UNIVERSITY OF CALIFORNIA
RIVERSIDE

Top Quark Pair Production in Early CMS Data

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Physics

by

Shih-Chuan Kao

August 2011

Dissertation Committee:
Dr. Robert B. Clare, Chairperson
Dr. Stephen J. Wimpenny
Dr. John Ellison

The Dissertation of Shih-Chuan Kao is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

I would like to give my special thanks to my advisers, Prof. Bob Clare and Prof. Steve Wimpenny. For all the patience, kindness and wisdom they have given to me. Bob Clare is the person who opened the door of high energy physics to me. He gave me the freedom to try out the ideas in the research but never forgot to drag me back to the right direction. Steve Wimpenny is the best guru and mentor. He guided me through the whole analysis. Without his help, I can't finish this program by now. Both of their knowledge and passion for physics inspire me. I am so honor to have this opportunity to work with these two greatest physicists.

I should also thanks to Prof. John Ellison, Dr. Dominique Fortin and Dr. Suharyo Sumowidagdo. All the discussions with them and their suggestion are invaluable aid for me.

Thanks to all the members in High Energy Physics group at the University of California, Riverside. They are the best colleagues. They make this team so enjoyable. In addition, I would thanks to my friends in riverside. This little city became lovely because of you.

The greatest thanks go to my parents for always being there for me. I gained my courage to face the difficulties by talking to them.

Finally, I am grateful to my wife, Yin-Ting for all her supports, understanding and tolerance. She is always the source of my happiness and encourages me when I was depressed. Thank her for everything she has done for me.

Dedicate to my parents and my wife.

ABSTRACT OF THE DISSERTATION

Top Quark Pair Production in Early CMS Data

by

Shih-Chuan Kao

Doctor of Philosophy , Graduate Program in Physics
University of California, Riverside, August 2011
Dr. Robert B. Clare, Chairperson

We present an alternative method to extract the $t\bar{t}$ cross-section from the muon plus jets channel in pp collisions at $\sqrt{s}$ = 7 TeV. In addition to a standard baseline selection, we also study the use of topological constraints to impove the measurement. Using the full 2010 dataset we obtain a result of 160.4 $\pm$ 15.0(stat) $^{+44.0}_{-42.0}$ (syst) $\pm$ 6.4(lumi) pb from our baseline analysis. By applying two different additional topological constraints this becomes 153.3 $\pm$17.3(stat) $^{+35.6}_{-30.7}$ (syst) $\pm$ 6.1(lumi) pb and 156.5 $\pm$23.7(stat) $^{+30.8}_{-30.5}$ (syst) $\pm$ 6.3(lumi) pb

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The top quark was discovered at the Fermilab Tevatron collider in 1995, and many of its properties have been measured by the D0 and CDF experiments. With the heavest mass among all known particles, the top quark has a significant role in the standard model and also in new physics searches.

At the center-of-mass energy of 1.96 TeV, top quark pair production has a cross-section of 6 pb. Comparing with its major background, W+jets, with a cross-section of 27 nb, the discovery of the top quark and the measurement of top quark properties required several years of data taking in order to have significant number of signal candidates. With 7 TeV energy in the early phase of LHC, various calculations predict top quark pair produciton cross-section to increase to around 157.5 pb with the W+jets cross-section rising to only about 84 nb. The signal-background ratio thus increases nearly 9 times compared to the

Tevatron. The abundance of top events itself is already the early physics in LHC and it allows the precision measurements of the properties of top quark such as spin, differential cross-section or single top production in the future.

The motivation of this study is to explore the kinematic features of $t\bar{t}$ events based on the current understanding of top quark. By employing the topological information, the method can extract $t\bar{t}$ events down to 1 pb^{-1}. This method not only can be applied to early cross-section measurement but also could benefit the mass measurement in the future.

The thesis is structured as follows: Chapter 2 briefly describes the standard model and discusses the kinematics and significance of the top quark in standard model. Chapter 3 gives the overview of the LHC and the CMS detectors. Chapter 4 discuss the CMS software including simulation and reconstruction. Chapter 5 explains the objects identification and event selection. Chapter 6 describes the data driven method for background estimation. Finally, chapter 7 discusses the analysis method, the uncertainty of the measurement and summarizes the results.

Chapter 2

The Top Quark and The Standard Model

2.1 Brief Discussion of The Standard Model

The Standard Model is a theory to explain how the elementary particles exist and interact. It contains 3 sectors which are 3 families of fermions, four gauge fields and the higgs mechanism. The first sector states that matter is made of fermions. The second sector describes the interactions between them. The last one, the Higgs sector, is still not confirmed by experiment but it is needed to fulfill the previous two sectors.

In the details of first sector, there are two groups of fermions, leptons and quarks. Each fermion group has 3 generations and each generation has two types, up and down. In the

quark family, the up type quarks have charge +2/3 and the down type quarks are $-1/3$. In addition, the different generations of quarks are mixed according to CKM matrix which was developed by Kobayashi and Maskawa in order to explain CP violation in the weak interaction and it also predicted 3 generations of quarks. Similar to the quark family, the charge difference between up type and down type leptons is 1. The up type leptons, neutrinos, are all neutral and the down type leptons carry -1 which are electron, muon and tau. After observing top quark in 1995 [54] and tau neutrino in 2000 [45], all the particles in this sector were discovered and they are listed in table 2.1

The second part of the standard model, the gauge sector, mainly describes the interactions between fermions. There are three kinds of fundamental forces in the standard Model: the strong, eletromagnetic and weak force. Each force is propagated by its own gauge bosons and is described by different dynamics. Table 2.2 shows the forces and their corresponding gauge bosons and theories.

	I	II	III
Leptons	ν_e	ν_μ	ν_τ
	e	μ	τ
Quarks	u	c	t
	d	s	b

Table 2.1: Table of elementary partilces of standard model

Force	Gauage boson	Coupling	Theory
Strong	gluon	$\alpha_s \sim 0.1$	QCD
Electromagnetic	photon	$\alpha \sim 1/137$	QED
Weak	W^+, W^- and Z	$\alpha_w \sim 4.2 \times 10^{-3}$	GSW Theory

Table 2.2: The 3 fundamental Interactions in the Standard Model

The gauge bosons are spin 1 and have all been discovered. The interaction of the gauge boson with fermions as well as their self interaction are described by the $SU(3) \times SU(2)_L \times U(1)$ gauge theory.

The Electromagnetic theory was developed starting with Maxwell. Quantum Eletrodynamics (QED) was constructed with the classic Electromagnetic theory and Relativistic Quantum Field theory. The work to combine these two theories was done by satisfying the local gauge invariance of electromagnetic theory in Lorentz covariant form. The symmetry involved in this process is called U(1) group. The physics interpretation of QED states that particles only can experience the electromagnetic force by interacting with a massless vector field, photon field.

QED is a very successful theory with accurate predictions. It has been used as a model to extend our understanding to other interactions. After Yang and Mills's efforts to migrate this theory from abelian U(1) group to non-abelian groups, QCD became the SU(3) ap-

plication for this theory. Comparing QCD with QED, the difference is the gauge boson of QCD, the gluon, can self interact which means the gluon itself also carries the color charge. The common point is the gauge bosons, no matter photon or gluons, are massless.

Although the Yang-Mills's theory successfully apply to QCD, it didn't work for weak interaction because of the requirement that the bosons are massless. In order to get the local gauge invariance for an non-abelian SU(2)xU(1) symmetry and retain the mass of W and Z bosons, a spin-0 scalar field was introduced and its symmetry was spontaneously broken consequently. The spin-0 scalar field is the last component of the Standard Model, Higgs.

Take U(1) gauge symmetry as an example. The idea of the spontaneous symmetry breaking is considering a scalar field ϕ in the Lagrangian

$$\mathcal{L} = (D_\mu \phi)^\dagger (D^\mu \phi) + \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 - \frac{1}{4}(F_{\mu\nu} F^{\mu\nu}) \tag{2.1}$$

and make the Lagrangian invariant under local gauge transformation. Therefore,

$$\phi \ \rightarrow \ e^{i\alpha(x)} \tag{2.2}$$

$$D_\mu = \partial - ieA_\mu \tag{2.3}$$

$$A_\mu \rightarrow A_\mu + \frac{1}{e}\partial_\mu \alpha \tag{2.4}$$

If $\mu^2 > 0$, the Lagrangian is just a spin-0 QED formula with mass μ which has no practical use in physics. Thus, choosing $\mu^2 < 0$ is the other way to go. It results in the spontaneous symmetry breaking of the ϕ field. The ϕ can be expanded as $h(x)$ and $\theta(x)$

$$\begin{aligned} \phi &= \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) && (2.5) \\ &= \frac{1}{\sqrt{2}}(v + h(x))e^{i\theta(x)/v} && (2.6) \\ A_\mu &\rightarrow A_\mu + \frac{1}{ev}\partial_\mu\theta && (2.7) \end{aligned}$$

with respect to minimum value $v = \sqrt{-\mu^2/\lambda}$ and 0 for ϕ_1 and ϕ_2. By gauging away $\theta(x)$, the Lagrangian is independent of $\theta(x)$ and the Lagrangian 2.4 becomes

$$\mathcal{L} = \frac{1}{2}(\partial_\mu h)^2 - \lambda v^2 h^2 - \frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial_\nu A_\mu - \partial_\mu A_\nu) + \frac{1}{2}e^2 v^2 A_\nu A^\nu \qquad (2.8)$$

The first and third term are the kinematics of h and vector field A. The second term gives the mass of scalar field h, $m_h = \sqrt{2\lambda v^2}$. The last term gives the mass of vector field A, $m_A = ev$. The implement of a scalar field is successful to provide a mechanism for generating the mass of gauge bosons.

The application of this idea in SU(2)×U(1) symmetry is the Electroweak theory where

$$D^{\mu} \rightarrow (\partial^{\mu} + \frac{ig\tau W^{\mu}}{2} + \frac{igB^{\mu}}{2}) \tag{2.9}$$

and the Lagrangian of the scalar field is

$$\mathcal{L} = (D_{\mu}\phi)^{\dagger}(D^{\mu}\phi) + \mu^2\phi^{\dagger}\phi - \frac{\lambda}{4}(\phi^{\dagger}\phi)^2 - \frac{1}{4}(F_{\mu\nu}F^{\mu\nu}) - \frac{1}{4}(G_{\mu\nu}G^{\mu\nu}) \tag{2.10}$$

$$F^{\mu\nu} = \partial^{\mu}W^{\mu} - \partial^{\nu}W^{\nu} - gW^{\mu} \times W^{\nu} \tag{2.11}$$

$$G^{\mu\nu} = \partial^{\mu}B^{\nu} - \partial^{\nu}B^{\mu} \tag{2.12}$$

By taking hyperchage Y = 1 and the third component of isospin T^3 = -1/2, the choice of the ground state of ϕ is

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_1 + i\phi_2 \\ \phi_3 + i\phi_4\end{pmatrix} = \begin{pmatrix}0 \\ \frac{1}{\sqrt{2}}(v + H(x))\end{pmatrix} \tag{2.13}$$

where ϕ_1, ϕ_2 and ϕ_4 are chosen to be zero. The remaining real scalar field $H(x)$ is the Higgs field. Thus, the Lagrangian becomes

$$
\begin{aligned}
\mathcal{L} \quad &= \frac{1}{2}\partial_\mu H \partial^\mu H - \mu^2 H^2 \\
&-\frac{1}{4}(\partial_\mu W_{1\nu} - \partial_\nu W_{1\mu})(\partial^\mu W_1^\nu - \partial^\nu W_1^\mu) + \frac{1}{8}g^2 v^2 W_{1\nu} W_1^\mu \\
&-\frac{1}{4}(\partial_\mu W_{2\nu} - \partial_\nu W_{2\mu})(\partial^\mu W_2^\nu - \partial^\nu W_2^\mu) + \frac{1}{8}g^2 v^2 W_{2\nu} W_2^\mu \\
&-\frac{1}{4}(\partial_\mu W_{3\nu} - \partial_\nu W_{3\mu})(\partial^\mu W_3^\nu - \partial^\nu W_3^\mu) - \frac{1}{4}G_{\mu\nu}G^{\mu\nu} \\
&+\frac{1}{8}v^2(gW_{3\mu} - g'B_\mu)(gW_3^\mu - g'B_\mu)
\end{aligned} \tag{2.14}
$$

The first term gives the mass of the Higgs $\sqrt{2\mu^2}$. From the second and third line of the equation 2.14, the mass of W_1 and W_2 are gv/2. The last two lines show the mixing of W_3 and B field. They can be decoupled as

$$Z^\mu \quad = \cos\theta_W W_3^\mu - \sin\theta_W B^\mu \tag{2.15}$$

$$A^\mu \quad = \sin\theta_W W_3^\mu + \cos\theta_W B^\mu \tag{2.16}$$

with

$$\cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}} \tag{2.17}$$

$$\sin\theta_W = \frac{g'}{\sqrt{g^2 + g'^2}} \tag{2.18}$$

and the last two lines can be re-written as

$$-\frac{1}{4}(\partial_\mu Z_\nu - \partial_\nu Z_\mu)(\partial^\mu Z^\nu - \partial^\nu Z^\mu) + \frac{1}{8}v^2(g^2 + g'^2)Z_\nu Z^\nu - \frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) \tag{2.19}$$

which gives the mass of $Z = \mathrm{v}\sqrt{g^2 + g'^2}/2 = M_W/\cos\theta_W$ and mass of $A = 0$. The massless A field is the photon.

As mentioned above, the spontaneously broken $SU(2) \times U(1)$ symmetry leads to mass terms of W and Z bosons in the Lagrangian with a proper choice of ground state of higgs field. With the same higgs isospin doublet, fermions also acquire mass by interacting with the higgs field. The existence of the higgs field not only provides the mass mechanism to particles but also ensures to the renormalizability of the theory. The process $\mathrm{W}^+\ \mathrm{W}^- \rightarrow \mathrm{W}^+\ \mathrm{W}^-$ is an example where divrgences occur in the high energy limit ($\sigma \propto s/M_W^2$) if only the diagrams in Figure 2.1 are taken into account. An additional diagram with the exchanging of a scalar particle (the Higgs boson) as Figure 2.2 will cancel the divergence.

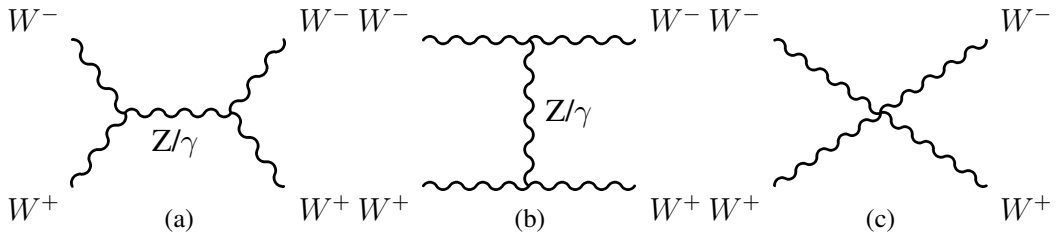


Figure 2.1: $W^+W^- \rightarrow W^+W^-$

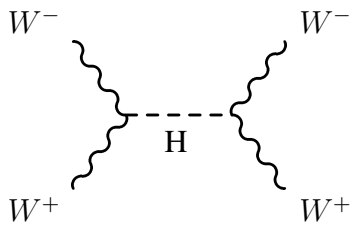


Figure 2.2: $WW \rightarrow H \rightarrow WW$

2.2 The Top Quark in The Standard Model

In the standard model, the top quark is in the third generation of the quark family, the weak isospin partner of the bottom quark. Like other up-type quarks, the third component of weak isospin I_3 of top quark is +1/2 and the charge is +2/3. However, its huge mass plays an significant role in radiation corrections which are listed below.

2.2.1 Mass of W and Z

The gauge bosons of the electroweak theory are one massless photon and three massive bosons, W^+, W^- and Z^0. They are predicted from the process of spontaneously SU(2)xU(1) symmetry breaking. The mass terms of $W^\pm$ and Z^0 are given from the Lagrangian 2.14

$$M_W = \frac{gv}{2} \tag{2.20}$$

$$M_Z = \frac{1}{2}v\sqrt{(g^2 + g'^2)} \tag{2.21}$$

$$= \frac{M_W}{\cos\theta_W} \tag{2.22}$$

In addition, the requirement that the massless boson be the photon yields

$$e = g\sin\theta_W \tag{2.23}$$

In order to get the prediction of W and Z^0's mass, the comparison of Fermi's V$-$A theory and the charged current of weak interaction provides the relation

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M_W^2} \tag{2.24}$$

Summarizing the above information, the mass of W boson is obtained by

$$M_W = \sqrt{\frac{\pi\alpha}{\sqrt{2}G_F}}\frac{1}{\sin\theta_W} \simeq \frac{37.28}{\sin\theta_W} \simeq 77.73\,\text{GeV}/c^2 \tag{2.25}$$

and the mass of Z^0 is 88.58 GeV. G_F is about 1.166×10^{-5} GeV $^{-2}$ and sinθ is around 0.23 which can be determined from experiment. The Z mass was measured by LEP and the W mass was determined by LEP and Tevatron. They are

$$M_Z = 91.1876 \pm 0.0021\,\text{GeV}/c^2$$

$$M_W = 80.451 \pm 0.061\,\text{GeV}/c^2$$

which are larger than the tree level predictions. The differences are attributed to higher order corrections such as the one loop corrections shown in Figure 2.3 and 2.4.

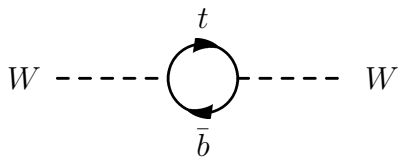


Figure 2.3: Loop correction from top to the W mass

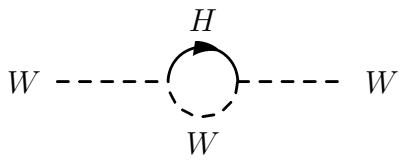


Figure 2.4: Loop correction from the Higgs to the W mass

The higher order correction for the W mass has the form

$$M_W^2 = \frac{\pi\alpha}{\sqrt{2}G_F}\frac{1}{\sin^2\theta_W(1-\Delta r)} \tag{2.26}$$

with

$$\Delta r \quad = \Delta r_0 - \frac{1-\sin^2\theta_W}{sin^2\theta_W}\Delta\rho + (\text{remaining terms}) \tag{2.27}$$

$$\Delta\rho \quad = \frac{3G_F(m_t^2-m_b^2)}{8\pi^2\sqrt{2}} \tag{2.28}$$

$$(\text{remaining terms}) \quad \approx \frac{11}{3}\frac{\sqrt{2}G_F M_W^2}{16\pi^2}[\ \ln(\frac{m_H^2}{M_W^2}) - \frac{5}{6}\ ] \tag{2.29}$$

which has the leading contribution from the square fermion mass and the secondary contribution from higgs mass because of its form of logarithm. Since the top mass is much larger than other quarks, it becomes the most important parameter in the radiative correction of W mass.

2.2.2 Decay Width of W Boson

The width of the W can be determined by direct measurement or by the ratio of R $\equiv$ $\sigma_W B(W \to e\nu)/\sigma_Z B(Z^0 \to e^+e^-)$.[36] It is expressed as

$$R \equiv \frac{\sigma_W B(W \to e\nu)}{\sigma_Z B(Z^0 \to e^+e^-)} = \frac{\sigma(pp \to WX)}{\sigma(pp \to ZX)} \frac{\Gamma(W \to e\nu)}{\Gamma(Z \to e^+e^-)} \frac{\Gamma(Z)}{\Gamma(W)} \tag{2.30}$$

The cross section ratio of W and Z production can be calculated from the theory as well as the partial width of W $\to$ $e\nu$. The theoretical uncertainties of the W and Z cross section mostly cancel since their production mechanisms are similar in a hadron collider. The partial and total width of the Z were already well measured at LEP. Hence the total width of the W can be determined.

By this mean, one can look at the branching ratio of W $\to$ $e\nu$. It has the sensitivity to top mass because of its constribution to the high order correction as its mass [43]. The Δr has a $m_t^2 - m_b^2$ term and it will have clear effect on W width if top mass is much greater than b quark's2.31.

$$\Gamma_{ij}^W = |V_{ij}|^2 \frac{G_\mu}{\sqrt{2}} \frac{M_W^3}{6\pi} [1 - \Delta r][1 + \frac{\alpha_s(M_W)}{\pi} D_i] C_i \tag{2.31}$$

V_{ij} is Cabibbo-Kobayashi-Maskawa matrix element; for leptons it is just δ_{ij}

$C_i/3 = D_i = 1$ for quark decays

$C_i = 1$, $D_i = 0$ for lepton decays

Figure 2.5 shows the experiment result in terms of W branching ratio and top mass assumption, suggested that the top mass is larger than 62 GeV/c^2 with 95% C.L. with the measured value $0.1094 \pm 0.0033(\text{stat}) \pm 0.0031(\text{syst})$ [36]. A followed direct measurement of W width and W/Z cross-section consists with the result and also exclude the possibility of new physics.[17, 55]

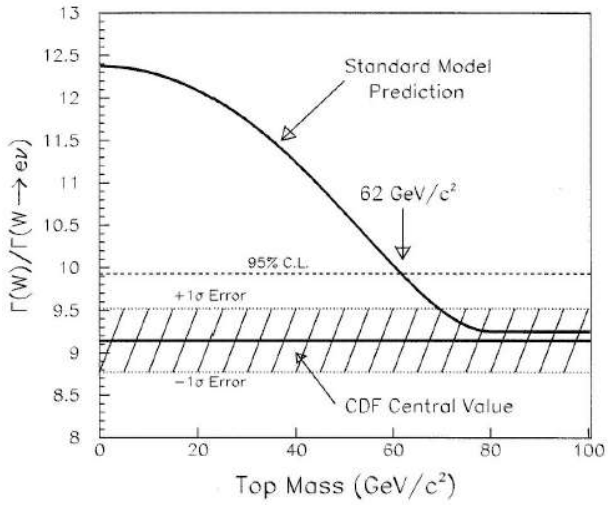


Figure 2.5: W Branching Ratio and Top Mass

2.2.3 Mass of Higgs

Since the higgs mass is one of the parameters in the high order correction of the W Mass as well as top mass are also highly correlation from this correction, thus higgs mass is con-

strained from the mass of top quark and W. According the current results from Tevatron and LEP in Figure 2.6, it seems to suggest the light higgs in standard model.

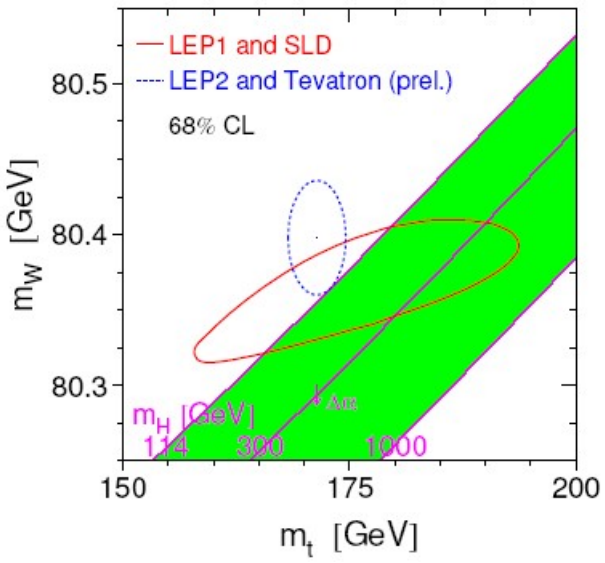


Figure 2.6: Higgs mass constraint from Top and W mass

2.3 Indirect Evidence of the Top Quark

2.3.1 $e^+e^- \to b\bar{b}$ forward-backward asymmetry

The forward-backward asymmetry of $e^+e^- \to b\bar{b}$ from the weak interactions [69] confirmed that the weak isospin of b quark was T_3 = 1/2. This means that the b quark is in an isospin doublet and that its partner T_3 = 1/2 will must exist.

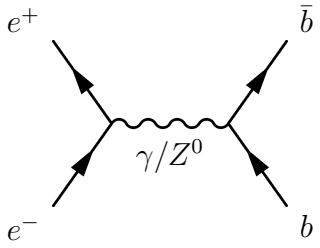


Figure 2.7: forward-backward asymmetry

2.3.2 Anomalies

Considering W boson scattering, in order to keep the theory renomalizable, at tree level Higgs boson is introduced to cure the divergence in high energy region. At higher order, a triangle loop with the W^+, W^- and Z^0 exists. This anomaly(Figure 2.8) would spoil the renomalizability of the theory. However this phenomenon can be cancelled in each fermion family doublet implying the existence of the top quark in the third quark family.

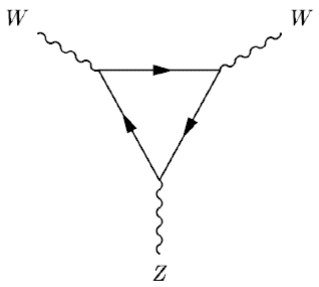


Figure 2.8: Top Anomalies

2.3.3 Absence of large tree level $B_d^0 - \bar{B}_d^0$ mixing and FCNC of B decay

Before the top quark was discovered, the explanation of $B_d^0 - \bar{B}_d^0$ mixing could be from the tree level diagram of F.C.N.C(Flavor Changing Neutral Current) and the box diagram of

charged current which are shown in Figure 2.9 and 2.10. Under the assumption of no top quark, the theoretical calculation indicates the major contribution from tree level diagram which is two orders of magnitude larger than the observation from experiment [52, 41]. This result strongly suggests the doublet state of third quark family. With the consideration of the top quark, the value of mixing parameter χ_d is 0.168 $\pm$ 0.024 given from experiments. This gives a lower bound on top mass of 60 GeV/c^2 at 95 $\%$ C.L. according to the definition from theory:

$$x_d = 0.15[\frac{\tau_B|V_{td}^2|}{3.3 \times 10^{-16}s}][\frac{B_B f_B^2}{(0.14GeV)^2}][\frac{M_t}{40GeV/c^2}]^2 \tag{2.32}$$

and

$$\chi_d = x_d^2/(1 + x_d^2)$$

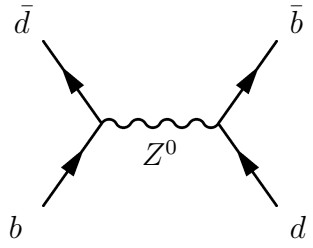


Figure 2.9: Tree diagram of $B_d^0 - \bar{B}_d^0$ mixing

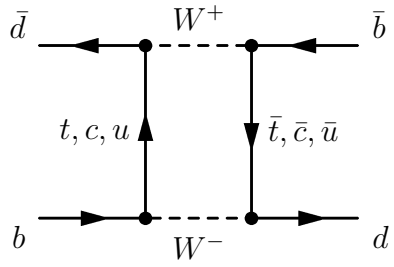


Figure 2.10: Box diagram of B^0-$\bar{B}^0_d$ mixing

2.3.4 $\mathbf{Z}^0 \to b\bar{b}$ decay width

If the b quark is an isospin singlet, the partial decay of Z $\to b\bar{b}$ is expected to be around 24 MeV. However, the precise measurements from LEP [67] has determined the value to be 384 MeV consistent with theoretical prediction with the value of 381 MeV [40]. This strongly indicates that the isospin partner of the b quark, the top quark, must exist.

2.4 The Discovery of Top quark

The direct search of top quark was performed in e^+e^- [44] by measuring the cross-section ratio with an expected value 5.

$$R = \frac{\sigma(e^+e^- \to q\bar{q})}{\sigma(e^+e^- \to \mu^+\mu^-)} = 5$$

or searching the resonances in the Z decay[50]. Similar searches also have been done at hadron colliders by looking at the decay from W or the direct decay from top such as

$$p\bar{p} \to W + X; W^+ \to t + \bar{b}; W^- \to \bar{t} + b$$

or

$$p\bar{p} \rightarrow t\bar{t} + X$$

and used the decay mode

$$t \rightarrow b + l + \nu_l$$

The lower bound of top mass was set to be greater than 69 GeV at 95% confidence level from the UA1 and UA2 experiments at the CERN Sp$\bar{p}$S Collider [65] [15]. With sensitivity to top pairt production, another experiments in Tevatron, D0 and CDF, has set the lower limit to 130 GeV which excluded the dynamics of W $\rightarrow$ tb [34] [53]. Therefore, the search focused on $p\bar{p} \rightarrow t\bar{t}$. The best signatures were the di-lepton channels where both Ws from the top decay leptonically ($e\mu, ee, \mu\mu$) and the semi-leptonic channels which only one W decay leptonically and the other decays hadronically (e+jets, μ+jets).

The analyses were based on cuts relying on the event signature to select isolated leptons, missing transverse energy ($\not\!\!E_T$) and jets. The major discriminator between background and signal was using H_T which was difined as the scalar sum of the E_T of the selected jets or added the E_T of the leading electron for $e\mu$ and ee channel. This quantity was also validated by comparing data and simulation for the shape of the H_T distribution.

In 1994, the CDF experiment at the Tevatron Collider showed the first evidence for top production [35] and in the next year, both experiments at the Tevatron, DO and CDF,

announced the discovery of the top quark.[54]

With an integrated luminosity of nearly 50 pb^{-1}, 17 events were observed with an expected background of 3.8 $\pm$ 0.6 events. The corresponding measured cross-section was 6.4 $\pm$ 2.2 pb and the measured mass was 199^{+19}_{-21}(stat) $\pm$ 22(sys) GeV.

2.5 Top Production in Hadron Collision

Top pair production in hadron colliders proceeds through $q\bar{q}$ annihilation (Figure 2.11) and gluon fusion (Figure 2.12). Based on the center-of-mass(C.M.) energy of the colliding beam and the parton distribution function(P.D.F.), the dominant production at the Tevatron is $q\bar{q}$ annihilation. At the LHC with 7 TeV center of mass, gluon fusion (Figure 2.12) is the dominant process for $t\bar{t}$ production.

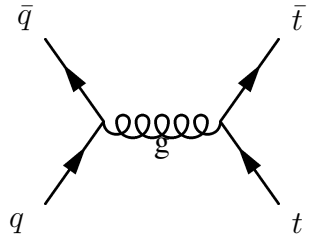


Figure 2.11: $q\bar{q} \rightarrow t\bar{t}$

In proton-proton collisions, the $t\bar{t}$ cross section is calculated by perturbative QCD with the factorization of hard scattering[19]. Similar to the Drell-Yan process, the cross-section can be written as

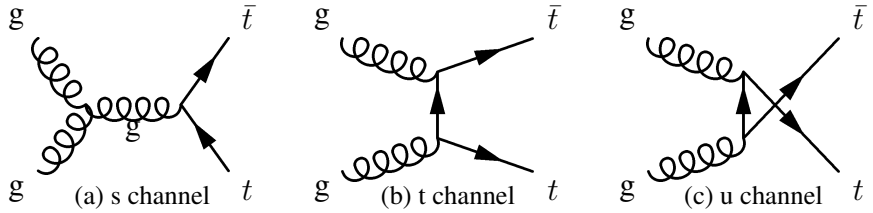


Figure 2.12: $gg \to t\bar{t}$

$$\sigma_{AB} = \int dx_a dx_b f_A(x_a, \mu_F^2) f_B(x_b, \mu_F^2) \hat{\sigma}_{ab\to X} \tag{2.33}$$

$$\hat{\sigma}_{ab\to X} = \hat{\sigma_0} + \alpha_S(\mu_R^2)\hat{\sigma_1} + \cdots \tag{2.34}$$

$\hat{\sigma}_{ab\to X}$ is the partonic cross-section which are $q\bar{q} \to t\bar{t}$ and $gg \to t\bar{t}$. f_A and f_B are parton distribution functions (P.D.F.), μ_F is the factorization scale which separates the perturbative short-distance physics (emitted parton with higher p_T) from non-perturbative long-distnace physics (emitted parton with $p_T < \mu$) and μ_R is the renormalization scale for the QCD running coupling constant. The conventional choice is $\mu_F = \mu_R =$ the top mass. The P.D.F.s are the solution of the DGLAP equations[6]:

$$f(x, \mu^2) = f(x) + \frac{\alpha_S}{2\pi} \int_x^1 \frac{dy}{y} f(y) P_{ij}(x/y, \alpha_S)[\ln(\frac{\mu^2}{m^2}) + C(x/y)] \tag{2.35}$$

$$P_{ij}(x, \alpha_S) = P_{ij}^0(x, \alpha_S) + \frac{\alpha_S}{2\pi} P_{ij}^1(x, \alpha_S) + \cdots \tag{2.36}$$

$$ij = qq, qg, gg, gp$$

where the splitting function P_{ij} is the probability density of finding parton j from parton i. The leading order (LO) prediction of the $t\bar{t}$ cross-section uses the first term in Figure 2.34 and Figure 2.36 ($\hat{\sigma_0}$ and $P_{ij}^0(x, \alpha_S)$) and the next-to-leading order (NLO) predition must include the second term in both equations. Table2.3 shows the NLO prediction for $t\bar{t}$ production at various C.M. energies at a hadron collider.

$\sqrt{s}$	2 TeV	7 TeV	10 TeV
$\sigma_{t\bar{t}}$	7.6 pb^{-1}	157.5 pb^{-1}	385 pb^{-1}

Table 2.3: $t\bar{t}$ cross-section prediction in NLO approximation

2.6 Decay of Top Quark

Because the top quark is much heavier than the W boson, it decays weakly into a real W plus a down type quark. According to CKM matrix, most of the down type quark is b quark and the decay rate is propotional to $|V_{tb}|^2$. The value of predicted width is around 1.7 GeV for top mass = 170 GeV [7] which also implies the lifetime of top quark is

$$\tau = \Gamma^{-1} \sim 4 \times 10^{-25}\ s$$

However, the time to form a hadron (with about 1 fm diameter), $T \sim 1\text{fm/c} = O(10^{-24}s)$, is greater than the lifetime of top so it only decays weakly (Figure 2.13) and its final decay mode is predominated by W boson.

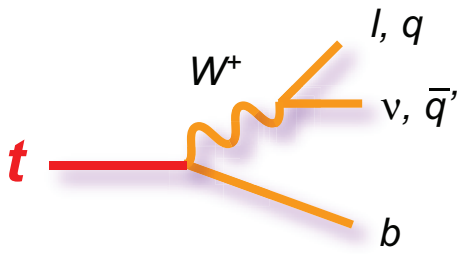


Figure 2.13: top decay

According to equation 2.31, the decay rate is proportional to $C_i|V_ij|^2$, the color factor times the corresponding element in CKM matrix. it also implies $|V_{id}| + |V_{is}| + |V_{ib}| = 1$, i = u or c. So the branching fraction from one W decay is 1 for each lepton, 3 for $u, \bar{d}'$ and 3 for $c, \bar{s}'$. Figure2.14 shows the breakdown of the decays of the $t\bar{t}$ system.

Since there are two tops in one event and top quarks decays to W boson and b quark, the topology of $t\bar{t}$ event can be classified by the decays of two W bosons which are semi-leptonic, di-leptonic and fully hadronic decay. Their definitions are listed below.

1. semi-leptonic channel: One of W decays into a lepton and the neutrino. The leptons considered are usually electrons or muons. Although the tau is the one of the decay mode, its short lifetime and its decays make the tau not as clear as the other leptons in the detectors. The other W decays into $q\bar{q}'$, either $u, \bar{d}'$ or $c, \bar{s}'$. The quarks fragment into jets which will have different feature than the b quark(jet) from top quark because of the longer lifetime of the b quark.

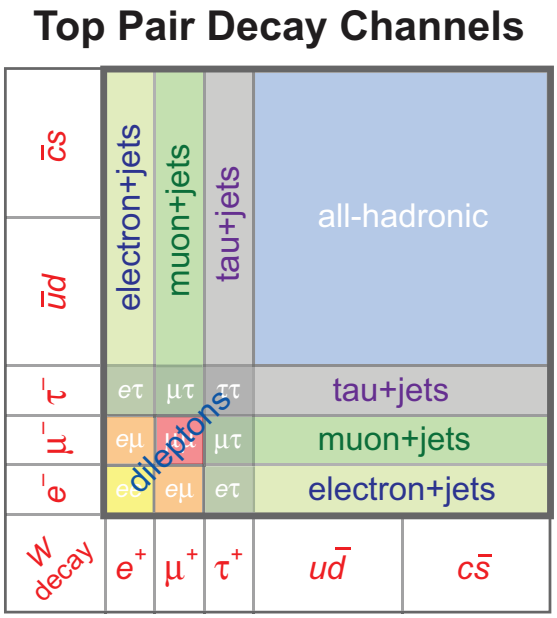


Figure 2.14: Branching ration of $t\bar{t}$ decay

2. di-lepton channel: Both W's decay into one lepton and a neutrino. The final state of these events contain two leptons plus two b jets so it is the channel with clearest signature. However the branching ratio is the smallest (about 5% without taking the tau mode into account)

3. fully hadronic channel: Both W's decay into quarks resulting in six jets in the events. Althogh this channel has biggest branching ratio (about 44%), it suffers from a very large QCD muti-jets background.

In this thesis, I will focus on the muon channel of semi-leptonic events and discuss the cross-section measurement in detail.

Chapter 3

LHC and CMS

3.1 The Performance of LHC

The Large Hadron Collider (LHC) is the present highest energy hadron collider in the world. The project started in 1994. After more than a decade of construction, LHC began its first run in 2008 and its first 7 TeV collisions in the spring of 2010. It provides much higher energy regime than previous experiments to perform the search of new physics. The goal of LHC is providing 14 TeV center of mass colliding energy with high luminosity. The luminosity is described by the formula with beam parameters

$$\begin{aligned} L &= \frac{N_b^2 n_b f_{rev}}{A} \\ &= \frac{N_b^2 n_b f_{rev} \gamma}{4\pi\varepsilon\beta^*} F \end{aligned}$$

where N_b is the number of proton per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency and A the cross section area of the bunch which is equal to $4\pi\varepsilon\beta^*/F\gamma$ where γ is the relativistic gamma, ε the normalized transverse beam emittance, β^* the beta function at the collision point and F the geometric luminosity reduction factor from the crossing angle at the interaction point.

The design performance of the machine is 14 TeV center of mass energy from two 7 TeV beams with 2808 bunches of 1.1 $\times 10^{11}$ protons each and 25 ns bunch crossing time. The expected instantaneous luminosity is $10^{34} cm^{-2} s^{-1}$. Due to the current status of magnets since the accident in the winter 2008, the beam energy is adjusted to 3.5 TeV for the early phase of LHC in 2010 and 2011. The layout of the accelerator is described in the following sections.

The LHC uses proton beams because the energy loss from an accelerating charged particle is propotional to mass^{-4} and proton is almost 2000 times heavier than an electron. However, the proton is not a point like particle, the actual rates of production also relates to the parton distribution function (P.D.F.) which gives the probability of a parton, either quark or gluon from the proton to have a given energy. Figure 3.1a shows that the probability in different x (the momentum fraction of a proton) for valence quarks, sea quarks and gluon. Thus, the production rate is deteremined according to the participants. The central mass

energy from the collision of two partons is

$$\hat{s} = (x_1 p_1 + x_2 p_2)^2 = x_1 x_2 (4 E_1 E_2) = x_1 x_2 S \tag{3.1}$$

$$\sqrt{\hat{s}} = \sqrt{x_1 x_2 S} \tag{3.2}$$

where x_i is the momentum fraction of parton i and p_i is the momentum of beam i which is approximate to E_i since the beam energy is much larger than proton mass. So $\sqrt{\hat{s}}$ is the actual collision energy from parton 1 and 2. Figure 3.1b shows that the potential to discover new physics and also shows that LHC could be a top quark factory because the threshold to produce top pair is about 350 GeV which is low x from 3.5 TeV beam energy. Therefore, the main production for $t\bar{t}$ events is gluon fusion.

