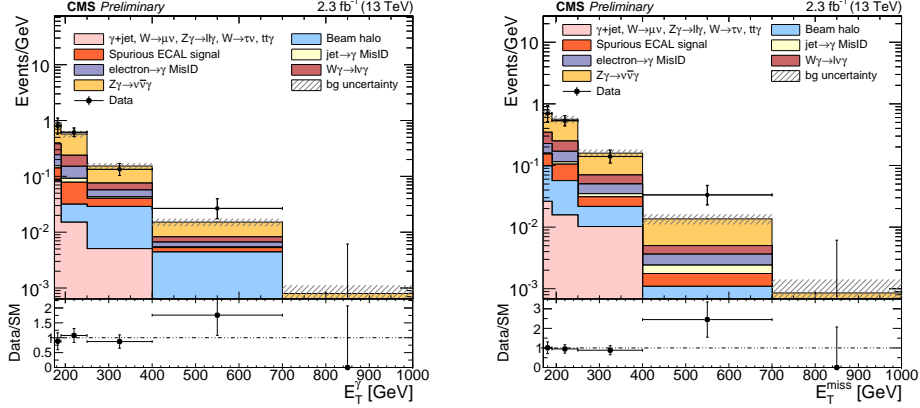
After applying the full selection criteria, 77 events in  $2.32 \text{ fb}^{-1}$  of data remain. Table 6.2 shows the estimated number of events and uncertainty from each background for the full 2015 run. The  $p_T$  spectrum and PF  $E_T^{\text{miss}}$  of the full combination of selected candidate events and estimated backgrounds can be seen in Figure 6.1 along with distributions of  $p_T/E_T^{\text{miss}}$  and the number of jets.

### 6.3.1 $pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma$ Cross Section Measurement

The  $pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma$  cross section for  $p_T^\gamma > 175 \text{ GeV}$  in the range  $|\eta| < 1.4$  is calculated using the formula

$$\sigma(pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma) = \frac{N_{data} - N_{BG}}{A \times \epsilon \times L} \quad (6.1)$$

**Figure 6.1:** The photon  $p_T$  and  $E_T^{\text{miss}}$  distribution for the candidate sample, compared with estimated contributions from SM backgrounds are shown. Here  $QCD\gamma$  refers to  $\gamma$ +jet background and the background uncertainty includes statistical and systematic error.



**Table 6.2:** Summary of estimated backgrounds and observed total number of candidates for  $2.32 \text{ fb}^{-1}$  of 2015 data. The category Others includes  $W \rightarrow \mu\nu$ ,  $Z \rightarrow \ell\ell\gamma$  and  $t\bar{t}\gamma$

Process	Estimate
$Z\gamma \rightarrow \nu\bar{\nu}\gamma$	$42.10 \pm 6.31$
$W\gamma \rightarrow l\nu\gamma$	$10.69 \pm 1.49$
$W \rightarrow e\nu$	$7.80 \pm 1.78$
$jet \rightarrow \gamma \text{ fakes}$	$3.36 \pm 1.13$
Beam halo	$5.9 \pm 4.7$
Others	$3.05 \pm 0.67$
Total Expectation	$72.9 \pm 8.30$
Data	77

where  $N_{data}$  is the observed number of events,  $N_{BG}$  is the number of estimated background events,  $A$  is the geometrical and kinematic acceptance of the selection criteria,  $\epsilon$  is the selection efficiency within the acceptance, and  $L$  is the integrated luminosity. The product of  $A \times \epsilon_{MC}$  is estimated from LO MADGRAPH simulation

and a correction factor,  $\rho$ , described in Section 6.2.1 is applied to account for the difference between the efficiency in the data and Monte Carlo:

$$A \times \epsilon = A \times \epsilon_{MC} \times \rho. \quad (6.2)$$

The product of  $A \times \epsilon_{MC}$  is estimated to be  $0.314 \pm 0.002$  (stat)  $\pm 0.048$  (syst) and rho is  $0.99 \pm 0.06$ .

