UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

MEASUREMENT OF THE TOP QUARK MASS IN THE ALL HADRONIC FINAL STATE AT THE DØ EXPERIMENT

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

AYESH JAYASINGHE
Norman, Oklahoma
2013

MEASUREMENT OF THE TOP QUARK MASS IN THE ALL HADRONIC FINAL STATE AT THE DØ EXPERIMENT

A DISSERTATION APPROVED FOR THE
HOMER L. DODGE DEPARTMENT OF PHYSICS AND ASTRONOMY

BY

Dr. Phillip Gutierrez, Chair

Dr. Brad Abbott

Dr. James Hawthorne

Dr. Chung Kao

Dr. Michael Santos

Acknowledgements

This thesis would not have been possible without the guidance and the help of many individuals who contributed and extended their valuable assistance in the preparation and completion of the analysis. First and foremost, my utmost gratitude to Dr. Phil Gutierrez, my supervisor for the thesis, who helped me in numerous ways to bring this to a success. His ample guidance, patience, prompt advises and cooperation that he bequeathed was inspirational. Specially, in my situation, I carried out the research and composed the thesis while residing away from the University of Oklahoma, on which Dr. Gutierrez's flexibility, altruism, encouragement and guidance was an abundant fortitude to make this effort a success. The members of the dissertation committee; Dr. Brad Abbott, Dr. James Hawthorne, Dr. Chung Kao and Dr. Michael Santos for their valuable inputs and immense support. The fruitful and lengthy discussions with Dr. Gianluca Petrillo and Petr Vokac from the DØ collaboration helped in numerous ways to make this analysis a success. And also, my sincere gratitude goes to the colleagues at the DØ collaboration for their valuable comments, guidance and support. Last but not least the motivation, encouragement and patience shown by my family and friends made this journey a pleasant and a memorable one.

Contents

List of Tables

List of Figures

Abstract

The top quark is the heaviest fermion observed to date. A precise measurement of its mass and W boson mass is important to indirect measurements of Higgs boson mass. Furthermore, the top quark mass, W boson mass and Higgs boson mass may test the Standard Model using the correlations between them. Here in this thesis, we present a measurement of the top quark mass in the all hadronic final state using the template method. This final state has the advantage of being fully reconstructed in the detector and having the largest branching fraction. The measurement is performed on 4033 candidate events collected using the DØ detector. The data is collected from $p\overline{p}$ collisions generated at $\sqrt{s}$ =1.96 GeV by the TEVATRON accelerator, Fermi National Accelerator Laboratory, Batavia IL. This is a two dimensional measurement formulated to extract the top quark mass as well as lower the systematic uncertainty due to the jet energy scale calibration. A kinematic fitter is employed to build the templates of signal and background for various input top quark mass points and jet energy scale variations. These templates are compared to data to obtain the fitted top quark mass, jet energy scale shift and their uncertainties. The obtained top quark mass is

$$m_t = 170.4 \pm 1.7 \text{ (stat)} \pm 2.9 \text{ (sys)} \text{ GeV}.$$

Chapter 1

Introduction

1.1 Standard Model

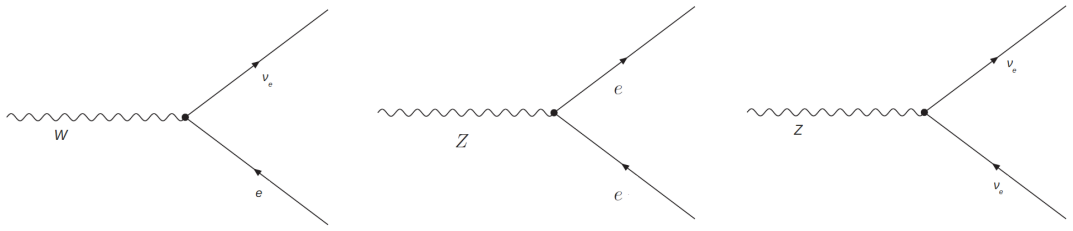
In 1964 Gell-Mann and Zweig independently proposed that hadrons (particles that interact through the strong force [12]) are not elementary particles but are composed of *quarks* [13][14]. To date, six quarks, up, down, charm, strange, top and bottom are experimentally observed; with the top quark being the most recently discovered. It was discovered at the Fermilab TEVATRON in 1995 [15].

Quarks are grouped into three generations according to their mass, charge (Q) and weak isospin (I) (see Table 1.1). All left-handed (direction of spin and

Table 1.1: Mass, charge and weak isospin of quarks [1].

Quark	Mass	Charge	I	I_3
Up	$2.3^{+0.7}_{-0.5}$ MeV	+2/3	1/2	+1/2
Down	$4.8^{+0.7}_{-0.3}$ MeV	-1/3	1/2	-1/2
Charm	$1.27^{+0.07}_{-0.11}$ GeV	+2/3	1/2	+1/2
Strange	95 ± 5 MeV	-1/3	1/2	-1/2
Top	173.2 ± 0.94 GeV	+2/3	1/2	+1/2
Bottom	$4.18+ \pm 0.03$ GeV	-1/3	1/2	-1/2

momentum are opposite to each other) fermions (quarks and leptons) form doublets with weak isospin quantum number 1/2 and $I_3 = \pm 1/2$, where I_3 is the third component of the weak isospin. The value of I_3 is +1/2 for *up type fermions* and -1/2 for *down type fermions*. The top quark (I_3= +1/2) and the bottom quark



(a) (b) (c)

Figure 1.1: (a) Interaction of an electron with an electron neutrino through the exchange of a W boson. (b) Interaction of an electron with an electron through the exchange of a Z boson. (c) Interaction of an electron neutrino with an electron neutrino through the exchange of a Z boson.

(I_3= -1/2) belong to the third generation of quarks and are the heaviest members of the quark family (see Table 1.1).

The Standard Model is a theory that describes the elementary particles and their interactions. The existence of the neutrino was experimentally confirmed in 1956 [16]. In 1962 it was experimentally verified that there are at least two types of neutrinos, one type associated with a muon and the other type associated with an electron [17]. The existence of the third type, a tau neutrino, was not predicted at that time. Furthermore, left handed electrons and left handed electron neutrinos interact through the exchange of a W boson (Fig. 1.1) and their states are represented by [18]

$$\psi_L = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L . \tag{1.1}$$

The electron itself couples to an electron through the exchange of a neutral Z boson and the electron neutrino also couples to itself through the exchange of a Z boson (Fig. 1.1) [19]. Therefore, the Standard Model has grouped these into a doublet (ν_e e)$_L$ [19]. Likewise, all the left-handed leptons and quarks form doublets. The symmetry this group shares is called $SU(2)$[1]. Right-handed fermions

[1]Special Unitary group of degree two represented by group of 2×2 matrices [20].

are singlets under $SU(2)$ and the Standard Model does not accommodate right-handed neutrinos. Interactions between these fermions are mediated by the force carriers, namely the photon, W and Z bosons, and gluons. The electromagnetic interactions are mediated by massless photons, the weak interactions are mediated by the massive W and Z bosons and the strong interaction is mediated by massless gluons. The electromagnetic interaction and the weak interaction are unified in the electroweak theory which is based on the gauge invariance of the group $SU(2)_L \otimes U(1)_Y$, where $U(1)_Y$ is the gauge group of weak hypercharge Y ($Y = 2(Q - I_3)$). The strong force is described by Quantum Chromodynamics (QCD). Symmetry this group share is denoted by the $SU(3)_C$ gauge group.

The masses of the elementary particles vary over a wide range. For example, the neutrinos have negligible masses while the top quark mass is 173.2 GeV. The Higgs mechanism successfully explains the masses of the gauge bosons. The Higgs field

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{1.2}$$

is a complex field with four real components. Three of these generate the masses of the $W^\pm$ and Z bosons and the remaining field is the observable Higgs boson. The elementary particles gain their masses by interacting with the Higgs field. The mass of each particle is determined by the interaction strength. For example, the electron mass is $\lambda_e \langle \phi^0 \rangle$, where λ_e is the electron Yukawa coupling[2] and $\langle \phi^0 \rangle$ is the vacuum expectation value of the Higgs field. Since, the top quark has the largest mass, it has the largest Yukawa coupling.

[2]The Yukawa interaction of the Higgs field couples the fermions to the Higgs field

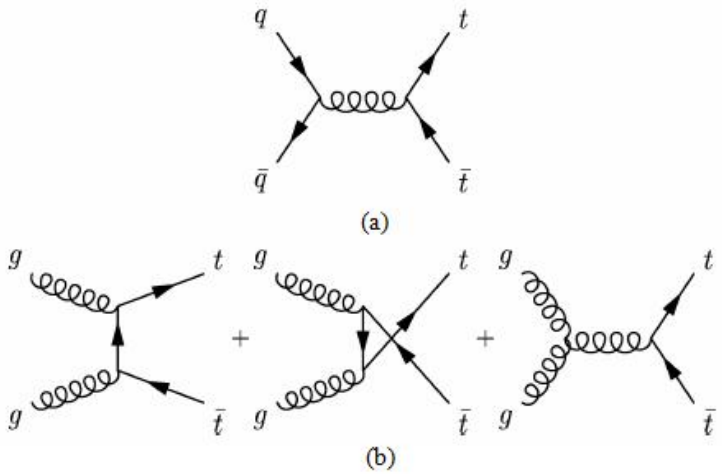


Figure 1.2: Feynman diagrams for $t\bar{t}$ production (a) quark anti-quark annihilation (b) gluon fusion [2].

1.2 Top Quark Production

The top quark can be produced either through the strong interactions as a $t\bar{t}$ pair or through the weak interaction as a single top quarks. At tree level, $t\bar{t}$ production proceeds through gluon fusion or the annihilation of a quark anti-quark pair (Fig. 1.2)[21]. Single top quark production proceeds via the weak interaction through one of three mechanisms depending on the virtuality of the W boson involved in the process [22], namely quark quark scattering involving a W boson with ${Q_W}^2 < 0$, quark anti-quark annihilation with a W boson of ${Q_W}^2 > 0$ and associated Wt production with a real W boson of ${Q_W}^2 = {M_W}^2$ (Fig. 1.3).

The process of quark anti-quark annihilation is the dominant process (85%) of top quark production at the TEVATRON [23]. This is determined by the Parton Distribution Functions (PDF) and center of mass energy of the collider. The fraction of momenta carried by each parton (quark, anti-quark, gluon) in the proton or anti-proton is denoted by x. Thus, the effective center of mass energy $\sqrt{s_{eff}}$ is lower than the actual center of mass energy $\sqrt{s}$ of the collider. They are

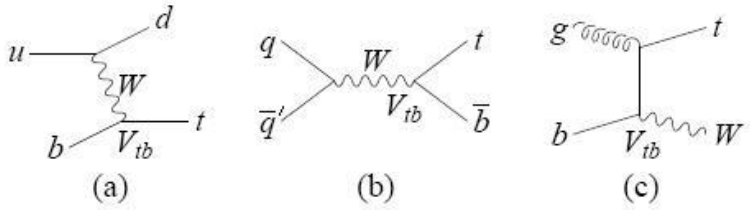


Figure 1.3: Feynman diagrams for single top quark production (a) quark-quark scattering (b) quark anti-quark annihilation (c) associated Wt production [3].

related by

$$s_{eff} = x_i x_j s, \tag{1.3}$$

where x_i (x_j) is the fractional momenta of the i^{th} (j^{th}) parton [2]. The threshold energy to produce a $t\bar{t}$ pair at rest is $2m_t$. From equation 1.3 we have $x_i x_j \geq 4m_t^2/s$. If we make an assumption that each parton in the proton or anti-proton carries the same momentum fraction x then $x \approx 2m_t/\sqrt{s}$. The value of x is 0.18 for the TEVATRON with $\sqrt{s}$= 1.96 TeV. According to the Fig. 1.4, it can be seen that for x= 0.18 the gluon density with threshold energy to produce $t\bar{t}$ pair is lower than that of quarks.

1.3 Top Quark Decay Modes

Since it is heavier than the W boson, the top quark can decay into a lighter quark and a W boson. The branching ratio of a top quark decaying into a W boson and a bottom quark is given by

$$Br(t \to Wb) = \frac{|V_{tb}|^2}{|V_{td}|^2 + |V_{ts}|^2 + |V_{tb}|^2}, \tag{1.4}$$

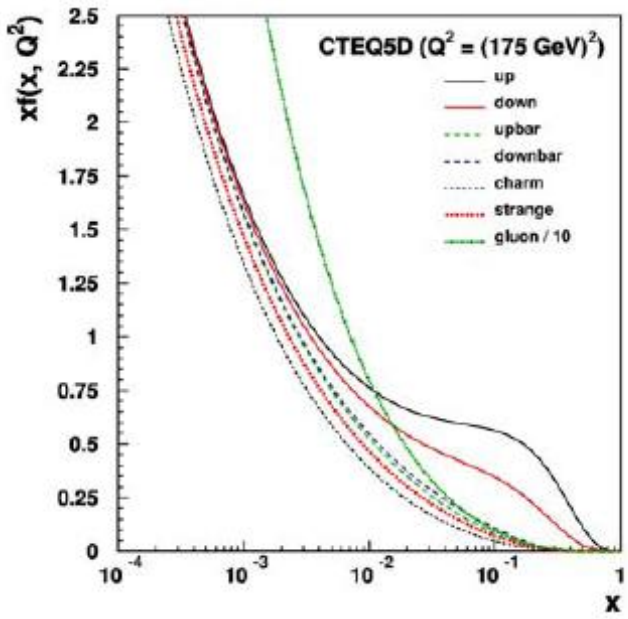


Figure 1.4: Parton momentum densities in the proton as a function of the longitudinal proton momentum fraction for Q^2=(175 GeV)^2 [2].

where the values of the CKM matrix elements[3] $|V_{td}|$, $|V_{ts}|$ and $|V_{tb}|$ are $0.00874^{+0.00026}_{-0.00037}$, 0.0407 ± 0.0010 and $0.999133^{+0.000044}_{-0.000043}$, respectively [1]. Therefore, the value $Br(t \to Wb) \approx 100\%$ guarantees that the top quark almost always decays into a W boson and a bottom quark. Thus, the reconstruction of the signal depends on the decay modes of the W boson.

The final state of the $t\bar{t}$ is classified by three possible modes depending on the decay modes of the W^+ and W^-. Both W bosons can decay into hadronic final state producing two quark anti-quark pairs. This channel is called the *all hadronic final state* (Fig.1.5). The decay mode of one W boson into hadrons and the other one into a charged lepton plus lepton neutrino pair is called the *lepton+jets final state*[4] (Fig. 1.5). The decay of both W bosons into leptonic final state gives rise

[3]Quarks participating in weak interactions are not pure flavor eigenstates but rotated by a mixing angle. The rotated eigenstates and the pure eigenstates are related by the Cabibbo Kobayashi Maskawa (CKM) matrix [18].

[4]For the top quark analysis an electron or a muon is referred as a lepton. Final states with taus are not considered in the lepton+jets analysis.

1. All jets channel

$$\begin{bmatrix} t \to W^+ b \to q_1 \bar{q}_2 b \\ \bar{t} \to W^- \bar{b} \to q_3 \bar{q}_4 \bar{b} \end{bmatrix} \Rightarrow q_1 \bar{q}_2 b q_3 \bar{q}_4 \bar{b}$$

2. Lepton+ jets channel

(a) $$\begin{bmatrix} t \to W^+ b \to q_1 \bar{q}_2 b \\ \bar{t} \to W^- \bar{b} \to l \overline{\nu_l} \bar{b} \end{bmatrix} \Rightarrow q_1 \bar{q}_2 b l \overline{\nu_l} \bar{b}$$

(b) $$\begin{bmatrix} t \to W^+ b \to \bar{l} \nu_l b \\ \bar{t} \to W^- \bar{b} \to q_1 \bar{q}_2 \bar{b} \end{bmatrix} \Rightarrow \bar{l} \nu_l b q_1 \bar{q}_2 \bar{b}$$

3. Dilepton channel

$$\begin{bmatrix} t \to W^+ b \to \bar{l} \nu_l b \\ \bar{t} \to W^- \bar{b} \to l' \overline{\nu_{l'} b} \end{bmatrix} \Rightarrow \bar{l} \nu_l b l' \overline{\nu_{l'} b}$$

Figure 1.5: $t\bar{t}$ decay channels. Here q denotes a quark and l denotes a lepton and ν_l denotes a lepton neutrino.

to a charged anti-lepton plus lepton neutrino pair and a charged lepton plus anti-lepton neutrino pair. Since this channel is associated with two charged leptons, it is called the *dilepton channel*[5] (Fig. 1.5).

The all hadronic final state has the largest branching ratio, $\approx 46\%$ [2] and is characterized by at least six jets, with two of these from bottom quarks. The most significant background for the all jets channel is multi-jet production via the strong interaction and is a few orders of magnitude larger than the signal [2].

The lepton+jets channel is characterized by at least four jets with two of these from bottom quarks, a muon or an electron with a large transverse momentum and large $\not\!\!E_T$ from neutrinos. The lepton+jets channel also has a large branching ratio of $\approx 30\%$ [2].

The dilepton channel has the lowest branching ratio, $\approx 4\%$, and is characterized by two oppositely charged high transverse momentum leptons (a muon or an

[5]For the top quark analysis an electron or a muon is referred as a lepton. Final states with taus are not considered in the dilepton analysis.

electron), large $\not\!\!E_T$ from neutrinos and at least two jets from the hadronization of bottom quarks [2].

The $t\bar{t}$ final states with one or two taus are identified and handled differently than the lepton+jets final state and dilepton final state due to the complexity of the decay modes of the tau. For, a $t\bar{t}$ event with an electron in the final state $t \to W \to \tau\nu_\tau \to e\nu_e$ looks like $t \to W \to e\nu_e$ hence is very difficult to distinguish from the electron+jets final state. Therefore, the recent measurements on this channel are performed only on the final states where the tau lepton decays into hadrons [24] [25].

1.4 Top Quark Mass

The top quark mass is a substantial parameter in the standard model. It is approximately 40 times heavier than the next heaviest fermion the bottom quark and is of the order of the vacuum expectation value of the Higgs field. Furthermore, the top quark mass as well as the Higgs boson mass contribute to the one-loop quantum mechanical corrections to the W boson mass (Fig. 1.6) [2]. Therefore, a precise measurement of the above serves as a test to the Standard Model or else measurement of any of the two parameters will lead to an indirect measurement of the third. Figure 1.7 shows the most up to date measurements of the W boson, top quark and Higgs boson masses. The Standard Model prediction for the correlation between the three masses is in accordance with that which is observed within the given uncertainties. Hence, it is important to reduce the uncertainties of the measurements in order to improve the comparison between experiment and the Standard Model.

The top quark mass has been measured to a precision of less than one percent using the data collected at the TEVATRON (Fig. 1.8). For the past few decades the measurement of the top quark mass has been updated using various techniques

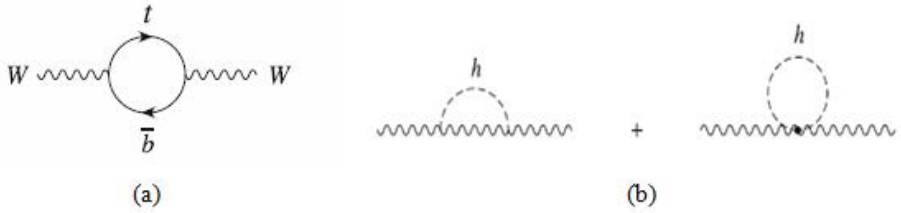


Figure 1.6: Contributions for the W boson mass from (a) virtual top quark loops (b) virtual Higgs boson loops [2].

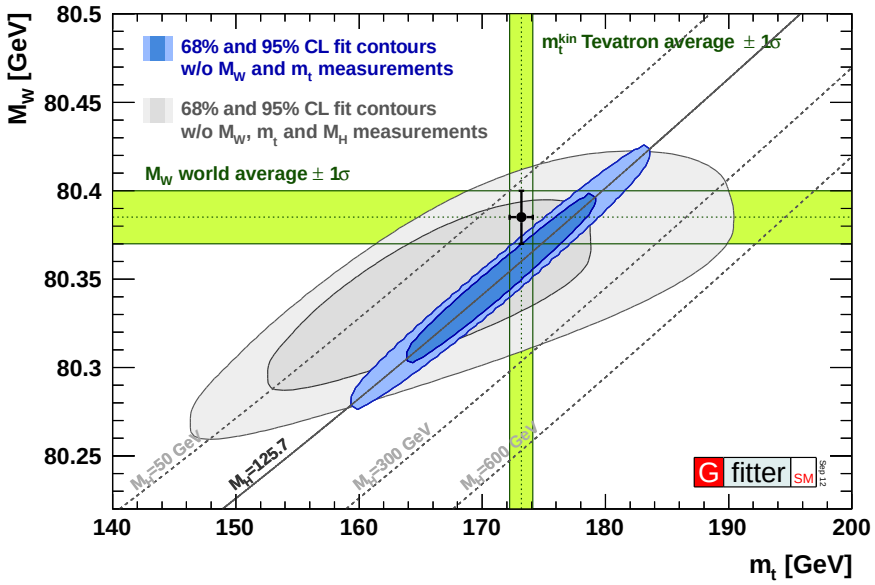


Figure 1.7: The blue shapes are the 68 % and 95 % confidence limit for m_W vs. m_t determined by measurement from LEP and SLD combining with the direct Higgs boson mass measurement. The gray oval is the same without using the direct measurement of the Higgs boson mass.

on different final states. The measurements performed in different final states and using various techniques lead to a better understanding of the systematic uncertainties hence, reduce the uncertainty in the combined result. The statistical uncertainties are reduced with the increase of amount of data collected by the each detector, CDF and DØ[6]. Yet, it is an ongoing effort to further improve the accuracy of the measurement with the advantage of having more data, better understood detectors and enhanced computing facilities such as better data processing and storage capabilities.

[6]CDF and DØ are the two detectors installed at the Fermi National Accelerator Laboratory TEVATRON.

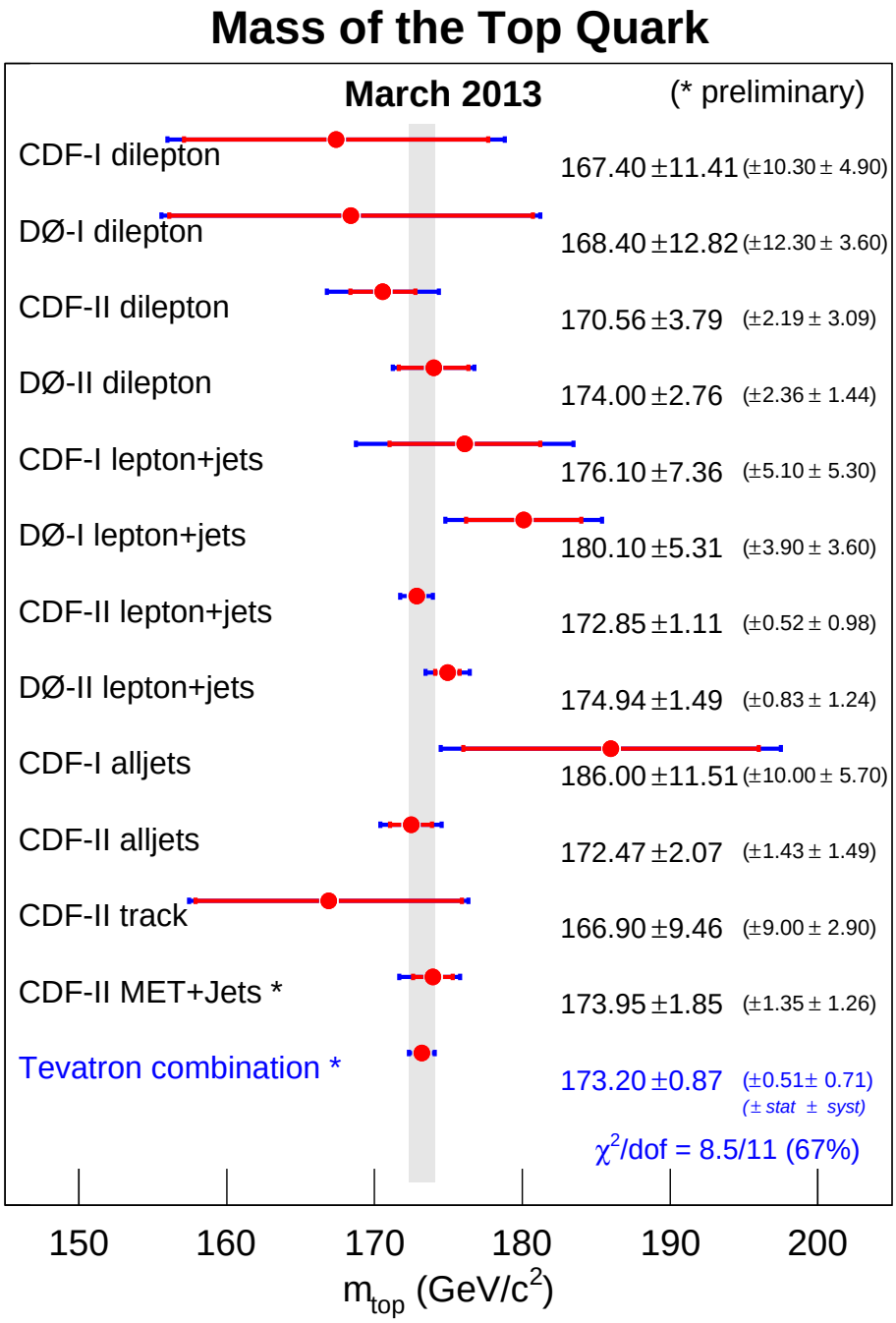


Figure 1.8: The top quark mass measurements made on various final states using the data collected at the TEVATRON and the combined mass as of March 2013.

Chapter 2

Experimantal Setup

2.1 Tevatron

The TEVATRON collider, housed at the Fermi National Accelerator Laboratory in Batavia, IL, is a high energy $p\bar{p}$ collider with a center of mass energy 1.96 TeV. The high energy protons and anti-protons are produced and set to collide at the points where the two detectors, CDF and DØ, are located. In this section, we discuss the production of high energy protons and anti-protons used by the TEVATRON.

The process of generating high energy protons and anti-protons, starts with a bottle of hydrogen gas, is achieved through a series of integrated equipment and accelerators (Fig. 2.1). To begin with, Hydrogen atoms are converted to H^- and then are accelerated to an energy of 750 keV. These accelerated H^- atoms are then transferred to the Linear Accelerator (LINAC) and further accelerated to 400 MeV. At the Booster Synchrotron Accelerator, electrons are removed from the H^- to obtain protons. These protons are further accelerated to 150 GeV at the Main Injector (MI) and a portion of those are used to establish the anti-proton beam. The rest of the protons are injected into the TEVATRON.

The proton beam at the MI goes through a series of steps before arriving as anti-protons at the TEVATRON. First, the high energy proton beam collides on a Nickel target. This produces a wide variety of particles including a few anti-protons. The anti-protons are selected using magnets and collected at the accumulator [4].

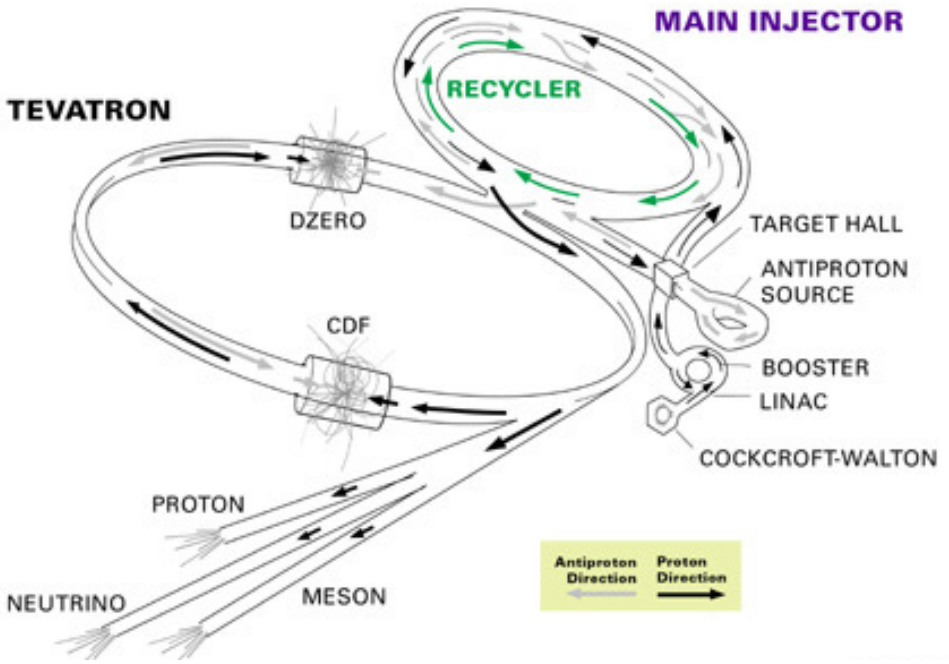


Figure 2.1: The TEVATRON accelerator chain [4].

A typical collider operation period at the TEVATRON is called a store which lasts for about 10-20 hours. At the beginning of each store, protons and anti-protons are injected into the TEVATRON and accelerated to their final energy of 980 GeV. The two beams are set to collide where the two detectors are located.

During RunIIB[1], the TEVATRON delivered about 9.8 fb^{-1} of total integrated luminosity [2] to the DØ detector. The collider operations of the TEVATRON were concluded on the 30^{th} of September 2011 and DØ collected about 9.0 fb^{-1} integrated luminosity.

2.2 The DØ Detector

In the following sections, the various detector components of the RunII DØ detector will be discussed (Fig. 2.2).

[1] RunIIB data set is collected from June 2006 to September 2011.

[2] The luminosity ($\mathcal{L}$) is defined as the number of interactions per unit area per unit time. Integrated luminosity is a measure of data collected over a specified time period.

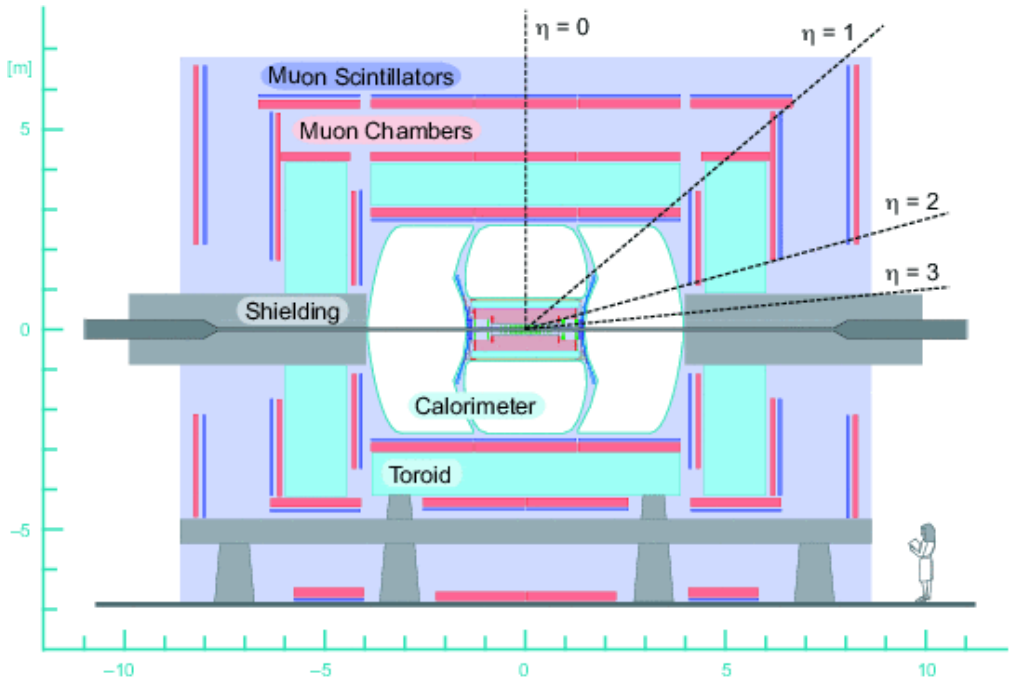


Figure 2.2: The upgraded DØ detector

For experiments of this nature, it is crucial to define a coordinate system to describe the properties of the particles detected. The center of the DØ detector is regarded as the origin of a right handed coordinate system, where the z-axis points in the direction of the proton beam and the y-axis points upwards. The polar angle is denoted by θ and the azimuthal angle is denoted by ϕ. Usually, pseudorapidity(η) is used to define the polar angle and is defined as $-\ln[\tan(\theta/2)]$ which approximates the true rapidity $y = 1/2\ln[(E + p_z c)/(E - p_z c)]$ in the limit of $(mc^2/E) \to 0$. The pseudorapidity is denoted by η_{det} when it is measured with respect to the origin of the detector and is denoted by η when measured with respect to the primary vertex of the event.

2.2.1 Central Tracking Detector

The central tracking system is the component of the DØ detector closest to the beam pipe. It is composed of the Silicon Microstrip Tracker, the Central Fiber

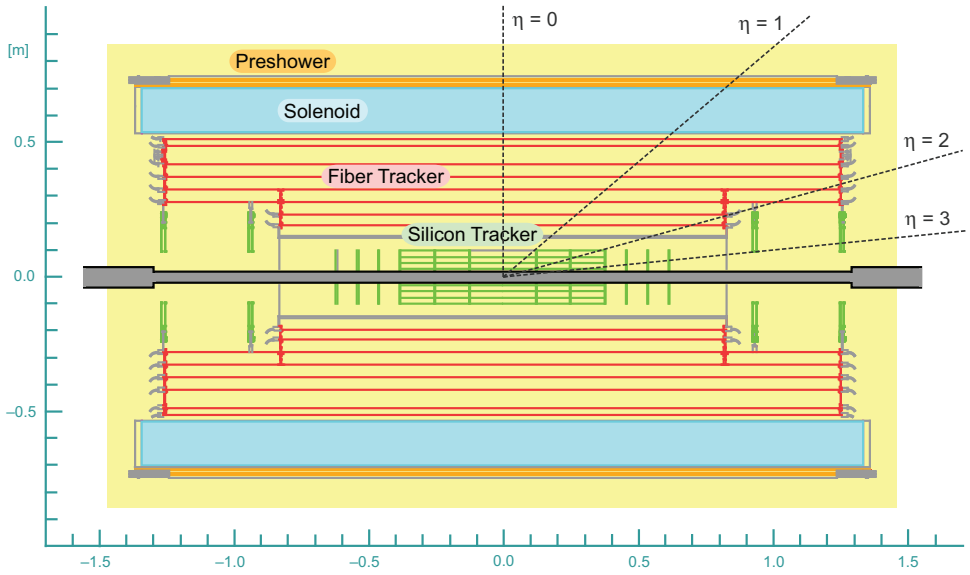


Figure 2.3: Cross sectional view of upgraded tracking system [5].

Tracker and a 2 T solenoid magnet (Fig. 2.3).

The Central Tracking Detector provides information to reconstruct the position of the primary vertex of the interaction and transverse momentum of charged particles [5]. It also provides information to identify heavy flavor jets using secondary vertices.

The Silicon Microstrip Tracker (SMT), which nearly covers the full detector pseudorapidity ($|\eta| < 3$)[5], employs semiconductor technology to obtain precise measurements of the position of a charged particle. The basic element in a semiconductor is a junction diode with a bias voltage. When a charged particle travels through this setup a voltage; signal, is created which will serve as a position measurement. These elements are combined to build the barrel assemblies and disk assemblies. The SMT detector is constructed from six barrel assemblies, 12 F disks and 4 H disks (Fig. 2.4) totaling 792 776 read out channels.

The Central Fiber Tracker (CFT), which extends radially from 20 cm to 52 cm from the center of the beam pipe, is built using scintillating fibers [5]. The signal,

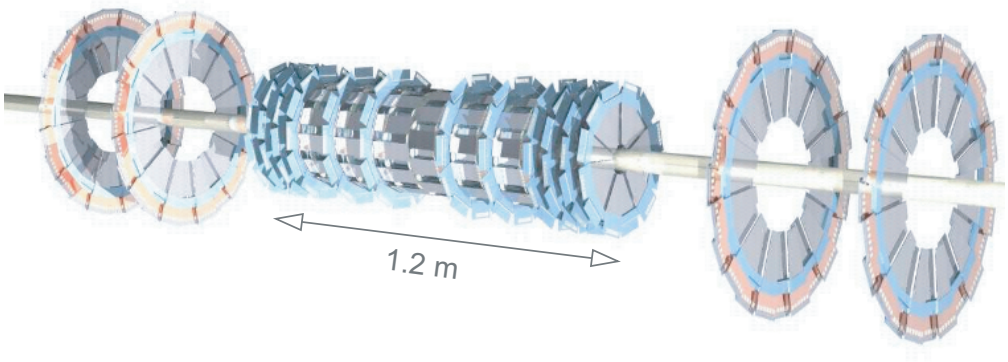


Figure 2.4: Silicon Microstrip Tracking system, disk and barrel design [5].

in the form of light, produced at the scintillating fibers is carried to the visible light photon counters (VLPC) via wave guides. At the VLPCs the light signal is converted to an electrical signal [5]. The CFT detector is assembled to provide two major services. The first, is to reconstruct tracks and measure the momentum of the charged particles combining the information from the SMT detector. Second, is to provide "Level 1" hardware triggering.