Although the initial state of the partons is unknown, their transverse momentum are still considered to be small due to large beam energy in z direction. So the physics quantities are still conserved in the transverse plane. Therefore, the variable, rapidity (y), is used rather than the polar angle θ because it is Lorentz invariant. It is definited as

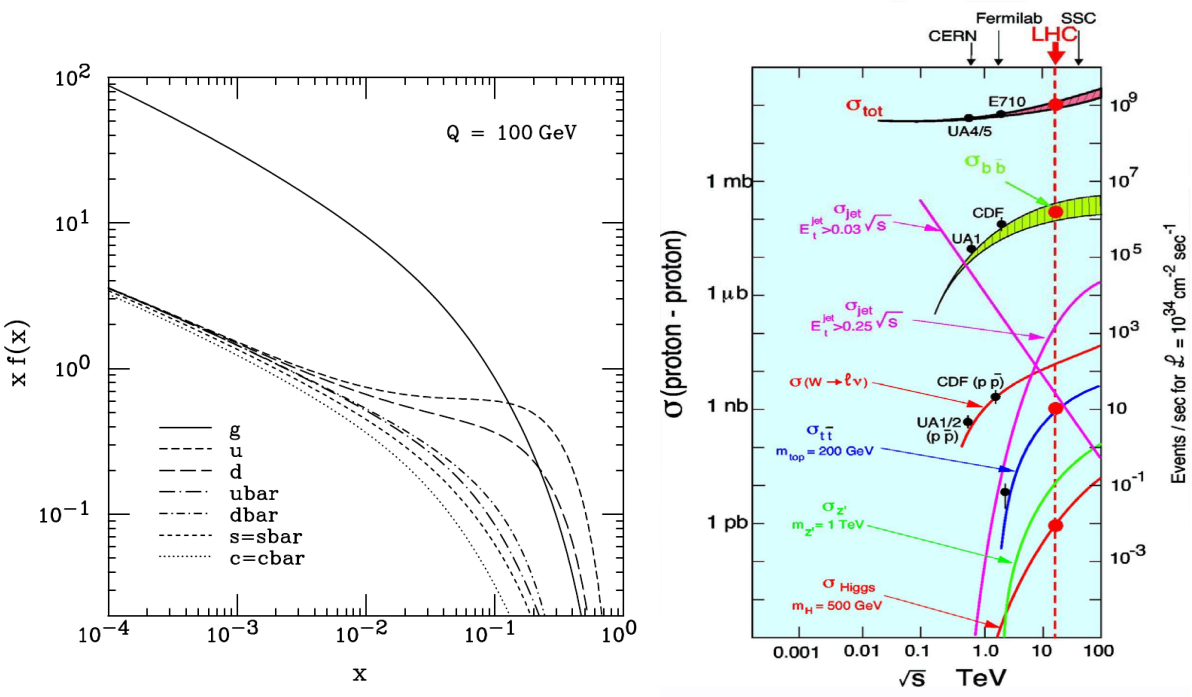
$$y = \frac{1}{2} \ln(\frac{E + P_L}{E - P_L}) \tag{3.3}$$

If the mass of the particle is small ($E \sim P$), the rapidity can be approximated by the

pseudo-rapidity, η

$$y \approx \eta = -\ln \tan \frac{\theta}{2} \tag{3.4}$$

which is conventionally used in CMS.



(a) P.D.F.

(b) Event cross-section in LHC

Figure 3.1: The P.D.F. measured by ZEUS and the cross-section of various production in term of colliding energy.

3.2 Main Components of The LHC

The major components of the LHC includes the magnet and the RF system. The magnets system include dipoles and quadrupoles magnets. Dipoles are used to bend the proton beams and quadrupoles are used for focusing the beam. In order to reach high energy

proton beams, 7 TeV by original design, the dipole magnet system needs to provide about 8.36 Tesla which is propotional to the required current and would result in a huge power consumption. Therefore, a superconducting magnet system was adopted. The material is classical NbTi superconducting material. However, the main challenge is the cooling of the magnets which use the superfluid helium to get the temperature below 1.8K. For the quadropoles, they are cooled to 4.5K. The magnets system of LHC overall contains 1232 dipoles and 858 quadrupoles.

The other crucial component is the RF system. RF system (cavities) are mainly used for capture, accelerating and storing particles (compensating the energy loss). There are two independent RF systems because the two colliding proton beams are in separate beam pipes. Each RF system includes two major parts, a main 400 MHz accelerating system (ACS) and a 200 MHz capture system(ACN).

In addition to the main 27 km LHC ring, the whole accelerator complex includes other smaller accelerators. The schematic of the complex is showed in Figure 3.2. The injection chains of beam particles start from Linac2 to the Proton Synchrotron Booster(PSB) where protons were produced and brought to 1.4 GeV. The protons then go to the Proton Synchrotron (PS) and then the Super Proton Synchrotron(SPS). The beam energy is ramped up from 25 GeV to 450 GeV. The final stage is the main LHC ring. It accelerates the protons

from 450 GeV to the desired energy.

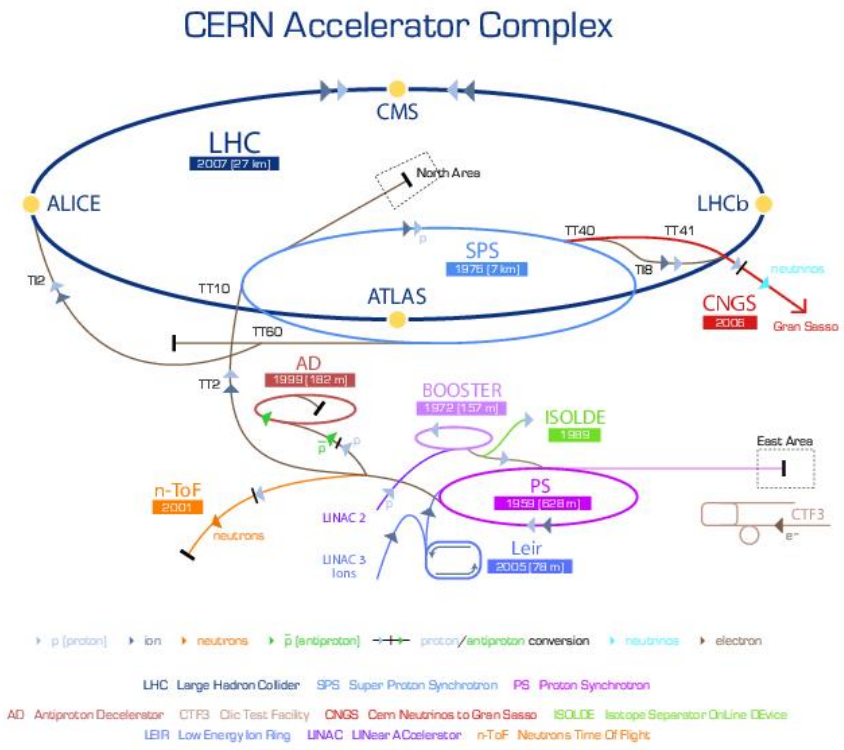


Figure 3.2: LHC Injection Chain

3.3 CMS Overview

CMS is one of four experiments at the LHC. With the aim to search for the Higgs and new physics, a general purpose detector was designed. The common signature of new physics is the high p_T objects in the events which brought the setup of the superconducting solenoid of 3.8 Tesla magnetic field in CMS in order to provide enough bending power for momentum measurement. The other general concepts of design is fine granularity of

detectors to improve resolution of momentum and energy measurements. Fine granularity also lowers the occupancy of read-out to reduce the effect from pile-up events but a large number of read-out channels requires effective on-line selection, massive data handling and synchronization. The detailed layout of the detectors is described in the following sections.

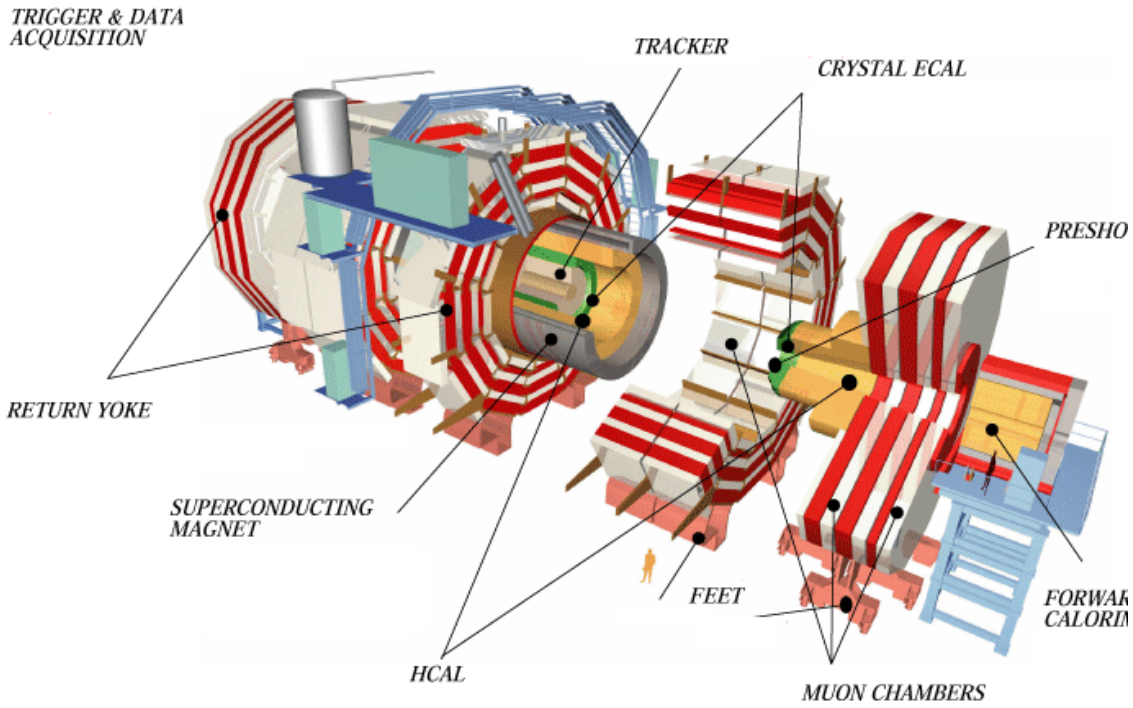


Figure 3.3: CMS Detector

3.3.1 Magnet

The CMS magnet system is a superconducting magnet with a 12.5 m long and 6 m diameter solenoid plus 10000 ton flux return yoke. The solenoid is designed to provide an axial 4 T field and its major feature is that the cold mass is made by winding 4 layers of NbTi

conductor. The flux return yoke is composed of 5 barrel wheels and 3 endcap disks on each end. In current operation, the CMS magnet provides 3.8 Tesla field to the experiment which mainly provides the bending power for central tracker. The return yoke conduct about 1 to 2 Tesla field which is used for the tracking of muon system. Figure 3.4 shows the map of magnetic field strength.

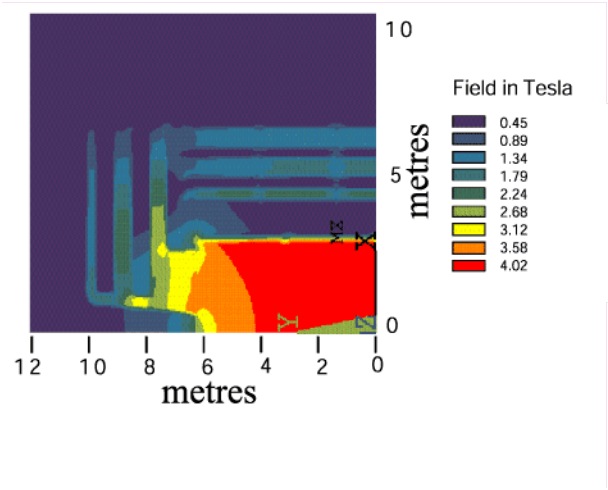


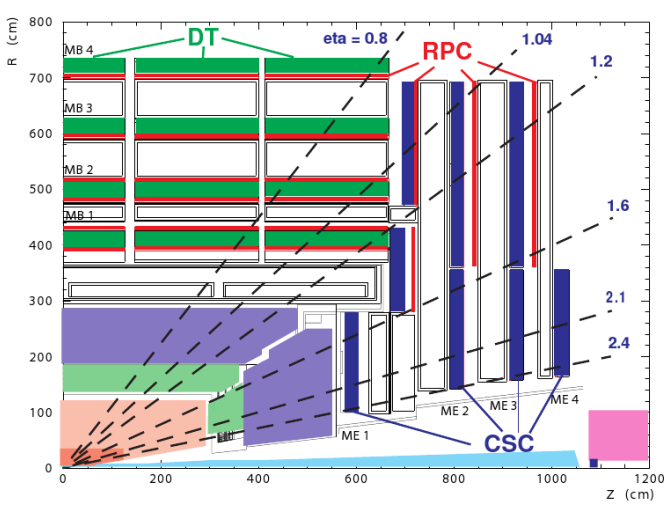
Figure 3.4: CMS magnetic field map

3.3.2 Muon System

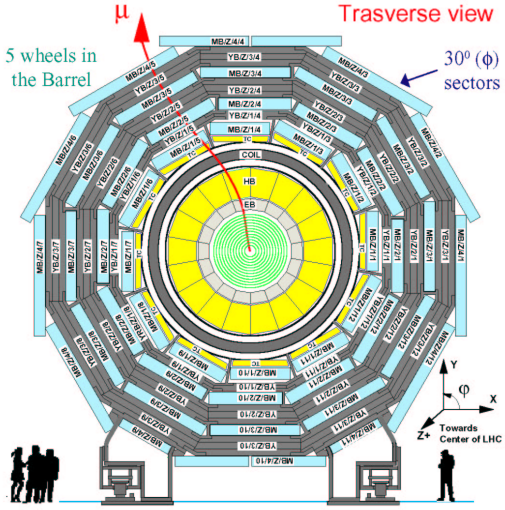
Muons have less interactions with material because they are heavier than electrons which results in fewer electromagnetic interactions and as they are leptons they don't interact via the strong interactions. So the muon chambers are placed at the outermost of CMS detector. By sandwiching muon chambers with the magnetic field return yoke, most punch-through particles can be stopped at the first layer. In addition, the muon chambers are in a 1 to 2 Tesla magnetic field which makes momentum measurements possible using only the muon

chamber system.

CMS has three types of muon detectors. They are Cathode Strip Chamber (CSC), Drift Tube (DT) and Resistive Plate Chamber (RPC). The full system covers the η range up to 2.4. The CSC's are used in the endcap region, the DT's are in the barrel and the RPC's are in both endcaps and barrel. The details of the hadware are shown in Figure 3.5a and described below.



(a) Muon system



(b) CMS transverse view

Figure 3.5: CMS muon system

Cathode Strip Chamber

Two endcap muon systems of CMS are assembled with 4 layers CSCs and 3 layers magnetic return yokes on both sides. The return yokes provide the bending power for muon tracks and mount all chambers. The structure of the system has 4 stations which are la-

beled as ME1, ME2, ME3 and ME4. ME2 and ME3 have 2 rings. ME1 has 3 and ME4 only has 1. There are total 7 types of CSCs. The geometry of the chamber is trapezoidal shape with a width from 60 cm to 150 cm, length from 160 cm to 330 cm and the height of 25 cm. (Figure 3.6shows the scheme of a CSC chamber).

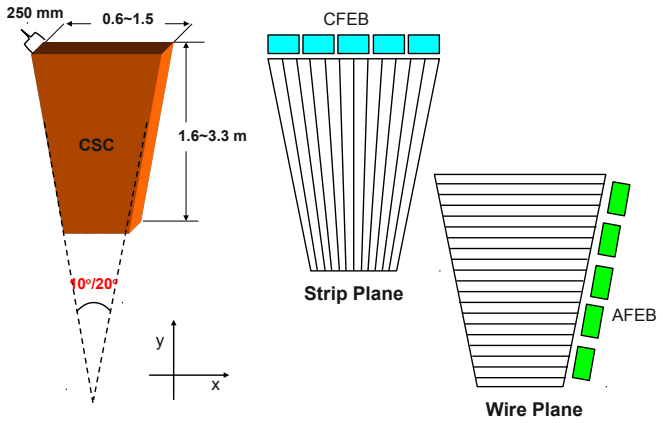


Figure 3.6: The Scheme of CSC

The technology of the CSC chambers is on the base of multiwire proportional chamber (MWPC). There are 6 layers in one chamber. Each layer consists of an horizontal anode wires layer and a radial cathode strips plane which are othogonal to the anode wires. When charged particles go through chamber, the ionization of the gas develops an avalanche toward the wire and induces a signal on the strip (Figure 3.7). Therefore it forms a 2-dimesional hit at one layer and at most 6 hits in one chamber for a single charged track stub. The induced charge distribution on the strips can be described by Gatti formula [47]

$$\Gamma(\lambda) = K_1 \frac{1 - \tanh^2(K_2\lambda)}{1 + K_3 \tanh^2 K_2\lambda} \tag{3.5}$$

$$\lambda = \frac{x}{h} \tag{3.6}$$

where x is the cooridinate perpendicular to the orientation of the anode wires. h is the spacing between anode wire plane and strip plane. K_1, K_2 and K_3 are parameters which are different for each design. After the fitting, the spatial resolution of strip plane is as good as 50μm. The signals on the strip plane are read out by CFEBs(Cathode Front-Ends Board). There are 5 CFEBs for each chamber except ME1/1 only have 4 CFEBs. The anode wires are grouped into several wire groups in one plane and each wire group is connected to a AFEB(Anode Front-End Board). The position from wire plane is given by the coordinate at the middle of the wire group. Therefore, the width of the wire group is the resolution.

Each chamber in ME2/1 , ME3/1 and ME4/1 covers 20 degree in ϕ. Other chambers cover 10 degree. The arrangement of chambers in each ring is slightly overlapped in order to get complete ϕ recoverage and the orientation is pointing to beam pipe which gives precise ϕ resolution from the intrinsic spatial resolution of strip pitch and its signal reconstruction.

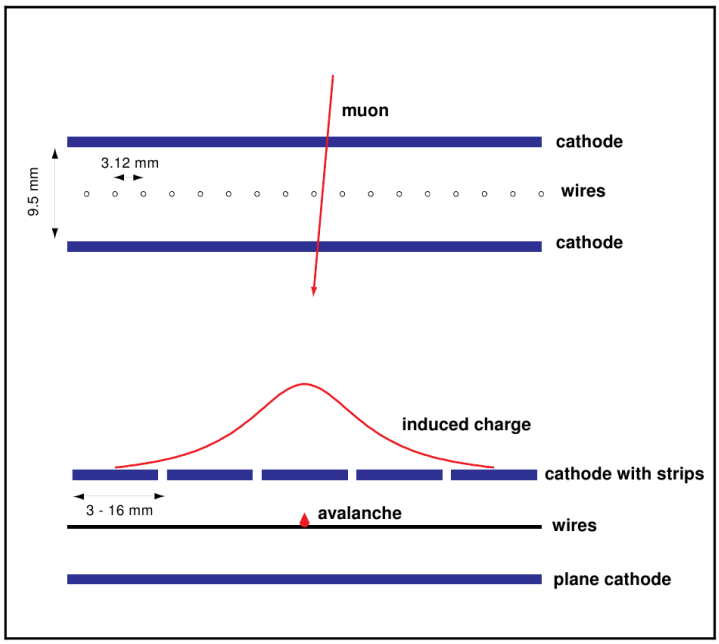


Figure 3.7: Principle of cooridinate measurement of CSC

Drift Tubes Chamber

The barrel muon system is made of four stations of drift tubes stacks (MB1 $\sim$ MB4). The DT chambers suround the beam line forming concentric cylinders with 12 sectors in one wheel (as shown in Figure 3.5b. There are in total 5 wheels in the barrel muon system which cover the η range from -1 to 1. The structure of each DT chamber consists of 3 super-layers which were made by DT tubes and 1 aluminum honeycomb support stucture. The outer two super-layers are used to measure ϕ and the central one is used to measure the z position. The fourth layer MB4 has only 2 super-layers without the z-position measurement.

Each drift tube cell contains a anode wire in the center, a field electrode on top and buttom and cathode strips at the two sides. The electric field in the tube cell is formed and shaped by charging the anode with +3.6 kV, the cathode strips -1.2 kV plus the +1.8 kV for the field strips. By stacking four layers of drift tubes, a super-layer is formed and provides a 2-D muon segment. Two ϕ segments plus one z segment from 3 superlayers (1 DT chamber).

Resistive Plate Chamber

Unlike the DT and the CSC, the RPC is not a wire chamber. It is formed by two high resistive planes(Bakelite) separated by a 2 millimeter gas gap and a strip plane is placed between two resistive planes. It combines adequate spatial resolution and good time resolution with fast response which can differentiate muons in two consecutive bunch crossing (BX). So the main purpose of the RPCs is providing the infomation of muon trigger with the assignment of BX and therefore they are placed in both endcap and barrel region.

3.3.3 Hadron calorimeter

The goal of the hadron calorimeter is to measure the hadronic activity in the events. It is designed to stop the hadrons and measure their energy, the jets and $\not\!\!E_T$ in particular. There are two major parts inside the solenoid, the barrel (HB) and the endcap (HE), which have the η range of $0 < |\eta| < 1.4$ and $1.3 < |\eta| < 3.0$. A pair of forward calorimeters (HF) are placed at outside of the endcap muon chambers which cover $|\eta|$ up to 5.2. The outer

barrel calorimeter (HO) is used to imporve the central shower containment in the region $|\eta| < 1.26$. An unit tower of the HCAL is formed by a $\Delta\eta \times \Delta\phi$ area from the scintillater tile and the depth of 17 layers of absorbers and scintillators sandwiches. The readout of the scintillator are from the attached wave length shiftting fiber (WLS). The detail geometry information of each part of the HCAL is listed below.

The HB

There are two half HB barrels (HB+ and HB-). Each of them contains 18 equal wedges in ϕ, 16 η sectors and 17 layers absorbers and scintillators. Each ϕ wedge is segmented into 4 scintillator tiles and each η sector fit one which implies the tower area is 0.087×0.087. The absorbers are mostly made of brass plates except the innermost and outermost layer use stainless steel for structural strength.

The HE

The HE is the endcap region of hadron calorimeter. The η range of HE covers 34% of the solid angle. The detector must withstand a high counting rate and the electronics should have sufficient radiation hardness. The tower area ($\Delta\eta \times \Delta\phi$)of HE is also 0.087×0.087 for $|\eta| < 1.7$ and is about 0.175×0.175 for $|\eta| > 1.7$.

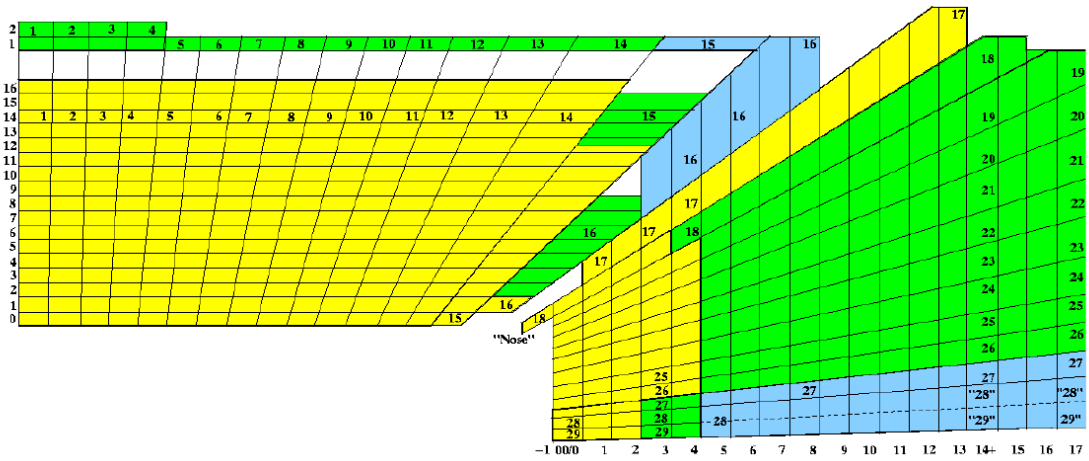


Figure 3.8: HCAL

The HO

In the central region ($|\eta| < 1.3$), the HB doesn't have enough depth to fully contain hadronic showers. So an extended layer, HO, was added outside the solenoid. The solenoid itself is an additional absorber which has equivalent interaction length about 1.4/sinθ. Since the HO is outside the main magnet, its geometry is contrained by the field return yoke.

The HF

The HF is a very forward detector which covers the range of $3 < |\eta| < 5$. It is divided into 36 wedges in ϕ and 12 segments in η that gives the tower area $\Delta\eta \times \Delta\phi \sim 0.175 \times 0.175$. On average, each proton-proton collision deposits about 760 GeV in HF (about 100 GeV in the others). Quartz fiber (fused-silicon core and polymer hard-cladding) is used for the scintillation material because of the requirement of radiation hardness in this kind of environment. The quartz fibers measure the Cherenkov light from charged particles which

results in a sensitivity of the HF to electromagnetic component of showers.

3.3.4 Electromagnetic calorimeter

The CMS electromagnetic calorimeter is made of the lead-tungstate ($PbWO_4$) crystals. Its major function is measuring the energy of electromagnetic showers and identification of photon and electrons. The design is very compact and hermetic in order to archieve the capability to detect the two photon decay from the postulated Higgs boson.

The $PbWO_4$ crystal has several characteristics. First, due to the high density (8.278 g/cm^3), it has an increased probability of initiating an electromagnetic shower from incident particles. Second, the short radiation length (X_0 = 0.89 cm) makes the size compact (the depth of the crystal is about $25.8X_0$). Third, the small Moliere radius(2.2 cm) results in a small radius of shower which confines the energy deposit and decreases the energy leaking to nearby crystals. Lastly, its scintillation decay time is of the same order of the LHC bunch-crossing time (80% of light emitted in 25 ns). The energy resolution of ECAL is parameterized by applying the formula

$$(\frac{\sigma}{E})^2 = (\frac{a}{\sqrt{E}})^2 + (\frac{\sigma_n}{E})^2 + c^2 \tag{3.7}$$

where a is the stochastic term, σ_n the noise and c the constant. The results from test beam is shown in Figure 3.9.

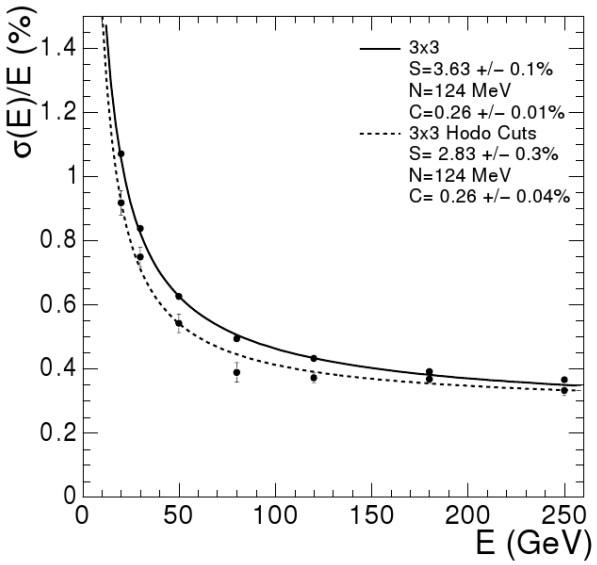


Figure 3.9: ECAL energy resolution

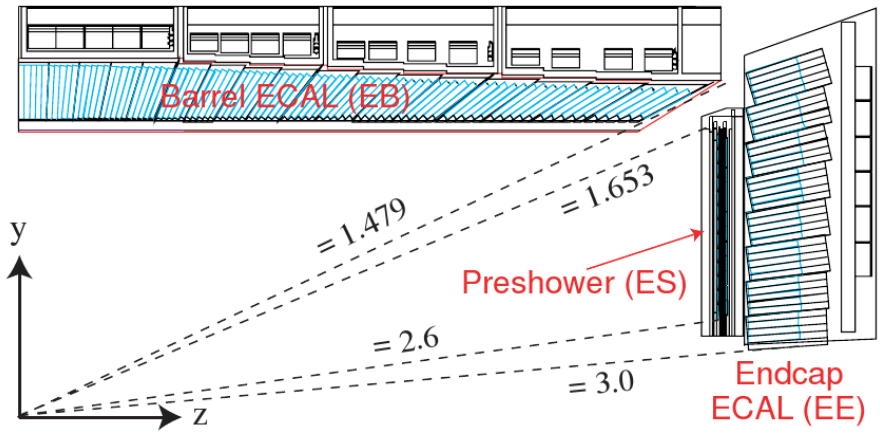


Figure 3.10: CMS ECAL

The CMS ECAL consists of three subsystem, barrel ECAL(EB), endcap ECAL(EE) and preshower(ES) shown in Figure 3.10. The EB covers pseudorapidity of $|\eta| < 1.48$. There are two barrels (EB+ and EB-). Each of them has 85-fold in η and each crystal covers 1 degree in azimuthal angle ϕ. So the every crystal has the cross-section area of 0.0174 $\times$ 0.0174 in $\Delta\phi \times \Delta\eta$. The avalanche photodiodes (APD) are used for the readout of the scintillation light.

There are two endcap in EE with the range of $1.48 < |\eta| < 3.0$. Each of them is divided into 2 halves (Dees) and each Dee consists 3662 crystals which are grouped in a unit of supercrystal (SC) with 5$\times$5 and contains 138 SCs and 18 partial SCs. The $\Delta\phi \times \Delta\eta$ area is about 0.05 $\times$ 0.05. The readout electronics use vacuum photottriodes (VPT). VPT is photomultiplier which is specially developed to use in 3.8T magnetic field of CMS.

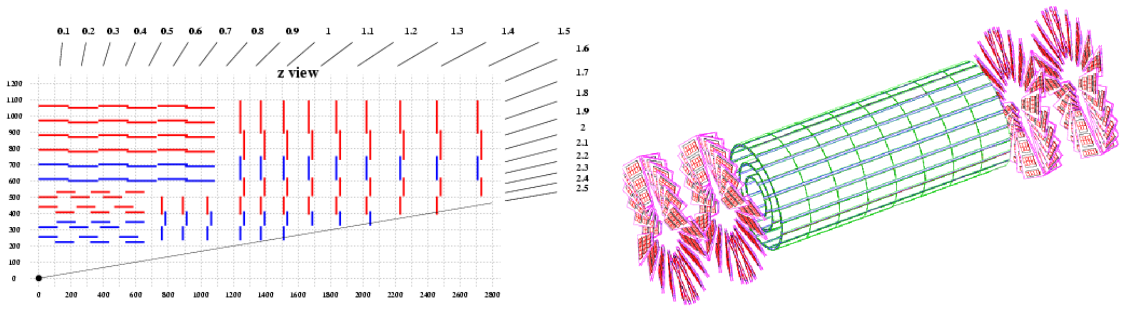
The third system is ES which is placed in front of the EE and covers the range of $1.653 < |\eta| < 2.6$. Unlike the other two systems with crystal base system, ES is a two layers sample detector. A lead radiator in fornt to initiate the showering process and the silicon strip sensors behind for measuring the energy deposit and the position. The reason to add preshower is to improve the identification of neutral pions and the electrons against minimum ionizing particles. Because the silicon strip sensors have better spatial resolution, the ES also improves the position resolution of eletrons and photon.

3.3.5 Tracker

The CMS tracker aims to provide a precise measurement of the trajectories and momentum of charged particles. To fullfill the physics requirements, a design of a silicon based tracker was adopted. It has a size with the length of 5.8 m and the diameter of 2.5 m and is immersed in a homogeneous 3.8 T magnetic field for the bending power. At the LHC design luminosity, an average of 1000 charged particles will be produced. So the fine granularity and large amount of front-end electronics are needed because the demand of fast response and efficiency to identify trajectories. All the criteria results in the techinical challenge regarding the minimum amount of material in order to avoid photon conversions, multiple scattering, bremsstrahlung as well as nuclear interactions and the radiation hardness for the lifetime of the experiment operation

The tracker system includes two parts, one is the strip tracker and the other is the pixel detector which is used for vertex measurements. The layout of strip tracker has 10 layers in barrel and 9 layers in endcap covering the pseudorapidity range of $\eta < 2.5$. It is categorized into five parts, 4 layers of the tracker inner barrel (TIB), 6 layers of the tracker outer barrel (TOB), 3 layers of the tracker inner disk (TID) at each end of TID and 9 layers of the tracker endcap (TEC) at each endcap. For the pixel, there are 3 layers in barrel (BPix) and 2 layers at each endcap(FPix). The detail is shown in Figure 3.11.

The goals of the CMS tracker are listed below.



(a) tracker (b) pixel

Figure 3.11: CMS tracker system

- High reconstruction efficiency and good momentum resolution. The efficiency for muons with p_T greater than 1 GeV is higher than 98% in the full η range of the tracker. Electrons also have over 90% efficiency. For track reconstruction in the jets or other hadronic environments, hadrons have an efficiency over 95% for $p_T > 10$ GeV or 85% with $p_T > 1$ GeV. The momentum resolution is a function of p_T.

$$\frac{\Delta P_T}{P_T} = \sqrt{(0.15 P_T)^2 + (0.5)^2}\ \% \tag{3.8}$$

with the unit of p_T in GeV/c. In the forward region $|\eta| > 1.6$, the momentum resolution degrades to

$$\frac{\Delta P_T}{P_T} = \sqrt{(60 P_T)^2 + (0.5)^2}\ \% \tag{3.9}$$

- Vertex reconstruction. This is used in primary vertex reconstruction, heavy flavour tagging and identifying tau decays. The impact parameter resolution in the transverse

plane is better than 35 μm and the longitudinal impact parameter is better than 75 μm in the full η range. For the b-tagging efficiency, the b jets with p_T between 50 GeV to 200 GeV have 50% tagging efficiency with 1% mistagging rate. In addition, the vertex reconstruction also helps to identify photon conversion in the tracker. About 22% photon convert in the tracker and about half of them can be identified from e^+e^- pairs.

3.3.6 Luminosity Measurement

Luminosity is used for the overall normalization of physics analysis. According to the formula, $R = L\sigma$, R is the production rate and the σ is the cross-section of the production. On the other hand, the same formula can be used to determine the luminosity. In CMS, the luminosity measurement uses the "zero counting" method from zero bias events by the HF. The goal of real-time luminosity (instantenous luminosity) measurement is under 1% statistical accuracy with an update rate of 1 Hz.

The "zero counting" method is applying the Possion statistics

$$p(n;\mu) = \mu^n \frac{e^{-\mu}}{n!}$$

where μ is the mean of interaction($\sim$ 25 for the full luminosity of the LHC), n is the number of interaction. Therefore, counting the zero interaction events gives the $\mu = -\ln p(0)$ and it also equal to

$$\mu = \frac{L\sigma}{f_{BX}}$$

f_{BX} is the revolution frequency of bunch-crossing(BX), σ is 80 mb [3] from the calculation of the cross-section using the current understanding of proton-proton interaction. Thus the luminosity(L) can be obtained. However, once the instantenous luminosity is high, the rate of zero interaction event is low which limits the statistical accurancy. A modified method to improve the statistic of zero interaction events is counting the zero signal towers since each tower has the same occupancy in each bunch-crossing.

There are total 864 = 2 $\times$ 12 $\times$ 36 towers in the forward and backward HF (with 36 wedges in ϕ and 12 segments in η). The number of empty tower is binomial distribution denoted with the probability p_0. The zero signal towers are defined as the HF towers which are not above a E_T threshold cut. Thus, the average fraction of zero towers in the HF is

$$< f_0 >= \frac{1}{M}\sum_{k=0}^{\infty} M\frac{e^{-\mu}\mu^k}{k!}p_0^k \approx e^{\mu(p_0-1)} \tag{3.10}$$

M is the number of bunch crossing in the period of time for measurement. With the limit that p_0 is close to 1, $< f_0 >$ is approximate to p_0^{μ} and μ is then obtained by taking the logarithm of $< f_0 >$.

The drawback of this method is that E_T threshold can not be a complete separation between signal and noise. Thus, cutting off the signal is unavoidable. The other alternative method is measuring the total E_T deposited in the HF. Because the avarage E_T is also proportional to the avarage number of interaction μ by replacing the probability of zero tower with the probability of E_T distribution $p_{E_{\mathrm{T}}}$

$$< E_T >= \frac{1}{M}\sum_{k=0}^{\infty} M\frac{e^{-\mu}\mu^k}{k!}p_{E_{\mathrm{T}}} = a\mu + b \tag{3.11}$$

where a and b are the constant along with $p_{E_{\mathrm{T}}}$ which are from the probability density function of signal and noise of the HF towers [4].

3.3.7 Trigger and DAQ

The CMS Trigger and Data Acquisition (DAQ) systems control the preliminary event selection and data taking process. With the full luminosity LHC, approximately 20 inelastic interactions will happen every bunch crossing. This scenario results in almost 1 GHz event rate. With a mean event size of about 1 MB, this huge amount of data exceeds the ability of the computing and the capacity of data storage. Therefore, an effective DAQ system and the physics oriented trigger are the principles of the design. There are three major elements in the trigger and DAQ system, Level 1 trigger (L1), High Level Trigger (HLT) and the overall DAQ system which are described in the following sections.

Level 1 Trigger

The Level 1 Trigger (L1) is the first step to suppress the high event rate from bunch-crossings. It reduces the rate of an order of 10^7 to 100 kHz which is the input capability of the HLT. The maximum of L1 trigger output rate is set by the average time to read information for processing by the HLT and the average time for completion of the HLT algorithms. After the crossing, the data is stored in the pipeline buffer of the front-ends. The depth of the buffer is 128 BX which is equivalent to 3.2 μs. This is also the required time for L1 data read-in, transmit, calculation and decision propagation. After L1 acceptance (L1A) propagate to all detectors, the data in pipeline buffers as well as the trigger objects from L1 trigger system are then readout by the DAQ. In order to avoid deadtime, the trigger electronics must be able to accept data every 25 ns. The architecture of the trigger flow is

shown in Figure 3.12. After a L1 acceptance received by detector front-ends, data including the L1 trigger are readout.

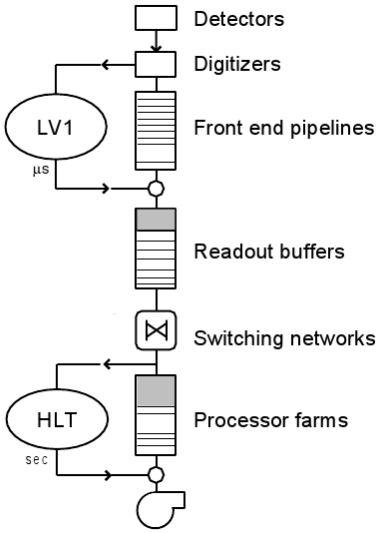


Figure 3.12: The Trigger Flow

Due to the high event rate, the L1 trigger only uses information from the muon systems and the calorimeters. It has four major components, Calorimeter trigger, Muon trigger, Global trigger and TTC (Timing, Trigger and Control System). The detail organization of L1 trigger is shown in Figure 3.13 and discussed below.

- The calorimeter trigger includes three subsystems, ECAL, HCAL and HF. The tower energy are summerized and sent to the Regional Calorimeter Trigger (RCT). The RCT find the candidates of photon, electrons, tau and jets and then passes them to the Global Calorimeter Trigger (GCT). After calculating the total transverse energy and missing energy, the GCT transmits the top 4 objects of each type to Global trigger.

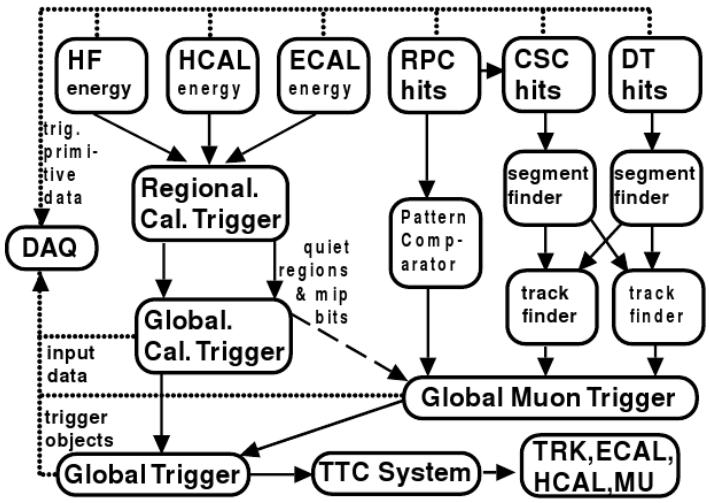


Figure 3.13: L1 Trigger

- The muon trigger also has three subsystems, CSC, DT and RPC. CSC and DT form the segments from each muon chamber locally and build the tracks in their track finder. The RPC uses a pattern comparator to form the segments and build tracks. The Global muon trigger receives the tracks from each system and sorts the tracks by p_T. The GMT sends the top 4 muon tracks to Global trigger.