The photon energy scale, jet and  $E_T^{\text{miss}}$  energy scale and resolution, and pileup related contributions are considered as sources of systematic uncertainty in the acceptance calculation. The uncertainty on the photon energy scale is about 1.5% and the uncertainty from variations in the  $E_T^{\text{miss}}$  energy scales is 5%. Contributions from the jet energy scale are accounted for in the uncertainty on the  $E_T^{\text{miss}}$ . The uncertainty on the integrated luminosity is 2.7% [33]. A summary of the systematic uncertainties are shown in Table 6.3.

**Table 6.3:** Summary of systematic uncertainties for signal and different background sources, shown in %.

Sources	$Z\gamma \rightarrow \nu\bar{\nu}\gamma$	$W\gamma \rightarrow \ell\nu\gamma$	$\gamma j$	$j$ faking $\gamma$	$e$ faking $\gamma$	Other bkg
Luminosity	2.7	2.7	2.7	-	-	2.7
PDF and Scale	5.37	8.9	-	-	-	-
EWK corrections	11	7	-	-	-	-
Jet background	-	-	-	30	-	-
Electron background	-	-	-	-	20	-
$j, E_T^{\text{miss}}, \gamma$ energy scale	6	6	6	-	-	6
Scale Factors	6	6	6	-	-	6

The measured cross section for  $pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma$  for photon  $p_T > 175$  GeV within



rapidity range  $|\eta_\gamma| < 1.4$  is

$$64.06 \pm 12.14(\text{stat}) \pm 12.88(\text{syst}) \pm 1.72(\text{lumi}) \text{ fb.} \quad (6.3)$$

The NNLO theoretical cross section is  $65.55 \pm 0.02$  fb where the uncertainty includes only the scale variations. The measured cross section agrees well with the NNLO theoretical cross section and this agreement with the SM prediction constrains possible DM models.

### 6.3.2 Limits on Dark Matter

Interpreting these results as setting limits on the cross section of a DM particle as a function of DM mass, Table 6.4 shows 90% confidence level (CL) upper limits on the production cross sections provided for the vector and the axial-vector model for a mediator mass of 10 TeV.

**Table 6.4:** Observed (expected) 90%CL upper limits on the DM production cross section assuming a vector or axial-vector mediator coupling, with mass  $m_M = 10$  TeV.

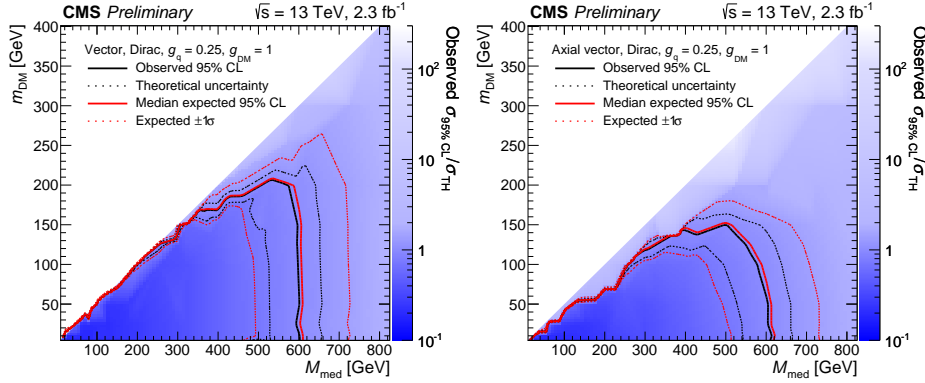
$m_\chi$ [GeV]	$\sigma(pp \rightarrow \gamma M \rightarrow \gamma\chi\bar{\chi})$ [fb]	
	$g_{Mq}$ vector	$g_{Mq}$ axial-vector
1	$3.821 \pm 3.242$	$3.782 \pm 3.211$
10	$3.820 \pm 3.244$	$3.785 \pm 3.213$
50	$3.827 \pm 3.249$	$3.793 \pm 3.213$
150	$3.826 \pm 3.254$	$3.754 \pm 3.192$
500	$3.588 \pm 3.052$	$3.488 \pm 2.961$
1000	$3.370 \pm 2.862$	$3.30 \pm 2.814$

Figure 6.2 shows the upper limits on the ratio of the cross section with respect to