This position information is used to reconstruct the primary vertex position, secondary vertices and track segments using advanced algorithms.

The superconducting solenoidal magnet, encloses the CFT and SMT detectors, provides enhanced tracking and momentum measurement capabilities.

2.2.2 Calorimeter

Primarily, the calorimeter measures the energy of electrons, photons and jets and aids in the identification of electrons, photons, jets and muons. Furthermore, the transverse energy imbalance of an event is calculated using the information from the calorimeter.

Conversion of the energy of a particle to a readable electrical signal is the essential functionality of the calorimeter. Hence, it is constructed using two types of materials, "passive material" and "active material". When the particles interact

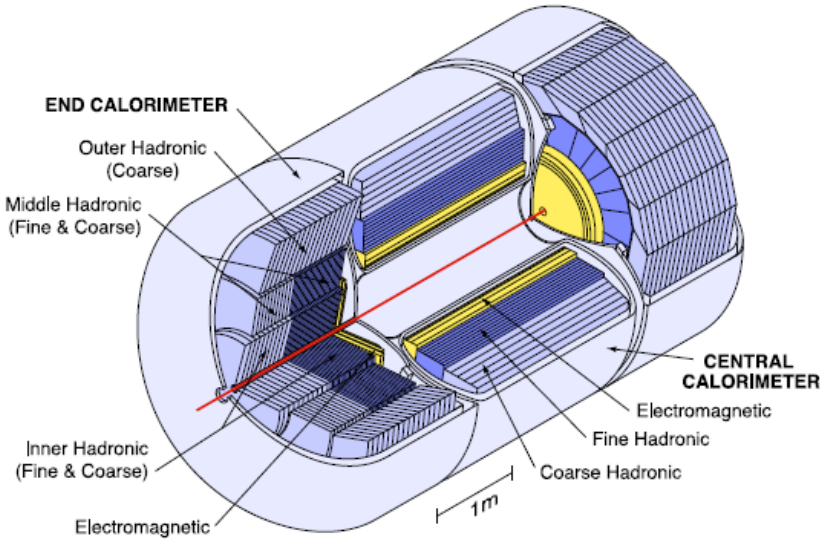


Figure 2.5: Three dimensional diagram of the calorimeter [5].

with the "passive material" they lose energy due to the creation of a shower of new particles. The "active material" produces a signal proportional to the number of particles in the resultant shower. Therefore, the produced electrical signal is proportional to the energy of the original particle.

The DØ calorimeter is composed of two components, the electromagnetic calorimeter and the hadronic calorimeter. The hadronic calorimeter consists of two parts, the Fine Hadronic (FH), which is constructed close to the beam pipe, and Coarse Hadronic (CH) calorimeter. The EM calorimeter and Hadronic calorimeter use liquid argon as the "active material", which yields an electric signal proportional to the number of particles in the cascade by ionization. In the EM calorimeter Uranium is used as the absorber plates while the FH calorimeter uses uranium-niobium alloy. The CH calorimeter absorber plates are built with copper in central calorimeter[3] (CC) region and with stainless steel in end cap[4] (EC) region.

In Fig. 2.5, a three dimensional diagram of the calorimeter, illustrates the

[3]The central calorimeter provides coverage in pseudorapidities ($|\eta|$) up to ≈ 1.1

[4]The End Cap calorimeters provides coverage in pseudorapidities ($|\eta|$) up to ≈ 1.1 to 4.2

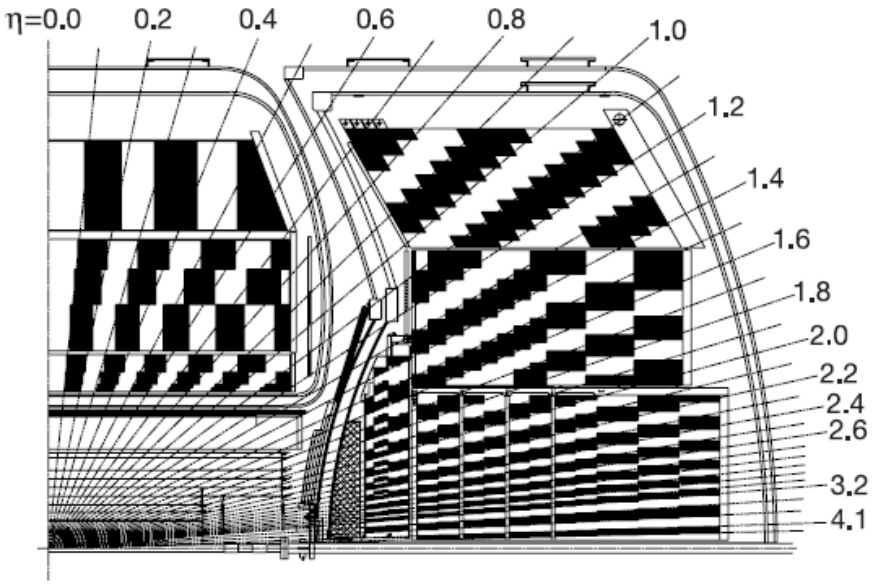


Figure 2.6: Diagram of the portion of the calorimeter. Shaded segments and white segments show the clustered cells to build towers [5].

arrangement of these components. The CC and EC calorimeters cover up to $\eta \approx 4$. The calorimeter is segmented into small virtual units (Fig. 2.7) called "read out cells" since this is the basic unit from which signal is being read. Calorimeter towers are built by clustering these cells together and are approximately $\Delta\phi \times \Delta\eta = 0.1 \times 0.1$ in size (Fig. 2.6).

The EM and Hadronic calorimeter dimensions are set such that all the energy from the particles except muons and neutrinos are well contained within the detector. An electron traveling through the EM calorimeter may lose energy through the bremsstrahlung process and $e\overline{e}$ pair production. The depth (20 radiation lengths[5]) of the EM calorimeter is set such that most of its energy is deposited in the EM calorimeter. The amount of energy a hadronic shower looses after traveling a nuclear interaction length (λ) is equal to $1 - e^{-1}$. Close to the CC region, the thickness of the hadronic calorimeter is about six nuclear interaction lengths.

Along with the readout electronic system the DØ calorimeter provides precise

[5]Radiation length is the distance an electron will travel while retaining $1/e$ of its energy.

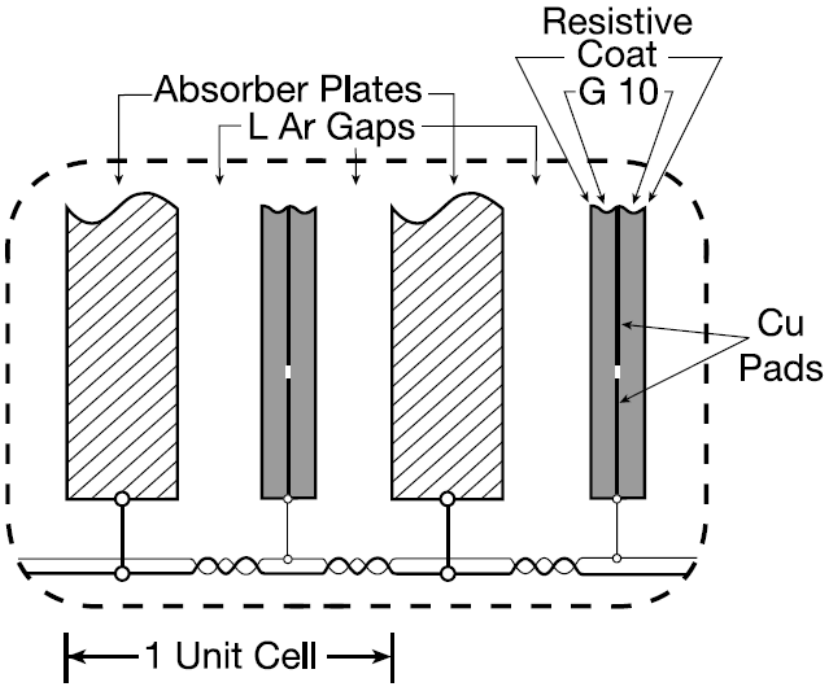


Figure 2.7: Schematic diagram of a calorimeter cell [5].

measurements of energy while supporting object identification process.

2.2.3 Muon System

The essential objective of the muon system in the DØ detector is to identify muons which escape the detector leaving only tracks. In addition, the muon system provides timing information to reject cosmic background and aids in momentum measurement.

The Muon system, which covers up to $|\eta| \approx 2.0$, consists of the central muon system, the forward muon system and the toroidal magnets (Fig. 2.8) [5]. The central muon system is constructed with Proportional Drift Tubes (PDTs) and scintillation counters while the forward muon system is constructed with Mini Drift Tubes (MDTs). Each muon system, forward and central, consists of three layers, A, B and C layers (Fig. 2.8).

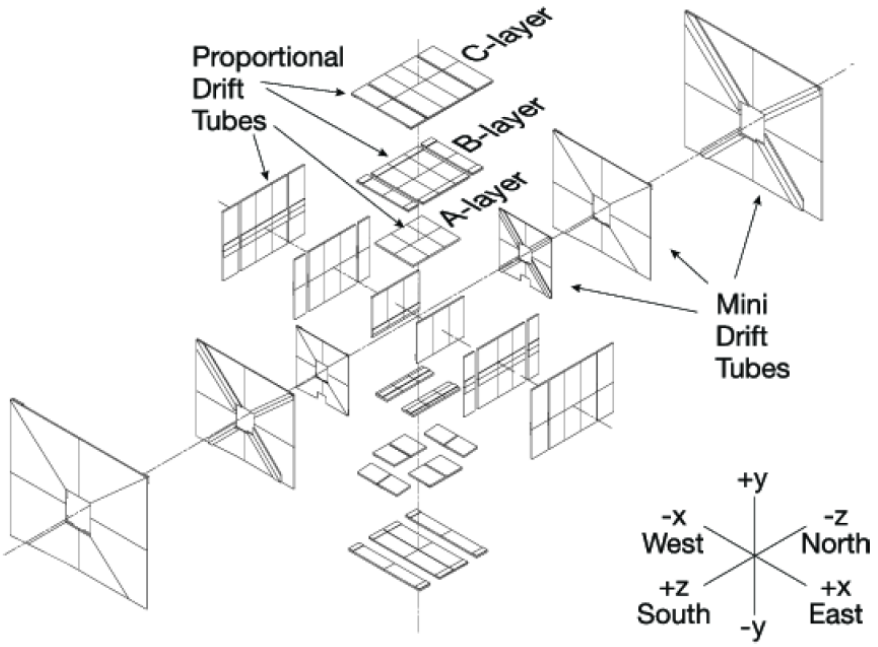


Figure 2.8: Diagram of the muon drift chambers [5].

The drift tubes are constructed from a hollow tube through which a thin wire is attached. The wire and the wall of the tube are kept at a voltage difference (wires are kept at a higher voltage than the walls). Due to the muons, the gas inside the tube is ionized. The voltage difference between wire and the wall cause the electrons to drift towards the wires and a electrical signal is generated.

Scintillator counter produces a signal in the form of visible light (photons) upon interaction with the radiation. This light signal is converted to an electrical signal by photo multiplier tubes [5]. Due their fast operation, scintillation counters (Fig. 2.9) are used in triggering [5].

The muon system along with the details from the tracking system helps to efficiently reconstruct the muons and reject the cosmic background, while providing fast information to Level 1 trigger.

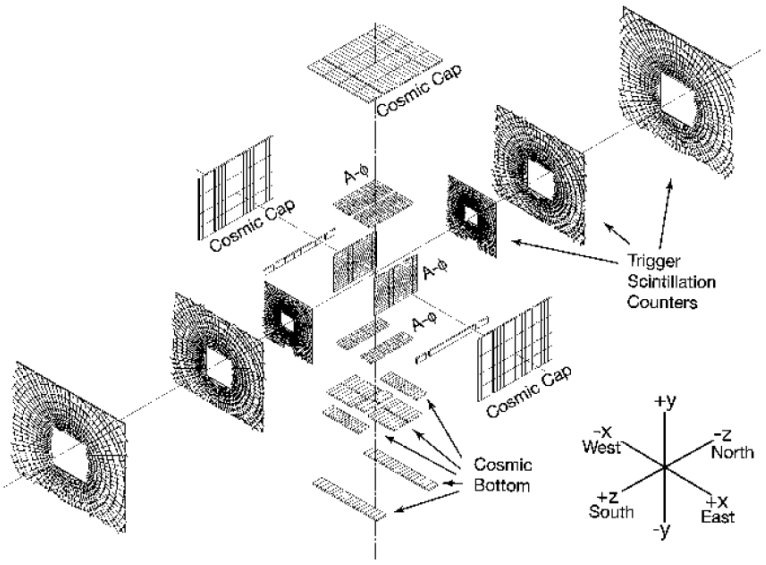


Figure 2.9: Diagram of the muon scintillator counters [5].

2.2.4 Luminosity Monitor

The TEVATRON luminosity at the DØ interaction region is an input to many physics analyzes. It provides the normalization for cross section measurements. The luminosity ($\mathcal{L}$) is defined as the number of interactions per unit area per unit time,

$$\frac{dN}{dt} = \sigma\mathcal{L} \tag{2.1}$$

where σ is the cross section. The luminosity detectors are located 140 cm from the center of the detector (Fig. 2.10) and are built from plastic scintillation counters (Fig. 2.11).

The luminosity monitor counts the number of inelastic $p\overline{p}$ collisions (N_{LM}) to assess the luminosity (Fig. 2.10) [5]. To model the effect of multiple interactions on N_{LM} fraction of beam crossing with no collisions are counted and Poisson statistics are used to estimate an average N_{LM} which is denoted by $\overline{N}_{LM}$. Then,

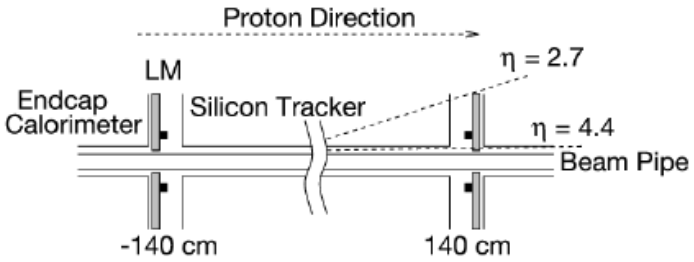


Figure 2.10: The position of the luminosity detector with respect to the beam pipe [5].

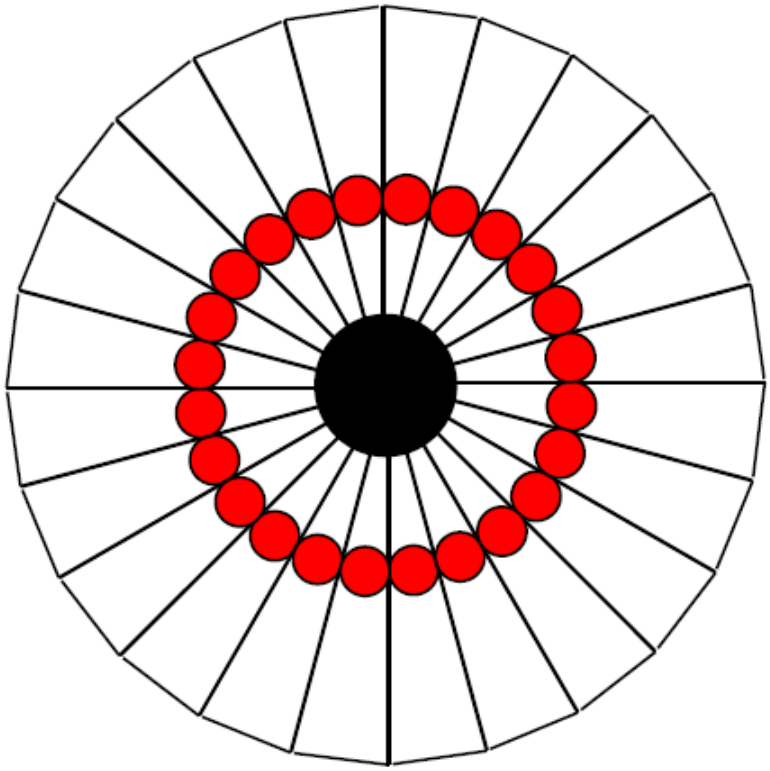


Figure 2.11: The arrangement of scintillator counters in the Luminosity Monitors [5].

the measured luminosity can be written as

$$L = \frac{f\overline{N}_{LM}}{\sigma_{LM}} \tag{2.2}$$

where, f is beam crossing frequency and σ_{LM} is effective cross section of the luminosity monitor. The effective cross section is calculated by taking into account the acceptance and efficiency of the luminosity detector [5].

2.2.5 Trigger System

At the DØ collision hall about 2.5 million events occur in one second [26]. It is impractical to record all these events onto storage disks for later analysis, due to time and resource constraints. The purpose of an exceptional trigger system is to determine the events of physics interest.

The DØ trigger system is a three level system of which level 1 (L1) is purely a hardware trigger system, level 2 (L2) is a combination of hardware and software trigger system and level 3 (L3) is purely a software based trigger system. At L1, L2 and L3 event rates are 2.5 kHz, 1kHz and 100 Hz respectively (Fig. 2.12). L1 and L2 buffers play an important role allowing more time in decision making for the subsequent triggers, hence minimize the experiment's dead time [5].

The L1 decisions should be made in a minimal time to avoid detector dead time. Therefore, it considers only the information from the main detector components such as the calorimeter, tracking system and muon system, named as L1CAL, L1CTT and L1MUO respectively (Fig. 2.13). The L1CAL makes decisions depending on the transverse energy deposits in the calorimeter, L1CTT and L1MUO triggers operate on the momentum information reconstructed using tracks.

The L2 system has an accept rate of 1 kHz and receives events from L1 at a rate of $\approx$ 2.5 kHz. The L2 make decisions depending on physics objects in contrast to L1. The L3 system reduces the input rate of $\approx$ 1 kHz from L2 to $\approx$ 100 Hz

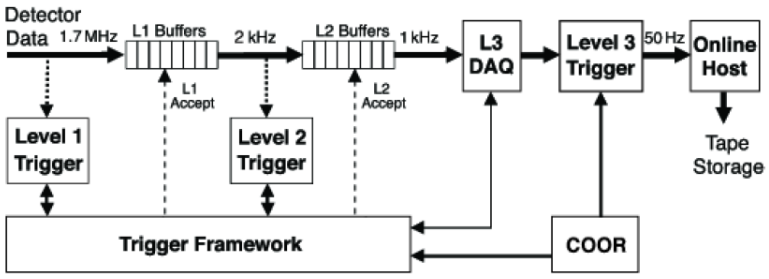


Figure 2.12: Block diagram of trigger and data acquisition systems [5].

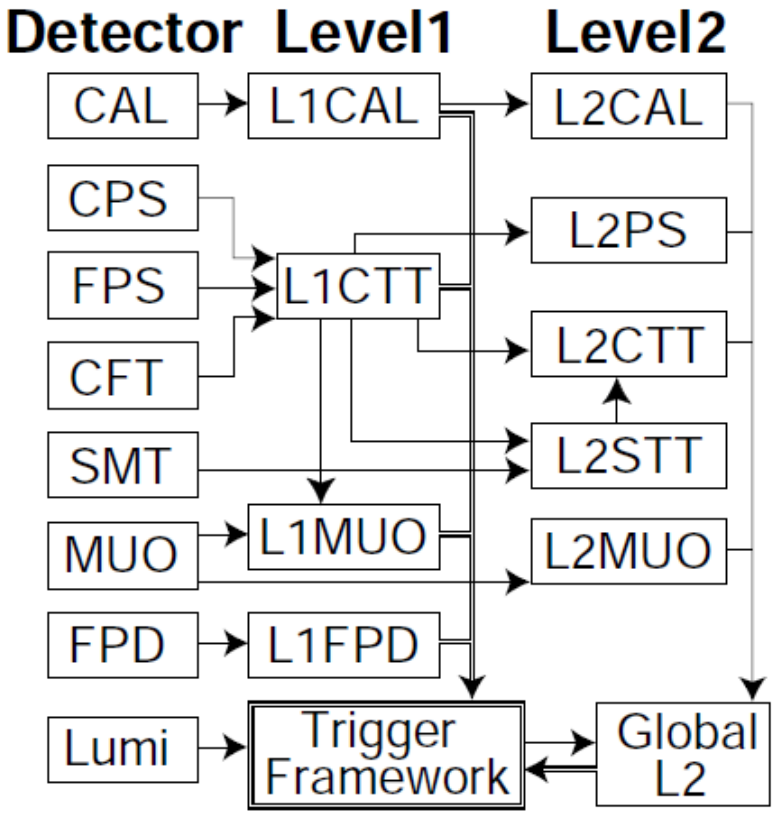


Figure 2.13: The layout of the L1 and L2 trigger systems [5].

making the decisions on partially reconstructed events. The accepted events are stored to be used in physics analysis.

Chapter 3

Object Reconstruction

The signatures left by the particles in each detector component should be combined to reconstruct physics objects such as electrons, muons, jets and taus (Fig. 3.1). These objects should be integrated in order to reconstruct a given event to use in the physics analyzes. These identified object energies might differ from their true energies at the point of collision due to imperfections in the detector as well as inefficiencies in the reconstruction algorithms. Hence, measured object energies should be calibrated before using in the analysis.

3.1 Track Identification

The charged particle trajectories provide information used to calculate momenta, identify primary vertices, secondary vertices and particles.

First, the tracks are identified and reconstructed from the information obtained from SMT and CFT tracking detectors. In DØ two tracking algorithms are used, the Histogramming Track Finder (HTF) and the Alternate Algorithm (AA).

The most common pattern recognition technique used in High Energy Physics is the Kalman filter [27]. The Kalman filter combines a set of measurements, which may include background from other processes, made over time to predict the current state of an unknown variable. Due to the very large number of hits on the tracking detector, $10^4 - 10^6$, it is computationally costly to use the Kalman

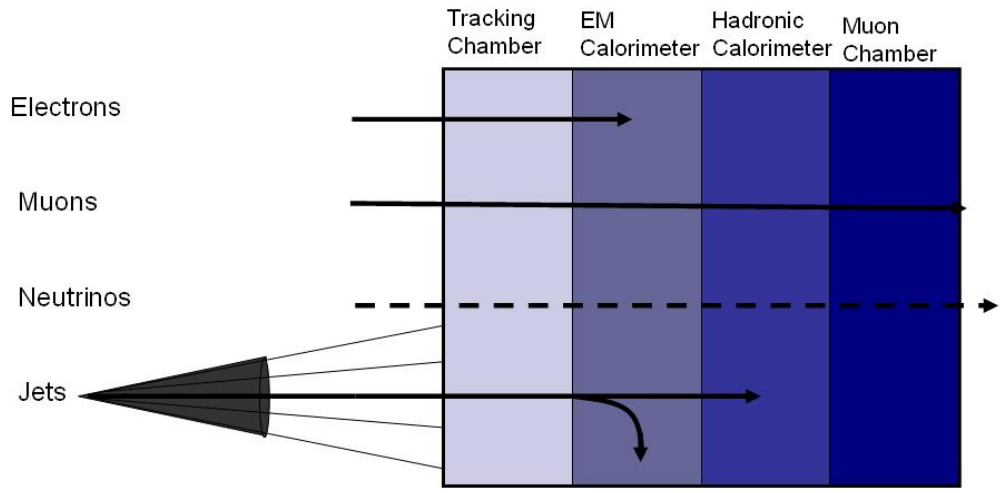


Figure 3.1: The way different particles interact with each detector component.

filter alone [6]. Hence, the HTF algorithm combines the Hough transformation method with the Kalman filter to reduce the amount of computation.

Due to the mostly uniform magnetic field generated by the 2 T solenoid magnet at the tracking detector, charged particles encounter a magnetic force of $q\vec{v} \times \vec{B}$, where q is the charge of the particle, $\vec{v}$ is the velocity and $\vec{B}$ is the magnetic field. In the ideal case, where there is no interference with the material, the momentum can be estimated using the curvature of the track. The position of this particle can be written as (ρ, d_0, ϕ), where, the curvature of the trajectory $\rho = q|\vec{B}|/p_T$, d_0 is the impact parameter and ϕ is the azimuthal angle [6]. For trajectories with small impact parameters, the parameter space (ρ, d_0, ϕ) reduces to (ρ, ϕ) hence knowledge of ρ and ϕ defines the position of a particle. This parameter space can be divided into small cells of $-\rho_0 < \rho < \rho_0, 0 < \phi < 2\pi$, where , $\rho_0 = qB/p_T^{min}$. The p_T^{min} is the minimum p_T of the tracks to be identified [6]. This creates a two dimensional grid which is depicted as a two dimensional (2-d) histogram. The bin content of the 2-d histogram is incremented when there is a hit. Hence, hits from the same track create a peak in the histogram while hits from different tracks will randomly fill the bins of the histogram. Therefore, a track is clearly distinguished

from the background hits. However, this needs N_h^2 operations to resolve an event where there are N_h hits. Thus, it is still computationally costly.

To resolve this issue, a Hough transformation is implemented [28]. The Hough transformation is used to transform the (x, y) coordinate system to (ρ, ϕ) parameter space. For a hit in the (x, y) coordinate system, many trajectories can be drawn such that they pass through the point of origin and the location of a hit. These trajectories transform to a line in the (ρ, ϕ) parameter space. Hence, all the hits from one charged particle will intersect at one point in the (ρ, ϕ) parameter space that will lead to a prominent peak in the 2-d histogram (Fig. 3.2) [6]. Then, the cells with less than N_h^{min} hits are removed to eliminate the background due to noise. Finally, the Kalman filter is applied on these selected cells to identify the tracks [6].

The Alternate Algorithm starts with three hits in the SMT detector that corresponds to a track originated from the interaction point [29]. Then this track is extrapolated to the next layer by searching for hits. If the χ^2 between the hit and the track is smaller than a predefined value the hit is considered as a part of the track candidate. This procedure is followed for the rest of SMT layers and CFT layers to define the track.

The track information from the HTF and the AA are combined to improve the efficiency of the track reconstruction.

3.2 Primary Vertex

"Primary Vertex", the point where proton and anti-proton collide. The resultant particles from this collision trace back to the primary vertex hence, it is necessary to reconstruct this with good precision to measure the kinematic properties of such particles. Furthermore, precise estimation of the primary vertex position is important as a discriminator from secondary vertices which originates from the

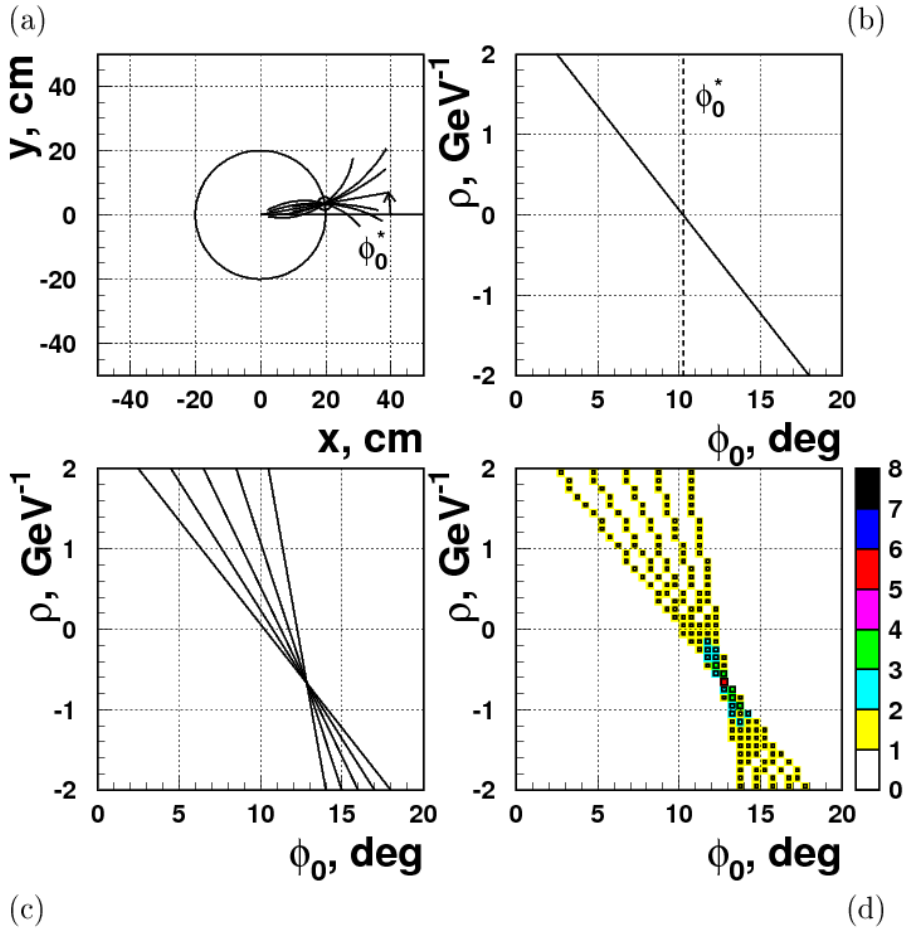


Figure 3.2: (a) Hit at 20 cm in the(x, y) coordinate system. (b) When all the possible trajectories in (a) is transformed to (ρ, ϕ) parameter space. (c) For 5 hits from the same charged particles in the (ρ, ϕ) parameter space. (d) 2-d histogram in the (ρ, ϕ) parameter space for these 5 hits [6].

decay of heavy flavor particles and reject background due to cosmic rays.

The Adaptive Vertex Fitting algorithm employs a two step procedure to reconstruct a primary vertex; first find the vertex and then implement the fitting [30]. To find the vertex all the tracks that satisfy a loose selection criteria, $p_T > 0.5$ GeV and two or more SMT hits, are fitted to find a common point. Then the tracks are removed with large impact parameters to establish an improved reconstructed position. To further increase the accuracy, tracks with the largest χ^2 contributions to the vertex are removed and continued. This procedure is repeated until the χ^2 per degree of freedom is reduced to 10 or less [30]. This procedure will lead to a list of primary vertices since there are many interactions per event. One primary vertex from this list is chosen to be from the hard scattering process using the fact that the transverse momenta of tracks from this process is higher compared to that coming from the remaining primary vertices due to the underlying events.

The primary vertex reconstruction algorithm provides a precise position measurement which is necessary for this analysis since there are long lived particles (b-quarks) in the final state.

3.3 Electrons

Electrons leave tracks in the tracking detector and deposit their energy in the electromagnetic calorimeter (Fig. 3.1). Hence, the reconstruction of an electron is performed by searching for electromagnetic clusters in the calorimeter and finding a matching track in the tracking detector [31].

The reconstruction of an electron starts with finding electromagnetic clusters in the calorimeter. The electromagnetic towers of cone radius $R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.2$ are grouped together around the highest energy tower to build an electromagnetic cluster [31]. The collective information from the EM cluster and the tracking detector is employed to reconstruct a true electron and thereby reject the

background.

Photons, pions and the instrumental effects can also mimic the signatures of electrons in the detector. The neutral pions that shower in the calorimeter associated with a track from a nearby charged particle could fake an electron [31]. Furthermore, photons after converting to electron positron pairs also can mimic electrons [31]. Hence, the electron reconstruction algorithm is tuned to address the above mentioned backgrounds.

True electrons are expected to deposit most of their energy in the EM calorimeter and have a distinctive shower shape compared to the background. The following variables are used to construct a discriminant to identify electrons.

EM fraction : The fraction of energy deposited in the EM calorimeter is defined as

$$f_{EM} = \frac{E_{EM}}{E_{tot}}, \tag{3.1}$$

where, E_{EM} is the energy deposited in EM calorimeter within $\Delta R < 0.2$ and E_{tot} is the total energy deposited within$\Delta R < 0.2$.

Electron isolation : Electron isolation is defined as

$$f_{iso} = \frac{E_{tot}(\Delta R < 0.4) - E_{EM}(\Delta R < 0.2)}{E_{EM}(\Delta R < 0.2)}. \tag{3.2}$$

Objects with smaller isolation (f_{iso}) values guarantee that most of its energy is deposited in the EM calorimeter and distinguish from showers due to hadronic objects since they are expected to have wider showers which diffuse to the hadronic calorimeter.

- **HMatrix :** Employs the distinguish shape of the shower due to an electron in the calorimeter to discriminate from hadronic showers.

- **Track Match Probability :** The closest track for an identified EM cluster is considered as the matching track for the electron. Then a probability is calculated based on the χ^2 for this track to be from the identified electron.

- **Electron Likelihood :** The electron likelihood discriminator is built to distinguish electrons from fake electrons which deposit a comparable amount of energy in the EM calorimeter such as photons and neutral pions which decay in to two photons [32]. The likelihood combines several variables which discriminate the signal from the fake electron background to form a discriminant such as fraction of energy deposited in the EM calorimeter, HMatrix, electron transverse energy divided by transverse momentum, shortest distance to the selected track from the primary vertex (along the z axis through this point), total number of tracks in the $\Delta R = 0.05$ cone and sum of transverse momenta of all the tracks (excluding the original candidate track) within $\Delta R = 0.4$ [32].

For this analysis events with one or more identified electrons are vetoed as described in the reference [33].

3.4 Muons

Muons are minimally ionizing particles (MIP) with a mass about 200 times the electrons. Hence, energy loss due to the radiation is very small compared to an electron[1]. They escape the detector leaving traces in the tracking detector, calorimeter and muon detector.

The reconstruction of muons is based on the tracking information from the tracking detector and hits in the muon detector. The muons identified based only

[1]synchrotron radiation, energy loss of a charged particle curving in a electric or magnetic field, is inversely proportional to m^4 where, m is the mass of the particle.

on the muon detectors are called "local muons". A local muon with a matched track in the tracking system is called a "central track-matched muon" [34].

The muons reconstructed from the above criterion are categorized (how likely a reconstructed muon is a real muon) using two criteria, "muon type" and "muon quality". The muon type, parameterized by the variable "nseg", depends on the availability of a matched track in the tracking system to the identified muon. If there is a track associated with the identified muon nseg is assigned with positive values, else assigned with negative values [35]. The absolute value of nseg represents the layers in the scintillating counter (A, B and C layers) being hit [34].

The background due to the cosmic muons is rejected using the timing information from the scintillator detector. For all three layers of the scintillating counter the hit times are required to be less than 10 ns between each layer [34].

In this analysis, we veto events if one or more muons satisfy the following requirements.

- **transverse momentum:** transverse momentum of the identified muon $>$ 20 GeV.

- `mediumnseg3:` at least two A layer wire hits, at least one A layer scintillator hit, at least two BC layer wire hits and at least one BC layer scintillator hits.

- `tracknewmedium:` magnitude of the distance from the extrapolated muon track from the reconstructed primary vertex (dca) is required to less than 0.2 cm (if associated with a SMT hit |dca| $<$ 0.04 cm) and χ^2 per degree of freedom $<$ 9.5.

