Doctoral Dissertation

Measurement of $\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj)$ in pp collisions at $\sqrt{s} = 8$ TeV

Department of Physics
Graduate School, Chonnam National University

Sanghyeon Song

August 2013

Contents

List of Figures

List of Tables

Measurement of $\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj)$ in pp collisions at $\sqrt{s} = 8$ TeV

SONG Sanghyeon

Department of Physics
Graduate school Chonnam National University
(Supervised by professor KIM Jaeyool)

(Abstract)

The cross-section ratio $\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj)$ is measured in the dilepton decay mode, using a data sample corresponding to an integrated luminosity of 19.6 fb^{-1} collected in pp collisions at 8 TeV with the CMS detector at the LHC. The cross-section ratio $\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj)$ is measured in the visible phase space corresponding to the detector acceptance, and corrected to particle level. The measurement is performed by means of a fit to be measured b-jet discriminator distribution of first and second additional jets of dileptonic top quark pair candidates with at least four reconstructed jets. The result of the measurement, 2.6 ± 0.4(stat.) $\pm$ 0.3(syst.)% at $\sqrt{s} = 8$ TeV.

Chapter 1

Introduction

The curiosity to study the ultimate constituents of matter and the nature of interactions between them evolved with the evolution of universe for mankind. Many of the theoretical and experimental insightful aspects of nuclear and particle physics have accomplished during the twentieth century and the period is considered as the 'golden period' for these ancestor physics to the High Energy Physics (HEP). It all started with 'J.J Thomson' who, in 1897, identified electron as a particle. People knew that atoms existed (or at least they conjectured that they did, as chemists had made quite clear) and were considered as the smallest and indivisible constituent of matter but they were surprised to find out that they were divisible. Thomson showed that electron was about 2,000 times lighter than hydrogen ion,, which was the lightest thing around. In 1911, the amazing results of the existence of atomic nucleus came up from the α-particle scattering experimental set up by 'Ernest Rutherford'. Then, in 1913, 'Niels Bohr' came up with his atomic model, where electrons move about the nucleus in circular orbit, separated from each other like rungs in a ladder.

As the experiments became more sophisticated and probed matter at higher and higher energies, all sorts of elementary particles started to show up. A remarkable trade off between energy and matter was at play, the expression of Einstein's famous $E = mc^2$ formula: if particles are accelerated to very high energies, making the head on collisions, the energy of their motion can transform and result into production of new particles which do not exist in nature in ordinary conditions. The rules that control these matter transmutations are the most basic laws of nature, in the sense that total amount of conserved quantity remains the same before and after the collision: conservation of energy, conservation of electric charge, and a bunch of other conserved quantities. In a sense, these laws constitute the being of modern physics, while the myriad material transmutations constitute the becoming. By the 1970s, a model capturing almost that was known in particle physics emerged, known as the 'Standard Model' (SM).

The SM describes the fact that all the hundreds of particles discovered during the twentieth century are made of only twelve fundamental particles, six quarks (that make up protons, neutrons and hadrons) and six leptons (the electron, the muon, the tau and their partner neutrinos respectively), and the interactions between those particles are mediated by the gauge bosons. The SM predictions are verified experimentally at particle accelerators with ever increasing accuracy. However, there is still a part of the puzzle missing. The SM is essentially a massless theory and it does not explain the mass of the particles. In 1964, the Belgian physicists 'Robert Brout' and 'Francios Englert' and the Scotsman 'Peter Higgs' proposed a mechanism in which particles of the SM gain mass by interaction with so called the Higgs field [1, 2]. The observation of this Higgs boson is one of the main goals of Large Hadron Collider (LHC), built by European Centre for Nuclear Research (CERN), in Geneva.

The LHC recorded the first proton-proton collisions at center of mass energy of $\sqrt{s} = 7$ TeV on March 30^{th}, 2010. Tow general purpose experiments were built at the point of proton-proton interaction, A Toroidal LHC Apparatus (ATLAS) and Compact Muon Solenoid (CMS), to record and reconstruct collisions with very high precision. These two experiments have been designed to cover a large spectrum of signatures in the LHC environment and the search for the Higgs boson is the major guider criterion followed to define the requirements and performances of detectors. Both the Higgs hunting experiments have observed an excess in their data, with a level of certainty worth a 'discovery'. On July 4^{th}, 2012, CERN announced the discovery of new boson of mass 125

GeV/c^2 decaying to two Z bosons, and two photons and hence with spin different from one, and therefore called a Higgs-like particle [3, 4]. And this particle was announced certain Higgs boson in March 14^{th}, 2013.

In the SM, the Higgs boson is expected to couple to the top quark with a strength of almost 1. Therefore, it is important to measure the direct coupling to top quark and check the consistency of the SM Higgs sector. One of the most promising channels in the search for direct coupling of the Higgs boson to top quark is the $t\bar{t}H$ channel. According to the mass of Higgs boson is about 125 GeV/c^2, it is expected to decay mostly to $b\bar{b}$, finally $t\bar{t}b\bar{b}$ state. However, this final state has an irreducible background from the $t\bar{t}b\bar{b}$ through normal QCD process (gluon splitting) and this final has not been observed yet. Therefore, in this thesis, it is outline a strategy that measuring this process and perform a measurement of cross-section ratio $\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj)$.

Chapter 2

Theoretical motivation

2.1 The standard model of elementary particles

2.1.1 Overview

The standard model of particle physics (SM) is a theory describing all known elementary particles and all their interactions except for the gravitational one, which is anyway irrelevant at microscopic scales.

The SM is formulated as a quantum field theory: matter is described in terms of spin 1/2 fermions, whose interactions are mediated by spin 1 bosons; finally, there is a scalar Higgs boson field, also interacting with fermions and gauge bosons. The three interactions mediated by the vector bosons are the nuclear strong, electromagnetic and nuclear weak interaction, mediated respectively by gluons, photons and W and Z bosons.

Three generations of fermions exist, each composed by a pair of quarks with electrical charges +2/3, -1/3, one leptons of charge -1, and one almost-massless neutrino with no electrical charge; each fermion has an associated anti-fermion with opposite electrical charge.

Quarks The three doublets of quarks are: up and down (u, d), stable constituents of the atomic nuclei, with masses of a few MeV/c^2: charm and strange (c, s), with masses of about 1.3 and 0.1 GeV/c^2; top and bottom (t, b), with masses of about 172 and 4.2 GeV/c^2. Quarks have only been observed in hadrons, bound states of either three quarks (baryons), three anti-quarks (anti-baryons), or one quark and one anti-quark (mesons); an exception is the top quark, whose extremely short lifetime prevents the formation of bound states.

Quark flavor is always conserved by strong and electromagnetic interactions, but not by the weak interactions mediated by W bosons. Decays through weak interactions of hadrons composed of u, d and s quarks are characterized by long lifetimes, with $c\tau$ values ranging from centimeters to meters, while for hadrons containing c and b quarks lifetimes are $\mathcal{O}$(100 μm) and $\mathcal{O}$(500 μm) respectively. Decys mediated by strong or electromagnetic interactions have lifetimes too short to be detectable.

Leptons The three generations of leptons are the electron e, the muon μ and the tau τ, with masses of about 0.5 MeV/c^2, 0.1 GeV/c^2 and 1.8 GeV/c^2; while only the electron is stable, also the muon can be considered stable in the context of collider experiments due to its long lifetime $c\tau \sim 660$m). Because of the shorter tau lepton lifetime, $c\tau \sim 87\mu$m, taus are observed only through their decay products.

Three neutrinos exist, each associated to one of the charged leptons, ν_e, ν_μ, ν_τ. A lepton number is defined in each generation as the number of leptons, plus the number of leptons and neutrinos, minus the number of anti-leptons and anti-neutrinos; in the SM, th lepton flavor is conserved separately for each generation, except for neutrino oscillation phenomena which conserve only the total lepton number. Neutrinos interact with other particles only through the nuclear weak force, and are therefore undetectable except in dedicated experiments.

In the contemporary formulation accounting for neutrino masses, usually a heavy neutrino is included in each fermion generation, but these neutrinos play little or no role in the high energy physics phenomenology.

For spin 1/2 fermions, a chirality can be defined as the projections of the spin along the momentum of the particle, which is a relativistic invariant for massless fermions; in the massless limit, chirality is also preserved in the interactions between fermions and vector bosons.

2.1.2 The standard model as gauge theory

The framework used to describe the interactions in the SM is that of gauge theories, quantum field theories where particles are endowed with a local symmetry, i.e. where the transformation parameters can be function of the space-time position; these symmetries act on some internal degrees of freedom of the particles, and not their space-time degrees of freedom.

A generic gauge theory is formulated by selecting a symmetry group $\mathcal{G}$, arranging all fields in representations of this group, and defining a gauge invariant action $S = \int \mathcal{L} d^4x$, which is most often achieved by making the Lagrangian itself invariant.

Fields transform linearly under the action of the gauge symmetry as $\phi'^i(s) = U^{ij}(x)\phi^j(x)$, but space-time derivatives of the fields do not since U^{ij} is dependent on x, so a covariant derivative is introduces as

$$\mathcal{D}_\mu \phi^i(x) = \partial_\mu \phi^i(x) - igt_a^{ij} A_\mu^a(x)\phi^j(x) \tag{2.1}$$

where $A_\mu^a(x)$ are the gauge vector fields, t_a^{ij} the generators of $\mathcal{G}$ in the representation of the fields ϕ^i and g is the charge. If all ordinary derivatives are replaced with covariant ones then a Lagrangian invariant under the global symmetry is made invariant also under gauge transformations.

The beautiful feature of gauge theories is that the interaction terms in the Lagrangian $\mathcal{L}$ arise naturally from the free Lagrangian just by replacing the ordinary derivatives in the kinetic terms with the covariant ones. This means that the most general gauge theory with vectors and fermions can be synthetically written as

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu}^a F^{\mu\nu a} - \bar{\Psi}(iD - m)\Psi \tag{2.2}$$

where the field strength tensor for the gauge bosons $F_{\mu\nu}^a$ is given by

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - igf^{abc} A_\mu^b(x) A_\nu^c(x) \tag{2.3}$$

the last term introduces trilinear and quadrilinear terms in the Lagrangian, corresponding to self-interaction between the gauge bosons, except in the case of abelian symmetries, for which there is a single gauge boson and f^{abc} is zero. The symmetry forbids the presence of mass terms for the gauge bosons in the Lagrangian.

As all interactions are determined by the free Lagrangian, gauge theories are very constrained: everything is fixed by the choice of the group, the representation of the fermions (gauge bosons are always in the fundamental representation), and the value of the charges; for non abelian group (i.e $SU(2)$) there is just one charge, while abelian $U(1)$ can have a different charge for each particle.

2.1.2.1 The standard model gauge group

In order to describe the strong, electromagnetic and weak interactions of the standard model, the gauge group of the SM is taken to be the direct product $SU(3)_C \times SU(2)_L \times U(1)_Y$.

Strong: $SU(3)_C$ $SU(3)$ is the group implementing the strong interactions among quarks and gluons, described by the quantum chromodynamics (QCD). The group has 8 generators, corresponding to the 8 massless gluons, and is non abelian, so the gluon are charged ("colored"), and have self-interactions.

Quarks fill the simplest non trivial representations of the group, 3 and $\bar{3}$, from the product of which the singlet bilinear $q\bar{q}$ and the octet $\bar{q}\gamma^\mu\lambda^a q$ are obtained, corresponding to the gauge-invariant mass term, and the covariant vector current couplet to gluons.

- **Strong coupling and confinement**
  At low energies the interactions among quarks are very strong, leading to a complex dynamic that cannot be described as a perturbative expansion around a free theory, although theoretical predictions can still be obtained with other techniques such as lattice calculations. The energy scale at which these effects become important is $\Lambda_{QCD} \sim 250$ MeV, which is roughly the mass scale of light hadrons.

It is an experimental fact that all free particles are "colorless" $SU(3)$ singlet: all mesons are in the scalar $\bar{q}^a q^a$ state, and baryons are in the antisymmetrical $\epsilon^{abc} q^a q^b q^c$. The intuitive explanation for this is that among colored particles there is an attractive force which increases for increasing distance, but there is yet to no well proved and quantitative explanation of this behavior from a theoretical point of view. In the lack of a full understanding of confinement from first principles, predictions for hadron formation are obtained using phenomenological models tuned on experimental results.

- **Asymptotic freedom**

  In the high energy regime the behavior of QCD changes: quantum corrections from vacuum polarization cause the interaction to become weaker with decreasing distance or increasing energy. This behavior is well understood in the context of renormalization, and is extremely important because it allows a perturbative treatment of the strong interactions.

Electroweak: $SU(2)_L \times U(1)_Y$ The electromagnetic and weak interactions emerge from the breaking of a $SU(2) \times U(1)$ symmetry, as explained later in section 2.2. in this section, the dynamics of these interactions will be described as if the symmetry were not broken.

From the point of view of the electroweak theory, fermions of different chirality are considered different fields, and have different interactions. Because of this, an unbroken $SU(2)_L \times U(1)_Y$ also forbids the presence of fermions mass terms in the Lagrangian.

The two components of the electroweak gauge group are associated to the weak isospin and weak hypercharge.

- **Weak isospin**

  The weak isospin is described in terms of the $SU(2)_L$ group, with three generators corresponding to a triplet of gauge bosons ($\vec{W}$ or W^i). Just like in the $SU(3)$ case, the SM fermions fill in the simplest representations of $SU(2)$: fermions of left-handed chirality are in the doublet $T = \frac{1}{2}$

representation, while right-handed fermions are in the trivial singlet $T = 0$ representation (i.e. they do not interact with the $\vec{W}$ bosons).

- **Weak hypercharge**

  $U(1)_Y$ is the abelian group associated to the weak hypercharge, corresponding to the interactions mediated by a neutral B boson. In principle, the symmetry would allow for arbitrary values of the charges for each particle, but in the SM constraints among the charges are enforced by the requirements that the symmetry is preserved at quantum level (anomaly cancellation).

The particle content of the SM, with their gauge group representation and hypercharges, including the right-handed neutrino, is summarized in figure 2.1 and table 2.1. The electrical charge is given by $Q = T_3 + Y$, the sum of the weak hypercharge and the weak isospin ($\pm\frac{1}{2}$ for doublets, 0 for singlets).

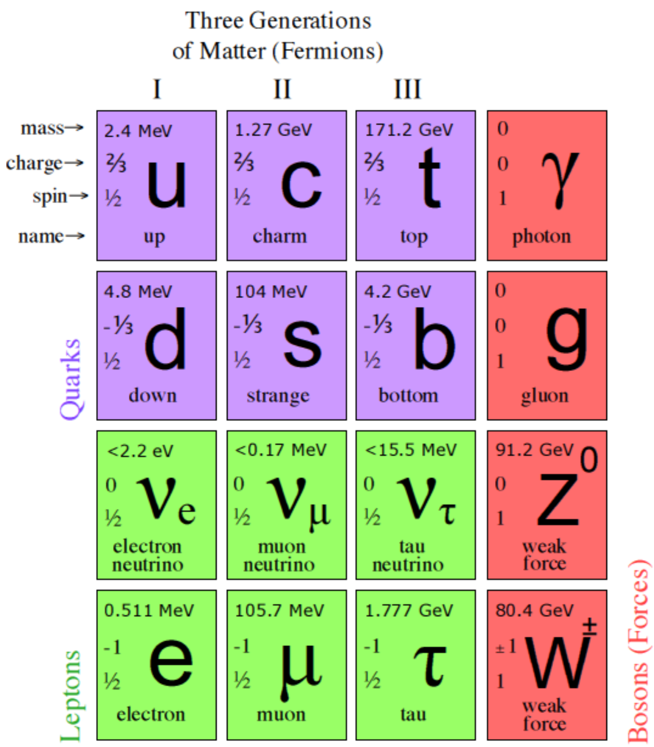


Figure 2.1: Schematic view of the SM elementary particles.

		$SU(3)$	$SU(2)$	$U(1)$
leptons	$L = \begin{pmatrix} \nu \\ e_L \end{pmatrix}$	1	2	-1/2
(x3 gen.)	e_R	1	1	-1
	N	1	1	0
quarks	$Q = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$	3	2	+1/6
(x3 gen.)	u_R	3	1	+2/3
	d_R	3	1	-1/3
gluons	g	8	1	0
W bosons	$\vec{W}$	1	3	0
B bosons	B	1	1	0

Table 2.1: Particle content of the SM for a gauge point of view, except for the Higgs sector. The L and R subscript refer to left-handed and right-handed chiralities; it is omitted for light neutrinos ν (only left-handed) and the possible heavy neutrinos N (only right-handed). The three generations of fermions have exactly the same gauge interactions so they are not stated separately.

2.2 Spontaneous symmetry breaking and the Higgs boson

The SM described in section 2.1.2.1, with only gauge interactions, is far from satisfying. First, the electroweak gauge symmetry requires the fermions and gauge bosons to be massless, which sharply contrasts with the observational evidence. However, an explicit breaking of the symmetry by adding extra mass terms in the Lagrangian yields a theory with fundamental problems in the high and low energy limits.

Moreover, gauge interactions are identical for all three generations of particles, i.e. that the Lagrangian is invariant under a full $U(3)$ group of unitary rotations in generation (or flavor) space, so that flavor quantum numbers for quarks in each generation must be separately conserved: this forbids transitions between fermions of different doublets such as those in the observed $K \to \pi\pi$ decays ($s \to du\bar{u}$ at quark level).

2.2.1 Phenomenology of spontaneous symmetry breaking

The spontaneous symmetry breaking is a peculiar feature of infinite dimensional systems like quantum field theories, in which a symmetric Lagrangian can produce a physics which is not symmetric; beyond particle physics, this feature is manifest in a plethora of different environments, from the rotational instability of fluids to condensed matter and solid state theories (superfluids, superconductor). The breaking happens in theories for which the vacuum state of minimum energy is not symmetric: the physical observables are vacuum expectation values of some functions of the fields, $\langle 0|\phi_1(x_1)\cdots\phi_n(x_n)|0\rangle$, which won't be symmetric if the vacuum $|0\rangle$ is not symmetric.

The phenomenology arising from the spontaneous breaking of a gauge symmetry can be introduced first with a minimalistic model: a charged scalar particle represented by the complex field ϕ coupled to a $U(1)$ gauge boson A_μ, for which the Lagrangian is

$$\mathcal{L} = -\frac{1}{4}(F_{\mu\nu})^2 + (D_\mu\phi)^\dagger(D^\mu\phi) + \mu^2\phi^\dagger\phi - \frac{\lambda}{4}(\phi^\dagger\phi)^2, \tag{2.4a}$$

$$(D_\mu = \partial_\mu - ieA_\mu) \tag{2.4b}$$

The potential depends only on $\rho^2 = \phi^\dagger\phi$, $V(\rho^2) == \mu\rho^2 + \frac{\lambda}{4}\rho^4$, and for $\mu > 0$ it has minimum on the circumference $\rho^2 = \frac{2\mu}{\lambda}$, not in the origin $\phi = 0$: this means that potential will be minimized for a vacuum expectation value the field $\nu = \langle\phi\rangle$ different from zero. The perturbative expansion of the theory is done using $\varphi = \phi - \nu$ as dynamical fiedl, and ν can be taken to be along the real axis. After the reparametrization $\varphi = \rho e^{i\theta}$ to express ϕ in terms of two real fields ρ, θ, the covariant derivative of ϕ becomes

$$D_\mu\phi = [\partial_\mu\rho + i\rho(\partial_\mu\theta - eA_\mu) - ieA_\mu(\rho - \nu)]\, e^{i\theta} \tag{2.5}$$

Under a gauge transformation of parameter Λ, fields transform as $\theta \to \theta + e\Lambda$ and $A_\mu \to \partial_\mu\Lambda$, so an appropriate choice of $\Lambda(x)$ can rotate away θ from the Lagrangian completely leading to

$$\mathcal{L} = -\frac{1}{4}(F_{\mu\nu})^2 + (\partial_\mu\rho)^2 + e^2(\rho - \nu)^2 A_\mu A^\mu + V(\rho) \tag{2.6}$$

This Lagrangian appears much different from the original one: there is a single, neutral scalar particle (ρ), interacting with the gauge boson through trilinear and quadrilinear vertices ρA^2 and $\rho^2 A^2$ and a mass term for the gauge boson $e^2\nu^2 A_\mu A^\mu$.

2.2.2 Electroweak symmetry breaking in the standard model

In the standard model, the Higgs field before symmetry breaking is a $SU(2)_L$ doublet Φ with weak hypercharge $\frac{1}{2}$; its covariant derivative can be written as

$$D_\mu\Phi = \partial_\mu\Phi - \frac{1}{2}ig\vec{W}_\mu\vec{\sigma}\Phi - \frac{1}{2}ig' B_\mu\Phi \tag{2.7}$$

In the parametrization where the upper and lower components of Φ are separately denoted as ϕ^+ and ϕ (both charged under $U(1)_Y$), and working in the unitary gauge for which $\langle\Phi\rangle$ is real, then the expansion of the field around the vacuum state of the broken symmetry is

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi \end{pmatrix} \xrightarrow{SSB} \begin{pmatrix} 0 \\ \nu \end{pmatrix} + \begin{pmatrix} \xi^+ \\ \xi \end{pmatrix} \tag{2.8}$$

Except for a more complex algebra, the conclusions are similar to those of the simplified $U(1)$ model: the terms quadratic in ν will provide mass to the gauge bosons through terms

$$\mathcal{L} \supset \frac{1}{8}\nu^2 \left(g^2|W^1|^2 + g^2|W^2|^2 + (-gw^3_\mu + g^{'}B^2_\mu\right) \tag{2.9}$$

which can be diagonalized to obtain four mass eigenstates

$$W^{\pm}_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu), \qquad m_W = g\frac{\nu}{2} \tag{2.10a}$$

$$Z^0_\mu = \frac{1}{\sqrt{g^2 + g^{'2}}}(gW^3_\mu - g^{'}B_\mu), \qquad m_Z = \sqrt{g^2 + g^{'2}}\frac{\nu}{2} \tag{2.10b}$$

$$A_\mu = \frac{1}{\sqrt{g^2 + g^{'2}}}(gW^3_\mu + g^{'}B_\mu), \qquad m_A = 0 \tag{2.10c}$$

The first two fields describe the W and Z bosons mediating weak interactions, and the last one the photon. Often four components of the Φ field, only one is left after the gauge transformation, the neutral scalar H referred to as the Higgs boson.

Fermion masses and flavor quantum numbers The same Higgs field that gives masses to the gauge bosons can also provide masses to the fermions: gauge invariance permits the following Yukawa terms to appear in the CM Lagrangian

$$\mathcal{L} \supset \lambda_u\bar{Q}\cdot\Phi^\dagger u_R + \lambda_u\bar{Q}\cdot\Phi d_R + \lambda_e\bar{L}\cdot\Phi e_R + \lambda_\nu\bar{L}\cdot\Phi^\dagger N \tag{2.11}$$

where the dot product is the scalar product in $SU(2)$ space.

The Yukawa couplings λ_x, 3×3 matrices because of three generations, result in mass matrices $m = \nu\lambda$ when Φ acquires a vacuum expectation value; the last term provide neutrino masses, but otherwise plays a negligible role in SM physics. In the case of leptons, when neglecting neutrino masses it is possible to redefine the fields so that the three lepton generations are simultaneously eigenstates of the mass matrix an of the $SU(2)$ interactions. This is not possible in the case of quarks, since there is one extra matrix to diagonalize; it is conventional to define the quark generations in the mas eigenstate basis, leaving a unitary mixing matrix V_{CKM} in the interaction part of the Lagrangian.

The introduction of Yukawa couplings also drastically reduces the global symmetry of the theory, as generations are no longer identical. With no neutrino masses, the total symmetry is $U(1)^4$: one from the global phase of all quarks, baryon number conservation, and one for each lepton flavor (which are separately conserved). If neutrino masses are added, but with no Majorana mass for N, it is only $U(1)_B \times U(1)_L$, and the total lepton and baryon number (L, B) are conserved; in the most general case another phase is constrained by the additional $m_N \bar{N}^c N$ term, and L is no longer conserved.

2.3 $\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj)$ analysis

2.3.1 Standard model Higgs boson production and decay

In the standard model, Higgs boson production in proton-proton collisions can happen through four main modes: gluon fusion $gg \to H$, vector boson fusion $qq \to H + 2jets$, associated production of a Higgs boson with a W or Z boson, and associated production with a $t\bar{t}$ pair. The hierarchy of the cross sections and their dependence on the Higgs boson mass is shown in figure 2.2. In general all production cross-sections decrease with increasing Higgs boson mass except for $gg \to H$.

Gluon fusion

The main Higgs boson production mechanism at the LHC is through gluon fusion, for which the leading Feynman diagram involves a quark loop (figure 2.3).

Vector boson fusion

With a production cross-section of about one tenth of the gluon fusion one, vector boson fusion is the second largest production mode relevant at the LHC. The three leading order diagrams for the process in the t, u and s channels are

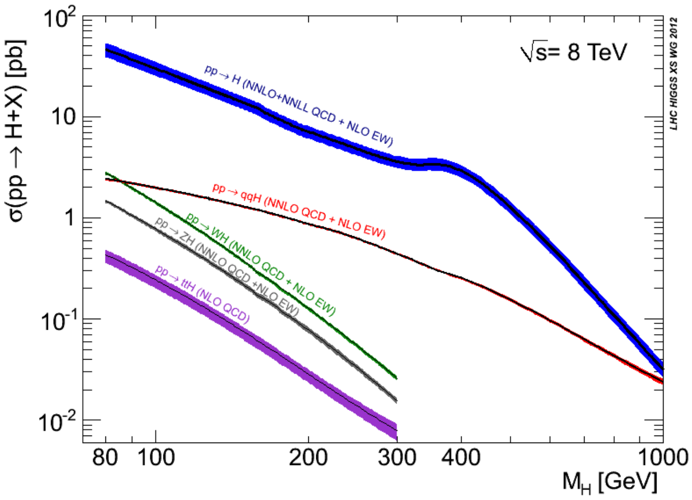


Figure 2.2: Theoretical predictions for the Higgs boson production cross-section in proton-proton collisions at $\sqrt{s} = 8$ TeV. The five production cross-section, in decreasing order, are gluon fusion, vector boson fusion, WH, ZH and $t\bar{t}H$.