- The global trigger receives the information from the calorimeter and muon trigger. It synchronizes the data from each sub-system and makes the decision to accept or reject the data. All the trigger objects contain the coordinates (η, ϕ) and p_T and are read out by DAQ system after L1A.

- The TTC communicates with the detector front-ends and the L1 trigger. It receives the decisions from the global trigger. It distributes a precise 40 MHz bunch crossing clock and the L1A signal to the sub-systems if the data is accepted.

Because the most important goal for the trigger is selecting potential events with the signature of specific physics interests, the L1 trigger has the following physics requirements.

- The general requirements for all trigger objects are within the pseudorapidity range of $|\eta| < 2.5$. The selection efficiency is greater than 95%.
- For single leptons, the p_T threshold is greater than 40 GeV.
- For di-leptons, the p_T threshold is 20 GeV and 15 GeV for the first and the second lepton respectively.
- For single photons or di-photons, the criteria is the same as the single lepton and di-lepton.
- For single and multi-jets, $|\eta| < 5$ is required in order to reconstruct jet spectrum that overlap with the data of lower energy collider like Tevatron.
- The $\not\!\!E_T$ trigger is set with a threshold of 100 GeV.

HLT Trigger

After the L1 accept is distributed, the data is read out from 512 front-end buffers to the buffers of the readout units. Then the network delivers the data to the processor farms where the events are assembled and the HLT is run. Although the event rate has been reduced to 100 kHz (75 kHz in the startup condition), a pre-selection filter is still performed

to refine the events in order to integrate the data from tracker and the full granularity of the calorimeters for futher use and more accurate object selections. In this thesis, the muon plus jets final state from $t\bar{t}$ events are used. Thus, the HLT trigger path, HLT_Mu9 and HLT_Mu15, are both applied for the data analyzed in this thesis so the discussions of the performance will focus on the muon object.

There are two steps in HLT muon identification. The first is using the stand-alone muon systems and the next step is integrating with tracker and calorimeters for a better precision of the measurements. The parameters used in the HLT are p_T, E, η, ϕ and isolation. According to the general signatures of the physics, the muon HLT selection is classified into two types, single-muon selection and di-muon selection.

- The single-muon selection has to pass the following criteria in different levels. From level 1, the low quality CSC tracks must match with RPC tracks by the Global Muon Trigger to ensure the p_T measurement. At the level 2, the muons must contain the standalone muon component and be extrapolated to a valid collision vertex. The barrel muons must have at least one DT segment. In addition, the sum of DT segments and RPC hits must be greater than three. At the level 3, muons must have more than 5 tracker hits from pixel or strip sensors.

  The performance is examined by evaluating the efficiency in terms of various parameters. Figure 3.14 shows the efficiency in term of η. The events were generated in a

flat p_T interval between 5 to 100 GeV/c with $|\eta| < 2.1$. The dips at $\eta \sim 0.3$ and 0.8 are caused by the gap between DT wheels. Minor effects can also be found between CSC rings at $\eta \sim 1.2$ and 1.6.

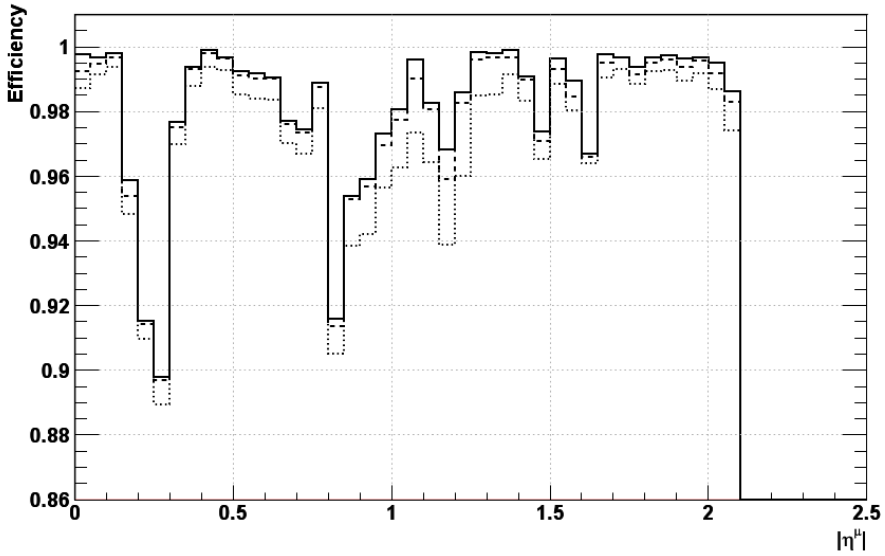


Figure 3.14: The HLT efficiency for single muons which pass Level-1(solid), Level-2(dashed) and Level-3(dot-dashed) trigger level as a function of η. The Level-3 trigger is corresponding to HLT.

The efficiency is also a function of muon p_T. It affects the trigger threshold which is defined at 90% efficiency. In order to evalute the turn-on curve for different trigger thresholds, the flat p_T samples are used. The results are shown in figure3.15. Four different p_T thresholds, 10 GeV/c, 20 GeV/c, 30 GeV/c and 40 GeV/c, are shown in sub-figures a,b,c and d respectively. Also shown are the different turn-on curves for the Level-1 (solid line), Level-2 (dashed) and Level-3 (dotted) selection.

Another important parameter for HLT efficiency is isolation. The isolation has strong effect to the trigger rate especially when lower p_T threshold are applied. The reason

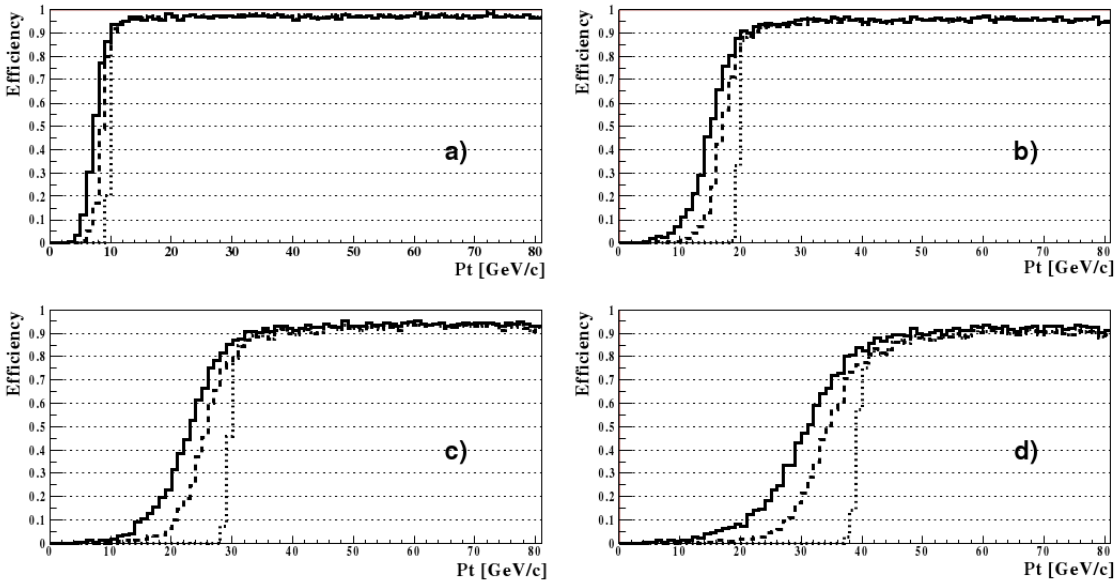


Figure 3.15: The single muon HLT efficiency as a function of p_{T} in three different trigger levels, Level-1(solid), Level-2(dashed) and Level-3(dot-dashed) trigger. The Level-3 trigger is corresponding to HLT.

is that the source of muons has larger fraction in K, π and heavy flavor (b,c) decays at low p_T and those decays are mostly involved with hadronic activities which accompany the muon. Figure 3.16 shows the level-3 trigger rates from various source of muons. It also reflects that the muons from W and Z decay are well isolated and dominate the single muon trigger at high p_T threshold.

The efficiency of W $\rightarrow$ $\mu\nu$ and $t\bar{t}$ $\rightarrow$ μ+ X are also examined since these two processes are the most important channels of single muon events to study the standard model physics. The Figure 3.17 shows the results. Below a p_T threshold 10 GeV/c, both type have a efficiency greater than 90%. Thus this trigger has a selection efficiency in the start-up scenario of the LHC.

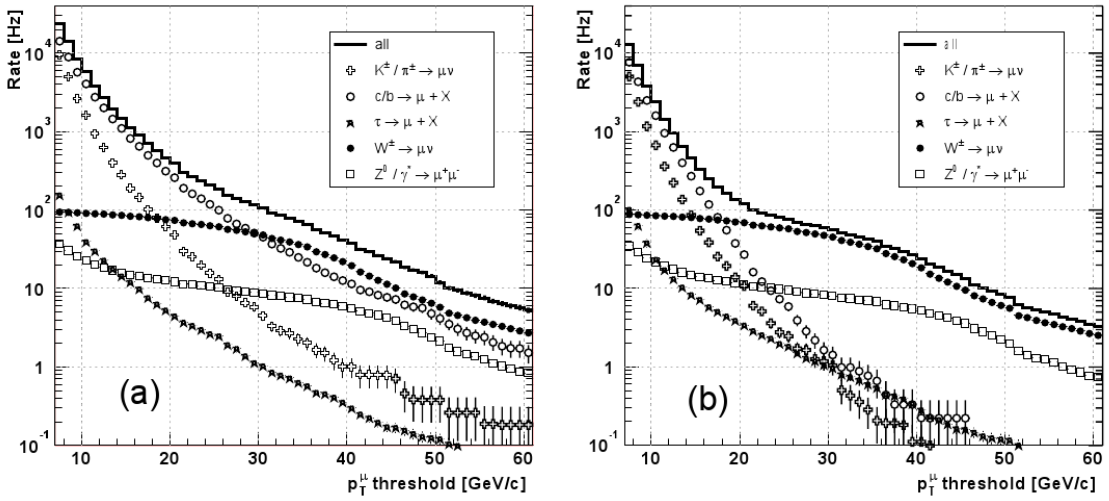


Figure 3.16: The single muon HLT efficiency before (a) and after (b) the isolation cut

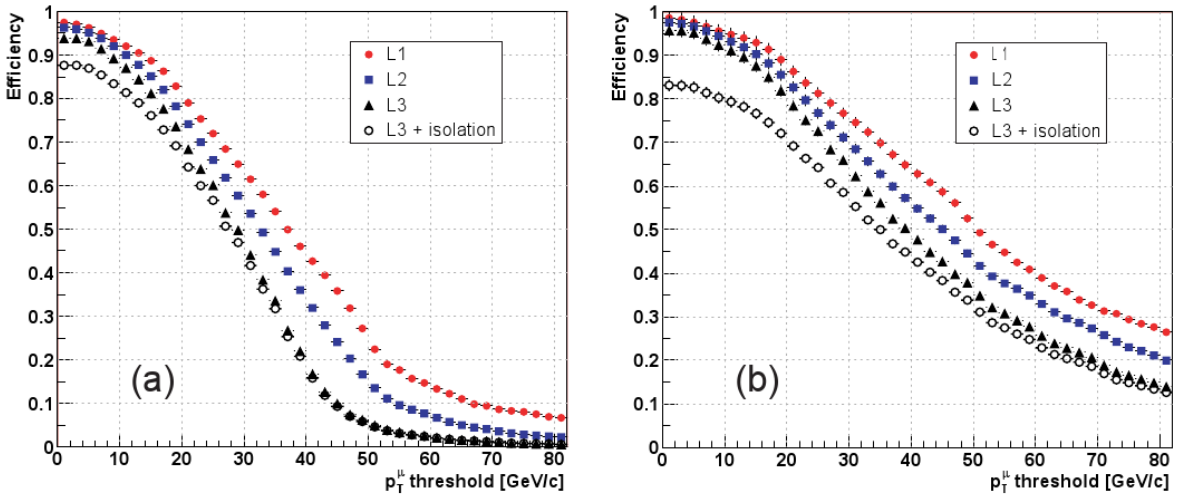


Figure 3.17: The single muon trigger efficiency of (a) W $\rightarrow \mu\nu$ and (b) $t\bar{t}$ events in different trigger level. The L3 is corresponding to HLT.

- The di-muon selection has same selection requirements as the single-muon channel except only one of muons needs to be isolated. Another requirement is that both muons should come from the same vertex within 5 mm in z coordinate and they have to be separated by $\Delta\phi > 0.05$, $|\Delta\eta| > 0.01$ and $\Delta p_T > 0.1$ for rejecting duplicated fake tracks.

Data Acquisition

In addition to the L1 trigger and HLT, the DAQ contains the following elements that interact with L1 triggers, manage the data flow, build the complete events and execute the HLT algorithms. The architecture is shown in Figure 3.18 and the function of each unit is summarized below.

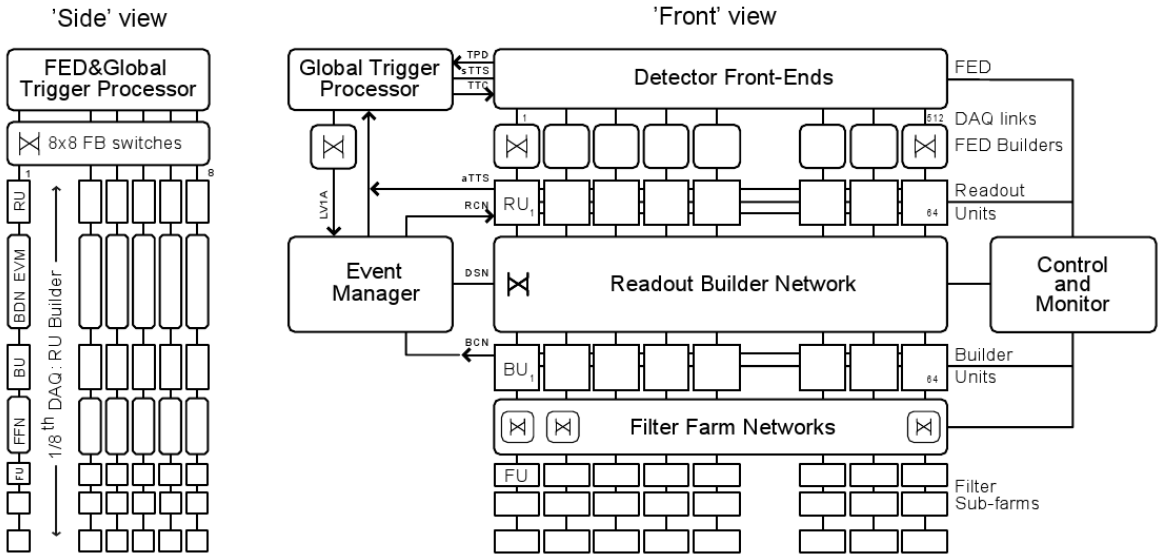


Figure 3.18: The DAQ Architecture

- There are approximately 700 detector front-ends modules. They store the raw data in the pipeline buffers and wait the decision to accept or reject the data from L1 trigger.
- Readout system: Readout the data from the front-end modules.
- Builder network: a switch based network which connects the readout and filter system.

- Filter system: including builder units and filter farm network and system. It runs the HLT algorithms and builds event fragments into full event buffers.

- Event manager: controls the data flow from readout units to filter farms. It also synchronizes the overall DAQ system with the L1 trigger.

- Run controls and monitor: responsible for the configuration, control and monitoring of all elements as well as supplying the user interface.

- Computing serving: the processors and networks which receive events from filter farms.

Chapter 4

CMS Software and Reconstruction

4.1 Software Framework

CMS software (CMSSW) has a structure with following major elements.

Framework and Event Data Model (EDM)

The framework and EDM are the backbones of CMSSW. The concept of the framework is to modularize the software which allows the development of each component independently. It requires that every module should have a well defined event-processing functionality. The modules are not allowed to communicate directly with each other but only interact through the event.

The Event Data Model describes the process between data and the event. Events are passed through a sequence of modules. The modules can read the data from the events or

add more data back to the events. The event data architecture is also modular similar to the framework. Different data branches in the event can be configured or used by different applications.

There are six types of modules which are defined in the framework.

- Pool source. It is the source of events which can be read in from a ROOT file (for example, the output of global DAQ) or be empty events which will be filled by a data simulation generator.

- EDProducer. The EDProducer will read data from event, produce a new data format and then put it back to the event. One example is the reconstruction of physics objects. These EDProducers read the information from the detector such as calorimeter towers or track segments and then produce physics objects such as jets or charged tracks.

- EDFilter. It is a module which determines whether to stop or continue the processing of the event by reading the relevant data.

- EDAnalyzer. Individual studies of data can be done by using an EDAnalyzer. It is not a module for data processing as it doesn't have the ability to write data back to the event. The output of the EDAnalyzer can be an independent root file from the analysis.

- EDLooper. A module is used to control "multi-pass" looping over the data of input source.

- Output module. Once all the scheduled paths have been executed then output the final product.

Simulation

There are two major simulation tasks in the CMS experiment. One is event generation and the other is detector simulation. In order to fulfill the goal of various studies, CMSSW has a generator interface which incorporates different kinds of generators. In this thesis, the general purpose generator, Pythia6[58] and the matrix element (ME) calculation one, Madgraph[8] are both used for the method development and efficiency calculations.

At the part of the detector simulation, CMS uses GEANT4 to describe the CMS detector. In addition, the sensitivity of detector behavior, track selection mechanisms, hit collection and digitization which simulates the electronic readouts are also implemented.

Reconstruction

The reconstruction is the process to construct the physics quantities from the information collected by the detector. In CMSSW, the reconstruction is modularized by EDProducer so the whole reconstruction chain is a series of independent objects. It can be classified in three major parts: local reconstruction, global reconstruction and combined physics objects. The local reconstruction is the process which uses the Digis (the output from DAQ) as input and only constructs objects from an individual detector. For example, muon lo-

cal reconstruction produces RecHits and Segments, calorimeter produces CaloTowers. The global reconstruction use the objects built by local reconstruction across different detectors from the same subsystem but not combining the local reconstruction from different subsystems. The next level, the combined physics objects are built from the output of global reconstruction.

Analysis Tools

There are several different analysis tools developed by different physics groups in order to fit their analysis requirements. This results in multiple similar efforts which has the same purpose in different Physics Analysis Groups (PAG). Therefore, the Physics Analysis Toolkit (PAT) was created by Physics Analysis Groups to be a general purpose product. The PAT is a high-level analysis layer where the ID algorithms and reconstructed objects developed by Physics Objects Groups (POG) are included. Physics objects can be selected and cleaned in this process to eliminate the objects with poor quality.

4.2 Simulation

Simulation provides a way to study the analysis method [60]. There are two uses for it: to understand the behavior of the underlying physics and to understand the detector response for that physics process. Event generators serve the first purpose and detector simulation does the second job. According to the QCD factorization theorem [19], the event gener-

ator also can be decomposed into two major components, the Parton Distribution Function(P.D.F.) and the Hard scattering processes.

$$\sigma_{AB} = \int dx_a dx_b f_A(x_a, Q^2) f_B(x_b, Q^2) \hat{\sigma}_{ab\to X} \tag{4.1}$$

$$\hat{\sigma}_{ab\to X} = \hat{\sigma_0} + \alpha_S(Q^2)\hat{\sigma_1} + \alpha_S^2(Q^2)\hat{\sigma_2} + \cdots \tag{4.2}$$

$\sigma_{AB\to X}$ is the total cross-section of X production and $\hat{\sigma}_{ab\to X}$ is the cross-section for a particular hard scattering process $ab \to X$. It can be expanded perturbatively to the sum of the tree level cross-section, $\hat{\sigma_0}$ and other higher order terms, $\hat{\sigma_1}$, $\hat{\sigma_2}$ and etc. Therefore, f_A and f_B are P.D.F for proton A and B. x_a and x_b are the momentum fraction for parton a and b from proton A and B. α_S is the strong coupling constant. However, it doesn't provide the description for final observables. Before the detector response can be simulated, the result of the hard scattering process must be hadronized. In detail this is described in the following.

Parton Distribution Function (P.D.F.)

The first step is to calculate the probability of the participant from the incoming protons. The PDF's are the functions of the longitudinal momentum fraction x and the virtuality Q^2. The distributions are solved perturbatively from the DGLAP equations 2.35 where the Q^2

dependance is determined. However, the x dependence has to be obtained by the global fits from the data of deep inelastic scattering (DIS) experiments, Drell-Yan (DY) process and jet production. At leading order (LO), the measurements from DIS experiments and the DY processes from hadron-hadron collisions provides the main source of information on quark distributions($f_q(x, Q^2)$) and the jet production in hadron-hadron collisions supplies the information for gluon distribution($f_g(x, Q^2)$). For the consistency of the calculation, the expansion of the perturbation from DGLAP should be the same as the hard process term. Thus, a full NLO or NNLO approach is demanded in this stage since most hard processes are discribed to NLO or NNLO level. This makes the $f_{q/p}(x, Q^2)$, $f_{g/p}(x, Q^2)$ and even $\alpha_S(Q^2)$ mixed in the global fits. Currently, CTEQ[51] and MRST [46] are the two major groups which fit the data.

With current large DIS and DY datasets, the systematic error for quark distributions are only about 3% with wide range of x ($10^{-4} < x < 0.75$). However, the gluon distribution can not be determined in a wide range. The low x gluon distribution can be determined indirectly by measuring the scaling violation in the quark distribution but the moderate to high x gluon distribution only can be obtained directly from jet production. This fact results in large uncertainties on the gluon distribution. There are two principle techniques to estimate pdf uncertainty, the Lagrange Multiplier [64] and Hessian method [51, 63], which has been used by CTEQ and MRST group. The Largrange Multiplier method is

more useful for probing the pdf uncertainty of a given process and Hessian method has a more general framework to estimate pdf uncertainty for any production.

Hard Scattering Process and Matrix Element(M.E.)

From the second term of equation 4.1, the partonic cross section can be expanded perturbatively as equation 4.2. Mostly, the leading-order dominates the contribution of cross-section. The actual calculation has the following formalism:

1. Identify the LO partonic process.
2. Calculate the matrix element for $\hat{\sigma_0}$.
3. Convolute with all possible pdfs of the initial state.
4. Make a choice of Q^2.
5. Perform a numerical integration over x_a, x_b, and the phase-space variables which have to be restricted in order to avoid the divergence of the matrix element. Most of time, the p_{T} of the final state particle is chosen for the phase-space variable.

However, LO is not accurate enough to estimate the cross-section especially in LHC due to the unphysical choice of the scales. Some extra partonic processes may contribute to the cross-section only going beyond the scale. So the next-to-leading order (NLO) is needed for better theoratical uncertainty. The NLO calculation contains the following issues:

- Virtual and real radiation: NLO calculation requires the consideration of all diagrams with an additional strong coupling, a quark or a gluon. It can be categorized into two

types, virtual (loop) and real radiation. When performing the calculation of the cross-section, the integral over the p_{T} spectrum of the extra parton could cause divergence if it is too soft for the real radiation. For the virtual processes, divergences also occure but with the opposite sign. So the final cross-section converges by summing all diagrams. After this step, the differential cross-section can be expressed as the form

$$\frac{d\sigma}{dQ^2 dy dp_{\mathrm{T}}^2} \sim \frac{\log\left(s/p_{\mathrm{T}}^2\right)}{p_{\mathrm{T}}^2} \tag{4.3}$$

where p_{T} is from the object of the concerned process. It diverges when the p_{T} is close to zero when performing the integral for total cross-section calculation.

- Scale Dependence: An observable which is perturbatively expanded to the order α_S^n is independent of the choice of either renormalization or factorization scale, up to the next higher order in α_S. For example, an inclusive jet production can be formulated as

$$\frac{d\sigma}{dE_{\mathrm{T}}} = \alpha_S^2(\mu_R)\sigma_0 \otimes f_1(\mu_F) \otimes f_2(\mu_F) \tag{4.4}$$

$$\begin{aligned}\frac{d\sigma}{dE_{\mathrm{T}}} &= [\alpha_S^2(\mu_R)\sigma_0 + \alpha_S^3(\mu_R)(\sigma_1 + 2b_0 L\sigma_0) + \alpha_S^3(\mu_R)K] \\ &\quad \otimes f_1(\mu_F) \otimes f_2(\mu_F)\end{aligned} \tag{4.5}$$

$$L = \log(\mu_R/E_{\mathrm{T}}) \tag{4.6}$$

with equation 4.4 at LO and can be expanded as the equation 4.5 at NLO level. In this case, the cross-section is independent of the choice of the scale (μ_R or μ_F) for the terms of α_S^2 and α_S^3. Although the scale dependency is still in the terms of order α_S^4, the overall effect from the scales are reduced.

- k-factor: It is defined as the ratio of the NLO and LO cross-section. In principle, the k-factor may be very different for various kinematic regions of the same process. In practice, the variation is usually small. The k-factor depends on the pdfs which means the LO pdfs is used for evaluating LO cross-section and NLO pdfs is for NLO cross-section. In addition, the k-factor also depends on the phase space. If the analysis has a restriction on phase space, the applied k-factors should be different from the value for the full phase space.

In this thesis study, the matrix element generator Madgraph is used to generate the signal events ($t\bar{t}$) and part of background events (W+jets, Z+jets and single top).

Parton Showering(P.S.)

Parton showering is an all-order approach for the fragmentation of a parton. It is the process which allows the evolution for partons from high energy scale to a soft scale like Λ_{QCD} where a non-perturbative model can be used to provide the transition from the low energy partons to the hadrons which can be observed in experiments.

The evolution uses the DGLAP formalisim and the solution of the DGLAP equations is rewritten under Sudakov form factor. The Sudakove form factor is expressed as [19]

$$\Delta(t, t_0) = \exp\{-\int_{t_0}^{t} \frac{dt'}{t'} \int \frac{dz}{z} \frac{\alpha_S}{2\pi} P(z) \frac{f(x/z, t)}{f(t)}\} \tag{4.7}$$

$$\Delta(t, t_0) = \exp\{-\int_{t_0}^{t} \frac{dt'}{t'} \int \frac{dz}{z} \frac{\alpha_S}{2\pi} P(z)\} \tag{4.8}$$

where t is the hard scale(usually it's the Q^2 or p_{T}^2 of the parent parton), t_0 is the cutoff for soft scale, z is the momentum fraction and $P(z)$ is the splitting function. The equation 4.7 is for initial state radiation and the equation 4.8 is for final state radiation without the pdfs weighting in the last term of equation 4.7. The Sudakov form factor gives the non-branching probability for a parton evolving from high energy scale to the hadronization scale without emitting another parton with energy harder than the hadronization scale. With the help of the Sudakov form factor, the soft and collinear gluon emission has well-defined prediction. However, the effects above the non-singular threshold are not included such as larger energy or wide angle gluon emission.

In CMSSW, the simulation performs the parton showering via the Pythia generator which use Lund String scheme for the splitting function in fragmentation [9, 59].

ME-PS Matching

Parton showering provides a good description for soft and collinear gluon emission and the Matrix Element gives the precise fixed order calculation. Both of them are needed for multijet processes. However, there is no smooth transistion from one region of phase space to the other and this causes the possibility of double-counting in the overlap region. In order to avoid this situation, an interface btween PS and ME was specified from the Les Houches Workshop on Collider Physics in 2001 and two major efforts, CKKW [21] and MLM [42] were developed. The CKKW matching process is described below:

- Define a resolution parameter d_{ini} to divide the phase space into two regions for ME and PS.
- Generate multi-partons events with cuts on d_{ini}.
- Cluster the event by k_T algorithm until the LO is reached.
- Use $d_i \sim k_T^2$ in each vertex(i) as scale for α_S.
- Weight events with Sudakov factors $\Delta(d_j, d_{ini})/\Delta(d_i, d_{ini})$ between vertices i and j for each parton line.
- Use parton showering to shower the parton in the events. The emission is only allowed if $k_T < d_{ini}$.

With same initial steps to define the phase space for matrix element, MLM has different approach for the matching with parton showering. The process is

- Generate multi-partons events with cuts on jet $p_{\mathrm{T}min}$, η_{max} and ΔR_{min} which guarantees events are in the phase space free of singularity of the matrix element calculation.
- Cluster events and use $d_i \sim k_T^2$ for α_S scale as CKKW technique.
- Use parton showering to shower the partons in the events.
- Collect showered partons by cone-jets algorithm with same ΔR_{min} and $p_{\mathrm{T}cut} > p_{\mathrm{T}min}$.
- Keep the events with every jet matched to one generated parton.
- Extra jets are allowed with the p_{T} less than $p_{\mathrm{T}min}$ of the parton for highest jet multiplicity case.

These two matching algorithms are both embedded in Pythia and the NLO generator Madgraph contains a interface with MLM. The combination of these two generators provides an complete physics simulation in CMSSW.

Detector Simulation and GEANT

In order to evaluate the response of detectors and the efficiency of the reconstruction, it is crucial to use Monte Carlo method to simulate the detector response. It requires the detectors have to be described in terms of size, shape, material type and position. On the other hand, reconstruction demands the clear geometry description and the conversion between local and global coordinates in order to define hits and tracks. Thus a geometry subsystem

is created in CMSSW which is based on the XML language and the detector simulation is from the software GEANT4 [5]. These two packages plus the CMS reconstruction software provides a complete detector simulation in CMSSW. In additon to the full simulation, a fast version of simulation called FASTSIM [1] was also developed which parameterizes the the detector reponses and reconstruction.

4.3 Muon Reconstruction

Following the architecture of CMSSW, the muon reconstruction can be classified in three stages. The first is local muon reconstruction, the second stand-alone muon reconstruction and the third global muon reconstruction and final muon physics object creation.

4.3.1 Tracking of Charged Particles

Muons can penetrate most material with little energy loss so the outer-most muon system provides a clear identification while other particles are stopped by calorimeters. The 3.8 Tesla superconducting magnet of the CMS detector not only provides a strong magnetic field to bend the trajectory of charged particles inside the solenoid, but also gives significant strength of magnetic field in the return yoke. Therefore, tracking is the major technique to identify muon and the tracking algorithm [31] in CMS is discussed below.

- Track Parameters. The equation of motion given by the Lorentz force describes the trajectory of charged particles [13]. Under the assumption of absence of electric field,

the equation can be written as

$$\frac{d^2\vec{r}}{ds^2} = \frac{q}{p}\frac{d\vec{r}}{ds} \times B(\vec{r}) \tag{4.9}$$

where the second derivative term is the curvature of the trajectory, the first derivative is the direction of the charged particle at a given position, q is the charge and p is the momentum of charged particle. With the choice of a reference surface and the trivial identity $ds^2 = dx^2 + dy^2 + dz^2$, there are only five free parameters, $\{x, y, dx/ds, dy/ds, q/p\}$. Thus, the trajectory can be solved given a known magnetic field(set $\vec{B} = B\hat{z}$ as CMS coordinate)

$$x(s) = x_0 + R_H[\cos(\phi_0 + hs\cos(\lambda/R_H) - \cos\phi_0] \tag{4.10}$$

$$y(s) = y_0 + R_H[\cos(\phi_0 + hs\cos(\lambda/R_H) - \cos\phi_0] \tag{4.11}$$

$$z(s) = z_0 + s\sin\lambda \tag{4.12}$$

$$R_H = \frac{p\cos\lambda}{qB} \tag{4.13}$$

where R_H is the radius of the helix, ϕ_0 is the azimuthal angle of the reference point with respect to the helix axis and the λ is the slope angle($= \arcsin(dz/ds)$).

- Material Effects. A major effect for a moving charged particle in material is random Coulomb scattering. Therefore, its direction of motion is smeared out. The deflection angle is a gaussian distribution with the center of the original trajectory and this

results in the uncertainty of position and direction at the next detector layer. The energy loss from Coulomb scattering for those charged particles heavier than electrons or positrons is described by the Bethe-Bloch formula [32]

$$-\frac{dE}{dx} = \frac{D}{\beta^2}\{\ln(\frac{2mc^2\beta^2\gamma^2 T_{max}}{I^2}) - \beta^2 - \frac{\delta(\beta\gamma)}{2}\} \tag{4.14}$$

where D is the constant for a specific material, I is the mean ionization potential of the atom averaged over electrons, T_{max} is the maximum kinetic energy which can be imparted to a free electron and m is the mass of electron.

- Tracking Algorithm. The tracking algorithm contains four stages which are seeding, trajectory building, trajectory cleaning and trajectory smoothing. Seeding is the first stage to determine the initial state of track finding. The muon seeding algorithm begins with the segment collection of local muon reconstruction. It uses the measured direction and position of the segment from the inner-most muon chambers as the starting point and then applies a matching ΔR cone to collect possible candidate segments. The cone is defined as $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$. Thus, the estimated p_{T} of muon seed is given by using the formula

$$P_{\mathrm{T}} \times \Delta\phi = \int Bds \tag{4.15}$$

where $\Delta\phi$ is the difference of azimuthal angle from two candidate segments in the different stations. The field integral is determined by parameterization from MC simulation in different combinations of muon station and η. The final estimated p_{T} is the weighted average of all possible combinations. The final seeding result returns the estimated muon p_{T}, starting position and direction as well as their uncertainties. In addition, the candidate segment collection is also passed to the next stage for further use.

The second step is building the trajectory from the the candidate segments. The seeds are the inputs of trajectory building algorithm. The building modules use a combinatorial Kalman filter [37] to fit the whole set of the track parameters at each detector layer and apply them to perform the track finding in the next detector surface. The Kalman filter uses an iterative method to update the track parameters $\tilde{P}$ and its covariance matrix $\tilde{C}$ through a propagator. At a given state, $\tilde{P}_i$ and $\tilde{C}_i$ is propagated to the next detector surface by using the known equation of motion with consideration of the material effects and magnetic field. CMS software has three different propagators: the analytic with material propagator, the Runge-Kutta propagator and stepping-helix propagator. The analytic with material propagator has the assumption of uniform magnetic field aligned with the z-axis. The propagation from one detector surface to the next is ideal and the material effects are only introduced at the end point of the propagation. The Runge-Kutta propagator not only considers the material effects between two detector surfaces but also takes the inhomogeneity

of magetic field into account by solving the fourth order Runge-Kutta formula. The third one is the stepping helix propagator. It has the features which proceeds a propagation step with the finite helix length (5 cm) as long as no material or magnetic volume boundary is crossed or the destination is reached. For each propagation step, it updates the magnetic field and material effect values at the middle point of propagation. The material effects take energy loss and multiple scattering into account. The energy loss is using the fitted function from iron

$$dE/dx = -(11.4 + 0.96|\ln 2.8p| + 0.033p(1 - p^{1/3}))\ \text{MeV}/cm \tag{4.16}$$

and the fractional value of iron for non-iron materials. The effects from multiple scattering depends on the radiation length and the momentum of muon. It uses the provided formula [32]

$$\theta_0 = \frac{13.6\,\text{MeV}}{\beta cp}\sqrt{x/\chi_0}(1 + 0.38\ln(x/\chi_0)) \tag{4.17}$$

θ_0 is the width of the scattered angle distribution which is approximate to Gaussian distribution. β and p are the velocity and momentum of the charged particle. x/χ_0 is the thickness of the scattering medium in radiation length (χ_0). The stepping helix propagator is used in muon system and the first two propagators are used in tracker system only.

The third stage of the tracking algorithm is trajectory cleaning. Many of the track candidates from trajectory building share a large fraction of recHits. The cleaning algorithm resolve the ambiguities of the possible candidates and keeps the maximum amount of them. The last step is trajectory smoothing. The Kalman filter can perform the backward fitting which allows all the updated states to be refitted.

4.3.2 Muon Local Reconstruction

As mention in the previous section, local muon reconstruction contains three parts from three subsystems of the CMS muon spectrometer, CSC, DT and RPC.

Each CSC chamber has 6 layers. Each layer contains a strip plane and a wire plane. A CSC recHit is built at the intersection of a cluster of strips and wire groups. Then the recHits from different layers are used to build a CSC segment by fitting at least 3 aligned recHits from 4 different layers in one chamber. The DT local reconstruction builds 1-D recHits in each drift tube by using the drift velocity to calculate the distance from wires. After fitting 4 aligned DT recHits, a 2-D muon segment in each superlayer is created and a final 3-D segment is made by combining information from three 2D segments. The RPC has different purpose than the CSC and DT detectors. From the design of the RPC, the only local reconstruction is recHit built by fitting a cluster of fired strips. Its main goal is providing the timing for the trigger rather than spatial information for tracking.

According to the tracking parameters, the final muon segment from CSC and DT provides the position and the direction of muon trajectory in the reference surface of an indivisual chamber. After local reconstruction, a stand-alone muon track is built. Stand-alone muon tracks are reconstructed using the muon systems only. The reconstruction consists of three main stages:

- The initial canidates for stand-alone muon are a pair of muon segments. It starts from the inner-most layer of the muon system and collects the relavent segments in a region of interest with a suitable $\eta-\phi$ range defined by the error of segment direction.

- The seeding algorithm performs the p_{T} estimation accroding to equation 4.15 for all permutations from the segment collection and then averages all the values with weighting from the error of p_{T} estimation. The seed is built according to the estimated p_{T} and the segment collection.

- Starting from the seed, the stand-alone muon tracking uses Kalman filter to update the selected segment collection.

- The final stage is track cleaning. After reconstruction, some of tracks will share the same subset of segments. The cleaning step will choose the one with the best χ^2 and containing the most segments in order to ensure the quality of track.

4.3.3 Global Muon Reconstruction

The global muon reconstruction combines the information from tracker tracks and stand-alone muon tracks. The process starts from stand-alone muon tracks and matches tracker tracks to it. Due to large track multiplicity in the tracker, a pre-selected subset of tracker tracks is necessary. The first step is defining a region of interest in an $\eta - \phi$ region where the corresponding tracks are considered. Then the matching algorithm loops over the tracks to find the best match to be combined with the stand-alone muon track.

The region of interest is defined as follows:

- the origin: the primary vertex or beam spot is used.
- Δz: the allowed z spread with respect to the origin.
- ΔR: the allowed r spread where r is transverse distance with respect to origin.
- Direction: the vector with respect to the origin from stand-alone muon.
- minimum p_{T}: the minimum p_{T} is given by 60% of the stand-alone muon p_{T}.
- $\Delta\phi$ and $\Delta\eta$: the $\eta - \phi$ size of the tracking region which are estimated from the uncertainty of η and ϕ of the stand-alone muon.

The next step is matching process. The matching for two tracks is done by comparing the trajectory-state-on-surface of two tracks at the same reference plane. The reference

surface is chosen by requiring the minimizing the covariant error matrix of the propagated track parameters as well as the error from misalignment and reducing the number of matches per stand-alone muon. After propagating both track to the same surface, the matching process will compare the χ^2 value set from the five tracking parameters.

After the selection from matching process, global refit of silicon hits and muon hits are performed. In this stage, no additional pattern recognition is needed since it has been performed during the tracker tracking and stand-alone muon reconstruction. The global fit simply combines the hits from the track track and the stand-alone muon track. The Global refit algorithm will iterate all possible pairs of tracker track and stand-alone muon track to find the best combination from the χ^2 value of global refit. For a certain energetic muon situation such as showering effect, only subset of the recHits from stand-alone muon track is included for the global fit. It would control the momentum resolution for TeV muon cases.

As additonal muon reconstruction is the tracker muon algorithm. It is used for the cases in which the stand-alone muon reconstruction fails. This mainly concerns muons which have p_{T} below 6 GeV, which typically don't have enough energy to leave clear signal in the muon system. The tracker muon algorithm starts from tracker tracks and uses the stepping helix propagator to find the compatible energy deposit in calorimeters and segments in muon system. No combined tracker hit and muon hit fit is performed since

there is not a good stand-alone muon. So the mometum of tracker muon is given by the tracker track.

4.4 Jets Reconstruction

The 7 TeV pp collision of LHC creates a huge amount of QCD reactions. This makes jet reconstruction significant for most physics studies. For the study of this thesis, $t\bar{t}$ production, there are at least four jets expected from the muon + jets toplogy. Thus the jet reconstruction algorithm play an important role. Since a jet is the phenomenon of fragmentation of partons, the main constituents are hadronic products and some decay photons and electrons. Thus the jet algorithm relies on the calorimeters (HCAL and ECAL).