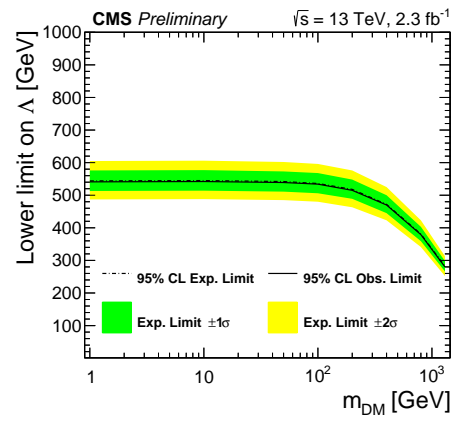
the theoretical predictions ( $\mu = \sigma^{95\%}/\sigma_{Th}$ ) for the vector and axial-vector mediator scenarios on the  $m_\chi$ - $m_M$  plane. The solid red and black curves are the expected and observed exclusion contours. The uncertainty on the expected upper limit includes the experimental uncertainties. For the simplified DM model considered, a mediator mass of up to 600 GeV is excluded for  $m_\chi < 10$  GeV.

For DM EFT model with a contact interaction of type  $\gamma\gamma\chi\bar{\chi}$ , upper limits are placed on the production cross section, which are then translated into the lower limits on the suppression scale  $\Lambda$ . The 95% CL observed and expected lower limits on  $\Lambda$  as a function of dark matter mass  $m_\chi$  are shown in Figure 6.3. Values of  $\Lambda$  up to 540 GeV are excluded at 95% CL.

**Figure 6.2:** 95% CL upper limits on  $\mu = \sigma/\sigma_{Th}$  in the  $m_\chi$ - $M_M$  plane for vector and axial-vector mediator, assuming couplings  $g_q = 0.25$  and  $g_\chi = 1$ . The solid red and black curves are the expected and observed exclusion contours. The dotted black contours around the observed limit and the dotted red contours around the expected limit represent the one standard deviation theoretical uncertainties in the cross section and the combination of the statistical and experimental systematic uncertainties, respectively.



**Figure 6.3:** (a) The 95% CL observed and expected lower limits on  $\Lambda$  for a dimension-7 operator EFT model with a contact interaction of type  $\gamma\gamma\chi\bar{\chi}$  as a function of dark matter mass  $m_\chi$ .



# Chapter 7

## CONCLUSIONS AND FUTURE

## PROSPECTS

In this thesis, two analyses of data collected by the CMS collaboration using  $pp$  collisions provided by the LHC are presented.

The SM process  $pp \rightarrow Wb\bar{b} \rightarrow \ell\nu b\bar{b}$  is studied at  $\sqrt{s} = 8$  TeV using a data sample that corresponds to an integrated luminosity of  $19.8 \text{ fb}^{-1}$ . The  $W$  boson is identified by an isolated lepton ( $\mu$  or  $e$ ) with  $p_T^\ell > 30$  GeV and  $|\eta^\ell| < 2.1$ . Backgrounds from  $pp \rightarrow t\bar{t}$  and Drell–Yan processes are reduced by rejecting events with a second lepton within  $p_T > 10$  GeV and  $|\eta| < 2.4$ . Exactly two  $b$ -tagged jets with  $p_T > 25$  GeV and  $|\eta| < 2.4$  are required to be present in selected events, to remove contamination in the signal region from charm and light flavor jets. To reduce the contribution from  $pp \rightarrow t\bar{t}$  events, events with a third jet with  $p_T > 25$  GeV and  $|\eta| < 4.7$  are rejected.

Fits are performed in sidebands dominated by  $t\bar{t}$  events to adjust the simulated jet energy scale as well as the scale factor associated with the difference in efficiency between data and simulation for the identification of  $b$  quarks. After making these adjustments, a fit is performed in the signal region and the cross section is extracted as  $\sigma(pp \rightarrow W(\ell\nu)+b\bar{b}) = 0.64 \pm 0.03(\text{stat}) \pm 0.10(\text{syst}) \pm 0.06(\text{theo}) \pm 0.02(\text{lumi}) \text{ pb}$ . This cross section is compared with four SM predictions made using MCFM and

MADGRAPH +PYTHIA with varied PDFs and is found to be compatible.