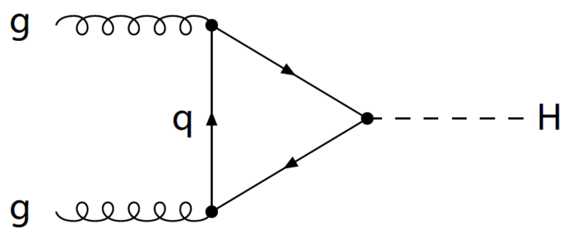


Figure 2.3: Leading diagram for $gg \to H$ production.

shown in figure 2.4).

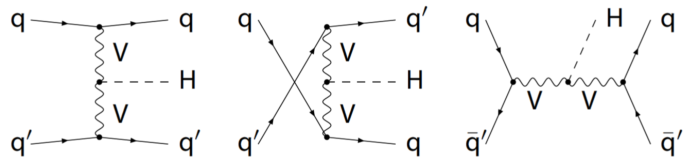


Figure 2.4: Leading order Feynman diagrams for $qq \to H + 2jets$ production throught vector boson fusion.

Associated production

The associated production of a Higgs boson and a W or Z boson is characterized by an event smaller production cross-section than vector boson fusion. This mode is nonetherless experimentally viable when considering Higgs boson decays to bottom quark pair, since leptons and neutrinos from the W or Z decay can provide handles to select event, while the all-hadronic final states from $gg \to H \to b\bar{b}, qq \to qqH \to qqb\bar{b}$ suffer larger backgrounds from QCD multijet production.

Associated $t\bar{t}H$ (figure 2.6) is an important channel, allowing for a direct measurement of the Yukawa coupling between top quark and Higgs boson ,but the very low production cross-section and the complex final state with a large number of hadronic jets make it very challenging to reconstruct properly. The cross-section has been computed to NLO in QCD only.

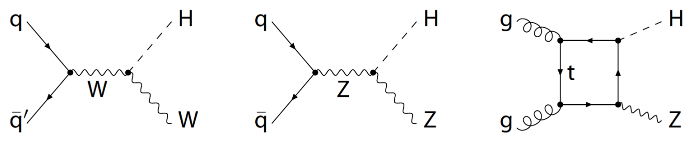


Figure 2.5: Leadnig order diagrams for WH and ZH associated production, and higher order contribution to ZH with gluon in the initial state.

The standard model Higgs boson is an unstable particle with immeasurably short lifetime, so it can be detected only through its decay products. As for any unstable particle, the branching ratio are determined by the partial widths of the decays into each final state

$$BR(H \to X) = \frac{\Gamma(H \to X)}{\sum_Y \Gamma(H \to Y)} \tag{2.12}$$

where each partial width depends onoly on the square of the coupling of the Higgs boson to those specific decay products and on kinematic factors.

At leading order in the SM couplings, the Higgs boson can decay into pairs of heavy fermions through Yukawa interactions (figure 2.7, left), and into pairs of W or Z bosons through $SU(2)_L$ interactions (figure 2.7, right).

The dependency of the branching ratio on the Higgs boson mass is shown in figure 2.8.

2.3.2 Top quark decay

An important feature of the top quark is its lifetime $\sim 10^{-24}$ ps, about 10 times shorter than the characteristic formation time of hadrons [15]. It means that the top quark decays before it hadronizes. The lifetime of a $t\bar{t}$ pair is in any case too small to allow a proper definition of a bound state with sharp binding energy. Even if the t and $\bar{t}$ quarks are likely to interact with the remnant proton partons which don't take part in the hard interaction to form light hadrons (early fragmentation process), this effect is negligible at the LHC energy. Hence we can assume that the top quarks are produced and decay like free quarks. The decay models $t \to d + W^+$ and $t \to s + W^+$ are CKM suppressed relatively to $t \to b + W^+$ by factors of $|V_{td}|^2$ and $|V_{ts}|^2$. Since these values are small compared to $|V_{tb}|^2$ ($0.9990 < |V_{tb}|^2 < 0.9992$) only the decay $t \to b + W^+$ is considered, as shown in Figure 2.9. Focusing our attention on the $t\bar{t}$ pair production, several decay channels can be investigated. They correspond to the decay mode of the W boson (hadronic or leptonic) originating from the top decay.

- **di-lepton channel ($t\bar{t} \to W^+b, W^-\bar{b} \to l_1\nu_1 b, l_2\nu_2\bar{b}$):**
  In this case both W^+ and W^- bosons stemming from the t and $\bar{t}$ quarks decay in a lepton (e, μ, τ) and the corresponding neutrinos (ν_e, ν_μ, ν_τ). The final state contains two isolated leptons with high p_T, two jets coming from the b quark hadronization (b-jet), and high missing transverse energy due

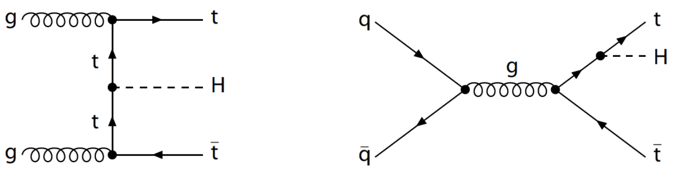


Figure 2.6: Leadnig order diagrams for $t\bar{t}H$ associated production wotj gluon and quarks in the initial state.

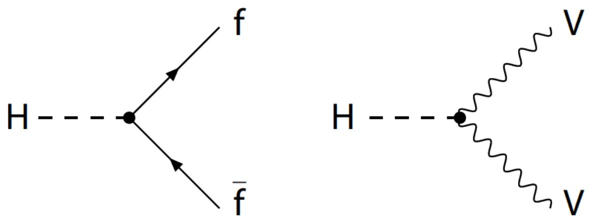


Figure 2.7: Leading order Feynman diagrams for Higgs boson decays into pairs of fermions (left) and gauge boson (right).

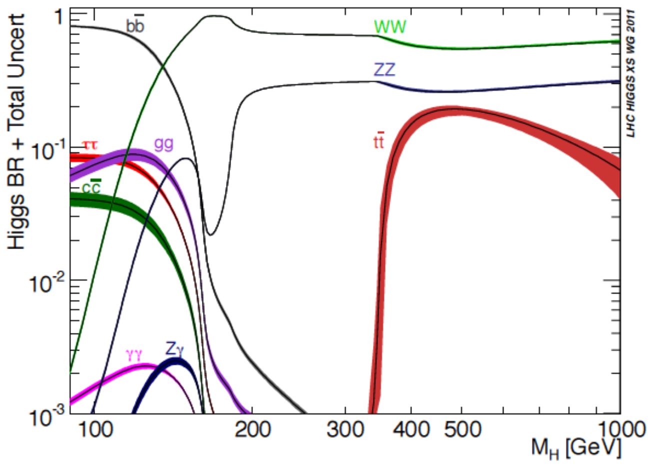


Figure 2.8: Predicted branching ratio for the standard model Higgs boson decays, as function of the Higgs boson mass.

to the neutrinos presence. The branching ratio for this channel is about 10%.

- **lepton + jets channel ($t\bar{t} \rightarrow W^+b, W^-b \rightarrow q_1q_2b, l_1\nu_1\bar{b}$):**
  It corresponds to the case in which one of the two W bosons decays into a lepton and the other one into a quark pair. So the final state contains one isolated high energy lepton, some missing transverse energy due to the presence of one neutrino, and 4 jets (two of them are b-jets). Each channel, electron + jets, muon + jets and tau + jets, has a branching ratio of about 15%.

- **fully hadronic channel ($t\bar{t} \rightarrow W^+b, W^-b \rightarrow q_1q_2b, q_3q_4\bar{b}$):**
  In this case both W bosons decay into a quark pair. Therefore the final state consists of 6 jets, two of them being b-jets. Even if this channel is the one with the largest branching ratio, 45%, it is challenging o investigate due to the important multijet background.

A summary of the branching ratios for the different channels is presented in Figure 2.10

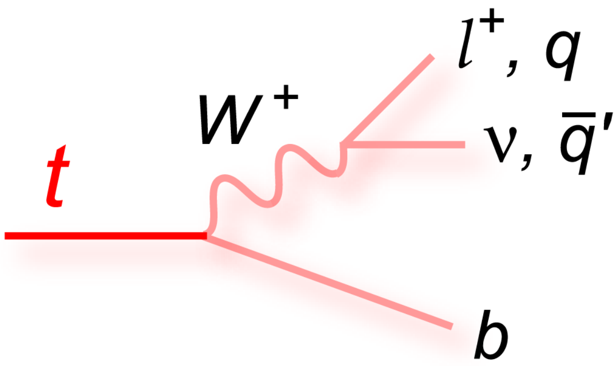


Figure 2.9: Main decay mode of the top quark.

2.3.3 Motivation of $\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj)$ measurement

In the SM, the Higgs boson is expected to couple to the top quark with a strength of about unity. Therefore, it is important to measure the direct coupling of the Higgs boson to the top quark and to check the consistency of the SM Higgs sector described in section 2.3.1.

One of the most promising channels for a direct measurement of the top quark Yukawa coupling is the the Higgs boson is produced in association with a $t\bar{t}$ pair ($t\bar{t}H$). In this case, it is expected to the Higgs boson decay mostly to $b\bar{b}$, leading to a $t\bar{t}b\bar{b}$ final state because the Higgs boson was discovered whose mass of about 125 GeV/c^2 (the Higgs boson decay branching ratio is shown in figure 2.8). However, this final state has an irreducible non-resonant background from the production of a top quark pair in association with a $b\bar{b}$ pair, as predicted by higher order QCD. This final state has not yet been observed. But it can be measured that cross-section ratio $\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj)$. The reason absolute cross-section measurement is difficult is due to the large differences in the expected cross-sections between $t\bar{t}jj$ and $t\bar{t}b\bar{b}$. One of the primary motivations for measuring the cross-section ratio rather than the absolute cross-section is that many experimental uncertainties are expected to cancel. Moreover, many kinematic distributions are expected to be similar in $t\bar{t}jj$ and $t\bar{t}b\bar{b}$, leading to similar acceptances and eventually reduced systematic uncertainties.

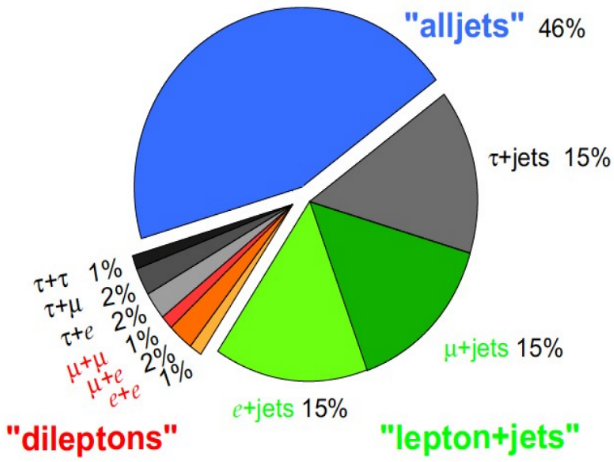


Figure 2.10: The branching ratios of $t\bar{t}$ decay channels.

Chapter 3

The CMS Experiment at the LHC

In this chapter the main characteristics of the LHC will be reviewed and the CMS experiment will be described.

3.1 The Large Hadron Collider

The Large Hadron Collider(LHC) is the most powerful proton-proton(p-p) collider in the world[16][17]. It was constructed in the European Organization for Neuclear Research(CERN) which is located in Swizerland-Frnace border near Geneva. The LHC is contained in a tunnel, with a circumference of 27 km, at a depth 50m to 175m underground shown in Figure 3.1. The tunnel was previously occupied by the Large Electron Positron(LEP).

A schematic description of the LHC accelerator complex and its services is shown in Figure 3.2. In the LHC, the acceleration is performed in several stages. The proton injection starts at the duo-plasmatron, which is the pronton source. A linear accelerator(LINAC) boosts the protons to energy of 50 MeV using Radio Frequency Quadruples and feeds to the Proton Synchrotron Booster(PSB). Then the protons are accelerated to 1.4 GeV and injected into the Proton Synchrotron(PS). The LHC bunch train starts in this Proton Synchrotron(PS). In the PS the energy increases again up to 25 GeV. Then the protons are acceler-

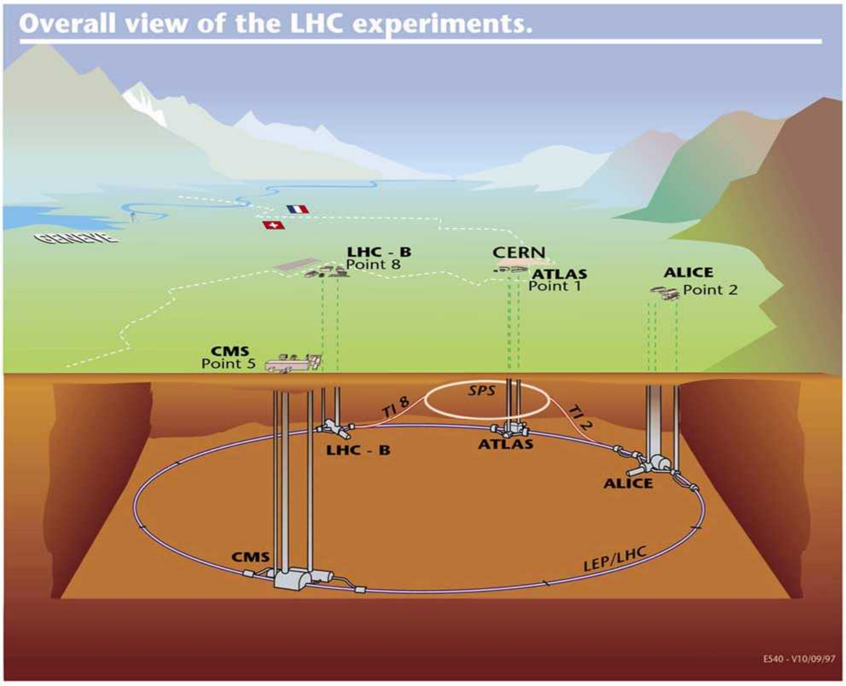


Figure 3.1: Computer graphic of the LHC facility showing clockwise, starting from CMS, LHCb, ATLAS and ALICE.

ated up to 450 GeV by the Super Proton Synchrotron(SPS) and they are finally injected into the LHC. In the LHC ring, the acceleration continues until the protons reach their moninal energy. The two beams collide in four interaction points.

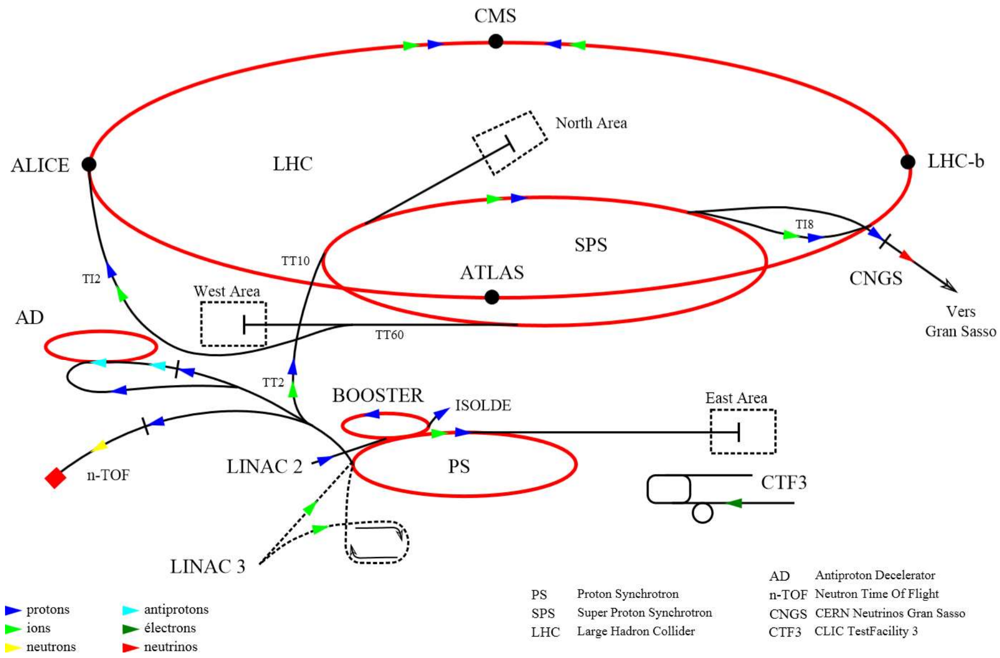


Figure 3.2: The LHC accelerator complex.

The LHC has 4 main experiments, each experiment is distinct and characterized by its unique subdetectors. A Toroidal LHC AppartatuS(ATLAS) and Compact Muon Solenoid(CMS) are general purpose detectors and their main goal is to search for Higgs boson which is predicted to be responsible for the electroweak symmetry breaking mechanism.Searching for new physics, such as supersymmetric particles, new massive vector bosons, extra dimensions, quark compositeness etc., at the TeV scale is an important for ATLAS ans CMS. Research programs with these detectors also aim to understand already discovered the Standard Model particles deeply and precisely. Having two independently designed detectors is for cross-confirmation of any new discoveries. Two medium size experiments are operated at LHC. One is LHCb and the

LHC Parameters	Nominal Values
Circumference	26.659 km
Center of mass energy($sqrts$)	14 TeV
Nominal Luminosity(L)	10^{34} cm^{-2}s^{-1}
Luminosity life-time	15 hrs.
Time between tow bunch crossings	24.95 ns
Distance between tow bunches	7.48 m
Longitudinal max. size of a bunch	7.55 cm
Number of bunches (n_b)	2808
Number of protons per bunch (N_b)	1.15×10^{11}
beta function at impact point ($\beta*$)	0.55 m
Transverse RMS beam size at impact point ($\alpha*$)	16.7 μm
Dipole field at 7 TeV (B)	8.33 T
Dipole temperature (T)	1.9 K

Table 3.1: List of the nominal LHC parameters, for p-p collisions, relevant for the detectors.

other is ALICE. The LHCb is a dedicated studying CP voilation and rare decays of heavy quarks, looking for new physics beyond the dominant Cabibbo-Kobayashi-Maskawa paradigm whithin the Standard Model. The ALICE is designed to exploit the unique physics potential of nucleus-nucleus interactions at LHC energies in the heavy-ion collisions The aim is to study the physics of strongly interacting matter at extreme energy densities, where the formation of a new phase of matter, the quark-gluon plasma, is expected.

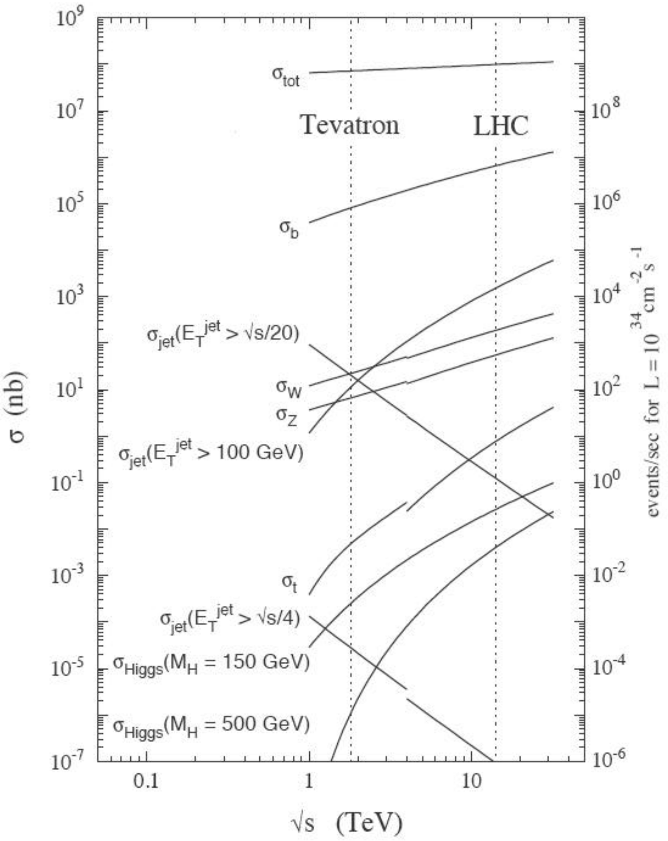


Figure 3.3: Expected cross section for several process as a function of the center of mass energy of pp collision.

3.2 The Compact Muon Solenoid

The Compact Muon Solenoid(CMS) is one of general purpose detector in the LHC[18]. The CMS is located approximately about 100 m underground in Cessy, France. The main goals of the CMS experiment are:

- To explore physics at the TeV scale
- To discover the Higgs boson
- To look for evidence of physics beyond the Standard Model, such as supersymmetry or extra dimensions
- To study aspects of heavy ion collisions

Figure 3.4 shows the schematic of the CMS detector, which has a diameter of 15m, a length of 28.7m and weight of 14000 tons. It is composed by a central barrel and two closing endcaps. It consists of several sub-detectors to identify precisely the different particles produced in the collisions that leads to the full event reconstruction. In this section, an overview of the CMS detector components is presented which leads to the better understanding of operation of the detector.

3.2.1 The coordinate system

The CMS used a right-handed coordinate system with the interaction point is chosen as the center of the coordinate system, the z-axis along the counterclockwise-beam axis, the y-axis is vertically upward and the x-axis is pointing to the center of the LHC ring, as illustrated in Figure 3.5. The aximuthal angle ϕ is measured with respect to the z-axis in the x-y plane and the polar ange θ is defined with respect to the z-axis. Suppose the four-momentum of a particle has the coordinates (E, p_x, p_y, p_z), then the longitudinal component is given as p_z and the transverse component is given by $p_T = \sqrt{p_x^2 + p_y^2}$. Hence the rapidity is given by

$$y = \frac{1}{2} \cdot ln \frac{E + p_z}{E - p_z} \tag{3.1}$$

If a particle is ultra-relativistic ($p \gg m$), its rapidity can be approximated by the pseudo-rapidity η, given by

$$\eta = -log\left(tan\frac{\theta}{2}\right) \tag{3.2}$$

where we have $\eta = 0$ for the particles moving perpendicular to the z-axis and $\eta = \pm\infty$ is in the $\pm z$-direction. Both $\Delta\eta$ and $\Delta\phi$ of two particles are independent of Lorentz boosts along the z-axis, therefore the distance between two particle can be measured ina third Lorentz invariant variable:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{3.3}$$

3.2.2 Tracker

The tracker is the innermost part of the CMS detector and it id designed to reconstruct the tracks of the charged particles with high efficiency and momentum resolution. It ensures the reconstruction of primary as well as the secondary vertices, making the detector highly hermetic. It covers a pseudo-rapidity range of $|\eta| < 2.5$. Since the particle flux decreases as $1/r^2$ within the detector, the tracking system is required to be as close as possible to the interaction point, depending on the radiation hardness of the material that is used to make it. The CMS inner tracker system is 5.8 m in length and of 2.5 m in diameter, around the interaction point placed inside the magnet system. The tracker is further subdivided into tow parts, the pixel tracking detector and the strip tracking detector. Where the former is made highly sophisticated to have the high momentum resolution and extreme accuracy to track the paths of emerging particles. A highly sophisticated electronics and powerful cooling are the obvious requirements for such operation. Figure 3.6 gives the general layout of the tracker. Each line represents a detector module. The double lines indicate the back-to-back mounted modules tilted by 100 mrad, and deliver the stereo hits. The transverse momentum resolution of the inner tracker is given by

$$\frac{\Delta p_T}{p_T} = C_1 \cdot p_T \otimes C_2 \tag{3.4}$$

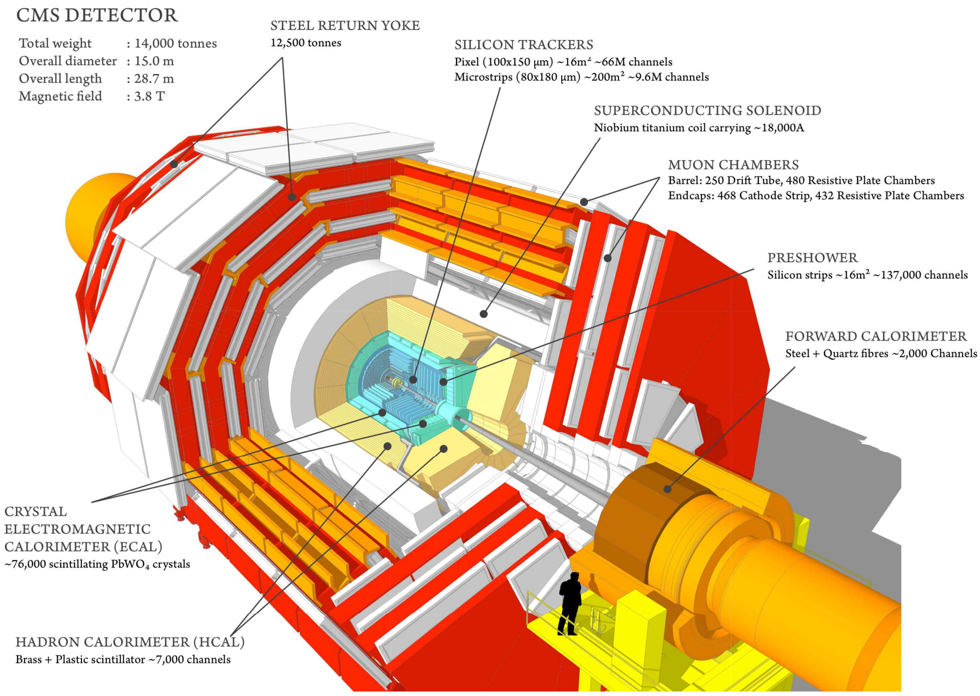


Figure 3.4: A view of CMS detector.