- `TopScaledTight:` the scalar sum of transverse momenta of all the tracks within $\Delta R < 0.5$ around the muon (except muon p_T) $I^{trk}_{\Delta R<0.5} < 0.1$ and scalar sum of transverse energies within the cone $0.1 < \Delta R < 0.4$ $I^{trk}_{\Delta R<0.5} < 0.1$. This

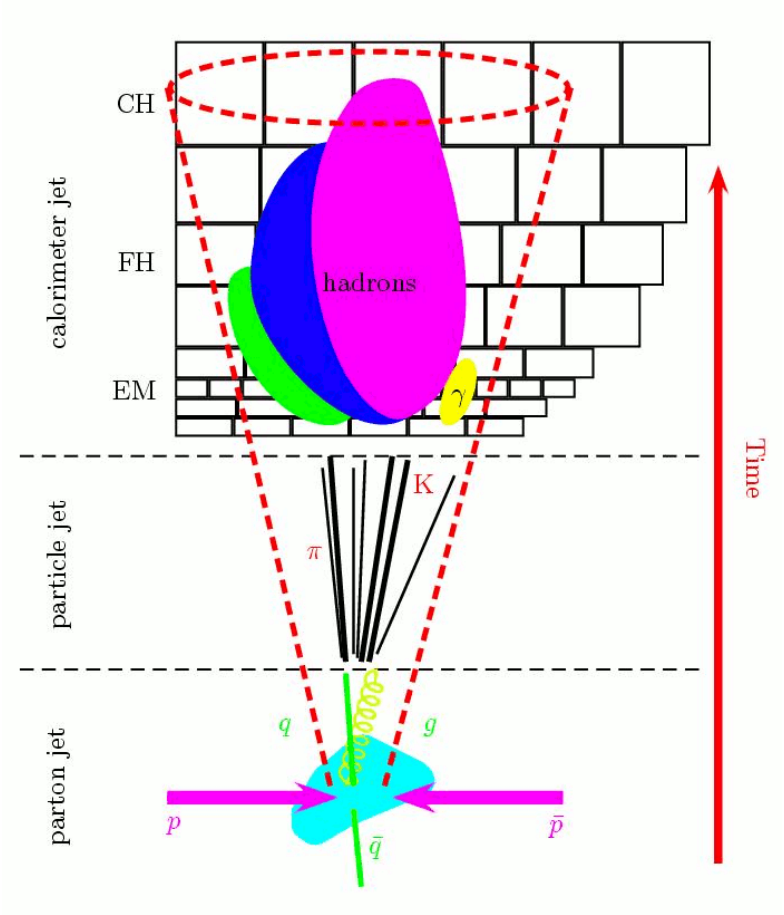


Figure 3.3: The evolution of a strongly interacting parton with the time in the DØ detector [7].

guarantees that the muon is isolated from secondary muons from heavy flavor quark decays.

3.5 Jets

Quarks and gluons hadronize creating a shower of particles and fragments (hadrons), which are called *jets*. Since, these gluons and quarks are boosted away from the $p\overline{p}$ collision point the resultant particle showers are identified as clusters of particles (Fig. 3.3).

A good jet reconstruction algorithm should be straightforward, efficiently use computing resources and provide maximum reconstruction efficiencies [36]. Both

experiments at the TEVATRON (CDF and DØ) use cone jet algorithms to reconstruct jets. A calorimeter cell is treated as a massless object with four momentum $\mathbf{p}^{cell} = (E^{cell}, p^{cell})$ [36]. To avoid cells that contain only electronic noise, a threshold is applied to each cell. NADA[2] algorithm is used to remove isolated cells and `t42` (see reference [38] for details) algorithm is used to remove cells with no prominent neighboring cells; hence cells due to electronic noise are further removed [36]. The cone jet algorithm cluster the selected cells to build pseudo-projective towers with a predefined radius R in $\eta \times \phi$ space. The four momentum of these towers are calculated as

$$\mathbf{p}^{\mathbf{tower}} = (E^{tower}, p^{tower}) = \sum_{i=cells \in tower} (E^i, p^i). \tag{3.3}$$

The DØ RunII cone algorithm uses the preclusters from the Simple Cone Algorithm as seeds to reconstruct the jets [36]. The Simple Cone Algorithm starts with the list of items, calorimeter towers in this case, which are p_T ordered. The towers are required to have $p_T > 0.5$ GeV. The tower with the highest p_T is selected as the precluster seed in the first iteration and removed from the list. In the next iteration, items that have $p_T > 1$ MeV and within 0.3 of ΔR from the precluster seed are combined with the precluster seed and removed from the list. This process is continued until there are no items left with the above requirements. These preclusters along with the calorimeter towers are then used as the input to the DØ RunII cone algorithm.

The preclusters from the Simple Cone Algorithm are the seeds for the RunII cone algorithm and the lists of items are used to build clusters of items which are called "proto-jets". The radius of the jet cone R_{cone} is selected according to the requirements of specific analysis. Currently there are two cone sizes which are being used, $R_{cone} = 0.5$ (JCCB jets) and $R_{cone} = 0.7$ (JCCA jets). The algorithm

[2]NADA algorithm is defined in the reference [37]

loops over p_T ordered preclusters to find the closest proto-jet. If the ΔR between precluster and the proto-jet is $\Delta R < R_{cone}/2$ the precluster is added to the proto-jet and algorithm proceed with the next precluster, else, it is used as a seed for a proto-jet candidate [36]. This is repeated until a stable (position of the axis of the jet is not changed from one iteration to the next) proto-jet is found.

Finally, the proto-jets are merged or split accordingly to reconstruct the jets to be used in the analysis. If one jet shares one or more items with another neighboring jet, the two jets are merged if the sum of p_T of the items shared is greater than 50 % of the highest p_T jet, left as separate jets if it is less than 50 % (for this case the shared items are assigned to the closest jet in ΔR).

The reconstructed jets are calibrated to their particle energies before being used in the physics analyzes. The jet energy scale calibration process will be discussed in detail in the next few sections.

3.5.1 Jet Energy Scale Corrections

The partons generated at the collision point are detected and reconstructed as explained above. The measured energies of the jets at the detector are different from the real particle level energy due to many factors such as noise from the calorimeter, not being able to include all the particles from the original parton in the jet cone and including particles not belong to the original parton in the jet cone. The process of correcting the jet's energy measured at the detector (detector level) to the energy of the parton generated at the collision point (particle level) is achieved via jet energy scale corrections.

The particle jet energy (E_{jet}^{ptcl}) can be written in terms of measured jet energy(E_{jet}^{meas}) [39]

$$E_{jet}^{ptcl} = \frac{E_{jet}^{meas} - E_0}{R_{jet} S_{jet}} \tag{3.4}$$

where, E_0 is the energy offset arising due to the noise and overlapping of other

$p\overline{p}$ events, R_{jet} is the response of the calorimeter and S_{jet} is the correction factor (showering correction) for energy deposited outside the jet cone and energy from particles that do not belong to the original parton deposited inside the jet cone. The measured jet energy can be written as

$$E_{jet}^{meas} = \sum_{i \in ptcljet} E_i^{meas} S_i + \sum_{i \notin ptcljet} E_i^{meas} S_i + E_0 \tag{3.5}$$

where, S_i is the fraction of the energy of the i_{th} particle of the jet included in the jets cone. With the estimated offset correction $\hat{E}_0$, response correction $\hat{R}_{jet}$ and showering correction the corrected jet energy can be written as

$$E_{jet}^{corr} = \frac{E_{jet}^{meas} - \hat{E}_0}{\hat{R}_{jet}\hat{S}_{jet}} \tag{3.6}$$

Due to the biases in the estimated values this is further corrected using Monte Carlo. The final expression for corrected jet energy is

$$E_{jet}^{corr} = \frac{E_{jet}^{meas} - \hat{E}_0 k_0}{\hat{R}_{jet} k_R \hat{S}_{jet}} \tag{3.7}$$

where, k_0 is bias correction for offset and k_R is bias correction for response.

In the following sections the estimation of each of these corrections are discussed in detail.

Jet Energy Offset

The energy deposited in the jet cone is a result of the actual jet and many other sources arising from the collision as well as in the detector. Electronic noise and uranium noise, due to the uranium decay, cause an energy offset at the detector level. The additional $p\overline{p}$ interactions and pile-ups also cause an energy offset.

At the TEVATRON, each bunch contains about 10^{10} protons and anti-protons.

Hence, there are many $p\overline{p}$ interactions for the same bunch. This is called Multiple Interactions(MI). Due to the lag of calorimeter electronics in transmitting the collected data compared to the bunch crossing frequency the information from the previous bunch crossing may be overlapped with the current causing the Pile-up effect. Both of these create an energy offset to the actual energy of the jet.

The offset energy can be written in terms of estimated contributions from Noise, Pile-up (NP) and Multiple Interactions as

$$\hat{E}_O^{ring}(i\eta, n_{PV}, \mathcal{L}) = \hat{E}_{NP}^{ring}(i\eta, \mathcal{L}) + \hat{E}_{MI}^{ring}(i\eta, n_{PV}, \mathcal{L}) \tag{3.8}$$

where, $\hat{E}_O^{ring}(i\eta, n_{PV}, \mathcal{L})$ is average offset energy per ring summed over all towers in $i\phi$, $\hat{E}_{NP}^{ring}(i\eta, \mathcal{L})$ is the estimated contributions from noise and pile-ups and $\hat{E}_{MI}^{ring}(i\eta, n_{PV}, \mathcal{L})$ is the estimated contributions from Multiple Interactions. The Pile up contribution mainly depends on the instantaneous luminosity ($\mathcal{L}$) and the Multiple Interactions component depends on the number of primary vertices (n_{PV}).

Zero Bias(ZB)[3] events are used to estimate the contributions from Noise and Pile-up ($\hat{E}_{NP}^{ring}(i\eta, \mathcal{L})$) after excluding events which have additional primary vertices. From this sample the average energy density for the i^{th} η ring is calculated.

To estimate the offset due to Multiple Interactions Minimum Bias(MB) (this trigger enforces very loose requirements on the final state of an event) events are used. The MB trigger demands simultaneous hits in both luminosity monitors located at z $\pm$ 140 cm from the center of the detector (The energy deposited in the calorimeter from elastic collisions is negligible compared to that of inelastic collisions). The average energy for MB events per ring is estimated in terms of $\mathcal{L}$ and n_{PV}. The average offset energy per ring is estimated as the difference between average energy of the MB event with exactly one primary vertex and average energy

[3]Zero Bias events are collected making no requirements of the final state.

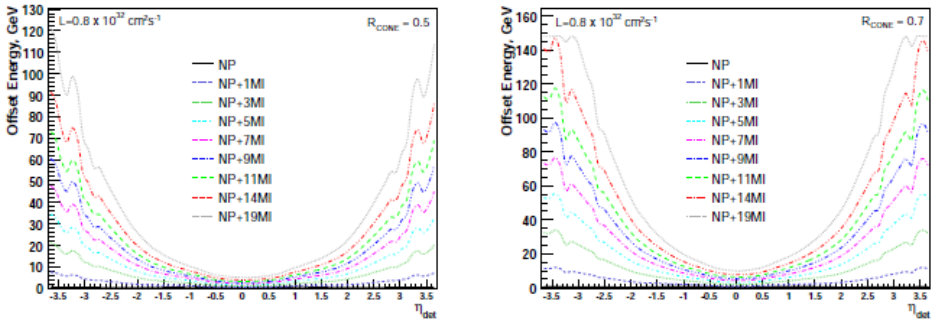


Figure 3.4: Estimated total offset jet energy (in GeV) as a function of detector η_{jet}^{det} , for jets with $R_{cone} = 0.5$(left) and $R_{cone} = 0.7$(right) for Run IIB-3.

of the MB event with any number of primary vertices as

$$\hat{E}_{MI}^{ring}(i\eta_{jet}^{det}, n_{PV}, \mathcal{L}) = \hat{E}_{MB}^{ring}(i\eta, n_{PV}, \mathcal{L}) - \hat{E}_{MB}^{ring}(i\eta, n_{PV} = 1, \mathcal{L}). \tag{3.9}$$

Finally, the total offset energy is the summation of the the NP and MI terms over all η rings within the jet cone

$$\hat{E}_O(i\eta_{jet}^{det}, n_{PV}, \mathcal{L}) \sim \sum_{i\eta \in R_{cone}} \hat{E}_O^{ring}(i\eta, n_{PV}, \mathcal{L}). \tag{3.10}$$

The estimated total offset correction for RunIIB-3[4] data set is shown in the Fig. 3.4.

Jet Energy Response

The total energy of the original parton is not deposited in the detector as the jets energy in the jet cone. There are numerous reasons that cause this. The particles emerging from the point of collision traverse the tracking detector and electronics before reaching the calorimeter. Hence, they lose energy due to these interactions.

[4]RunII data set is subdivided in to 5 sets, as RunIIA, RunIIB-1, RunIIB-2 and RunIIB-3, according to the period data is collected.

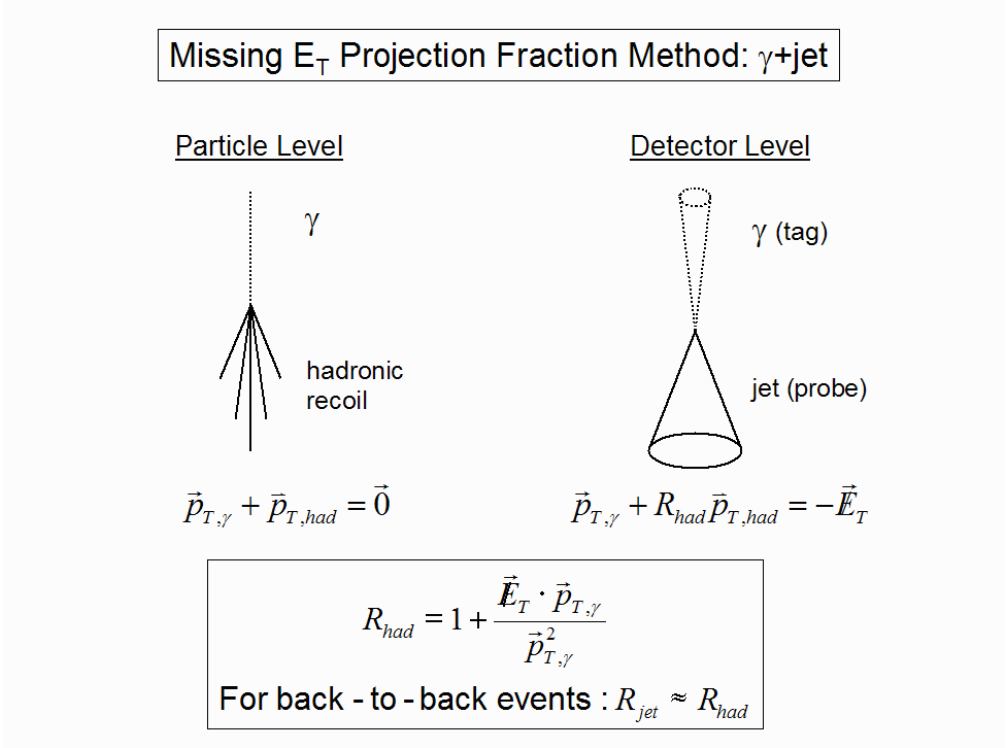


Figure 3.5: Two body process. Here, probe object is the jet for which response is being estimated. Tag object can be a γ, Z or a jet

Furthermore, particles originating from the initial hadron may bend due to the magnetic field and will not be included in the jets cone. Moreover, calorimeter response to hadrons is not linear. The above mentioned issues are accounted for the response correction, which is the largest contribution to the jet energy scale corrections.

To estimate the jet response, two body processes are used (Fig. 3.5). For a two body process, as illustrated in Fig. 3.5, at the particle level, the transverse momenta of the tag object[5] should be equal to that of the hadronic recoil

$$\vec{p}_{T_{tag}} + \vec{p}_{T_{recoil}} = 0. \tag{3.11}$$

[5]The object of which energy is measured to a better precision compared to the object of which energy is calibrated (probe object).

At the detector level, measured transverse momentum difference of the probe $\vec{p}^{meas}_{T_{recoil}}$ and tag $\vec{p}^{meas}_{T_{tag}}$ objects are equal to the missing transverse energy measured

$$\vec{p}^{meas}_{T_{tag}} + \vec{p}^{meas}_{T_{recoil}} = -\vec{\not{E}}^{meas}_{T} \tag{3.12}$$

The response of the detector is defined as $R_{Probe} = \vec{p}^{meas}/\vec{p}^{part}$ where, $\vec{p}^{part}$ is the transverse momentum at the particle level. This is defined in terms of the quantities measured at the detector as

$$\frac{R_{recoil}}{R_{tag}} = 1 + \frac{\vec{\not{E}}^{meas}_{T} \cdot \vec{n}_{Ttag}}{p^{meas}_{T_{tag}}} \tag{3.13}$$

where, $\vec{n}_{Ttag}$ is the direction of the tag object. Given the response of the tag object, the response of the hadronic recoil can be estimated. This method of estimating the response of the hadronic recoil is called missing E_T fraction (MPF) method.

The estimation of the response for the central calorimeter is called the Absolute Response Correction. Photon+jet events with photons in the central calorimeter ($|\eta|$<1), exactly one jet with $|\eta| < 0.4$, exactly one reconstructed primary vertex and with back to back photon and a jet in the $r-\phi$ plane are selected ($\Delta\phi(\gamma, jet)>$ 3 radians). Assuming photon is already corrected to the particle level ($R_{tag} = 1$) the jet energy response can be calculated. The calculated absolute response is shown in the Fig. 3.6 for jet algorithm JCCB. This information is used to derive the correction factors for the jets in the region $|\eta| > 0.4$.

The relative response is the ratio of MPF responses measured at the central calorimeter to that of anywhere else in the calorimeter. The relative MPF response is derived using the γ+$jets$ and dijet events. For γ+$jets$ the relative MPF response is written as

$$F^{\gamma+jets}_{\eta} = \frac{R^{\gamma+jets}_{MPF,\eta}}{R^{\gamma+jets}_{MPF,CC}}. \tag{3.14}$$

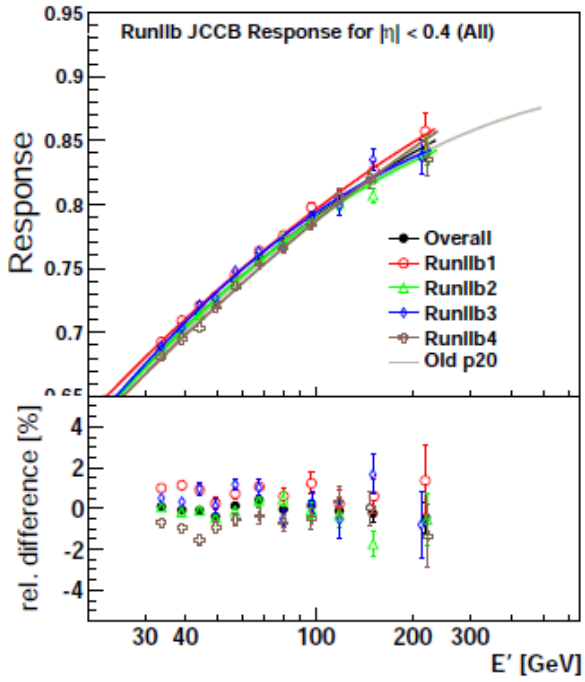


Figure 3.6: Absolute MPF response as a function of E' for the jet cone JCCB ($E' = p_{T_\gamma}^{meas} \cosh(\eta_{jet})$).

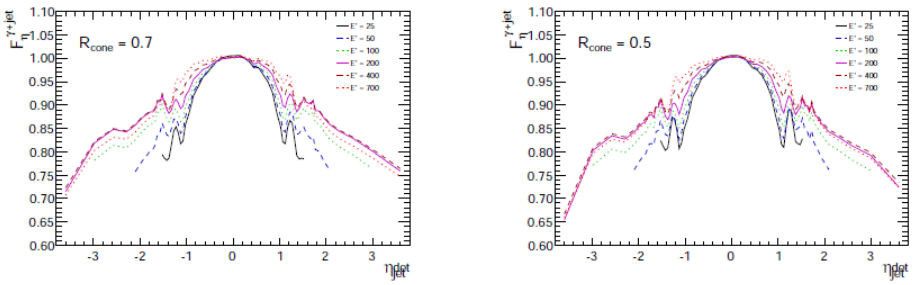


Figure 3.7: Relative MPF response as a function of η_{jet}^{det} for RunIIB1 data for $\gamma + jets$

Figure 3.7 shows the relative MPF response for $\gamma + jets$.

Showering Correction

The showering correction to the jet accounts for energy gained from the underlying event and lost due to its predefined cone size.

The showering correction is assessed using a γ+1jet samples for data and Monte Carlo. These events are required to have only one jet and exactly one reconstructed primary vertex. For Monte Carlo, the showering correction is estimated directly from the information available in the simulation. This correction is called the "true showering correction" and defined as

$$S_{jet} = \frac{\sum_{i \in ptcljet} E_i^{meas} S_i + \sum_{i \notin ptcljet} E_i^{meas} S_i}{\sum_{i \in ptcljet} E_i^{meas}} \tag{3.15}$$

where, E_i^{meas} is the visible energy in the calorimeter from the i^{th} particle and S_i is the fraction of energy contained in the jet cone [40]. To assess the showering correction for data, an annulus is defined in terms of the jet cone radius $\delta R < R_{cone}$, where R_{cone} is the jet cone radius [40]. The energy distribution of the jet with respect to the annulus radius is defined as the energy profile of the jet. Then

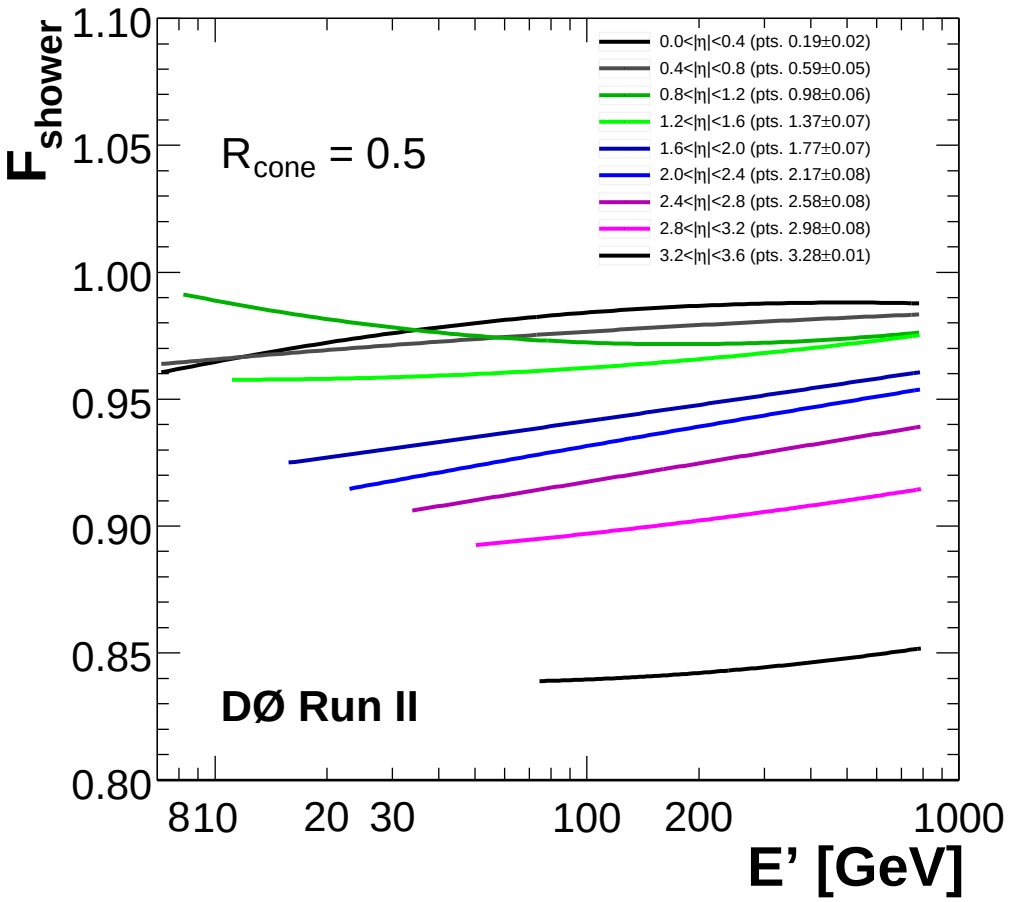


Figure 3.8: Showering correction for the jet cone R_{cone}=0.5 (JCCB algorithm) [8].

the showering correction for data is defined as

$$\hat{S}_{jet} = \frac{E^{MC}_{from-jet}(\delta R < R_{cone})}{E^{MC}_{from-jet}} + \frac{\alpha}{\beta}\frac{E^{MC}_{non-jet}(\delta R < R_{cone})}{E^{MC}_{from-jet}} \tag{3.16}$$

where, $E_{from-jet}$ is the energy profile of the particles belonging to the jet, $E^{MC}_{non-jet}$ is the energy profile of the particles not belonging to the jet and α and β are Monte Carlo to data scale factors. The showering correction obtained for jets with cone radius R_{cone}=0.5 is shown in the Fig. 3.8.

Global Closure Tests in MC and Data

The closure tests are designed to validate the jet energy scale corrections and its uncertainties. The Monte Carlo closure tests are a direct comparison of particle level energies to the corrected energies. Since, there is no such information available for data, corrected jet energies of data is compared to that of Monte Carlo.

The closure variable for Monte Carlo is defined as the ratio of corrected jet energy of the reconstructed jet E_{jet}^{corr} to matched particle jet energy E_{jet}^{ptcl}

$$D = \frac{\langle E_{jet}^{corr} \rangle}{\langle E_{jet}^{ptcl} \rangle}. \tag{3.17}$$

The particle jet is said to be matched if it is the closest jet to the reconstructed jet and ΔR between two jets is less than $R_{cone}/2$.

The closure tests for Monte Carlo are performed using $\gamma + jets$ events and the results are shown in the Fig. 3.9.

Closure variable for data is defined as

$$D = \frac{\langle E_{jet}^{corr,data} \rangle}{\langle E_{jet}^{corr,MC} \rangle} \tag{3.18}$$

where, $\langle E_{jet}^{corr,data} \rangle$ is the average corrected jet energies of data and $\langle E_{jet}^{corr,MC} \rangle$ is the average corrected jet energies of Monte Carlo. To compensate the effects on data due to the dijets background, Monte Carlo events are chosen as a mixture of $\gamma + jets$ events and dijets events weighted according to the purity. The closure plots for data is shown in the Fig. 3.10.

The closure tests prove that the latest derived jet energy scale calibration performs well hence, is a good estimation of the particle level energies of the jets.

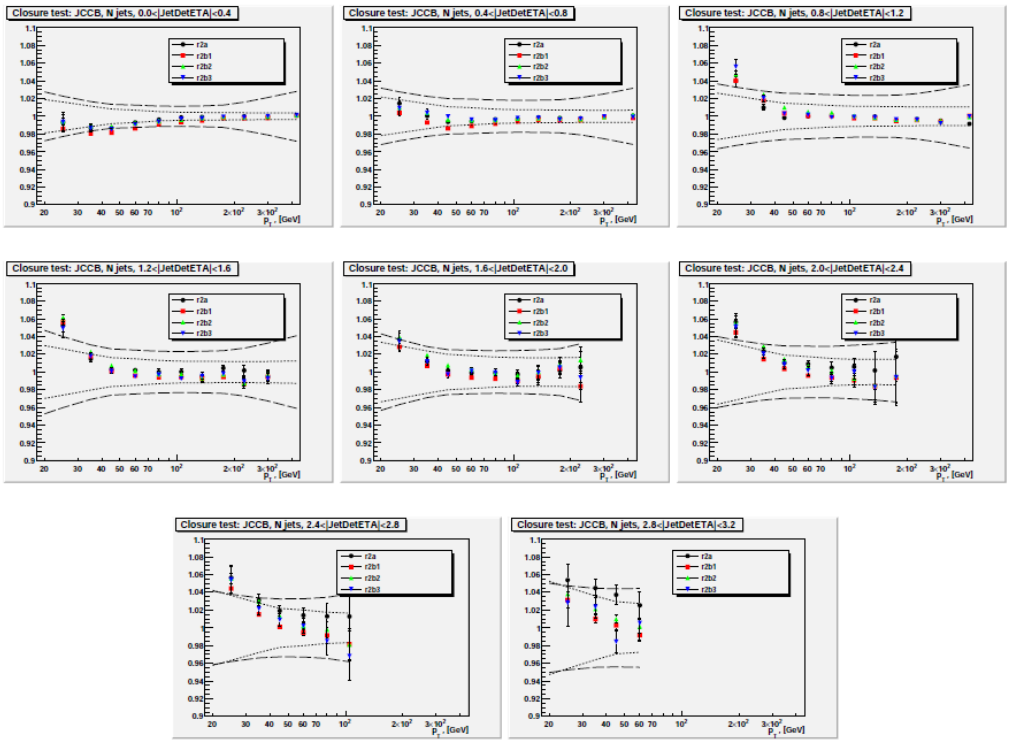


Figure 3.9: Direct closure tests for Monte Carlo for the JCCB jet algorithm

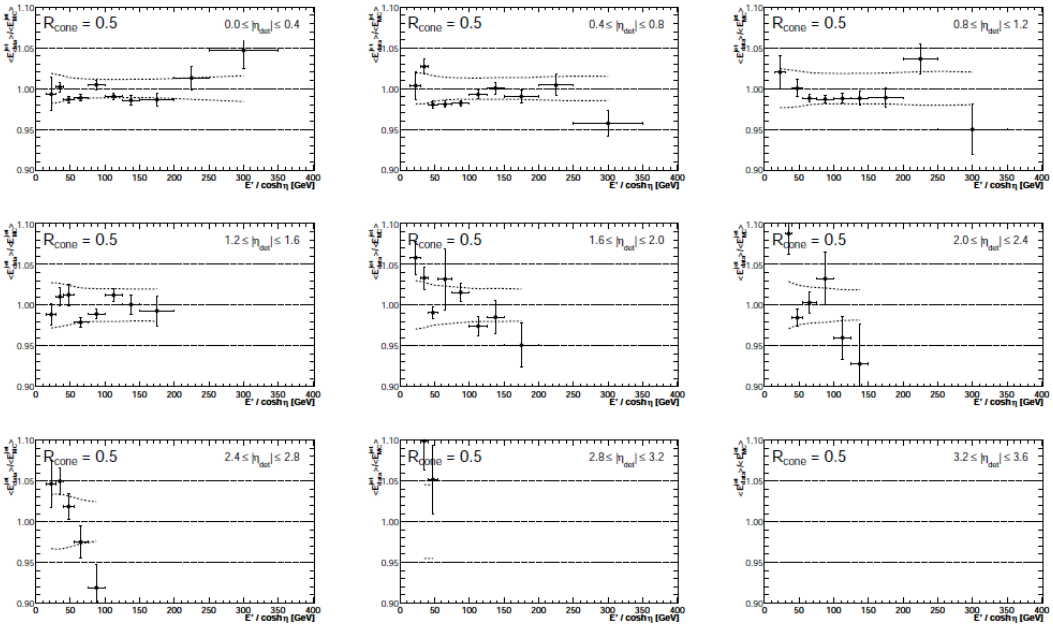


Figure 3.10: Direct closure tests for data for the JCCB jet algorithm.

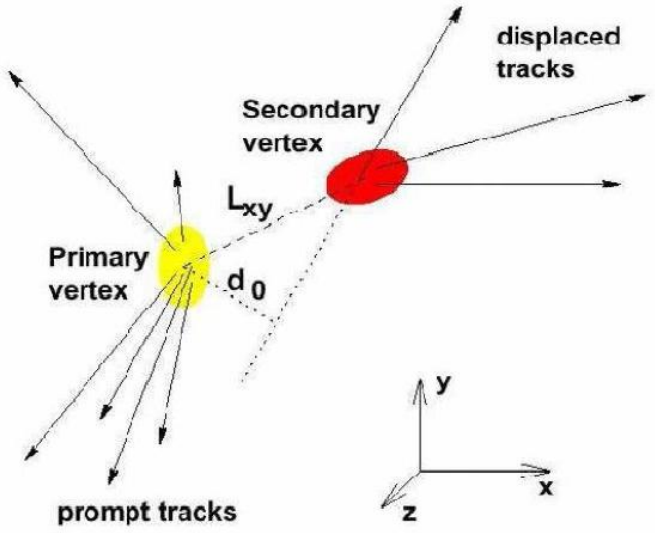


Figure 3.11: A displaced secondary vertex from the primary vertex due to a particle with a finite life time [9].

3.6 *b*-jets

The identification of the bottom quark from the top quark decay is very important since it can be used to distinguish the signal from background.

The reconstruction of the bottom quark is called *b*-tagging. Bottom quarks produced at the primary vertex hadronize into clusters of particles including B hadrons. These B hadrons have a relatively long life time hence can travel a few millimeters (due to time dilation) and decay into a cluster of particles (Fig. 3.11). The point at which the decay of the B hadron occurs is called the *secondary vertex*. Furthermore, about 20% of *b*-jets contain a muon within the jet cone. The bottom quark is identified through these unique features.

For this analysis, *b*-jets are identified using the MVA BL *b*-tagger developed by the *b*-ID group using a multivariate method used to combine suitable variables that distinguish *b*-jets from other jets [41]. The variables used are chosen from two classes of variables. The first category of variables are based on the impact

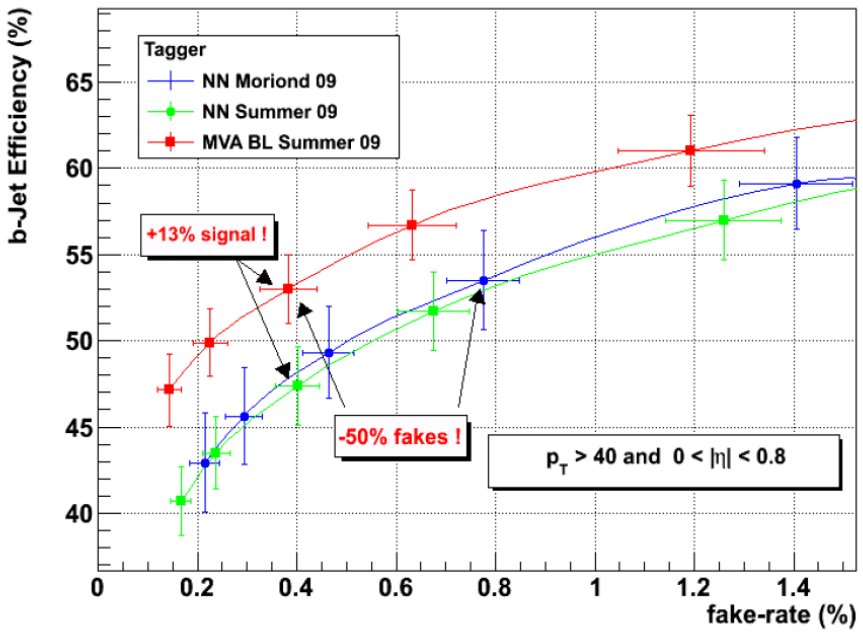


Figure 3.12: b-tagging efficiencies and fake rates for the MVA BL tagger and previous Neural Network tagger.

parameter(IP). This includes the "Jet Lifetime Probability" where the probability of tracks matched to a specific jet originating from a primary vertex is calculated. The lower the calculated probability, the more likely the jet is a b-jet [41]. The second category is based on the secondary vertex. These variables include the number of tracks from the secondary vertex, fraction of transverse momentum carried by the secondary vertex, decay length along the z direction of the detector. The MVA BL tagger combines these variables using Boosted Decision Trees (see section 5.1.2 for a description of Boosted Decision Trees) to build a single discriminant. The efficiency of this tagging algorithm is compared to the previous Neural Network based method in Fig. 3.12. Twelve operating points for MVA BL tagger are provided. In this analysis, we require two b-jets identified using the MVA BL medium operating point.

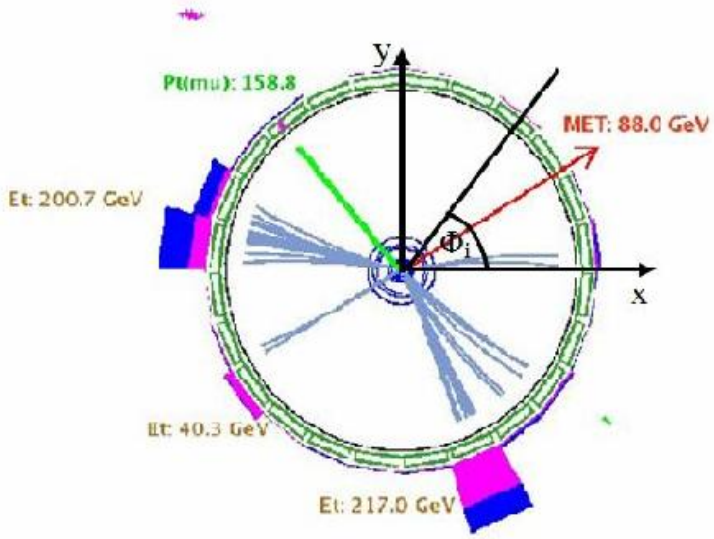


Figure 3.13: Illustration of method for calculating $\not\!\!E_T$. The xy is the transverse plane of the event and ϕ_i is the angle of the i^{th} calorimeter tower (Comments added) [10].

3.7 Missing Transverse Energy

Neutrinos do not interact with the detector material and leave no tracks or make energy deposits in the calorimeter. Therefore, events associated with the neutrinos are detected using the imbalance in transverse momentum. The transverse energy is calculated by taking the vector sum of transverse momenta of the calorimeter towers of the event ($\vec{p}_{T,i}$).