4.4.1 Tower Definition

The tower is the input of the jet reconstruction. Because the granularity of a HCAL cell($\Delta\phi \times \Delta\eta = 0.087 \times 0.087$) is much coarser than a ECAL cell(0.0174×0.0174), a calorimeter tower is defined by the addition of signals of HCAL and ECAL in an $\Delta\phi - \Delta\eta$ bin of the size of a HCAL cell. The energy deposit associated with a tower is the sum of all corresponding readout cells which pass the on-line zero suppression threshold and additional noise and pedestal threshold. In the jet reconstruction algorithm, the typical applied energy cut is either $E_{\mathrm{T}} > 0.5$ GeV or $E > 0.8$ GeV plus $E_{\mathrm{T}} > 0.5$ GeV. Each indivisual tower is

treated as a massless particle. The energy of the particle is the same as tower energy and its direction is given by the vector from the interaction point to the center of tower.

4.4.2 Anti-k_T cone jet clustering algorithm

There are two types of jet clustering algorithms: the sequential recombination (k_T [20] and Cambridge/Aachen [30]) and cone jet[14, 10] algorithms. The difference between these two types involves the issue concerning the regularity of the boundaries of the resulting jets. This issue is sensitive to the non-perturbative effects of the hadronization process and underlying event contamination.

The cone jet algorithm has been studied for more than two decades. It is an infrared and collinear (IRC) safe algorithm but it has the property that soft radiation can provoke irregularities in the boundaries of the final jets. In CMS, two cone jets algorithms have been developed: midpoint cone and siscone jet algorithm [56].

The sequential recombination type algorithm can simplify the theoretical calculation and eliminate some parts of momentum-resolution loss caused by underlying events and pileup effects. The basic idea is expressed by

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} \tag{4.18}$$

$$d_{iB} = k_{ti}^{2p} \tag{4.19}$$

where d_{ij} is the distance between entities i and j.

d_{iB} is the distance between entities i and beam B.

$\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$.

k_{ti}, y_i and ϕ_i are transverse momentum, rapidity and azimuthal angle of particle i.

R is the default radius parameter, usually R = 0.5 is set.

p is the parameter to govern the relative power of the energy versus geometrical scales(Δ_{ij}).

The principle procedure loop over all constituents in the detector (calorimeter) by identifying the entities with the smallest distances d_{ij} and d_{iB}. For d_{ij}, the entities i and j are combined and the process is continued. If d_{iB} is the smallest, then entity i would be recognized as a jet and be removed from the list. The whole process is repeated until no entities are left. In the case of p = 1, it is called inclusive k_T jet algorithm, for p = 0 is the Cambridge/Aachen algorithm and p = -1 is called anti-k_T algorithm [18] which is the current official jet algorithm for CMS jet reconstruction.

The functionality of the anti-k_T algorithm is described by considering an event contains a few well-separated hard constituents with transverse momentum k_{t1}, k_{t2} ... and some soft constituents. According to the algorithm,

$$d_{1j} = \min(k_{t1}^{-2}, k_{tj}^{-2}) \frac{\Delta_{1j}^2}{R^2}$$

is the distance between hard constituent 1 and soft constituent j. Clearly, d_{1j} is dominated by the transverse momentum of the hard constituent and Δ_{1j} and the d_{ij} among those soft constituents will appear to be larger. Thus, the softer constituents will tend to be clustered by the hard ones before they cluster between each other. With the scenarios of the presence of second hard constituent, the behaviors of the algorithm are listed below.

- $\Delta_{12} > 2R$: the algorithm simply accumulates all the soft constituents within a circle of radius R resulting in a perfect conical jet.

- $R < \Delta_{12} < 2R$: the algorithm will form two jets but only the harder one becomes the conical jet. If both constituents have similar transverse momentum, then neither one can forms a perfect conical jet. The boundary of two jets will be clipped by a boundary line b which defines $\Delta_{1b}/k_{t1} = \Delta_{2b}/k_{t2}$.

- $\Delta_{12} < R$: then the constituent 1 and 2 will form a single jet.

The key feature of this algorithm is that the soft constituents will not modify the shape of the jet. In other words, the jet boundary in this algorithm is resilient with respect to soft radiation, but flexible with respect to hard radiation.

4.4.3 Jet Energy Correction and Jet Energy Scale

A jet is reconstructed from the output of detectors which are supposed to record the energy from the corresponding particles. However, due to the following reasons, the reconstructed jets have a potential energy mismatch with the corresponding cascaded particles from original partons.

- The non-uniform and non-linear response of the calorimeters
- Electronics noise
- Pile-up effects which lead to extra energy deposite
- Unclustered energy deposits

Therefore, the jet energy correction is an necessary calibration after a jet is reconstructed. Its main goal is to relate, on average, the energy measured in the detector to the energy of the corresponding particle jets from parton level which is decribed as

$$P_i^{corrected} = C(P_T^{raw}, \eta) P_i^{raw} \tag{4.20}$$

P_i^{raw} is the four-momentum of an uncorrected jet. $C(P_T^{raw})$ is the multiplicative correction factor which is a function of uncorrected jet p_{T} and η. The typical jet energy calibration has three levels: Monte-Carlo truth, relative jet energy scale v.s. η and absolute jet energy scale v.s. p_{T}. In CMS, there are three approaches for jet energy calibration, Calorimeter-Base, Jet-Plus-Ttrack and Particle-Flow (PF) [27]. All three cases use anti-k_T clustering algorithm with $R = 0.5$. In this thesis, the general method is discussed and the Particle-Flow jets is adopted for the use of analysis.

Monte-Carlo truth Calibration

Since the source of a jet is a cascade of decayed particles, the MC simulation can simulate this process. One can use the MC truth information to correct the response of the algorithm. In CMS, the events are generated with Pythia generator and use the same jet reconstruction algorithm to build a particle jet called "gen jet" in CMSSW terminology. Thus, "reco jet" refers to the jet built from detector output.

In this level of calibration, gen jets are spatially matched to the reco jet in a $\eta - \phi$ space by requiring

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.25$$

Then the distribution of a quantity $p_{\mathrm{T}}^{RecoJet}/p_{\mathrm{T}}^{GenJet}$ in bins of p_{T}^{GenJet} is used to calculate the correction factors for different $p_{\mathrm{T}}^{RecoJet}$ and η. The result varies between different jet

algorithms. For PF jets with $p_{\mathrm{T}} > 30$ GeV, the correction factor is 1.1 in the central region of $|\eta| < 1.3$ and increases to 1.2 toward the higher η regions. The correction factor for all jet types, outside the tracker coverage ($|\eta| > 2.6$) is similar, ranging from 1.2 at p_{T} = 20 GeV to 1.05 at p_{T} = 200 GeV.

Relative Jet Energy Scale vs η

In order to understand the relative corrected jet energy response on η, a data-driven method, the dijet p_{T} balance technique, is employed [23]. An ideal back-to-back dijet event has the feature of p_{T} conservation. By taking a well-defined jet as a tag, the p_{T} of the other jet can be calibrated accordingly. The central region ($\eta < 1.3$) is chosen to be the reference because of its uniform response of the detectors. The detail event selection criteria are listed:

- the event must pass HLT with average uncorrected p_{T} of two jets

$$p_{\mathrm{T}}^{dijet} = (p_{\mathrm{T}}^{barrel} + p_{\mathrm{T}}^{probe})/2$$

- the separation of the two jets in the azimuthal plane must satisfy $\Delta\phi_{jj} > 2.7$ to ensure a back-to-back dijet event.

- If there is a third jet in the event, its p_{T} should only be a small fraction of the dijet average p_{T}, with $p_{\mathrm{T}}^{3rdJet}/p_{\mathrm{T}}^{dijet} < 0.2$ for reducing the fluctuation from radiation.

Then, there are two variables defined for the evaluation of the correction. The inbalance quantity is

$$B = \frac{(p_{\mathrm{T}}^{probe} - p_{\mathrm{T}}^{barrel})}{p_{\mathrm{T}}^{dijet}} \tag{4.21}$$

and the relative response of the probe jet with respect to the barrel jet is calculated by the expression

$$r = \frac{2+ < B >}{2- < B >} \tag{4.22}$$

Both quantities are recorded in bins of p_{T}^{dijet} and η. For the small bins of p_{T}^{dijet}, r is equivalent to $< p_{\mathrm{T}}^{probe} > / < p_{\mathrm{T}}^{barrel} >$. Thus, the relative correction can be derived in each $p_{\mathrm{T}}^{dijet} - \eta$ bin and be fitted with a second order polynomial of the logarithm of p_{T}^{probe}.

$$Rel(\eta, < p_{\mathrm{T}}^{probe} >) = \frac{1}{r(\eta, < p_{\mathrm{T}}^{probe} >)} \tag{4.23}$$

$$= a_0(\eta) + a_1(\eta) \ln p_{\mathrm{T}}^{probe} + a_2(\eta) \ln^2 p_{\mathrm{T}}^{probe} \tag{4.24}$$

Absolute Jet Energy Scale v.s. p_{T}

The absolute jet energy scale can be also calibrated by tagging a well-measured object. Compton scattering ($q + g \rightarrow q + \gamma$ and $q + q \rightarrow \gamma + g$) in pp collisions provides the processes which have transverse momentum balance of a photon-jet system and the photon

can be accurately measured in the ECAL. These events contain an isolated photon and a jet. The selection criteria are

- Both the photon and the jet are in the central region $|\eta| < 1.3$.
- The events are triggered by HLT single photon with 10 GeV and 15 GeV p_{T} threshold.
- p_{T} of reconstructed photon must be greater than 15 GeV.
- The photon candidate must be isolated in HCAL, ECAL and tracker and it must has a shower shape consistent to a photon.
- The separation in azimuthal plane satisfies $\Delta\phi_{jet,\gamma} > 2.7$.
- Additional jets with p_{T} greater than 0.2 p_{T}^{γ} and outside the ΔR = 0.25 cone around the photon
- The selected γ+jet sample covers p_{T}^{γ} range from 15 GeV to 200 GeV.

which ensure a clear back-to-back γ+jet event in a region with flat detector response and free of the iterference from radiation. There are two different approaches, MPF(missing E_{T} projection fraction) method and the p_{T} balance method for this calibration.

The idea of MPF method bases on the fact that there is no $\not\!\!E_T$ in these events; in the other words, the photon should be perfectly balanced with the recoiling jet. Therefore, it

can be formulized by equation 4.25 and equation 4.26 with the consideration of detector response.

$$\vec{p_{\mathrm{T}}^{\gamma}} + \vec{p_{\mathrm{T}}^{jet}} = 0 \tag{4.25}$$

$$R_{\gamma}\vec{p_{\mathrm{T}}}^{\gamma} + R_{recoil}\vec{p_{\mathrm{T}}}^{jet} = -\vec{\not{E}}_T \tag{4.26}$$

where R_{γ} and R_{recoil} are the detector responses for the photon and the jet respectively. Usually, R_{γ} is well understood in early data taking phase by using $Z \rightarrow e^{+}e^{-}$ or π^0/ $\eta \rightarrow \gamma\gamma$ events. Thus, the R_{recoil} can be derived as

$$R_{recoil} = R_{\gamma} + \frac{\vec{\not{E}}_T \cdot \vec{p_{\mathrm{T}}}^{\gamma}}{(\vec{p_{\mathrm{T}}}^{\gamma})^2} \equiv R_{MPF} \tag{4.27}$$

In general, other secondary jets or unclustered energy may exist. The $R_{coil} = R_{leadingJet}$ holds true if those secondary particles or leaking energy have similar response to the recoiled one or if these particles are perpendicular to the photon axis.

The p_{T} balance method uses the ratio between the jet and the photon p_{T} as an estimation of the jet response which is $R = p_{\mathrm{T}}^{jet}/p_{\mathrm{T}}^{\gamma}$. [24]. However, the MPF method is less sensitive to various systematic uncertainties than the p_{T} balance method and it is particularly well suited for PF jet because of the superior resolution of Particle-Flow MET. Thus, MPF is the main method to measure the energy resolution in CMS.

4.4.4 B Tagging

B jets play an important role in various physics. Since top quark decays to a W boson and a b quark, the identification of b jets is also a cruical topic for background rejection or event unfolding. There are several b-tagging algorithms that have been developed in CMS. They all use the character of long lifetime of b hadrons or high track multiplicity from the hadronization of b quarks. Tracks are the most important ingredient for b-tagging since most information of b-tagging are from them. So some general quality requirements [22] are imposed by

- total number of silicon hits(strip + pixel) ≥ 8
- number of pixel hit ≥ 2
- transverse impact parameter $d_{xy} < 0.2$ cm
- longitudinal impact parameter $d_z < 17$ cm
- transverse momentum > 1 GeV/c
- $\chi^2/ndof$ of the track fit < 5
- distance to the jet axis ($\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$) < 0.5

The concept of b-tagging is using the impact parameter(IP) or secondary vertex to discriminate b jets from the other light flavor jets [25]. The IP method is defined and explained in the following list.

- The IP is the distance of a track between its closest approach point and the primary vertex.
- It is signed by the scalar product of the IP segment and jet axis.
- In CMS, the IP is the calculated in 3D because the pixel provides good z position measurement.
- A better observable parameter is made by IP/σ_{IP} called IP significance.

In CMS, the application of IP method is Track Counting(TC) tagger. The tagger ranks the IP significance tracks in the jet and count the number of jets (N) exceeding the discriminating threshold. If set N = 2 for the discriminator, then the tagger is called Track Counting High Efficiency(TCHE) and N = 3 case is called Track Counting High Purity(TCHP).

For the secondary vertex method, it adopts the same algorithm from the primary vertex finding and applies it to the tracks associated with the jet. The vertices with at least 65$\%$ of tracks shared with primary vertex are not considered. The tagger named Simple Secondary Vertex (SSV) uses the significance of the 3D flight distance as a discriminating variable. Similar to the TC algorithm, a minimum number of tracks (N) attached to the vertex is required. The N = 2 case is called Simple Secondary Vertex High Efficiency(SSVHE) and N = 3 case is called Simple Secondary Vertex High Purity(SSVHP).

4.5 Missing Transverse Energy

Neutrinos and other hypothetical weakly interacting particles cannot be directly detected as they don't interact with detector material. However, the information of these particles can be extracted by exploiting the conservation of total transverse momentum of a event. It implies that

$$\sum_{i=1}^{N} \vec{p_{\mathrm{T}}^{i}} = 0$$

Thus, the vector balanced with a non-zero total p_{T} is called the missing transverse momentum and its magnitude is called the missing transverse energy ($\not\!\!E_T$). In semi-leptonic $t\bar{t}$ events, $\not\!\!E_T$ is an important component. It represents the transverse momentum of the neutrino from the W boson. With a W boson mass constraint, the longitudinal component of the momentum can be solved even though there is a two-fold ambiguity. So the resolution of $\not\!\!E_T$ plays an important role in unfolding the kinematics of $t\bar{t}$ events.

The original concept of $\not\!\!E_T$ [16, 33] is calculated by the vector sum of all the transverse energy deposit in calorimeter towers.

$$\vec{\not\!\!E}_T = (\sum_{i=1} -E_i \sin\theta_i \cos\phi_i)\hat{x} + (\sum_{i=1} -E_i \sin\theta_i \sin\phi_i)\hat{y} \tag{4.28}$$

$$= \not\!\!E_T^x \hat{x} + \not\!\!E_T^y \hat{y} \tag{4.29}$$

However, the missing transverse energy measurement at LHC will be complicated by the presence of pile-up collision. In CMS, the $\not\!\!E_T$ measurement is also degraded by the difference between photon and pion response in the ECAL and HCAL and by the bending of charged particles in the CMS magnetic field.

The $\not\!\!E_T$ correction in CMS has two levels [39], type I and type II correction. Type I correction consider the effects from muon and tau leptons and Type II correction take out-of-cone and unclustered energy into account. The three corrections are described in the following subsection.

4.5.1 Muon Correction

Since muons are minimum ionizing particles, they only leave a small fraction of energy in calorimeters. This results in under-estimated total transverse momentum. Nevertheless, their momentum can be measured accurately by tracker and muon system. Therefore, a compensation process is done by removing the muon contribution in the calorimeter and adding the muon momentum into the total p_{T} calculation.

$$\vec{\not\!\!E_T} = -\vec{E_{\mathrm{T}}}^{\,calo} - (\vec{p_{\mathrm{T}}}^{\,\mu} - \vec{E_{\mathrm{T}}}^{\,\mu-in-calo}) \tag{4.30}$$

$$= \vec{\not\!\!E_T}^{\,uncorrect} - \vec{p_{\mathrm{T}}}^{\,\mu} + \vec{E_{\mathrm{T}}}^{\,\mu-in-calo} \tag{4.31}$$

where the $E_{\mathrm{T}}^{\mu-in-calo}$ is the muon energy deposition in the calorimeter which is matched with a muon track. The muon should also pass the following criteria

- the muon must be Global muon
- p_{T} is > 10 GeV/c
- impact parameter of silicon hit, corrected for beam spot < 2 mm
- number of valid hits in tracker > 10
- $\chi^2/ndof <$ of global fit < 10

4.5.2 Tau Correction

Hadronic tau decays have similar signatures as a jet. However, tau decays have a lower particle multiplicity and fairly energetic products while a standard jet with the same energy has higher multiplicity with a larger fraction of low energy particles. These differences cause incorrect $\not\!\!E_T$ calculation if the tau is treated as a standard jet. Thus the correction for the events with taus can be made by defining a cone region where covers all the energy deposited from the tau and calculates the difference between the real tau energy and the response from the jet algorithm. In addition, underlying events and pile-up effects are also

taken into account. These can be described by

$$\Delta\vec{\not{E}}_T = \sum_{reg} \vec{E}_{\mathrm{T}}^{\,calo} - \sum_{reg} \vec{E}_{\mathrm{T}}^{\,UE} - \sum_{reg} \vec{E}_{\mathrm{T}}^{\,PU} - \vec{E}_{\mathrm{T}}^{\,\tau} \tag{4.32}$$

$$= \sum \vec{E}_{\mathrm{T}}^{\,cone5jet} - \vec{E}_{\mathrm{T}}^{\,PF\tau} \tag{4.33}$$

The correction can be simplified to 4.33 since the cone size of jet $\Delta R = 0.5$ is always wide enough to cover the tau energy deposit.

4.5.3 Type II Correction

The type II correction mainly considers the effects from unclustered and out-of-cone energy for every jet. It is done on top of the type I correction and can be described by

$$\vec{\not{E}}_T^{\,TypeII} = \vec{\not{E}}_T^{\,TypeI} + c(p_{\mathrm{T}j}^{\,cone5j}, \eta) \times \sum_j [\vec{p_{\mathrm{T}}}_j^{\,cone7J} - \vec{p_{\mathrm{T}}}_j^{\,cone5J}] \tag{4.34}$$

where $\vec{p_{\mathrm{T}}}_j^{\,cone7J}$ and $\vec{p_{\mathrm{T}}}_j^{\,cone5J}$ are p_{T} of jet with cone size 0.7 and 0.5 respectively. The correction is done by matching cone 5 jet and cone 7 jet within $\Delta R < 0.1$ with respect to jet axis. The correction function was parameterized by fitting the mean energy leaking, $< p_{\mathrm{T}}^{cone7J} - p_{\mathrm{T}}^{cone5J} >$, in bins of p_{T}^{cone5} and η.

4.6 Electrons and Photons

Electrons and photons shower in the ECAL. Approximately 94% of the incident energy of an electron or photon is deposited in a 3x3 crystal array and 97% in a 5x5 crystal array. Therefore, the reconstruction of electrons and photons rely on the reconstruction of ECAL signal. However, due to the significant amount of tracker material in front of the ECAL, this causes electron bremsstrahlung and photon conversion which results in different shower shapes in the ECAL. In addition, the CMS magnetic field also causes the radiated energy loss to spread in ϕ when electrons travel through the tracker.

The main reconstruction algorithm in ECAL is clustering the energy deposit from an electromagnetic shower fired by electron or photon. Two superclustering algorithms are created since there are several different shower shapes corresponding to different scenarios. The "Hybrid" algorithm is used in ECAL barrel and the "Island" algorithm is used in endcaps. The showers appear as a local maximum energy deposit area in a crystal array. The general search concept of the superclustering begins with a seed, a single crystal with energy deposit above a certain threshold. Then it collects the adjacent energy deposits from the same shower but avoids those from nearby particles and noise.

Although ECAL is not a tracking device, its fine granularity still provides the position of incident particle. The position measurement of ECAL can be determined from the

distribution of shower energy. However, the energy density decreases approximately exponentially with lateral distance from the shower core, a simple energy weighted mean position of the crystals in a cluster gives a biased positon toward the core of the shower. The cure is taking the logarithm of the crystal energy, the weighted mean then gives better position estimation.

$$x = \frac{\sum x_i W_i}{\sum W_i} \tag{4.35}$$

$$W_i = W_0 + \log \frac{E_i}{\sum_j E_j} \tag{4.36}$$

where the x_i is the position of crystal i and W_i is the weight of the crystal i. W_0 is the smallest fractional energy that a crytal can contribute to the position measurement.

4.6.1 Electron Reconstruction and Identification

The reconstruction of electrons has three major steps. First of all is initiated by the electromagnetic superclustering. A tracking process is followed and then a final matching for cluster and track is performed. The detail procedures for tracking and matching are described below.

- The tracking process is the same as that for muons. It begins with the seed generator. The trajectory builder builds the track inside out from the seed. Then the trajectory cleaner will clean the ambiguities among all possilbe tracks and maximum number

of tracks are kept. The last step is trajectory smoother which uses all the collected hits and re-evaluates the track parameters through a backward fit.

- The reconstruction is seeded by two different methods, tracker driven seeding and ECAL driven seeding. The first one is suitable for low p_{T} electrons which may not reach the ECAL or the electrons inside jets. The ECAL driven seeds start from ECAL superclusters with $E_{\mathrm{T}} > 4$ GeV. It is optimized for isolated electrons in the p_{T} range relevant for Z and W decays down to about 5 GeV/c.

- The seed is created when two hits are compatible with a given beam spot in the pixel detector. The ECAL driven seeding requires the matching of the seed with a supercluster. It lowers the fake rate of reconstruction, increases the reconstruction efficiency and purity of electron candidates.

- Electron tracks suffer from non-Gaussian fluctuations due to bremsstrahlung. Gaussian Sum Filter(GSF) and Bethe Heitler modeling for electron energy loss are used intead of Kalman filter for forward and backward filter [68]. The process stops when the last tracker layer is reached or no hit is found in two subsequent layers. The final best two candidates (with smallest χ^2) are kept.

- Track-Clustering matching: In addition to the seeds matched with ECAL superclusters, there are more criteria used for the matching in order to improve the identification and classification of electrons.

- Energy-momentum matching between the supercluster and the track, $E/P < 3$ is required. E is the energy from supercluster and P is the momentum from track.
- η, ϕ geometrical matching: $|\Delta\eta_{in}| = |\eta_{sc} - \eta_{trk}^{extrap}| < 0.1$ and $|\Delta\phi_{in}| = |\phi_{sc} - \phi_{trk}^{extrap}| < 0.1$. where η_{sc} and ϕ_{sc} is the η position of the supercluster. η_{trk}^{extrap} and ϕ_{trk}^{extrap} is the η and ϕ of track at the closest position to the supercluster position.
- H/E < 0.2: A ratio of the energy deposited in the HCAL tower which is just behind the ECAL seed cluster
- The combined electron energy is determined by the measurement of ECAL supercluster and eletron track. The rules are listed below.
  - $|E/p - 1| < 2\sigma_{E/p}$: the weighted mean of E and p by their sigma.
  - $E/p > 1 + 2\sigma_{E/p}$: E alone is used
  - $E/p < 1 - 2\sigma_{E/p}$ and E > 15 GeV: E alone is used
  - $E/p < 1 - 2\sigma_{E/p}$ or E < 15 GeV: p alone is used

- Since the radiated energy loss is caused by the material in the tracker and magnetic field, the energy spread is η and p_{T} dependent. Other effects are also taken into account such as that the fraction of energy in a fixed array varies as a function of the shower position with respect to the cluster boundary or the energy leakage due to the crack at the borders of the intermodules.

4.6.2 Photon Reconstruction and Identification

Since a photon is a neutral particle which doesn't trigger the ionization process directly, the photon reconstruction is only built from ECAL using the same superclustering algorithm. The measurement of energy deposit from photons is evaluated by the R9 variable. R9 is defined as the ratio of the energy in 3x3 array of crystals centered on the highest deposit to the total supercluster energy. Since R9 $>$ 0.943 contains the best 70% of photons, the photon candidates with R9 $>$ 0.94 in barrel region (0.95 in endcaps region) are selected. The photon energy is assigned by the energy in 5x5 crystal array centered to the highest crystal. Otherwise, the energy of supercluster is used. The R9 threshold in endcap is higher than the one in barrel because the module in endcap is larger. In order to exclude the possibility from charged particles, the supercluster is required not to match pixel hits consistent with a track from the IP region.

The dominated effect for the photon measurement is photon conversion in which a photon is converted into an electron-positron pair. The conversions occur due to the material in front of the ECAL which is about one radiation length on average in CMS. The photon conversion finder is based on the standard track finding algorithm. The track seed is initiated by a supercluster with outside-in direction. All clusters in the supercluster satisfied with $\Delta\eta < 0.015$ and $\Delta\phi < 0.25$ from the position of supercluster are used to build the track seeds. The p_{T} of seed is given by the cluster energy. The tracking starts from the outermost layer of the tracker and propagates inward to search for possible hits in the next

layer. It finishes with a list of tracks with biggest amount of opposite charge tracks. Then the hits from the track are used as the seed for a standard inside-out tracking process. The final track collection is then fed into the converted photon vertex finding algorithm.

4.7 Particle Flow Algorithm

A specific physics object is reconstructed by exploiting its characteristics and the detector responses. Therefore, the development is an individual efforts without any conjunction with other object reconstruction. This fact results in overlapping reconstructions for the same objects. For example, an electron with bremsstrahlung photons could be reconstructed as a jet from the clustering algorithm. Another problem that arises from an independent reconstruction is that the reconstruction relies on the signature from a specific detector, like muon from muon chambers, jets from calorimeter towers, etc.. Lacking a connection mechanism between different types of detectors also causes incorrect reconstruction and mis-identification.

In order to solve the above issues, the Particle-Flow (PF) algorithm is adopted for the study of this thesis. It decribes every event with five ingredients: muons, electrons, charged hadrons, neutral hadrons and photons. The identification of these particles is based on the following elements, Iterative Tracking, Calorimeter Clustering and Linking Algorithm.

Iterative Tracking

At the first stage, tracks are seeded and reconstructed by using the standard CMS tracking software but under a very strict criteria. It thus gives high quality tracks with moderate efficiency but negligible fake rate. The hits associated with those tracks are then removed from the hit collection. The next step is re-running the same tracking processes for the residual hits with looser criteria. The hits removal step can retain the low fake rate because the combinatorics of hits are reduced and the iteration will repeat several times in order to increase the efficiency. At the fourth and fifth iterations, the constraints on the origin vertex is relaxed. This allows the reconstruction of secondary particles from certain processes such as photon conversion or nuclear interactions with tracker material.

Calorimeter Clustering

Different from the clustering algorithm for jet reconstruction, the PF Calorimeter clustering aims to identify hadrons and photons. There are four parts in PF clustering algorithm. The first is to measure the energy and direction of neutral partilces. The second is categorizing the energy deposits from charged particles and from neutral particles. The third is to reconstruct and to identify electrons and all accompanying bremsstrahlung photons. The last is to help the energy measurement from the charged particles with associated poor track quality.

The algorithm is performed separately in different subsystems such as ECAL barrel, ECAL endcaps, HCAL barrel, HCAL endcaps, ES first layer and ES second layer. It consists of following processes:

- Cluster seeds are identified as local calorimeter cell energy maxima above a certain threshold.

- A topological cluster is grown from a seed by aggregating other cells with one side attached to the seed cell and exceeding a given threshold. It is possible to have several seeds in one topological cluster.

- The position of a cluster is obtained from the seed and the weight average of its four or eight adjacent cells.

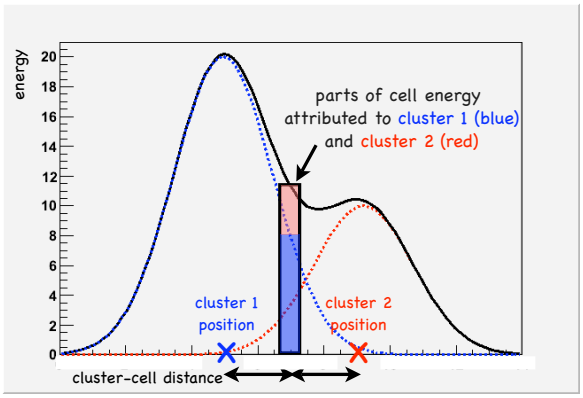
$$x = \sum_i (w_i x_i) / \sum_i (w_i) \tag{4.37}$$

  where $w_i = \ln(E_i/E_{threshold})$, E_i and x_i are the energy and position of the cell i in the cluster.

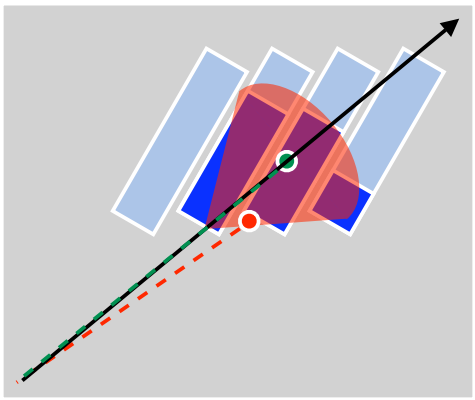
- For the case of multiple seeds in one topological cluster, the energy sharing between the sub-cluster (PF cluster) is determined by assuming the shower shape is gaussian and applying the cell-cluster distance(Fig. 4.1a). The cell-cluster distance is defined as ΔR in the $\eta - \phi$ plane by exploiting the calorimeter granularity.

- The depth Correction is only applied to the ECAL clusters because the ECAL crystal is not directly pointing to the IP. In order to avoid a bias in η position, the position

at a specific depth is used for computing the cluster position.(Fig. 4.1b) The depth is estimated by parameterization of the electromagnetic shower with the formula $depth = a(b + \ln E)$.



(a) Energy sharing between two sub-clusters



(b) The depth of a showering

Link Algorithm

In order to exploit the information from detectors for object reconstruction, PF algorithm includes a linking mechanism between tracks and clusters. The link algorithm has following elements:

- The charged track is extrapolated from its last hit to the following clusters:
  - Two layers of ES.
  - ECAL at a depth corresponding to the expected maximum of a typical longitudinal electron shower profile.
  - HCAL at a depth corresponding to one interaction length, typical of a hadron shower.

- The track and a given cluster is linked if the extrapolated position in the corresponding calorimeter is within the cluster boundaries.

- The cluster envelope is enlarged by up to the size of a cell in each direction to account for the presence of gaps between calorimeter cells, the cracks between calorimeter modules, the uncertainty of shower maximum and the effect of multiple scattering for low-momentum charged particles.

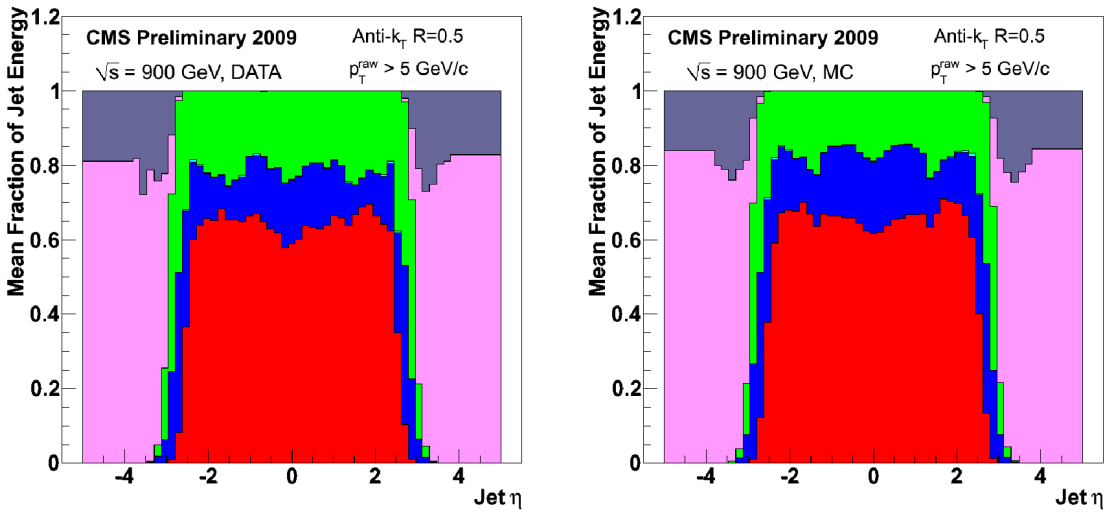
- The link distance is defined as the distance of ΔR in the $\eta - \phi$ plane between the extrapolated track position and the cluster position. It is used to quantify the quality of the linking.

The linking is done recursively. First of all, the algorithm builds a list of PF elements from the input objects which are tracks and clusters. The algorithm starts from moving one element into a PF block and then searches the associated elements from the rest of the list. If one is found, then it is removed from the list and added into that PF block. The process is continued until the PF block is complete. The whole process starts over again to build the next PF block until the list of PF elements is empty.

The Improvements

The main inmprovements from applying the PF algorithm are on jets and missing E_{T}. Unlike the jet reconstruction from calorimeter towers, the PF algorithm integrates all available

information from the detectors. It improves not only the resolution of the energy but also the direction. For the PF jet, the constituents in jets are clustered with the same anti-k_T algorithm with ΔR cone size 0.5. The jet constituents are categorized into seven types: 1. charged hadrons, 2. neutral hadrons, 3. electrons, 4. photons, 5.muons, 6. E.M. energy deposit in HF and 7. hadronic energy deposit in HF. Items 6 and 7 are additional to the main ingredients of PF because the PF clustering algorithm is not applied in HF. For each reconstructed jet, the jet-componenet energy E_X is carried by the particles of type X in the jet and $R_X = E_X/E_{jet}$ is the fraction of the energy of type X. According to the data from 2010 commissioning, it was found that the charged hadron are 65$\%$, neutral hadron are 20 $\%$ and photons are 15$\%$ in the barrel region and about 80$\%$ hadronic energy deposit, 20$\%$ E.M. energy deposit in HF. The M.C. shows good agreement with the data [26].



(a) (b)

Figure 4.1: Mean fraction of reconstructed jet energy, left plot is the result from data and right plot is the result from M.C.

Another variable $\sum E_T$ is also used to evaluate the PF algorithm. Unlike $\not\!\!E_T$, $\sum E_T$ is the scalar-sum of all particle E_{T}. Thus, all detector effects will be enhanced by the summation but they might be cancelled out by the vector-sum. From the M.C. study it shows that particle-base $\sum E_T$ is closer to the true generated $\sum E_T$ than calorimeter-base $\sum E_T$. The reasons are that the PF algorithm accounts for those effects from charged hadrons and very low enegy particles which are down to about 100 MeV/c for charged hadrons and about 200 MeV for photons. It brings the correct energy scale for the event and leads to a better $\not\!\!E_T$ reconstruction.

Chapter 5

Event Selection

5.1 Data and MC Samples

This study is based on the data collected in 2010 with integral luminosity 36.1 pb^{-1}. The corresponding MC were mostly generated using the NLO generator, Madgraph except the QCD and WW samples were produced by Pythia. The samples used in this analysis are listed in table 5.1.

In order to ensure the quality of data, two cleaning steps are applied; the first one called beam scraping is to cut on the fraction of high purity tracks in the event (at least 25% of good tracks are required) and the second is "HBHE Noise Filter" for cleaning the anomalous HCAL noises where the noises results from the electronic pedestal. Another one is cleaning the anomalous ECAL signals which are characterized by single ECAL channels

with sizeable pulse and without any other surrounding activities. Thus, a requirement is applied

$$\frac{e_4}{e_1} > 0.05$$

for a real electron. e_1 is the energy of the seed crystal and e_4 is the energy sum of the four crystals adjacent to the seed crystal.

Channel	σ (pb)	Acceptance + HLT Efficiency	σ (pb) LO	k-factor
$t\bar{t}$, Inclusive	157.5	0.2468	94	1.676
W+Jets, leptonic decayed	31314	0.2025	24380	1.284
Z+Jets, leptonic decayed	3048	0.2835	2289	1.332
single top t-channel	64.6	0.327	63.3	1.021
single top tW-channel	10.6	0.2247	10.56	1.004
QCD Muon-enriched	84679.3	0.8247	N/A	N/A
WW, Inclusive	43	0.1581	N/A	N/A

Table 5.1: MC samples used in the analysis

5.2 Baseline Selection

Events containing $t\bar{t}$ decays are classified in terms of the decay products of the two W-bosons and the two b-jets that they yield. The semi-muon $t\bar{t}$ channel is defined to be the channel in which the two W-decays are $\mathrm{W} \to \mu\nu$ and $\mathrm{W} \to \mathrm{q\bar{q}'}$. The signature of this type

of event is one isolated muon, missing transverse energy ($\not\!\!E_T$) and at least four jets from the decay of four quarks. The baseline object selection for this analysis follows the suggestion of the Top Lepton plus Jets group [11]. This is summarized in Table 5.2 and the details are described below.

Trigger Selection

Since the topology of muon plus jets is the target channel, the presence of a muon and jets in the event are the signatures. However, jet activities in pp collision are overwhelmed by QCD interactions. Thus the muon trigger is most efficient filter to suppress the enormous QCD backgrounds and retain the signal events. The high level trigger (HLT) object is used rather than L1 because it has better precision. In this thesis, HLT_Mu9 and HLT_Mu15 are used. The HLT muon object is defined in section 3.3.7. HLT_Mu9 is the HLT filter which requires at least one HLT muon object with p_{T} greater than or equal to 9 GeV/c in the event. HLT_Mu15 is defined as the same way with a 15 GeV/c threshold. The early data, run 2010A and partial run 2010B, are triggered by HLT_Mu9 and the rest of run 2010B data are triggered by HLT_Mu15 due to the prescale of HLT_Mu9 resulting from the higher instantaneous luminosity. Although the dataset contains two different trigger thresholds, it didn't cause noticeable efficiency difference because the final muon selection is based on reconstructed muon objects with p_{T} greater than 20 GeV/c.

The muon trigger efficiency was studied by the top group [61] using the tag and probe method. The study selected $Z \to \mu^+\mu^-$ events with both muon passing the muon selection

criteria. One of the muons matched to an HLT object is the "tag", while the other muon has no trigger requirement in order to probe the HLT efficiency. The events are selected in the window of reconstructed Z mass. Thus, the efficiency is defined as

$$\epsilon = \frac{N^{probe}(\text{fitted \& triggered})}{N^{probe}(\text{fitted})} \tag{5.1}$$

$N^{probe}(\text{fitted \& triggered})$ is the number of probed muon fitted in Z mass window and also fire HLT. $N^{probe}(fitted)$ is the total number of probed muon in the Z mass window. The results shows that the HLT_Mu9 and HLT_15 efficiencies are almost the same in the same data taking phase(2010B) especially for muon with p_{T} above 20 GeV/c (see Fig. 5.1). HLT_Mu9 is 5% lower in 2010A than in 2010B.