The other analysis is of the monophoton signature and is performed using data corresponding to  $2.3 \text{ fb}^{-1}$  at  $\sqrt{s} = 13 \text{ TeV}$ . The data are selected requiring one isolated photon with  $p_T^\gamma > 175 \text{ GeV}$  and  $|\eta| < 1.44$  and events are vetoed if they contain a charged lepton (an electron or a muon) with  $p_T > 10 \text{ GeV}$  that is separated from the photon by  $\Delta R > 0.5$  radians. The monophoton signature is one where the photon recoils from the interaction with some particle(s) that do not leave a trace in the detector so events are required to have  $E_T^{\text{miss}} > 170 \text{ GeV}$ . To ensure that the main source of  $E_T^{\text{miss}}$  is not photon energy mismeasurement, the azimuthal opening angle between the candidate photon and  $E_T^{\text{miss}}$  is required to be greater than 2 radians. Jet energy mismeasurement can also give rise to  $E_T^{\text{miss}}$  so, events are rejected if the minimum azimuthal opening angle between  $E_T^{\text{miss}}$  and up to four leading jets ( $\min\Delta\phi(E_T^{\text{miss}}, j)$ ) is less than 0.5 radians.

Interpreting these results as a measurement of the SM cross section for invisible decays of the  $Z$  boson the cross section is measured as  $\sigma(pp \rightarrow Z\gamma \rightarrow \nu\bar{\nu}\gamma)64.06 \pm 12.14(\text{stat}) \pm 12.88(\text{syst}) \pm 1.72(\text{lumi}) \text{ fb}$  which is in agreement with the theoretical value calculated at NNLO of  $65.55 \pm 0.02 \text{ fb}^{-1}$ . These results are also interpreted in the context of a search for DM using simplified models with a vector or axial-vector mediator and as an EFT coupling vertex  $\gamma\gamma\chi\bar{\chi}$ . which allows for DM production via the channel  $pp \rightarrow \gamma \rightarrow \gamma\chi\bar{\chi}$ . No evidence for DM has been found, and limits on the parameters in these models are set.

In the simplified model assuming a DM mass  $m_\chi < 10 \text{ GeV}$ , the mediator mass is found to be  $m_M \not\leq 600 \text{ GeV}$  assuming either vector or axial-vector couplings. In the

EFT model, lower limits are on the coupling strength suppression scale are presented as a function of  $m_\chi$  and are  $\Lambda < 540$  GeV is excluded at 95% CL.

The LHC continues to provide  $pp$  collisions at  $\sqrt{s} = 13$  TeV which are presently being collected and analyzed by the CMS collaboration. Only  $2.3 \text{ fb}^{-1}$  were analyzed in the monophoton analysis presented in this thesis, and data continues be collected. The statistical uncertainty presented in this monophoton analyses is comparable with the systematic uncertainty and will decrease with more data as the number of expected events scales linearly with the integrated luminosity. With  $30 - 40 \text{ fb}^{-1}$  of data projected to be collected by the end of 2016, this corresponds to an expected 1000-1300 events clearly identified in the monophoton final state.

With the discovery of the Higgs boson at the LHC, all fundamental particles predicted by the SM have now been observed. Searches for physics beyond the SM and for DM in particular are therefore an exciting field of study and could possibly lead to new understandings of the material hypothesized to constitute the majority of mass throughout the universe.