Theoretically, the vector sum of the transverse energy is equal to zero. However, this is not observed experimentally when the event is associated with particles which poorly interact with matter (such as neutrinos) and instrumental sources such as noise from the calorimeter. This deficit is called the missing transverse energy ($\not\!\!E_T$) and is calculated using (Fig. 3.13)

$$\not\!\!E_T = -\sum_i (\vec{P}_{T,i}). \tag{3.19}$$

At DØ raw $\not\!\!E_T$ is calculated using the energy deposits in the calorimeter cells except the cell in the coarse hadronic calorimeter due to the noise [42]. This is

being corrected for the physics objects in the event by subtracting raw energy of calorimeter objects and adding corrected energies and adding muon corrections. Hence, $\not\!\!E_T$ calculation is required to be the final piece in the object reconstruction.

Chapter 4

Samples and Event Selection

The template method employs the technique of comparing the signal and background templates to data. Thus, it requires having simulated events that represent the real signal events and multijet background events. The $t\bar{t} \rightarrow bW^{+}\bar{b}W^{-}$ signal template events are generated through the use of Monte Carlo event generators. The background template events are generated using data due to the complexity of multijet events which is not handled properly by Monte Carlo generators. The data collected using the DØ detector and the simulated events are required to pass certain quality requirements and primary criteria, such as triggers fired, number of jets, number of b-jets and detector η, before being considered for the analysis. This chapter details the simulation process for signal and background events and the selection procedure for the data, signal and background events used in the analysis.

4.1 Data Events

The data sample used in this analysis was collected during a time spam of about five years, from June 2006 to September 2011 (Fig. 4.1). The data collected during this time period is referred to as the RunIIB data set. During the RunIIB period, the TEVATRON delivered about 9.8 fb^{-1} of integrated luminosity and DØ collected about 9.0 fb^{-1} of data (the reference [43] provides a good explanation of

the calculation of integrated luminosity and its uncertainty for RunIIB).

The data quality requirements are necessary for the selection of valid data for an analysis. During data taking, the DØ detector may encounter technical difficulties, which requires the setting of a quality flag on the data taken that depends on the performance of the detector. A flag is set for each sub-detector giving its performance as "bad", "reasonable" or "good". Depending on the affected sub-detector component, the DØ data quality group provides the necessary information to the analyzers to select optimum events to analyze. The effective integrated luminosity is 8.6 fb^{-1} after the quality requirements are applied.

Table 4.1: Integrated luminosity collected using the DØ detector in each run period.

Run Period	Integrated luminosity (fb^{-1})
RunIIB-1	1.3
RunIIB-2	3.2
RunIIB-3	2.0
RunIIB-4	2.5
Total	9.0

4.2 Signal Event Generation

As mentioned above, this analysis requires signal events generated at various mass points to construct templates. Furthermore, signal samples are needed with certain variations to assess systematic uncertainties.

The $t\bar{t}$ signal is simulated through a series of steps. The hard scattering process is simulated first. This is followed by the showering of the partons and then the detector simulation. The hard scattering process is generated using the ALPGEN program which uses the Leading Order (LO) matrix element calculation. It provides a good description of W+jets production, Z+jets production and $t\bar{t}$ pro-

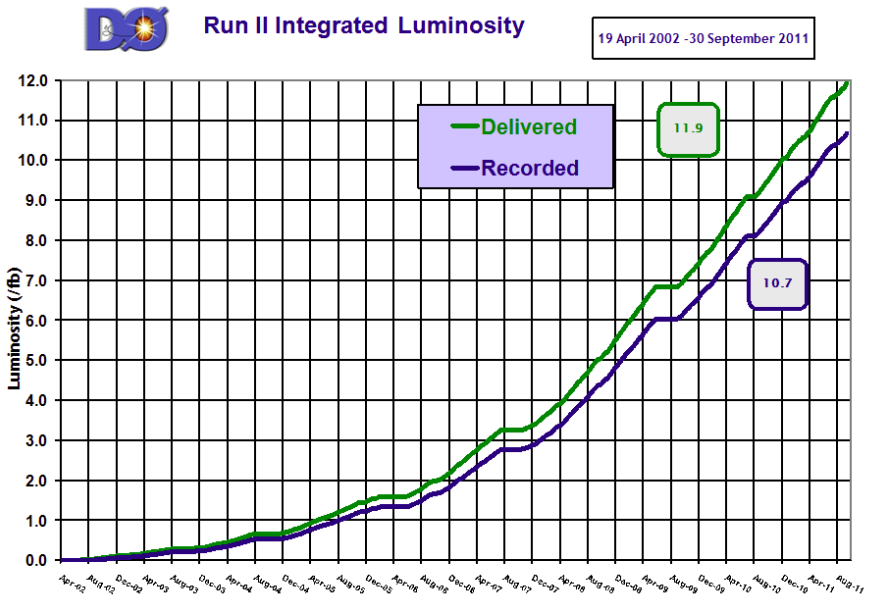


Figure 4.1: Integrated luminosity delivered by TEVATRON during RunII and that of collected by DØ [11].

duction [44]. The Parton Density Functions used in the simulation of the hard scattering process are those derived from data by the CTEQ collaboration [45]. The $t\bar{t}$ events are produced with exclusive parton multiplicity at tree level such as $t\bar{t}$ plus zero light partons ($t\bar{t} + 0lp$), $t\bar{t}$ plus one light parton ($t\bar{t} + 1lp$) and $t\bar{t}$ plus two light partons ($t\bar{t} + 2lp$) . The showering (evolution of quarks and gluons over the time) of the generated partons to final state particles is simulated using the PYTHIA program. The latest version of PYTHIA is capable of simulating hard scattering process, initial and final state parton showering and particle decays [46]. These particles are processed by a detector simulator, GEANT program which describes the interactions of particles with the matter. GEANT is adapted to include the detector information such as detector geometry, material and the magnetic fields generated by the solenoid and toroid magnets. Then the electronic responses in the detector, including such effects as electronic noise is simulated. The final step of the simulation is the DØ three level trigger system, which is

simulated using the TrigSim package. This final product of these steps is similar to the events collected by DØ detector.

Afterwards, this simulated data is processed through the object reconstruction algorithms. This analysis uses Monte Carlo samples generated at the mass points 150, 160, 165, 170, 172.5, 175, 180, 185, 190 and 195 GeV.

4.2.1 Monte Carlo Corrections

Even though, all known effects have been accounted for in the simulation, there are still small disagreements between the simulation and data that are accounted for by correction factors.

As discussed above, the simulated $t\bar{t}$ events are generated using the ALPGEN matrix element event generator and processed through PYTHIA to simulate the showering process. This may result in adding a parton by PYTHIA which is already being generated in the matrix element calculation using ALPGEN. For example, $t\bar{t} + 1lp$ event generated with ALPGEN will be duplicated by a $t\bar{t} + 0lp$ event, where an additional parton is created in the showering process from PYTHIA. The MLM technique is used to avoid this double counting. First, each jet in the event is required to have a minimum transverse energy. Then, the partons before the showering process are matched to the jets based on their separation in the (η,ϕ) space. If each jet in the event is matched to a parton, the event is retained, else is discarded [47].

The peak instantaneous luminosity at the DØ collision hall changed over time. From 2006 to 2011 the instantaneous luminosity changed from about 1.5×10^{32} $\text{cm}^{-2}\text{sec}^{-1}$ to about 4×10^{32} $\text{cm}^{-2}\text{sec}^{-1}$. The higher instantaneous luminosities increase the rate of multiple interactions. Since, the Monte Carlo generators do not properly model the Zero Bias events, the event is modelled by overlaying the Monte Carlo hard scatter event with Zero Bias events from data [48]. After the

data quality and preliminary selection is applied, a discrepancy between the data and Monte Carlo instantaneous luminosity distributions is observed. Hence, the instantaneous luminosity distribution of the Monte Carlo events are reweighted to follow that of data.

The Monte Carlo simulation assumes the interaction region to be Gaussian distributed along the z axis and centered at the origin of the DØ coordinate system. On the other hand the interaction region observed at DØ is not a Gaussian along the longitudinal direction. Therefore, a weight is applied to the Monte Carlo events such that the distribution of the primary vertex position along the z axis agrees with data.

The data events are selected by the DØ three level trigger system. This is simulated in Monte Carlo by the TrigSim package, which does not exactly match the data. The probability for an event to be selected (P) from all three triggers can be written as

$$P = P(L_1) \times P(L_2|L_1) \times P(L_3|L_1, L_2) \tag{4.1}$$

where, $P(L_1)$ is the probability that the event is selected by the L1 trigger, $P(L_2|L_1)$ is the conditional probability that it is selected by the L2 trigger given that it is selected by the L1 trigger and $P(L_3|L_1, L_2)$ is the conditional probability that it is selected by the L3 trigger given that it is selected by the L1 and L2 triggers. These probabilities are calculated using the jet trigger efficiencies, usually called trigger turn on curves, are measured using data collected by the detector on a few variables such as jet transverse momentum and pseudorapidity. The calculated probability is added as a weight to the event to correct for the trigger inefficiencies of Monte Carlo.

The six jets of each event are required to be vertex confirmed. A jet is said

to be vertex confirmed if two tracks within the jet point to the primary vertex. Monte Carlo jets always fulfill this requirement, but some of the jets from data do not (about 15 % of events with four or more jets with $p_T > 15$ GeV contain at least one non-vertex confirmed jet). Hence, the vertex confirmation weight is introduced to match the difference between data and Monte Carlo.

The b quark fragmentation of the Monte Carlo events is simulated using the Bowler fragmentation function [49]. The Bowler fragmentation function used in PYTHIA does not agree with the results from the LEP[1] experiments [50]. Therefore, the parameters of the Bowler fragmentation function are recalculated using the results from LEP data. Therefore, each b-jet is reweighted to bring the default Monte Carlo generated by PYTHIA into agreement with the LEP tunes. The final event weight (fragmentation weight) from this is the product of weights for each b-jet in the event.

The jets from the Monte Carlo simulation have a higher efficiency of being reconstructed and better energy resolutions compared to data due to approximations made by the showering programs and the detector modelling. These effects are corrected using JSSR (Jet Shifting Smearing and Removing) to incorporate the jet reconstruction efficiencies and jet resolution. The corrections are calculated comparing the transverse momentum imbalance of Z+jets events [51]

$$\Delta S = (p_T^{jet} - p_T^Z)/p_T^Z. \tag{4.2}$$

where, p_T^{jet} is the transverse momentum of the jet and p_T^Z is the transverse momentum of the Z boson. In the ideal situation the distribution of ΔS should be a Gaussian distribution centered at zero. But, this is not observed experimentally due to the above mentioned issues with the jets. Hence, ΔS is described using a

[1]Large Electron-Positron Collider

Gaussian distribution combined with an error function [51]

$$f(\Delta S) = Aexp[-(\Delta S - \langle \Delta S \rangle)^2][1 + erf(\frac{\Delta S - T}{2\sigma_T})]. \quad (4.3)$$

The variables $\langle \Delta S \rangle$, σ_T and T represent the jet shifting, resolution and threshold respectively. Hence, each of these contributions are calculated by fitting the ΔS distributions of data and Monte Carlo. The corrections are applied to the Monte Carlo events generated using ALPGEN and PYTHIA. The uncertainties are obtained by adding the data and Monte Carlo uncertainties (uncertainties on the fits performed using data and Monte Carlo) in quadrature.

Due to the approximations made in the simulation chain, the b-tagging probabilities are higher compared to data. To correct this effect the b-tagging efficiency (taggability scale factors) from data is applied to the simulated events. A random number is generated between zero and one and if this number is less than the taggability scale factor the jet is considered as taggable.

After the above corrections, the generated Monte Carlo events exhibit the same properties as data, hence, Monte Carlo is treated as data in the rest of the analysis.

4.3 Background Event Generation

The multijet events are a major background to the all hadronic final state. This analysis requires six jets in the final state including two b-jets. Hence, any background to this final state will have at least six jets. The simulation programs such as ALPGEN and PYTHIA do not provide a good description of these events. Therefore, background events for this analysis are generated using the collected data.

A data driven background such as this should fulfill certain requirements to ensure that it will not cause any biases in the measurement. First, the background

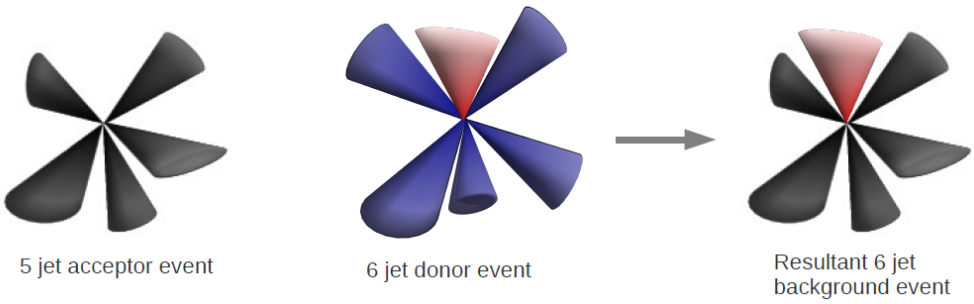


Figure 4.2: Modeling of the background starting from a five jet event.

should be minimally contaminated by $t\bar{t}$ signal events. Then, the background events should be able to describe the collider data. To exclude the $t\bar{t}$ signal events in the background, the background model is built starting from a five jet data sample with two or more identified b-jets (usually $t\bar{t}$ signal is expected to have six or more jets in the final state). This selection also helps to preserve the correlations between the b-jets in the event. Hence, it is expected to describe the variables associated with the b-jets well.

The additional jet required for the above mentioned five jet event is obtained from a six jet data event with two or more identified b-jets. The five jet event is called an "acceptor" event and the six jet event is called a "donor" event. Each donor event is compared to all the available acceptor events to identify the best match. The matching criteria is defined as follows. The matching observable Q is defined as

$$Q^2 = \frac{1}{N_{jets}} \sum_{j=0}^{N_{jets}} [(\Delta p_T^j)^2] \tag{4.4}$$

where, N_{jets} is the number of jets in the acceptor event and Δp_T^j is the transverse momentum difference between the donor event and the acceptor event for the j^{th} jet. The acceptor events with a Q^2 less than a predefined value are selected as the matched events. The sixth jet from the donor event is added to the acceptor event if ΔR between it and all five jets in the acceptor event is greater than 0.5

(this requirement ensures that there is no overlapping between the jets in the final background event) resulting in a six jet event (Fig. 4.2). The performance of this background model is tested by comparing the distributions for the background with signal subtracted data. To avoid any biases, only the variables with good agreement are chosen to be used in the multivariate discriminator; Boosted Decision Trees in this analysis.

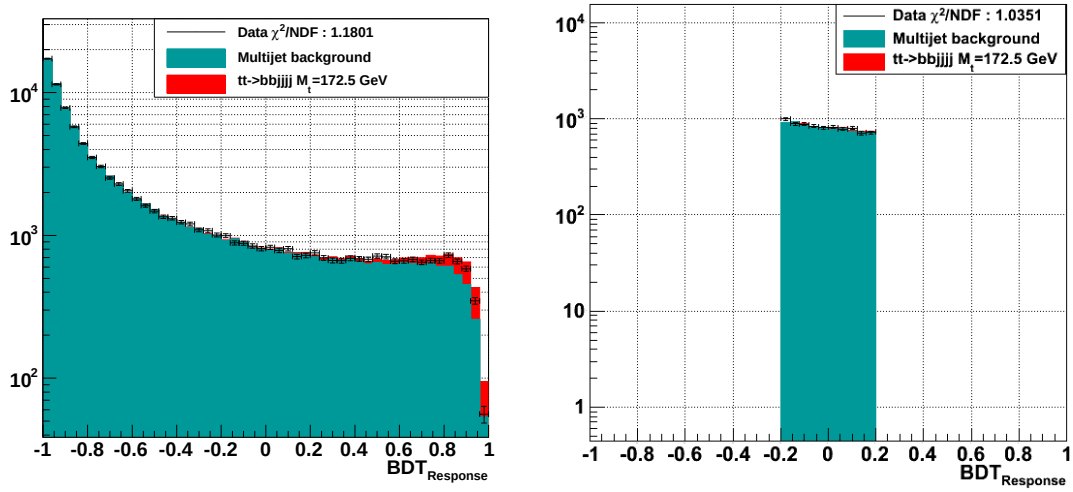
4.3.1 Background Validation

As described above, the background sample for this analysis is generated from a data sample and not by a simulation method. Hence, it is necessary to evaluate the performance of the background model.

The process of background validation for this analysis is developed based on the important variables associated with the method. The distributions of these variables are compared in control regions where the events do not enter the final measurement and events from the region where the analysis is performed. These regions are defined in terms of BDT response when constructing the top quark mass templates and W boson mass templates (Fig. 4.3)(see chapter 5 for more details).

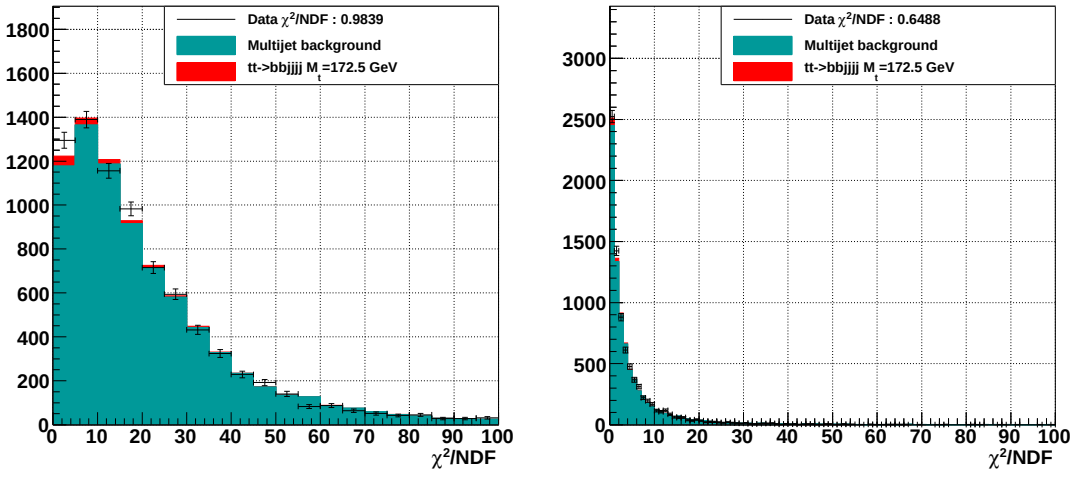
The comparison plots are built with signal, background and data. Signal is normalized to the expected cross section and background is normalized to the difference between data and signal. Figure 4.4 shows the χ^2 per degree of freedom plots for the BDT response region -0.2 to 0.2 and Fig. 4.5 shows the top quark mass and W boson mass templates for the same BDT response region (Fig. 4.3 shows the full BDT response distributions for signal, background and data).

To further affirm the performance of the background model, top quark mass templates and W boson mass templates are also compared for the events from the selected BDT region (the region where events are used to build the final templates.



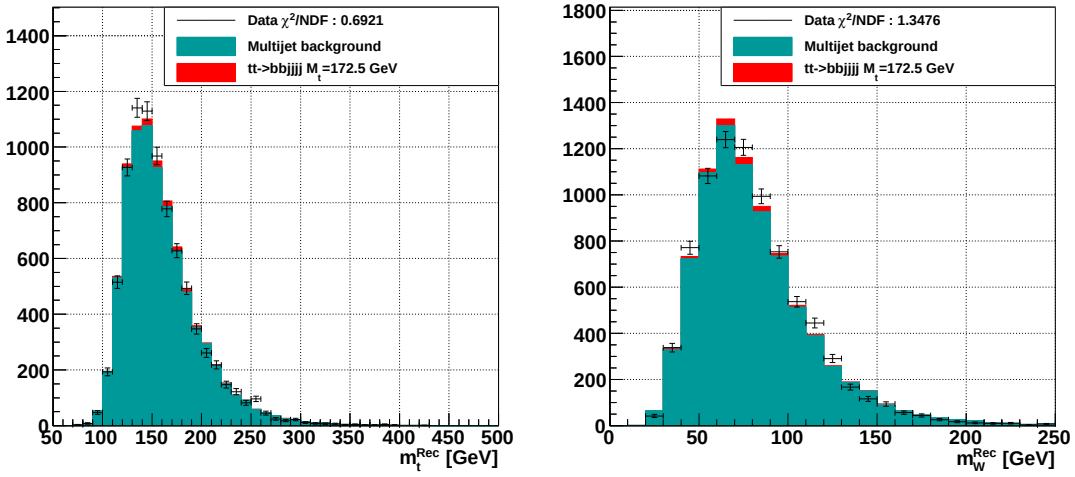
(a) (b)

Figure 4.3: BDT response distribution for signal, background and data (a) Full region. (b) Selected region (-0.2 < BDT_R < 0.2) to validate the background.



(a) (b)

Figure 4.4: (a) χ^2/NDF for top quark mass templates (b) χ^2/NDF for W boson mass templates for the BDT response region -0.2 to 0.2

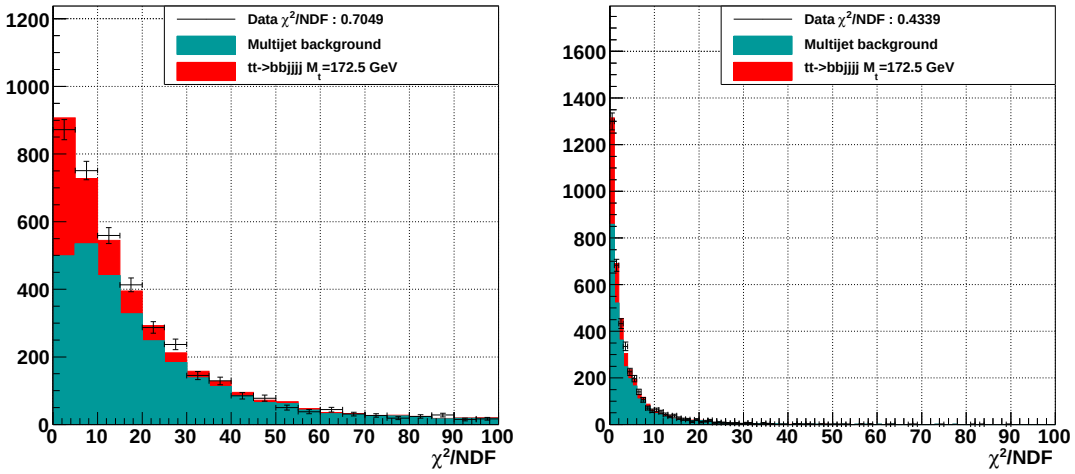


(a) (b)

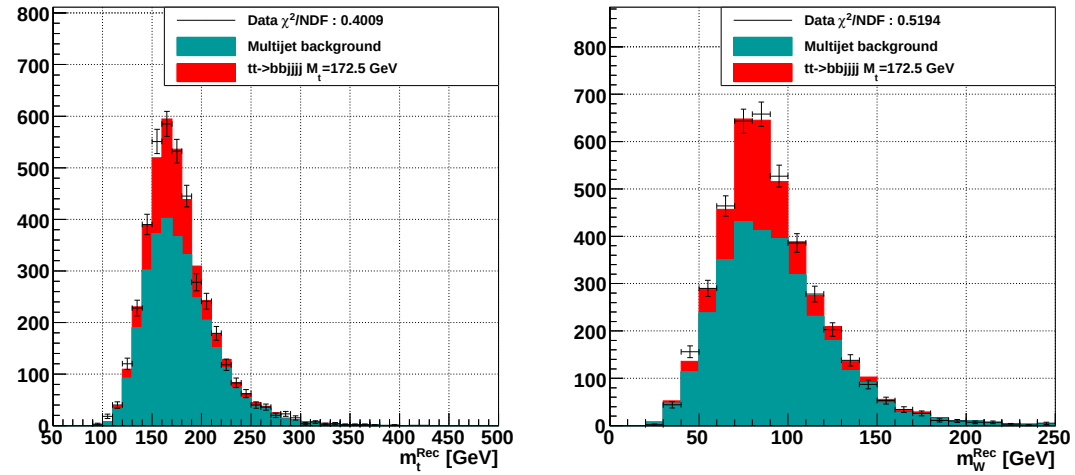
Figure 4.5: (a) Reconstructed top quark mass template (b) Reconstructed W boson mass template for the BDT response region -0.2 to 0.2.

Fig. 4.6). Most of the plots demonstrate a valid agreement between the data and background. A systematic uncertainty will be assessed for the variable (used in the BDTs) which shows the worst agreement (see chapter 6 for more details).

The conformity of data, signal and background suggest that the proposed background model is well suited for this method.



(a) (b)



(c) (d)

Figure 4.6: (a) χ^2/NDF for top quark mass templates (b) χ^2/NDF for W boson mass templates (c) Reconstructed top quark mass template (d) Reconstructed W boson mass template for the selected events for the analysis (events in the correct BDT region).

Chapter 5

Top Quark Mass Extraction

The top quark mass is extracted for the all hadronic final state applying the likelihood technique on data and the top quark mass templates. In order to reduce the uncertainty due to the jet energy scale, the mass is extracted from a 2-d fit to the top quark mass and the jet energy scale.

The procedure for the measurement is as follows: As mentioned earlier, the multijet background is a few orders of magnitude greater than the $t\bar{t}$ signal. Therefore, selection criteria are applied to maximize the statistical significance of the measurement. First, a pre-selection based on each variable such as jet transverse momentum and pseudorapidity is applied. Then, the signal fraction relative to the background is further improved using Boosted Decision Trees which combine information from many variables to generate a single discriminant. The top quark mass and W boson mass templates for signal and background are built using simulated events that pass the selection. Finally, data events are compared with these templates using the likelihood method to obtain the final measurements.

We will discuss the steps given above in detail in the next few sections.

5.1 Event Selection

The event selection criteria are formulated to extract the $t\bar{t} \to bbjjjj$ events from data, since the bulk of the events collected are not $t\bar{t}$ events. The selection pro-

cedure is based on the difference of kinematic and topological variables between $t\bar{t} \to bbjjjj$ and multijet background events.

5.1.1 Pre-selection

The purpose of the pre-selection is to remove the bulk of the events that are not of the interest. Following the pre-selection the signal fraction compared to the background is further increased using a multivariate discriminant.

Events are selected such that they satisfy the 3JT or 4JT triggers (see appendix A for details) with the z position of the primary vertex within ± 35 cm of the center of the interaction region. Furthermore, events are required to have exactly six jets including two b-jets identified with the b-ID operating point MVA BL medium. Each jet in the event is required to be vertex confirmed (two or more identified tracks associated with the primary vertex of the event [7]) and within a detector η of ± 2.5. Events with one or more isolated leptons are vetoed following the criteria described in the reference [52]. After the pre-selection has been applied, the signal to background ratio is 1:50.

5.1.2 Boosted Decision Trees

As stated above, the simple selection criteria based on individual variables has the disadvantage of removing large number of signal events while not substantially reducing the background contamination. There are various multivariate techniques that are in current use, such as the Likelihood, the Artificial Neural Network and the Boosted Decision Trees. These have the advantage of combining many observables to develop a single discriminant based on the correlations between these observables.

This analysis uses the Boosted Decision Trees (BDT) as the multivariate discriminant. A decision tree is schematically represented as a flow chart (Fig. 5.1)

where, a node represents a test on an observable. Decision trees are trained and tested using simulated signal and background events. First, these events are sorted on one observable and the value that gives the best separation between signal and background is chosen as the splitting value [53]. The signal like events are assigned to the signal node and background like events are assigned to the background node creating two new nodes (Fig. 5.1). Then the procedure is to go to one of the new nodes and repeat the above with another observable. This procedure is repeated until the *last node* (leaf) is pure signal or background, or certain user defined criteria is reached (these will be discussed later in this section) [53]. The boosting technique is introduced to avoid the instabilities in the decision tree technique (small changes in the training sample may result on large changes in the final discriminant [53]). An event misclassified in a tree (signal event assigned to a background leaf or a background event assigned to a signal leaf) is given a higher weight and a new tree is created with a different combination of observables. This procedure will result in a set of trees. The BDT response (BDT_R) is calculated summing the individual responses of each event. The decision tree response of an event can take two values, +1 if the event ended up in a signal leaf, -1 if the event ended up in a background leaf. The weighted sum of these responses for all the trees is the final response of the discriminant, where events with values close to +1 guarantees that it is more signal like. To test the validity of this discriminant, it is tested using a distinct set of signal and background events from the training sample. Good agreement between training and testing samples is required to avoid any over training.

For this analysis the BDTs are built utilizing the TMVA package for 31 observables [54]. These observables are chosen from a pool of about 100 kinematic and topological variables (see appendix B). First, the variables that exhibit poor agreement between the data, signal and background are removed. For variables

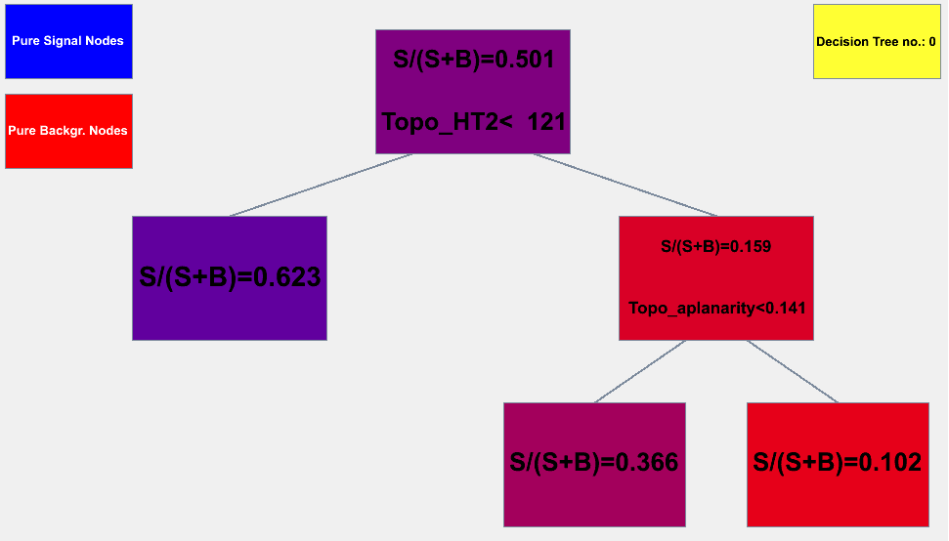


Figure 5.1: An example of a BDT used in the analysis.

that exhibit a large degree of correlation, the variable with the largest separability is retained where separability is defined as

$$\langle S \rangle = \frac{1}{2} \int \frac{(\hat{y_s} - \hat{y_b})^2}{\hat{y_s} + \hat{y_b}} dy \tag{5.1}$$

where, $\hat{y_s}$ and $\hat{y_b}$ are signal and background probability density functions.

The set of variables included in the BDT training along with the signal and background distributions are included in appendix B. The TMVA package allows the analyzer to set the configuration variables such as number of decision trees, maximum number of nodes and maximum number of levels in a tree. In the analysis, BDTs are trained with 2500 decision trees and the maximum number of tree nodes is set to five (all other parameters are set at their default values [54]). This configuration is selected as having the best discriminating power while having good agreement between training and testing samples (Fig. 5.2). Events that have a BDT response of ≥ 0.7 are selected to build the templates. This operating point is selected after performing a series of peudo-experiments by varying the BDT response selection. The point at which the statistical uncertainty of the

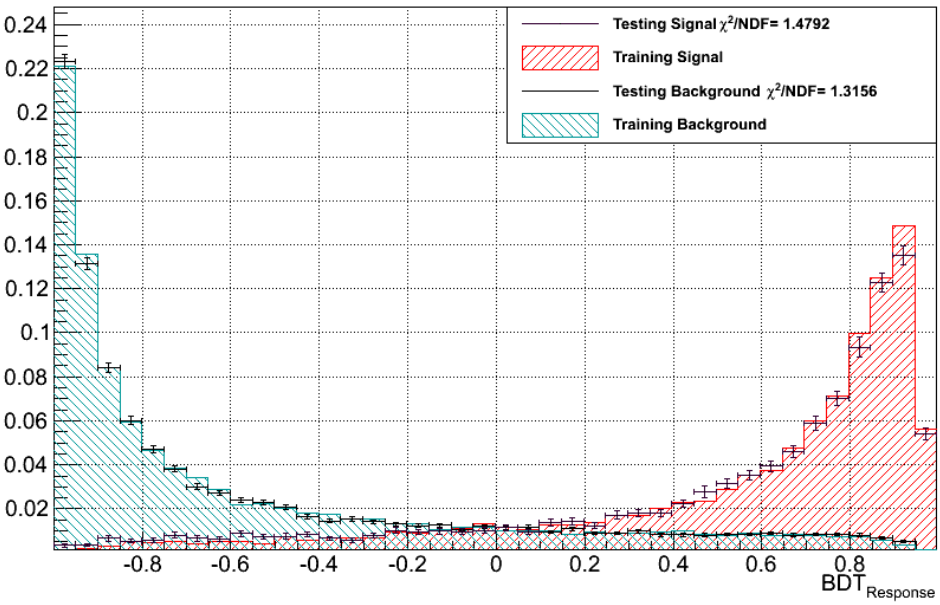


Figure 5.2: BDT training and testing samples overlayed.

measurement is minimized is BDT_R =0.7. The BDT response of data, signal (M_t=172.5 GeV) and background is shown in the Fig. 5.3. After the BDT selection, the signal to background ratio is about 1:3.

5.2 Top Quark Mass and W Boson Mass Templates

The template method employs a variable that is sensitive to the quantity to be measured to build the template distributions. This selected variable may not exactly represent the quantity of interest. But, the correlation between these two variables allows one to obtain a estimator for the variable on which the measurement is performed. For this analysis, the reconstructed top quark mass m_t^{Rec} (top quark mass obtained by fitting the kinematics of the six jets in the final state) is selected as the variable to build the templates to extract the top quark mass. The reconstructed W boson mass (W boson mass obtained by fitting the kinematics of the four light jets in the final state) is selected to reduce the uncertainty due to the jet energy scale calibration.

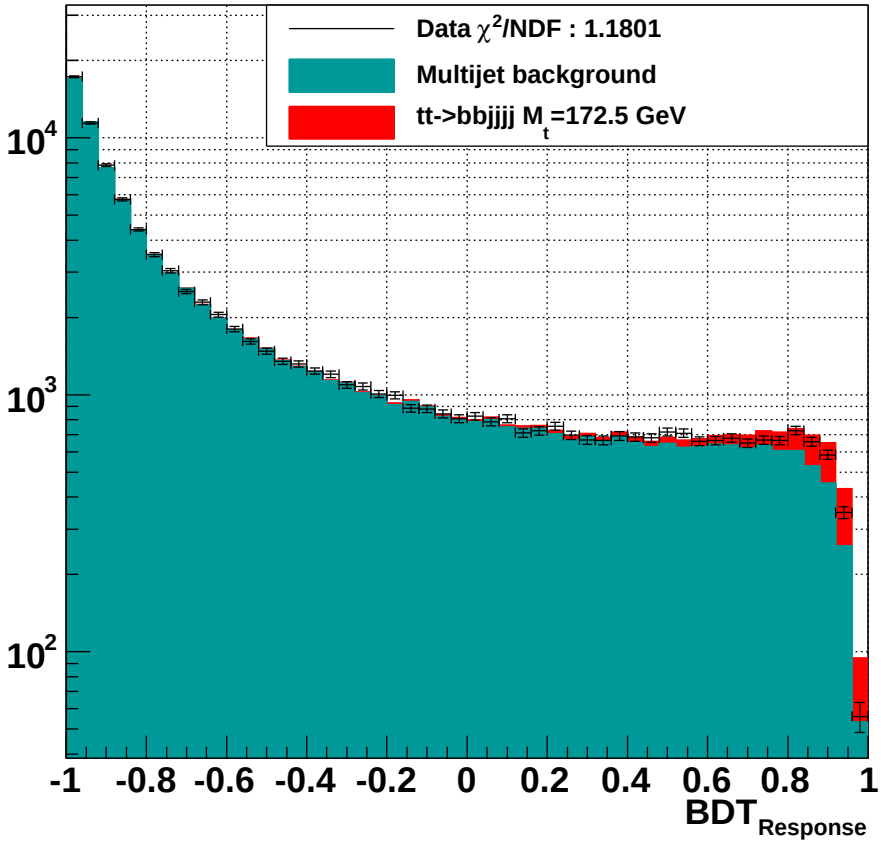


Figure 5.3: BDT response distribution for signal, background and data.