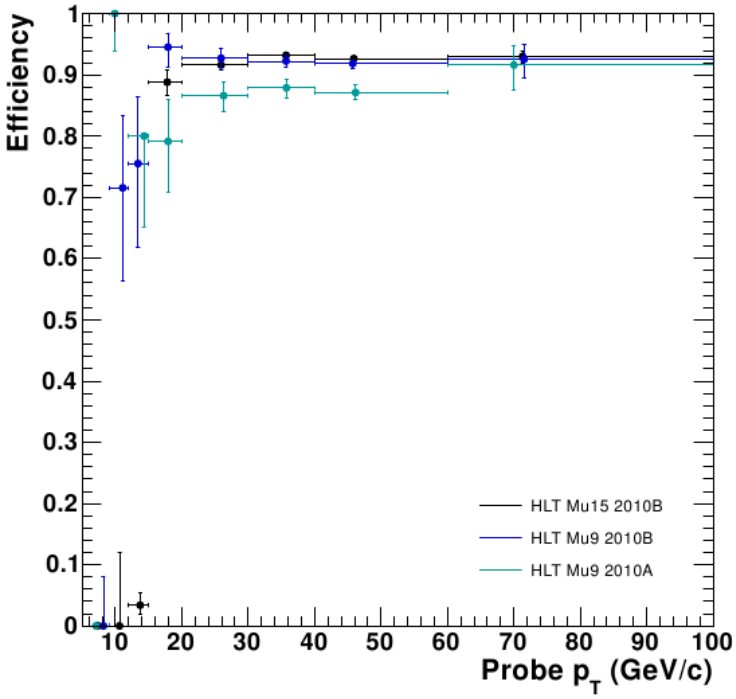


Figure 5.1: The HLT efficiency from different data-taking phase

Since the HLT efficiency for $t\bar{t}$ events is evaluated by MC, the difference between data and MC is also studied. The result(in Fig. 5.2) shows the uniformity at $p_{\mathrm{T}} > 20$ GeV/c and the consistency in $\eta - \phi$ space. The scale factor for MC is 0.968 ± 0.002 which is used in this analysis.

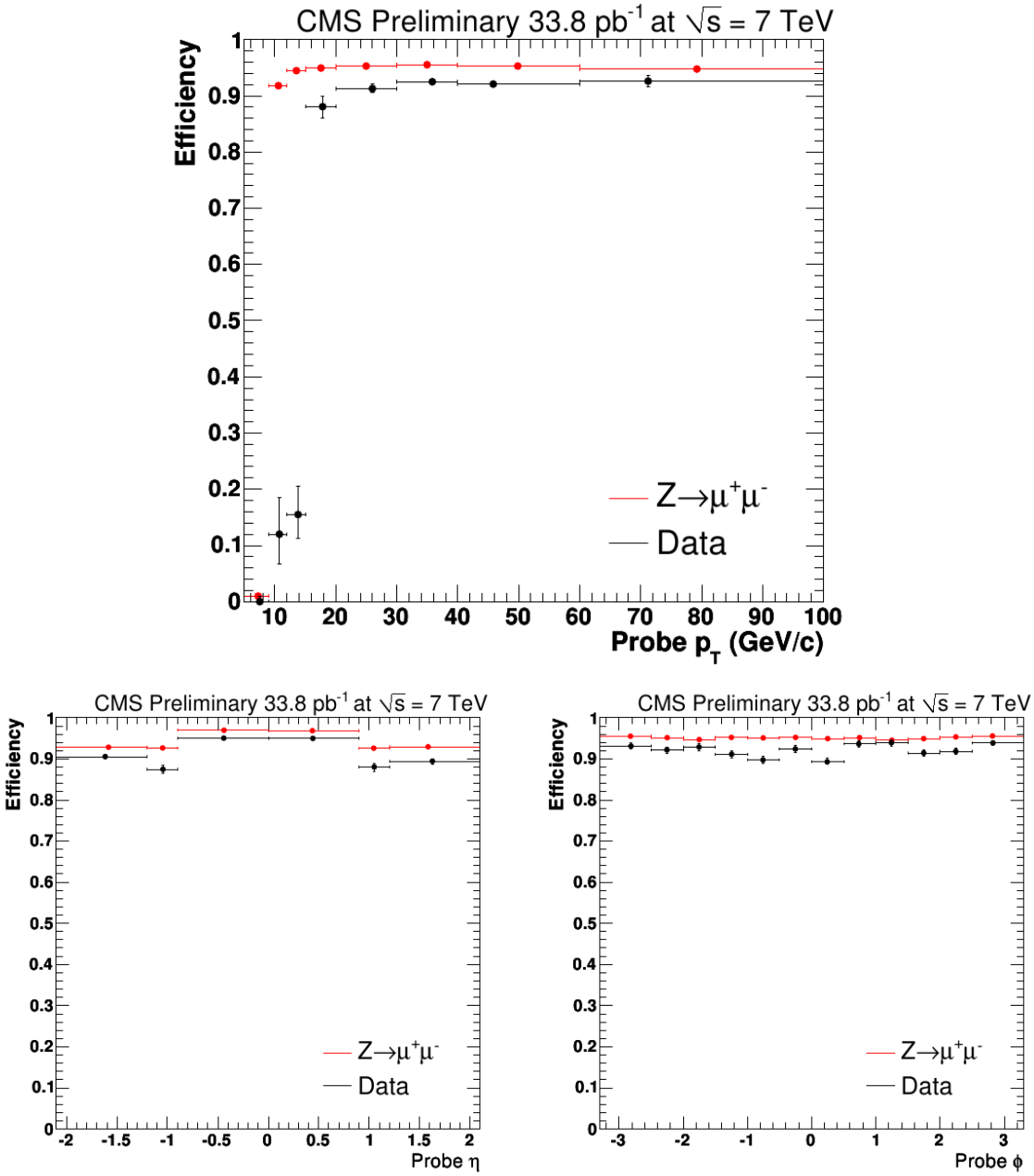


Figure 5.2: The HLT efficiency of data and MC

Primary Vertex

Due to the high instantaneous luminosity pp collision, mulitple interactions can happen in one single bunch crossing and objects from different interactions can be recorded in a single events. These kind of events are called pile-up events. Since the objects in pile-up events are mixed together, it causes event mis-reconstruction such as incorrect $\not\!\!E_T$ or incorrect topology. To reduce the effect from pile-up, a requirement of good primary vertex is set to satisfy ndof > 4, nodf is the number of degree of freedom corresponding to the weighted sum of number of tracks used for the construction of the primary vertex. In addition, the primary vertex must be located within the central region of $|z| < 24$ cm and $\rho < 2$ cm with respect to the interaction point.

Muon Selection

The muons from $t\bar{t}$ events are decayed from the W bosons which are the decayed products of top quark. So the muons from $t\bar{t}$ events are well separated from other objects in the event topology and carry significant amount momentum since W boson mass is 80 GeV/c^2. Except those cuts used to ensure the quality of muon objects, muon isolation and p_{T} threshold are two major selection criteria.

The concept of isolation is examining the energy or momentum deposit in the isolation cone. As the reference of Fig. 5.3, the vector of muon momentum at vertex is the cone axis and the cone radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ is defined. A smaller veto cone along with the

muon track is also included because the contribution from the muon has to be subtracted. In this thesis study, the isolation cone size is set to 0.3 and the veto cone size is 0.01. The variable used to determine isolation is defined by "$RelIso$" as

$$RelIso = \frac{Iso^{Track} + Iso^{ECAL} + Iso^{HCAL}}{{p_{\mathrm{T}}}^{\mu}} \tag{5.2}$$

Iso^{Track} is the momentum sum of all tracks in the isolation cone from tracker. Iso^{ECAL} and Iso^{HCAL} are the total energy depositions in the defined isolation cone from ECAL and HCAL. A well isolated muon should have minimum energy in the isolation cone and it leads to a small $RelIso$ value. The $RelIso$ value for muon selection in this thesis is set to be 0.1.

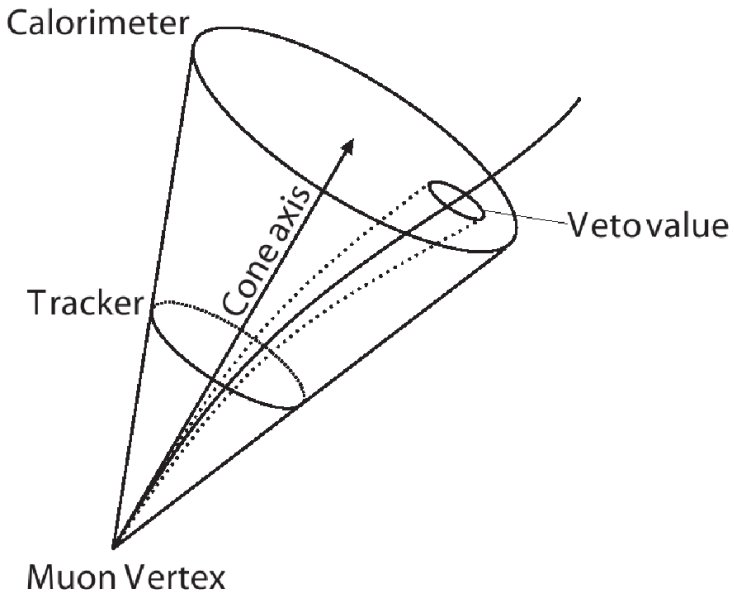
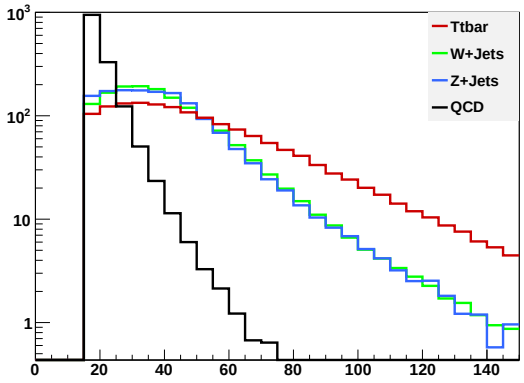
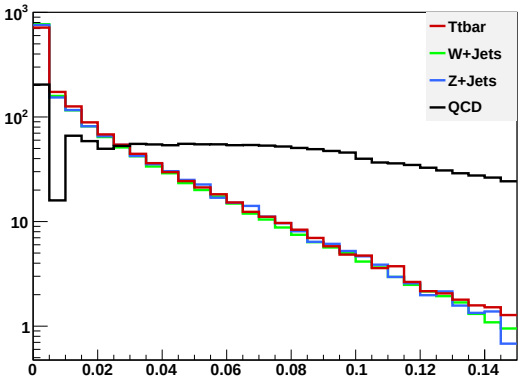


Figure 5.3: The scheme of the isolation cone

In addition to isolation, the muon p_{T} is also required to pass a 20 GeV/c threshold. From Fig. 5.4a and Fig. 5.4b, they show that the QCD has a different muon p_{T} spectrum and isolation distribution because the muons from QCD processes usually come from b or c quark decay. These two cuts effectively suppress the enormous cross-section of QCD events.



a. Muon P_{T} Distribution



b. Muon Isolation Distribution

Jet Selection

For the muon plus jets channel of $t\bar{t}$ events, there are at least four jets from the topology of events. Two b jets are directly from the top decays and the other two jets are from the hadronic W decay. Unlike background processes such as W/Z+jets and QCD events where the jets are from the QCD bremsstrahlung, the jets from $t\bar{t}$ events represent the products of decayed partons from high mass objects like W bosons or top quarks and thus a significant p_{T} is expected. A 25 GeV/c p_{T} threshold is imposed on the jet selection and the minimum

jet multiplicity is set to be four. Fig. 5.4 shows the different population of jet multiplicity and p_{T} spectrum.

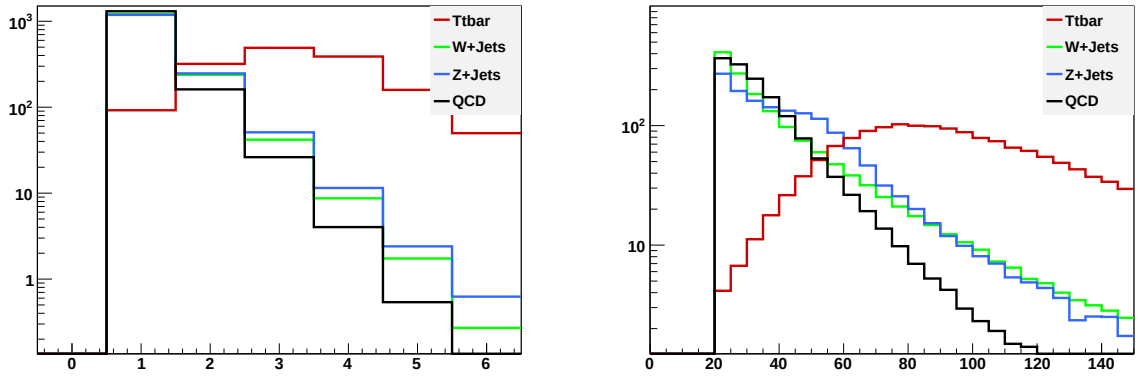


Figure 5.4: Jet Multiplicity and the Leading Jet P_{T} for $t\bar{t}$, W/Z+Jets and QCD events

Additional Jet-ID cuts for particle flow jets provided by Jet/MET group are also applied [57]. They are

- number of constituents > 1
- Charged EM energy fraction (CEF) < 0.99
- Neutral hadronic energy fraction (NHF) < 0.99
- Neutral EM energy fraction (NEF) < 0.99
- if $|\eta|$ of the jet < 2.4, charged hadronic energy fraction (CHF) > 0
- if $|\eta|$ of the jet < 2.4, charged multiplicity (NCH) > 0

which are used to exclude the fake jets.

Second Lepton Veto

Since this study focuses on the topology of muon plust jets channel, the di-lepton channel of $t\bar{t}$ events has to be rejected because it has similar characters of the decayed products such as harder jet p_{T} spectrum and isolated leptons. With 2 b jets in the event topology, there is a possibility to see a second muon or another lepton in the event. However, the leptons from b decay usually are not isolated. In order to retain most of the signal events and also reject the events with more than one isolated muon such as di-lepton channel of $t\bar{t}$ events or Z+Jets events, a second lepton veto criteria are required. On the balance of high rejection efficiency for possible mimicked backgrounds and allowing the b decayed muon, a loose isolation requirements (RelIso < 0.2) is set for a second muon or electron. If there is a second lepton passing the loose selection in the event, the event will be skipped.

Cut	Cut discriptions
Trigger	HLT_Mu9 and HLT_Mu15
Primary Vertex	At least 1 primary vertex with Number of d.o.f. > 4 $\mid z$ position $\mid < 24$ cm and $\mid \rho \mid < 2.0$ cm
Muon	Exactly one muon Must be GlobalMuon(GlobalMuonPromptTight) • normalized $\chi^2 < 10.0$ • number of valid muon hits > 0 Must be TrackerMuon $-$ number of valid hits > 10 $p_T > 20$ GeV , $\mid \eta \mid < 2.1$, RelIso < 0.1 $\Delta(\mu, jet) > 0.3$ Absolute 2D impact parameter with respect to beam spot < 0.02 cm
2nd Muon Veto	Must be GlobalMuon $p_T > 15$ GeV/c, $\mid \eta \mid < 2.5$, RelIso < 0.2
Electron Veto	$E_T > 15$ GeV , $\mid \eta \mid < 2.5$, RelIso < 0.2
Jet Selection	$P_T > 25$ GeV/c, $\mid \eta \mid < 2.4$ charged E.M. Energy Fraction < 0.99 neutral HadronEnergyFraction < 0.99 neutral E.M. EnergyFraction < 0.99 if $\mid \eta \mid < 2.4$, charged Hadron Energy Fraction > 0 if $\mid \eta \mid < 2.4$, charged Multiplicity > 0

Table 5.2: The baseline selection cuts

5.3 Leptonic Transverse Mass

The leptonic transverse mass is defined as the transverse mass of the muon and the $\not\!\!E_T$ using the formula $M_{\mathrm{T}} = \sqrt{2\not\!\!E_T P_{\mathrm{T}}(1-\cos\phi)}$, where ϕ is the azimuthal angle between $\not\!\!E_T$ and the muon direction. Events containing a leptonic W-decay($t\bar{t}$, W+jets) will have a Jacobian peak in the M_T distribution whose shape depends on the P_{T} of W boson. At higher values of the W transverse momentum, the W has a higher probability to have smaller M_T due to the boost in the transverse direction and thus the peak is smeared out [62]. For an event without real leptonic decayed W, the main source of $\not\!\!E_T$ is from off-shell W decay and imperfect reconstruction of event. The $\not\!\!E_T$ tends to have a small angle with respect to the muon which results in small M_T (see Fig. 5.5). So the M_T distribution of this type of events is a falling curve from 0. Fig. 5.6 shows the comparison of $t\bar{t}$, V+jets(W and Z plus jets production) and QCD events as a function of jet multiplicity. From this it is clear that a cut on leptonic transverse mass can be used to remove a significant part of the QCD background. However, QCD has relative small amount of events in the 4-jet bin and above. Thus, this cut is only applied to 2-jet events for extracting control samples.

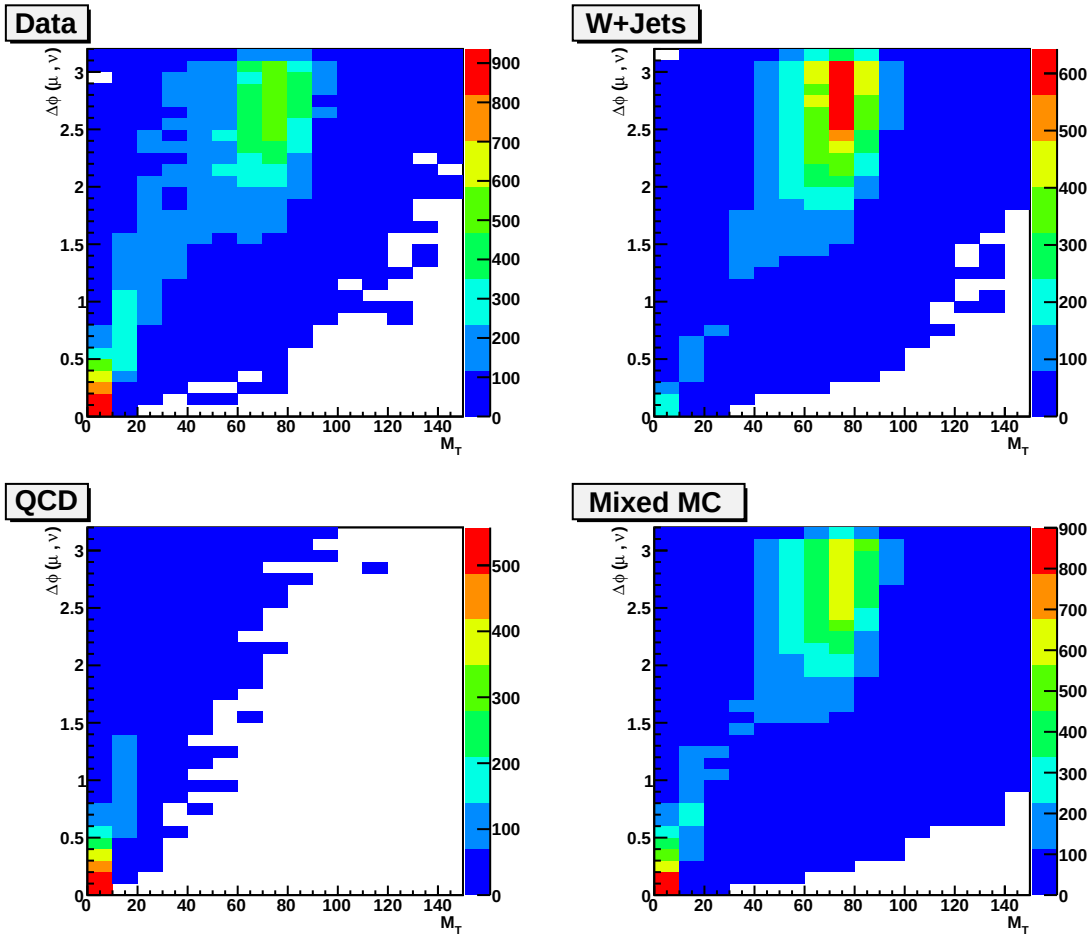
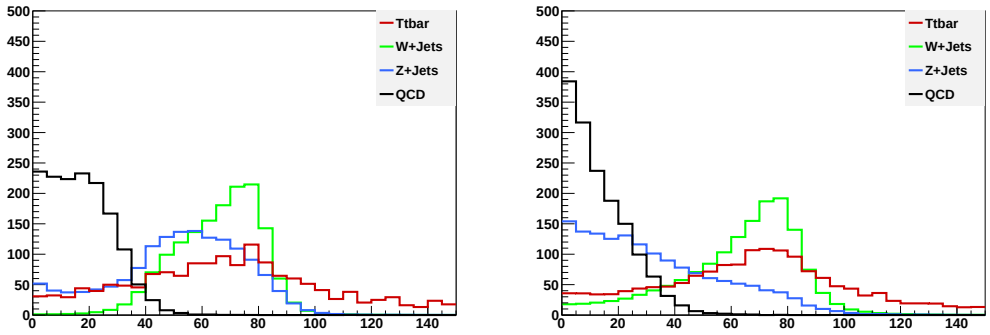


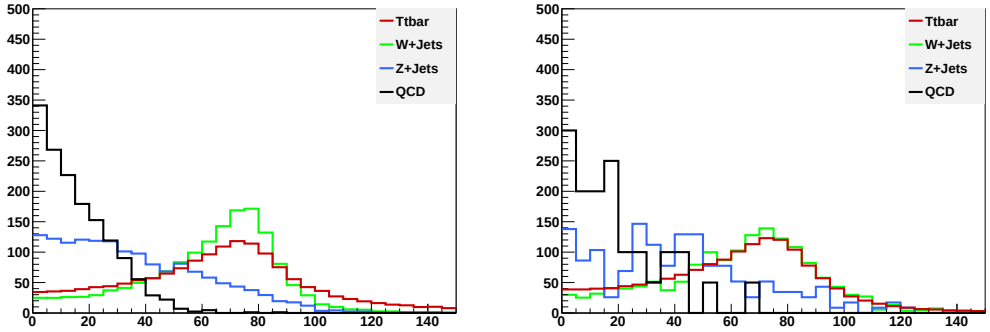
Figure 5.5: M_T v.s. $\Delta\phi(\mu, \not\!\!E_T)$ from the events with ≥ 1 jets

5.4 Two-body ($M2_h$) and Three-body ($M3_h$) Mass Permutations

Topogically, there are at least four jets in the final state of $t\bar{t}$ semi-leptonic events. So the jet permutations from the hadronic portion of the event contain the mass information of top and W if it is a $t\bar{t}$ event. Thus, two variables, $M2_h$ and $M3_h$ are adopted for examining the permutations. In this and subsequent sections we use the following $M2_h$ and $M3_h$ definitions:



0 jets events 1 jets events



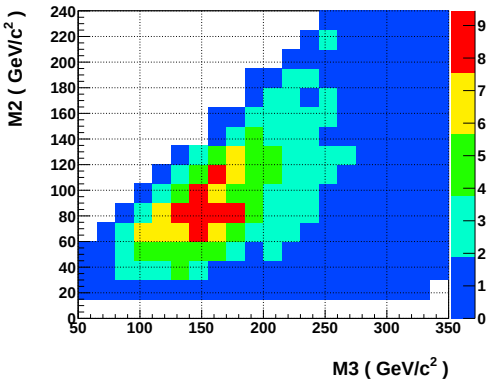
2 jets events $\geq$4 jets events

Figure 5.6: Transverse mass of various processes in different jet multiplicity. The muon p_T threshold is $16GeV/c$ and each sample is normalized to 1000 events

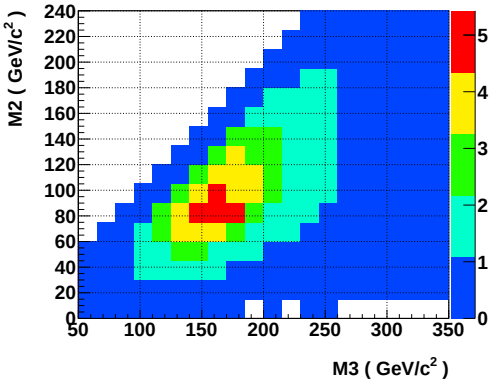
- $M2_h$: Di-jet mass associable to the hadronic W-decay.
- $M3_h$: 3-body mass corresponding to the hadronic top decay. ie. $M2_h$ + the b-jet from the same parent top.
- $M3_l$: 3-body mass corresponding to the leptonic top decay. ie. lepton + $\not\!\!E_T$ + the b-jet from the same parent top.

For a correctly reconstructed $t\bar{t}$ event, $M2_h$ and $M3_h$ should have values of 80.4 GeV and 172.4 GeV, respectively and there should be at least one jet-permutation which gives

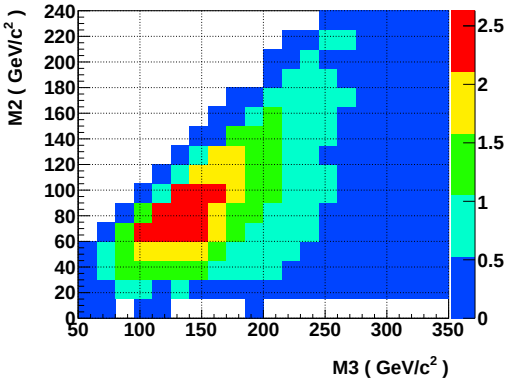
the correct solution. Fig. 5.7 shows the comparison of $M2_h$ vs $M3_h$ for the data and various MC samples. There is a clear difference in population between the $t\bar{t}$ signal and the V+jets and QCD backgrounds. Firstly, a rectangular window with respect to the expected values in the $M2_h - M3_h$ plane is created. The size of the window is determined according to the variation of jet energy scale.(Fig. 5.9) For a $t\bar{t}$ event, there should be at least one permutation matched in the window for 4-jet events, two matched permutations for 5-jet events and so on. If no matching permutation is found, then the event is discarded. Three jet events are also rejected because the missing hadronic component results in a biased $M2_h - M3_h$ measurement. Once an event is accepted, all of the permutations from the event are also accepted. For events with multiple permutations, each permutation contributes to the $M2_h - M3_h$ measurement with a weight equal to 1 over the number of permutations for the event, so that the total event weight is one. No kinematic preference is used to select the permutations from the event and no weighting on event counting. These criteria keep the permutations un-biased when passing them to subsequent topological selections. Consequently, each permutation in a higher jet multiplicity event has less weighting than the one in the events with fewer jets.



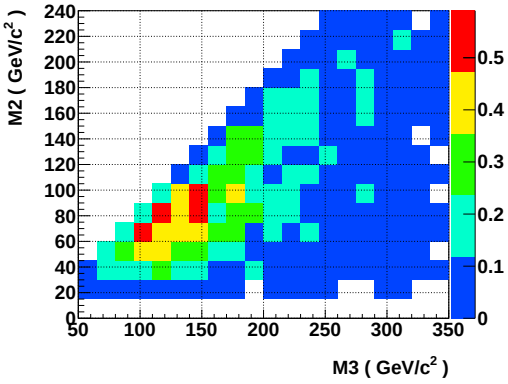
36.15 pb^{-1} data



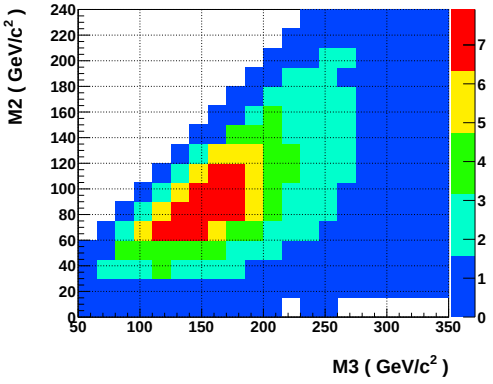
MC $t\bar{t}$ events



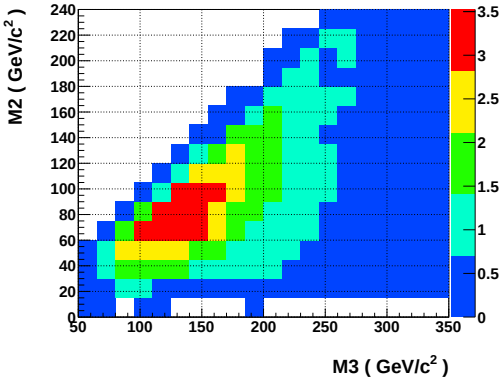
MC W+jets events



MC QCD events



All MC events



All MC background events

Figure 5.7: $M2_h$ vs. $M3_h$ distribution before applying any topological constraints for data and MC. The MC samples are normalized to a luminosity of 36.15 pb^{-1}

5.5 Three-body($M3_l$) Mass Permutations

From the event topology, the only measurable corresponding to the neutrino is the $\not\!\!E_T$. By applying the W mass as a constraint, the z-component of the neutrino momentum can be obtained with a two-fold ambiguity. This allows the determination of $M3_l$ which also has a two-fold ambiguity. For the cases without physical(real) p_z solution, p_z = 0 is assigned because this kind of event usually has M_T larger than W boson mass. According to M_T distribution from MC studies(Fig. 5.6), the events with M_T greater 80.4 GeV/c mostly are still from leptonic decayed W bosons. With this assumption of neutrino p_z solution, it can increase the efficiency of $t\bar{t}$ event selection. Since there are two tops in each event and their mass should be equal, the difference between $M3_h$ and $M3_l$ can be the minimized to improve the yielding selection of the correct jet permutation and neutrino solution. Fig. 5.8 shows the population of $M3_h - M3_l$ for each event. Similar to $M2_h - M3_h$, a $M3_h - M3_l$ window can also be used in the yielding event selection by constraining the permutations. However, the range of permutations from the hadronic part increases the probability to match the $M3_h - M3_l$ criteria. Therefore, a combination of $\Delta M3$ cut, ($|M3_h - M3_l|$) and $M2_h - M3_h$ is created. The permutations passed the $M2_h - M3_h$ cut must also satisfy the $\Delta M3$ requirement.

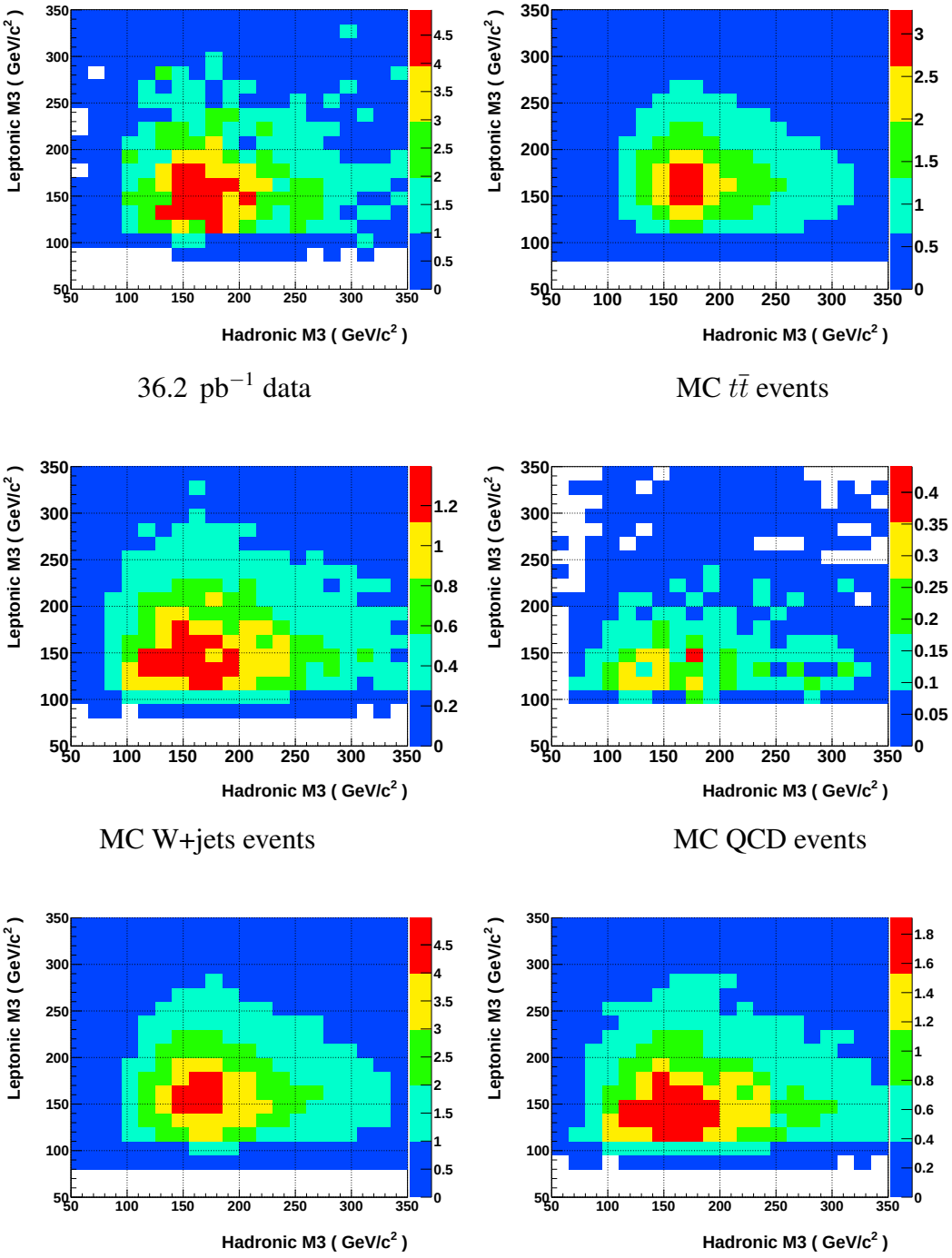


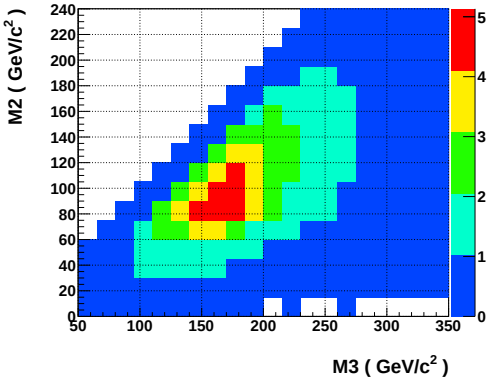
Figure 5.8: Event distributions in the $M3_h - M3_l$ plane with no additonal selection applied. The data and MC distributions are both for 36.15 pb^{-1}

5.6 Topological Constraints

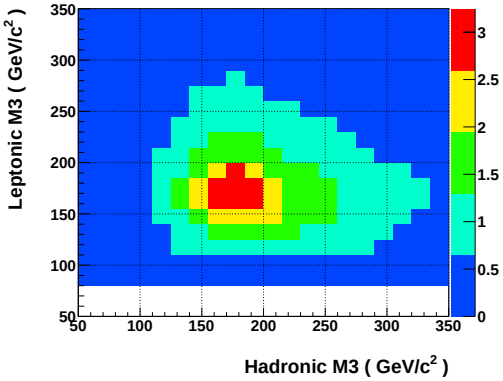
The decision of topological constraints are under following considerations. First, the uncertainty of JES. Jet energy scale changes the spectrum of jet P_{T} and jet multiplicity. So the invarint mass of two jets and three jets change. We checked the population of M2M3 and M3M3 with scale up and down JES(Fig. 5.9) and set the initial window for $M2_h$ and $M3_h$ as 10 GeV/c and 20 GeV/c respectively according to the visual effect from the $M2_h - M3_h$ plot.

The next step, the individual 1-D distributions are checked under the constraint of the other variables. The $M2_h$ is divided into 3 regions with $M3_h < 140$ GeV, $M3_h > 200$ GeV and between. Similar to $M2_h$, $M3_h$ has three regions with $M2_h < 60$ GeV, $M2_h > 100$ GeV and between. From the Fig. 5.10 in the potential signal region($60 \leq M2_h \leq 100$ GeV/c and $140 \leq M3_h \leq 200$ GeV/c), it shows the enhancement of the signal and deviation of the background distributions.

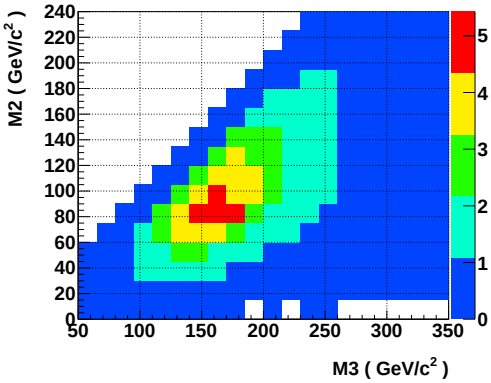
A more quantitative analysis is a grid scanning on the $M2_h - M3_h$ plane. $M2_h$ and $M3_h$ window size are varied from 10 GeV/c to 40 GeV/c. Several indicated values were evaluated and shown in Fig. 5.11. It's clear that the efficiency for both $t\bar{t}$ and other background channels drop when the cuts become tighter. Consequently, the S/B ratio is higher for tighter cuts. However, signal significancy is higher when the cuts get looser which suggest



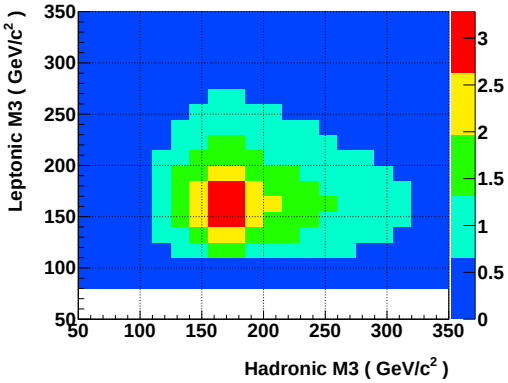
M2M3 - scale-up JES



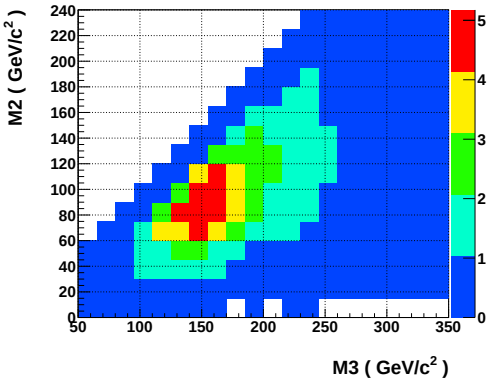
M3M3 - scale-up JES



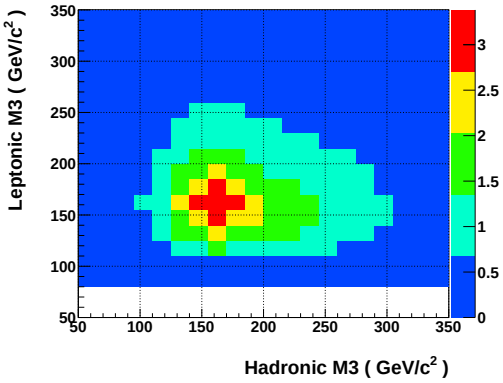
M2M3 - no JES scale



M3M3 - no JES scale



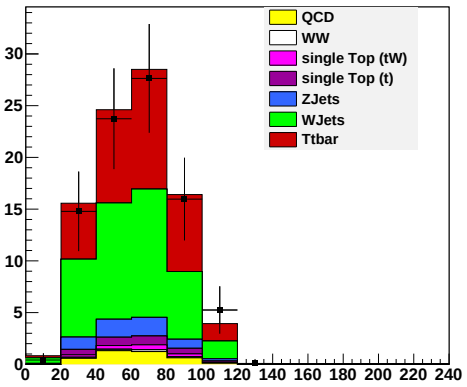
M2M3 - scale-down JES



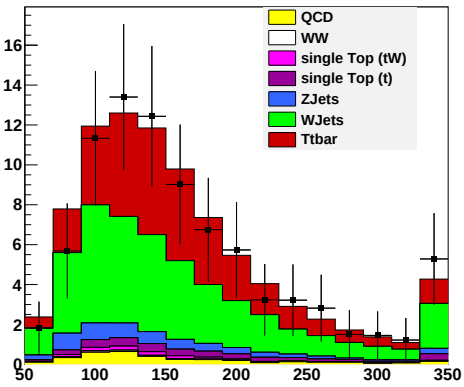
M3M3 - scale-down JES

Figure 5.9: Event distributions of $t\bar{t}$ MC in the $M2_h - M3_h$ and $M3_h - M3_l$ plane. The top row is the results with scaled-up JES. The bottom row is from scaled-down JES. The middle one is with the original JES setup.