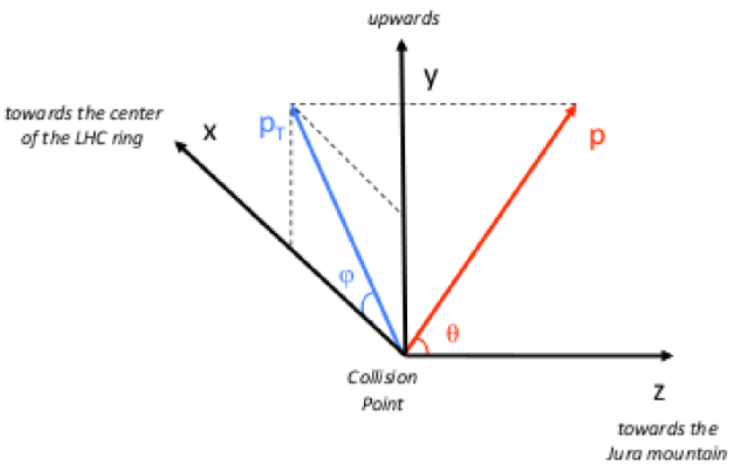


Figure 3.5: The CMS coordinate system.

where the term C_2 contains the multiple Coulomb scattering effects, which are dominant for low energy particles. Whereas C_1 depends on the detector geometry, in particular on the number of hits n, used to reconstruct a track, on its length L, and on the resolution on the single point (hit) measurement, σ_x, given as

$$C_1 \propto \frac{\sigma_x}{\sqrt{n} \cdot B \cdot L^2} \tag{3.5}$$

therefore C_1 is minimal for tracks made of many hits and traveling a long path within the tracker volume.

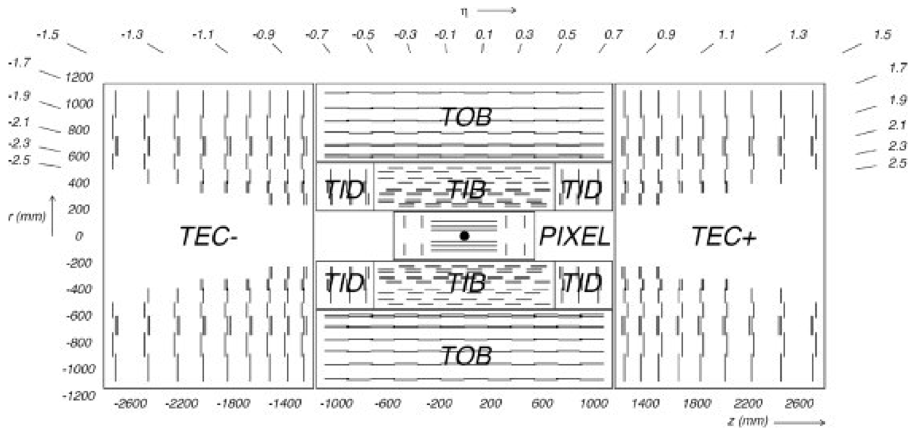


Figure 3.6: Schematic of the CMS inner tracking system in the $r - z$ plane. It is also shown that the η ranges of the different sub-systems.

3.2.2.1 Pixel Detector

The pixel tracker is made of silicon material, so-called silicon pixel vertex detector. It consists of three sylindrical layers pattern in the barrel region at the radial distance of 4.4 cm, 7.3 cm, 10.2 cm, with a total length of 53.5 cm, and tow disks in each endcap region, at $|z| = 34.5$ cm and 46.5 cm, with radio ranging from 6 to 15 cm, as shown in figure 3.7. It consists of 65 million of pixels in total with dimensions of 100×150 μm^2; 47923200 in the barrel and 17971200 in the endcap region. Hits in the pixel detector provides the true space points with considerable benefits for pattern recognition, and hence leads to the precise detection of primary vertex. The pixel detector delivers the spacial resolution of about 10 μm along the $r - \phi$ coordinate and of about 20 μm in the $r - z$ plane for such true space points.

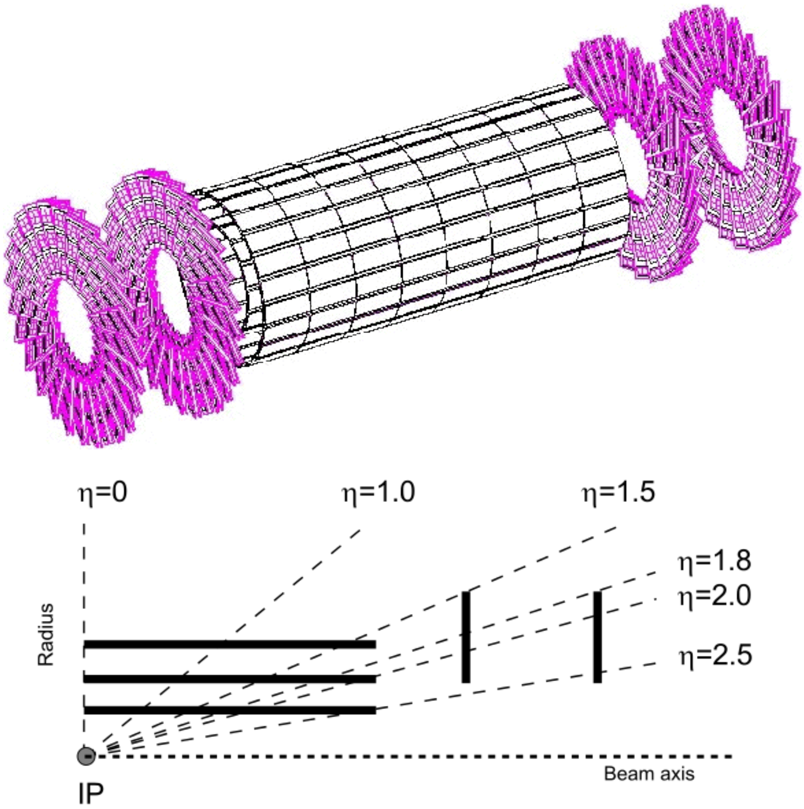


Figure 3.7: The layout of Silicon pixel detector.

3.2.2.2 Strip Detector

After the pixels and on their way out of the tracker, particles pass through the strip detectors which is reaching out to a radius of 130 cm. The strip detector is also made of silicon material and contains two collections of barrel modules for a pseudo-rapidity range of $|\eta| < 1.6$, the Tracker Inner Barrel (TIB) and the Tracker Outer Barrel (TOB). In the endcaps, it contains two collections of modules for a pseudo-rapidity range of $1.6 < |\eta| < 2.5$, the Tracker Inner Discs (TID) and the Tracker EndCaps (TEC). The silicon strip detector layout as shown in Figure 3.6.

The silicon strip detector consists of 10 active silicon layers which ontains

9.3 million of strips, covering 198 m^2 of area. The resolutions for the different parts of the silicon detector are different which depend on there position with respect to the interaction point: 23-35 μm for TIB/TID, 35-53 μm for TOB and 230-530 μm for TEC.

3.2.3 Electromagnetic calorimeter

The electromagnetic calorimeter is located just outside the tracker and surroundding it. The ECAL is made of scintillating crystals which provides precise identification and measurements on electrons and photons. The ECAL capability of identifying electomagnetic objects with excellent precision. The quarter of the CMS detector are shown Figure 3.8. The CMS ECAL is composed of 75848 lead tungstate($PbWO_4$) crystals, organized in a modular structure. The central barrel(EB) is structured in both halves as 18 identical "super-modules", each covering 20° in ϕ and a pseudo-rapidity of $0 \leq |\eta| \leq 1.479$, with an innermost radius of 1.29m. Blocks of 5 × 2 crystals compose submodules, which are in turn gathered into modules in the number of 40 or 50, surrounded by a metallic cross-plate. The 1700 crystals in the 85 × 20 modules structure from a super-module. Given the presence of 36 super-modules, the whole EB part consists of 61200 crystals. Each crystal exposes a front face cross of 22 × 22mm^2 and 23cm of length, corresponding to $25.8X_0$. In order to avoid particles channeling, crystals are off-pointing by a 30 angle with respect to a straight line from the interaction point, both in η and ϕ. Blocks of 5 × 5 crystals are the basic elements of the ECAL data readout and are referred to as trigger towers. The two endcaps(EE) are 3.17m separated from the vertex and cover a pseudo-rapidity range of $1.479 \leq |\eta| \leq 3$, with precision energy measurements provided up to $|\eta| = 2.6$ and crystals being installed up to $|\eta| = 3$, in order to improve the energy flow measurement in the forward direction. Each endcap is structured in two "Dees", demicircular aluminum plates composed by units of 5 × 5 crystals, known as "super-crystals". The endcap crystals off-point from the nominal vertex position and are arranged into an $x - y$ grid. Crystals used here have a front face cross section of 8.62 × 28.62mm^2 and are slightly shorter(22cm) than those in the EB, given the support coverage provided by the preshower detector, which is a sampling structure made of lead and silicon strips of $3X_0$ in thickness. The preshower detector is located just before the endcaps and its main purpose is to help identifying neutral point within the fiducial region $1.653 \leq |\eta| \leq 2.6$. It also improves the position determination of electrons and photons, due to its superior granularity.

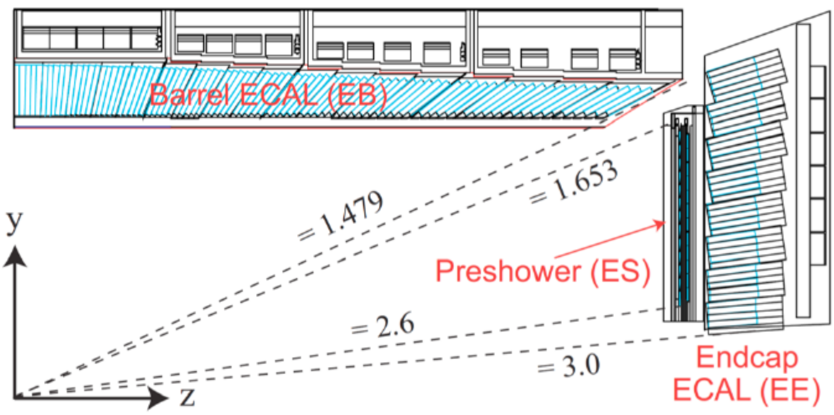


Figure 3.8: The layout of the CMS Electromagnetic Calorimeter system.

3.2.4 Hadronic Calorimeter

The CMS Hadron Calorimeter (HCAL) is used together with ECAL to measure energy and direction of chaged and neutral hadrons, and the energy imbalance in the transverse plane E_T^{miss}.

The central barrel and endcap HCAL subdetectors completely surround the ECAL and are fully immersed within the high magnetic field of the solenoid. The barrel (HB) and endcap (HE) are joined hermetically with the barrel extending out to $|\eta| = 1.4$ and the endcap covering the overlapping range $1.3 < |\eta| < 3.0$. The forward calorimeters (HF) are located 11.2m from the interaction and extend the pseudo-rapidity coveraged overlapping with the endcap from $|\eta| = 2.9$ down to $|\eta| = 5$. The forward calorimeters are specifically designed to measure energetic forward jets optimized to discriminate the narrow lateral shower profile and to increase the hermeticity of the missing transverse energy measurement. Central shower containment in the region $|\eta| < 1.26$ is improved with an array of scintillators located outside the magnet in the outer barrel hadronic calorimeter (HO).

HCAL is made of four subdetectors (shown by Figure 3.9):

- **Barrel Hadronic Calorimeter (HB)** : It is placed inside the magnetic coil and it covers the central pseudo-rapidity region, up to $|\eta| = 1.3$. The barrel hadron calorimeter consists of two half barrels, each composed of 18 identical 20° wedges in ϕ. Each wedge is composed of flat brass absorber plates parallel to the beam axis and interleaved with 3.7 mm thick plastic scintillators. The signal is readout through wavelength-shift fibers and hybrid photodiodes (HPD). The granularity is of the order of $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$. The HB has an energy resolution for single pions of approximately $120\%/\sqrt{E}$. The minium depth is about 5.8 λ_0.

- **Endcap Hadronic Calorimeter (HO)** : It is located as well inside the magnetic coil and it is made of two endcaps extending the angular coverage up to $|\eta| = 3$. The HE has the same architecture as HB with the same granularity. The signal is read through wavelength-shift fibers and hybrid photodiodes.

- **Outer Hadronic Calorimeter (HO)** : It is placed in the barrel region, and was added outside the magnetic coil in order to extend the depth of the calorimeter in terms of nuclear interaction length. The total depth in the central region is thus exended to about 11.8 λ_0. It consist of two scintillator layers with the same granularity than the one of the HB.

- **Forward Hadronic Calorimeter (HF)** : It consists of two units placed outside the magnetic coil, at ± 11.2 m from the interaction point along the beam direction. It extends the pseudo-rapidity coverage up to $|\eta| = 5$. The material of the two units, steel absorbers and embedded radiation hard quartz fibers, provide a fast collection of Cherenkov light. The granularity of HF is $\Delta\phi \times \Delta\eta = 0.17 \times 0.1745$.

In order to obtain a reference calibration and to measure the characteristics of the different parts of the HCAL, these were exposed to beams of electrons, pions, protons and muons. An ECAL module was also included in the test beam setup. The hadronic energy resolution of the barrel HCAL and ECAL combination is parameterized as $\sigma^2/E^2 = a^2/E + b^2$, where a corresponds to a stochastic term and b to a constant term. The energy resolution in the endcaps is similar to that of the barrel.

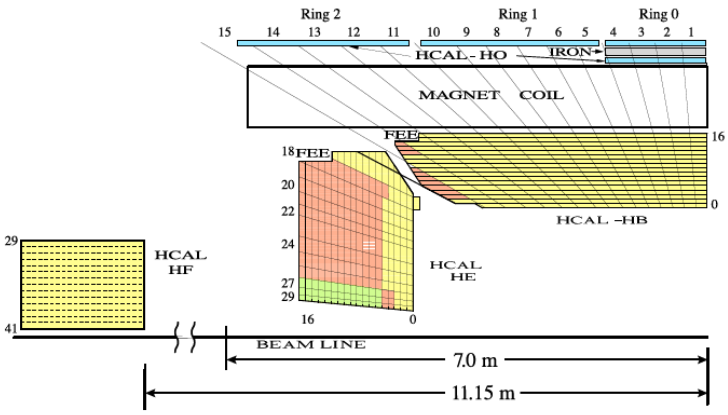


Figure 3.9: A schematic view of the CMS hadronic calorimeter (HCAL).

3.2.5 Magnet

The CMS magnet is the central device around which the experiment is built, with a 4 Tesla magnetic field that is 100,000 times stronger than the Earth's. The CMS has a large solenoid magnet. This allows the charge/mass ratio of particles to be determined from the curved track that they follow in the magnetic field. It is 13 m long and 6 m in diameter, and its refrigerated superconducting noibium-titanium coils were orginally intended to produce a 4 T magnetic field. The operating field was scaled down to 3.8 T instead of the full design strength in order to maximize longevity[19].

The inductance of the magnet is 14 H and the nominal current for 4 T is 19,500 A, giving a total stored energy of 2.66 GJ, equivalent to about half-atonne of TNT. There are dump circuits to safely dissipate this energy should the magnet quench. The circuit resistance (essentially just the cables from the power converter to the cryostat) has a value of 0.1 mΩ which leads to a circuit time constant of nearly 39 hours. This is the longest time constant of any circuit at CERN. The operating current for 3.8 T is 18,160 A, giving a stored energy of 2.3 GJ.

The job of the big magnet is to bend the paths of particles emerging from high energy collisions in the LHC. The more momentum a particle has the less its path is curved by the magnetic field, so tracing its path gives a measure of momentum. The CMS began with the aim of having the strongest magnet possible because a higher strength field bends paths more afnd, combined with high precision position measurements in the tracker and muon detectors, this allows accurate measurement of the momentum of even high energy particles.

The trackerand calorimeter detectors(ECAL and HCAL) fit snugly inside the magnet coil whilst the muon detectors are interleaved with a 12-sided iron structure that surrounds the magnet coils and contains and guides the field. Made up of three layers this "return yoke" reaches out 14 meters in diameter and also acts as a filter, allowing through only muons and weakly interacting particles such as neutrinos. The enormous magnet also provides most of the experimet's structural support, and must be very strong itself to withstand the forces of its own magnetic field.

3.2.6 Muon system

The CMS Muon System is dedicated to measure high p_T muons, in combination with the tracker. The Muon System is placed outside the magnetic coil, embedded in the return yoke, to fully exploit the 1.8 T return flux.

The system consists of three independent subsystems shown as Fig. 3.10

- Drift Tubes(DT) are placed in the barrel region, where the occupancy is relatively low ($< 10Hz/cm^2$).
- Cathode Strip Chambers(CSC) are in the endcaps, where the occupancy

is higher ($> 100Hz/cm^2$).

- Resistive Plate Chambers(RPC) are both in the barrel and endcaps.

The Drift Tube(DT) is made of chambers consisting of twelve layers of drift tubes each, packed in three independent substuctures called *super-layer*, for a total of four chambers with three super-laypers per chamber. In each chamber two super-layers have anode wires parallel to the beam axis, and one has perpendicular wires. Thus, each chamber(Fig. 3.11) can provide two measurements of the $r - \phi$ coordinate and one measurement of the z coordinate of the track hit positions. Each chamber is made of two parallel aluminium plates jointed with "I" shaped spacer cathodes. The position resolution is about 100 μm in both $r\phi$ and rz.

Cathode Strip Chabers(CSC) are multi-wire proportional chambers with segmented cathodes(Fig. 3.12). Each chamber can provide both hit position coordinates. The chambers are filled with a gas mixture of Ar(40%), CO_2(50%), CF_4(10%). The chamber spatial resolution is about 80-85 μm.

Resistive Plate Chambers(RPC) are made of parallel bakelite planes(Figure 3.13), with a bulk resistivity of $10^{10} - 10^{11}\Omega$cm. The gap between the plates is filled with a mixture of $C_2H_2F_4$(96.2%), iso-C_4H_{10}(3.5%) and SF_6 plus water vapour(0.3%). They operate in avalance mode. Those chambers have limited spatial resolution, but they have exellent timing performances; they are used for bunch crossing identification and for trigger purposes.

3.3 Forward Detector

3.3.0.1 CASTOR

The Centauro And Strange Object Research (CASTOR) detector is instrumented in a very forward region of the CMS detector with the purpose of search for exotic events in central Pb-Pb collisions at the LHC. It is a Cherenkov effect based, quartz (acting material) and tungsten (absorber material) calorimeter (45°degrees inclination with respect to the beam axis) placed at a distance of

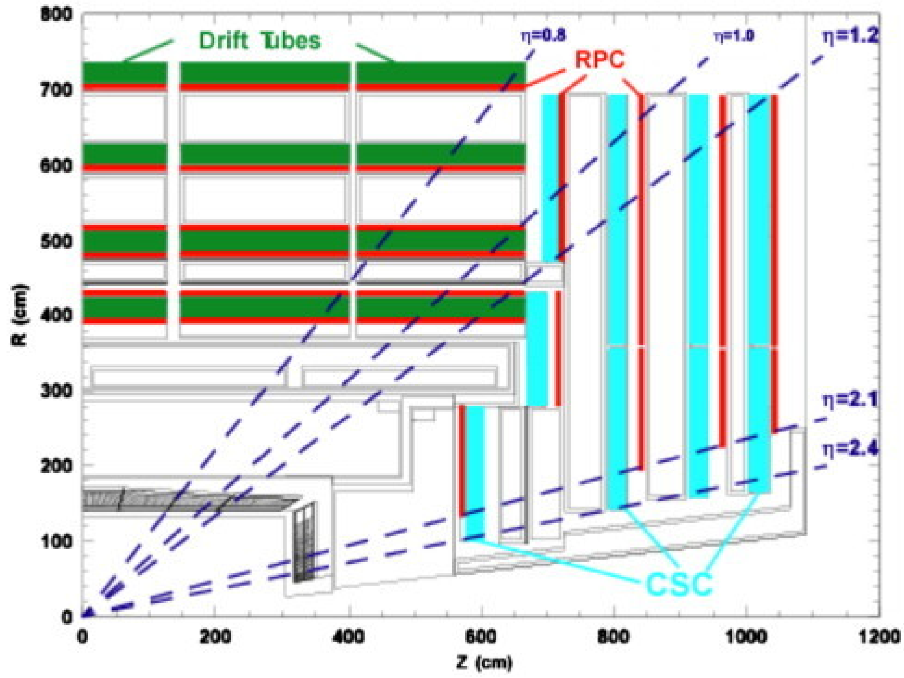


Figure 3.10: Layout of one quarter of the CMS muon system.

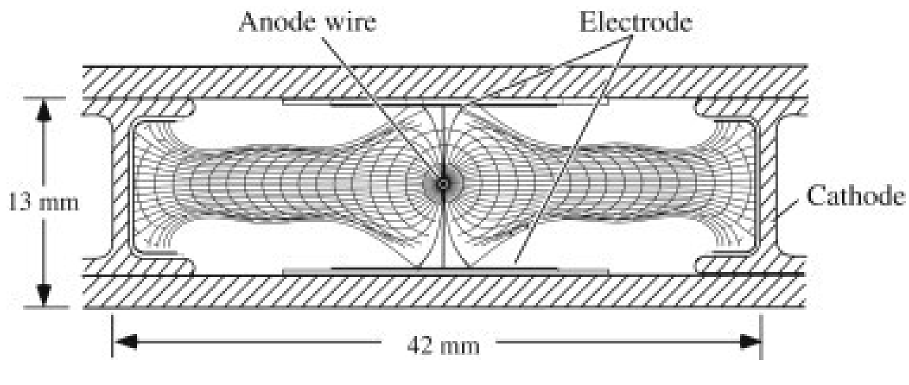


Figure 3.11: A schematic representation of a drift tube. Drift lines in presence of magnetic field are also shown.

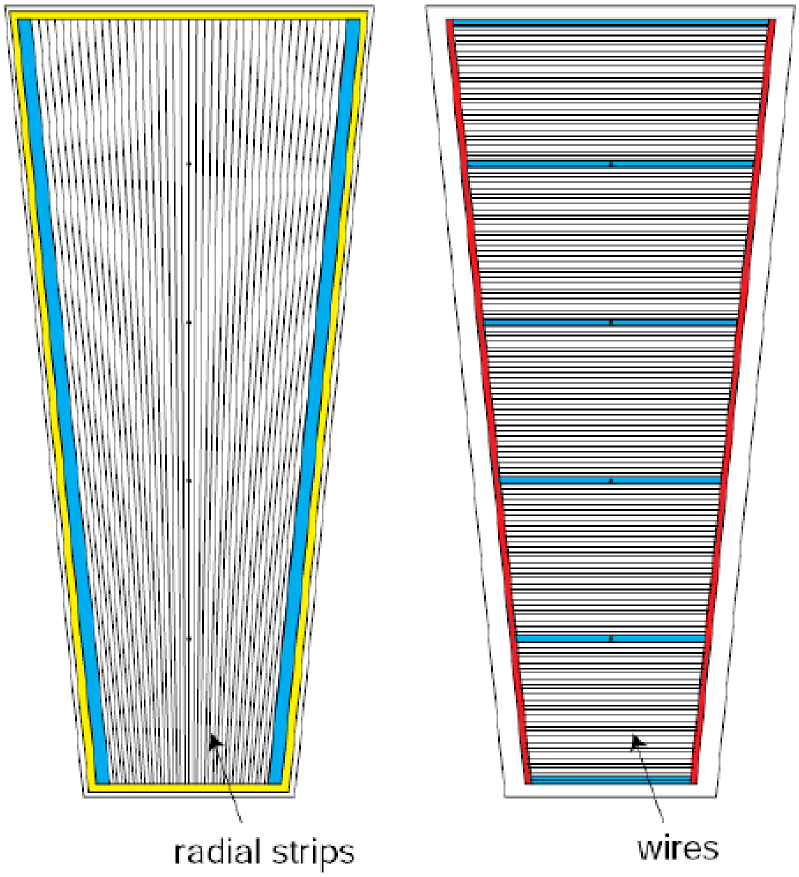


Figure 3.12: A schematic representation of the CSC cathode panel (on the left) and anode panel (on the right).

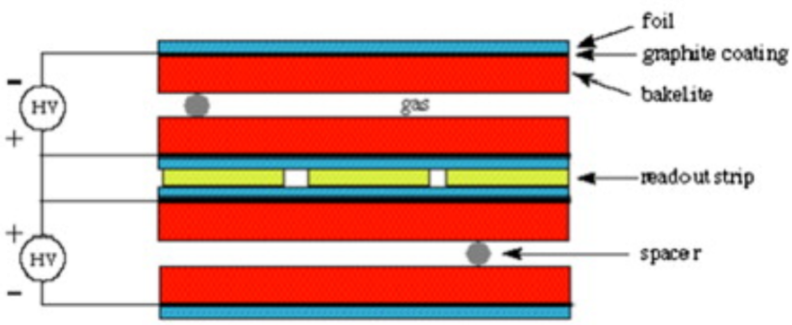


Figure 3.13: Schematic view of the RPC double-gap structure. The read-out strips in the Barrel chambers run along the beam direction.

14.37 m from the interaction point and provides the pseudo-rapidity coverage to the region $5.2 < |\eta| < 6.6$. It is cylindrical in shape with inner and outer radii of 3.7 cm and 14 cm respectively, with depth of 10.5 λ_I in terms if interaction length. The light signal originated while the passage of charged particle is transmitted to photo multiplier tubes by the means of air-core light guides. The detector is composed of a 20.12 X_0 thick electromagnetic section and of a 9.5 X_0 thick hadronic section. The whole unit is azimuthally divided in 16 sectors and longitudinally segmented in 12 reading units provides the energy resolution of $\sim 1\%$. Figure 3.14 gives the schematic of CASTOR calorimeter.