As mentioned in section 4.2, signal events are generated for top quark masses of 150, 160, 165, 170, 172.5, 175, 180, 185, 190 and 195 GeV. In addition to the above, the samples needed to set a limit on the uncertainty due to the jet energy scale calibration is obtained by shifting the jet energies of the signal Monte Carlo events by ± 2.5% of its energy. The method employes eleven jet energy scale shifted samples (-4,-,3,-2,-1,-0.5,0,+0.5,+1,+2,+3,+4 where +1 represents the increase of the jet energy by 2.5% , -1 represents a decrease of jet energy by 2.5% etc.). This shift of ± 2.5% is quoted as σ_{JES} in the rest of this thesis. For example $+1\sigma_{JES}$ represents a sample of which energies of the all the jets are shifted by 2.5%.

The template method is developed for events with exactly six jets where two of them are identified as b-jets. These six jets generate six trijet ($b_1 j_1 j_2$ and $b_2 j_3 j_4$ where b_1 and b_2 are b-jet and j_i are light jets) combinations (Fig. 5.4). Each combination is minimized with respect to m_t^{rec}

$$\chi^2_{top} = \frac{(m^1_{jj} - 80.4)}{\sigma^2_{jj}} + \frac{(m^2_{jj} - 80.4)}{\sigma^2_{jj}} + \frac{(m^1_{bjj} - m^{rec}_t)}{\sigma^2_{bjj}} + \frac{(m^2_{bjj} - m^{rec}_t)}{\sigma^2_{bjj}} \qquad (5.2)$$

where, m_{jj} , m_{bjj} are dijet and trijet invariant masses, and σ_{jj} , σ_{bjj} are their resolutions obtained using Monte Carlo. Here, 80.4 GeV is the measured mass of the W boson. This figure being already measured to a better accuracy enhances the resolution of the method by constraining the dijet mass. This is apparent in the χ^2 distribution (Fig. 5.5), where it is evident that the signal events are peaked towards the low χ^2 region while the background show no such pattern. The reconstructed top quark mass for the permutation with minimum χ^2/NDF enters the top quark mass templates.

The W boson mass templates are constructed by minimizing

$$\chi^2_W = \frac{(m^1_{jj} - m^{rec}_W)}{\sigma^2_{jj}} + \frac{(m^2_{jj} - m^{rec}_W)}{\sigma^2_{jj}} + \frac{(m^1_{bjj} - m^{rec}_t)}{\sigma^2_{bjj}} + \frac{(m^2_{bjj} - m^{rec}_t)}{\sigma^2_{bjj}} \qquad (5.3)$$

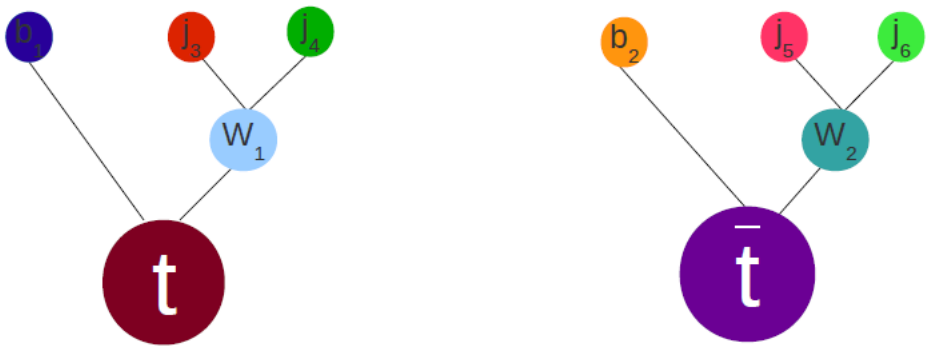
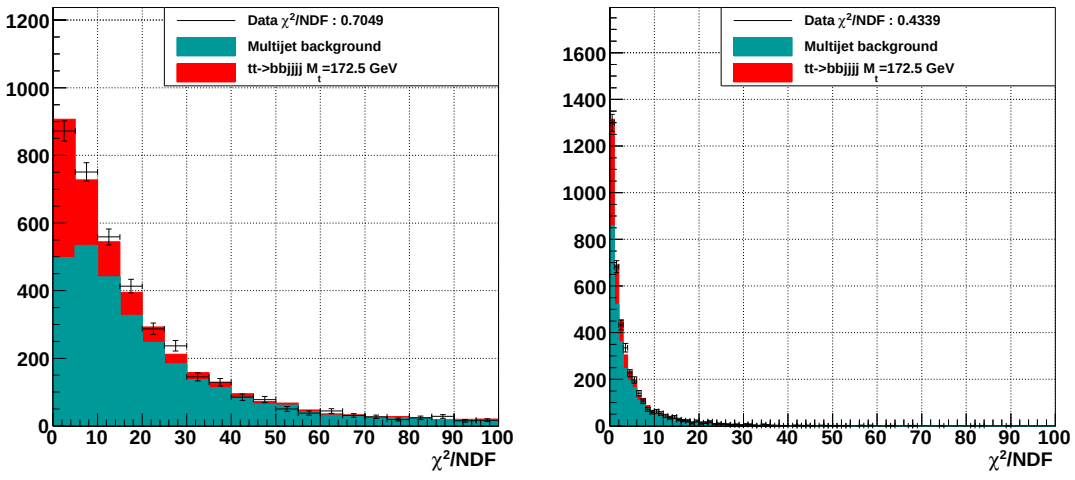
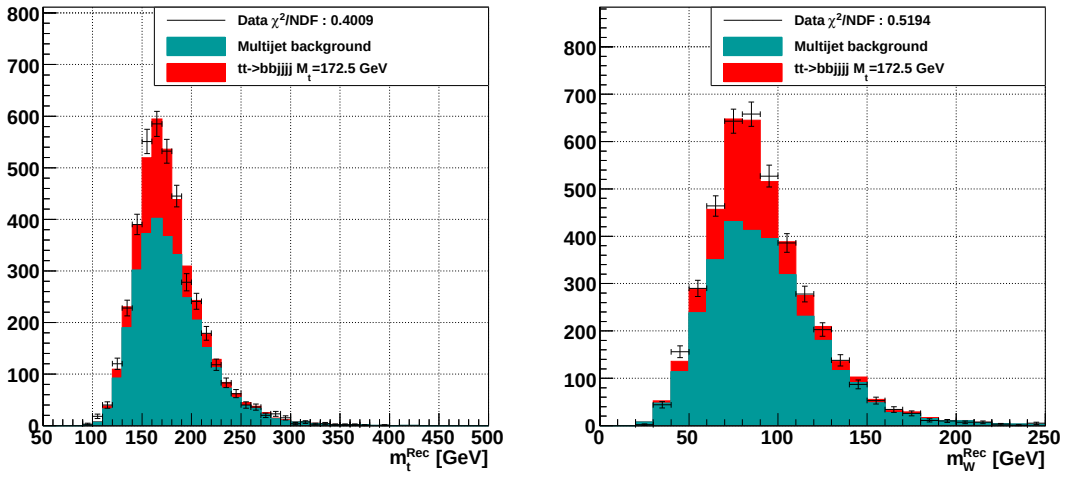


Figure 5.4: One of the combination out of six combinations to assign six jets to $t\bar{t}$.



(a) (b)

Figure 5.5: (a) χ^2/NDF for top quark mass templates. (b) χ^2/NDF for W boson mass templates



(a) (b)

Figure 5.6: (a) top quark mass template of signal and background plotted with data. (b) W boson mass template of signal and background plotted with data.

with respect to both m_t^{rec} and m_W^{rec} . For this instance, the reconstructed W boson mass (m_W^{rec}) of the combination with minimum χ^2 enters the W boson mass templates. Figure. 5.6 shows the signal and background top quark mass and W boson mass templates plotted with data. Figure. 5.7 and 5.8 show the signal and background templates along with the smoothed probability density functions.

5.3 Likelihood Technique

The unbinned likelihood technique compares the signal and background probability distribution functions to data. It allows a series of measurements to be combined in order to estimate the most probable value of an observable known as the estimator. Hence, this method can be applied to the data collected by the DØ detector to extract the top quark mass.

The probability density function for a variable x which depends on a set of parameters $\overline{\theta}$, can be written as $P(x|\overline{\theta})$. The joint probability density function for N independent sets of observables of x $(x_1, x_2, x_3, ..., x_N)$ is the product of the

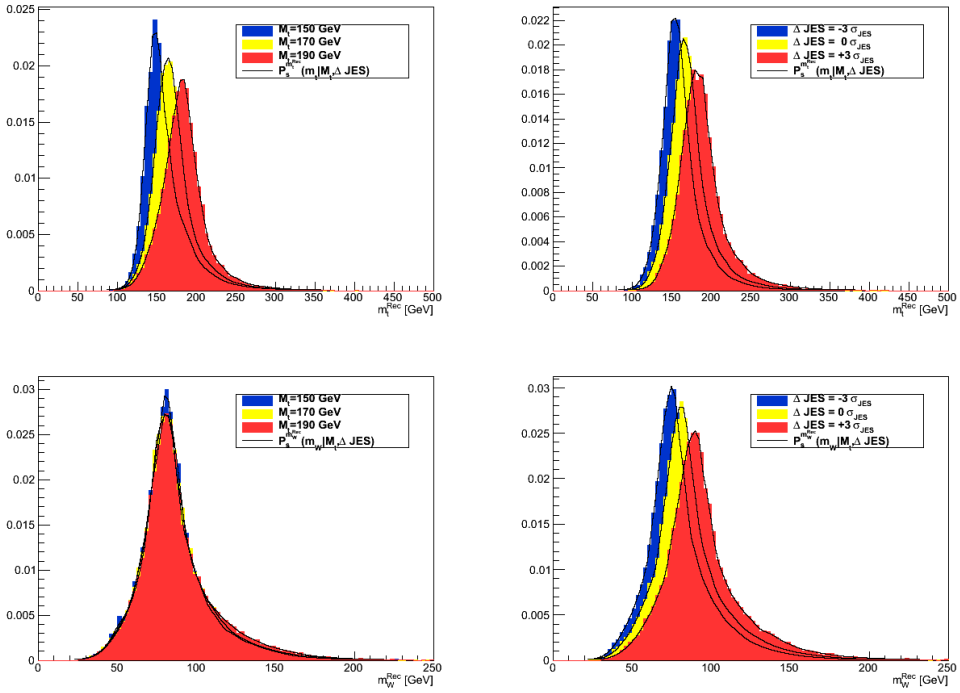


Figure 5.7: Top left: probability density functions of reconstructed top quark mass for the best χ^2 for different input top masses ($\Delta JES = 0\ \sigma_{JES}$). Top right: probability density functions of reconstructed top quark mass for different input ΔJES for the best χ^2 (M_t=172.5 GeV). Bottom left: probability density functions of reconstructed W boson mass for the best χ^2 for different input top quark masses($\Delta JES = 0\ \sigma_{JES}$). Top right: probability density functions of reconstructed W boson mass for different input ΔJES for the best χ^2 (M_t=172.5 GeV).

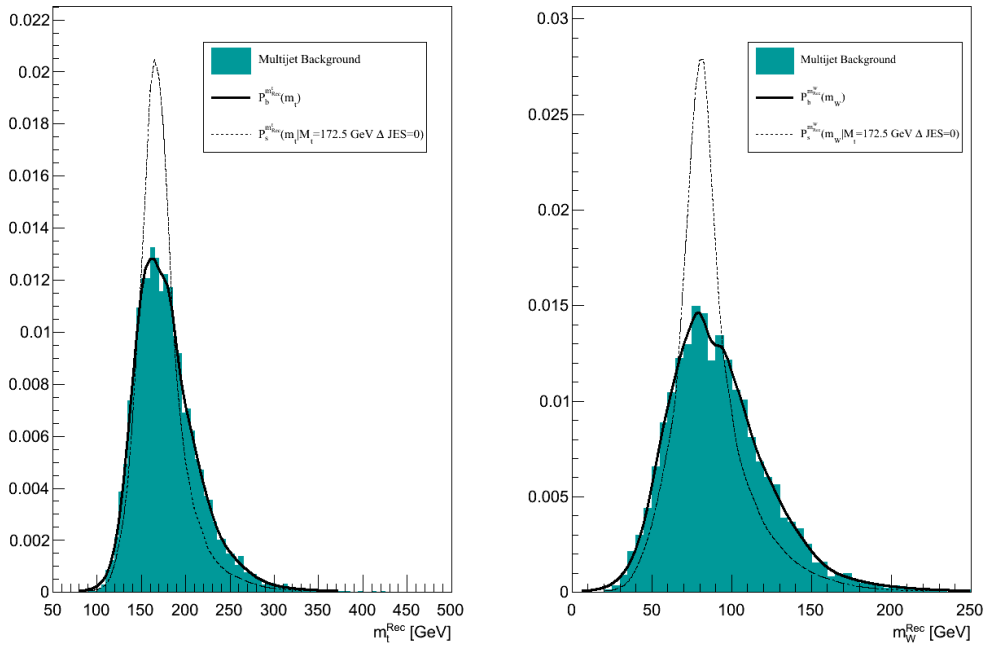


Figure 5.8: Left plot :Probability density function of the reconstructed top quark mass for background built from the best χ^2. Right plot:Probability density functions of the reconstructed W boson mass for background built from the best χ^2. The templates drawn with the dashed lines are for signal (M_t= 172.5 GeV and ΔJES= 0 σ_{JES}) to compare the differences in shape for signal and background.

probability density of each variable ($P(x_1, x_2, x_3, ..., x_N|\overline{\theta}) = \prod_{i=0}^{N} P(y_i|\overline{\theta})$)). For the instance, where x_i are a set of observed data, the likelihood function is written as

$$L(x_1, x_2, x_3, ..., x_N|\overline{\theta}) = \prod_{i=0}^{N} P(y_i|\overline{\theta}). \tag{5.4}$$

The data events collected by the DØ detector are independent from each other. Hence, applying the same hypothesis to the measurement, the likelihood to observe a number of top quark mass ($m_{t,i}$) measurements N given the top quark mass (M_{top}) and jet energy scale shift (ΔJES) in data can be written as

$$L_{top} = \prod_{i=0}^{N} (n_s P_s^{m_t^{rec}}(m_{t,i}|M_{top}, \Delta JES) + (N - n_s) P_b^{m_t^{rec}}(m_{t,i}))/N \tag{5.5}$$

where, P_s and P_b are the signal and background probability density functions. Similarly, the likelihood to observe N number of W boson mass ($m_{W,i}$) measurements given the top mass (M_{top}) and jet energy scale shift (ΔJES) in data can be written as

$$L_{JES} = \prod_{i=0}^{N} (n_s P_s^{m_W^{rec}}(m_{W,i}|M_{top}, \Delta JES) + (N - n_s) P_b^{m_W^{rec}}(m_{W,i}))/N. \tag{5.6}$$

In addition, to improve the accuracy of the measurement, a Gaussian constraint is applied to the expected number of signal events

$$L_{n_{sig}} = e^{-\frac{(n_s - n_{cons})^2}{{\sigma_{n_s}}^2}} \tag{5.7}$$

where, n_{cons} =952 is set to the expected number of signal events. This is the number of signal events for M_t= 172.5 GeV left after the selection. This number may be influenced by the uncertainties due to the measured integrated luminosity

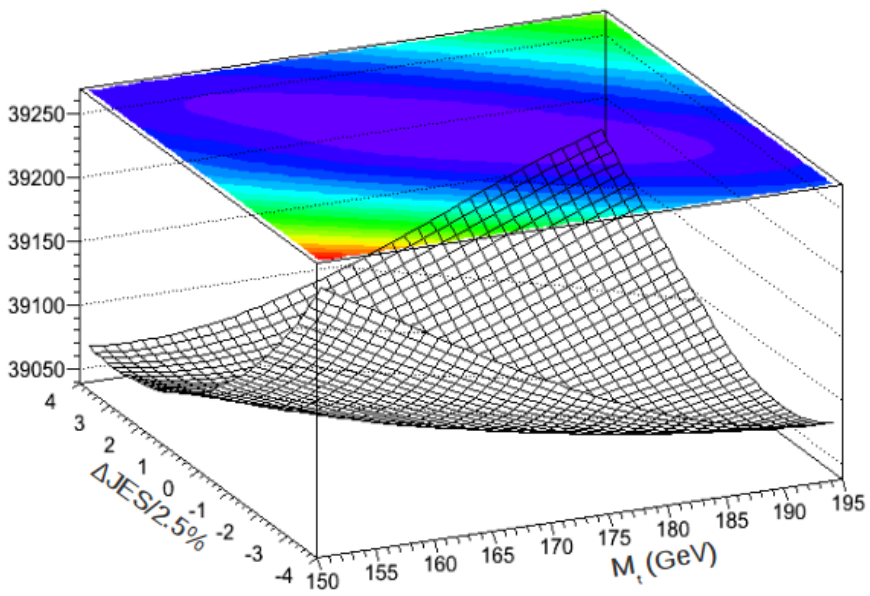


Figure 5.9: The three dimensional view of the 2-d graph constructed using the $-\log(L_{total})$ for each mass point and ΔJES point.

and the theoretical cross section. Hence, a systematic uncertainty will be assessed. The σ_{n_s} of the constraint is set to 10% of the expected number of signal events. The total likelihood is the product of the above three terms

$$L_{total} = L_{top} \times L_{JES} \times L_{n_{sig}} \tag{5.8}$$

The value of L_{total} is maximized or equivalently $-\log(L_{total})$ is minimized to find the estimators for M_{top} and ΔJES and their uncertainties in a two step procedure. First $-\log(L_{total})$ is minimized with respect to expected number of signal events n_s using MINUIT. Then a two dimensional graph is created using the points obtained from the first step (Fig. 5.10). This graph is fitted by slices to obtain the estimators for true top quark mass and relative jet energy scale shift. The statistical uncertainties are obtained by estimating the variation of the measured quantities when $-\log(L_{total,min}) + 1/2$.

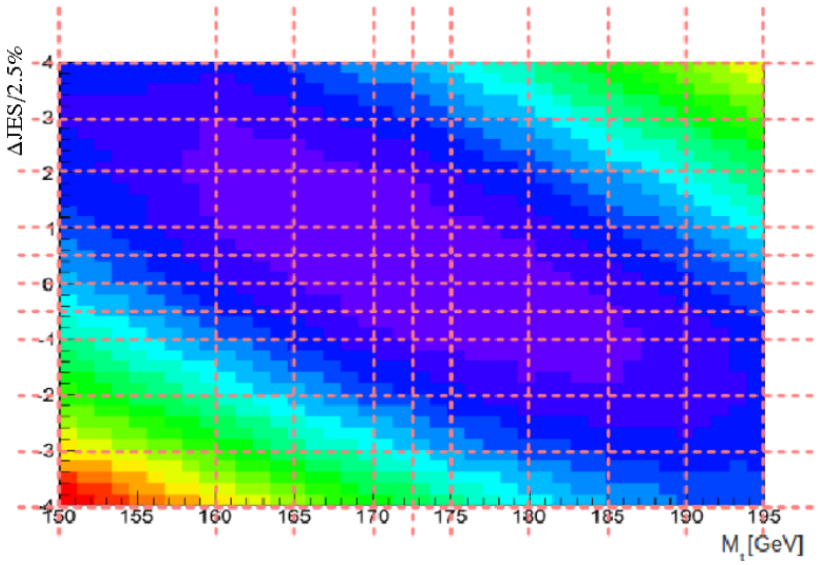


Figure 5.10: The two dimensional view of the 2-d graph constructed using the $-\log(L_{total})$ for each mass point and ΔJES point. Graph is constructed with 110 points, where dotted pink lines cross, M_t = 150, 160, 165, 170, 172.5, 175, 180, 185, 190, 195 GeV and ΔJES= -4,-3,-2,-1,-0.5,0,+0.5,+1,+2,+3,+4 σ_{JES}.

5.4 Validation of the Method and Calibration

The validity of the method is evaluated by performing a large number of simulated pseudo-experiments for different input top quark masses and ΔJES points. This approach of testing the validity of the method is called ensemble testing.

Each pseudo-experiment is implemented by comparing a pseudo-data set to the available templates using the likelihood technique as stated in section 5.3. The pseudo-data set is constructed with simulated signal and background events such that the total number of pseudo-events is equal to the number of data events. The number of signal events are chosen from a Poisson distribution centered at the expected number of signal events to accommodate statistical fluctuations. The pseudo-signal events are chosen in such a manner that the probability of an event with higher weight entering a pseudo-data set is higher than that of an event with a lower weight. The pseudo-background events are selected randomly. The fitted

mass, ΔJES, their uncertainties and the pull (Equation 5.9) is calculated via the likelihood method. The pull distribution shows the variation of the fitted quantity from its expected true value and is defined as

$$pull = \frac{m_t^{meas} - M_t}{\sigma_{m_t}}. \tag{5.9}$$

The results of the ensemble tests for 250 pseudo-experiments for fitted mass, fitted jet energy scale shift, uncertainties and pull distributions are shown in Fig. 5.11-5.27.

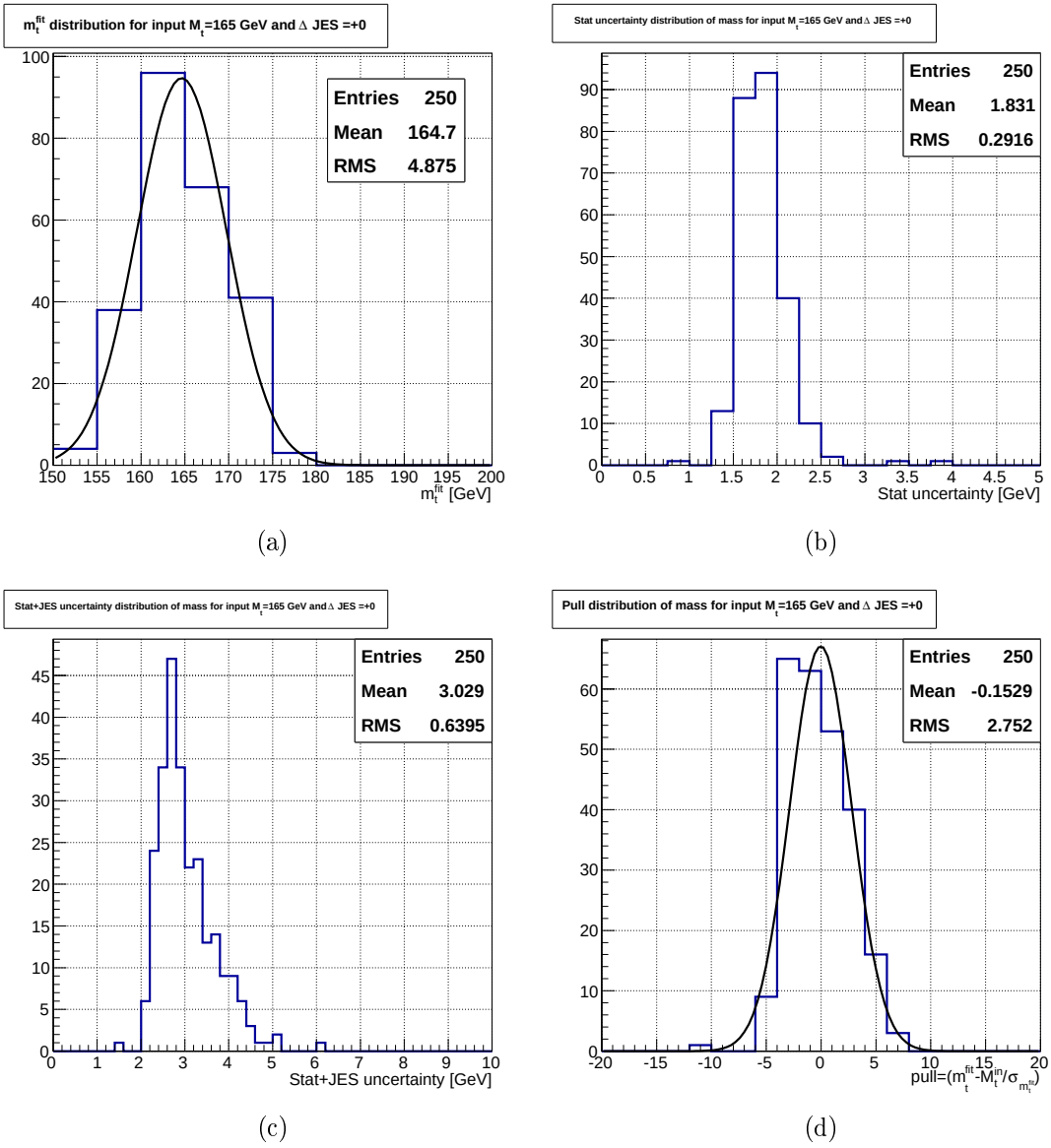


Figure 5.11: (a) Fitted mass distribution (b) Statistical uncertainty of measured mass (c) Stat+JES uncertainty of measured mass (d) Pull distribution of measured mass for input mass $M_t = 165$ GeV and $\Delta JES = 0$ σ_{JES}.

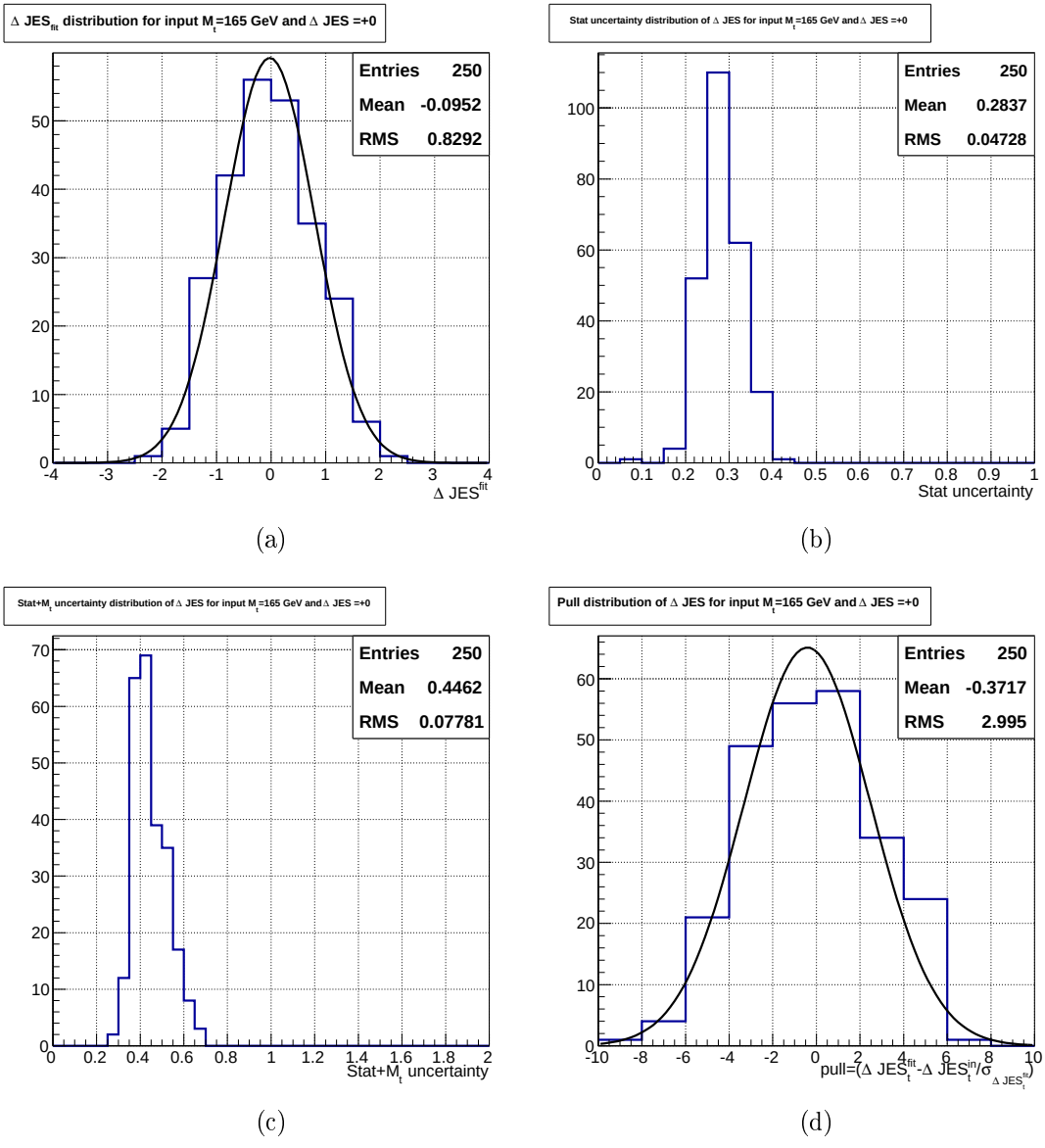


Figure 5.12: (a) Fitted JES distribution (b) Statistical uncertainty of measured JES (c) Stat+M_t uncertainty of measured JES (d) Pull distribution of measured JES for input mass $M_t = 165$ GeV and $\Delta JES = 0\ \sigma_{JES}$.

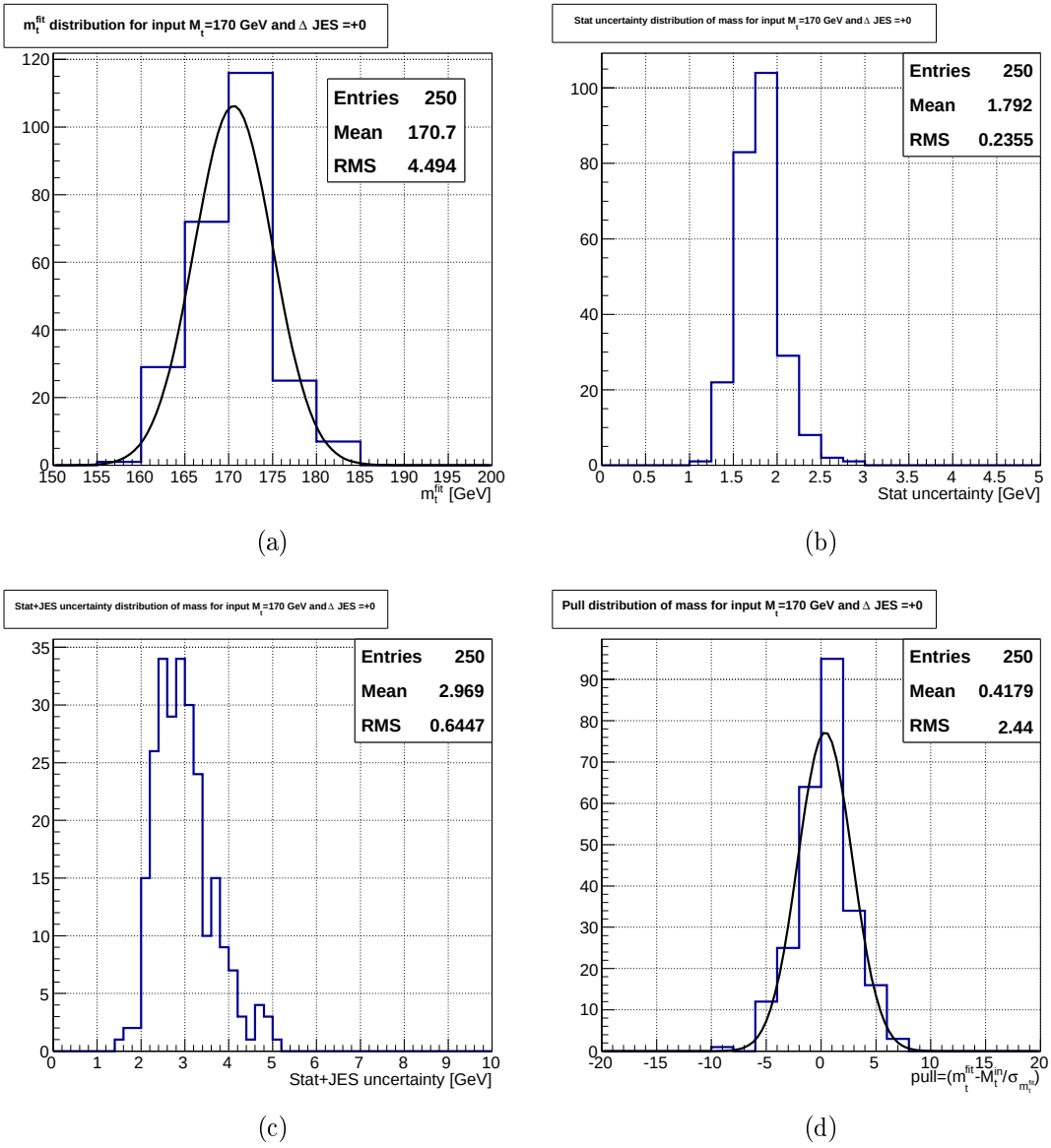


Figure 5.13: (a) Fitted mass distribution (b) Statistical uncertainty of measured mass (c) Stat+JES uncertainty of measured mass (d) Pull distribution of measured mass for input mass $M_t = 170$ GeV and $\Delta JES = 0$ σ_{JES}.

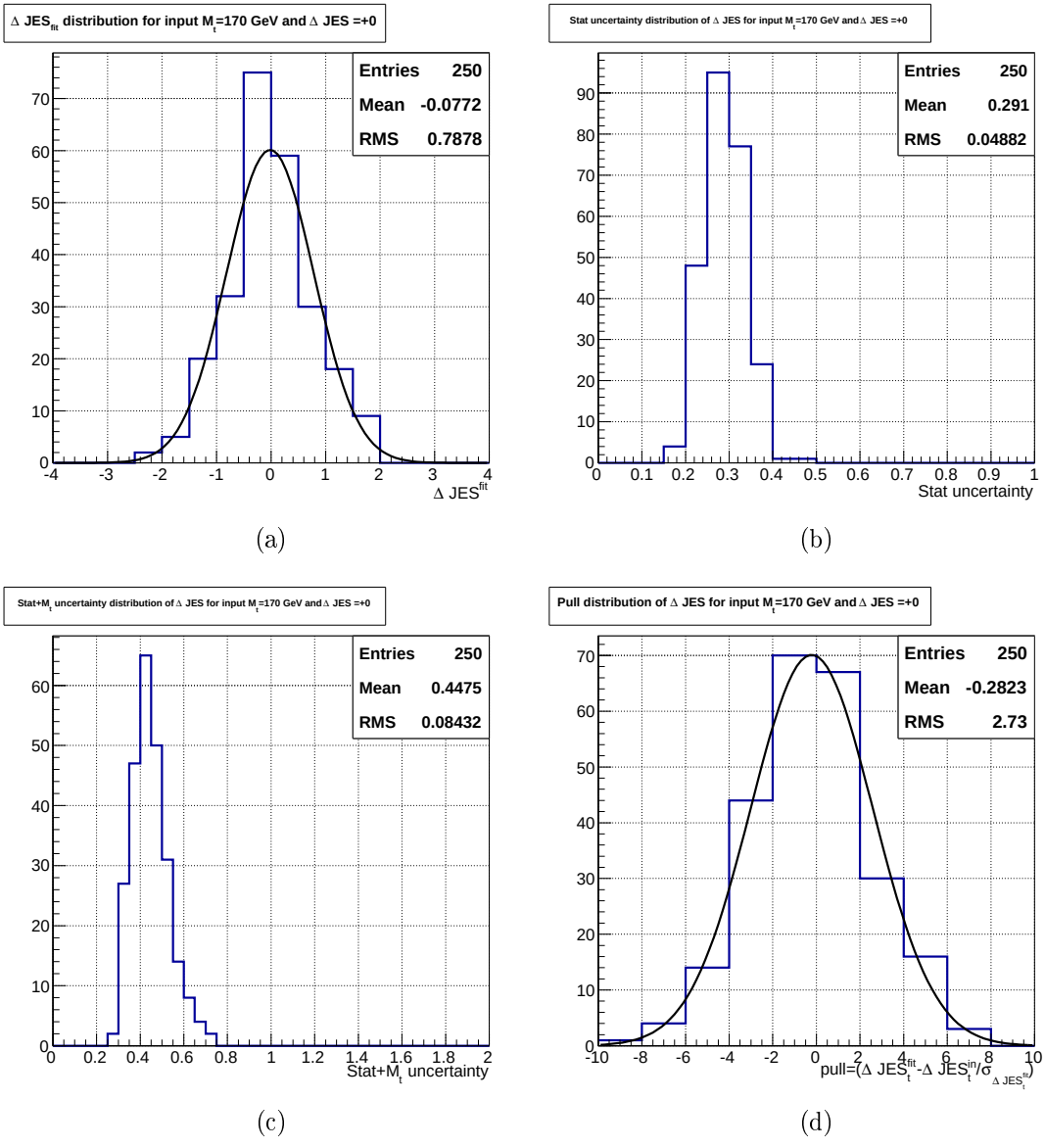
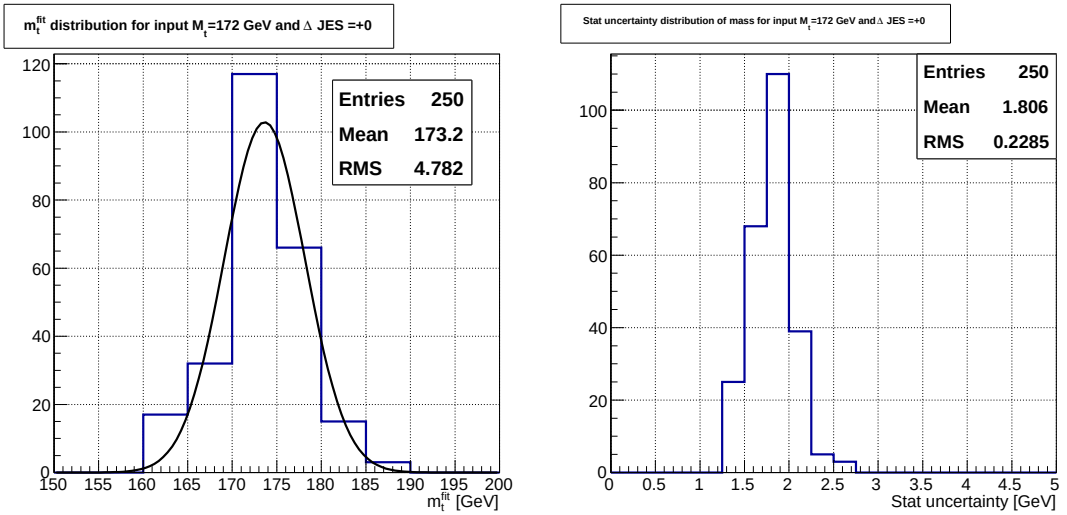


Figure 5.14: (a) Fitted JES distribution (b) Statistical uncertainty of measured JES (c) Stat+M_t uncertainty of measured JES (d) Pull distribution of measured JES for in put mass $M_t = 170$ GeV and $\Delta JES = 0$ σ_{JES}.