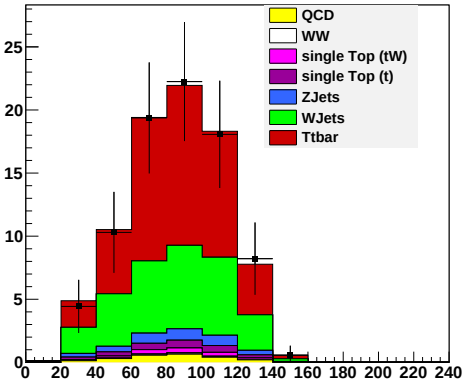
that better statistic error if no cut is applied. Similar test was also done with additional criterion, $\Delta M3 < 10$. The results suggest the same conclusion (Fig. 5.12).



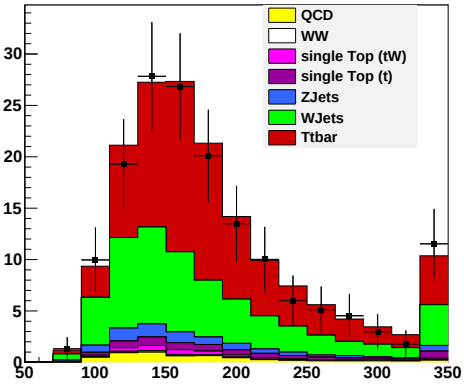
$M2_h$ with $M3_h < 140$ GeV/c



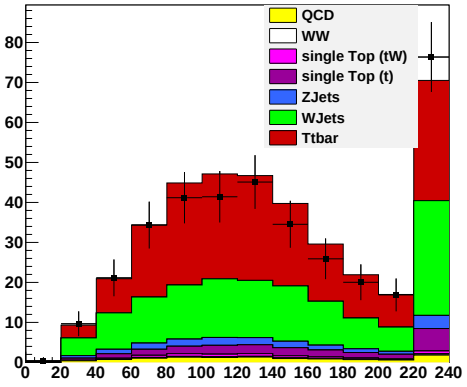
$M3_h$ with $M2_h < 60$ GeV/c



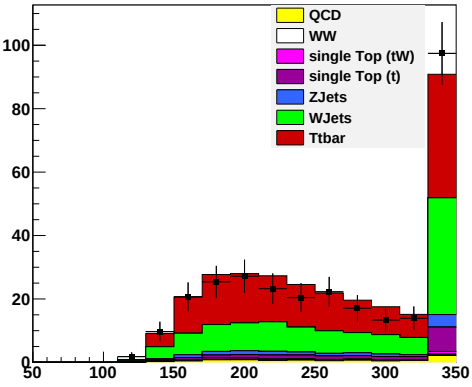
$M2_h$ with $140 \leq M3_h \leq 200$ GeV/c



$M3_h$ with $60 \leq M2_h \leq 100$ GeV/c

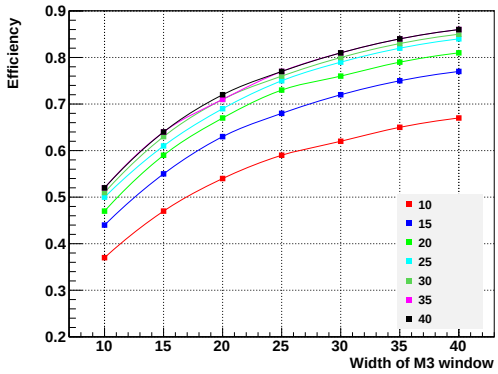


$M2_h$ with $M3_h > 140$ GeV/c

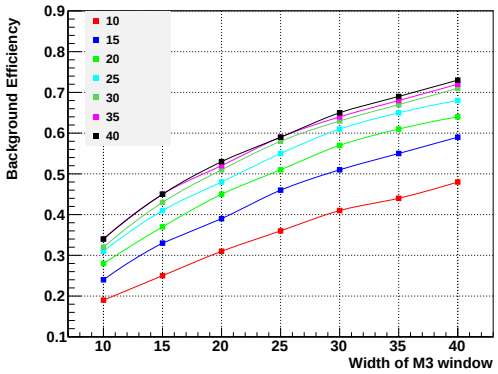


$M3_h$ with $M2_h > 60$ GeV/c

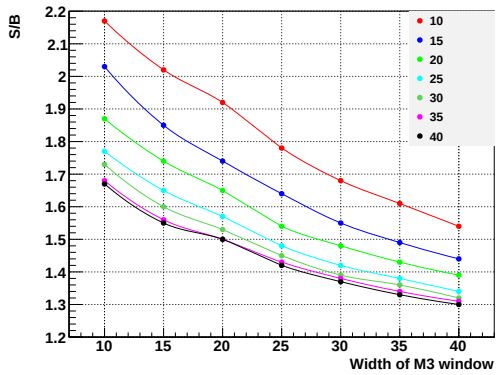
Figure 5.10: Solo $M2_h$ and $M3_h$ distribution under different cosntraints.



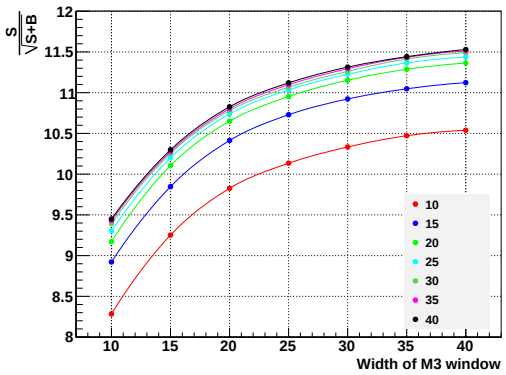
$t\bar{t}$ Efficiency



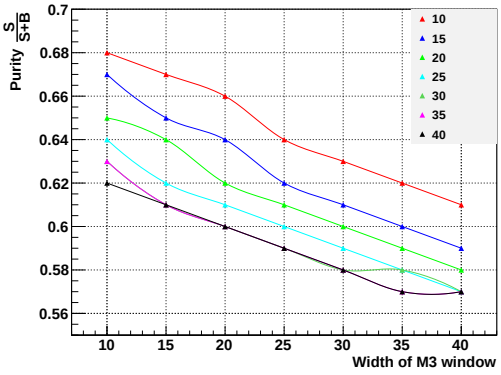
Background efficiency



S/B ratio

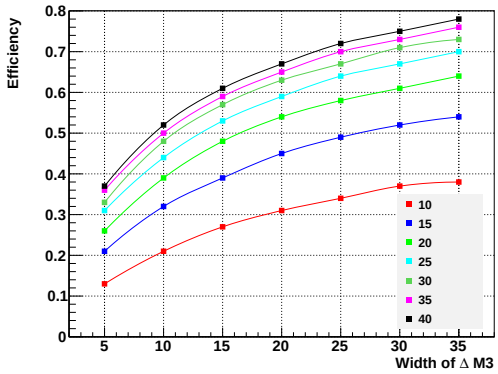


Signal Significancy

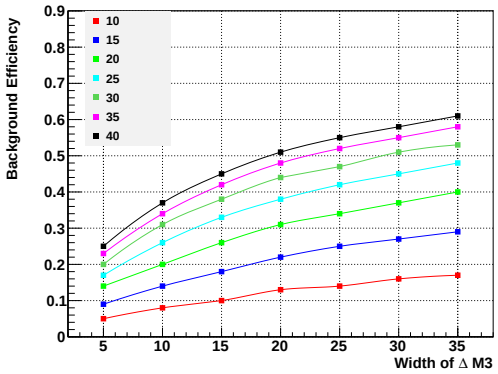


Purity

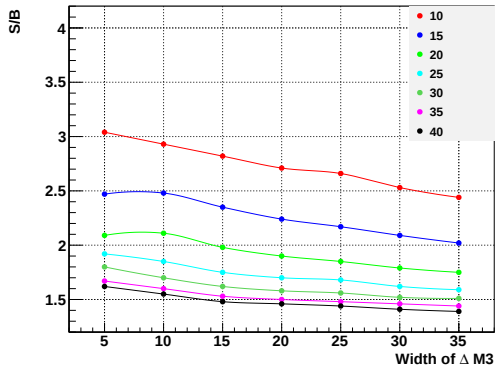
Figure 5.11: The x-axis is the width of $M3_h$. The variations of $M2_h$ width are shown in different color.



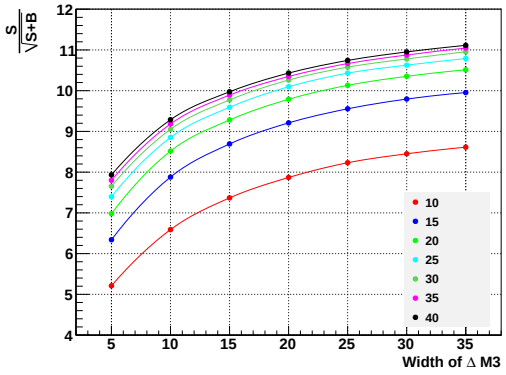
$t\bar{t}$ Efficiency



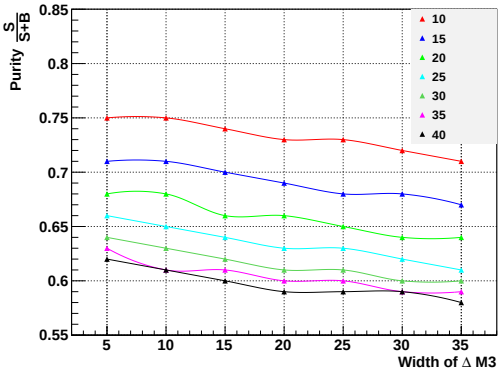
Background efficiency



S/B ratio



Signal Significancy



Purity

Figure 5.12: The x-axis is the width of $\Delta M3$. The variations of $M2_h$ width are shown in different color. $M3_h$ width is 1.5 times of $M2_h$ width.

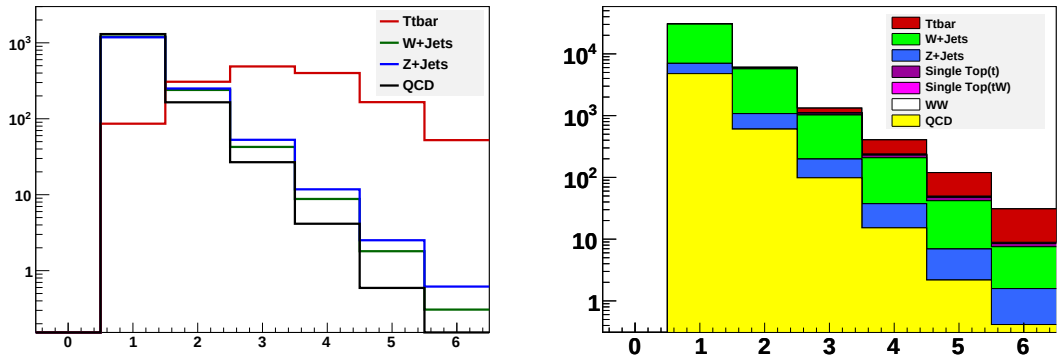
Chapter 6

Background Analysis

6.1 Data-Driven Background Estimation

For semi-leptonic $t\bar{t}$ events, there are at least four partons from the top quark decay whereas for the backgrounds, most of the jet multiplicity comes from initial state radiation. This is shown in Fig. 6.1a) and 6.1b) which demonstrate that the shapes of the signal and backgrounds as a function of multiplicity are very different. The background, both QCD and V+jets (W+jets and Z+jets), shows the characteristics of a Berends[12] scaling behavior as number of jet increases, whereas the $t\bar{t}$ signal declines much slower. This behavior is also confirmed by CMS electroweak physics group[48].

To calculate the QCD and V+jets background in the 4-jet bin and above we make use of the fact that the 2-jet bin is essentially signal free. By taking the ratio of the number of 4-jet events(and above) to 2-jet events from the background MC and scaling this by the



(a) Normalized to the same number of events (b) Normalized to the same luminosity

Figure 6.1: Jet multiplicity distribution for each event type. (a) normalized to the same number of event. (b) normalized to the same luminosity.

number of 2-jet events seen in the data, we can obtain a data-normalized prediction for the inclusive 4-jet background. Thus

$$\text{Number of estimated background} = \frac{N_{bg}^{MC\,\geq 4jets}}{N^{MC\,2jets}} \times N_{data}^{2jets} \times \varepsilon^{background} \tag{6.1}$$

$N_{bg}^{MC\,4jets+}$ is number of $\geq$ 4-jet events from the MC, $N^{MC\,2jets}$ is number of 2-jet events from MC, N_{data}^{2jets} is the number of 2-jet events measured from data and $\varepsilon^{background}$ is the topological cut efficiency for background channels. The 4-jet:2-jet ratio can be expressed as

$$R(\geq 4jet : 2jet) = \frac{N_{bg}^{MC\,\geq 4jets}}{N^{MC\,2jets}} = \frac{fN_{VJets}^{4jets+} + gN_{QCD}^{4jets+}}{fN_{VJets}^{2jets} + gN_{QCD}^{2jets}} \tag{6.2}$$

$N_{QCD}^{4/2jets}$ and $N_{VJets}^{4/2jets}$ are the number of events from QCD and other processes predicted by the MC. The coefficients f and g are the Data-MC normalization for V+jets and QCD processes used to correct the fraction of QCD and V+jets composition in MC according to data. The method basically relies on the correct prediction for the slope of the jet multiplicity in each channel. However, the normalization does not attempt to fix the jet multiplicity of each process which is sensitive to the modeling in MC. Thus, it doesn't couple with the theoretical systematic uncertainty because the normalization for 4-jet events and 2-jet events in each proecess are actually cancelled. According to the systematic studies, the factorization scale, Q^2, is the largest uncertainty for this method.

A second application of Berends scaling to estimate the background can be derived from equation 6.2

$$R(\geq 4jet : 2jet) = \frac{N_{bg}^{\geq 4jet}}{N^{2jet}} = \frac{N_{VJets}^{\geq 4jet} + N_{QCD}^{\geq 4jet}}{N^{2jets}} \tag{6.3}$$

$$= \frac{N_{QCD}^{2jet}}{N^{2jet}} \frac{N_{QCD}^{\geq 4jets}}{N_{QCD}^{2jet}} + \frac{N_{VJets}^{2jet}}{N^{2jet}} \frac{N_{VJets}^{\geq 4jets}}{N_{VJets}^{2jet}} \tag{6.4}$$

$$= f^{QCD} Q(4:2) + f^{VJets} V(4:2) \tag{6.5}$$

where the

$$Q(4:2) = \frac{N_{QCD}^{\geq 4jet}}{N_{QCD}^{2jet}} = (\frac{N_{QCD}^{\geq 4jet}}{N_{QCD}^{4jet}})(\frac{N_{QCD}^{4jet}}{N_{QCD}^{3jet}})(\frac{N_{QCD}^{3jet}}{N_{QCD}^{2jet}}) \quad (6.6)$$

$$= k_q(Q(4:3)Q(3:2)) \quad (6.7)$$

and

$$V(4:2) = \frac{N_{VJets}^{\geq 4jet}}{N_{VJets}^{2jet}} = (\frac{N_{VJets}^{\geq 4jet}}{N_{VJets}^{4jet}})(\frac{N_{VJets}^{4jet}}{N_{VJets}^{3jet}})(\frac{N_{VJets}^{3jet}}{N_{VJets}^{2jet}}) \quad (6.8)$$

$$= k_v V(4:3)V(3:2) \quad (6.9)$$

Thus the ratio4:2 can be rewitten as

$$R(\geq 4jet:2jet) = f^{QCD}k_q Q(4:3)Q(3:2) + f^{VJets}k_v V(4:3)V(3:2) \quad (6.10)$$

$$= f^{QCD}k_q\ Q(2:1)^2 + f^{VJets}k_v\ V(2:1)^2 \quad (6.11)$$

with application of Berends scaling

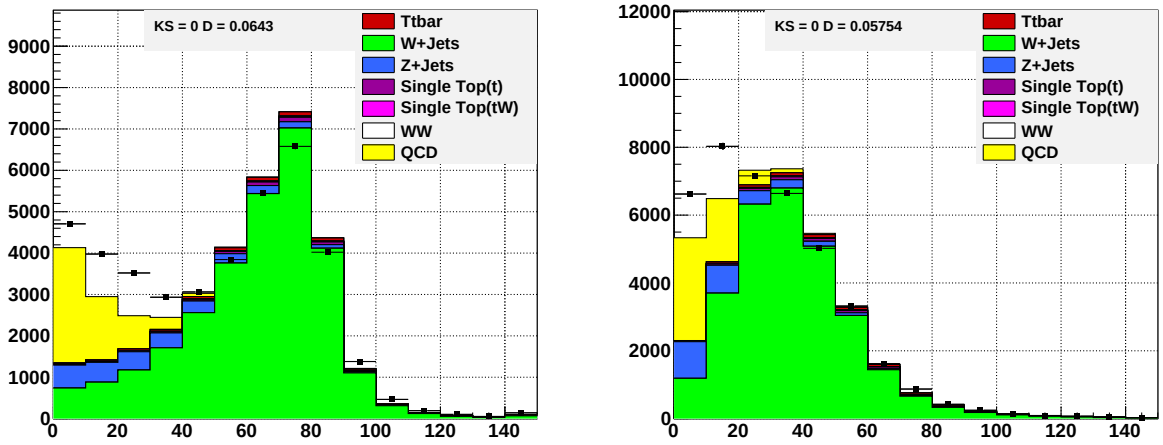
$$Q(2:1) = \frac{N_{QCD}^{2jet}}{N_{QCD}^{1jet}} \simeq \frac{N_{QCD}^{3jet}}{N_{QCD}^{2jet}} \simeq \frac{N_{QCD}^{4jet}}{N_{QCD}^{3jet}} \tag{6.12}$$

$$V(2:1) = \frac{N_{VJets}^{2jet}}{N_{VJets}^{1jet}} \simeq \frac{N_{VJets}^{3jet}}{N_{VJets}^{2jet}} \simeq \frac{N_{VJets}^{4jet}}{N_{VJets}^{3jet}} \tag{6.13}$$

where f^{QCD} and f^{VJets} are the fraction of QCD/V+jets in the 2-jet events. Q(2:1) and V(2:1) are the 2jet:1jet ratio for QCD and V+jets events and they are equivalent to $\alpha_s(Q^2)$ with corresponding Q^2. k_v and k_q are the ratio for inclusive and exclusive 4-jet events for V+jets and QCD process. If assuming the validity of the Berends scaling, the Njet:N-1jet ratio remains nearly constant across different jet multiplicity. Under this assumption, 4-jet:2-jet ratio can be replaced by 2-jet:1-jet ratio with equation 6.12 and equation 6.13. The advantage of this method is that the slope of jet multiplicity is determined from data. In addition, ratio 2:1 is nearly signal($t\bar{t}$) free and is well measured with higher statistic. The fraction of QCD and Vjets component needed in the 2-jet bin can be extract from M_{T} distribution but the low M_{T} region (QCD dominated region) has larger uncertainty. The ratios of inclusive and exclusive 4-jet event (k_q and k_v) still rely on MC prediction and both are sensitive to theoretical uncertainties which are not well determined. Moreover, the possible correction at high jet multiplicity for Berends scaling has been seen in the study from CMS electroweak group [48]. It also brings the uncertainties for this method.

6.2 Data-MC Comparison

From 2010 data analysis it was clear that there was a discrepancy between MC samples and data. As an example of the problem, in Fig. 6.2 we show the M_T distribution for events containing at least one particle flow jet above 25 GeV. There is a clear discrepancy, not only in the region dominated by QCD, but also in the region dominated by V+jets.



M_T distribution $\not\!\!E_T$ distribution

Figure 6.2: M_T and $\not\!\!E_T$ distribution of the events with ≥ 1 jets in MC and Data

First we study the consistency of QCD and V+jets prediction in exclusive jet multiplicity bins. We obtain a good separation of the two backgrounds by selecting events in restricted regions of the $\not\!\!E_T$ vs M_{T} plane. The $\not\!\!E_T - M_{\mathrm{T}}$ plane is used because M_{T} and $\not\!\!E_T$ both have distinguishable difference and separation between QCD events and V+jets events (see Fig. 6.3). Therefore, We define two control regions:

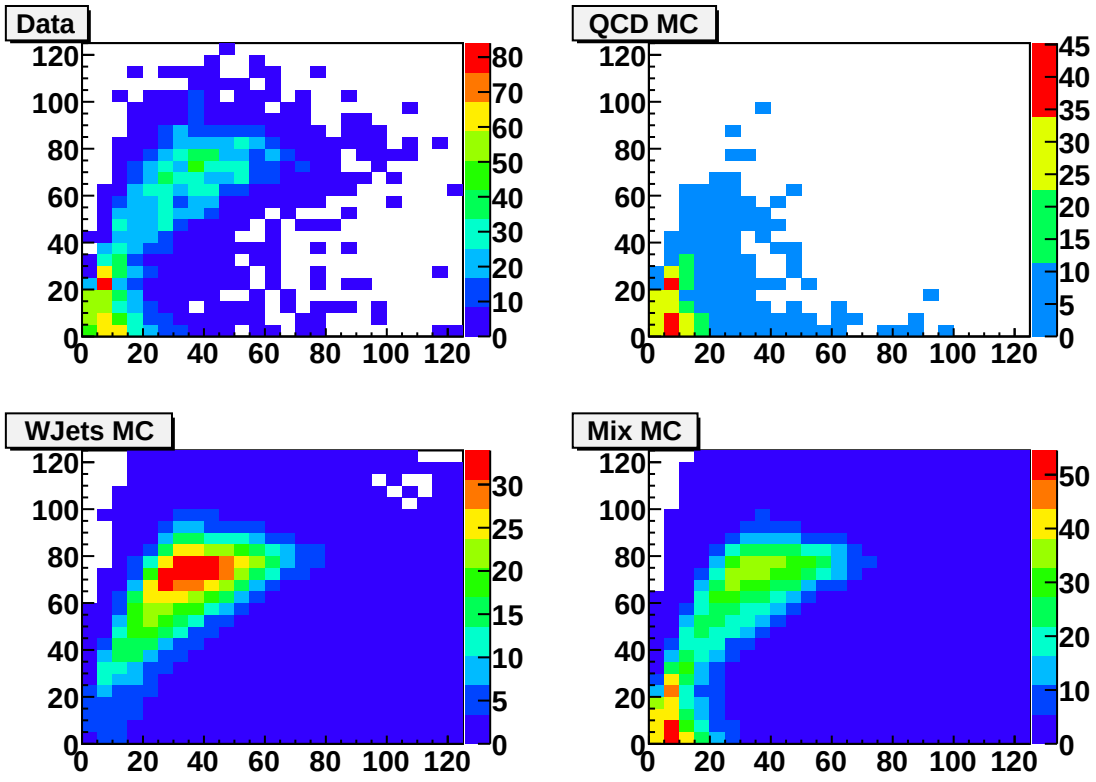


Figure 6.3: $\not\!\!E_T$ and M_T distributions in data and MC samples. The x-axis is $\not\!\!E_T$ and the y-axis is M_T

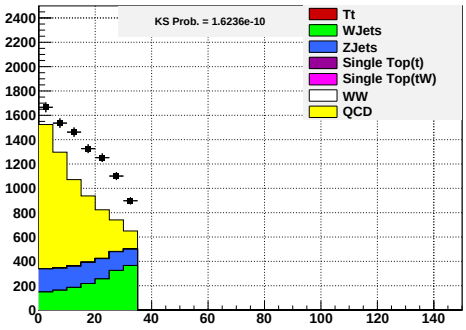
- QCD control region- $M_T < 35$ GeVand $\not\!\!E_T < 20$ GeV
- V+jets control region- $M_T > 50$ GeV and $\not\!\!E_T > 30$ GeV

which are used to extract two different types of control samples. M_{T} distribution is then applied to both samples in order to determine the normalization in equation 6.2. Figure 6.4 shows the comparison between the data and MC in the two control regions for the 1 and 2-jet exclusive bins. Clear discrepancies can be seen in all four cases. The V+jets control region is very clean in the 1-jet bin but has a 2.5 $\%$ contamination from $t\bar{t}$ events in the 2-jet bin. The QCD control region has $\sim 20\%$ contamination from V+jets for both

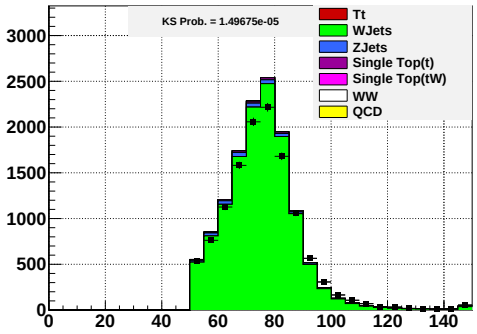
jet bins but they can be removed after the normalization of V+jets. We start by doing a bin-by-bin normalization as a function of M_T to determine the correction needed to the MC V+jets predictions in the V+jets. The values are determined by fitting the χ^2 curve of different normalizations. The error of the normalization is the half width of the χ^2 curve at the minimum plus 0.5. The results, which are consistent with single factors for each jet bin, are listed in Table 6.2. Next we apply these factors to the V+jets background in the QCD control region and repeat the process to find the corrections needed for the QCD predictions. The M_T distribution is chosen because it has distinguishable shape difference between QCD type and V+jets type which has been discussed in section 5.3.

The normalization processes doesn't apply to 3-jet bin and above events. As shown in Fig.6.1, $t\bar{t}$ have significant amount in 3-jet bin and the above events. The normalization will be biased by $t\bar{t}$ events because $t\bar{t}$ is the signal to be measured.

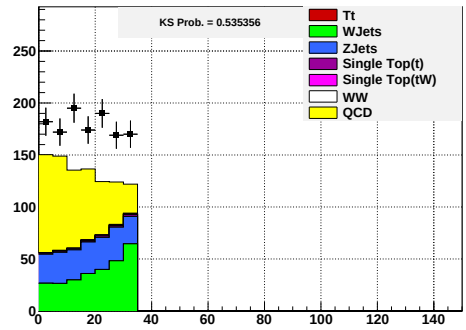
Table 6.2 shows the 1- and 2-jet normalization results. For the V+jets MC we find that we need -8% and -9% corrections for the 1- and 2-jet bins, respectively and +57% and +76% for the same bins of the QCD MC. As we also need to correct the MC prediction for higher jet multiplicities, we apply the factors from the 2-jet analysis to these bins. We illustrate the effect of these corrections in Figure 6.5 which shows the comparison between data and MC predictions for the jet multiplicity distribution (Njets $\geq$ 1) both before and after correction. A quantitative examination, the Kolmogorov-Smirnov test, is performed



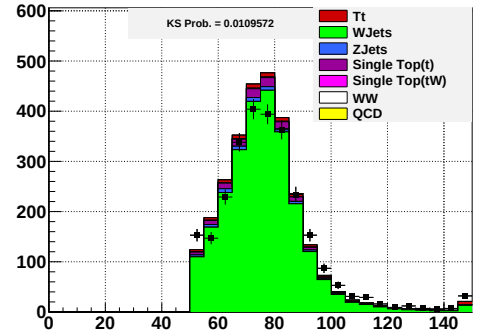
QCD control region : 1-jet bin



V+jets control region : 1-jet bin



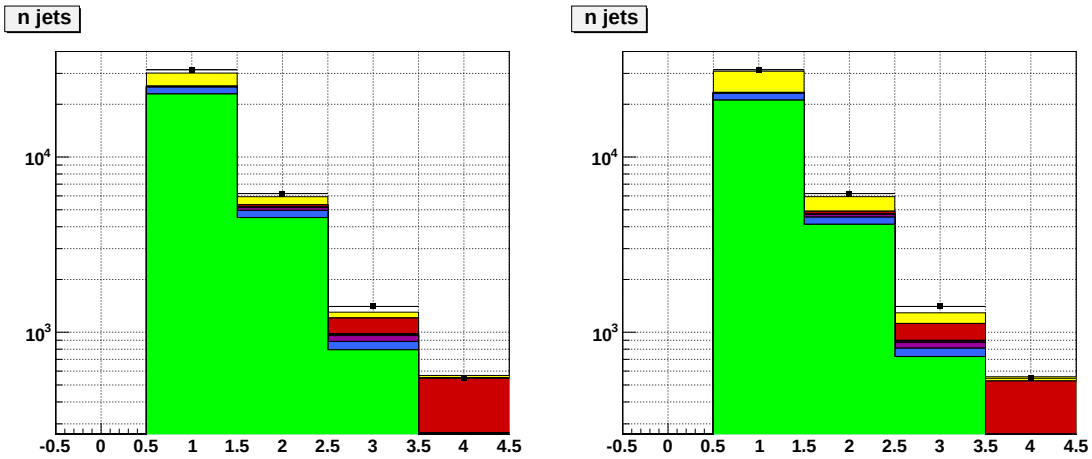
QCD control region : 2-jet bin



V+jets control region : 2-jet bin

Figure 6.4: Comparision of data and MC for M_T in the control regions

for other kinematic variables in 2-jet and 4-jet events. From the results(Table 6.1 and plots in Appendix B), it shows the significant improvement for the consistency of kinematic variables after normalization is adjusted. In addition, the effects on topological variables are also checked shows no significant shape change in these distributions (Appendix A.1).



a. before normalization b. after normalization

Figure 6.5: Jet multiplicity before and after normalization. The 0 jet events are excluded

	Before normalization 2 Jet Events	normalization 2 Jet Events	Before normalization $\geq$ 4 Jet Events	normalization $\geq$ 4 Jet Events
muon P_T	0.21	<0.01	0.72	0.98
muon η	0.90	1	0.98	0.98
muon Iso	<0.01	0.03	0.68	0.95
M_T	<0.01	0.03	0.84	0.96
HT_{lep}	<0.01	<0.01	0.97	1
HT_{tot}	0.03	<0.01	0.79	0.92
MET	<0.01	0.10	0.27	0.57
Jet1 P_T	<0.01	<0.01	0.71	0.82
Jet2 P_T	0.08	<0.01	0.83	0.91
Jet1 η	0.64	0.74	1	1
Jet2 η	0.86	0.92	1	1

Table 6.1: KS Probabilities for kinematic variables in 2 jet events and in $\geq$4jet events

	1-jet bin	2-jet bin
QCD sample	1.594 ± 0.022	1.761 ± 0.072
V+jets samples	0.915 ± 0.009	0.923 ± 0.020

Table 6.2: MC normalization corrections for 1- and 2-jet exclusive bins

6.3 Cross Check

As test of the robustness of the correction and prediction procedure, we do a closure test on the ratio of the number of 2-jet events to the number of 1-jet events and compare the results predicted by our corrected MC and the data. This ratio is chosen because it is essentially free from signal contamination.

Table 6.3 shows a comparison of two MC results obtained using the direct MC truth and the MC events in the two control regions. The first column gives the results for the QCD and V+jets ratios, calculated directly from the sample itself without no $\not\!\!E_T$ and M_T selection. The second is what we obtain from the mixed MC samples in our two control regions. The results are calculated before any normalization corrections have been applied. The agreement shows that there is not an inherent bias introduced by the use of the cuts to define the control regions. Table 6.4 shows the results after the normalizatiions are applied. Although the ratio from MC truth is slightly changed, it is still consistent between the results from MC truth and the one from control region.

In Table 6.5 we show the comparison between normalized MC and data by using the two control regions. Column 1 is the MC samples with normalization corrections and column 2 is the result from observed data. For both the QCD and V+jets regions we find excellent agreement between the predictions and the data.

	Using MC truth	Control region
QCD control region	0.126 ± 0.002	0.134 ± 0.001
V+jets control region	0.212 ± 0.001	0.213 ± 0.001

Table 6.3: Ratio of 2-jets:1-jet for QCD and V+jets MC. The two columns correspond to the direct ratio from MC truth (column1) and the ratio obtained for the control regions.(column2)

	Using MC truth	Control region
QCD control region	0.140 ± 0.002	0.134 ± 0.001
V+jets control region	0.214 ± 0.001	0.215 ± 0.001

Table 6.4: Ratio of 2-jets:1-jet for QCD and V+jets MC after applying the normalization

	M.C. Prediction	Data
QCD control region	0.134 ± 0.001	0.135 ± 0.004
V+jets control region	0.215 ± 0.001	0.218 ± 0.005

Table 6.5: Comparison of predicted and observed 2-jet:1-jet ratios using the control region.

Another cross-check that can be done is using the equation 6.5. From the equation 6.5 and table 6.6, f^{QCD} and f^{VJets} are detetermined from the normalized MC templates in 2-jet bin. k_q and k_v also can be estimated from $\geq$ 4-jet and exclusive 4-jet events of the normalized MC samples. The ratio2:1 for QCD or VJets are obtained by using the control region to data. Thus, the ratio4:2 from second application of Berends scaling is

$$
\begin{aligned}
R(\geq 4jet : 2jet) \quad &= f^{QCD}\, k_q\, Q(2:1)^2 + f^{VJets}\, k_v\, V(2:1)^2 \\
&= 0.177 \times 1.148 \times 0.135^2 + 0.823 \times 1.249 \times 0.218^2 \\
&= 0.0562
\end{aligned}
$$

Comparing with the ratio4:2 from the first method, 0.0468 $\pm$ 0.009 (syst.), both values are consistent.

	Data	MC	$t\bar{t}$	W+jets	Z+jets	t	tW	WW	QCD
1Jet	32017	31376	37	21416	2046	183	7	56	7631
2Jet	6390	6148	131	4289	437	171	14	40	1066
3Jet	1444	1326	208	762	92	67	13	11	173
4Jet	419	402	170	157	21	20	6	2	27
$\geq$4Jet	562	550	263	195	26	25	7	3	31

Table 6.6: Cut flow table for data and MC processes(after normalized to data)

Chapter 7

Cross Section Calculation

The $t\bar{t}$ cross-section is calculated using

$$\sigma = \frac{N_{obs} - N_{bg}}{\int L \times A \times \varepsilon} \tag{7.1}$$

where N_{obs} is the number of observed events, $\int L$ is the integrated luminosity, A is the acceptance, ε is the efficiency of the analysis for signal and N_{bg} is the number of background events. ε_{bg} is the efficiency of the analysis for the background events. The A and ε are evaluated by using MC simulation. Thus, the correctness of MC is an important issue which is directly related to the acceptance and efficiency.

One known problem is the branching ratio of W boson in the MadGraph generator. The branching ratio of $W \rightarrow \mu\nu$ is set to its leading order value 1/9. However, the current

world average is 0.1080 $\pm$ 0.0009 according to PDG (Particle Data Group) [49]. This also implies that the ratio of hadronic final state is 0.676. Therefore, a re-weighting of simulated events is needed for events containing two real W bosons, $t\bar{t}$ and single top tW channel in this study (WW production is generated by Pythia). For a single lepton channel, its branching ratio has to be corrected from $1/9 \times 6/9$ to 0.108×0.676. Thus, a re-weighting value for single lepton channels is $0.108 \times 9 \times 0.676 \times 9/6$. The same adjustment also applies to other decay channels. $0.108 \times 9 \times 0.108 \times 9$ is used for di-lepton channels and $0.676 \times 9/6 \times 0.676 \times 9/6$ is used for hadronic channels. The other correction to MC simulation is the scale factor for the efficiency. This is discussed in the next section.

7.1 Efficiency

The efficiency in the cross-section calculation has two major components. One is the efficiency for pre-selection. In this study, the efficiency for the isolated muon selection belongs to this part. The other is the efficiency for the rest of the final selection, including topological cuts. The muon efficiency can be factorized into the following three components.

$$\varepsilon = \varepsilon_{ID}\ \varepsilon_{Iso}\ \varepsilon_{Trigger} \tag{7.2}$$

$$= (\frac{\mu^{ID}}{\mu^{reco}})(\frac{\mu^{Iso+ID}}{\mu^{ID}})(\frac{\mu^{Trigger}}{\mu^{Iso+ID}}) \tag{7.3}$$

where ε_{ID} is muon identification efficiency, ε_{Iso} is muon isolation efficiency and $\varepsilon_{Trigger}$ is the trigger efficiency from HLT. Each component is evaluated progressively as the definition in equation 7.3. ε_{ID} is the ratio of the offline reconstructed muons which pass the identification criteria. The number of muons passing the identification criteria is the denominator of ε_{Iso} and its numerator is the denominator for $\varepsilon_{Trigger}$. Although the efficiency can be estimated from MC, a direct measurement from data is still needed in order to correct the discrepancy between data and MC.

Therefore, a tag and probe method is adopted to directly measure the muon efficiency by using $Z \rightarrow \mu^+\mu^-$ events from data. The idea of the tag and probe method is examining a candidate object from the control events which are tagged with a specific signature. In the case of the muon efficiency study, the control sample is selected with two muons with invariant mass of Z boson within 15 GeV/c^2. The two muons must have opposite charge. One of them must be satisfied with tag muon criteria which is the same as the muon selection requirement for this study and also fire HLT. The probe muon must be Global and Tracker muon and then be examined with ID and Iso or trigger criteria. The spectrum of the invariant mass of the di-muon is used. The Z mass peak is fitted with Gaussian distribution. Exponential function is used to fit the background shape for substraction. The efficiency is defined to be the ratio of the integral of final Z peak before and after the probing as below.

$$\varepsilon = \frac{\text{event with probed muon passed}}{\text{total events with at least one muon tagged}}$$

This process is applied to data and MC samples. Thus the efficiency ratio of data and MC from Z $\rightarrow \mu\mu$ is obtained and it is assumed to be the same as the ratio of $t\bar{t}$ efficiency with muon plus jets topology.

$$\frac{\varepsilon_{t\bar{t}}^{data}}{\varepsilon_{t\bar{t}}^{MC}} = \frac{\varepsilon_{Z,T\&P}^{data}}{\varepsilon_{Z,T\&P}^{MC}} \tag{7.4}$$

where $\varepsilon_{t\bar{t}}^{data}$ is the efficiency of $t\bar{t}$ events in data, $\varepsilon_{t\bar{t}}^{MC}$ is the efficiency of $t\bar{t}$ events in MC. $\varepsilon_{Z,T\&P}^{data}$ is the efficiency of Z events measured by tag and probe method in data. $\varepsilon_{Z,T\&P}^{MC}$ is the efficiency of Z events measured by tag and probe method in MC. From the ratio of tag and probe Z events, the efficiency of $t\bar{t}$ event in data can be derived since the efficiency of $t\bar{t}$ event in MC is known. Therefore, the ratio of $Z \rightarrow \mu\mu$ from tag and probe serves as the scale factor for the efficiency correction of $t\bar{t}$ events.

The scale factors for muon efficiency used in top analysis have been determined by an official efficiency group in Top physics group[61]. A cross-check analysis was also performed with a different isolation cut [29]. Since this analysis uses the same looser isolation

defination as the cross check analysis, the values from the cross check group are used and listed in table 7.1.

	Trigger	ID&ISO	ISO $\geq$ 2jets
Data	0.918	0.957	0.97
MC	0.931	0.944	0.957
Scale Factor	0.987	1.013	1.01

Table 7.1: Scale factor(SF) of muon efficiency for muon isolation ≤ 0.1

In addition to the muon efficiency, the efficiency of the rest of the event selection in chapter5 is calculated by

$$\varepsilon = \frac{\sum w_i n_i^+}{\sum w_i n_i} \tag{7.5}$$

where n_i is the total number of event from channel i, n_i^+ is the number of event passed the selection and w_i is its weighting for the channel i. According to the rule of error propagation, the error of the efficiency for the weighted events is

$$\delta\varepsilon = \sqrt{\sum (\frac{\partial\varepsilon}{\partial n_i^+})^2 n_i^+ + (\frac{\partial\varepsilon}{\partial n_i^-})^2 n_i^-} \tag{7.6}$$

where n_i^- is $n - n_i^+$, the number of event fail to pass the selection criteria for channel i.

Therefore, its final formula can be expressed as

$$\delta\varepsilon = \frac{\sqrt{(\sum w_i n_i^+)^2(\sum w_i^2 n_i^-) + (\sum w_i n_i^-)^2(\sum w_i^2 n_i^+)}}{(\sum w_i n_i)^2} \tag{7.7}$$

7.2 Method Validation

In order to validate the cross-section measurement, there are two approaches that have been done for this purpose. The first one is the closure test and the second is pseudo experiments. The closure test is done by using half of MC samples with normalization of 36.1 pb^{-1}. Every other event is selected in order to cover all phase space. The Closure test is a simple test using large statistic from each channel.