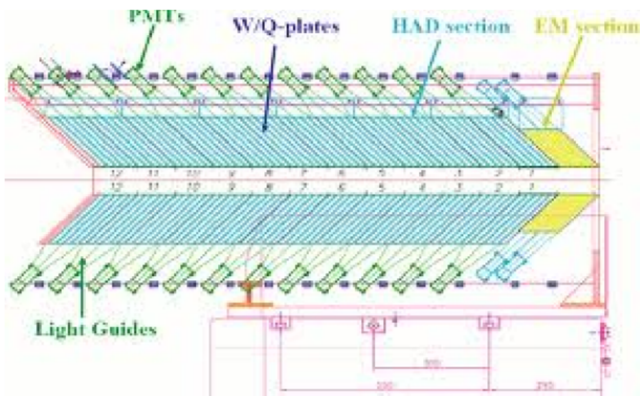


Figure 3.14: The CASTOR detector.

3.3.0.2 The Zero Degree Calorimeter

The Zero Degree Calorimeter (ZDC) is designed to catch particles (neutrons and very forward photons) very close to the beam axis, in particular, for heavy ion collisions. In addition, it works for beam tuning and luminosity monitoring. The design of the ZDC includes two independent calorimeter sections, an electromagnetic section and a hadronic section. Cherenkov effect based sampling calorimeters using tungsten and quartz fibers have been chosen for the energy measurements. It is installed 140 cm far from the interaction point on the both sides, ZDC+ and ZDC-, between the two LHC beam pipes with the pseudo-rapidity coverage of $|\eta| \geq 8.5$. The mechanical and optical design is shown in

Figure 3.15.

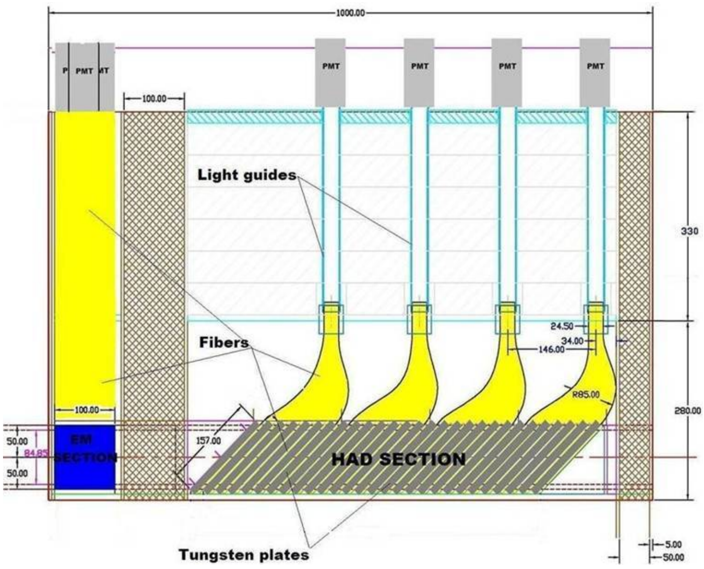


Figure 3.15: Schematic view of the Zero Degree Calorimeter.

3.4 Trigger and Data Acquisition

At the nominal operational conditions of the LHC, one expect the bunch crossing rate of 40 MHz corresponding to 25 ns of bunch crossing time, at the CMS interaction point. Each bunch crossing give rise to about 20 proton-proton collisions and hence we get 600 million collisions per second. One raw event needs $\sim$ 1 MB of memory to be recorded on the tape and therefore the final memory requirement is 40 TB per second. Since the Data AcQuisition system (DAQ) of the CMS experiment can cope with 100 GB of data per second, a drastic rate reduction is needed. In other words, a precise selection of useful events by physics purpose is required. The CMS trigger system helps in this selection which takes decision in a short time of 25 ns as per the requirements. It consists of two independent levels, the Level-1 (L1) and the High Level Trigger (HLT).

3.4.0.3 The Level-1 Trigger

The L1 trigger is completely hardware based system capable of bringing down the event rate from the initial 40 MHz to 100 kHz. The L1 schematic is given in Figure 3.16. Since the latency between the bunch crossing and the L1 to accept the signal is 3.2 μs, the L1 takes the decision on the basis of the calorimeter and the muon system information. The tracker information do not participate because the track reconstruction time exceeds the time limits of L1 decision. The accepted events are passed to the HLT.

The Calorimeter Trigger selects the four 'best' candidates of each of the following categories: electrons and photons, central jets, forward jets and jets identified on the basis of the shape of the deposited energy. These candidates are handed over to the Global Calorimeter Trigger (GCT), along with the measured missing transverse energy.

The Muon Trigger is performed independently by DTs, CSCs and RPCs. The DT and CSC triggers carry out a local muon reconstruction by comparing the slopes of track segments built in subsequent detector layers. The RPC trigger compares a given muon track with predefined hit patterns depending on the track p_T. The four best muon candidates are passed to the Blobal Muon Trigger system which discards the low quality tracks after matching those from DTs and CSCs with those from RPCs. The L1 electronics is installed partly directly on the detectors, partly in the underground control room about 20 m far from wht experimental cavern.

3.4.0.4 The High Level Trigger

The High Level Trigger (HLT) is a software system implemented in an Event Filter Farm which is single processing farm containing 1000 CPUs processes data from $\sim$ 700 front-end electronics. It reduces the event rate down to the final output of $\sim$ 100 Hz that can be written on tape.

The HLT is divided in 2 'virtual' trigger levels, Level-2 and Level-3 triggers (L2 and L3). Since L1 trigger performs the reduction in rate therefore the HLT can have access to the complete read-out data and capable to perform complex calculations similar to those performed at the analysis level. At Level-2 the information from the muon system and the calorimeters is used and more refined objects are reconstructed using the L1 track as seeds. It defines a region in the $\eta - \phi$ space in which a seed for L3 objects is found. The Level-3 trigger matches the L2 seeds with the tracker tracks which leasds to the availability of the full event information. Further selection is completely dependent on the physics analyses requirements.

3.4.0.5 The Data Acquisition System

The Data Acquisition (DAQ) perform the transportation of the data from about 650 front-ends at the detector side, to the 'filter units' for processing of complete events. The central DAQ runs the online software on about 3000 PC used for intelligent buffering and processing of the event data. The principle components of the DAQ system of the CMS are shown in Figure 3.17. The detector front-ends are read out through a builder network with a bandwidth of 100 GB per second. Complete events are fed to the filter systems stems from the design choice of the CMS to build the full event already after the L1 trigger instead of building partial events as in traditional multi level trigger systems.

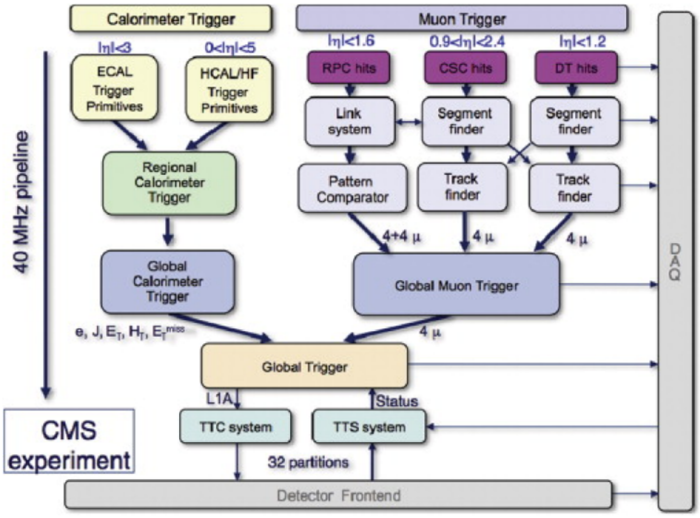


Figure 3.16: Schematic representation of the CMS Level-1 trigger system.

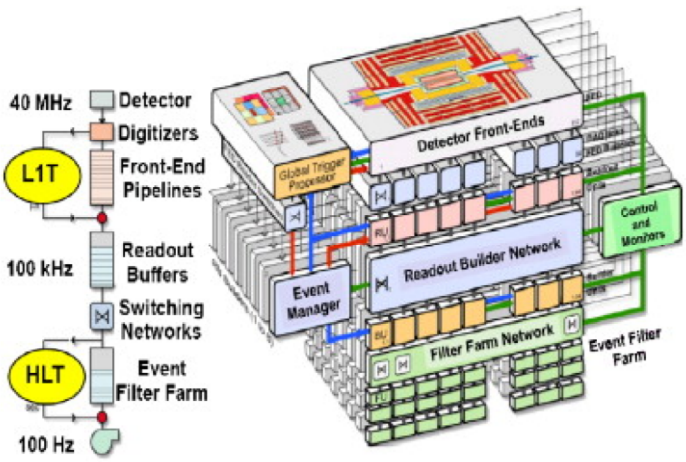


Figure 3.17: The principal components of the DAQ system of the CMS detector.

Chapter 4

Event Simulation and Reconstruction

4.1 Event generation

Proton-proton (pp) collisions allow the search for SM Higgs boson as well as the new physics due to large scales of momentum transfer involved. But the understanding of final states of high energy particle collisions such as those at the LHC is an extremely challenging theoretical problem. There are several basic phases of the process that are needed to be simulated:

- a primary hard subprocess
- the parton showers associated with the incoming and the outgoing colored participants in the subprocess
- the non-perturbative interactions that convert the showers into outgoing hadrons, so-called hadronization, and connect them to the incoming beam hadrons
- the secondary interactions that give rise to the underlying event, and the decays of unstable particles that do not escape from the detector

Of course, not all these steps are relevant in all processes. In particular, the majority of events that make up the total hadron-hadron cross-section are of soft QCD type and rely more on phenomenological models. At the other extreme the simulation of events such as the SM Higgs production and the SM backgrounds, rely on essentially all of the components.. A typical in-elastic pp collision at the LHC is given in Figure 4.1

Over last thirty years, a crucial tool of factorization has been developed which allows the separation of many processes of interest into differentregimes, according to the scales of momentum transfer involved [20]. The hard and soft scale regimes describe the hard subprocess and hadronization process respectively. The hard and soft regimes are distinct but connected by an evolutionary process that can be calculated in principle from perturbative QCD. One consequence of this scale evolution is the production of many additional partons in the form of initial and final state parton showers, which eventually participate in the low scale process of hadron formation. All the regimes are eminently suited to computer simulatioin using Monte-Carlo techniques. Moreover, the omportant components involved in the event generation process can be described as following.

4.1.1 Parton distribution functions

The Parton Distribution Funtions (PDFs) play an important role in event generations [20, 21], for the simulation of hard process, parton showers and underlying evnets. They are defined as the momentum distribution functions of the partons within the proton (when the spin derection of the partons is not considered), since the protons accelerated to high energies can be considered as a stream of partons, each carrying a fraction x of the lonitudinal momentum. In other words, they represent the probability densities to find a parton carrying a momentum fraction x at a squared energy scale. Thus, the choice of PDF set influences both cross-sections and event shapes.

4.1.2 Hard subprocesses

The simulation normally begins with a hard subprocess [20] in which constituents of the colliding protons interact at a high momentum scale to produce

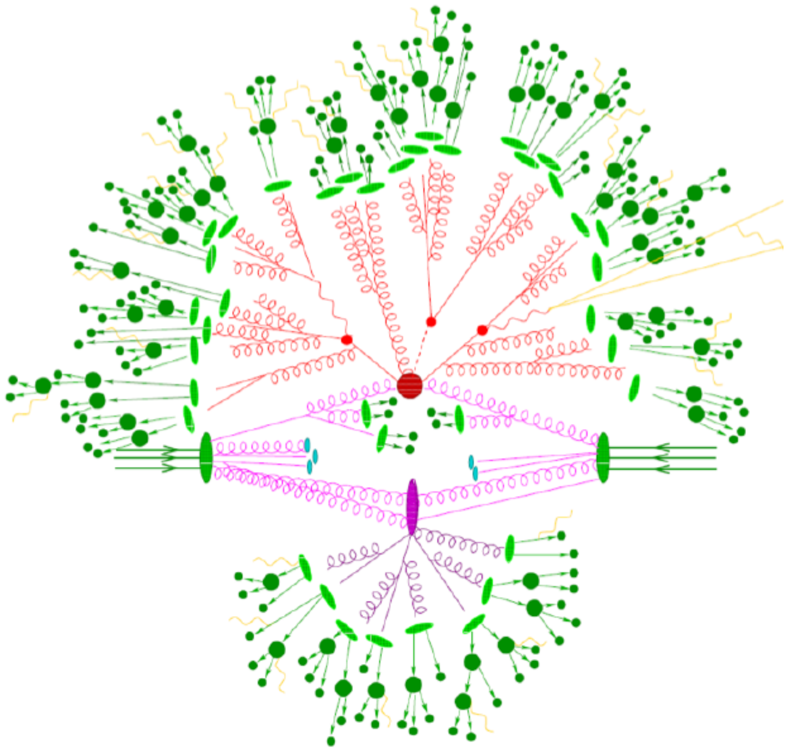


Figure 4.1: Scheme of a proton-proton collision. Two partons of the incoming proton interaction (red), while the proton remnants (magenta) provide the underlying event. The partons created in the hard interaction hadronization (light green) and unstable hadrons decay further to stable particles (dark green).

a few outgoing fundamental object: SM quarks, leptons and/or gauge or Higgs bosons, or hypothetical particles of some new theory. The momenta of the colliding constituents are selected by sampling the PDFs of the proton at the energy scale of the subprocess. These distributions have been measured at lower energies in other processes and are evolved to higher scales using the QCD evolution equations for parton densities. Convolution with the differential cross-section of the subprocess and integration over phase space gives the relevant production cross-section.

4.1.3 Parton shower

The hard subprocess, by definition, involves large momentum transfers and therefore the partons involved in it are violently accelerated. Just as accelerated electric charges emit QED radiation (photons), the accelerated colored partons emit QCD radiation in the form of gluons. Unlike the uncharged photons, the gluons themselves carry color charges and can therefore emit further radiation, leading to parton showers [20]. In principle, the showers represent the higher order corrections exactly, instead, an approximation cheme is used, in which the dominant contributions are included in each order of perturbation theory describing the hard subprocess. These dominant contributions are associated with collinear parton splitting or soft (low energy) gluon emission.

Using the Monte-Carlo method to generate the values of splitting parameters for each splitting, a parton shower is developed from each colored parton of the hard subprocess.

4.1.4 Hadronization

At some point, the evolution perturbation theory becomes invalid and the dynamics enter in a non-perturbative phase, which leads to the formation of the observed final state hadrons. This hadronization process is not amenable to the currently available non-perturbative techniques for calculation, and therefore event generators have to rely on models based on general features of QCD.

4.1.4.1 String model

The string model for hadronization [20], is based on the observation, from lattice simulations of QCD, that at large distances the potential energy of the color sources, such as a heavy quark-antiquark pair, increases linearly with their separation, corresponding to a distance-independent force of attraction. This is thought to be due to the self-attraction of the gluonic field, causing it to collapse into a string or tube configuration with thickness of the order of 1 fm when the separation of the sources becomes much larger than this.

4.1.4.2 Cluster model

The cluster hadronization model [20] is based on the so-called preconfinement property of QCD which describes that at the evolution scales much less than the hard subprocess scale, $q << Q$, the partons in a shower are clustered in colorless groups with an invariant mass distribution that is independent of the nature and scale of the hard subprocess, depending only on q and the fundamental QCD scale Λ. It is then natural to identify these clusters at the hadronization scale Q_0 as proto-hadrons that decay into the observed final state hadrons.

4.1.5 Underlying event

In hadron collider events that contain a hard subprocess, there is extra hadron production that cannot be ascribed to showering from the colored partons participating in the subprocess. Furthermore this extra activity, known as the underlying event is greater than that in so-called minimum-bias event. The underlying event [20] is believed to arise from collisions between those partons in the incoming hadrons that to not directly participate in the hard subprocess.

4.1.6 Hadron and τ decays

Many of so-called primary hadrons, originating directly from string breaks and/or cluster decays, are unstable and undergo further decay, until a set of particles is obtained that can be considered stable on time scales relevant to the given measurement [20]. In case of tau leptons, it undergoes the further semi-leptonically decay to a tau neutrino and either hadron or lepton. The decay modeling can therefore have a significant impact on final particle yields and spectra, especially for the lowest-lying hadronic states, which receive the largest relative contributions from decays.

4.1.7 Validation and tuning

Validation in the context of Monte-Carlo generators means confronting a model with all relevant data that it claims to be able to describe [20]. It is essential that the validation is global, because the model should describe the underlying physics and not just parameterize the data, otherwise it would not have any predictive power. In this sense, validation is important for developing models as well as for debugging both code and physics models. Tuning means adjusting the free parameters of the model within their allowed ranges and as per the detector simulation to improbe the description of the relevant data.

4.1.8 Event generators

4.1.8.1 PYTHIA

The general-purpose event generator, PYTHIA [22], has probably been the most used generator for LHC, during the last 20 years. The PYTHIA program is a standard tool for the generation of high energy collisions, including a coherent set of physics models for the evolution from a few body hard process to a comples multi hadronic final state. PYTHIA treats an extensive list of hardcoded subprocesses, over 200, that can be switched on individually. These are mainly two body to one body decay, but no multiplicities higher than that. Consecutive resonance decays may, of course, lead to more final state particles, as well the parton showers. The subprocess cross-sections have to be convoluted with PDFs to obtain the event rates.

4.1.8.2 POWHEG

The other generators is POWHEG [23] used at CMS. The main idea of the POWHEG method is that the hardest emission (the one with the highest p_T)

is simulated according to the exact NLO cross-section. Then during the parton shower, the hardest emission is excluded and subsequent emissions are vetoed if they are harder than the hardest emission. The POWHEG method, compared with PYTHIA, provides a much better description of basic processes such as vector boson and Higgs production and is used explicitly to describe final states with low multiplicity such as inclusive W/Z production.

4.1.8.3 MadGraph

The MadGraph generator [24] has the ability to identify the relevant sub-processes of a given process. It generates both the amplitudes and the mappings needed for an efficient integration over the phase space. Once the events are generated, MadGraph is interfaced to PYTHIA for the parton shower and hadronization procedures.

4.1.9 K-factors

Next-to-leading order (LNO) cross-section calculations are imperative for experimental analysis at the LHC since some highly interested processes may involve large logarithms that need to be resumed or extra partonic processes may contribute only when going beyond the leading order (LO) approximation. Some of the information from a NLO calculation can be encapsulated in the K-factor given as the ratio of the NLO to LO cross-section for a given process, including the values of the renormalization and factorization scales, as well as PDFs used at LO and NLO. For example, in case of the $t\bar{t}b\bar{b}$ production in pp collision:

$$K = \frac{\sigma_{HO}(pp \to t\bar{t}b\bar{b})}{\sigma_{LO}(pp \to t\bar{t}b\bar{b})} \tag{4.1}$$

where HO stands for the higher order cross-section.

The QCD corrections to the transverse momentum, p_T, and rapidity distributions are also available in the case of vector-boson fusion and gluon-gluon fusion. In the latter case, that is the one of main interest for this work, the resummation of the large logarithms for the p_T distribution has been performed at the next-to-next-to-leading-logarithm (NNLL) accuracy. In the gluon-gluon fusion mechanism the calculation of the cross-section at next-to-next-to-leading order (NNLO) is also necessary.

4.2 Detector simulation

The complexity of the CMS detector requires very sophisticated simulation to properly reproduce the detector behavior in the presence of particles from proton collisions. The CMS detector simulation is based on the GEANT4 toolkit [25]. GEANT4 relies on the accurate detector description including the full geometry, the materials of the detecting devices and also taking into accounts the 'sensitive' parts (those furnished with a readout system) opposed to 'dead materials', to simulate the particle response. It takes as input particles from the event generator and then propagates them through the detector taking into account the measured magnetic field map (for charged particles) and any interactions between particle and material such as bremmstralung, multiple scattering and photon photon conversions. At the final stage, GEANT4 produces a set of simulated hits in the active material such as the energy loss of a particle in the sensitive volume of a detector. Subsequently, there is the digitization step to model the response of the detector readout electronics: signal collection and electronic effects are computed, noise is added and pile-up events are superimposed. The following step in the analysis is called reconstruction and can be applied independently from the origin of input data (simulation or real data).

4.3 Reconstruction of physics objects

In the reconstruction phase, collected collision information from all subdetectors of the CMS experiment, stored in 'raw' data format, is used to reconstruct high-level objects such as jets, tracks, vertices and leptons etc. It involves the primary step of building up the 'RecHits' collections corresponding to the different parts of the detector. RecHits objects contain several useful information such as energy releases, 3D-position and collected charge of the particle which is interacting with the sub-detector matter. RecHits collections are further used by high level reconstruction algorithms to reconstruct the physics objects. The reconstruction of the physics objects involved in this analysis is presented in this section.

4.3.1 Track reconstruction

Tracks are the fundamental objects need to be reconstructed as precisely as possible since the reconstruction of high level objects in highly dependent on it. In addition, the Particle Flow (PF) algorithm [28] is strongly dependent on sophisticated track reconstruction, which is used to reconstruct taus, jets and missing transverse energy, E_T^{miss}.

The track reconstruction at the CMS is performed by the combinatorial track finder (CTF) [32]. The reconstruction [33] starts from the hit formation by clustering the silicon pixel and strip tracker information where the hit position provides the seed to the track seed generators. Track seed generators are responsible for building up the initial track using the triplets or the pairs of hits. The constraint from the beam spot or a vertex is also taken into account. Further propagation of the initial track takes place which results in the association of one additional compatible hit corresponding to each silicon strip. The propagation continues until either the limit of the tracker is reached or no more compatible hits can be found. Moreover, in multiple seeds per track or multiple tracks per seed situations (which results in more than one reconstructed trajectories), the iterative reconstruction process is performed which can introduce some ambiguities. To remove the ambiguities, the fraction of hits that are shared between 2 trajectories is calculated using the formula

$$f_{shared} = \frac{N_{shared}}{min(N_1, n_2)} \tag{4.2}$$

where N_{shared}, N_1 and N_2 give the number of shared hits, number of hits of first and second track candidates, respectively. For $f_{shared} > 0.5$, the track with least number of hits is rejected. The track with the highest χ^2 is discarded for $N_1 = N_2$. In the final step, this collection of hits is fit to obtain the best estimate of the track parameters.

4.3.2 Vertex reconstruction

The reconstruction of the vertices in the event starts from the tracks collection. Prompt tracks are selected based on the transverse impact parameter along with uncertainty, number of hits, and the normalized track χ^2. The selected tracks are then clustered, using Deterministic Annealing (DA) clustering [34], in z-direction and the cluster is fit with an adaptive vertex fit [35], where tracks in the vertex are assigned a weight between 0 and 1 based on their compatiblity with the common vertex. The sum of weights roughly corresponds to the effective number of tracks accepted by the adaptive vertex fitter. This sum is directly related to the number of degrees of freedom assigned to the vertex through $N_{dof} = 2\Sigma w_i - 2$ for an unconstrained fit and $N_{dof} = 2\Sigma w_i$ for a fit with beam constraint. Tracks assigned to the (unconstrained) vertex because the weight of every track is always (slightly) smaller than 1.

Multiple vertices are possible and the vertices in the output are sorted according to the sum of the square of the transverse momenta of the tracks in the track cluster. The vertex corresponding to the highest sum is considered as primary vertex (PV). Rest are related to multiple pile-up interactions.

4.3.3 Muon reconstruction

Muons are reconstructed using the tracker and the muon detectors information. Tracks are first reconstructed independently in the inner tracker (tracker track) and in the muon system (standalone muon track) of the CMS experiment [36]. Based on these objects, two approaches are used for the muon reconstruction, given in Figure 4.2.

4.3.3.1 Standalone muon reconstruction

The first step for muon identification at the CMS is the reconstruction using the muon system only, to produce track segments or full tracks which can be matched to the tracks reconstructed in the inner detector. The algorithm used is hierarchical: segments are first reconstructed in the individual stations, and then combined to produce tracks.

- **DT segments**

  As the magnetic field is almost completely confined in the steel return joke of the magnet, with in each DT station the muon tracks are well approximated by straight lines. Segments are first reconstructed independently in the transverse and longitudinal planes ($r - \phi$ and $r - z$ respectively), through pattern matching and linear fitting of the hits, solving also the left/right ambiguities; such 2d segments provide a measurement of one coordinate and one angle.

  Afterwards, the two views are combined to produce full “4d” segments, which carry information about two position coordinates and two angles.

- **CSC**

  In each CSC layer, hits are reconstructed by combining the information from anode wires and cathode strips to obtain a 2d point in the transverse plane. Segments are then formed from hits in all six layers of each station.

- **Tracks**

  In order to reconstruct full muon tracks, the segments are extrapolated from one DT or CSC to the next taking into account the magnetic field and the energy loss and multiple scattering in the steel; if a compatible segment is found in the other station the two measurements are combined and the procedure is repeated for the other stations. Hits reconstructed in the RPC system are also included in the tracks at this stage. The beam spot position in the transverse plane is also used in the standalone muon reconstruction: combined with the measurement of the track angle at the entrance of the muon system, this information allows a muon p_T measurement exploiting the large bending power of the magnetic field inside the solenoid, and not just in the return joke.

4.3.3.2 Global muon reconstruction (outside-in)

For each standalone muon track, a matching tracker track is found by comparing parameters of the two tracks propagated onto a common surface. A global muon track is fitted combining hits from the tracker track and standalone muon track, using the Kalman filter technique [37]. At large transverse momenta, $p_T \geq 200$ GeV, the global muon fit can improve the momentum resolution compared to the tracker only fit [36][38]

4.3.3.3 Tracker muon reconstruction (inside-out)

In this approach, all tracker tracks with $p_T > 0.5$ GeV and total momentum $p > 2.5$ GeV are considered as possible muon candidates and are extrapolated to the muon system taking into account the magnetic field, the average expected energy losses, and multiple Coulomb scattering in the detector material. If at least one muon segment (i.e. a short track stub made of DT or CSC hits) matches the extrapolated track, the corresponding tracker track qualifies as a Tracker Muon. Track-to-segment matching is performed in a local (chamber) coordinate system, where local x is the best measured coordinate (in $r - \phi$ plane) and local y is the coordinate orthogonal to it. The extrapolated track and the segment are considered to be matched if the distance between them in local x is less than 3 cm or if the value of the pull for local x is less than 4, where the pull defined as the difference between the position of the matched segment and the position of the extrapolated track, divided by their combined uncertainties [36].