(a) (b)

(c) (d)

Figure 5.15: (a) Fitted mass distribution (b) Statistical uncertainty of measured mass (c) Stat+JES uncertainty of measured mass (d) Pull distribution of measured mass for input mass $M_t = 172.5$ GeV and $\Delta JES = 0$ σ_{JES}.

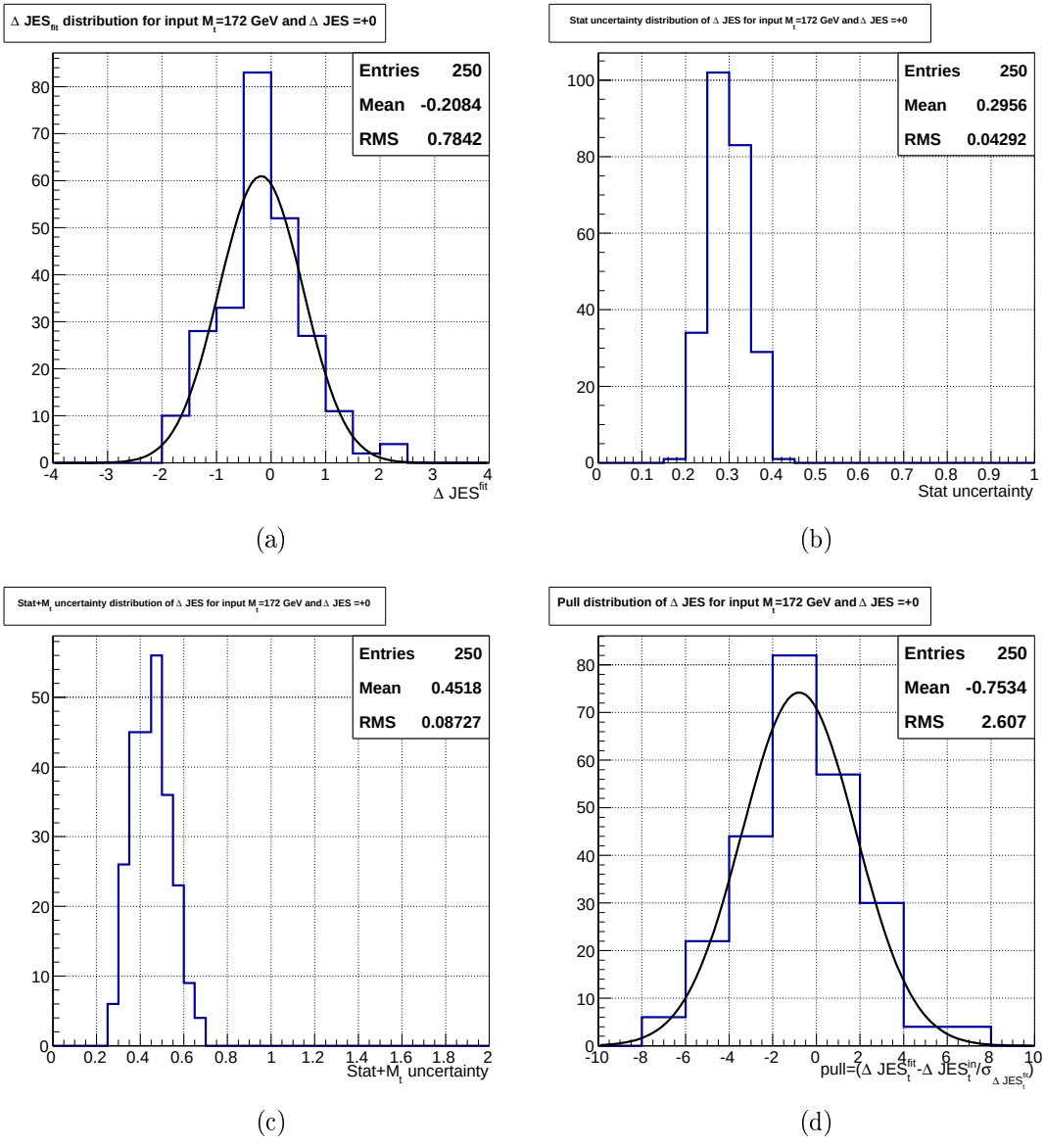


Figure 5.16: (a) Fitted JES distribution (b) Statistical uncertainty of measured JES (c) Stat+M_t uncertainty of measured JES (d) Pull distribution of measured JES for input mass $M_t = 172.5$ GeV and $\Delta JES = 0\ \sigma_{JES}$.

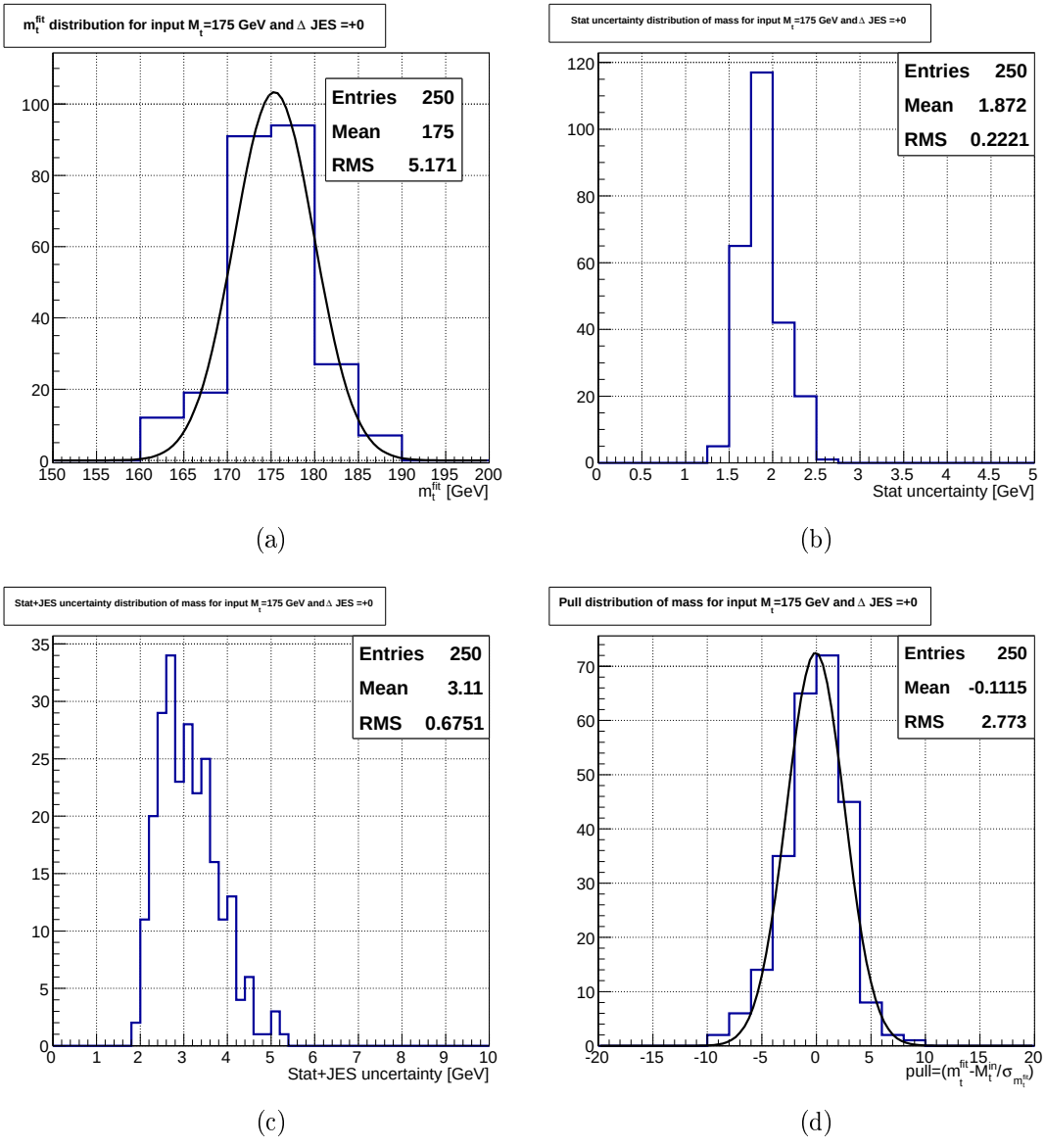


Figure 5.17: (a) Fitted mass distribution (b) Statistical uncertainty of measured mass (c) Stat+JES uncertainty of measured mass (d) Pull distribution of measured mass for input mass $M_t = 175$ GeV and $\Delta JES = 0$ σ_{JES}.

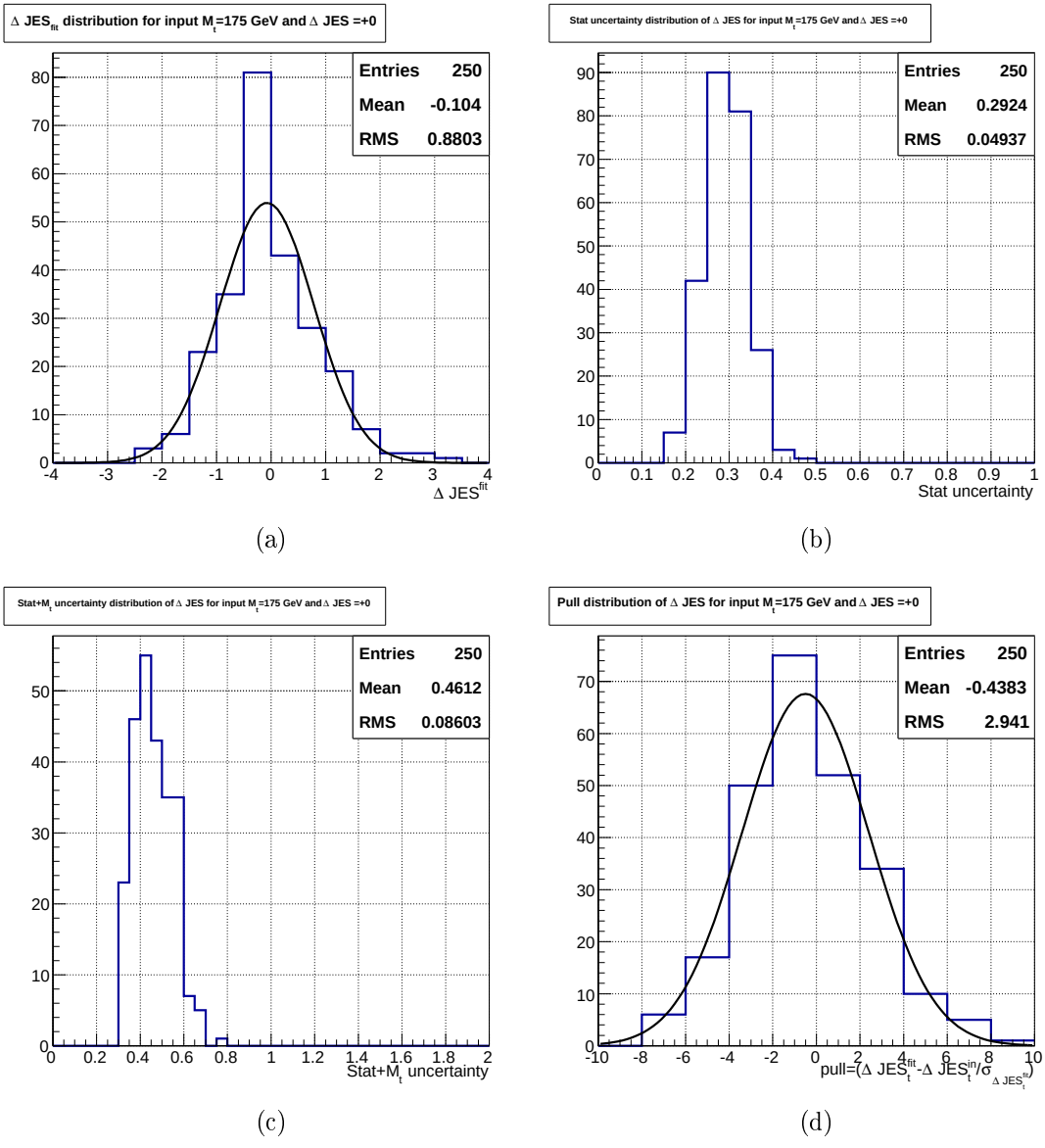


Figure 5.18: (a) Fitted JES distribution (b) Statistical uncertainty of measured JES (c) Stat+M_t uncertainty of measured JES (d) Pull distribution of measured JES for in put mass $M_t = 175$ GeV and $\Delta JES = 0$ σ_{JES}.

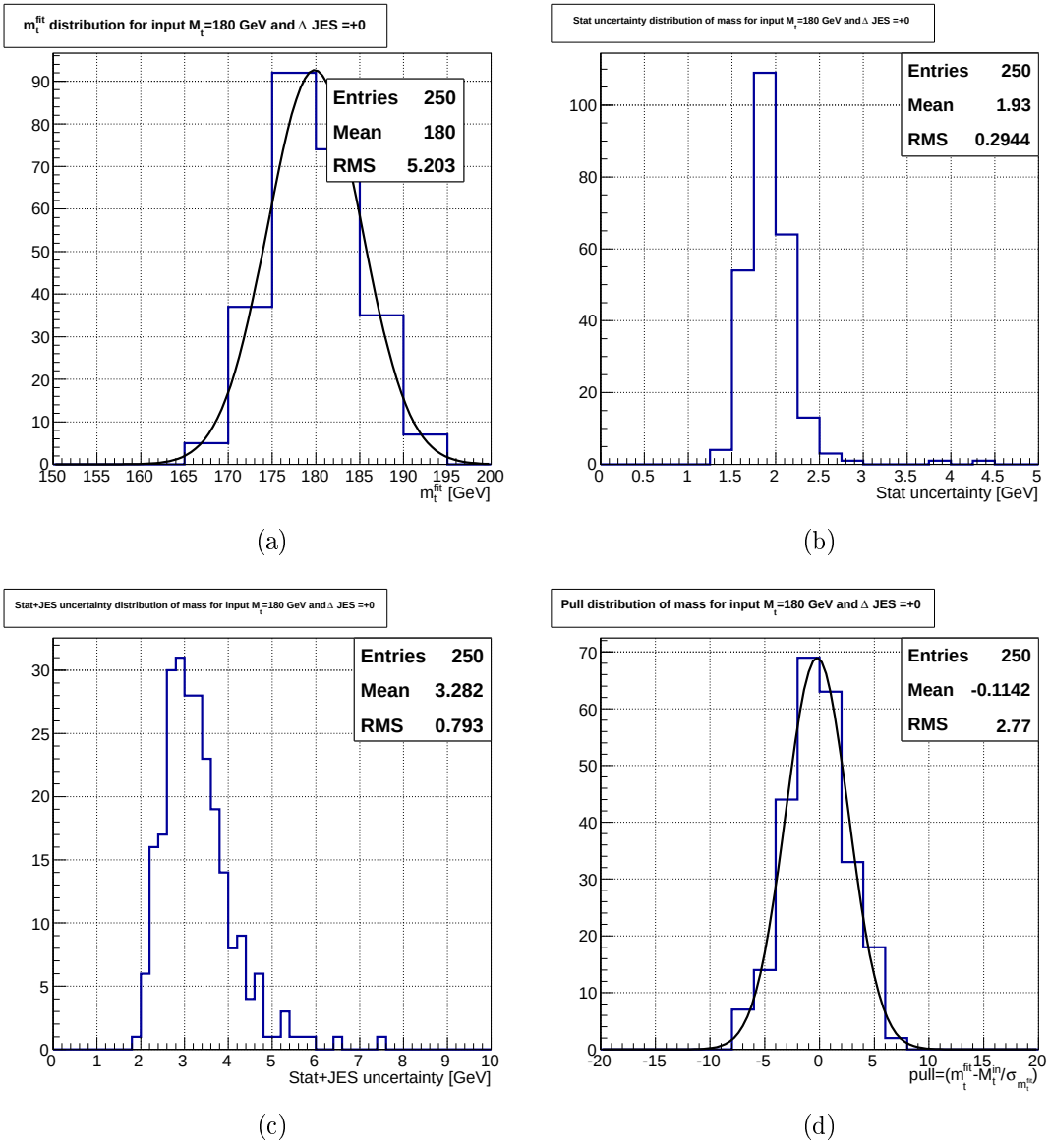


Figure 5.19: (a) Fitted mass distribution (b) Statistical uncertainty of measured mass (c) Stat+JES uncertainty of measured mass (d) Pull distribution of measured mass for input mass $M_t = 180$ GeV and $\Delta JES = 0$ σ_{JES}.

For the set of pseudo-experiments, the calibration curve along the mass axis for nominal ΔJES is shown in the Fig. 5.29. This shows very good agreement between the input top quark mass and the measured top quark mass. The calibration curve along the ΔJES axis is shown in Fig. 5.31 and is also in good agreement with the input ΔJES.

5.5 Measurement on the Data

The above described procedure is applied to data in order to extract the top mass and the jet energy scale shift. After the final selection there are 4033 data events. The signal and background top quark and W boson mass templates fitted to data are shown in Fig. 5.6.

The fitted top quark mass and the jet energy scale shift for data is found to be

$$M_t^{fitted} = 170.7 \pm 2.5 \text{ (stat+ JES) GeV}$$
$$\Delta JES^{fitted} = \text{-}0.2 \pm 0.44 \text{ (stat+ } M_t) \ \sigma_{JES}$$

Fitted values with contours corresponding to one, two and three standard deviations are shown in the Fig. 5.33.

The final calibrated top quark mass, jet energy scale shift and their uncertainties are

$$M_t^{calibrated} = 170.4 \text{ GeV} \pm 2.5 \text{ (stat+ JES) GeV}$$
$$\Delta JES^{calibrated} = \text{-}0.18 \pm 0.47 \text{ (stat+ } M_t) \ \sigma_{JES}$$

After separating the pure statistical uncertainty from the total uncertainty

$$M_t^{calibrated} = 170.4 \text{ GeV} \pm 1.7 \text{ (stat)} \pm 1.9 \text{ (JES) GeV}$$
$$\Delta JES^{calibrated} = \text{-}0.18 \pm 0.36 \text{ (stat)} \pm 0.30 \ (M_t) \ \sigma_{JES}$$

The observed jet energy scale shift is less than 0.5% of jet energy. Hence, this confirms the validity of the default jet energy scale calibration.

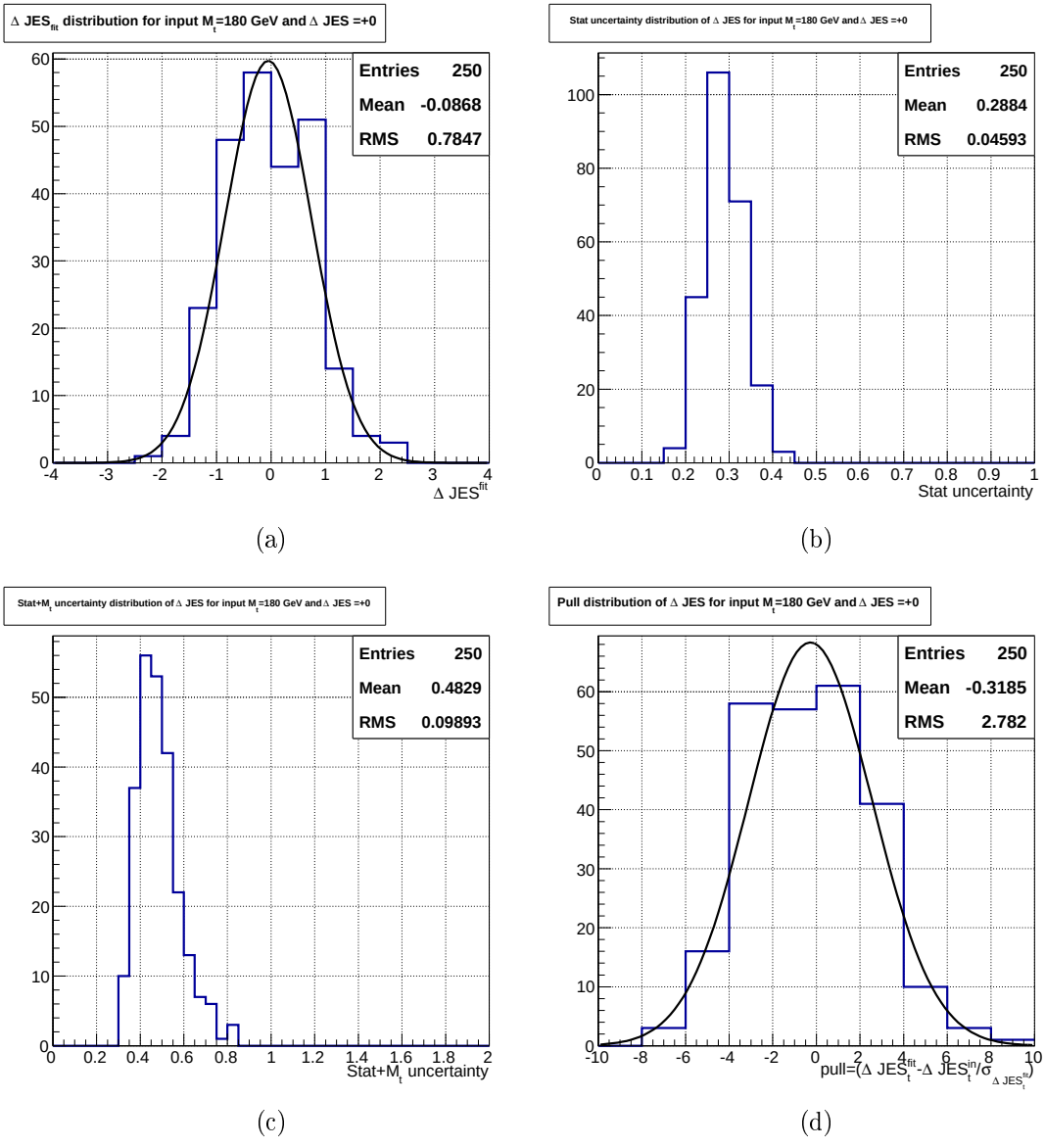


Figure 5.20: (a) Fitted JES distribution (b) Statistical uncertainty of measured JES (c) Stat+M_t uncertainty of measured JES (d) Pull distribution of measured JES for input mass $M_t = 180$ GeV and $\Delta JES = 0$ σ_{JES}.

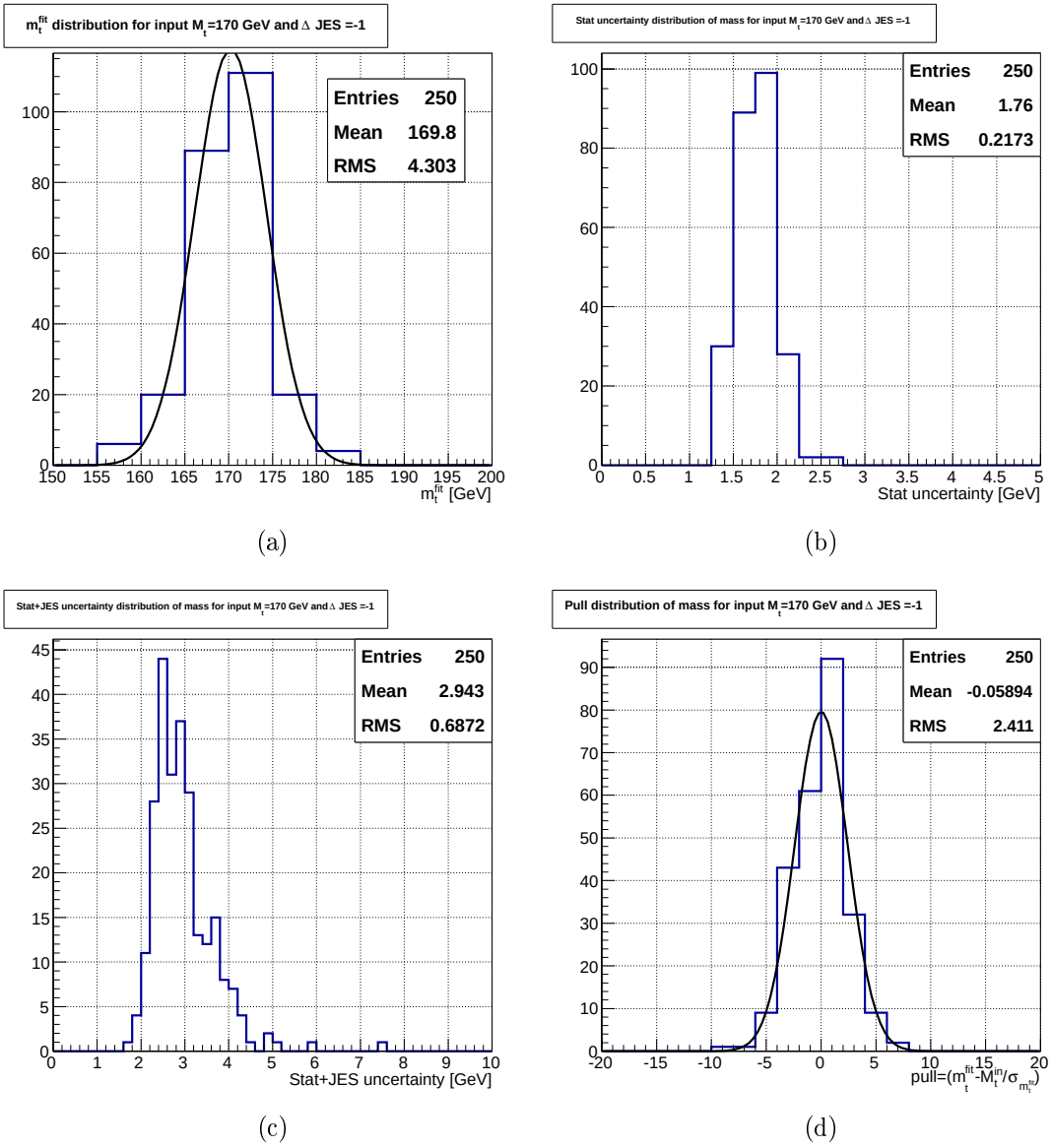


Figure 5.21: (a) Fitted mass distribution (b) Statistical uncertainty of measured mass (c) Stat+JES uncertainty of measured mass (d) Pull distribution of measured mass for input mass $M_t = 170$ GeV and ΔJES =-1 σ_{JES}.

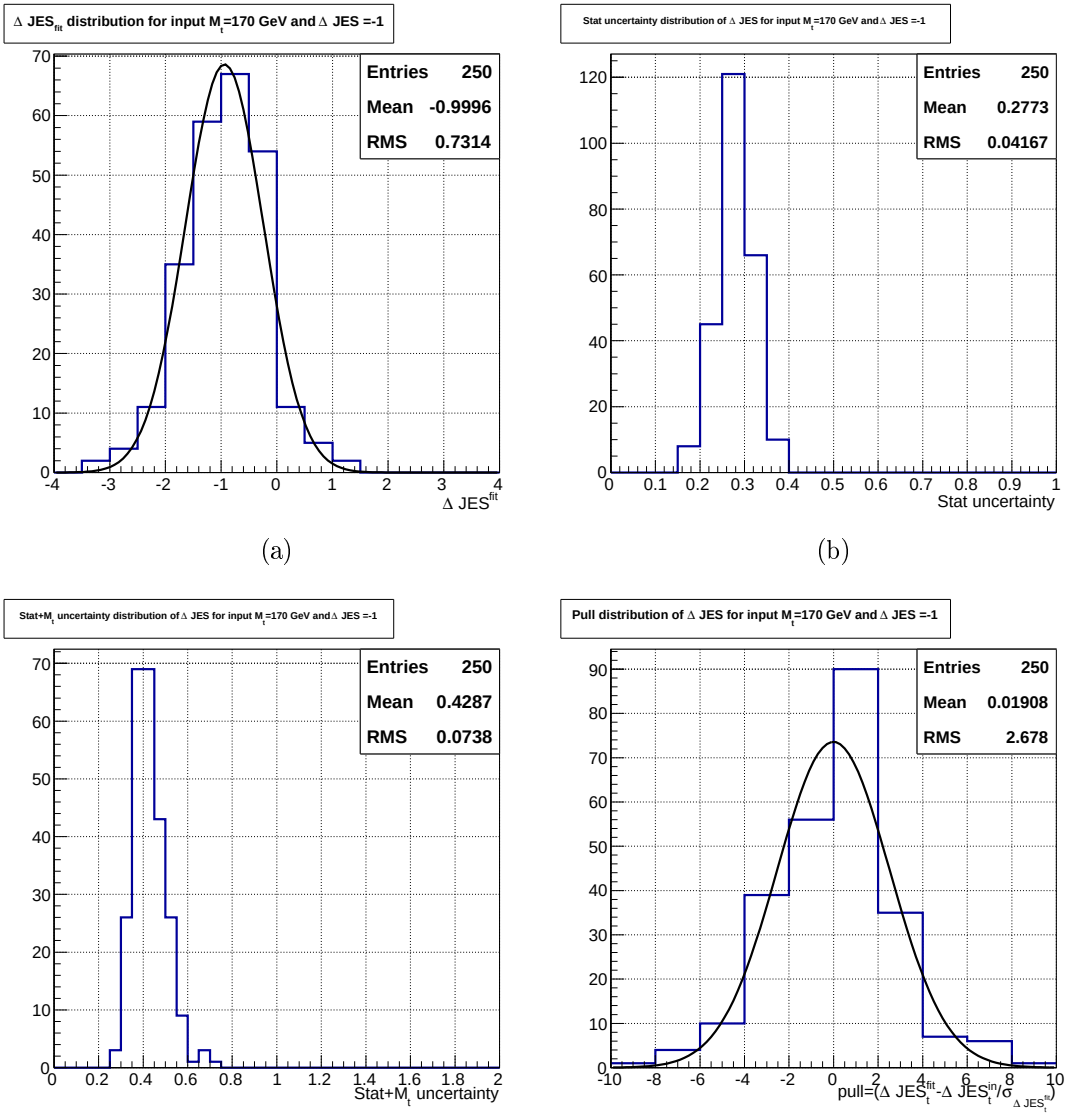


Figure 5.22: (a) Fitted JES distribution (b) Statistical uncertainty of measured JES (c) Stat+M_t uncertainty of measured JES (d) Pull distribution of measured JES for input mass $M_t = 170$ GeV and $\Delta JES = \text{-}1\ \sigma_{JES}$.

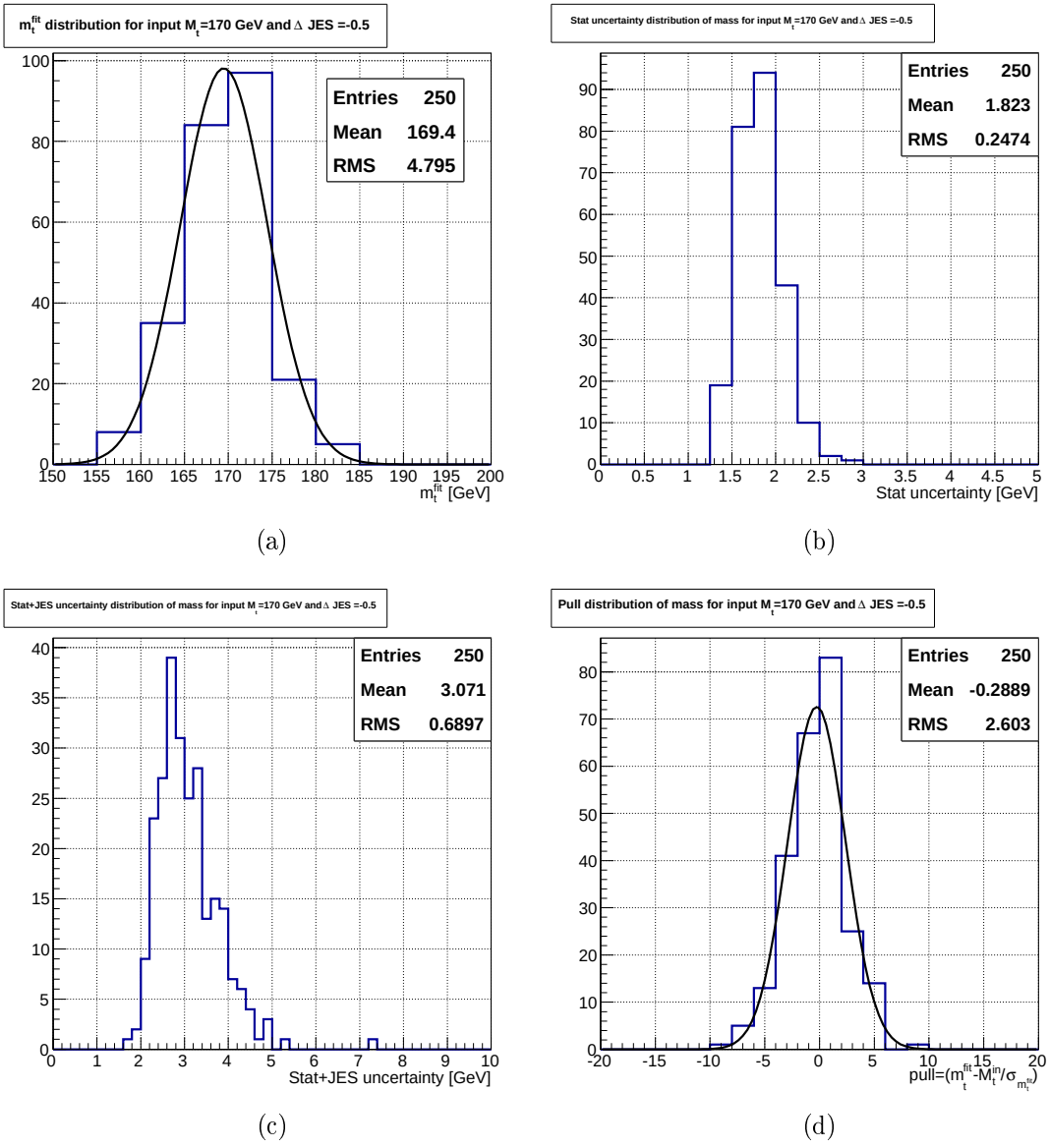


Figure 5.23: (a) Fitted mass distribution (b) Statistical uncertainty of measured mass (c) Stat+JES uncertainty of measured mass (d) Pull distribution of measured mass for input mass $M_t = 170$ GeV and $\Delta JES =$ -0.5 σ_{JES}.