In Table 7.2 we present the comparison of a series of cross-section results obtained using no topological selection (top row) and with two different topological selections (rows 2-3). In each case the result is quoted together with its statistical error. The first column gives the results obtained via the MC closure test with 36.1 pb^{-1} equivalent luminosity. The input cross-section of 157.5 pb is recovered reliably for each selection and the errors suggest that we should expect a statistical uncertainty from 11% to 19 %. The second column is the cross-section derived from the data. The measured cross-section and statistical

error are found to be consistent.

	σ(pb) MC Closure Test	σ(pb) Data	A ε_{tt}	ε_{bg}
Baseline selection	157.5 $\pm$ 14.5	160.4 $\pm$ 15.0	0.1872 $\pm$ 0.0104	1.0 $\pm$ 0
$70 < M2_h < 90$, $150 < M3_h < 190$	159.5 $\pm$ 17.2	153.3 $\pm$ 17.3	0.1007 $\pm$ 0.0080	0.3139 $\pm$ 0073
$\lvert M3^{lep}$ - $M3^{had}\rvert < 10$, $70 < M2_h < 90$ $150 < M3_h < 190$	160.9 $\pm$ 23.5	156.5 $\pm$ 23.7	0.0469 $\pm$ 0.0056	0.0995 $\pm$ 0.0048

Table 7.2: Closure test and data analysis of cross-section measurement

Another approach is performing a pseudo experiment. Each pseudo experiment (ensemble) is made of events from every channel with a randomized number of events. The number of events in each channel is given randomly according to a Poisson function with the mean value at 36.1 pb^{-1}. After running a large amount of pesudo experiments, the mean value from all ensembles represents the most probable measured cross-section and the width of the distribution is the statistical error. The algorithm of the pseudo experiments is constructed as follow.

1. Shuffle the events for each MC sample.

2. For each channel, a specific number of events will be chosen randomly according to a Poisson distribution with the mean value which comes from the closure test based on 36.1 pb^{-1}.

3. Pick the events in each channel with the number of the events determined from step 2. Once the MC sample pools run out of events, step 1 will start over again and the event in the sample pools will be re-used.

4. Since current amount of MC sample is not enough to cover a large amount of pseudo experiments, the events must be re-used. In addition to step 1, a re-used event will be processed by phase space smearing:

   - The whole system of lepton, jets and $\not\!\!E_T$ are slightly boosted. The boosting vector is determined by a flat random function and is only along the z direction.
   - The whole system is also rotated around z axis randomly.
   - For jets, the four momentum will be smeared randomly according to a Gaussian distribution. The width of the Gaussion function is the uncertainty of the jet energy scale.
   - For muons, the four momentum of the muon is smeared as for jets but the width of the Gaussion is 1% of the total momentum. An M_T constraint is imposed to get the correct $\not\!\!E_T$ resolution.
   - For $\not\!\!E_T$, the smearing is the same as for jets with the same width. Before smearing, the $\not\!\!E_T$ is re-calculated according to the original M_T constriant after the muon is smeared.

5. Mix all the chosen event samples for 2-jet and 4-jet events respectively and calculate the cross-section.

The result from the pseudo experiments is shown in Figures7.1, 7.2 and 7.3. The mean value from the distribution is 158 pb which agrees very well with the input cross section, 157.5 pb and the expected statistic uncertainty is 15 pb for baseline selection and 18 pb and 25 pb for topological cut1 and cut2.

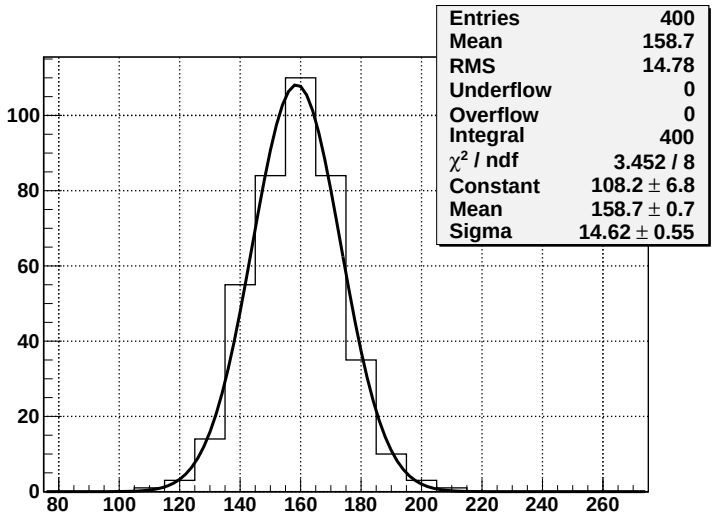


Figure 7.1: The pseudo experiment using the baseline selection

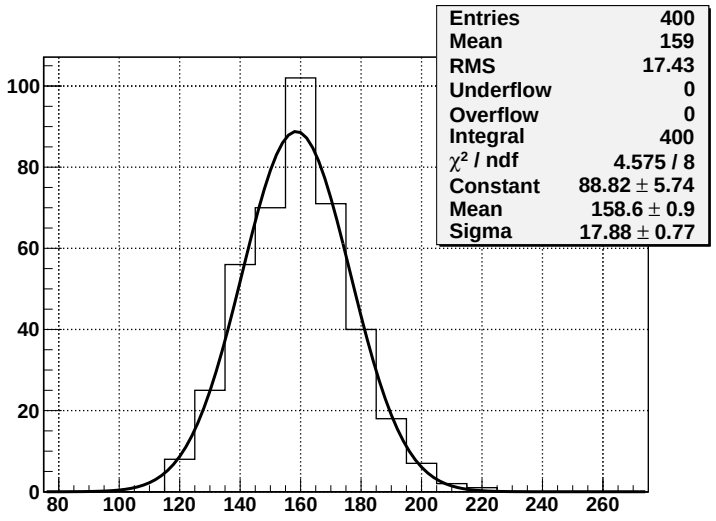


Figure 7.2: The pseudo experiment using topological cut 1

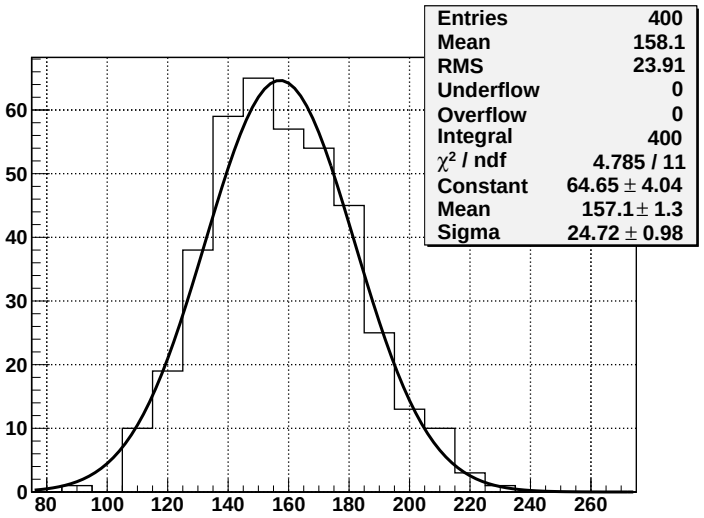


Figure 7.3: The pseudo experiment using topological cut 2

7.3 Systematic Uncertainty

The consideration of systematic uncertainty includes theorectical modeling and experiment setup. The samples used in the systematic studies are listed in table 7.3. Their effects are evaluated separately and summed in quadrature. In the method of this study, the systematics mainly affect 4-jet:2-jet ratio for background estimation and propagate to final cross-section measurement. The uncertainties of 4-jet and 2-jet ratio are shown in table 7.4. The impacts to cross-section of each systematic are listed in tables 7.5, 7.6 and table 7.7 with different topological cut scenarios. The details are discussed in the following sections.

Channel	σ (pb^{-1})	Number of generated event	Modification
$t\bar{t}$, Inclusive	121	1.5M	Z2 tuned
$t\bar{t}$, Inclusive	94.6	1.5M	smaller ISR/FSR
$t\bar{t}$, Inclusive	100.5	1.5M	larger ISR/FSR
$t\bar{t}$, Inclusive	105.9	1M	Matching Efficiency up($\times$ 2)
$t\bar{t}$, Inclusive	111.1	1M	Matching Efficiency down($\times$ 0.5)
$t\bar{t}$, Inclusive	75	1M	$Q^2\times$ 2
$t\bar{t}$, Inclusive	186.6	1M	$Q^2\times$ 0.5
W+Jets, leptonic decayed	25950	10M	Matching Efficiency up($\times$ 2)
W+Jets, leptonic decayed	24990	3.1M	Matching Efficiency down($\times$ 0.5)
W+Jets, leptonic decayed	27230	6.2M	$Q^2\times$ 2
W+Jets, leptonic decayed	26530	5.5M	$Q^2\times$ 0.5
Z+Jets, leptonic decayed	2366	1.5M	Matching Efficiency up($\times$ 2)
Z+Jets, leptonic decayed	2205	1.5M	Matching Efficiency down($\times$ 0.5)
Z+Jets, leptonic decayed	2583	1.5M	$Q^2\times$ 2
Z+Jets, leptonic decayed	2541	1.5M	$Q^2\times$ 0.5

Table 7.3: MC samples used for systematic studies

JES (down,up)	-6.84 %	+7.48 %
JER (+10%,-10%)	-0.43 %	+0.64 %
Unclustered Energy (-10%,+10%)	-2.35 %	+2.14 %
QCD (+50%,-50%)	-0.21 %	+0.21 %
Single top and WW (+30%,-30%)	-2.35 %	+2.35 %
W+jets Matching-Threshold (higher,lower)	-4.49 %	+9.4 %
W+jets Q^2 (up,down)	-16.88 %	+12.82 %
Z+jets Matching-Threshold (higher,lower)	-1.07%	$<$ 0.1 %
Z+jets Q^2 (up,down)	-2.99 %	+3.21 %
total	-19.32 %	+18.15 %

Table 7.4: Systematic impacts to 4-jet:2-jet ratio from different sources using the baseline selection.

$t\bar{t}$ MC model: D6T:Z2	N/A	-1.6 %
$t\bar{t}$ ISR/FSR (smaller,bigger)	+0.3%	-3.8 %
$t\bar{t}$ Q^2 (up,down)	+5.0%	-6.3 %
$t\bar{t}$ Matching-Threshold (higher,lower)	+2.8%	-6.7 %
W+jets Matching-Threshold (higher,lower)	+5.1%	-10.7 %
W+jets Q^2 (up,down)	+19.2%	-14.6 %
Z+jets Matching-Threshold (higher,lower)	+1.2%	$<$ 0.1 %
Z+jets Q^2 (up,down)	+3.4%	-3.6 %
QCD (+50%,-50%)	+0.2%	-0.2 %
Single top and WW (+30%,-30%)	+2.7%	-2.7 %
JES (down,up)	+16.9%	-14.8 %
JER (+10%,-10%)	+0.7%	-0.9 %
Unclustered Energy (-10%,+10%)	+2.7%	-2.4 %
PDF	+3.0%	-3.0 %
Pile-Up	+0.5%	-0.5 %
MC Statistic	+1.5%	-1.5 %
total	+27.4%	-26.2 %

Table 7.5: Systematic uncertainties from different sources using the baseline selection. The nominal measured cross-section is 160.4 pb.

$t\bar{t}$ MC model: D6T:Z2	0.7%	N/A
$t\bar{t}$ ISR/FSR (smaller,bigger)	-0.1%	-4.6 %
$t\bar{t}$ Q^2 (up,down)	+3.6%	-4.5 %
$t\bar{t}$ Matching-Threshold (higher,lower)	+3.9%	-8.0 %
W+jets Matching-Threshold (higher,lower)	+3.6%	-9.4 %
W+jets Q^2 (up,down)	+11.8%	-6.7 %
Z+jets Matching-Threshold (higher,lower)	+1.8%	-0.1 %
Z+jets Q^2 (up,down)	+2.1%	-1.9 %
QCD (+50%,-50%)	+0.1%	-0.1%
Single top and WW (+30%,-30%)	+1.7%	-1.6 %
JES (down,up)	+18.0%	-11.8 %
JER (+10%,-10%)	+2.6%	-0.1 %
Unclustered Energy (-10%,+10%)	+1.6%	-1.4 %
PDF	+3.0%	-3.0 %
Pile-Up	+0.5%	-0.5 %
MC Statistic	+2.3%	-2.3 %
total	+23.2%	-20.0 %

Table 7.6: Systematic uncertainties from different sources using topological cut1. The measured cross-section is 153.3 pb.

$t\bar{t}$ MC model: D6T:Z2	1.3%	N/A
$t\bar{t}$ ISR/FSR (smaller,bigger)	-0.8%	-6.5 %
$t\bar{t}$ Q^2 (up,down)	+3.4%	-4.9 %
$t\bar{t}$ Matching-Threshold (higher,lower)	+4.6%	-9.2 %
W+jets Matching-Threshold (higher,lower)	-1.9%	-6.6 %
W+jets Q^2 (up,down)	+6.1%	-4.4 %
Z+jets Matching-Threshold (higher,lower)	+1.1%	+0.4 %
Z+jets Q^2 (up,down)	+1.7%	-2.2 %
QCD (+50%,-50%)	+0.1%	+0.4 %
Single top and WW (+30%,-30%)	+1.3%	-1.3 %
JES (down,up)	+15.3%	-9.7 %
JER (+10%,-10%)	+3.6%	-5.5 %
Unclustered Energy (-10%,+10%)	+5.3%	-4.4 %
PDF	+3.0%	-3.0 %
Pile-Up	+0.5%	-0.5 %
MC Statistic	+3.6%	-3.6 %
total	+19.5%	-19.7 %

Table 7.7: Systematic uncertainties from different sources using topological cut2. The measured cross-section is 156.5 pb.

7.3.1 Jet Energy Scale

The jet energy scale (JES) is one of the major systematic uncertainties in this study. It is sensitive to the jet multiplicity which is directly related to the event selection and background estimation as well as other kinematic variables. Figure 7.4 shows the changes of jet multiplicity and M_T distribution in 2-jet events with the jet energy shifted up and down with respect to the nominal JES value. The uncertainty of JES is a function of jet parameters. It has been studied and provided by the CMS Jet/MET group [27]. The total uncertainty includes following consideration:

1. A $p_{\mathrm{T}} - \eta$ dependent uncertainties retrieved from database.

2. Pile-up uncertainty is taken into account. $E_{PU} \times JA \times AvgPU/p_{\mathrm{T}}$ where E_{PU} is the pedestal from pile-up effect. Without appling L1 (Pile-up) correction to jet energy correction, E_{PU} is 0.75 GeV. JA is jet area of 0.8 for AK5 cone jet, $AvgPU$ is the average pile-up for the data range. 1.2 for Run 2010A and 2.2 for Run 2010B and p_{T} is the jet p_{T}.

3. For those jets identified as b-jets in the MC, a b-jet energy scale uncertainty is added. The b-jets with $50 < p_{\mathrm{T}} < 200$ GeV and $|\eta| < 2.0$ have a 2 $\%$ uncertainty; all others have a 3$\%$ uncertainty.

4. The total uncertainty from the above considerations is added in quadrature.

The study is done by varing the scale up and down on the MC samples which results in a different number of 4-jet events and 2-jet events. The consequences are that the difference of events propagate to 4-jet:2-jet ratio and selection efficiency as seen in table 7.8. These different 4-jet:2-jet ratios and selection efficinecies are then applied to the data and the resulting cross-section differences are used as the systematic error from JES.

	JES scale down	nominal JES	JES scale up
Ratio 4-jet:2-jet	0.0436 ± 0.0008	0.0468 ± 0.0008	0.0503 ± 0.0007
signal Efficiency	0.1726 ± 0.0101	0.1872 ± 0.0104	0.2010 ± 0.0107

Table 7.8: Changes of ratio 4-jet:2-jet and signal efficiency

7.3.2 Jet Energy Resolution

Jet energy resolution (JER) is measured by a data-driven method from di-jet events. Events are selected with the criteria where two leading jets are back-to-back in ϕ and the third jet doesn't exceed a maximun p_{T} cut. The resolution is extracted from the asymmetry variable A

$$A = \frac{{p_{\mathrm{T}}}^{jet1} - {p_{\mathrm{T}}}^{jet2}}{{p_{\mathrm{T}}}^{jet1} + {p_{\mathrm{T}}}^{jet2}} \tag{7.8}$$

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = \sqrt{2}\sigma_A \tag{7.9}$$

with assumption of $p_{\mathrm{T}} = {p_{\mathrm{T}}}^{jet1} = {p_{\mathrm{T}}}^{jet2}$ and $\sigma_{p_{\mathrm{T}}} = \sigma_{p_{\mathrm{T}}}^{jet1} = \sigma_{p_{\mathrm{T}}}^{jet2}$. According to the studies from Jet/MET group [2], the jet energy resolution is about 10% worse in data compared to MC. Thus, a +10% bias correction is imposed in MC samples for the central value. An additonal 10% upon the bias correction and the nominal value are considered to be the 10% fluctuation. The results show that JER only contributes a minor effect to the measurements because it only contributes small deviation in jet multiplicity. (Fig. 7.5)

7.3.3 Unclustered Energy

Because $\not\!\!E_T$ is the measurement of the inbalance energy and momentum in the transverse plane, it is also sensitive to the JES or JER. Consequently, the variations from JES and JER are also propagated to $\not\!\!E_T$ and taken into account in the above studies. Therefore, an independent source of uncertianty from unclustered energy has to be considered and it has been done by varying all clustered energy with $\pm$ 10% of its total vectorial sum. In our studies, it changes the $\not\!\!E_T$ spectrum and thus shifts the M_T distribution(Fig. 7.6). Thus, it causes large deviations of in the Data-MC normalization(Table 7.9). Although it alters the relative fraction of QCD/V+jets componenet, the ratio 4-jet:2-jet partially cancels out the effect and the effect from QCD component is also relatively small. Therefore, unclustered energy is a minor systematic effect.

	scale down	no scale	scale up
QCD	2.249	1.761	1.411
V+jets	0.856	0.923	1.037
QCD/J+jets	2.627	1.908	1.361
Ratio4:2	0.0457	0.0468	0.0478

Table 7.9: Changes of Data-MC normalization in 2-jet bin with variations of total clustered energy. The relative fraction of QCD/V+jets shows the significant changes. However, ratio4:2 is more stable

7.3.4 Factorization Scale and Radiation

There are serveral crucial uncertainties from the modeling of the MC generator. Since the strong coupling constant, α_S, is a function of the factorization scale, Q^2 as described in equation 7.10, the setup of the scale in MC generator will affect the coupling and jet multiplicity will be altered consequently. The impact has been studied by scaling the Q^2 to 0.5 and 2 with respect to the current MC setup.

$$\alpha_s(Q^2) = \frac{12\pi}{(33 - 2n_f)\log(Q^2/\Lambda^2)} \tag{7.10}$$

The results from table 7.5 shows that the variation of Q^2 of W production is the largest uncertainty. The scaling up Q^2 results in $\sim 19\%$ uncertainty and scaling down Q^2 gives $\sim 15\%$ without any topological constraint. This is due to the change of the kinematics of W production which can be seen from Figures 7.7. After the topological constraints are

applied, it clearly shows smaller uncertainty (12% and 6%, see table 7.6 and table 7.7). Z+jets events is similar to W+jets events but has smaller cross section. For the signal, $t\bar{t}$ events, the main source of jet multiplicity is from weak decay of top quark. The influence is major on the higher jet multiplicity($\geq$ 5 jets) where the extra jet is from the gluon radiation. Similar test is performed by increasing or decreasing the initial and final state radiation on $t\bar{t}$ events.

7.3.5 Matching Threshold

As mention in chapter 4.2, ME-PS matching is applied in order to avoid the double counting of the simulation at the overlapping phase space between fixed order calculation and soft, collinear simulation. Thus, the matching threshold, which is the cut to separate two regions can effects the population of jet multiplicity. The variation were done by double and half the threshould. The impact is minor as can be seen in Figure 7.9.

7.3.6 Lepton and Jet Selection

Another source of systematic uncertainty is from the muon selection. After scanning the P_{T} threshold from 16 GeV/c to 24 GeV/c with 2 GeV/c increment, we found no significant systematic uncertainty from muon P_{T} threshold. The muon isolation is also examined by changing the isolation values from 0.05 to 0.15 with 0.025 increment. We observe no

significant dependancy in either case. Figure 7.10 shows the relative statistic uncertainty with respect to nominal setup.

$$\Delta\sigma = \frac{\sqrt{\Delta N_{obs}^2 + \Delta N_{bg}^2}}{\int L \times A \times \varepsilon_0} \tag{7.11}$$

where ΔN_{obs} and ΔN_{bg} are the difference of observed 4-jet event and predicted 4-jet background between varied cut and nominal cut. The acceptance A and efficiency ε_0 are the ones from the nominal cuts. The resulting cross-section difference $\Delta\sigma$ reflects the statistic significant for the difference of the selected events. In figure 7.10, it's clear that the cross-sections from different cuts(muon p_{T} and relIso) are slightly changed however they are within the relative statistical uncertainties. It implies that no dependence between the muon selection and the final cross section measurement. Similar test is also applied to p_{T} cut of jet selection. Figure 7.11 shows large uncertainties when jet p_{T} cut is changed. This is not only due to the change of the selected event but also the efficiency of $t\bar{t}$. Nevertheless, topological constraints stablize the efficiency and prove the independency between jet p_{T} cut and cross section.

7.3.7 Parton Distribution Function, Luminosity and Pile-Up

The production of $t\bar{t}$ and its major background, W+jets are mainly produced at low x at the LHC, With the implication that the main production mechanism is from gluons. As already

discussed in chapter 4.2, the gluon PDF distribution has the largest uncertainty. Therefore, it effects the cross-section of signal ($t\bar{t}$) and backgrounds processes. Since CTEQ[51] is used in CMSSW, a large matrix(22$\times$22) with dimension equal to the number of free parameters in the fit has to be diagonalized. The result is 22 orthonormal eigenvectors which gives 44 PDF errors from the plus and minus direction for event re-weighting. Besides, LHAPDF[38] provides the interface to access PDF information from a grid of x and Q^2 and is also used in CMSSW. Therefore, the re-weighting values are given from LHAPDF package. Other systematic such as luminosity[28] is 4$\%$ given from the official luminosity measurement. Pile-up events are also evaluated by using a special generated samples. These uncertainty are done by other groups [29][66] and their result are quoted in the table7.10 below.

Sytematic Type	$\%$ Error
PDF	$\pm$ 3
Pile-Up	$\pm$ 0.5
Luminosity	$\pm$ 4

Table 7.10: Systematic uncertainties from PDF, Luminosity and Pile-up events

7.4 Conclusion

Using 36.1 pb^{-1} data collected in 2010 at the LHC , the $t\bar{t}$ cross-section in 7 TeV proton-proton collision from our baseline selection is determined to be

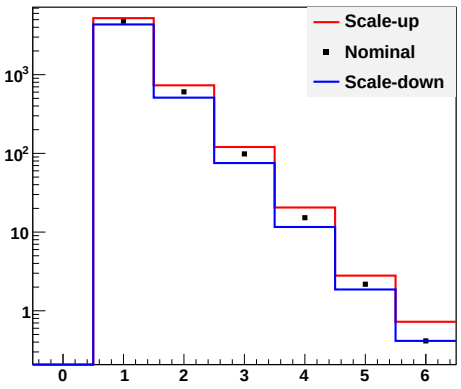
$$160.4 \pm 15.0(\text{stat}) \, ^{+44.0}_{-42.0}(\text{syst}) \, \pm 6.4(\text{lumi}) \text{ pb}.$$

By applying the topological constraints(cut1 and cut2), we obtain results of

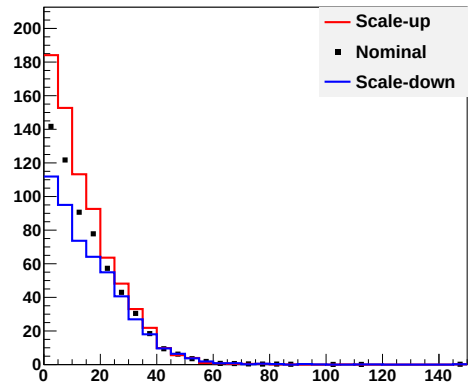
$$153.3 \pm 17.3(\text{stat}) \, ^{+35.6}_{-30.7}(\text{syst}) \, \pm 6.1(\text{lumi}) \text{ pb}.$$

$$156.5 \pm 23.7(\text{stat}) \, ^{+30.5}_{-30.8}(\text{syst}) \, \pm 6.3(\text{lumi}) \text{ pb}.$$

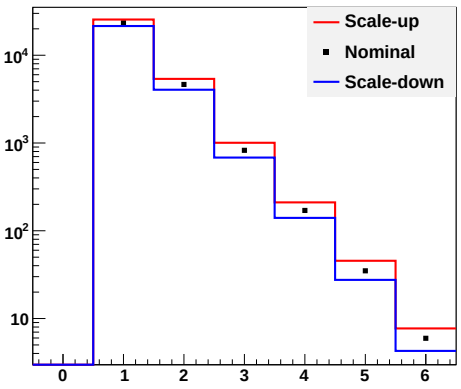
where the systematic uncertainty has been reduced after applying topological cuts. The results are all well consistent with the Standard Model prediction($157.5 \, ^{+23.2}_{-24.4}$ pb^{-1}).



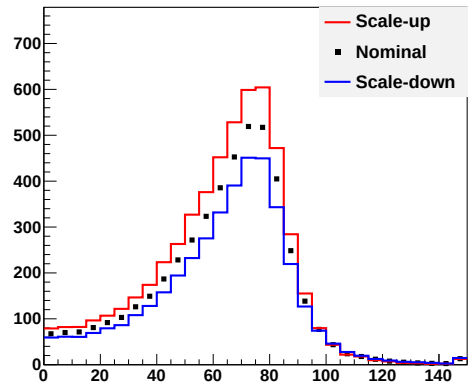
QCD Njets distribution



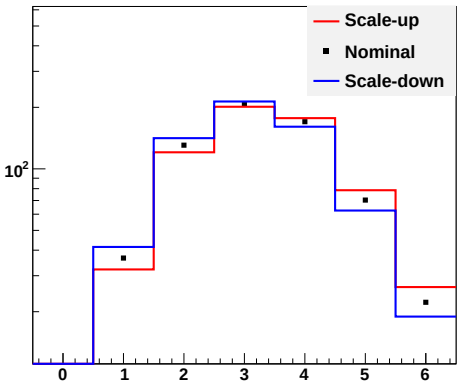
QCD M_T distribution in 2-jet event



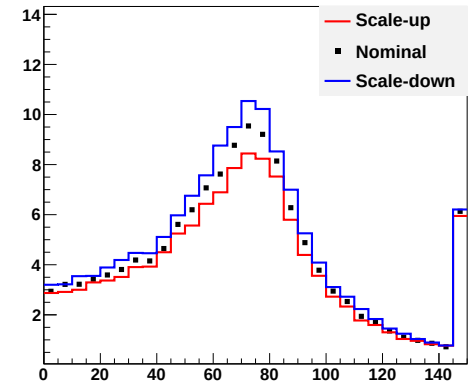
Wjets Njets distribution



Wjets M_T distribution in 2-jet event

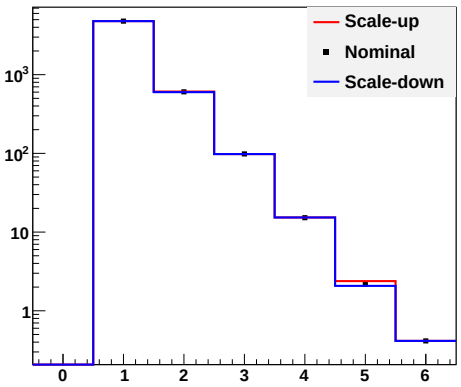


$t\bar{t}$ Njets distribution

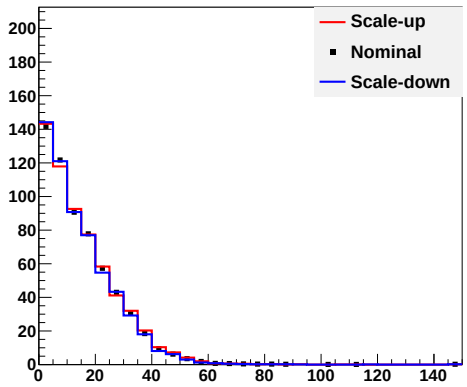


$t\bar{t}$ M_T distribution in 2-jet event

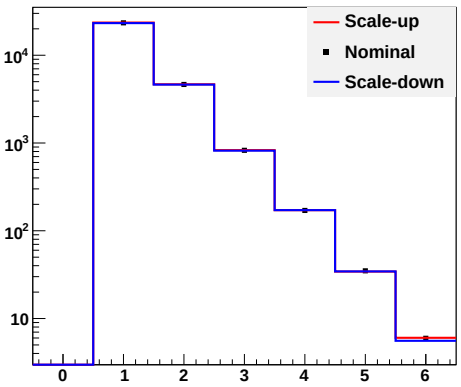
Figure 7.4: Njets and M_T distribution with JES scaled up, scaled down and nominal JES for $t\bar{t}$, W+jets and QCD MC samples.



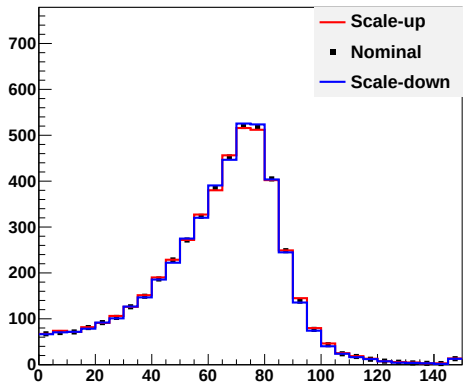
QCD Njets distribution



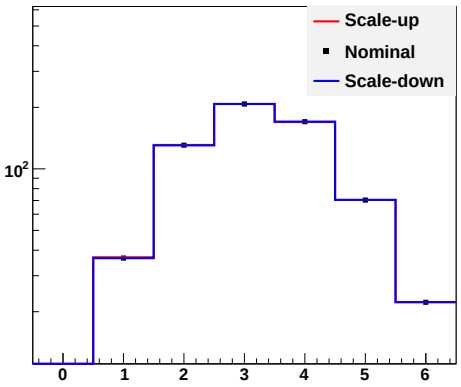
QCD M_T distribution in 2-jet event



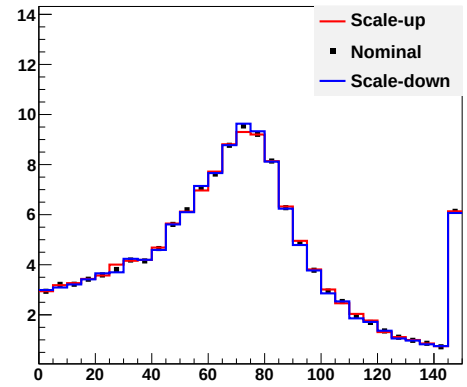
QCD Njets distribution



Wjets M_T distribution in 2-jet event

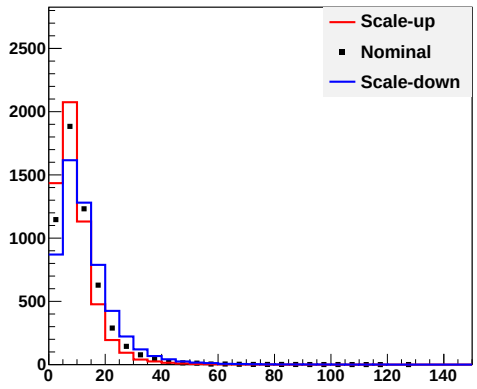


$t\bar{t}$ Njets distribution

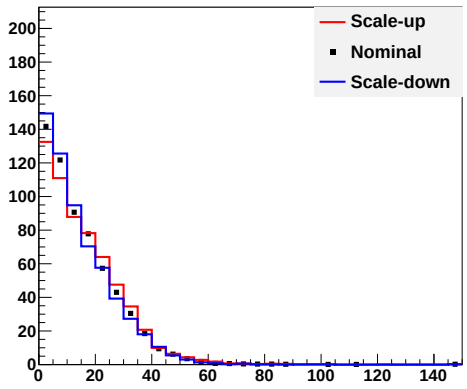


$t\bar{t}$ M_T distribution in 2-jet event

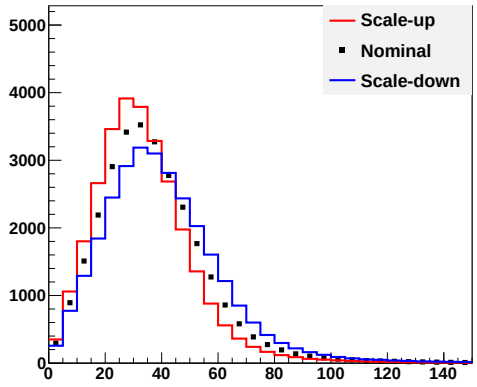
Figure 7.5: Njets and M_T distribution with JER scaled up, scaled down and nominal JER for $t\bar{t}$, W+jets and QCD MC samples. All of them are normalized to the same amount(1500)



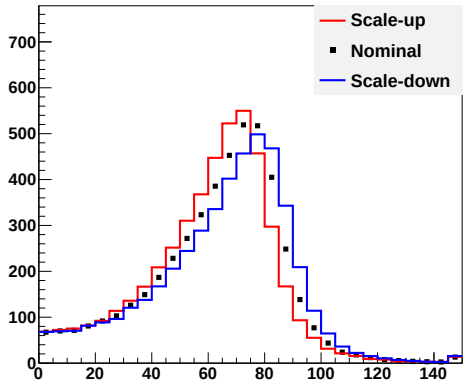
QCD $\not\!\!E_T$ distribution



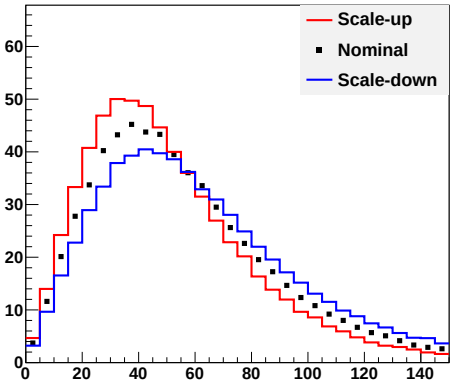
QCD M_T distribution in 2-jet event



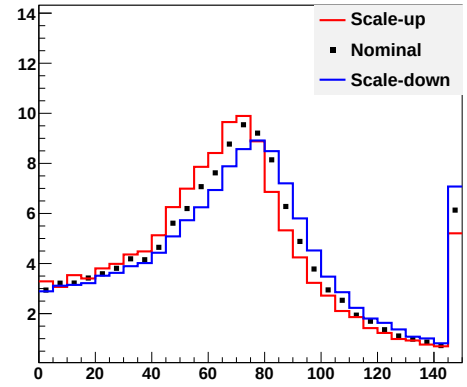
QCD $\not\!\!E_T$ distribution



Wjets M_T distribution in 2-jet event

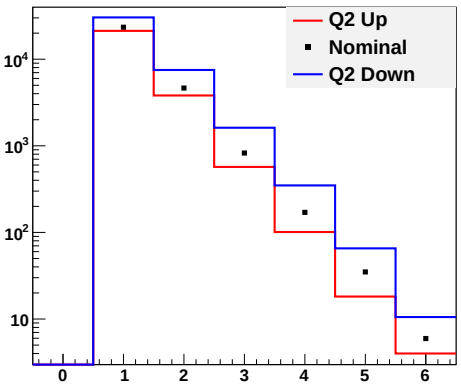


$t\bar{t}$ $\not\!\!E_T$ distribution

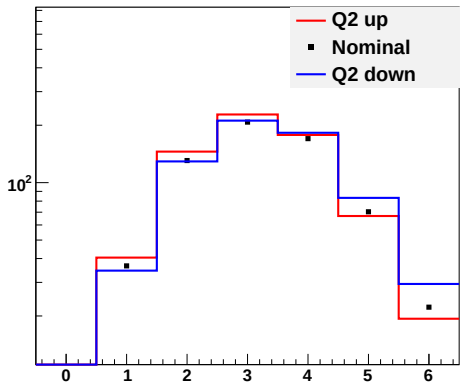


$t\bar{t}$ M_T distribution in 2-jet event

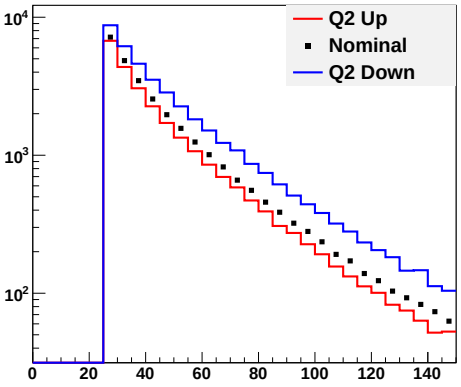
Figure 7.6: $\not\!\!E_T$ and M_T distribution with unclustered energy scaled up, scaled down and nominal one for $t\bar{t}$, W+jets and QCD MC samples.