Tracker Muon reconstruction is more efficient than the Global Muon reconstruction at low momenta, $p \leq 5$ GeV/c, because it requires only a single muon segment in the muon system, whereas Global Muon reconstruction is designed to have high efficiency for muons penetrating through more than one muon station and typically requires segments in at least two muon stations. About 99% of muons produces in pp collisions within the geometrical acceptance of the muon system and having sufficiently high momentum are reconstructed either as a Global Muon or a Tracker Muon, and very often as both. Candidates found both by the Global Muon and the Tracker Muon approaches that share the same tracker track are merged into a single candidate.

For a small fraction of the cases ($\sim 1\%$) where both approaches fail, the track reconstructed at the muon chamber is considered as the muon candidate, so called 'standalone muon'. If the standalone muon tracks are not included in a global muon and share a muon segment with the tracker muon, they are merged with a Tracker Muon.

4.3.4 Electron reconstruction

Two complementary approaches are used at the CMS for electron reconstruction: 'ECAL driven' and 'Tracker driven' seeding.

4.3.4.1 ECAL driven Seeding

The ECAL driven altorithm starts from the reconstruction of ECAL 'superslusters' [39] of transverse energy $E_T > 4$ GeV that is optimized requirement for isolated electrons in the p_T range relevant for Z or W decays, down to $p_T \sim 5$ GeV. Supercluster is a group of one or more associated clusters of energy deposits in the ECAL constructed using an algorithm which takes into account their characteristic narrow width in the η coordinate and their characteristic spread in ϕ due to the bending in the magnetic field of electrons radiating in the tracker material. As a first filtering step, superclusters are matched to track seeds (pairs or triplets of hits) in the inner tracker layers, and electron tracks are built from these track seeds. Trajectories are reconstructed using a dedicated modeling of the electron energy loss and fitted with a Gaussian Sum Filter (GSF) [40]

4.3.4.2 Tracker driven Seeding

In the case of electrons in jets, pollution of the supercluster energy by particles produced near the electron degrades the energy measurement. Tracker driven algorithm helps in this case. The algorithm starts from all reconstructed tracks and electromagnetic clusters. From the inner track position, the bremstrahlung hypothesis is tested at the point on the calorimeter surface calculated by extrapolating a straight line from the track position and momentum vector at the corresponding detector layer, given in Figure 4.3. The process is repeated for all layers and a supercluster is defined by summing all linked electromagnetic cluster deposits.

Furthermore, a preselection is applied based on a multivariate analysis for candidates found only by the tracker driven seeding algorithm [28]. For candidates found by the ECAL driven seeding algorithm, the preselection is based on the matching between the GSF track and the supercluster in η and ϕ [41]. The few ECAL driven electron candidates ($\sim 1\%$ for isolated electrons) not accepted by these matching cuts but passing the multivariate preselection are also kept.

4.3.5 Particle flow reconstruction

The CMS particle-flow algorithm aims at an exclusive reconstruction and identification of all stable particles in the event, i.e. electrons, muons, photons, charged hadrons and neutral hadrons, by means of an optimized combination of information from all subdetectors. The algorithm is described in detail in reference [28], and information on its commissioning with early data are provided in references [29][30][31].

The core of the particle-flow reconstruction technique is the algorithm used to link the signals in the individual subdetectors, i.e. tracks or calorimetric clusters, whenever they can possibly originate from the same particle. The association used in the linking stage is purely geometrical: tracks are linked to calorimetric clusters if their trajectory intersects one of the calorimetric cells of the cluster; and likewise clusters in the ECAL preshower, ECAL and HCAL are linked if the cluster position measured in the finer granularity subdetector lies within the envelop of the cluster in the coarser granularity sbdetector. In order to account for uncertainties from multiple scattering in the track extrapolation and on the estimated position of the shower maximum in the calorimeters, a geometrical tolerance of the size of one calorimeter cell is included when defining links; this tolerance can also account for gaps and cracks in the calorimeters. By design, the linking algorithm is simple and robust, as it does not rely on the precise knowledge of the position recollecting bremsstrahlung clusters in the case of

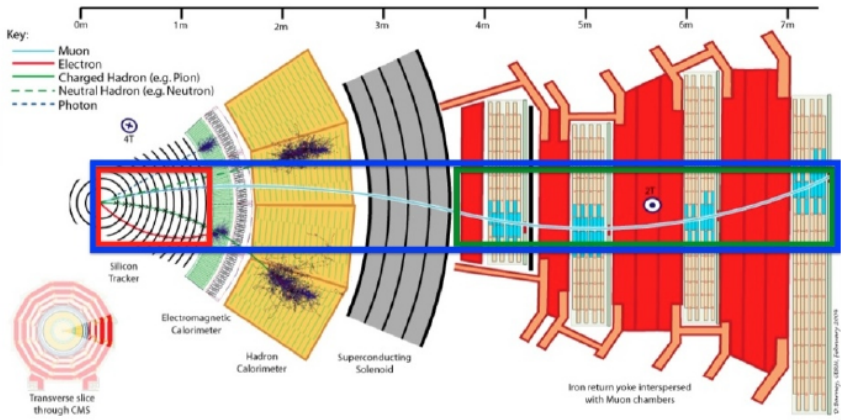


Figure 4.2: Reconstruction of muon objects at the CMS experiment. Tracker muon (red box), standalone muon (green box) and global muon (blue box) are shown.

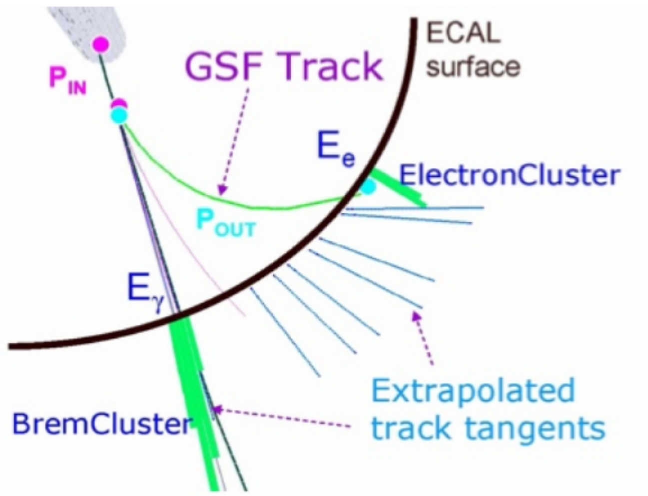


Figure 4.3: Reconstruction of electron objects at the CMS experiment.

electrons by considering tangents to electron trajectories at the crossing points with the tracker layers, as described in section 4.3.4.

Blocks of one or more linked objects are then processed to identify and reconstruct particle candidates. Isolated electrons and muons are selected first, and reconstructed using the dedicated algorithms developed for them, as described earlier in sections 4.3.3 and 4.3.4; similarly, non-isolated tracks which satisfy tight muon identification criteria are immediately identified as muons. Charged hadrons are identified as tracks in the inner tracker, normally linked to calorimetric deposits if the particle p_T is sufficient for the trajectory to reach the calorimeters. If the momentum measurements from the track and calorimeter are compatible, after accounting for non-linearities and zero suppression effects, the best energy determination is obtained as a combination of the two. If the track momentum significantly exceeds the measured calorimetric energy, the particle is identified as muon if it satisfies very loose muon identification criteria; otherwise, tight track quality requirements are applied to reject misreconstructed tracks. If instead an excess of calorimetric energy deposition is found with respect to the momentum of the associated track, or if the same cluster is linked to multiple ones, e.g. in the case of collimated hadronic jets, the residual energy is identified as a photon or a neutral hadron. Additional photons and neutral hadrons are also identified from calorimetric deposits not linked to any track.

4.3.6 Jet reconstruction

Four types of jets (from which the quark and gluon energies and directions are inferred) are reconstructed at the CMS, which differently combine individual contributions from sub-detectors to form the inputs to the jet clustering algorithm: calorimeter jets, Jet-Plus-Track (JPT) jets, Particle-Flow (PF) jets and track jets. The calorimeter jets are completely based on calorimeter information and JPT jets are the improvements to the calorimeter jets that improves the measurement by exploiting the associated tracks. Trach jets are reconstructed using only the well measured tracks. At the CMS, the mostly used jets are the

PF jets, which are reconstructed combining the information from all the sub-detectors. PF jets are used to reconstruct as well as identify the tau objects, and of higher interest from the point of view of the thesis work. The following text is describing the PF based reconstruction of jets.

The PF jet algorithm uses the PF event reconstruction [28], starts from the reconstructing and identifying all stable particles in the event such as photons, charged and neutral hadrons etc, with a thorough combination of all the CMS sub-detectors towards an optimal determination of their direction, energy and type. Charged hadrons, in particular, are reconstructed from the tracks in the central tracker. Everngy clusters in the ECAL and HCAL lead to the reconstruction of the photons and the neutral hadrons respectively. Clusters separated from the extrapolated position of tracks in the HCAL constitute a clear signature of these neutral particles. Furthermore, the precise study of the excess in the calorimeter energy ensures the overlapping to the charged particles. These individual particles are then clusterd into jets using the 'anti-k_T' algorithm with the size parameter $R = 0.5$ in $\eta - \phi$ space [42]. The PF algorithm gives better performance for momentum and spatial resolutions with respect to calorimeter jets, as it uses the tracking detectors and of the excellent granularity of the ECAL. The PF algorithm therefore allows to resolve and precisely measure the charged hadrons and photons inside jets that constitute the $\sim 90\%$ of the jet energy, given in Figure **??**. While the lose of 10% corresponds to neutral hadrons that are affected by the poor HCAL resolution, especially for low neutral hadron E_T.

4.3.7 E_T^{miss} reconstruction

The missing transverse energy [43] used in this analysis has been computed using PF algorithm [28]. It is estimated in terms of momentum imbalance in the plane perpendicular to the beam direction. The missing transverse momentum vector is computed as the opposite of the transverse momentum sum of all PF particles reconstructed in the event.

$$\vec{E}_T^{miss} = -\sum_i \vec{P}_T^i \tag{4.3}$$

The magnitude of this vector is the missing transverse energy E_T^{miss}.

4.3.8 b-tagging algorithm

In the decay of top quark pairs jets that arise from bottom quark hadronization and decay (usually denoted as b-jets) are present. The ability to accurately identify b-jets is crucial for reducing the otherwise overwhelming background to this channel, which mainly comes from processes involving jets from gluons and u, d, s quark fragmentation (light flavor) and also from charm quark hadronization.

The properties of bottom quarks are used to identify the hadron jets into which they fragment. These properties include their hard fragmentation functions due to the relatively large mass and the long lifetime of the heavy flavor hadrons. Their semi-leptonic decays can be exploited as well. The CMS detector, with its precise charged particle tracking system, robust lepton identification and finely segmented calorimetry, is well matched to the task of b-jet identification (b-tagging).

There are several b-tagging algorithm ("taggers") developed by the CMS. Each produces as output a numerical "discriminator", which can be used to select jets according to the best analysis dependent operating point.

The b-jet algorithm in the CMS relies on the long lifetime, high mass and large momentum fraction of b hadrons produced in b-quark jets, as well as on the presence of soft leptons from semi-leptonic b decays.

The b-tagging algorithms rely on the measure of two main variables:

- The impact parameter (IP). It is defined as the distance in space between tracks (linearized around the point of closest approach to the primary vertex: PV) and the jet axis at the point of closest approach th the primary interaction vertex (figure 4.4). To account for resolution effects, we used the IP divided by its error: IP/σ(IP). The impact parameter can be signed as positive/negative if the associated tracks are produced downstream/upstream with respect to the primary interaction vertex.

- The measure of the position of the secondary vertex (SV). It is the point where the b-hadron decays (figure 4.4). It is possible to measure the position of this point thanks to the high resolution of the CMS tracker system. The secondary vertex reconstruction is performed using adaptive vertex fitter algorithm [44]. The resulting list of vertices is then subject to a cleaning procedure, rejecting SV candidates that share 65% or more of their tracks with the PV

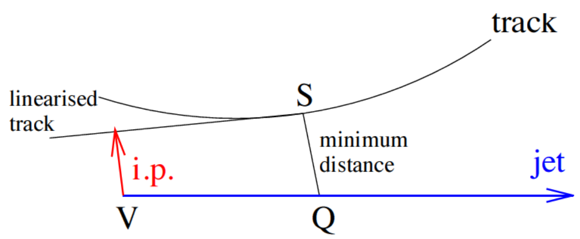


Figure 4.4: Geometric meaning of the impact parameter (IP). The sign is positive/negative when the angle between the impact parameter direction and the jet axis is smaller/larger than 90°. The primary vertex (PV) is shown in green while the secondary vertex, the point where the b hadron decays, in red.

Starting from these variables the different algorithms provide the discriminators, i.e. output variables on which the users can cut on to select b-jets in different regions of the efficiency versus purity phase space. Varying the cuts on the discriminator we obtain different efficiencies of the tagger. We establish standard operating points as, loose (L), medium (M) and tight (T), being the value at which the tagging of light jets is estimated from MC to be 10%, 1% or 0.1%, respectively, for jets with p_T of about 80 GeV.

Chapter 5

Measurement of $\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj)$

5.1 Monte-Carlo samples

The complete set of MC simulation samples used for the analysis are summarized in Table 5.1 along with the informations about cross-section values and the event generators. The Summer12 official samples have been used for this analysis.

5.1.0.1 Signal samples

$t\bar{t}$ decay into dileotonic process normalization in this analysis, the approximate NNLO $t\bar{t}$ production cross-section, 23.6 pb at 172.5 GeV. The acceptance for selecting $t\bar{t}b\bar{b}$ and $t\bar{t}jj$ events is modeled using the MadGraph event generator with matrix elements (ME) providing up to four additional partons including b-quark.

5.1.0.2 Background samples

The W+jets and z+jets samples are simulated in MadGraph. The total inclusive NNLO cross-section is calculated with the FEWZ program [26]. The diboson productions are simulated with PYTHIA. The single top production is simulated in POWHEG with a cross-section calculated at approximate NNLO [27]. The QCD background contribution is obtained directly from data.

Sample	$\sigma \times Br$ [pb]	$\mathcal{L}$ [pb^{-1}]
/TTJets_FullLeptMGDecays_8TeV-madgraph	23.6196	1.63351e+07
/TTJets_MassiveBinDECAY_TuneZ2star_8TeV-madgraph-tauola	23.6196	6.92375e+06
/T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola	11.1	497658
/Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola	11.1	493460
/DYJetsToLL_M-10To50filter_8TeV-madgraph	860	7.13222e+06
/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball	3503.17	3.04344e+07
/WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball	36257.2	5.718e+07
/WW_TuneZ2star_8TeV_pythia6_tauola	54.8	1.00004e+07
/WZ_TuneZ2star_8TeV_pythia6_tauola	32.3	1.00003e+07
/ZZ_TuneZ2star_8TeV_pythia6_tauola	7.7	9.79585e+06

Table 5.1: Monte-Carlo simulation samples used for the signal and background process.

5.2 Data samples

This analysis is based on the complete data set collected with the CMS detector at the LHC in 2012 corresponding to the run range 190456–208686. DoubleMu, DoubleElectron and MuEG primary dataset is used. The dataset name and corresponding run range are listed in the Table 5.2. The good runs and luminosity blocks flagged defined by CMS data quality and physics validation teams are used in this analysis. The total integrated luminosity is calculated to be 19.6 fb^{-1}.

The triggers used which triggers are listed in Table 5.3. Events required to pass the muon/electron trigger are taken from the DoubleMu/DoubleElectron primary dataset. In the $e^{\pm}\mu^{\mp}$ final state, events are taken from the MuEG primary dataset and required to pass the cross trigger.

5.3 Signal definition

The $t\bar{t}b\bar{b}$ signal is defined at generator level in a phase space region with two leptons and at least 4 jets. The leptons from W boson in top decay are selected with transverse momentum, $p_T > 20$ GeV/c and absolute pseudo-rapidity, $|\eta| < 2.5$. In addition, the leptons from tau decay are included. Jets are obtained recombining all final state particles excluding neutrinos from resonances such as W and Z bosons with the anti-k_T clustering algorithm [] with clustering parameter $\Delta R = 0.5$. Thus summarize

$$p_T > 20GeV/c, |\eta| < 2.5, \Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} > 0.5 \tag{5.1}$$

where ΔR represents the distance

At particle level, the b and c-jets in this analysis are defined with gen-jets as following:

- ***b*-jet**

  The gen-jet, "ak5GenJets" was used. This jet excludes only neutrinos from resonances such as W and Z bosons. From this gen-jet collection,

run range	dataset
190456-193621	/DoubleMu/Run2012A-13Jul2012-v1/AOD
	/DoubleElectron/Run2012A-13Jul2012-v1/AOD
	/MuEG/Run2012A-13Jul2012-v1/AOD
190782-190949	/DoubleMu/Run2012A-recover-06Aug2012-v1/AOD
	/DoubleElectron/Run2012A-recover-06Aug2012-v1/AOD
	/MuEG/Run2012A-recover-06Aug2012-v1/AOD
193834-196531	/DoubleMu/Run2012B-13Jul2012-v4/AOD
	/DoubleElectron/Run2012B-Jul132012-v4/AOD
	/MuEG/Run2012B-13Jul2012-v4/AOD
198022-198523	/DoubleMu/Run2012C-24Aug2012-v1/AOD
	/DoubleElectron/Run2012C-24Aug2012-v1/AOD
	/MuEG/Run2012C-24Aug2012-v1/AOD
198934-203002	/DoubleMu/Run2012C-PromptReco-v2/AOD
	/DoubleElectron/Run2012C-PromptReco-v2/AOD
	/MuEG/Run2012C-PromptReco-v2/AOD
201191-201191	/DoubleMu/Run2012C-EcalRecover_11Dec2012-v1/AOD
	/DoubleElectron/Run2012C-EcalRecover_11Dec2012-v1/AOD
	/MuEG/Run2012C-EcalRecover_11Dec2012-v1/AOD
203773-208686	/DoubleMu/Run2012D-PromptReco-v1/AOD
	/DoubleElectron/Run2012D-PromptReco-v1/AOD
	/MuEG/Run2012D-PromptReco-v1/AOD

Table 5.2: The dataset name and corresponding run range.

the b-jets are identified by looping over the jet constituents and looking for ancestors from B hadron. Furthermore, if two gen-jets com from the same B hadron, the closest gen-jet with ΔR is selected as b-jet.

- ***c*-jet**

  The c-jet id identified in particle level jet same as b-jet but matching to C hadrons. However, if the gen-jet is matched to B hadrons already, then this gen-jet is not considered as c-jet.

In this analysis, we define the $t\bar{t}jj$ process at generator requiring at least four jets and at least two b-jets in dilepton decay mode and the $t\bar{t}b\bar{b}$ process is required to have two additional b-jets and $t\bar{t}c\bar{c}$ to have at least two 4c4-jets. The remaining events are considered as $t\bar{t}LF$ events, where the extra two jets are from light flavor (LF) such as gluon, u, d or s quark.

5.4 Lepton identification and isolation

5.4.1 Muon identification

The reconstructed muons in Sectionsec:mureco collection contains a significant amount of (un-decayed) charged hadrons, such as kaons and pions, which are mis-identified like muons. In order to have a pure sample of muons, the identification requirements must be applied. The particle flow (PF) identification has been used for the muon identification in this analysis, which is based on the PF event reconstruction [28]. The muons are required to be global or tracker muons as well. The PF identification uses the measurement of energy released by the tracks in the calorimeter to identify the muons inside the jets with high efficiency and low mis-identification rate.

Each global muon gives rise to a PF identified muon, if its combined momentum of all the PF elements is compatible with that determined from the sole tracker within three standard deviations.

On the other hand, the other muons which have a track momentum significantly larger than the corresponding energy deposit in the calorimeter thereby making them incompatible with a charged hadron hypothesis, can also be recovered. Firstly, all the global muons, not already selected by the algorithm and for which an estiate of the momentum exists with a precision better than 25%, are treated as PF identified muons. The redundancy of the measurements in the tracker and the calorimeters thus allows a few more muons to be found without increasing the mis-identification rate. This redundancy is further exploited by progressively removing the tracks from the block, ordered according to their measured p_T uncertainty. The process stops when either all the tracks with the p_T uncertainty in excess of 1 GeV have been examined, or the removal of a track would render the total track momentum smaller than the calibrated calorimetric energy. Less than 0.3% of the tracks are concerned by this procedure.

Furthermore, to ensure the high efficiency of the reconstruction in the tracker acceptance, a preselection is applied requiring the $p_T > 20$ GeV/c and $|\eta| < 2.5$. In addition, other sources of muon-like signature in the detector are the cosmic, decay-in-flight and punch-through muons. The decay-in-flight muons are the real muons which originate from the decay of kaons, and pions and give good hits in the muon chambers. On the other hand, the punch-through muons are not real muons instead are the particles (hadrons) which do not undergo nuclear interactions upstream of the muon system. The selection steps are as following:

- The global track should have at least one good muon chamber hit, to suppress the hadronic punch-through and the decay-in-flight muons.
- The global track should match to at least two muon stations, to suppress the punch-through and the accidental track-to-segment matches. It also provides the consistency with the muon trigger, which requires the segments in at least two muon stations to obtain a meaningful estimate of the muon p_T. Moreover, it ensures the muons arbitration i.e. a track is uniquely associated to a muon segment.
- The tracker track should fire at least five tracker layers, to suppress the decay-in-flight muons. It also guarantee a good p_T measurement.

5.4.2 Electron identification

After the reconstruction of electrons in Section 4.3.4, a further set of requirements is introduced to distinguish between the prompt electrons and the charged pions mis-identified like the electrons and the electrons produced from photon conversions. The main handles on reducing the mis-identified electron rate are the angular difference between the track and the supercluster, the ratio of HCAL to ECAL energy associated with the supercluster, and the ECAL shower shape. The identification of electrons relies on a Boosted Decision Tree (BDT) multivariate technique. In this analysis, we used this MVA electrons with BDT discriminator is larger than 0.5 which MVA electron is means Multi Variate Analysis trained [46] electron. In the MVA training twenty variables have been used, and this values are listed as following:

5.4.2.1 track-ECAL matching variables:

- E_{sc}/E_{in} - the ratio of the supercluster energy and the measured electron track momentum atthe innermost track position.
- E_{ϵ}/P_{out} - the ratio of the energy of the closest cluster to the electron track extrapolation to ECAL and the measured electron track momentum at the outermost track position.
- $\Delta\eta_{in}$ - the difference in η between the electron track and the supercluster at the vertex, and at the calorimeter surface.
- $\Delta\phi_{in}$ - the difference in ϕ between the electron track and the supercluster at the vertex.
- $IoEmIoP$ = (1/energy of supercluster)-(1/momentum of electron).

5.4.2.2 electron-pion discrimination variables, based on the calorimeter energy and the shower shape:

- H/E - the ratio of the supercluster HCAL energy to the ECAL energy.

- the ratio of the pre-shower energy and the uncorrected energy of the supercluster.
- $\sigma_{i\eta i\eta}$ - the RMS of energy along the η direction within the supercluser.
- $\sigma_{i\phi i\phi}$ - the RMS of energy along the ϕ direction within the supercluster.
- the width of the supercluster in the η and ϕ directions.
- $1 - E_{1\times5}/E_{5\times5}$ = 1-(energy inside 1 $\times$ 5 grid of crystals in $\eta \times \phi$ around the seed crystal)/(energy inside 5 $\times$ 5 grid of crystal in $\eta \times \phi$ around the seed crystal).
- $R9$ - the ratio of the energy in the 3 $\times$ 3 grid of crystals around the seed crystal to the uncorrected energy of the supercluster.

5.4.2.3 pure tracking observables for further improvement of electron-pion discrimination:

- the measured bremsstrahlung fraction from electron track: (measured electron track momentum at the innermost track position - measured electron track momentum at the outermost track position)/measured electron track momentum at the innermost track position.
- the normalized χ^2 of the closest CTF track to the electron, and of the electron track.
- the number of track layers fired by the closest CTF track to the electron.

5.4.3 Lepton isolation

In the LHC environment, a large number of particles are produced in the high energy pp collision. So the measurement of the isolated behavior of the objects is very challenging at the CMS. Muons and electrons are typically well isolated, while leptons from heavy flavor decays and decays in flights are expected to be poorly isolated due to be inside the jets. Moreover, muons and electrons coming from the decay of taus are also expected to be isolated in the detector. Therefore, the proper computation of the lepton isolation is a must.

The PF based approach has been used for the isolation computations for both the electrons and muons. The tracker and the calorimeter information of PF particles have been used to calculate the isolation deposits around the muons and electrons, so-called PF combined isolation. All the charged particles are considered in the isolation calculations, while photons and neutral hadrons are required to have $E_T > 0.5 GeV$ to be considered in the isolation sum. When the calculation isolation the $\Delta R < 0.3$ with respect to the lepton axis cut have to applied where $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$. The p_T weighted PF combined isolation is used in the analysis given by the following formula:

$$Iso_{rel}^{PF} = \frac{\sum(p_T^{CH} + p_T^{NH} + p_T^{PH})}{p_T} \tag{5.2}$$

where CH is the charged hadron, NH is neutral hadron and PH is photon.