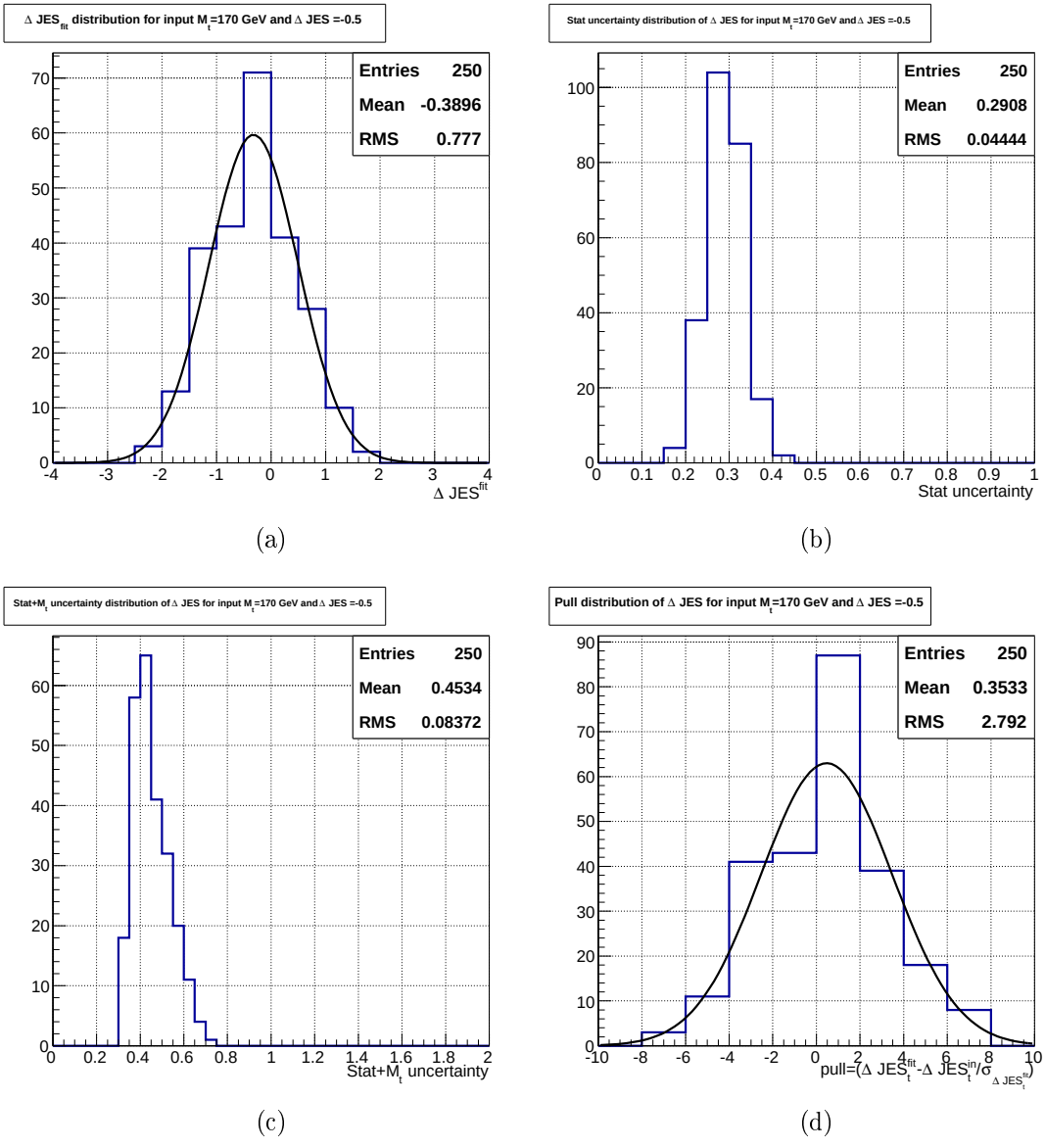


Figure 5.24: (a) Fitted JES distribution (b) Statistical uncertainty of measured JES (c) Stat+M_t uncertainty of measured JES (d) Pull distribution of measured JES for input mass M_t = 170 GeV and ΔJES = -0.5 σ_{JES}.

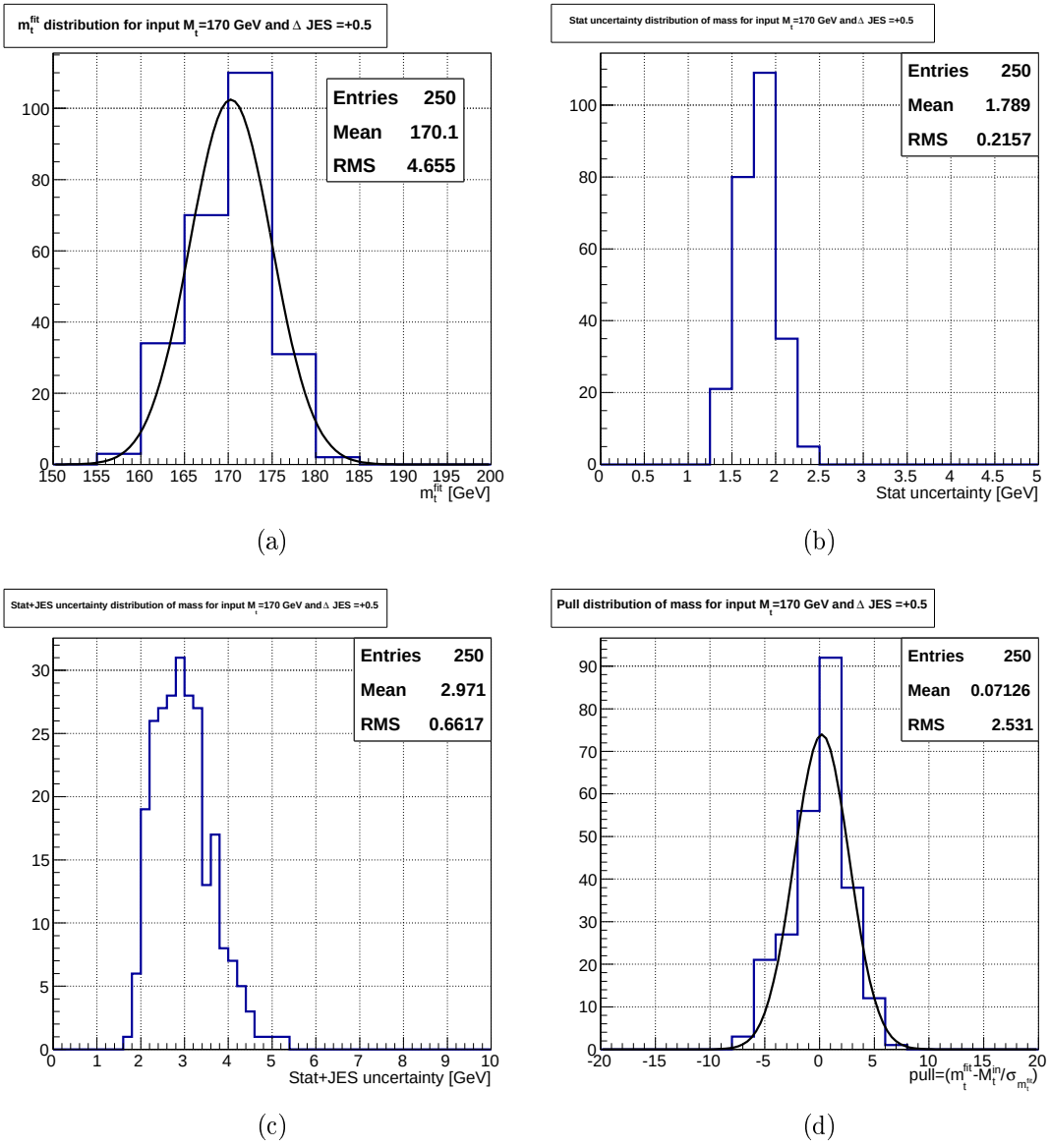


Figure 5.25: (a) Fitted mass distribution (b) Statistical uncertainty of measured mass (c) Stat+JES uncertainty of measured mass (d) Pull distribution of measured mass for input mass $M_t = 170$ GeV and $\Delta JES = +0.5\ \sigma_{JES}$.

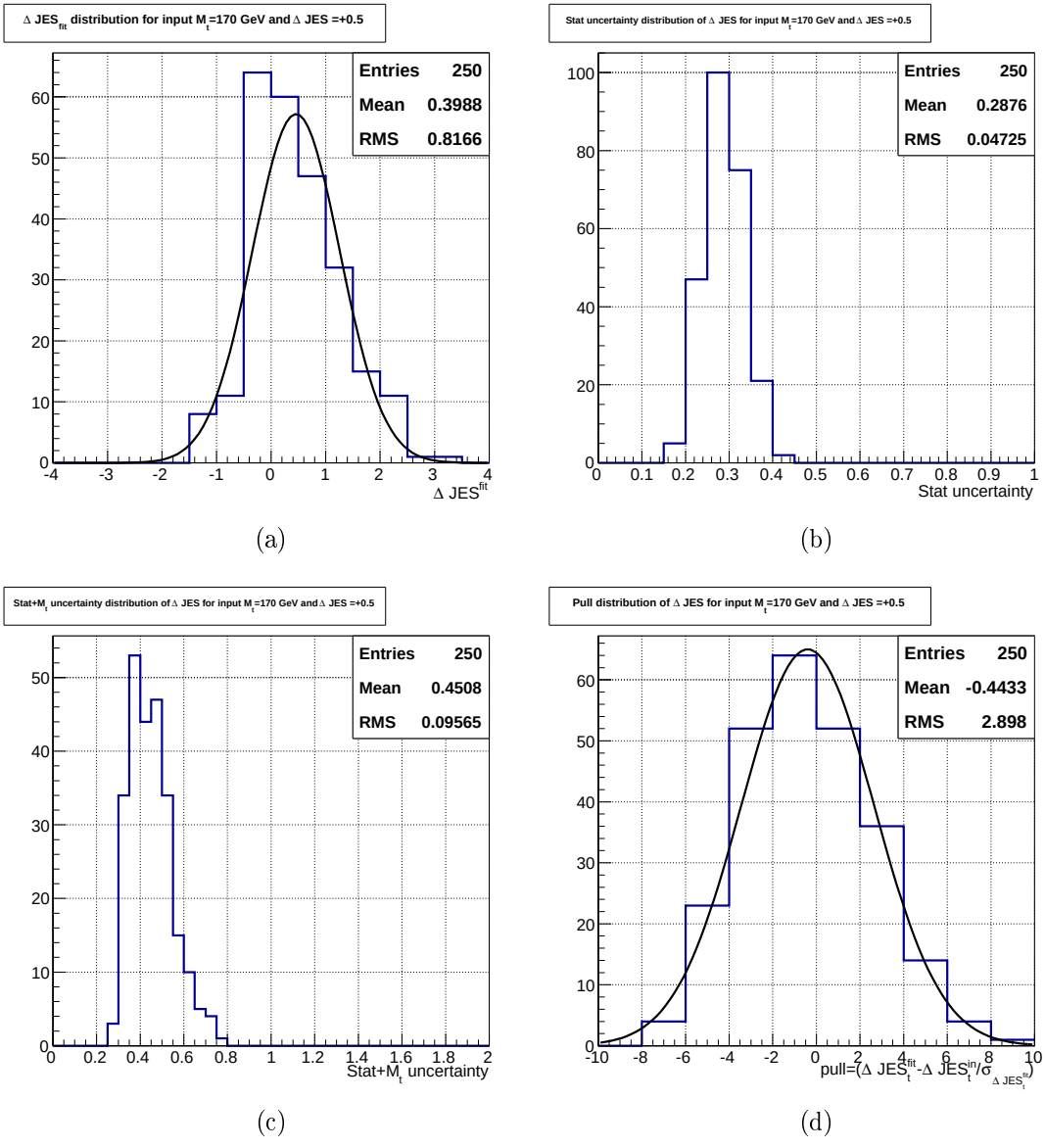


Figure 5.26: (a) Fitted JES distribution (b) Statistical uncertainty of measured JES (c) Stat+M_t uncertainty of measured JES (d) Pull distribution of measured JES for input mass $M_t = 170$ GeV and $\Delta JES = +0.5\ \sigma_{JES}$.

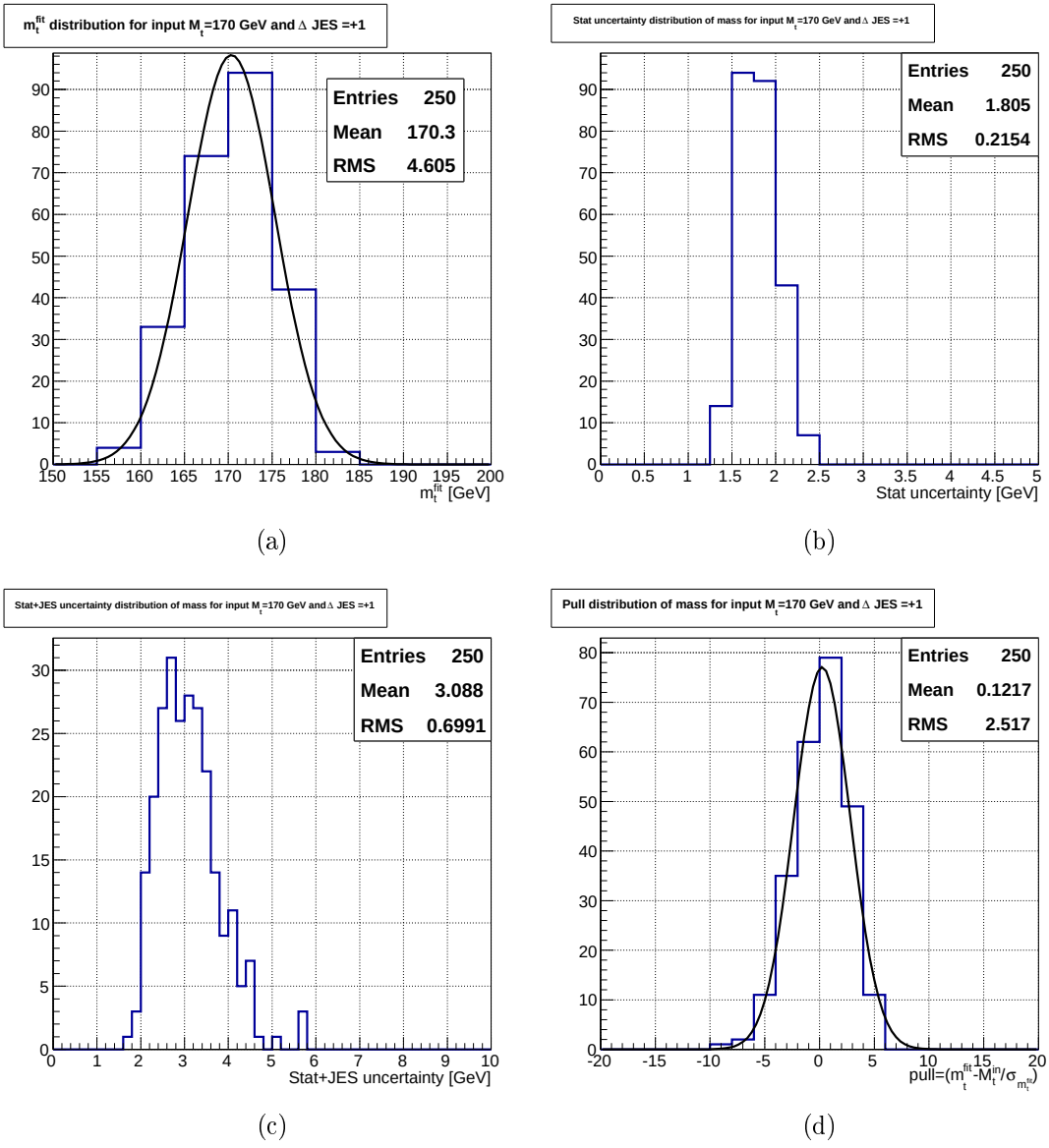


Figure 5.27: (a) Fitted mass distribution (b) Statistical uncertainty of measured mass (c) Stat+JES uncertainty of measured mass (d) Pull distribution of measured mass for input mass $M_t = 170$ GeV and ΔJES=+1 σ_{JES}.

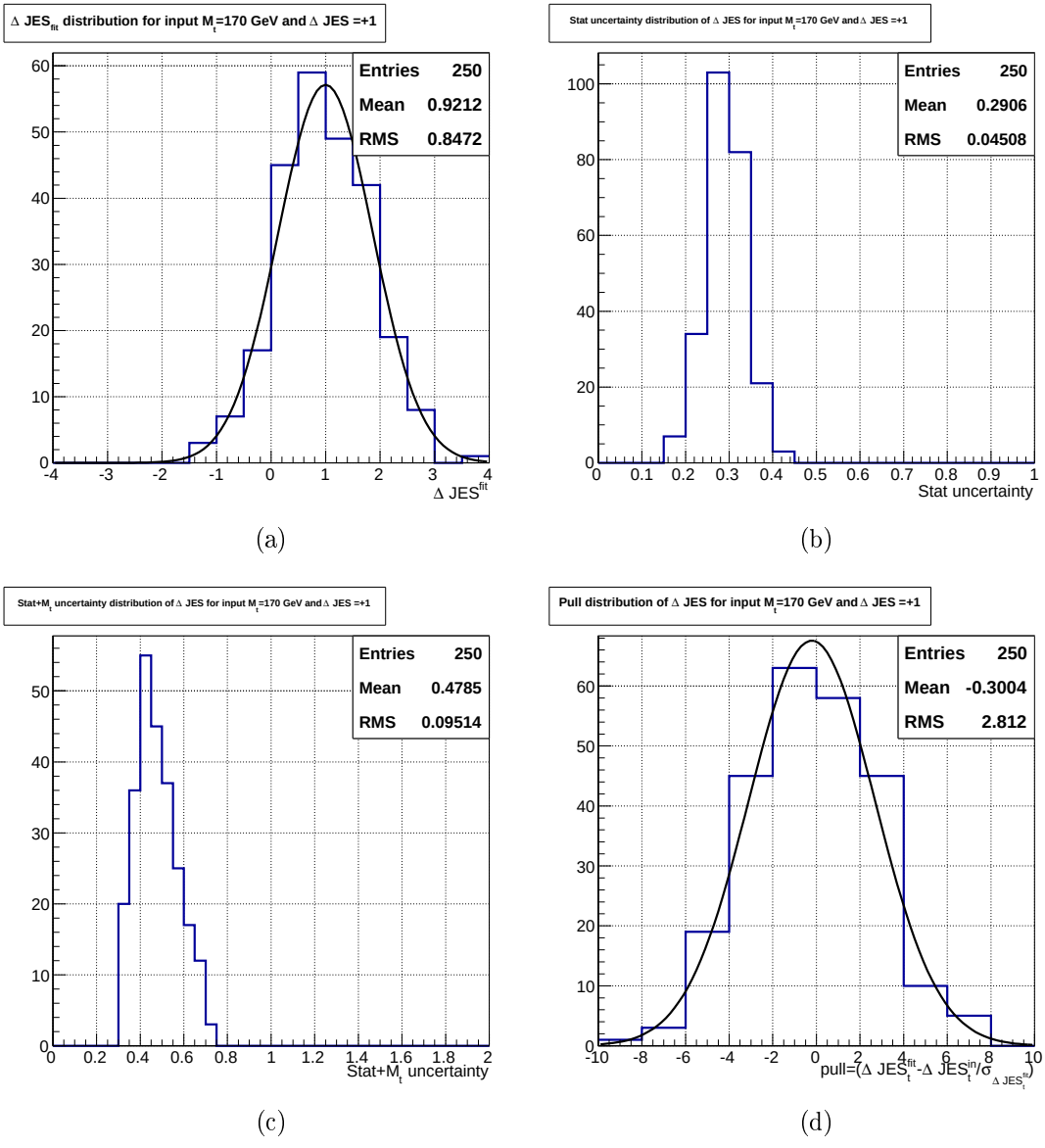


Figure 5.28: (a) Fitted JES distribution (b) Statistical uncertainty of measured JES (c) Stat+M_t uncertainty of measured JES (d) Pull distribution of measured JES for input mass $M_t = 170$ GeV and $\Delta JES = +1\ \sigma_{JES}$.

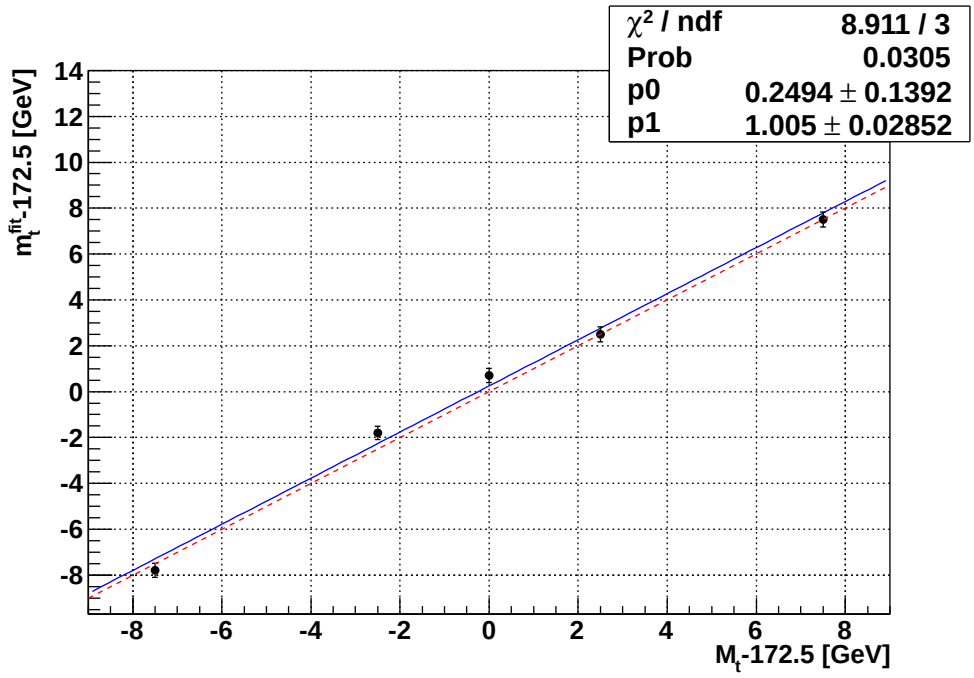


Figure 5.29: The calibration curve for mass for the nominal Jet Energy Scale (ΔJES= 0 σ_{JES}).

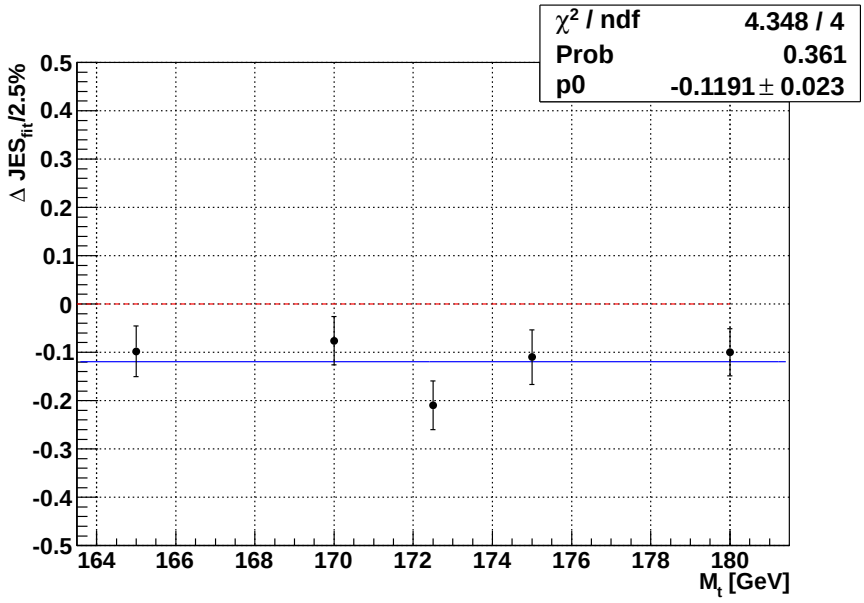


Figure 5.30: The fitted ΔJES for each mass point (input ΔJES =0 σ_{JES})

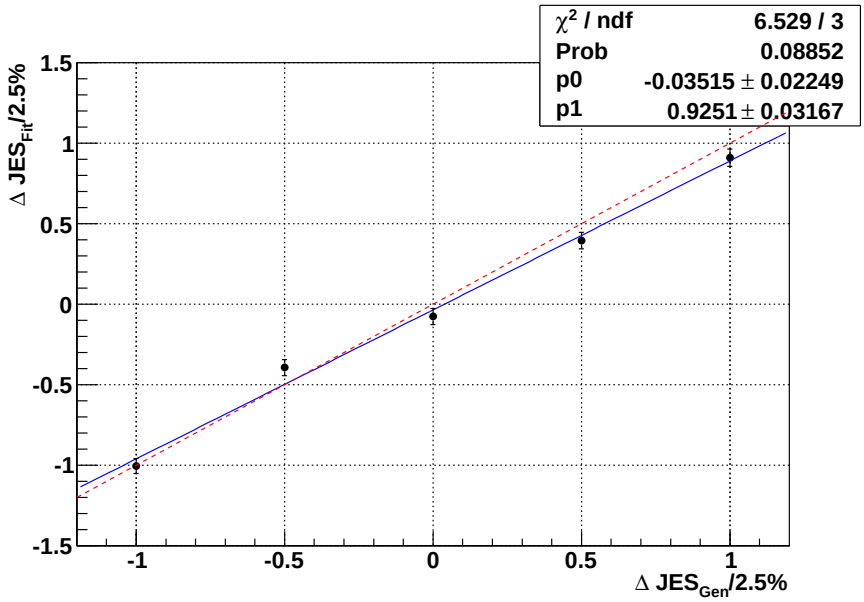


Figure 5.31: The calibration curve for ΔJES (input top mass M_t= 170.0 GeV).

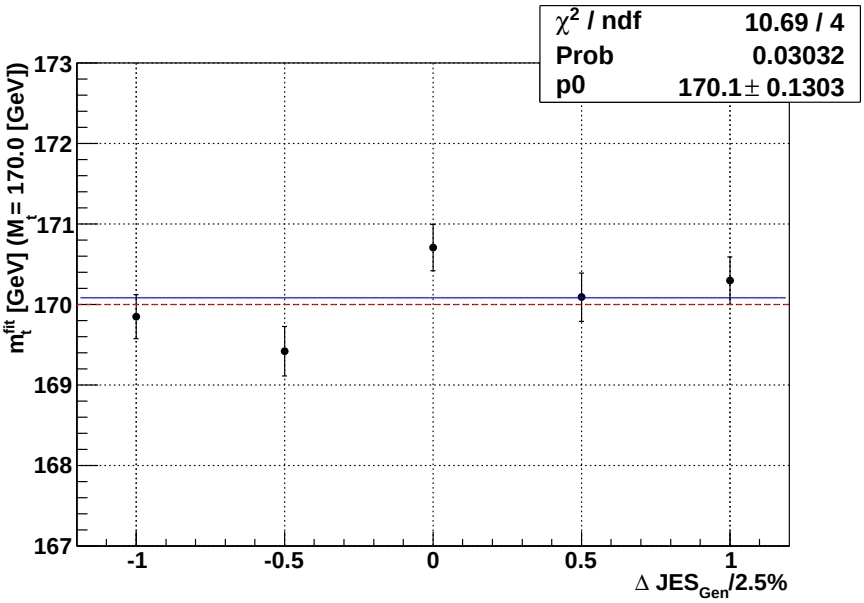


Figure 5.32: The fitted m_t for each ΔJES point (input top mass M_t= 170.0 GeV)

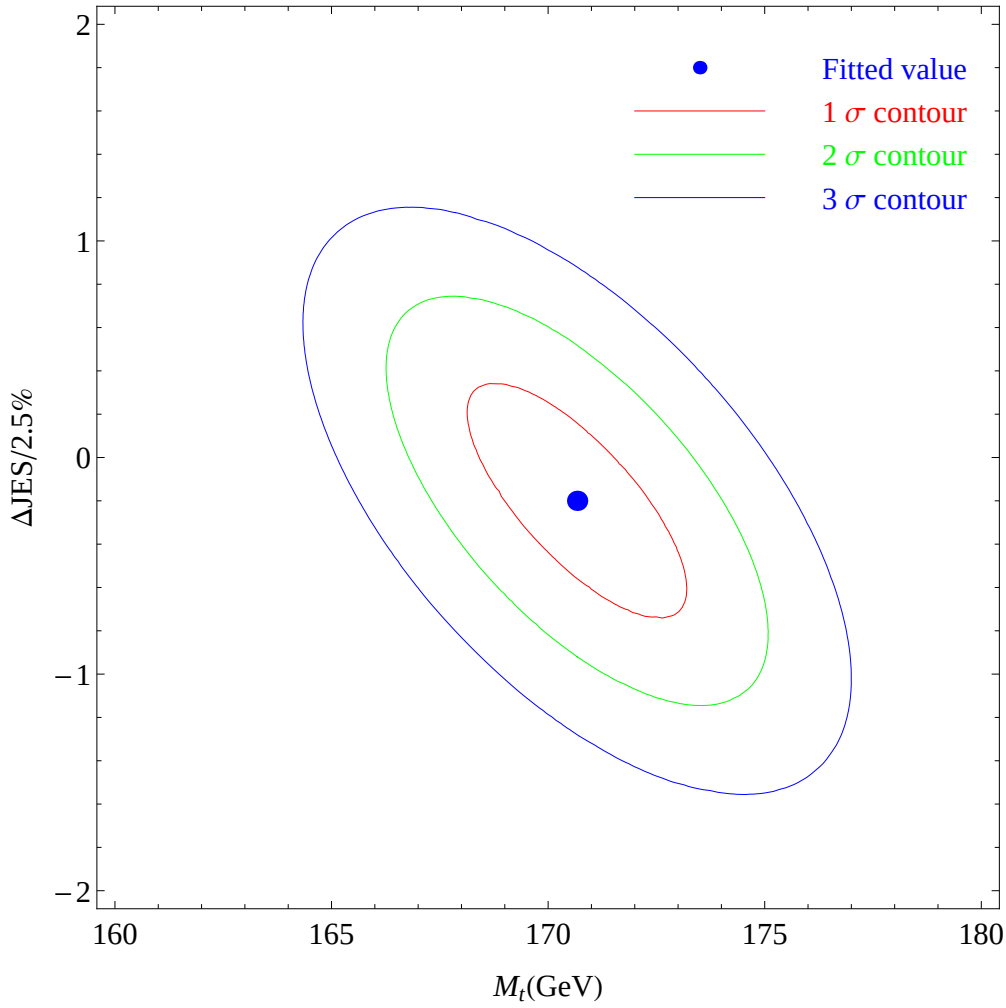


Figure 5.33: Fitted top mass and jet energy scale shift for data and contours corresponds to one, two and three standard deviations.

Chapter 6

Systematic Uncertainties

The estimated results may be influenced by numerous biases that are method specific and sample based. In this chapter, the origin of the possible biases and the procedure of assessing them is discussed.

The possible biases are assessed using pseudo-experiments by changing the templates and pseudo-data sets accordingly. For the expected biases arising due to the formulation of the method such as usage of BDTs and templates, the method is varied to obtain systematic uncertainties. Expected uncertainties due to the signal and background event generation are assessed by varying the parameters of the generators within their known uncertainties.

The signal sample of M_t= 172.5 GeV with default jet energy scale correction is used to assess all the systematic uncertainties. This sample is regenerated with desired variations and compared with the available templates to assess the systematic uncertainties. The systematic uncertainties are quoted based on the method the calculations are performed. For the uncertainties calculated using a different model or a few different models, the maximum difference between the default measurement and the desired variations are quoted as the systematic uncertainty due to the respective effect. The uncertainties assessed using samples generated with variations of $+\sigma$ and $-\sigma$ of a variable is quoted as the average between the two measurements or the maximum difference between the default measurement and

the variations [55].

6.1 Hadronization

The systematic uncertainty due to the usage of PYTHIA for simulation of the hadronization process is assessed using samples generated with HERWIG. The differences of the measurements between default (ALPGEN+PYTHIA) and the ALPGEN+HERWIG sample are quoted as the systematic uncertainties.

6.2 Higher Order Corrections

The ALPGEN generator uses the Leading Order Matrix element calculations to model the hard scattering. This is then input to PYTHIA for the hadronization. The higher order effects are modelled by the Next to Leading Order generator (NLO) MC@NLO[1]. The samples generated with the higher order effects are compared to the default template to calculate the shift in measured top mass and ΔJES. The systematic uncertainty due to the higher order effects are quoted as the difference between these measurements and the measurements obtained using ALPGEN+HERWIG sample to avoid double counting.

6.3 Color Reconnection

The color connections between the hard scattering process and the underlying event[2] are called color reconnections [55]. The systematic uncertainty due to color reconnections are assessed using PERUGIA[3] 2011 NOCR and PERUGIA 2011 CR sample. The difference between the two measurements for top quark mass and the ΔJES are quoted as the systematic uncertainties.

[1] MC@NLO uses HERWIG instead of PYTHIA to simulate the showering. Hence, Additional steps are necessary to correct for this secondary effect.

[2] the soft interactions in the collision that includes additional particles to the event.

[3] PERUGIA tunes are introduced to include systematic variations suggested by the theory (detailed description is in the reference [56]).

6.4 Initial State and Final State Radiation

The effect of having jets from gluons due to the initial and final state radiation is estimated using the PYTHIA samples generated by varying the ISR and FSR parameters up and down by their uncertainties [55]. Then, half of the difference between these variations is quoted as the systematic uncertainty due to the ISR/FSR radiation.

6.5 Parton Density Functions

The kinematics of an event is dependent on the Parton Density Functions hence the top quark mass may depend on the choice of the PDFs. To asses the systematic uncertainty due to the choice of PDFs the events are reweighted to accommodate possible variations provided by 20 PDFs. The total PDF systematic is calculated by summing these effects using

$$\delta m_t = \frac{1}{2}(\sum_{i=1}^{20}[\Delta M^+ - |\Delta M^-]^2)^{1/2}, \tag{6.1}$$

where ΔM^+ and ΔM^- are mass shifts for positive and negetive variations of these 20 PDFs [57].

6.6 b-fragmentation

The b-fragmentation systematic uncertainty represents the possible mismodelling of the b-quark fragmentation. Incorrect modelling of b-quark fragmentation may cause shifts in the measured top quark mass. The Bowler fragmentation function parameters are recalculated using the data from the SLD[4] [58] data. The difference between this measurement and the default is quoted as the systematic uncertainty.

[4]Stanford Linear Accelerator (SLAC) Large Detector

6.7 Multiple Hadron Interactions

As mentioned in section 4.2 the instantaneous luminosity profile of the Monte Carlo events are reweighted to match that of data. To estimate the systematic uncertainty due to this modification, a sample is generated without the luminosity profile reweighing and compared with the available templates to measure the shift in the top quark mass and jet energy scale.

6.8 Sample Dependent JES

The $t\bar{t}$ all hadronic final state contain light jets as well as b-jets. The jet energy scale corrections are developed using Z+jets events and γ+jets events. Hence, it does not contain specific corrections depending on the initial parton composition of an event. The different parton composition may result in differences in the calorimeter response. For example, b-jet response is different from the light jet response in the calorimeter. The double ratio of particle level to the detector level response for b-jets and light jet is calculated

$$\frac{(p_T^{detector}/p_T^{particle})_{b-jet}}{(p_T^{detector}/p_T^{particle})_{light-jet})} \tag{6.2}$$

and the observed difference is applied to rescale the response of b-jets [55]. The systematic uncertainty due to this effect is assessed using the samples produced by shifting the nominal response of the b-jets by the calculated amounts [55].

6.9 Jet Energy Resolution

The difference of the resolution of the jets from Monte Carlo is corrected by JSSR. The systematic uncertainty due to this procedure is assessed shifting the parameters of the jet smearing functions by their uncertainties.

6.10 Jet Identification Efficiency

The particle jets from Monte Carlo show a higher probability of being reconstructed at the detector compared to that of data. Hence, scale factors are applied to Monte Carlo jets to compensate for this effect. The systematic uncertainty is assessed by varying these scale factors by their uncertainties [55]. The samples are produced by lowering the Jet Identification Efficiency by one standard deviation in the CC region, ICR region and in both CC and ICR regions. Then these samples are compared with the available templates and shifts are calculated. The maximum shift from the above three is taken as the systematic uncertainty.

6.11 b-tagging Efficiency

The taggability is estimated for data and applied to Monte Carlo as a weight in order to get the correct efficiencies. The uncertainties due to taggability are estimated by raising and lowering the taggability weight by one standard deviation and comparing it to the default.