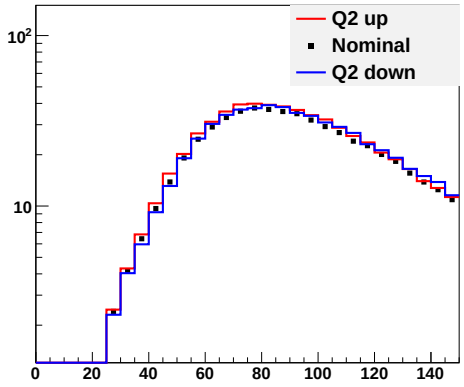
Njets distribution of W+jets events



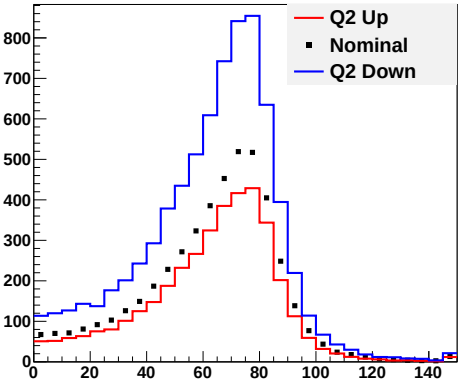
Njets distribution of $t\bar{t}$ events



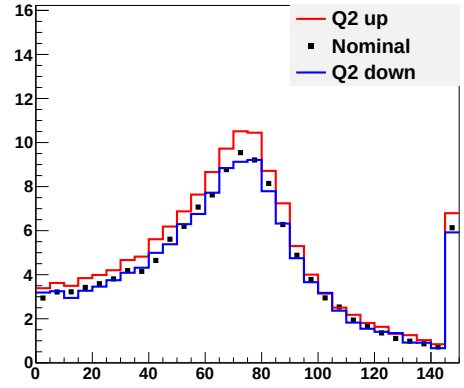
Leading jet p_{T} of W+jets events



Leading jet p_{T} of $t\bar{t}$ events

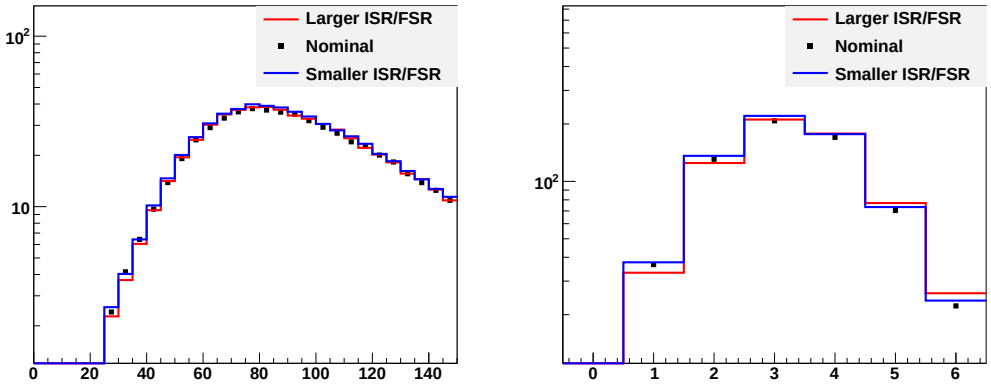


M_T distribution of W+jets event



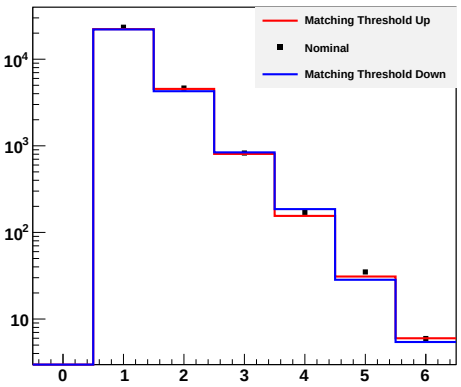
M_T distribution of $t\bar{t}$ events

Figure 7.7: Njets, leading jet p_{T} and M_T distribution with the variation of Q^2 in W+jets and $t\bar{t}$ events

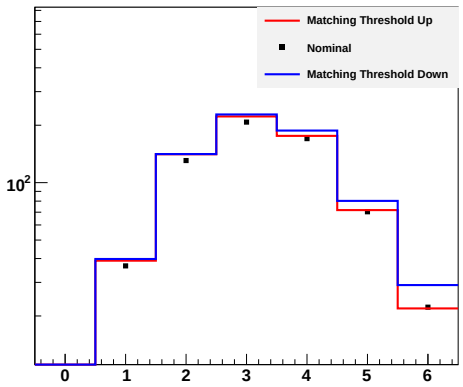


Leading jet p_{T} of W+jets events Njets distribution of $t\bar{t}$ events

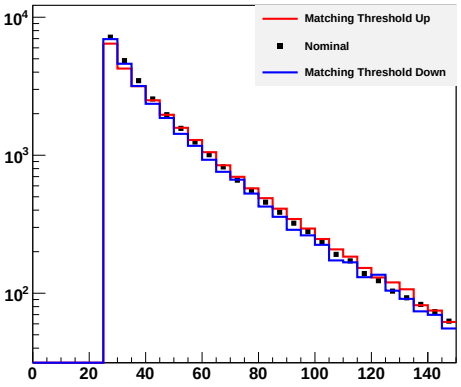
Figure 7.8: Njets and leading jet p_{T} distribution with larger or smaller ISR/FSR in $t\bar{t}$ events



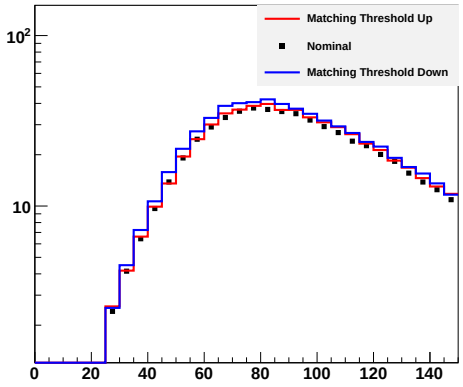
Njets distribution of W+jets events



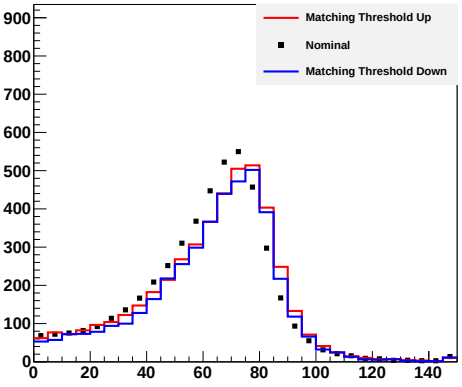
Njets distribution of $t\bar{t}$ events



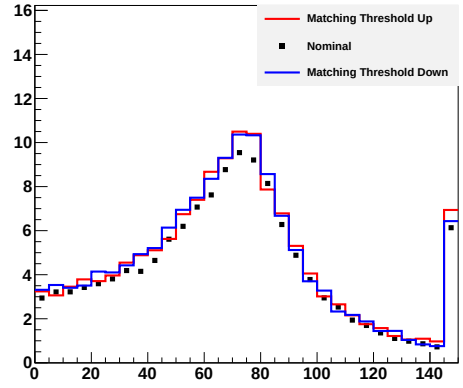
Leading jet p_{T} of W+jets events



Leading jet p_{T} of $t\bar{t}$ events

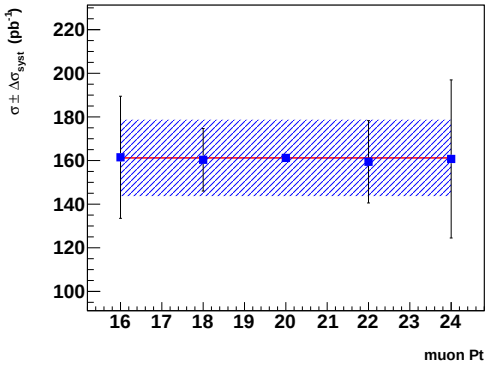


M_T distribution of W+jets event

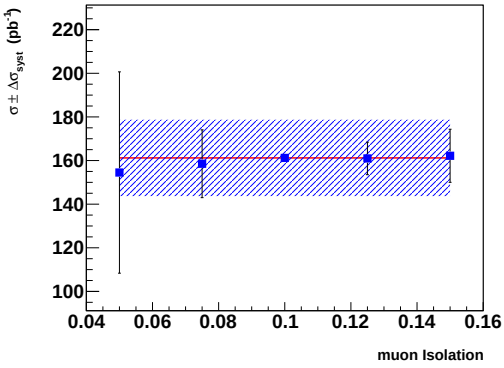


M_T distribution of $t\bar{t}$ events

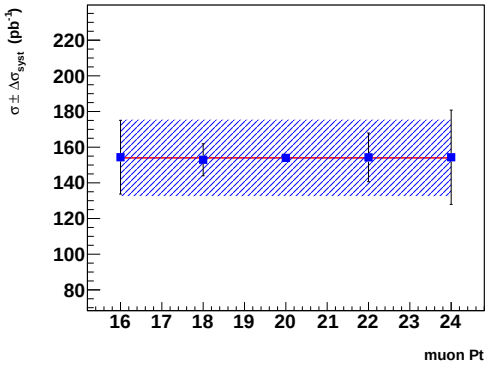
Figure 7.9: Njets, leading jet p_{T} and M_T distribution with the variation of matching threshold in W+jets and $t\bar{t}$ events



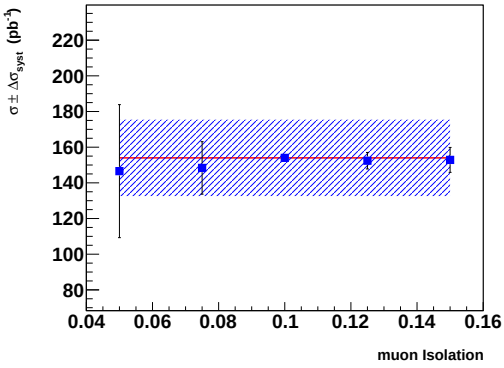
muon P_{T}, cut0



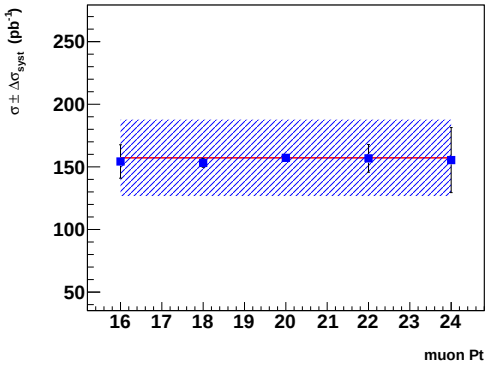
muon isolation, cut0



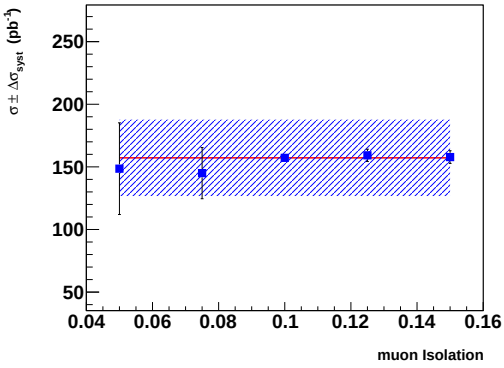
muon P_{T}, cut1



muon isolation, cut1

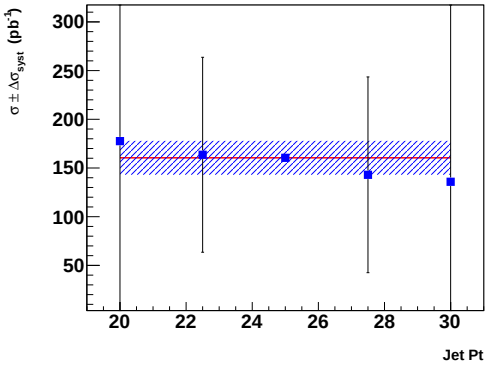


muon P_{T}, cut2

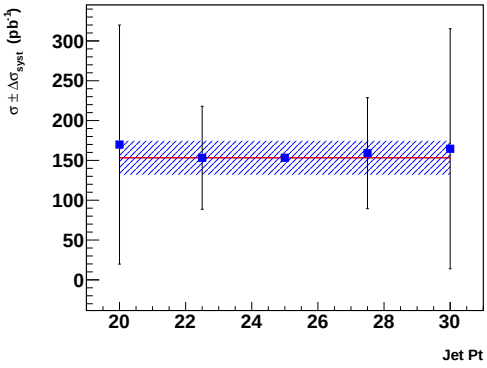


muon isolation, cut2

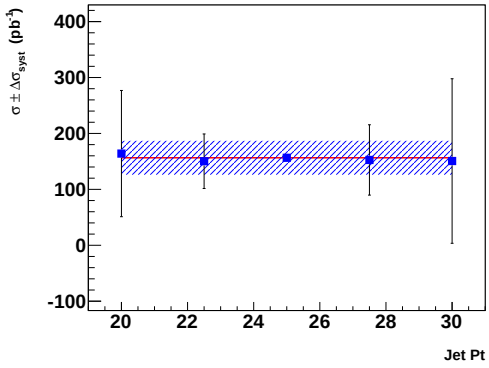
Figure 7.10: Relative statistic uncertainty with respect to muon P_{T} 20 GeV/c and muon isolation 0.1



Jet P_{T}, baseline selection



Jet P_{T}, topological cut1



Jet P_{T}, topological cut1

Figure 7.11: Relative statistic uncertainty with respect to jet P_{T} 25 GeV/c.

Bibliography

[1] CMSSW Fast Simulation Offline Guide. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideFastSimulation`.

[2] A. Smoron and L. Apanasevich and N. Varelas. Data Driven Method for Obtaining the Jet p_{T} Resoultion at CMS. *CMS Analysis Note*, CMS-AN-2009-061, 2009.

[3] N. Adam et al. Measurement of the Inelastic Cross Section using the Luminosity system. *CMS Internal Note*, CMS-IN-2007-046, 2007.

[4] N. Adam et al. The CMS Luminosity system. *CMS Internal Note*, CMS-IN-2007-030, 2007.

[5] Agostinelli et al. (GEANT4 Collaboration). Geant4: A simulation toolkit. *Nucl. Instrum. Methods Phys. Res., A*, 506(CERN-IT-2002-003. SLAC-PUB-9350. 3):250–303. 86 p, Jul 2002.

[6] G. Altarelli and G. Parisi. Asymptotic freedom in parton language. *Nucl. Phys. B*, 126(2):298, 1977.

[7] G. Altarelli and G. Parisi. The top width. *Nucl. Phys. B*, 358(1):46, 1991.

[8] Alwall et al. Madgraph/madevent v4: The new web generation. *J. High Energy Phys.*, 09(arXiv:0706.2334):028. 38 p, Jun 2007.

[9] B. Andersson. *The Lund Model*. Cambridge monographs on particle physics, nuclear physics, and cosmology. Cambridge Univ. Press, Cambridge, 1998.

[10] C. Arnison et al. Hadronic Jet Production at the CERN Proton - anti-Proton Collider. *Phys. Lett.*, B132:214, 1983.

[11] S. Beauceron et al. Selection of $t\bar{t}$ Candidates in Muons+Jets Channel. *CMS Analysis Note*, CMS-AN-2010-173, 2010.

[12] F. A. Berends et al. On the Production of a W and Jets at Hadron Colliders. *Nucl. Phys. B*, 357:32, 1991.

[13] R. K. Bock et al. Data analysis techniques for high-energy physics experiments. *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.*, 11:1–434, 2000.

[14] C. Bromberg et al. Observation of the Production of Jets of Particles at High Transverse Momentum and Comparison with Inclusive Single Particle Reactions. *Phys. Rev. Lett.*, 38:1447, 1977.

[15] C. Albajar et al (UA1 Collaboration) . Search for new heavy quarks in proton-antiproton collisions at $\sqrt{s}$ = 0.63 TeV. oai:cds.cern.ch:208048. *Z. Phys. C*, 48(CERN-EP-90-57):1–12. 24 p, May 1990.

[16] C. Albajar et al. (UA1 Collaboration). Events with Large Missing Transverse Energy at the CERN Collider: Search for Decays of W+- Into Heavy Leptons and of Z0 Into Noninteracting Particles. 2. *Phys. Lett.*, B185:241–248, 1987.

[17] C. Albajar et al. (UA1 Collabration). Direct measurement of w boson width. *Phys. Rev. Lett.*, 74(3):341, 1995.

[18] Cacciari et al. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[19] J. M. Campbell, J. W. Huston, and W. J. Stirling. Hard interactions of quarks and gluons: a primer for lhc physics. *Rep. Prog. Phys.*, 70(hep-ph/0611148):89. 118 p, Nov 2006.

[20] S. Catani et al. Longitudinally invariant K_t clustering algorithms for hadron hadron collisions. *Nucl. Phys.*, B406:187–224, 1993.

[21] S. Catani et al. QCD matrix elements + parton showers. *JHEP*, 0111:063, 2001.

[22] The CMS Collaboration. Algorithm for b Jet Identification in CMS. *CMS Physics Analysis Summary*, BTV-09-001, 2009.

[23] The CMS Collaboration. Determination of the Relative Jet Energy Scale at CMS from Dijet Balance. *CMS Physics Analysis Summary*, JME-08-003, 2009.

[24] The CMS Collaboration. Jet energy calibration with photon+jet events. *CMS Physics Analysis Summary*, JME-09-004, 2009.

[25] The CMS Collaboration. Commissioning of b-jet identification with pp collisions at pp collision at $\sqrt{s}$ = 7 TeV. *CMS Physics Analysis Summary*, BTV-10-001, 2010.

[26] The CMS Collaboration. Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV. *CMS Physics Analysis Summary*, PFT-10-002, 2010.

[27] The CMS Collaboration. Determination of the Jet Energy Scale in CMS with pp Collisions at $\sqrt{s}$ = 7 TeV. *CMS Physics Analysis Summary*, JME-10-010, 2010.

[28] The CMS Collaboration. Measurement of CMS Luminosity. *CMS Physics Analysis Summary*, EWK-10-004, 2010.

[29] D. Green et al. (CMS Collaboration). Measurement of the Top quark pair Production Cross Section Fitting a Kinematic Distribution in Muon+Jets+Missing Transverse Energy Events at pp collistions at 7 TeV. *CMS Analysis Note*, CMS-AN-2010-301, 2010.

[30] Dokshitzer et al. Better Jet Clustering Algorithms. *JHEP*, 08:001, 1997.

[31] G. Abbiendi et al. Muon Reconstruction in the CMS Detector . *CMS Analysis Note*, 097, 2008.

[32] Amsler et al. (Particle Data Group). Review of particle physics. *Phys. Lett.*, B667:1–1340, 2008.

[33] C. Albajar et al. (UA1 Collaboration). Events with Large Missing Transverse Energy at the CERN Collider: W $\rightarrow \tau-$neutrino Decay and Test of $\tau - \mu\ e$ Universality at $Q^2 = m(w^2)$. 1. *Phys. Lett.*, B185:233–240, 1987.

[34] F. Abe et al (CDF Collaboration). Lower limit on the top-quark mass from events with two leptons in p$\overline{p}$ collisions at $\sqrt{s}$ = 1.8 tev. *Phys. Rev. Lett.*, 68(ANL-HEP-PR-92-31. CDF-PUB-1527. FERMILAB-PUB-91-280-E):447–451. 18 p, Oct 1991.

[35] F. Abe et al. (CDF Collaboration). Evidence for top quark production in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.8 TeV. *Phys. Rev. Lett.*, 73(2):225, 1994.

[36] F. Abe et al. (CDF Collaboration). Measurement of the ratio $\sigma_W B(W \rightarrow e\nu)$ / $\sigma_Z B(Z^0 \rightarrow e^+e^-)$. *Phys. Rev. Lett.*, 73(2):220, 1994.

[37] R. Fruhwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth.*, A262:444–450, 1987.

[38] W. Giele et al. The QCD / SM working group: Summary report. 2002.

[39] L. Greg and M. Filip. MET Reconstruction, Performance, and Validation. *CMS Analysis Note*, AN-2008-089, 2009.

[40] A. Guido, K. Ronald, and V. Claudio, editors. *Workshop on Z Physics at LEP1*, 1989.

[41] H. Albrecht et al. (ARGUS Collabration). Observation of $B^0 - \bar{B}^0$ mixing. *Phys. Lett. B*, 192(1):245, 1987.

[42] S. Hche et al. Matching parton showers and matrix elements. oai:cds.cern.ch:926597. (hep-ph/0602031), Feb 2006.

[43] F. Jegerlehner. Vector boson parameters:scheme dependence and theoretical uncertainties. *Z. Phys. C.*, 32:425, 1986.

[44] K. Abe et al. (VENUS Collaboration). Measurement of r and search for new quark flavors decaying into multi-jet final states. *Phys. Rev. B*, 234(3):382, 1990.

[45] K. Kodama et al. (DONUT Collaboration). Observation of tau neutrino interactions. *Phys. Lett. B*, 504(hep-ex/0012035. FERMILAB-PUB-2000-335-E):218–224. 12 p, Dec 2000.

[46] A. D. Martin et al. Physical gluons and high e_t jets. *Phys. Lett. B*, 604(hep-ph/0410230. CAVENDISH-HEP-2004-27. DCPT-2004-128. IPPP-2004-64):61–68. 11 p, Oct 2004.

[47] E. Mathieson and J. S. Gordon. Cathode charge distributions in multiwire chambers ii. approximate and empirical formulae. *Nucl. Instrum. Methods*, 227(A):277, 1984.

[48] M. Mozer et al. Z $\rightarrow$ ee production in association with jets in proton-proton collisions at $\sqrt{s}$ = 7 TeV. *CMS Analysis Note*, CMS-AN-2010-342, 2010.

[49] K. Nakamura et al. Review of particle physics, 2010-2011. review of particle properties. *J. Phys. G*, 37(7A):075021, 2010.

[50] O. Adriani et al. (L3 Collaboration). Search for narrow vector resonances in the z mass range. *Phys. Lett. B*, 313:326, 1993.

[51] Pumplin et al.

[52] D. P. Roy and S. Umar. Upper limit on $B_d^0 - \bar{B}_d^0$ mixing as evidence for the existence of the top quark. *Phys. Rev. Lett. B*, 243(3):296, 1990.

[53] S. Abachi et al. (D0 Collaboration). Search for the top quark in $p\bar{p}$ collisions at $\sqrt{s}$ =1.8 TeV. *Phys. Rev. Lett.*, 72(14):2138, 1994.

[54] S. Abachi et al. (D0 Collaboration). Observation of the top quark. *Phys. Rev. Lett.*, 74(14):2632, 1995.

[55] S. Abachi et al. (D0 Collaboration). W and z boson production in $p\bar{p}$ collision at $\sqrt{s}$ = 1.8 TeV. *Phys. Rev. Lett.*, 75(8):1456, 1995.

[56] Salam et al. A practical Seedless Infrared-Safe Cone jet algorithm. *JHEP*, 05:086, 2007.

[57] N. Saoulidou. Particle Flow Jet identification Criteria. *CMS Analysis Note*, CMS-AN-2010-003, 2010.

[58] Sjostrand et al. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

[59] T. Sjstrand. An evaluation of high-p_T physics monte carlos. (LU-TP-88-12. LUTP-88-12):55 p, Jul 1988.

[60] T. Sjstrand. Monte carlo generators. oai:cds.cern.ch:999717. (hep-ph/0611247. CERN-LCGAPP-2006-06):23 p, Nov 2006.

[61] J. Slaunwhite et al. Top Lepton Plus Jets: Electron and Muon Efficiency Measurements for 2010 Dataset. *CMS Analysis Note*, CMS-AN-2010-362, 2010.

[62] J. Smith et al. The Transverse Mass and Width of the W Boson . *Phys. Rev. Lett.*, 50:1738, 1983.

[63] D. Stump et al. Inclusive jet production, parton distributions, and the search for new physics. *J. High Energy Phys.*, 10(hep-ph/0303013. MSU-HEP-2003-03-03):046. 50 p, Mar 2003.

[64] D. R. Stump et al. Uncertainties of predictions from parton distribution functions: 1, the lagrange multiplier method. *Phys. Rev. D*, 65(hep-ph/0101051. CERN-TH-2000-359):014012. 36 p, Jan 2001.

[65] T. Åkesson et al. (UA2 Collaboration). Search for top quark production at the CERN $\overline{p}$p collider. oai:cds.cern.ch:202869. *Z. Phys. C*, 46(CERN-EP-89-152):179–189. 30 p, Nov 1989.

[66] T. Chwalke et al. (CMS Collaboration). Measurement of the ttbar cross section in the electron+jets and the muon+jets channels. *CMS Analysis Note*, CMS-AN-2010-310, 2010.

[67] The LEP Electroweak Working Group The LEP Collaborations, editor. *Schaille D. In XXVII Int. Conf. on High Energy Physics*, 1995.

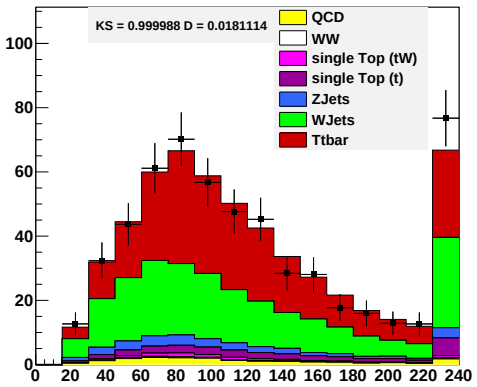
[68] A. Wolfgang et al. Reconstruction of electrons with the gaussian-sum filter in the cms tracker at the lhc. Technical Report CMS-NOTE-2005-001. CERN-CMS-NOTE-2005-001, CERN, Geneva, Jan 2005.

[69] S. L. Wu. e^+e^- interaction at high energies. *Nuclear Physics B (Proc. Suppl.)*, 3:39, 1988.

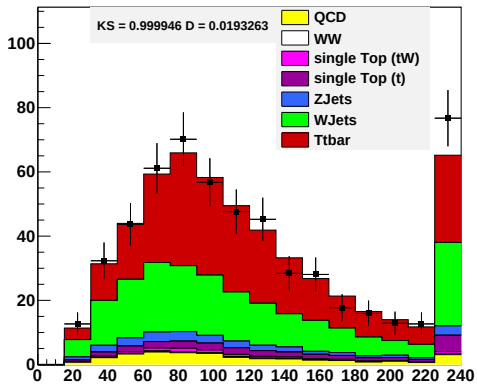
Appendix A

1-D M2M3 Plots

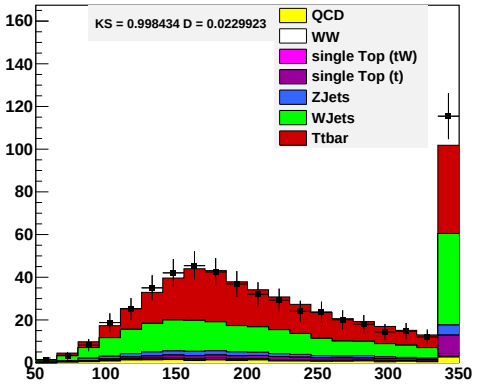
After applying the Data-MC normalization to MC samples, the topological parameters ($M2_h, M3_h and M3_l$) of the events are checked for the consistency with data. The results (see Fig. A.1) shows the agreement between data and normalized MC samples.



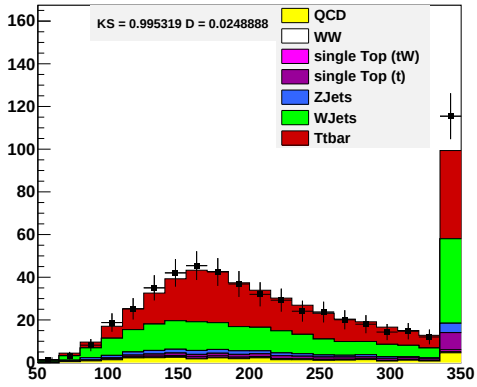
$M2_h$ distribution before normalization



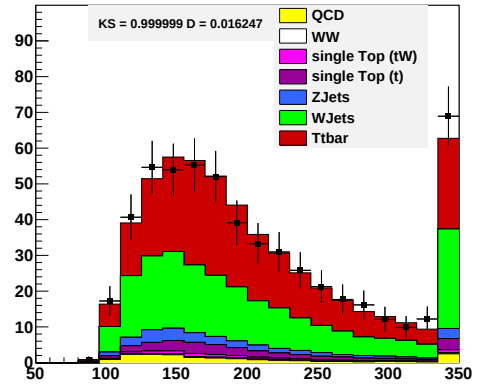
$M2_h$ distribution after normalization



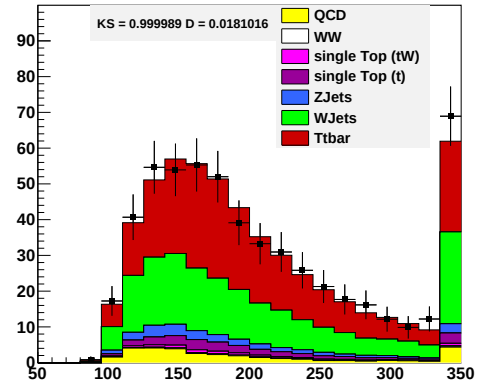
$M3_h$ distribution before normalization



$M3_h$ distribution after normalization



$M3_l$ distribution before normalization



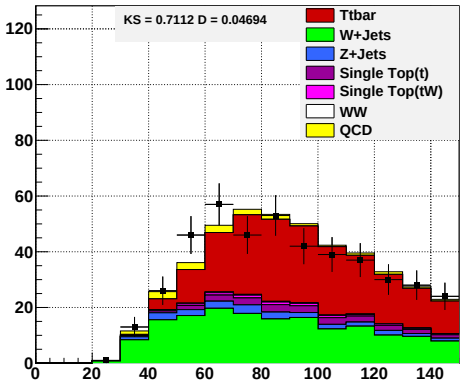
$M3_l$ distribution after normalization

Figure A.1: Mass information before(left) and after(right) normalization from events with ≥ 4 jets. The last bin in the histograms are the overflowed bin.

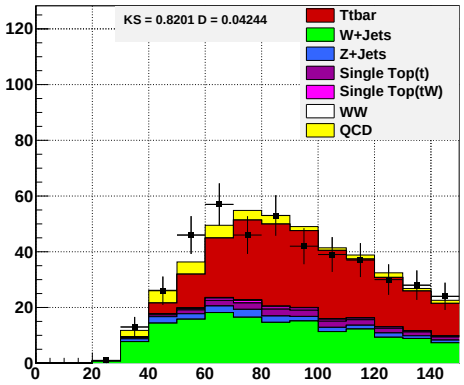
Appendix B

Kinematic Plots With Data-MC Normalization

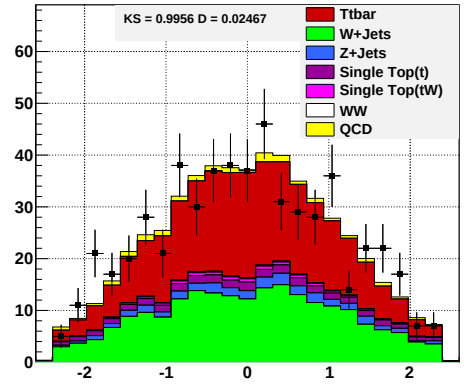
A cross check for kinematic variables is performed after Data-MC normalization since the normalization processes only use the M_{T} distribution in 2-jet events. The comparison before and after normalization in $\geq$ 4-jet events are shown below. The improvement of the agreement between data and MC samples is validated.



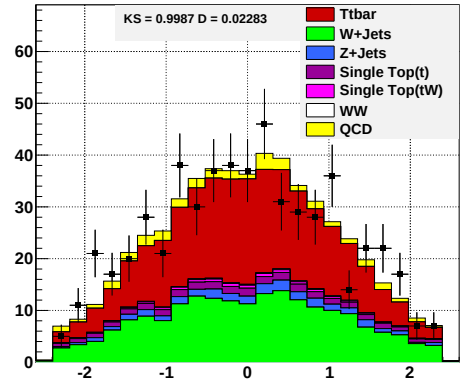
Jet1 P_T before normalization



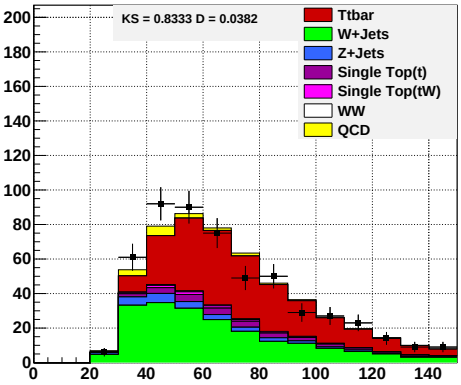
Jet1 P_T after normalization



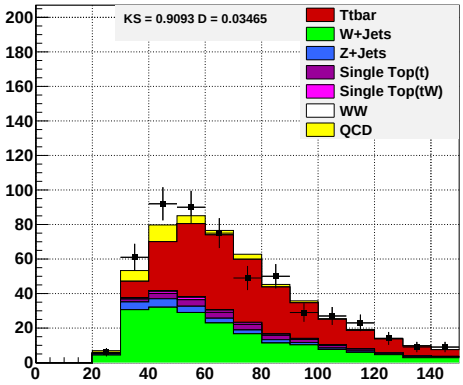
Jet1 η before normalization



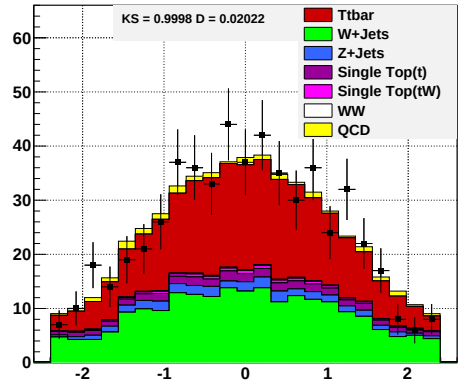
Jet1 η after normalization



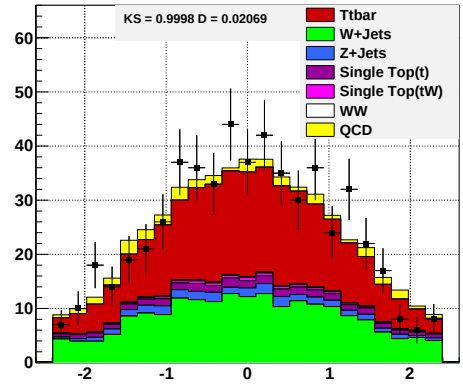
Jet2 P_T before normalization



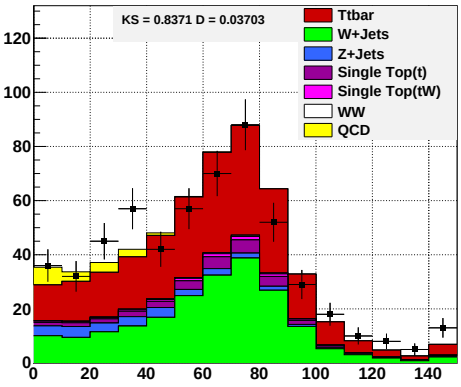
Jet2 P_T after normalization



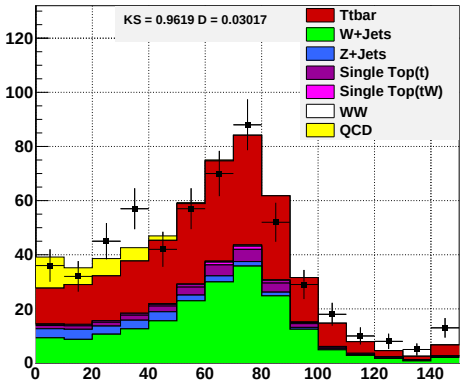
Jet2 η before normalization



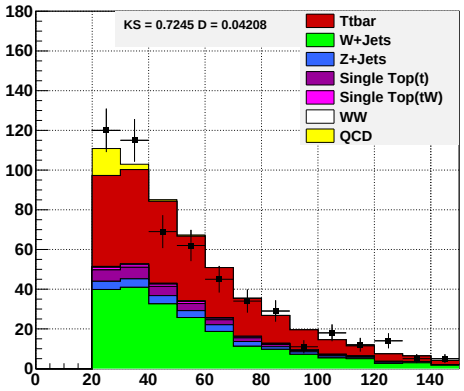
Jet2 η after normalization



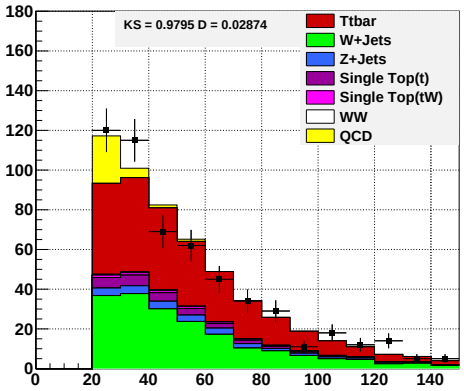
$M_T \geq$ 4Jets before normalization



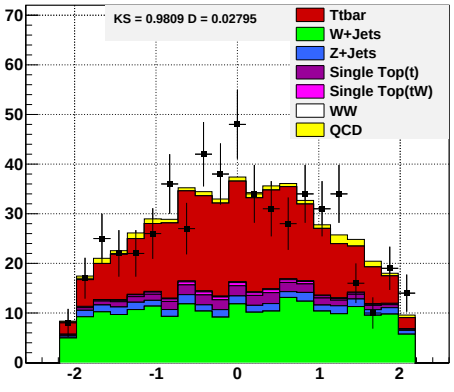
$M_T \geq$ 4Jets after normalization



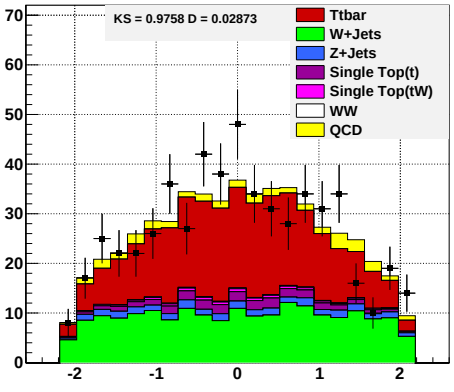
muon P_T before normalization



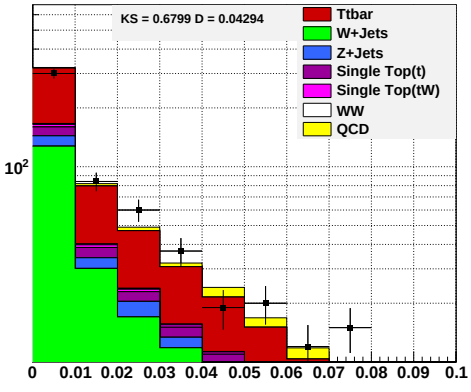
muon P_T after normalization



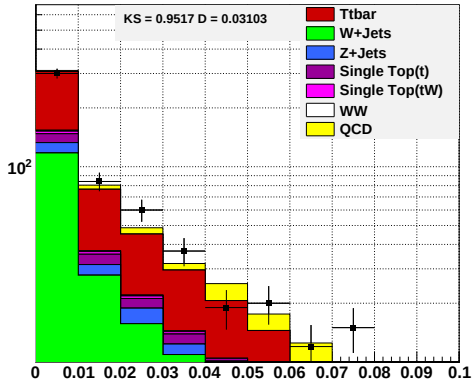
muon η before normalization



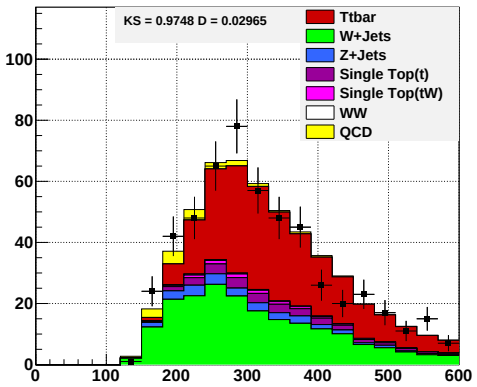
muon η after normalization



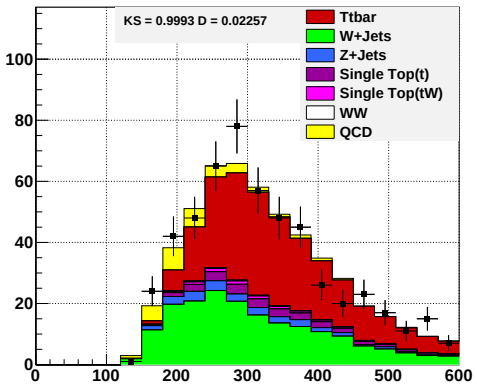
muon iso before normalization



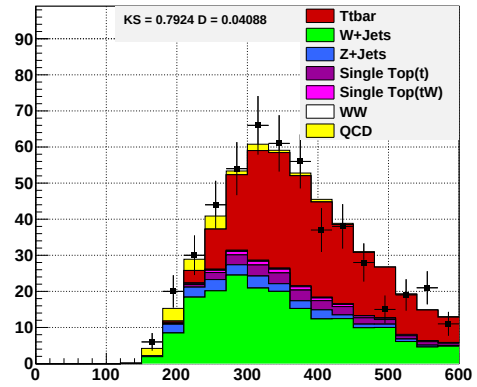
muon iso after normalization



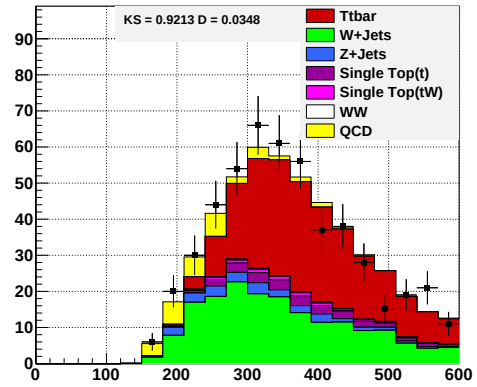
H_T^{lep} before normalization



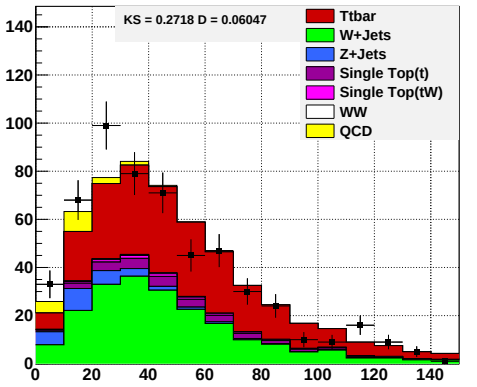
H_T^{lep} after normalization



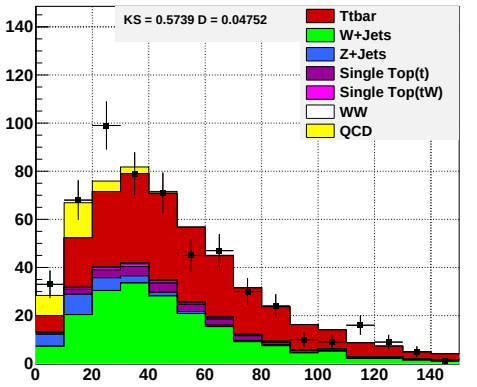
H_T^{tot} before normalization



H_T^{tot} after normalization



MET before normalization



MET after normalization