The isolation variables are sensitive by pileup (PU) effect. Pileup leads to the increase in mean energy deposited in the detector which further leads to the rise of the mean isolation values. Thus, the efficiency of a cut on isolation variables strongly depends on pileup conditions. The pileup effect is observed to be stronger in the calorimeters and quite feeble in the tracking system. This is due to the requirement that the tracks contributing to the isolation cone originate from a common vertex. In order to have a pileup robust analysis, the isolation variable has to be corrected. Among several correction methods, the one using the FastJet [47][48] energy density ρ in the event has been chosen to estimate the mean PU contribution within the isolation cone of a lepton for PF isolation. A ρ variable is defined for each jet in a given event and the median of the ρ distribution for each event is taken. The correction to the isolation variable is then applied according to the formula:

$$\frac{\sum p_T^{CH} + MAX\left(0.0, \sum p_T^{NH} + \sum p_T^{PH} - \rho \cdot A_{eff}(NH + PH)\right)}{p_T^{electrons}} \tag{5.3}$$

where effective area A_{eff} is the geometric area of the isolation cone times a correction factor which accounts for residual dependence of the isolation on pileup as a function of pseudo-rapidity. This effective area is shown in Table 5.4.

In muon case, $\Delta\beta$ corrected isolation is used instead of ρ correction. The $\Delta\beta$ corrected isolation formula is following:

$$\frac{\sum p_T^{CH} + MAX\left(0.0, \sum p_T^{NH} + \sum p_T^{PH} - 0.5 \sum p_T^{PU\ CH}\right)}{p_T^{muons}} \tag{5.4}$$

where $PU\ CH$ is charged hadron considered pileup effect.

5.4.4 Lepton efficiency

In this analysis, the lepton selection cut is following, details for this selection is shown in Section 5.4.1–5.4.3. For muons:

- particle-flow muon
- global muon or tracker muon
- $p_T > 20$ GeV/c
- $|\eta| < 2.4$
- particle-based relative isolation < 0.15 when $\Delta R = 0.3$

and for electrons:

- particle-flow electron
- MVA > 0.5
- $p_T > 20$ GeV/c
- $|\eta| < 2.5$
- impact parameter, $d_{xy} < 0.4$ cm
- particle-based relative isolation < 0.15 when $\Delta R = 0.3$

Mode	Trigger selection
ee	HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL
	_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v*
$\mu\mu$	HLT_Mu17_Mu8_v*
	HLT_Mu18_TkMu8_v*
$e\mu$	HLT_Mu17_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v*
	HLT_Mu8_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v*

Table 5.3: Used trigger path for each decay mode.

$\|\eta\|$	$A_{eff}(NH + PH)$
$0.0 < \|\eta\| < 1.0$	0.13 ± 0.001
$1.0 < \|\eta\| < 1.479$	0.14 ± 0.002
$1.479 < \|\eta\| < 2.0$	0.07 ± 0.001
$2.0 < \|\eta\| < 2.2$	0.09 ± 0.001
$2.2 < \|\eta\| < 2.3$	0.11 ± 0.002
$2.3 < \|\eta\| < 2.4$	0.11 ± 0.003
$\|\eta\| > 2.4$	0.14 ± 0.004

Table 5.4: Effective area estimated on $Z \to ee$ in data.

After lepton selection cut, the efficiency is measured for this leptons. In muon case, the efficiency values are measured to be in 91.3% to 98.0% for the real data, 92.0% to 97.5% for the MC as shown in the Figure 5.1 and 5.2 and scale factor is calculated in the Table 5.5. The definition of scale factor is: $\epsilon_{data}/\epsilon_{MC}$.

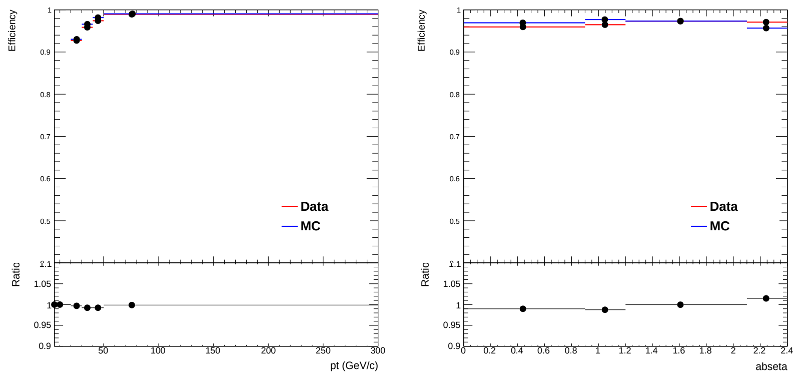


Figure 5.1: Muon efficiency as a function of p_T and $|\eta|$.

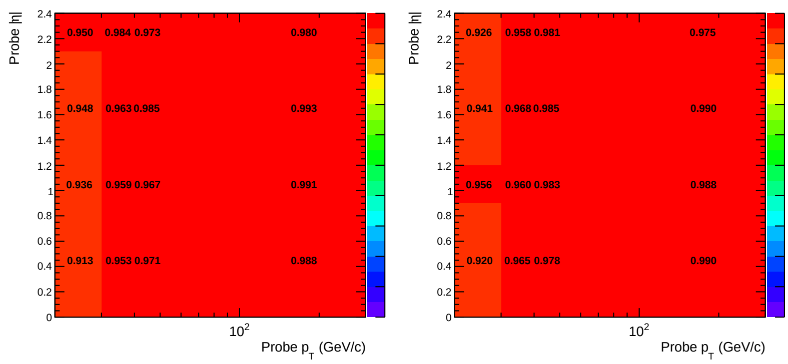


Figure 5.2: 2 dimensional muon efficiency for the real data (left) and the Monte-Carlo simulation (right).

In electron case, the efficiency values are measured to be in 72.8% to 95.8% for the real data, 82% to 97.4% for the MC as shown in the Figure 5.3 and 5.4 and scale factor is calculated in the Table 5.6. The scale factor is defined as $\epsilon_{data}/\epsilon_{MC}$ also.

	$20 < p_T < 30$	$30 < p_T < 40$	$40 < p_T < 50$	$50 < p_T < 300$
$0.0 < \|\eta\| < 0.9$	$0.980^{+0.014}_{-0.007}$	$0.988^{+0.009}_{-0.019}$	$0.982^{+0.002}_{-0.004}$	$0.992^{+0.002}_{-0.002}$
$0.9 < \|\eta\| < 1.2$	$0.972^{+0.009}_{-0.011}$	$0.994^{+0.007}_{-0.009}$	$0.979^{+0.008}_{-0.010}$	$0.988^{+0.004}_{-0.014}$
$1.2 < \|\eta\| < 2.1$	$0.996^{+0.007}_{-0.013}$	$0.982^{+0.007}_{-0.012}$	$0.986^{+0.006}_{-0.007}$	$1.000^{+0.004}_{-0.010}$
$2.1 < \|\eta\| < 2.4$	$1.019^{+0.016}_{-0.030}$	$1.018^{+0.018}_{-0.014}$	$1.000^{+0.004}_{-0.009}$	$1.024^{+0.010}_{-0.017}$

Table 5.5: The scale factor for muon selection

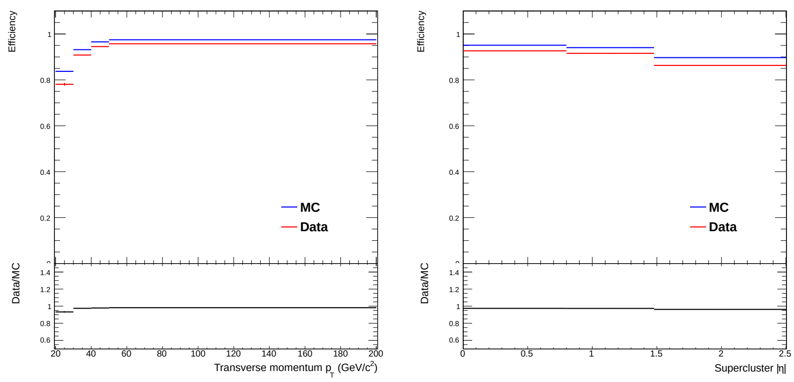


Figure 5.3: Electron efficiency as a function of p_T and $|\eta|$.

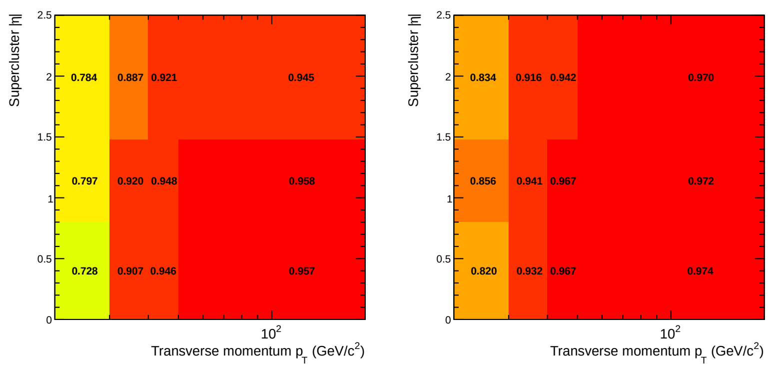


Figure 5.4: 2 dimensional electron efficiency for the real data (left) and the Monte-Carlo simulation (right).

5.5 Jet identification

Top quark almost decays into a b-quark in the standard model. There could be jets from the initial and final state radiation, which are soft enough to be removed by the kinematic requirement. The particle-flow jets are used which is described in section 4.3.5 and 4.3.6. The tau jets are not excluded from this jet collection. The momentum of jets is corrected with L1FastJet and relative L3 and absolute L3 corrections. The residual correction for different jet energy scale between data and MC is also applied only for data. Following selections are required

- particle-flow jet
- $p_T > 30$ GeV/c^2
- $|\eta| < 2.5$

5.5.1 b-jet identification

In this analysis, we used combined secondary vertex (CSV) algorithm for b-tagging that CSV algorithm is combined information of impact parameter and secondary vertex. These algorithms are described in section 4.3.8 respectively. By using these additional variables, the CSV algorithm provides discrimination even when secondary vertex is not found, so the maximum possible b-tagging efficiency is not limited by the secondary vertex reconstruction efficiency. In many cases, tracks with an IP significance > 2 can be combined in a so-called pseudo vertex, allowing for the computation of a subset of secondary vertex based quantities even without an actual vertex fit. The figure 5.5 is shown CSV discriminator distribution. In this figure, 0.898 point, so-called tight working point (CSVT), is used in this analysis for reduce uncertainty. discriminator value is 0.898.

	$20 < p_T < 30$	$30 < p_T < 40$	$40 < p_T < 50$	$50 < p_T < 150$
$0.00 < \|\eta\| < 0.80$	$0.887^{+0.012}_{-0.013}$	$0.973^{+0.002}_{-0.000}$	$0.978^{+0.001}_{-0.001}$	$0.982^{+0.000}_{-0.000}$
$0.80 < \|\eta\| < 1.48$	$0.932^{+0.007}_{-0.007}$	$0.978^{+0.001}_{-0.001}$	$0.980^{+0.001}_{-0.001}$	$0.985^{+0.003}_{-0.003}$
$1.48 < \|\eta\| < 2.50$	$0.940^{+0.005}_{-0.005}$	$0.968^{+0.002}_{-0.002}$	$0.978^{+0.001}_{-0.001}$	$0.974^{+0.013}_{-0.013}$

Table 5.6: The scale factor for electron selection.

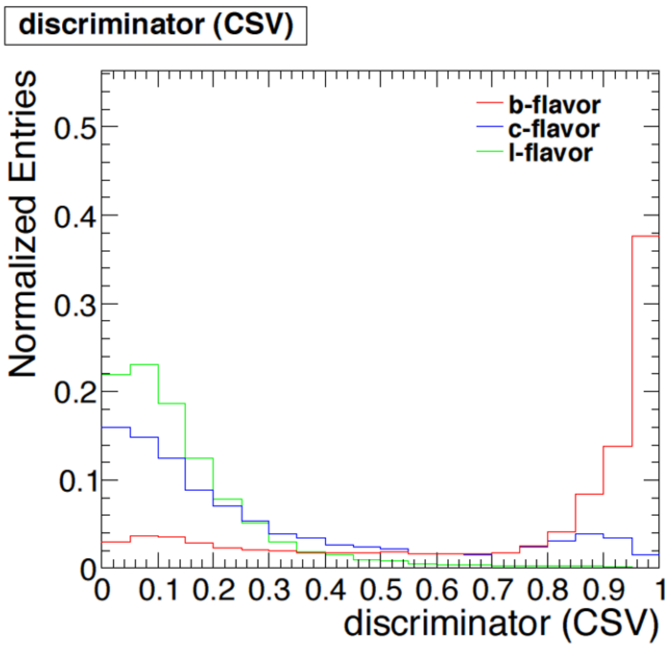


Figure 5.5: Distribution of CSV discriminator for jet flavor.

5.6 E_T^{miss} identification

The missing transverse energy (E_T^{miss}) indicates the presence of invisible particles, which are neutrinos in this analysis. W bosons decay to lepton with neutrino in the $t\bar{t}$ signal while Z/γ^* and QCD as the main backgrounds do not decay with neutrinos naturally. Therefore, the E_T^{miss} from these backgrounds are expected to be very small or ideally zero. The QCD background contribution is negligible once two isolated leptons are required. In order to suppress the remaining background contribution from Z/γ^*, the missing transverse energy is required as the final selection.

5.7 Event selection

Since the early collisions, the scrapping evets caused by beam halo or beam gas exist. For remove the scrapping events it must be required that at least 25% of high purity tracks if there are more than ten tracks in the event. The noise found in hadronic calorimeters is also removed. The event should have at least one good primary vertex which is required to pass following selections.

Further selections are applied which is following the 2012 TOP PAG reference analysis selection [45]. The event selection proceeds in five step, described in the following paragraphs.

- **S1: Low invariant mass veto, 2 isolated and opposite signed leptons**

  In this analysis, we consider only dileptonic channel, and this two leptons are decayed from W^+ and W^- respectively. Therefore at least two isolated lepton is required in the event and leptons must have opposite sign charge. And it is applied that low invariant mass region veto cut because of to remove dilepton event from Υ or J/ψ.

- **S2: Z mass veto**

  We also applied Z mass region veto cut because of to remove dilepton event from Z.

- **S3: Low missing transverse energy veto**

  Two neutrinos are appeared in this $W^{\pm} \to l\nu$ channel. Thus missing transverse energy is larger relatively.

- **S4: 4 jets requirement**

  In $t\bar{t}b\bar{b}$ event, there are at least four jets. Tow jets are from $t\bar{t}$ ant the other two jets are from $b\bar{b}$.

- **S5: 2 b-jet requirement**

  Two jets from t quark are almost b-jet, so there are at least two b-jet in the event.

5.8 Background estimation

5.8.1 Drell-Yan background

The Drell-Yan background outside veto window was estimated directly from data. The method is followed by Ref. [49]. The ratio $R_{out/in}$ of events outside and inside of the Z mass region is obtained from the simulation with pileup at each selections (S2-S5). The expected contribution, N_{out} outside of the Z mass region from Z/γ^* events is obtained by multiplying this ratio to the number of events, N_{in} inside of the Z mass region in data subtracting backgrounds. The background in data inside the Z mass region was estimated using the events, $N_{in}^{e^{\pm}\mu^{\mp}}$ in the $e\mu$ final state considering efficiency difference, k between electrons and muons. Following formula is used for the estimation in the $\mu^+\mu^-$ final state. A similar formula was used for the prediction of the contribution in the e^+e^- final state.

$$N_{out}^{\mu^+\mu^-} = R_{out/in}^{\mu^+\mu^-}\left(N_{in}^{\mu^+\mu^-} - 0.5N_{in}^{e^{\pm}\mu^{\mp}}k_{\mu\mu}\right), k_{\mu\mu} = \sqrt{\frac{N_{in}^{\mu^+\mu^-,S1}}{N_{in}^{e^+e^-,S1}}} \tag{5.5}$$

The scale factors used for this estimation are from the Ref. [50] and summarized in Table 5.7. The estimated number from data was used as Drell-Yan background.

5.8.2 QCD background

The QCD background contribution is estimated directly from the data. It has been shown that the QCD background is mostly from semileptonic decays of $b\bar{b}$, where a cascade decay of one of the b quarks can lead to a like-sign lepton pair. All estimations of QCD background are made at each selection. Differential distributions of QCD background are obtained by inverting the isolation cut (0.25) for both leptons. $N_{data}^{non-iso}$ is defined as the number of the events in this selection: the number of events at each selection with the two leptons not fulfilling the isolation criteria. The $N_{data}^{non-iso}$ sample. We extrapolate the QCD estimate from the non-isolated region to the isolated region of the phase space by applying a normalization factor f_{QCD}. The number of like-sign QCD events $N_{QCD}^{\pm\pm}$ is estimated from the excess of like-sign events $N_{data}^{\pm\pm}$ above the expected contribution $N_{MC}^{\pm\pm}$ from all standard model backgrounds except QCD: $N_{QCD}^{\pm\pm} = N_{data}^{\pm\pm} - N_{MC}^{\pm\pm}$. Here $N_{data}^{\pm\pm}$ and $N_{MC}^{\pm\pm}$ are defined as the number of data and MC events after each selection in like-sign events, respectively. Figure **??** shows the dileptonic invariant mass having same sign in the non-isolation region and the excess of like-sign events over the all standard model backgrounds except QCD sample is around 98%. The normalization factor f_{QCD} for the QCD sample is obtained by taking the ratio of the number of like-sign QCD events to the number of non-isolated events in the like-sign sample ($N_{data}^{non-iso,\pm\pm}$). The normalization factor f_{QCD} id defined as

$$f_{QCD} = \frac{N_{data}^{\pm\pm} - N_{MC}^{\pm\pm}}{N_{data}^{non-iso,\pm\pm}} = \frac{N_{QCD}^{\pm\pm}}{N_{data}^{non-iso,\pm\pm}} \tag{5.6}$$

The QCD background contribution before isolation requirement is obtained by multiplying f_{QCD} to the number of non-isolated events, $N_{data}^{non-iso,\pm\mp}$:

$$N_{QCD}^{\pm\mp} = f_{QCD} \times N_{data}^{non-iso,\pm\mp} \tag{5.7}$$

The isolation efficiency is obtained in the like-sign sample. Assuming that events remaining after subtracting Monte-Carlo from data are all QCD events, the isolation efficiency for QCD is defined as

$$\epsilon_{iso} = \frac{N^{iso,\pm\pm}_{data} - N^{iso,\pm\pm}_{MC}}{N^{\pm\pm}_{data} - N^{\pm\pm}_{MC}} \tag{5.8}$$

where $N^{iso,\pm\pm}_{data}$ and $N^{iso,\pm\pm}_{MC}$ are the number of data and MC events after the isolation requirement in like-sign sample, respectively. To get an estimation of the QCD contribution after applying the isolatioin requirement, the isolation efficiency from QCD is multiplied to the QCD contribution. The normalization factor and isolation scale factor are summarized in Table 5.8.

5.9 Comparison data and Monte-Carlo

The distribution of the di-lepton invariant mass, lepton transverse momentum and lepton pseudo-rapidity after the lepton identification, isolation and opposite sign pair requirement are shown in the Figure 5.6–5.10. When the two leptons are required to be isolated, most of the background from QCD are removed. The distributions of missing transverse energy after Z invariant mass veto are shown in the Figure 5.11. The distributions of jet multiplicity after missing transverse energy veto are shown in the Figure 5.12. Figure 5.13–5.16 shows the distributions of the jet transverse momentum after at least four jet requirement. The kinematic jet p_T distributions of *ttbb* and *ttll* are shown to be similar to each other. After four jet selection, the distributions of b-jet multiplicity of each working point are shown in Figure 5.17. The remaining number of events at each step after the di-lepton selection (S1) is shown in Table **??**–**??**. In the $e\mu$ final state, the Z mass veto and E_T^{miss} requirement are not applied for the final cross section measurement.

Selection	S2	S3	S4	S5
e^+e^-				
$R^{e^+e^-}_{out/in}$	0.095	0.099	0.141	0.111
Data/MC	1.052	1.346	1.365	1.888
$\mu^+\mu^-$				
$R^{\mu^+\mu^-}_{out/in}$	0.100	0.104	0.155	0.134
Data/MC	1.028	1.303	1.288	1.724

Table 5.7: The prediction from Drell-Yan in simulation compared with estimated contribution in data at 8 TeV

	e^+e^-	$\mu^+\mu^-$	$e^\pm\mu^\mp$
f_{QCD}	2.574	1.131	1.501
e_{iso}	0.238	0.025	0.041

Table 5.8: The normalization factor and isolation efficiency for multi jet events after each selection in the like-sign sample in ee, $\mu\mu$ and $e\mu$ decay. The data events remaining after subtracting Monte-Carlo backgrounds are considered to be due to QCD.

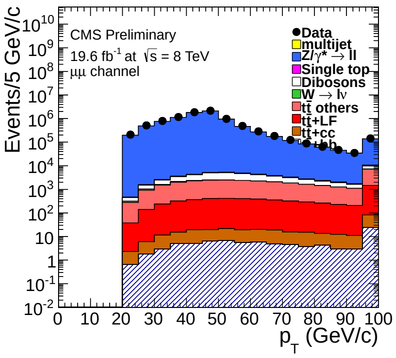
Final state(CSVT)	e^+e^-	$\mu^+\mu^-$	$e^\pm\mu^\mp$	All
$t\bar{t}+b\bar{b}$	11.5 ± 0.63	21.4 ± 0.81	45 ± 1.19	77.9 ± 1.57
$t\bar{t}+c\bar{c}$	15.2 ± 0.68	26.6 ± 0.87	53.9 ± 1.27	95.7 ± 1.69
$t\bar{t}+LF$	338 ± 3.21	481 ± 3.71	1102 ± 5.71	1921 ± 7.53
$t\bar{t}$ others	37 ± 1.81	44.6 ± 1.73	104 ± 2.01	185.6 ± 3.21
multi-jets	0 ± 0.00	1.4 ± 1.22	1.4 ± 1.19	2.8 ± 1.70
W+jets	0 ± 0.00	0 ± 0.00	0 ± 0.00	0 ± 0.00
VV	0 ± 0.00	0.3 ± 0.14	0.1 ± 0.11	0.4 ± 0.17
Single top	10.5 ± 2.01	16.2 ± 2.52	30.2 ± 3.67	56.9 ± 4.89
$Z/\gamma^* \to ll$	9.3 ± 4.62	11.7 ± 4.32	4.3 ± 3.21	25.3 ± 7.10
MC total	422 ± 6.31	604 ± 6.69	1341 ± 8.06	2367 ± 12.23
Data	367	506	1145	2018

Table 5.9: Expected number of Monte-Carlo simulation and number of observed events after final selection cut.

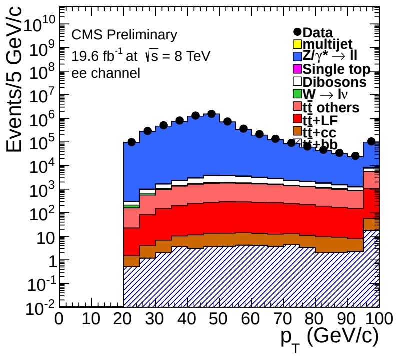
5.10 Pileup effect

As the LHC luminosity goes up, there could be more than one collision in the same bunch crossing. Those extra proton-proton interactions produce additional soft particles that can overlap with jets produced in the hard scatter collision, resulting in extra energy added to jets and more objects like charged hadrons, neutral hadrons and photons. Thus, the pileup can affect the particle based isolation and the selection of jets or missing transverse energy. To remove this pileup effect for isolation, we removed charged hadrons from the pileup vertex in PAT object processing. With this pileup removal sequence in PAT processing, the pileup effect due to charged hadrons in isolation was removed.

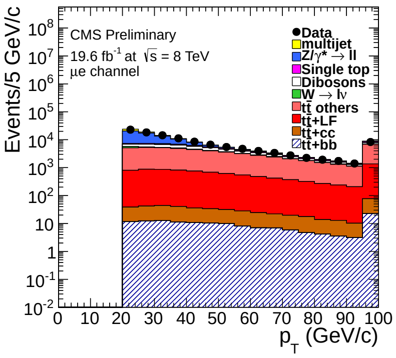
It is well known that the vertex distribution from MC simulation does not properly describe the data. Therefore, the events from simulated data was re-weighted as to represent the vertex distribution in real data. We followed the official pileup re-weighting tool described in [52]. The weight is obtained from



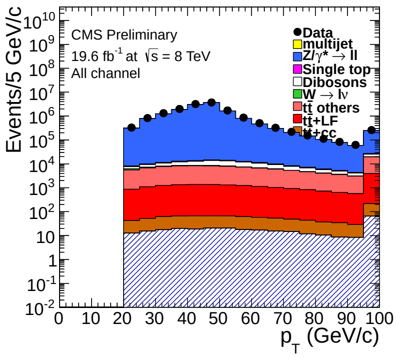
(a) $\mu^+\mu^-$ channel



(b) e^+e^- channel

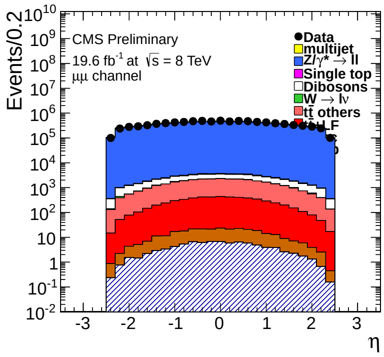


(c) $e^\pm\mu^\mp$ channel

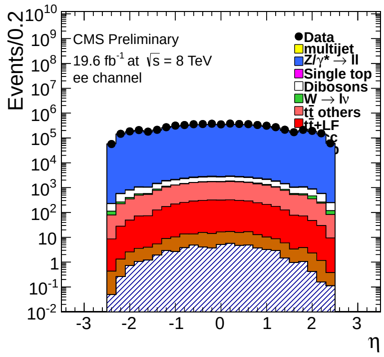


(d) Combined all channel

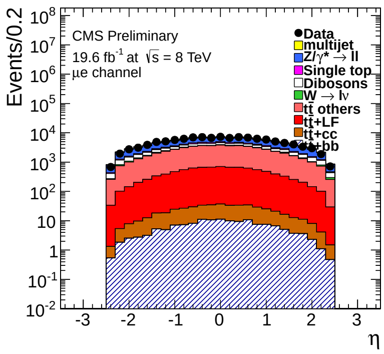
Figure 5.6: The distributions of the leading lepton's transverse momentum after the lepton identification, isolation and opposite sign pair requirement (S1).