# BIBLIOGRAPHY

- [1] S. 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.B*, vol. 735, p. 204, 2014.
- [2] G. 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*, vol. 06, p. 084, 2013.
- [3] “Measurement of the  $p\bar{p} \rightarrow W+b+X$  production cross section at  $\sqrt{s} = 1.96$  TeV,” *Phys.Lett.B*, vol. 718, p. 1314, 2013.
- [4] “First Measurement of the b-jet Cross Section in Events with a W Boson in p-pbar Collisions at  $\sqrt{s} = 1.96$  TeV,” *Phys.Rev.Lett.*, vol. 104, p. 131801, 2010.
- [5] V. Khachatryan *et al.*, “Search for new phenomena in monophoton final states in proton-proton collisions at  $\sqrt{s} = 8$  TeV,” *Phys. Lett.*, vol. B755, pp. 102–124, 2016.
- [6] M. Aaboud *et al.*, “Search for new phenomena in events with a photon and missing transverse momentum in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector.” submitted to JHEP, 2016.
- [7] S. Chatrchyan *et al.*, “Search for dark matter and large extra dimensions in monojet events in  $pp$  collisions at  $\sqrt{s} = 7$  TeV,” *JHEP*, vol. 1209, p. 094, 2012.
- [8] V. Khachatryan *et al.*, “Search for physics beyond the standard model in final states with a lepton and missing transverse energy in proton-proton collisions at  $\sqrt{s} = 8$  TeV,” *Phys. Rev.*, vol. D91, no. 9, p. 092005, 2015.
- [9] S. Chatrchyan *et al.*, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys.Lett.B*, vol. 716, p. 30, 2012.

- [10] G. 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.B*, vol. 716, p. 1, 2012.
- [11] R. Field, “Early lhc underlying event data - findings and surprises,” 2010.
- [12] H.-L. Lai, J. Huston, Z. Li, P. Nadolsky, J. Pumplin, D. Stump, and C.-P. Yuan, “Uncertainty induced by QCD coupling in the CTEQ global analysis of parton distributions,” *Phys.Rev.D*, vol. 82, p. 054021, 2010.
- [13] “Measurements of Inclusive W and Z Cross Sections in pp Collisions at  $\sqrt{s} = 7$  TeV,” *JHEP*, 2011.
- [14] “Performance of the CMS missing transverse energy reconstruction in pp data at  $\sqrt{s} = 8$ TeV,” *JINST*, vol. 10, p. P02006, 2015.
- [15] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- $k_t$  jet clustering algorithm,” *JHEP*, vol. 04, p. 63, 2008.
- [16] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, “MadGraph 5 : Going Beyond,” *JHEP*, vol. 1106, p. 128, 2011.
- [17] F. Maltoni and T. Stelzer, “Madevent: Automatic event generation with mad-graph,” *JHEP*, vol. 02, p. 027, 2003.
- [18] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 physics and manual,” *JHEP*, vol. 05, p. 026, 2006.
- [19] J. Alwall *et al.*, “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions,” *Eur. Phys. J.*, vol. C53, pp. 473–500, 2008.
- [20] J. Alwall, S. de Visscher, and F. Maltoni, “QCD radiation in the production of heavy colored particles at the LHC,” *JHEP*, vol. 02, p. 017, 2009.
- [21] P. Nason, “A new method for combining NLO QCD with shower Monte Carlo algorithms,” *JHEP*, vol. 11, p. 040, 2004.



- [22] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with parton shower simulations: the POWHEG method,” *JHEP*, vol. 11, p. 070, 2007.
- [23] S. Alioli, P. Nason, C. Oleari, and E. Re, “NLO single-top production matched with shower in POWHEG: s- and t-channel contributions,” *JHEP*, vol. 0909, p. 111, 2010.
- [24] S. Alioli, P. Nason, C. Oleari, and E. Re, “Single-top Wt-channel production matched with parton showers using the POWHEG method,” *Eur.Phys.J. C*, vol. 71, p. 1547, 2011.
- [25] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, “W Physics at the LHC with FEWZ 2.1,” *Comput.Phys.Commun.*, vol. 184, p. 208, 2013.
- [26] Y. Li and F. Petriello, “Combining QCD and electroweak corrections to dilepton production in FEWZ,” *Phys.Rev.D*, 2012.
- [27] J. M. Campbell and R. Ellis, “Mcfm for the tevatron and the lhc,” *Nucl.Phys.Proc.Suppl.*, vol. 205, p. 10, 2010.
- [28] S. Badger, J. M. Campbell, and R. Ellis, “Qcd corrections to the hadronic production of a heavy quark pair and a w-boson including decay correlations,” *JHEP*, vol. 03, p. 027, 2011.
- [29] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, “Parton distributions for the LHC,” *Eur.Phys.J.C*, vol. 63, p. 189, 2009.
- [30] C. Collaboration, “Combination of ATLAS and CMS top quark pair cross section measurements in the emu final state using proton-proton collisions at 8 TeV,” tech. rep., CMS, 2014.
- [31] G. Aad *et al.*, “Measurement of the  $t\bar{t}$  production cross-section using  $e\mu$  events with  $b$  -tagged jets in  $pp$  collisions at  $\sqrt{s} = 7$  and 8 TeV with the ATLAS detector,” *Eur. Phys. J. C*, vol. 74, no. 10, p. 3109, 2014.