6.12 Residual JES

The extraction of jet energy scale uncertainty is performed by shifting the jet energies by a constant factor. But, the jet energy scale uncertainties depend on many factors other than the jets energy, such as transverse momentum, detector pseudorapidity and other effects such as detector imperfections. To assess these higher order effects, a signal pseudo-data set is built by shifting the jet energies by jet energy scale uncertainties provided by the jet energy scale group [39]. This pseudo-data set is compared to the templates to measure the mass shift.

6.13 Template Statistics

The signal and background templates are constructed using events that pass the selection criteria. Hence, the templates are limited by statistics. The effect is significant for the background template, since there are approximately 8000 events remaining after the selection.

The systematic uncertainty due to the background statistics is assessed by resampling the background sample many times and performing the measurement for data. The width of the distribution is taken as the systematic uncertainty for mass and ΔJES.

The systematic uncertainty due to the signal template statistics is also calculated using the same procedure.

6.14 Background

The background events used in the analysis are built from the data events. The systematic uncertainties associated with the modelling of the background should be assessed.

6.14.1 Background Modelling

The variables of interest for the analysis show very good agreement between signal subtracted data and background. Nonetheless, there are a few variables that show minor disagreements, hence, creates possibility for biases. The systematic uncertainty due to the background modelling is accounted for by reweighting the events such that the agreement between signal subtracted data and background is perfect for the variable which has the worst agreement that is used in the BDTs. The shift in the fitted mass and ΔJES is taken as the systematic uncertainties of the measured top quark mass and ΔJES.

6.15 Signal Fraction

The total likelihood is minimized by constraining the expected number of signal events (n_s) to the observed number of signal events (for M_t=172.5 GeV, ΔJES=0 σ_{JES}) after the selection (section 5.3). But, this number may differ from the true number of signal events due to the uncertainties in the theoretical prediction and measurement in the integrated luminosity. To assess the uncertainty, the expected number of signal events is varied within its uncertainty.

Table 6.1: Systematic uncertainties calculated for each source for top quark mass and ΔJES. The total systematic uncertainty is the quadrature sum of each source.

Source	δM_t^{meas} (GeV)	$\delta \Delta JES^{meas}$ (σ_{JES})
Hadronization	1.4	0.50
Higher order effects	1.1	0.20
b-fragmentation	0.1	0.01
ISR/FSR	0.1	0.04
PDF	0.4	0.06
Multiple $p\overline{p}$ interactions	0.1	0.02
Color reconnection	0.3	0.02
Vertex confirmation weight	0.3	0.05
b-tagging efficiency	0.1	0.10
Beam position reweighing	0.1	0.03
Sample dependent correction	0.2	0.15
Residual JES	0.3	-
Jet identification	0.3	0.07
Jet Energy resolution	0.2	0.06
Template statistics	1.0	0.21
Background modelling	0.1	0.07
Signal fraction	0.2	0.03
Calibration	0.2	0.02
Total	2.2	0.63

Chapter 7

Discussion

In this thesis, the top quark mass measurement along with a constraint to the jet energy scale uncertainty for the all hadronic final state is presented. The measurement is performed on 4033 data events and the extracted top quark mass and its uncertainties are

$$m_t = 170.4 \pm 1.7 \text{ (stat)} \pm 2.9 \text{ (sys) GeV}.$$

The measurement shows agreement with previous results (Fig. 7.1).

The total uncertainty of this measurement is dominated by systematic uncertainties, mainly from the jet energy scale calibration, simulation of the signal events and background template statistics. The higher order corrections and the hadronization needs better understanding hence, some future studies may be implemented to address these issues. The other main contribution, background template statistics may be lowered by increasing the number of background events. As mentioned earlier the background model is data driven. Hence, statistics are limited by the available number of data events which are used to build the background. But, currently studies are underway to build a new background model that includes more events.

The statistical uncertainty of the measurement is 1.7 GeV. As mentioned in chapter 5 this analysis is performed on events with exactly six jets with two identified b-jets in the final state. It is possible to reduce the statistical uncertainty of the

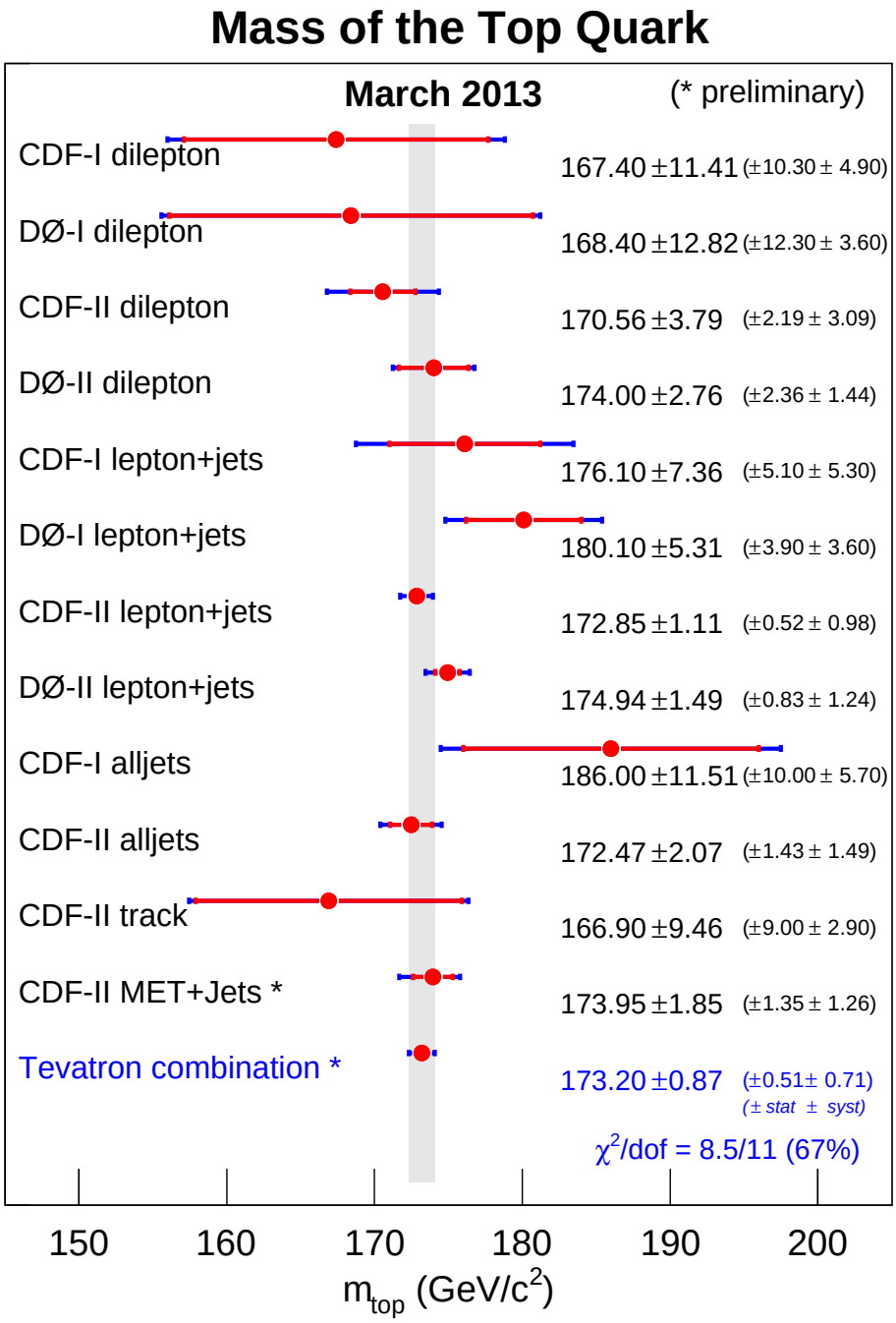


Figure 7.1: The top quark mass measurements made on various final states using the data collected at the TEVATRON and the combined mass as of March 2013.

measurement by including more events. For example, the number of signal events with one identified b-jet is twice that being used in this analysis. This sample is not used in this analysis since, there was no background model available for such a selection. But, building the one b-tag background model will help to reduce the statistical uncertainty significantly.

The application of W boson mass constrain to the measurement improved the systematic uncertainty from the standard DØ jet energy scale calibration. This 2-d method reduced the uncertainty due to the jet energy scale corrections by about 1.3 GeV[1].

In conclusion, the measured top quark mass is consistent with the previous measurements. Further enhancement of the measurement is possible using a larger data set and with a better understanding of the uncertainties.

[1]This is calculated by performing a one dimensional top mass measurement and calculating the Jet Energy Scale uncertainties on the measured value.

Bibliography

[1] J. Beringer et al. (particle data group). *Phys. Rev. D86, 010001*, 2012.

[2] A. Quadt. Top quark physics at hadron colliders. *The European Physical Journal C - Particles and Fields*, 48:835–1000(166), December 2006.

[3] Zack Sullivan. Understanding single-top-quark production and jets at hadron colliders. *arXiv:hep-ph/0408049v2*, 25 Oct 2004.

[4] Accelerator - fermilab's tevatron. `http://www.fnal.gov/pub/science/accelerator/`.

[5] V. Abazov. The upgraded DZero detector. *arXiv:physics/0507191*, February 2008.

[6] A. Khanov. HTF: histogramming method for finding tracks. the alogorithm description. *DZero Note 3778*, Spetember 2000.

[7] Jet identification. `http://www-d0.fnal.gov/phys_id/jets/jetid.html`.

[8] DZero final RunIIA Jet Energy Scale. `http://www-d0.fnal.gov/phys_id/jes/public_RunIIa/`.

[9] Tao Han. Collider phenomenology: Basic knowledge and techniques. *arXiv:hep-ph/0508097 v1*, August 9, 2005.

[10] A quantum diaries survivor. `http://dorigo.wordpress.com/2008/03/25`.

[11] The DZero experiment. `http://www-d0.fnal.gov/atwork/index.html`.

[12] Francis Halzen and Alan D. Martin. *Quarks and Leptons.* JOHN WILEY & SONS, 1984.

[13] M. Gell-Mann. A schematic model of baryons and mesons. *Physics Letters,*, 8, 1964.

[14] G. Zweig. An SU3 model for strong interaction symmetry and its breaking. *CERN Report 8419/Th 412*, 1964.

[15] S. Abachi et al. Observation of the top quark. *Phys.Rev.Lett. 74 2632-2637*, March 1995.

[16] F. Reines and C. Cowan. The neutrino. *Nature*, Sept 1, 1956.

[17] G. Danby, J-M. Gaillard, K. Goulianos, L. M. Lederman, N. Mistry, M. Schwartz, and J. Steinberger. Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos. *Phys. Rev. Lett.*, 9(1):36–44, Jul 1962.

[18] Donald H. Perkins. *Introduction to High Energy Physics.* Cambridge University Press, Cambridge., 2000.

[19] Gordon Fraser. *The Particle Century.* Institute of Physics Publishing, 1998.

[20] WIKIPEDIA. `http://en.wikipedia.org`.

[21] Nikolaos Kidonakis and Ramona Vogt. Theoretical top quark cross section at the fermilab tevatron and the cern lhc. *Physical Review D*, 78(7):074005, 2008.

[22] T. Stelzer, Z. Sullivan, and S. Willenbrock. Single-top-quark production at hadron colliders. *Phys.Rev. D58 (1998) 094021*, 2, 1998.

[23] A.P. HEINSON. Observation of single top quark production at the tevatron collider. *Phys.Lett. A25 (2010) 309-339*, February 2010.

[24] V. M. Abazov. Measurement of tt production in the $tau+jets$ topology using pp collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev. D 82, 071102*, October 27 2010.

[25] T. Aaltonen. Measurements of the top-quark mass and the $t\bar{t}$ cross section in the hadronic tau+jets decay channel at $\sqrt{s} = 1.96$ TeV. *Phys. Rev. Lett. 109, 192001*, 7 November 2012.

[26] R. Angstadt, G. Brooijmans, D. Chapin, M. Clementsa, D. Cutts, A. Haas, R. Hauser, M. Johnson, A. Kulyavtsev, S. E. K. Mattingly, M. Mulders, P. Padley, D. Petravick, R. Rechenmacher, S. Snyder, and G. Watts. The DZero level 3 data acquisition system. *IEEE TRANSACTIONS ON NUCLEAR SCIENCE*, 51, June 2004.

[27] R.E. Kalman. A new approach to linear filtering and prediction problems, transactions of the ASME. *Journal of Basic Engineering, Series D, 82 (1960), pp. 35–45.*, 1960.

[28] P.V.C. Hough. Machine Analysis of Bubble Chamber Pictures. *Conf.Proc. C590914 (1959) 554-558*, September 1959.

[29] G. Borissov. Technical details of aa tracking. *All DZero meeting*, February 2003.

[30] A. Schwartzman and C. Tully. Primary vertex reconstruction by means of adaptive vertex fitting. *DZero Note 4918*, Spetember 2005.

[31] J. Kozminski, R. Kehoe, H. Weerts, S. Park, A. Quadt, J. Gardner, and S. Jabeen. Electron likelihood in p14. *DZero Note 4449*, April 2004.

[32] A. Kumar, B. Choudhary, J. Kozminski, R. Kehoe, J. Hays, and J. Stark. Electron likelihood study. *DZero Note 4769*, July 2005.

[33] K. Augsten, V. Bazterra, O. Brandt, A. Garcia-Bellido, A. Jung, B. Lee, D. Meister, A. Schukin, and Y. Tsai. Selection of single and double top events in the lepton+jets channel with 9.7 fb^{-1}. *DZero note 6308*, May 2012.

[34] P. Calfayan, T. Gadfort, G. Hesketh, V. Lesne, M. Owen, R. Stroehmer, V. Sharyy, and B. Tuchming. Muon identification certification for p17 data. *DZero Note 5157*, February 2007.

[35] O. Brandt, S. Cho, M. Cooke, M. Eads, D. Hedin, J. Lim, S. Park, A. Santos, B. Tuchming, Y. Yatsunenko, and S. Youn. Muon identification certification for the summer 2009 extended dataset (Run IIB-1 and -2). *DZero Note 6025*, March 2010.

[36] E. Busato and B. Andrieu. Jet algorithms in DZero RunII software: Description and user's guide. *DZero note 4457*, August 2004.

[37] G. Bernardi, B. Olivier, B. Knuteson, and M.Strovink. Nada: A new event by event hot cell killer. *DZero note 3687*, October 2001.

[38] G. Bernardi, E. Busato, and J. Vlimant. Improvements from the T42 algorithm on calorimeter objects reconstruction. *DZero note 4335*, February 2004.

[39] D. Bandurin, N. Bartosik, A. Das, Y. Ilchenko, H. Liu, R. Kehoe, Georgy G., A. Verkheev, A. Jayasinghe, G. Petrillo, Y. Tsai, M. Wang, M. Wobisch, and Z. Ye. Jet energy determination at D0 Run II. December 2012.

[40] JES Group. Jet Energy Scale Determination at D0 Run II (final p17 version). *DZero note 6327*, May 2012.

[41] S. Greder and I. Ripp-Baudot. B-tagging with combined multivariate techniques. *DZero note 5936*, March 2011.

[42] K. Herner. Well MET: Understanding missing et at DZero. *University of DZero Lecture*, November 2011.

[43] B.C.K. Casey, M. Corcoran, K. DeVaughan, Y. Enari, E. Gallas, I. Katsanos, J. Linnemann, J. Orduna, R. Partridge, M. Prewitt, H. Schellman, G.R. Snow, and M. Verzocchi. The D0 Run IIb Luminosity Measurement. *DZero note 6317*, March 2012.

[44] G. Grenier. ALPGEN. `http://www-d0.fnal.gov/computing/MonteCarlo/generators/alpgen.html`.

[45] R. Scalise. The coordinated theoretical-experimental project on QCD. `http://www.phys.psu.edu/~cteq/`.

[46] T. Sjostrand, S. Mrenna, and Peter Skands. A brief introduction to PYTHIA 8.1. *arXiv:0710.3820*, October 2007.

[47] S. Hoche, F. Krauss, N. Lavesson, L. Lonnblad, M. Mangano, A. Schalicke, and S. Schumann. Matching parton showers and matrix elements. *arXiv:hep-ph/0602031v1*, February 2006.

[48] J. Hegeman. Luminosity determination and reweighting of Monte Carlo overlay luminosity for p17 hadronic top analyses. *DZero note 5561*, January 2008.

[49] Y. Peters, M. Begel, K. Hamacher, and D. Wicke. Reweighting of the fragmentation function for the D0 Monte Carlo. *DZero note 5325*, January 2007.

[50] The large electron-positron collider. `http://home.web.cern.ch/about/accelerators/large-electron-positron-collider`.

[51] D. Bandurin et al. Jet energy scale determination at DZero Run II. *DZero Internal*, May 2013.

[52] K. Augsten, V. Bazterra, O. Brandt, A. Garcia-Bellido, A. Jung, B. Lee, D. Meister, A. Schukin, and Y. Tsai. Selection of single and double top events in the lepton+jets channel with 9.7 fb^{-1}. *DZero note 6308*, May 2012.

[53] B. Roe, H. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor. Boosted Decision Trees as an alternative to Artificial Neural Networks for particle identification. *arXiv:physics/0408124*, November 2004.

[54] M. Backes, T. Carli, O. Cohen, A. Christov, D. Dannheim, K. Danielowski, S. Henrot-Versile, M. Jachowski, K. Kraszewski, A. Krasznahorkay Jr., M. Kruk, Y. Mahalalel, R. Ospanov, X. Prudent, A. Robert, D. Schouten, F. Tegenfeldt, A. Voigt, K. Voss, M. Wolter, and A. Zemla. TMVA 4 toolkit for multivariate data analysis with ROOT. *arXiv:physics/0703039*, November 2009.

[55] F. Deliot, A. Grohsjean, C. Schwanenberger, E. Shabalina, M. Wang, D. Wicke, and Z. Ye. Systematic uncertainties in top quark measurements. *DZero note 6024*, February 2010.

[56] P. Skands. Tuning Monte Carlo generators: The Perugia tunes. *Phys.Rev.D82:074018,2010*, May 2011.

[57] V. M. Abazov. Precise measurement of the top-quark mass from lepton+jets events at DZero. *Phys. Rev. D 84, 032004*, July 2011.

[58] Introduction to the SLD collaboration. `http://www-sld.slac.stanford.edu/sldwww/sld.html`.

Appendix A

3JT and 4JT Triggers

During the RunIIB data taking period two major trigger versions were used, v15 and v16. Following are the trigger definitions of the 3JT and 4JT triggers for these two versions.

- Level 1
  - One jet with $E_T > 30$ GeV and $|\eta| < 2.4$ and
  - Second jet with $E_T > 15$ GeV and $|\eta| < 2.4$ and
  - Third jet with $E_T > 8$ GeV and $|\eta| < 3.2$
- Level 2

  OR of the following 4 scripts
  - Script 1
    - One jet with $E_T > 20$ GeV and $|\eta| < 2.4$ and
    - No events with a pair of jets with $E_T > 5$ GeV which is back to back within a window of 11.25 degrees opening angle in ϕ and
    - Total H_T (for jets with $E_T > 6$ GeV within $|\eta| < 2.6$) is larger than 35 GeV and

* Missing E_T (calculated using all the jets with $E_T > 10$ GeV) is larger than 20 GeV.

- Script 2
  * Three jets with $E_T > 8$ GeV and $|\eta| < 2.4$ (only in v16) and
  * Two of the jets from above are required to have $|\eta| < 2.4$ and $E_T > 15$ GeV and 30 GeV and
  * $H_T > 100$ GeV (in v15. for v16 $H_T > 75$ GeV) for all jet objects with $E_T > 6$ GeV and $|\eta| < 2.6$ and
  * Also require one STT IP track with IPSIG >=3. and chi2 < 5.5 (only in v16).
- Script 3
  * Three jets with $E_T > 8$ GeV and
  * Two of the jets from above are required to have $|\eta| < 2.6$ and $E_T > 15$ GeV and 30 GeV and
  * $H_T > 75$ GeV (for jet objects with $E_T > 6$ GeV and $|\eta| < 2.6$) and
  * Missing $E_T > 10$ GeV(in v15. For v16 Missing $E_T > 20$ GeV) (for all the jets with $E_T > 10$ GeV).
- Script 4
  * Three jets with $E_T > 6$ GeV and
  * $H_T >$75 GeV for all the jets with $E_T > 6$ GeV and $|\eta| < 2.6$ and
  * Sphericity > 0.1 and
  * One STT IP track with IPSIG>=3. and chi2 < 5.5 (only in v16).
- Level 3
  * Primary vertex with $|z| <$ 35cm and

* At least three jets with E_T >15 GeV found using a simple cone algorithm and
* At least two jets with E_T >25 GeV found using a simple cone algorithm and
* IP B-event tag $<$ 0.4 (in v15.00-v15.07 it was 0.04).

Appendix B

Variables

Following are the full set of variables used in the analysis.

`Jet0_Pt` : Leading jet transverse momentum

`Jet1_Pt` : 2nd leading jet transverse momentum

`Jet2_Pt` : 3rd leading jet transverse momentum

`Jet3_Pt` : 4th leading jet transverse momentum

`Jet4_Pt` : 5th leading jet transverse momentum

`Jet5_Pt` : 6th leading jet transverse momentum

`Jet0_eta` :Leading jet pseudo rapidity

`Jet1_eta` : 2nd leading jet pseudo rapidity

`Jet2_eta` : 3rd leading jet pseudo rapidity

`Jet3_eta` : 4th leading jet pseudo rapidity

`Jet4_eta` : 5th leading jet pseudo rapidity

`Jet5_eta` : 6th leading jet pseudo rapidity

`Jet0_E` : Leading jet Energy

`Jet1_E` : 2nd leading jet Energy

`Jet2_E` : 3rd leading jet Energy

`Jet3_E` : 4th leading jet Energy

`Jet4_E` : 5th leading jet Energy

`Jet5_E` : 6th leading jet Energy

`Jet0_phi` : Leading jet azimuthal angle

`Jet1_phi` : 2nd leading jet azimuthal angle

`Jet2_phi` : 3rd leading jet azimuthal angle

`Jet3_phi` : 4th leading jet azimuthal angle

`Jet4_phi` : 5th leading jet azimuthal angle

`Jet5_phi` : 6th leading jet azimuthal angle

`Topo_aplanarity` : Event aplanarity $3/2\ \lambda 3$

`Topo_planarity` : Event planarity $\lambda 2$ - $\lambda 3$

`Topo_aveY` : p_T-weighted average of rapidities

`Topo_sphericity` : $3/2$*($\lambda 2$+$\lambda 3$)

`Topo_centrality` : H_T / H

`Topo_C` : $\sum 3/2\lambda_i \times \lambda_j$

`Topo_costhetastar` : $\cos(\theta)$ of the leading jet in all-good-jets rest frame

`Topo_D` : 27 * $\lambda 1$*$\lambda 2$*$\lambda 3$

`Topo_H` : Total energy of the jets

`Topo_HT` : Total transverse energy of the jets

`Topo_HT2` :Total transverse energy of the leading 2jets

`Topo_HT3` :Total transverse energy of the leading 3jets

`Topo_HT4` :Total transverse energy of the leading 4jets

`Topo_HT5` :Total transverse energy of the leading 5jets

`Topo_HTb` : Total transverse energy of the b-jets

`Topo_FWM0` : Fox-Wolfram moment order 0

`Topo_FWM1` : Fox-Wolfram moment order 1

`Topo_FWM2` : Fox-Wolfram moment order 2

`Topo_FWM3` : Fox-Wolfram moment order 3

`Topo_FWM4` : Fox-Wolfram moment order 4

`Topo_FWM5` : Fox-Wolfram moment order 5

`Topo_FWM6` : Fox-Wolfram moment order 6

`Topo_FWM7` : Fox-Wolfram moment order 7

`Topo_FWM8` : Fox-Wolfram moment order 8

`Topo_FWM9` : Fox-Wolfram moment order 9

`Topo_pTl0` : Transverse momentum of leading light jet

`Topo_pTl1` : Transverse momentum of 2nd leading light jet

`Topo_pTl3` : Transverse momentum of 4th leading light jet

`Topo_pTb0` : Transverse momentum of leading b jet

`Topo_pTb1` : Transverse momentum of 2nd leading b jet

`Topo_dPT01` : Transverse momentum difference between leading jet and 2nd leading jet

`Topo_dPT02` : Transverse momentum difference between leading jet and 3rd leading jet

`Topo_dPT03` : Transverse momentum difference between leading jet and 4th leading jet

`Topo_dPT04` : Transverse momentum difference between leading jet and 5th leading jet

`Topo_dPT05` : Transverse momentum difference between leading jet and 6th leading jet

`Topo_dPhib0b1` : Azimuthal angle difference between two b-jets

`Topo_dPhi01` : Azimuthal angle difference between two leading jets

`Topo_dPhi02` : Azimuthal angle difference between leading jet and 3rd leading jet

`Topo_dPhi12` : Azimuthal angle difference between 2nd leading jet and 3rd leading jet

`Topo_PZH` : Longitudinality

`Topo_PZ` : scalar sum of longitudinal momenta of jets

`Topo_dYb0b1` : rapidity separation of the two leading b-jets

`Topo_yl0` : rapidity of leading light jet

`Topo_yl1` : rapidity of 2nd leading light jet

`Topo_yl2` : rapidity of 3rd leading light jet

`Topo_yl3` : rapidity of 4th leading light jet

`Topo_yb0` : rapidity of leading b jet

`Topo_yb1` : rapidity of 2nd leading b jet

`Topo_minDRbl` : minimum dR between b-jets and light jets

`Topo_minDRll` : minimum dR between light jets

`Topo_minDRbb` : minimum dR between b-jets

`Topo_maxDRbl` : maximum dR between b-jets and light jets

`Topo_maxDRll` : maximum dR between light jets

`Topo_maxDRbb` : maximum dR between b-jets

`Topo_dRmin` : minimum ΔR between jets

`Topo_dRmax` : maximum ΔR between jets

`Topo_dYmax` : max ΔY between jets

`Topo_dYmax4` : max ΔY between leading 4 jets

`Topo_dY01` : rapidity separation of the two leading jets

`Topo_dY02` : rapidity separation of the leading jet and 2nd leading jet

`Topo_dY12` : rapidity separation of the 2nd leading jet and 3rd leading jet

`Topo_M2` : Invariant mass of first two jets

`Topo_M3` : Invariant mass of first three jets

`Topo_M4` : Invariant mass of first four jets

`Topo_M5` : Invariant mass of first five jets

`Topo_M6` : Invariant mass of first six jets

`Topo_Mb0b1` : Invariant mass of b-jets

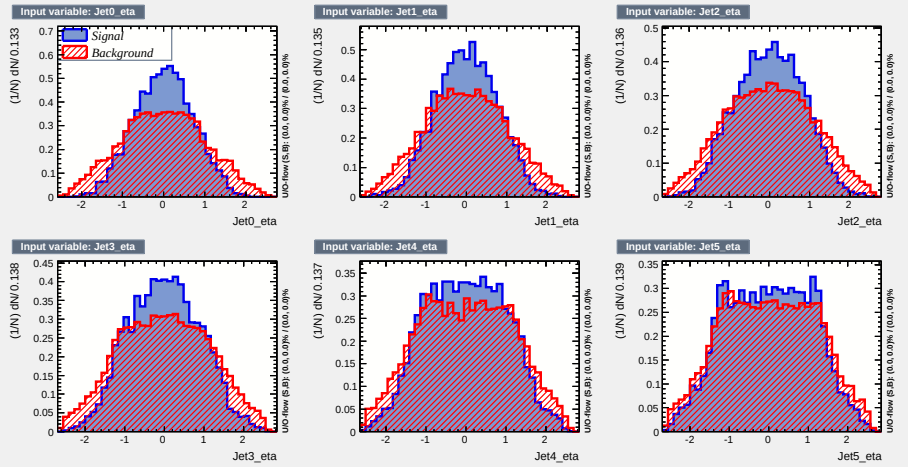


Figure B.1: Comparison plots of Jet0_eta, Jet1_eta, Jet2_eta, Jet3_eta, Jet4_eta and Jet5_eta for signal and background.

`Topo_Mb0b1overMall` : Invariant mass of b-jets/Invariant mass of all jets

`Topo_scaledMb0b1` : Invariant mass of b-jets/b-jets transverse momenta

`Topo_pT1overHT` : 2nd leading jet tarnsverse momenta divided by H_T

`Topo_pT0overHT` : leading jet tarnsverse momenta divided by H_T

`Topo_lambda1` : momentum tensor first eigen value

`Topo_lambda2` :momentum tensor second eigen value

`Topo_lambda3` : momentum tensor third eigen value

`Topo_dRb0b1` : Separation between two b-jets

`Topo_dR01` : Separation between leading jet and 2nd leading jet

`Topo_dR02` : Separation between leading jet and 3rd leading jet

`Topo_dR12` : Separation between 2nd leading jet and 3rd leading jet

`Topo_v1` : Average cos (average $\Delta\phi$)

`Topo_v2` : Average cos 2* (average $\Delta\phi$)

`Topo_v4` : Average cos 4* (average $\Delta\phi$)

The comparison plots of signal and background for the variables used in the Boosted Decision Trees are shown in Fig. B.1 to Fig. B.6.

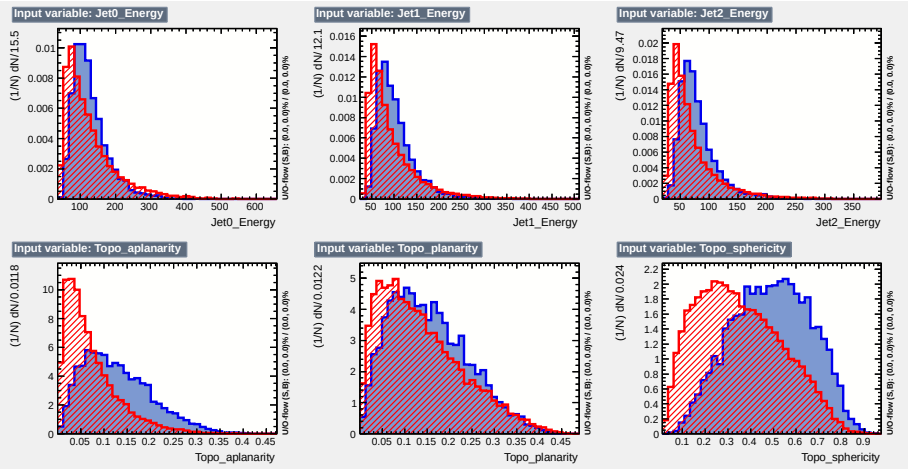


Figure B.2: Comparison plots of Jet0_Energy, Jet1_Energy, Jet2_Energy, Topo_aplanarity, Topo_planarity and Topo_sphericity for signal and background.

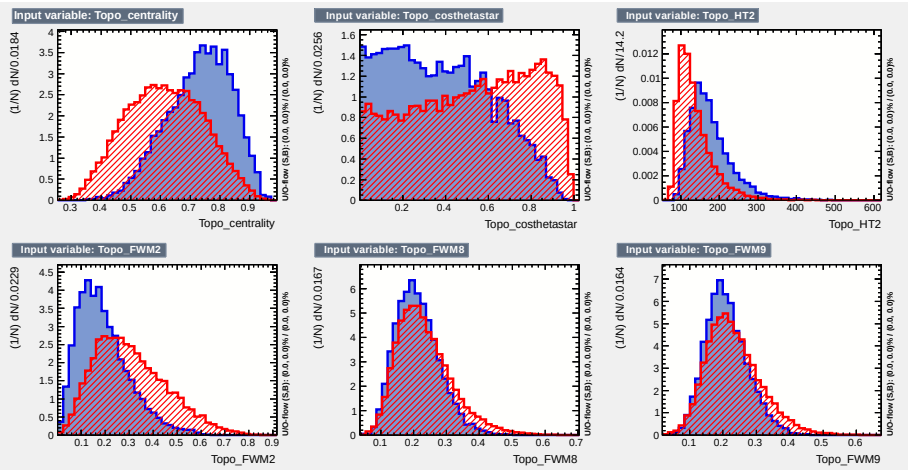


Figure B.3: Comparison plots of Topo_centrality, Topocos(θ^*), Topo_HT2, Topo_FWM2, Topo_FWM8 and Topo_FWM9 for signal and background.

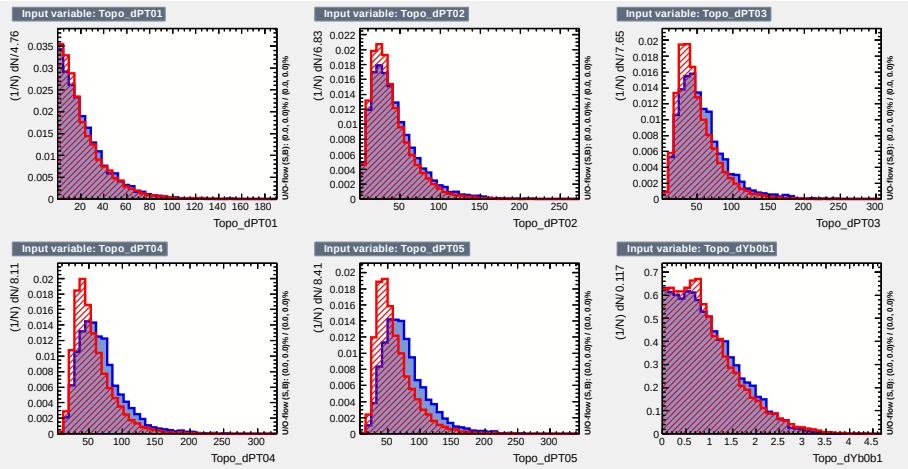


Figure B.4: Comparison plots of Topo_dPT01, Topo_dPT02, Topo_dPT03, Topo_dPT05, Topo_dPT05 and Topo_dYb0b1 for signal and background.

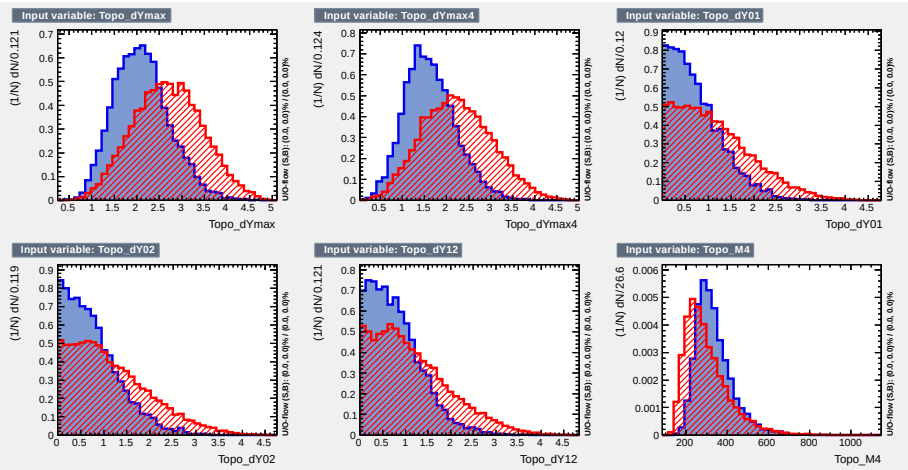


Figure B.5: Comparison plots of Topo_dYmax, Topo_dYmax4, Topo_dY01, Topo_dY02, Topo_dY12 and Topo_M4 for signal and background.

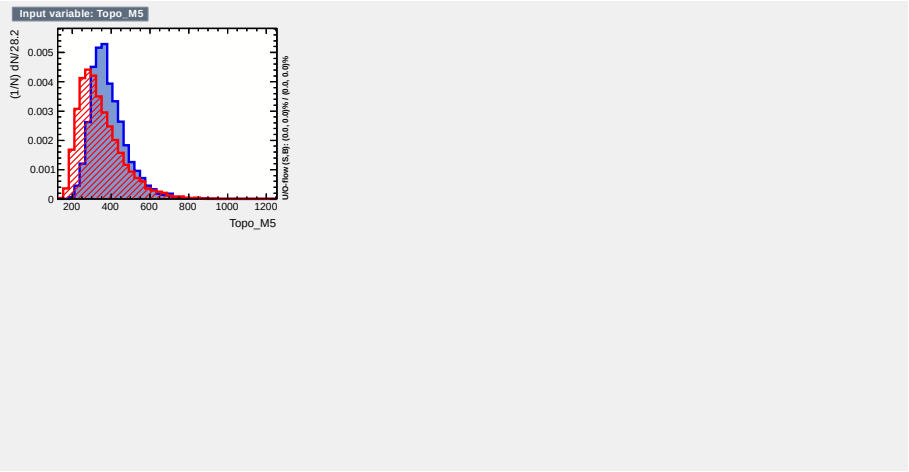


Figure B.6: Comparison plots of Topo_M5 for signal and background.