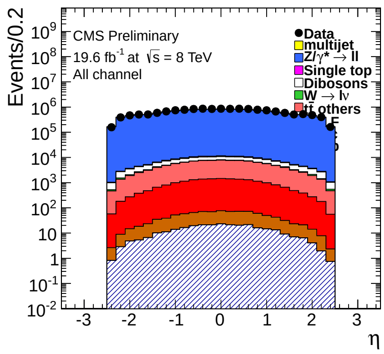
(a) $\mu^+\mu^-$ channel



(b) e^+e^- channel

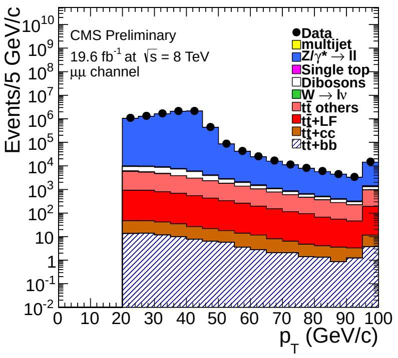


(c) $e^\pm\mu^\mp$ channel

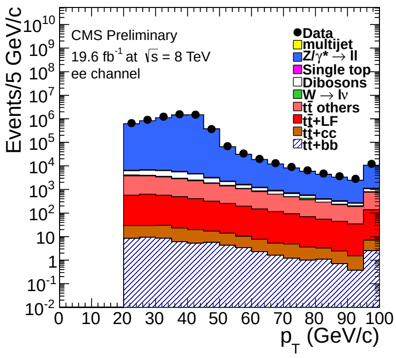


(d) Combined all channel

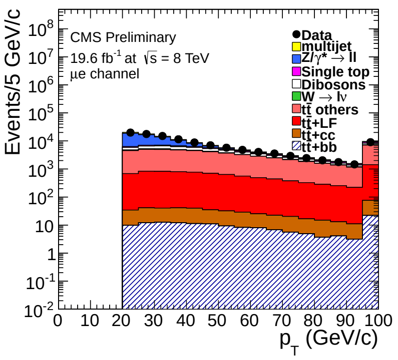
Figure 5.7: The distributions of the leading lepton's absolute pseudo-rapidigy after the lepton identification, isolation and opposite sign pair requirement (S1).



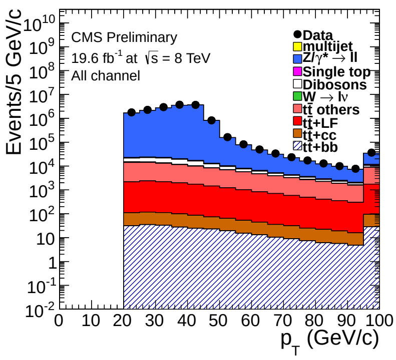
(a) $\mu^+\mu^-$ channel



(b) e^+e^- channel

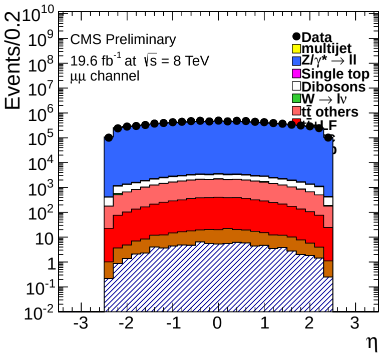


(c) $e^\pm\mu^\mp$ channel

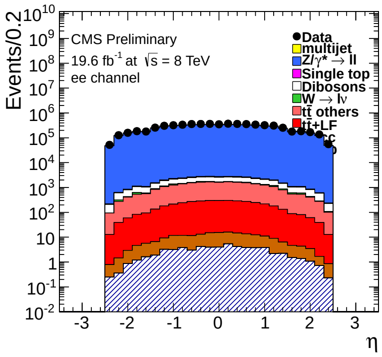


(d) Combined all channel

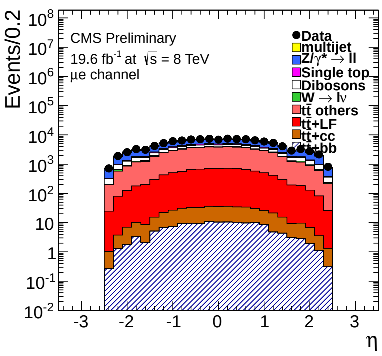
Figure 5.8: The distributions of the second leading lepton's transverse momentum after the lepton identification, isolation and opposite sign pair requirement (S1).



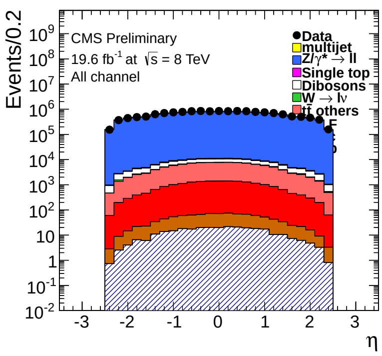
(a) $\mu^+\mu^-$ channel



(b) e^+e^- channel

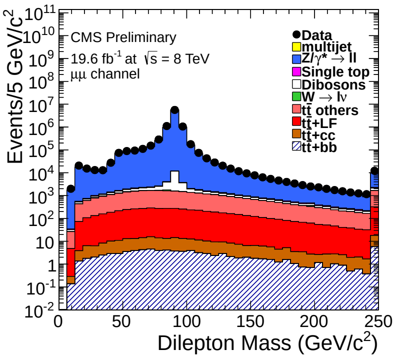


(c) $e^\pm\mu^\mp$ channel

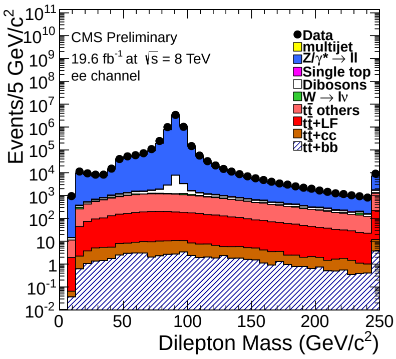


(d) Combined all channel

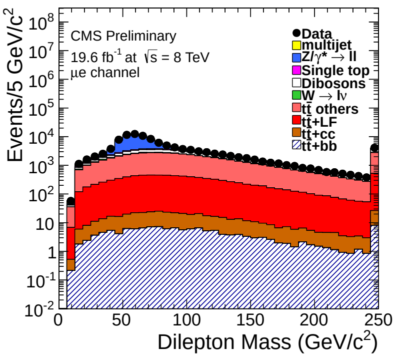
Figure 5.9: The distributions of the second leading lepton's absolute pseudorapidigy after the lepton identification, isolation and opposite sign pair requirement (S1).



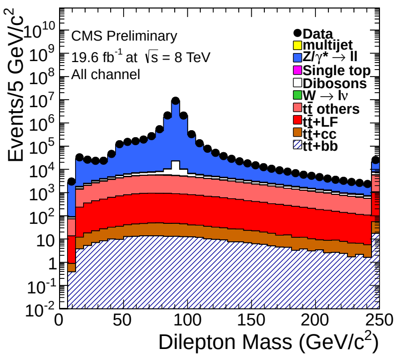
(a) $\mu^+\mu^-$ channel



(b) e^+e^- channel

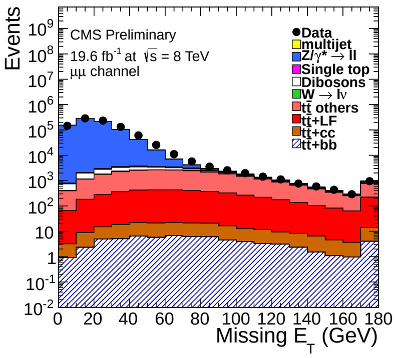


(c) $e^\pm\mu^\mp$ channel

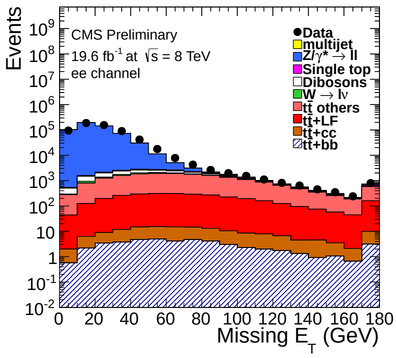


(d) Combined all channel

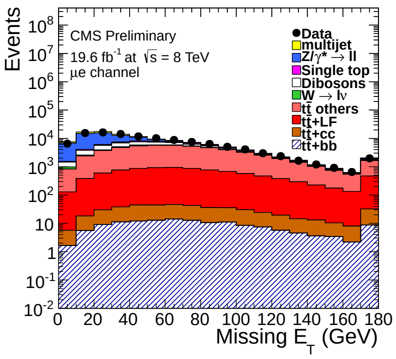
Figure 5.10: The distributions of the di-lepton invariant mass after the lepton identification, isolation and opposite sign pair requirement (S1).



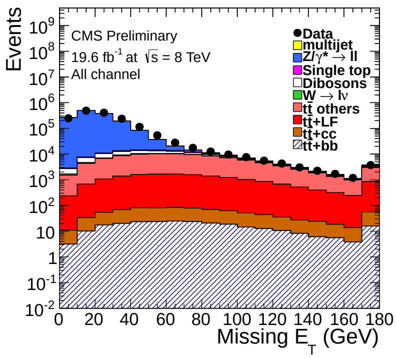
(a) $\mu^+\mu^-$ channel



(b) e^+e^- channel

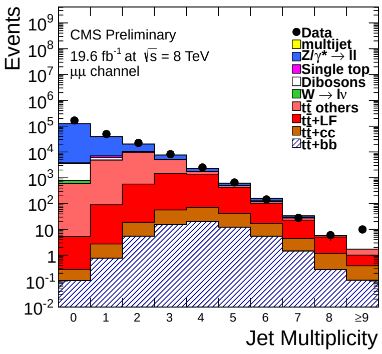


(c) $e^\pm\mu^\mp$ channel

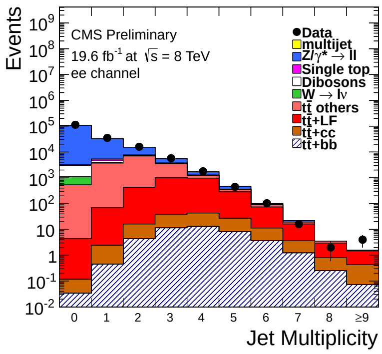


(d) Combined all channel

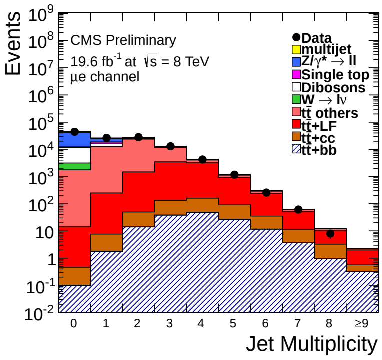
Figure 5.11: The distributions of the missing transverse energy after Z invariant mass veto (S2).



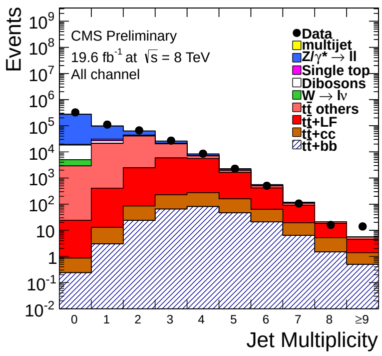
(a) $\mu^+\mu^-$ channel



(b) e^+e^- channel

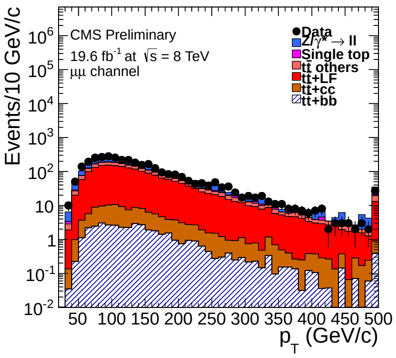


(c) $e^\pm\mu^\mp$ channel

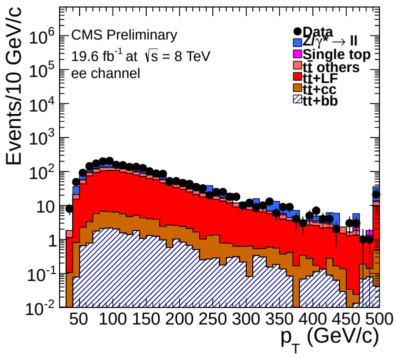


(d) Combined all channel

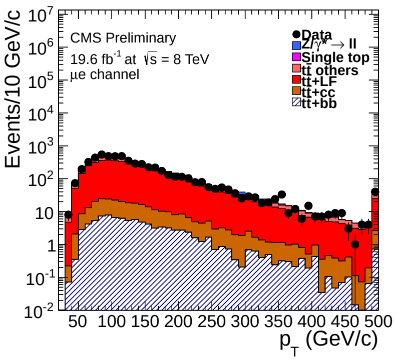
Figure 5.12: The distributions of the jet multiplicity after missing transverse energy veto (S3).



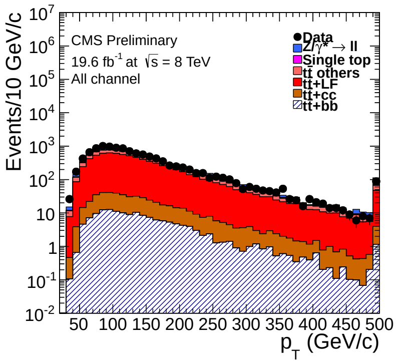
(a) $\mu^+\mu^-$ channel



(b) e^+e^- channel

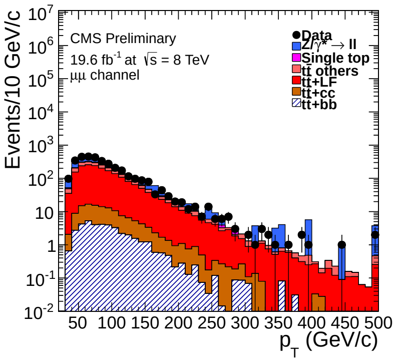


(c) $e^\pm\mu^\mp$ channel

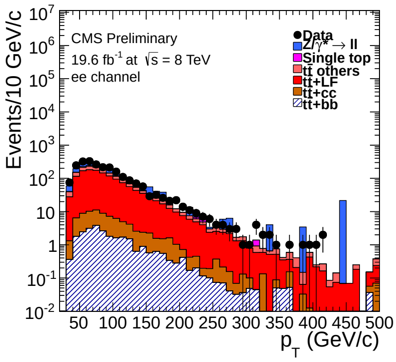


(d) Combined all channel

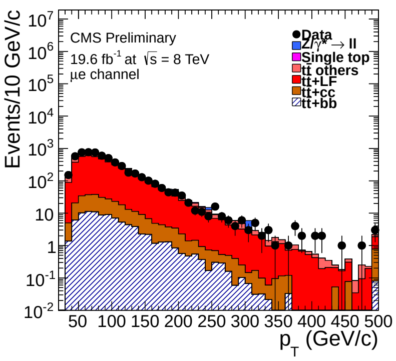
Figure 5.13: The distributions of the leading jet transverse momentum after at least four jet requirement (S4).



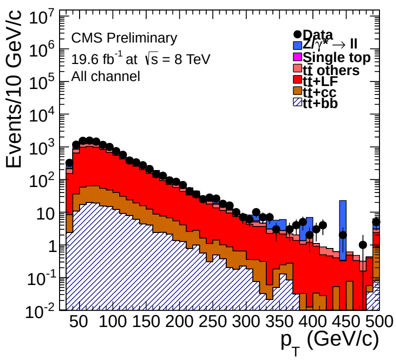
(a) $\mu^+\mu^-$ channel



(b) e^+e^- channel

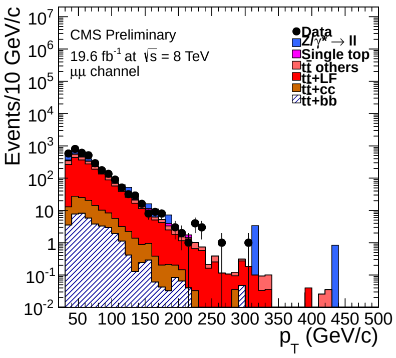


(c) $e^\pm\mu^\mp$ channel

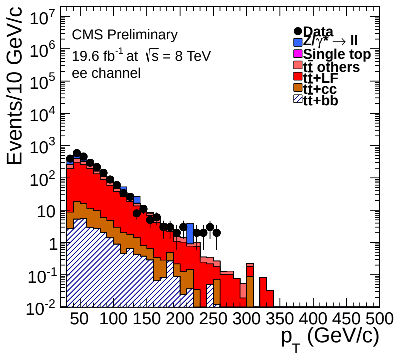


(d) Combined all channel

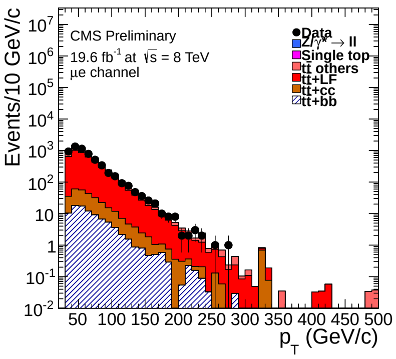
Figure 5.14: The distributions of the second leading jet transverse momentum after at least four jet requirement (S4).



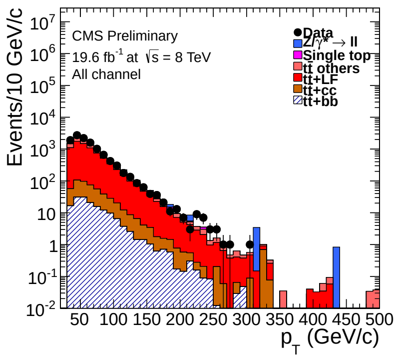
(a) $\mu^+\mu^-$ channel



(b) e^+e^- channel

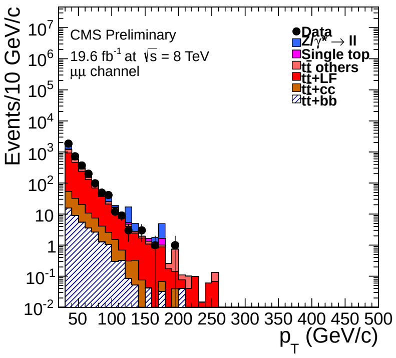


(c) $e^{\pm}\mu^{\mp}$ channel

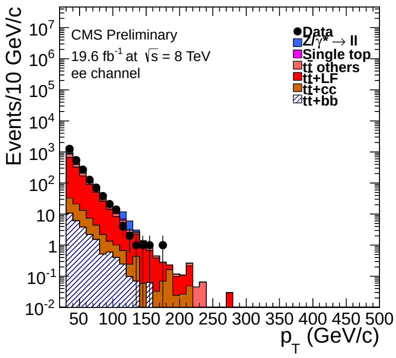


(d) Combined all channel

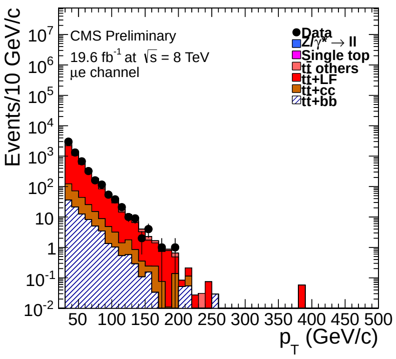
Figure 5.15: The distributions of the third leading jet transverse momentum after at least four jet requirement (S4).



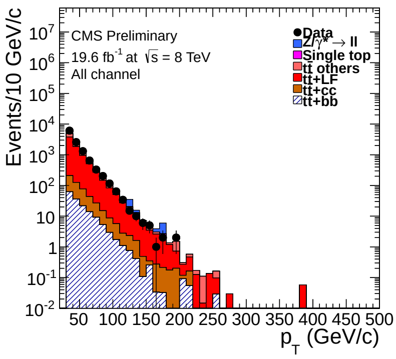
(a) $\mu^+\mu^-$ channel



(b) e^+e^- channel

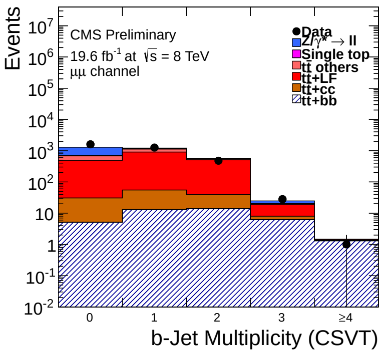


(c) $e^\pm\mu^\mp$ channel

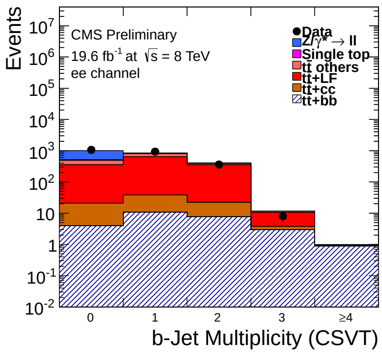


(d) Combined all channel

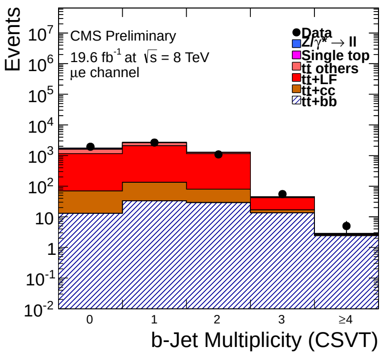
Figure 5.16: The distributions of the fourth leading jet transverse momentum after at least four jet requirement (S4).



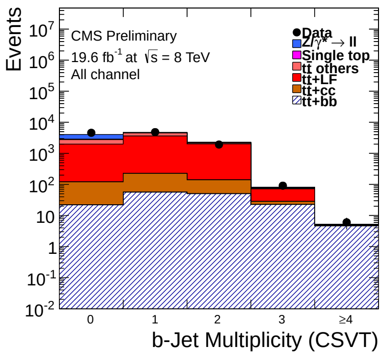
(a) $\mu^+\mu^-$ channel



(b) e^+e^- channel

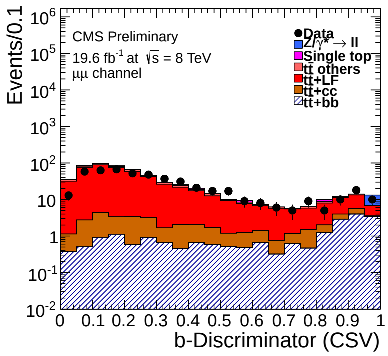


(c) $e^\pm\mu^\mp$ channel

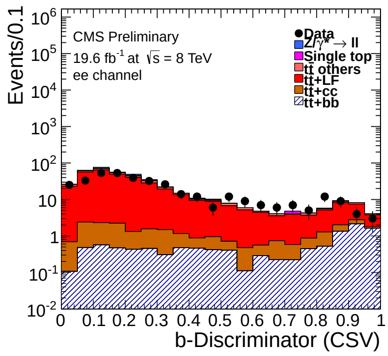


(d) Combined all channel

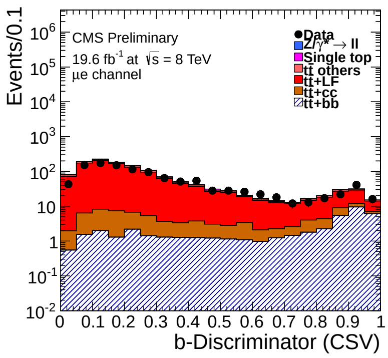
Figure 5.17: The distributions of the b-jet multiplicity after at least four jet requirement (S4).



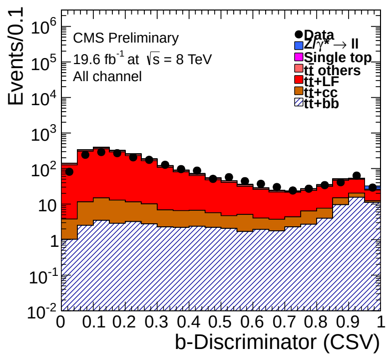
(a) $\mu^+\mu^-$ channel



(b) e^+e^- channel

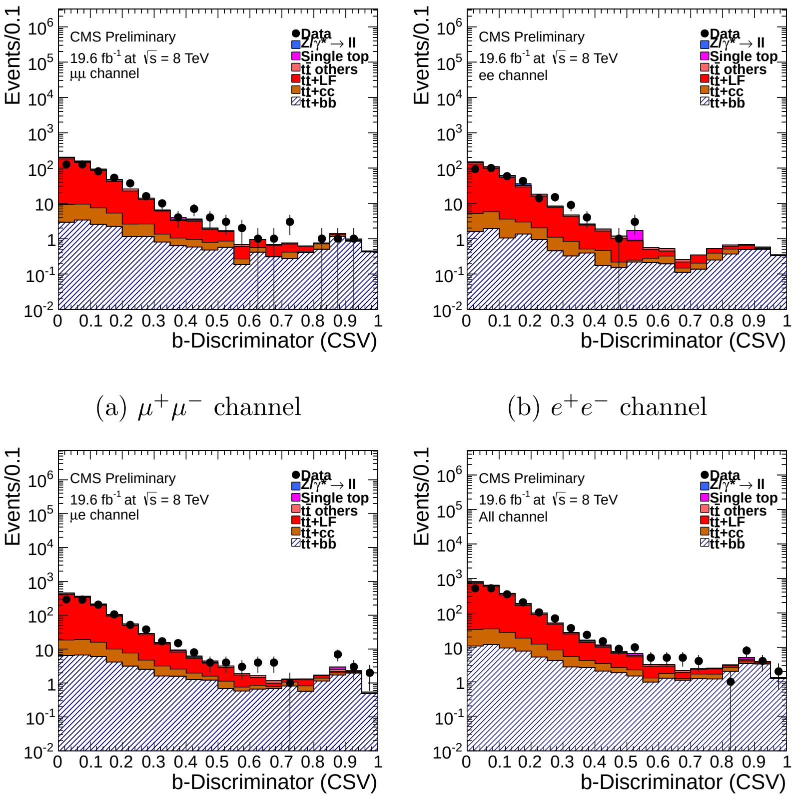


(c) $e^\pm\mu^\mp$ channel



(d) Combined all channel

Figure 5.18: b-discriminator values of first additional jet after at least two b-jet requirement (S5).