- [32] S. Chatrchyan *et al.*, “Measurement of the  $t\bar{t}$  production cross section in the dilepton channel in pp collisions at  $\sqrt{s} = 8$  TeV,” *JHEP*, vol. 02, p. 24, 2014. [Erratum: JHEP02,102(2014)].
- [33] CMS Collaboration, “Cms luminosity based on pixel cluster counting - summer 2013 update,” CMS Physics Analysis Summary CMS-PAS-LUM-13-001, CMS, 2013.
- [34] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. RÃ¼fenacht, M. Schönherr, and G. Watt, “LHAPDF6: parton density access in the LHC precision era,” *Eur. Phys. J. C*, vol. 75, p. 132, 2015.
- [35] M. Botje *et al.*, “The PDF4LHC Working Group Interim Recommendations,” 2011.
- [36] S. Alekhin, S. Alioli, R. D. Ball, V. Bertone, J. Blumlein, *et al.*, “The PDF4LHC Working Group Interim Report,” 2011.
- [37] R. D. Ball *et al.*, “Parton distributions with LHC data,” *Nucl. Phys.*, vol. B867, p. 244, 2013.
- [38] A. Buckley *et al.*, “LHAPDF6: parton density access in the LHC precision era,” *Eur.Phys.J.C*, vol. 132, p. 3, 2015.
- [39] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “A brief introduction to PYTHIA 8.1,” *Comp.Phys.Comm.*, vol. 178, p. 852, 2008.
- [40] S. Chatrchyan *et al.*, “Measurement of the production cross sections for a Z boson and one or more b jets in pp collisions at  $\sqrt{s} = 7$  TeV,” *JHEP*, vol. 06, p. 120, 2014.
- [41] S. Chatrchyan *et al.*, “Study of double parton scattering using W + 2-jet events in proton-proton collisions at  $\sqrt{s} = 7$  TeV,” *JHEP*, vol. 1403, p. 032, 2014.
- [42] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, “The automated computation of

- tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations,” *JHEP*, vol. 07, p. 079, 2014.
- [43] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, “An Introduction to PYTHIA 8.2,” *Comput. Phys. Commun.*, vol. 191, pp. 159–177, 2015.
- [44] S. Agostinelli *et al.*, “Geant4 - a simulation toolkit,” *Nucl. Instrum. Meth. A*, vol. 506, p. 250, 2003.
- [45] J. Allison *et al.*, “Geant4 developments and applications,” *IEEE Trans. Nucl. Sci*, vol. 53, p. 270, 2006.
- [46] S. Catani, D. de Florian, G. Ferrera, and M. Grazzini, “Vector boson production at hadron colliders: transverse-momentum resummation and leptonic decay,” *JHEP*, vol. 12, p. 047, 2015.
- [47] A. Denner, S. Dittmaier, M. Hecht, and C. Pasold, “NLO QCD and electroweak corrections to  $W+\gamma$  production with leptonic  $W$ -boson decays,” *JHEP*, vol. 04, p. 018, 2015.
- [48] A. Denner, S. Dittmaier, M. Hecht, and C. Pasold, “NLO QCD and electroweak corrections to  $Z + \gamma$  production with leptonic  $Z$ -boson decays,” *JHEP*, vol. 02, p. 057, 2016.
- [49] V. Khachatryan *et al.*, “Measurements of Inclusive  $W$  and  $Z$  Cross Sections in  $pp$  Collisions at  $\sqrt{s} = 7$  TeV,” *JHEP*, vol. 01, p. 080, 2011.
- [50] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torielli, and M. Zaro, “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations,” *JHEP*, vol. 07, p. 079, 2014.
- [51] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “Pythia 6.4 physics and manual,” *JHEP*, vol. 5, p. 26, 2006.