(a) $\mu^+\mu^-$ channel

(b) e^+e^- channel

(c) $e^\pm\mu^\mp$ channel

(d) Combined all channel

Figure 5.19: b-discriminator values of second additional jet after at least two b-jet requirement (S5).

this ratio corresponding to number of simulated pileup vertices in MC event by event. The reconstructed vertex distribution after applying weight is shown for all decay mode in Figure 5.20.

The remaining difference between data and MC has been taken into account by using data driven background estimation in Section 5.8.1.

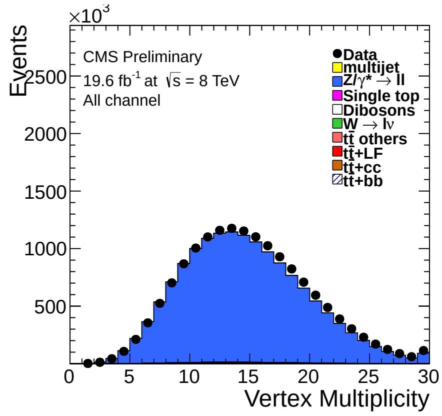


Figure 5.20: The distribution of vertex multiplicity after pre-selection for data and MC.

5.11 b-tag re-weight

The b-tagging efficiency from data and MC simulation are generally not the same [51]. In order to take into consideration of the difference, the MC simulation was re-weighted with an event weighting method by applying a scale factor to the MC b-tagging efficiency. Since the b-tagging efficiency is known to be different depending on event selection, we obtained the b-tagging efficiency after four jet selection (S4) as a function of p_T and η from $t\bar{t}jj$ simulation sample. figures 5.21 and 5.22 shows the b-tagging efficiency with CSVT as a function of p_T and η for different processes, $t\bar{t}b\bar{b}$, $t\bar{t}LF$ and $t\bar{t}jj$ and different source of b-jets. The b-tag efficiency between $t\bar{t}b\bar{b}$ and $t\bar{t}jj$ processes are observed to be not significantly different.

The data and MC simulation b-tagging probabilities for each event are obtained taking into account jet flavor dependency for the scale factor, b-tagging efficiency and number of tagged jets for each event in following formula:

$$P(MC) = \prod_{i=tagged} \epsilon_i \prod_{j=nottagged} (1 - \epsilon_j)$$
$$P(DATA) = \prod_{i=tagged} SF_i \epsilon_i \prod_{j=nottagged} (1 - SF_j \epsilon_j)$$

and the probabilities for number of tagged jets are shown in Table 5.10. where ϵ_i is the MC b-tagging efficiency, and SF_i and ϵ_i are, as usual, function of the jet flavor, jet p_T and jet η. The event weight is calculated as

$$w = \frac{P(DATA)}{P(MC)} \tag{5.9}$$

For this re-weighting, the scale factors are obtained based on Ref. [51]. The scale factors with their uncertainties as a function of p_T and η are shown in Figs. 5.23.

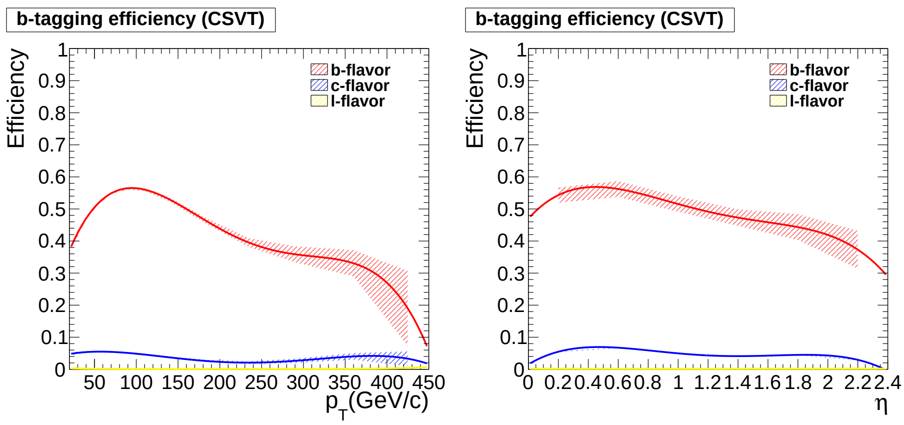


Figure 5.21: The efficiency of b-tagging for jet flavor as a function of jet p_T (left) and η (right) with CSVT.

5.12 Fitting method

The fraction of $t\bar{t}b\bar{b}$ events with respect to $t\bar{t}jj$ events was obtained from data by fitting the CSV discriminator for additional first jet and second jet.

The fit function used to extract the fraction can be written as following:

Probability	MC	Data
P (0 tag) =	$\prod_j^{n_{jets}} (1 - e_j^f)$	$\prod_j^{n_{jets}} (1 - SF_j^f e_j^f)$
P (1 tag) =	$\sum_i^{n_{jets}} e_i^f \cdot \prod_{j \neq i}^{n_{jets}} (1 - e_j^f)$	$\sum_i^{n_{jets}} SF_i^f e_i^f \cdot \prod_{j \neq i}^{n_{jets}} (1 - SF_j^f e_j^f)$
P (2 tag) =	$\sum_i^{n_{jets}} e_i^f \cdot \sum_{j>i}^{n_{jets}} e_j^f \cdot \prod_{j \neq i,j}^{n_{jets}} (1 - e_k^f)$	$\sum_i^{n_{jets}} SF_i^f e_i^f \cdot \sum_{j>i}^{n_{jets}} SF_j^f e_j^f \cdot \prod_{j \neq i,j}^{n_{jets}} (1 - SF_k^f e_k^f)$
P (3 tag) =	$\sum_i^{n_{jets}} e_i^f \cdot \sum_{j>i}^{n_{jets}} e_j^f \cdot \sum_{k>j}^{n_{jets}} e_k^f \cdot \prod_{l \neq i,j,k}^{n_{jets}} (1 - e_l^f)$	$\sum_i^{n_{jets}} SF_i^f e_i^f \cdot \sum_{j>i}^{n_{jets}} SF_j^f e_j^f \cdot \sum_{k>j}^{n_{jets}} SF_k^f e_k^f \cdot \prod_{l \neq i,j,k}^{n_{jets}} (1 - SF_l^f e_l^f)$
P ($\geq$ 4 tag) =	1 - P(0) - P(1) - P(2) - P(3)	

Table 5.10: b-tagged jet probabilities for number of b-jet.

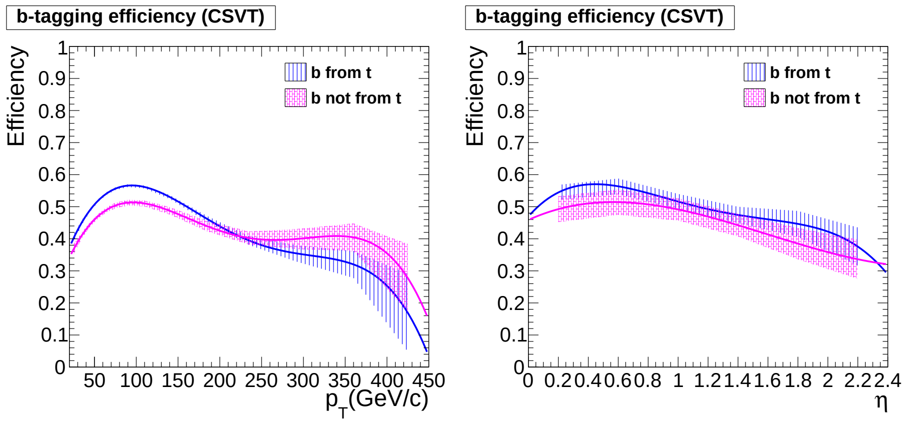


Figure 5.22: The efficiency of b-tagging for source of b-jet as a function of jet p_T (left) and η (right) with CSVT.

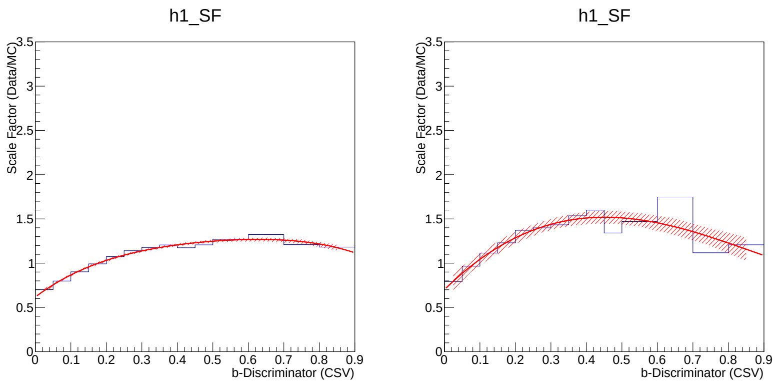


Figure 5.23: The scale factor of b-tagging as a function of CSV discriminator for 1^{st} (left) and 2^{nd} (right) additional jet with CSVT.

$$f(k,R) = k \cdot N_{t\bar{t}jj} \cdot [R \cdot N^{norm}_{t\bar{t}b\bar{b}} + (1-R) \cdot N^{norm}_{t\bar{t}LF/t\bar{t}c\bar{c}}] + k \cdot N^{MC}_{bkg} + N^{Data-driven}_{bkg} \quad (5.10)$$

where R is defined the fraction of $t\bar{t}b\bar{b}$ events with respect to the number of $t\bar{t}jj$ events at reconstruction level and $N_{t\bar{t}jj}$ is the number of events for $t\bar{t}jj$ and $N^{norm}_{t\bar{t}b\bar{b}}$ and $N^{norm}_{t\bar{t}LF}$ are the normalized distribution for $t\bar{t}b\bar{b}$ and for combined of $t\bar{t}LF$ and $t\bar{t}c\bar{c}$, respectively. The k is the parameter for normalization for MC. The data-driven background estimation was added to the fit function without the normalization parameter.

The likelihood fitting method was used CSV discriminator of 1^{st} and 2^{nd} additional jet. Figure 5.24 show the distributions after fitting. The ratio, R at reconstruction level was obtained to be 6.8 with CSVT b-tagging selection.

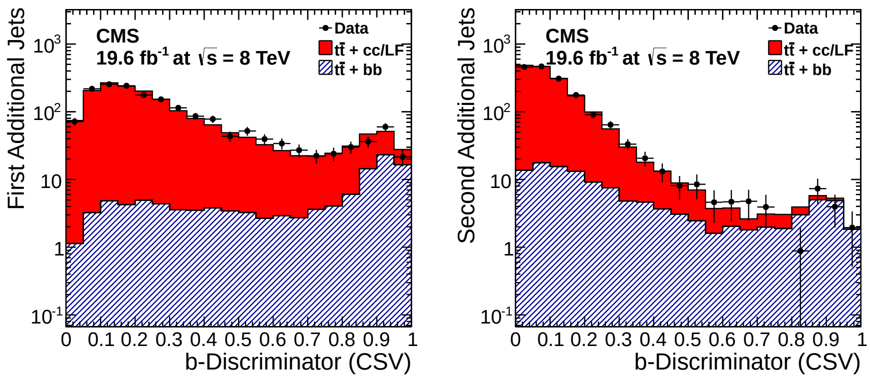


Figure 5.24: The fit CSV discriminator of 1^{st}(left) and 2^{nd}(right) additional jet for $t\bar{t}b\bar{b}$ and $t\bar{t}c\bar{c}/t\bar{t}LF$ in combined channel.

The ratio, R measured from the fitting is at the reconstruction level. In order to be able to compare with theory prediction or with other experiments, the correction factor to the defined visible phase space needs to be applied. The acceptance and efficiency ratio of $t\bar{t}jj$ and $t\bar{t}b\bar{b}$ should be applied to the measured ratio at reconstruction level. Since we measure the ratio in the visible phase space, we need to correct the selection efficiency. The efficiency $\epsilon_{t\bar{t}jj}$ is defined as the number of selected $t\bar{t}jj$ events divided by the number of $t\bar{t}jj$

events. and $\epsilon_{t\bar{t}b\bar{b}}$ as the number of selected $t\bar{t}b\bar{b}$ events divided by the number of $t\bar{t}b\bar{b}$ events in the visible phase space. The correction factor after the final selection (S5) was found to be 0.38 for CSVT selection. The correction factor was applied to the fraction, R at the reconstruction level as following:

$$\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj) = R \times \epsilon_{t\bar{t}jj}/\epsilon_{t\bar{t}b\bar{b}} \tag{5.11}$$

5.13 Systematic uncertainty

Normalization uncertainties, as the one related to luminosity, cancel out. The scale factor for lepton efficiency are very close to 1 and there would be no effect on b-tagged jet distribution. Since the kinematic distribution of jets in $t\bar{t}jj$ and $ttbb$ are similar, jet energy scale uncertainty would be relatively small in ratio measurement. Main uncertainty would be from the b-tagging scale factor which can change the shape of b-tag jet distribution and the acceptance ratio of $t\bar{t}jj$ and $t\bar{t}b\bar{b}$ from theory for correction. Following systematic uncertainties are considered.

Pileup

The number of interactions in the data, which is to be re-weighting target for pileup simulation MC samples, is estimated from the measured luminosity in bunch crossing times an average total inelastic cross-section. We vary +/- 5% the number of interactions to cover the uncertainties due to pileup modeling. The total inelastic cross-section of 68 mb is used as the central value.

Jet energy scale

The b-tagged jet multiplicity can be affected by jet energy scale variation. In order to also estimate the effect of the scale on the missing energy selection, the variation from the uncertainty dependent on the jet kinematics is propagated to the value of E_T^{miss}. The effects of the jet energy scale uncertainty on the ratio measurement are evaluated by varying the simulated signal and backgrounds. The systematic uncertainty on the jet energy scale was found to be relatively small compared to b-tagging scale factor variation.

b-tagging scale factor

The main uncertainty is mainly from b-tagging scale factor which can change b-tagged jet multiplicity. The mis-tag rate scale factor uncertainty turned out to be very large due to the contribution at the three and four tag bins. We followed the Ref. [51] to take into account the uncertainty from b-tagging scale factor. We varied up and down the b-tagging scale factor separately for heavy-quark (b/c) flavor and light-quark flavor within the scale factor uncertainty.

MC generator

The acceptance of $t\bar{t}jj$ and $t\bar{t}b\bar{b}$ depend on different generator. We used different MC generated based on MadGraph and on POWHEG for the acceptance ratio of $t\bar{t}jj/t\bar{t}b\bar{b}$. Since MadGraph is a LO generator and POWHEG is an NLO generator, the difference of acceptance ratio between these two samples will represent the uncertainties on different modeling.

Q^2 scale uncertainty

The $t\bar{t}b\bar{b}$ cross-section strongly depends on the renormalization and factorization scale factor. However, we expect that the dependency would be reduced in ratio measurement. We used the dedicated scale up and down samples to obtain the systematic uncertainty on this scale variation. However, the statistics for extracting $t\bar{t}b\bar{b}$ is limited so the assigned uncertainty should also represent the statistical fluctuation.

The uncertainties considered in this analysis are summarized in the Table 5.11

5.14 Result

In order to be able to compare with theory predictions or with other experiments, a correction facto to the particle level definition of the cross-section ratio within the visible phase space needs to be applied. The efficiency ratio of $t\bar{t}jj$ and $t\bar{t}b\bar{b}$ should be applied to the measured ratio at reconstruction level. The efficiency $\epsilon_{t\bar{t}jj}$ is defined as the nuber of reconstructed $t\bar{t}b\bar{b}$ events divided by the number of $t\bar{t}jj$ events at the particle level, both defined in the visible phase space. Similarly, the efficiency $\epsilon_{t\bar{t}b\bar{b}}$ is defined. The correction factor after the final selection (S5) is found to be 0.38 for CSVT selection. The correction factor is applied to the fraction, R at the reconstruction level in the following way: $\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj) = R \times \epsilon_{t\bar{t}jj}/\epsilon_{t\bar{t}b\bar{b}}$.

After correcting the acceptance ratio and taking into account the systematic uncertainty, we measure the cross-section ratio $\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj)$ in the visible phase space using tight working points of the CSV b-tagging algorithm. The measured cross-section ratio is

$$\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj) = 2.6 \pm 0.4(stat.) \pm 0.3(syst.)\% \tag{5.12}$$

at 8 TeV.

Source	$\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj)$ (%)
Pileup	0.5
Jet energy scale	1.0
b-tag (heavy flavor)	2.0
b-tag (light flavor)	9.0
MC generator	3.0
Q^2 scale	6.0
Total uncertainty	12

Table 5.11: Break-down of the systematic uncertainties on various sources contributing to the fraction of $\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj)$ at 8 TeV.

Chapter 6

Conclusion

The Higgs boson can be found couple with $t\bar{t}$ pair, $t\bar{t}H$, according to Standard Model. In march 2013, the Higgs boson was discovered in the LHC accelerator whose mass is around 125 GeV/c^2. When the Higgs boson have low mass like announced 125 GeV/c^2 the Higgs boson will decay into mostly $b\bar{b}$ dominantly, finally $t\bar{t}b\bar{b}$. However, this final state has an irreducible background from the $t\bar{t}b\bar{b}$ through normal QCD process. Therefore, we performed the cross-section ratio measurement of $\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj)$.

The data collected by the CMS experiment in year 2012 at $\sqrt{s} = 8$ TeV, has been used. It corresponds to the integrated luminosity of 19.6 fb^{-1}. To measure cross-section ratio, we used highly efficient physics objects reconstruction algorithms designed within the CMS collaboration and consider pile-up effect according to the increasing luminosities.

A selection criteria has been established by CMS TOP working group to select the signal-like events and validated by data and simulation comparisons after each selection step. To reduce uncertainties, we used tight b-tag selection.

Measurement of cross-section ratio $\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj)$ is performed by means of a fit to be measured b-jet discriminator distribution of first and second additional jets of dileptonic top quark pair candidates with at least four reconstructed jets and at least two b-jet.

The cross-section ratio is:

$$\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj) = 2.6 \pm 0.4(stat.) \pm 0.3(sys.)\% \tag{6.1}$$

Bibliography

[1] F. Englert and R. Brout, *"Broken Symmetry and the Mass of Gauge Vector Mesons"*, Phys. Rev. Lett. **13**, 321 (1964).

[2] P. W. Higgs, *"Broken Symmetries and the Masses of Gauge Bosons"*, Phys. Rev. Lett. **13**, 508 (1964).

[3] CMS Collaboration, *"A new boson with a mass of 125-GeV observed with the CMS experiment at the Large Hadron Collider"*, Science **338**, 1569 (2012).

[4] Atlas Collaboration, *"A particle consistnt with the Higgs Boson observed with the ATLAS Detector at the Large Hadron Collider"*, Science **338**, 1576 (2012).

[5] S.L. Glashow, *"Partial-Symmetries of Weak Interactions"*, Nucl. Phys. 22:579-588 (1961).

[6] S. Weinberg, *"A Model of Leptons"*, Phys. Rev. Lett 19:1264-1266 (1967).

[7] A. Salam, *"Weak and Electromagnetic Interactions"*, Conf.Proc. C680519:367-377 (1968).

[8] O. W. Greenberg, *"Spin and Unitary Spin Independence in a Paraquark Model of Baryons and Mesons"*, Phys. Rev. Lett 13:598-602 (1964).

[9] Wolfgang Wagner, *"Top quark physics in hadron collisions"*, Rept. Prog. Phys., 68:2409-2494, (2005).

[10] R. Brock *et al.* [CTEQ Collaboration], *"Handbook of perturbative QCD: Version 1.0"*, Rev. Mod. Phys. **67**, 157 (1995).

[11] CDF Collaboration, *"Observation of top quark production in $\bar{p}p$ collisions"*, Phys. Rev. Lett. **74**, 2626 (1995)

[12] Tevatron Electroweak Working Group and CDF and D0 Collaborations, *"Combination of CDF and D0 results on the mass of the top quark using up to 5.8 fb-1 of data"*, arXiv:1107.5255 [hep-ex].

[13] ATLAS Collaboration, *"Combination of ATLAS and CMS results on the mass of the top quark using up to 4.9 fb1 of data"*, ATLAS-CONF-2012-095.

[14] https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat8TeV

[15] W. Wagner, *"Top quark physics in hadron collisions"*, Rept. Prog. Phys. **68**, 2409 (2005).

[16] The LHC Study Group Collaboration, *"LHC-the Large Hadron Collider Accelerator Project"*, Technical report, CERN-AC-93-03 (2003).

[17] The LHC Study Group Collaboration, *"LHC-the Large Hadron Collider Conceptual Design"*, Technical report, CERN-AC-95-05 (2005).

[18] CMS Collaboration, *"The CMS Experiment at the CERN LHC"*, JINST **3**, S08004 (2008).

[19] CMS Collaboration, *"Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays"*, JINST **5**, T03021 (2010).

[20] Buckley et al., *"General-Purpose Event Generators for LHC Physics"*, Phys. Rept. **504**, 145 (2011).

[21] A.D. Martin et al., *"Parton Distributions for the LHC"*, Eur. Phys. J. C **63**, 189 (2009).

[22] T. Sjostrand, S. Mrenna, and P.Z. Skands, *"Pythia 6.4 Physics and Manual"*, JHEP **0605**, 026 (2006).

[23] S. Frixione, P. Nason and C. Oleari, *"Matching NLO QCD computations with Parton Shower simulations: the POWHEG method"*, JHEP **0711**, 070 (2007).

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

[25] S. Agostinelli *et al.* [GEANT4 Collaboration], *"GEANT4: A Simulation toolkit"*, Nucl. Instrum. Meth. A **506**, 250 (2003).

[26] K. Melnikov and F. Petriello, *"Electroweak gauge boson production at hadron colliders through O(alpha(s)**2)"*, Phys. Rev. D **74**, 114017 (2006) [hep-ph/0609070].

[27] N. Kidonakis, *"Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-"*, Phys. Rev. D **82**, 054018 (2010) [arXiv:1005.4451 [hep-ph]].

[28] CMS Collaboration, *"Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus and E_T^{miss}"*, CMS-PAS-PFT-09-001 (2009).

[29] CMS Collaboration, *"Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector"*, CMS-PAS-PFT-10-001 (2010).

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

[31] CMS Collaboration, *"Commissioning of the particle-.ow event reconstruction with leptons from J/Psi and W decays at 7 TeV"*, CMS-PAS-PFT-10-003 (2010).

[32] T. Speer, W. Adam, R. Fruhwirth, A. Strandlie, T. Todorov and M. Winkler, *"Track reconstruction in the CMS tracker"*, Nucl. Instrum. Meth. A **559**, 143 (2006).

[33] CMS Collaboration, *"Tracking and Primary Vertex Results in First 7 TeV Collisions"*, CMS-PAS-TRK-10-005.

[34] *"Deterministic Annealing, Clustering, and Optimization"*, Ph. D. Thesis, California Institute of Technology (1991).

[35] R. Fruhwirth, W. Waltenberger and P. Vanlaer, *"Adaptive vertex fitting"*, J. Phys. G **34**, N343 (2007).

[36] CMS Collaboration, *"Performance of muon identification in pp collisions at s**0.5 = 7 TeV"*, CMS-PAS-MUO-10-002.

[37] R. Fruhwirth, *"Application of Kalman filtering to track and vertex fitting"*, Nucl. Instrum. Meth. A **262**, 444 (1987).

[38] CMS Collaboration, *"Performance of CMS Muon Reconstruction in Cosmic-Ray Events"*, JINST **5**, T03022 (2010)

[39] CMS Collaboration], *"CMS physics: Technical design report"*, CERN-LHCC-2006-001.

[40] W. Adam, R. Fruhwirth, A. Strandlie and T. Todorov, *"Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC"*, eConf C **0303241**, TULT009 (2003), J. Phys. G **31**, N9 (2005).

[41] S. Baffioni et al., *"Electron reconstruction in CMS"*, Eur. Phys. J. C **49**, 1099 (2007).

[42] M. Cacciari, G. P. Salam and G. Soyez, *"The Anti-k(t) jet clustering algorithm"*, JHEP **0804**, 063 (2008).

[43] CMS Collaboration, *"Missing transverse energy performance of the CMS detector"*, JINST **6**, P09001 (2011).

[44] W. Waltenberger, *"Adaptive vertex reconstruction"*, CMS Analysis Note 2008/33 (2008).

[45] `https://twiki.cern.ch/twiki/bin/view/CMS/TWikiTopRefEventSel`

[46] S. Baffioni et al., it "Electron Reconstruction in CMS", Eur. Phys. Jou. C 49:1099-1116, (2007).

[47] M. Cacciari and G. P. Salam, *"Pile-up Subtraction Using Jet Areas"*, Phys. Lett. B659:119-126, (2008).

[48] M. Cacciari and G. P. Salam, *"The Catchment Area of Jets"*, JHEP 04:005, (2008).

[49] W. Andrews et al., *"A Method to Measure the Contribution of $DY \to l^+l^-$ to a di-lepton + MET Selection"*, AN-2009/023 (2009).

[50] S. Choi et al., *"Measurement of the $t\bar{t}$ differential cross section, $d\sigma/dM_{t\bar{t}}$ with 5 fb^{-1}"*, AN-2012/075 (2012).

[51] CMS Collaboration, *"b-Jet Identification in the CMS Experiment"*, PAS BTV-11-004 (2011).

[52] S. Choi et al., *"Measurement of the $t\bar{t}$ differential cross section, $d\sigma/dM_{t\bar{t}}$ with 5 fb^{-1}"*, AN-2012/075 (2012).