SEARCH FOR SUPERSYMMETRY IN THE JETS + MET + TAUS FINAL STATE USING THE CMS DETECTOR AT THE LHC

A Dissertation

by

ROY JOAQUIN MONTALVO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Teruki Kamon
Committee Members,	Bhaskar Dutta
	Stephen Fulling
	Alexei Safonov
	David Toback
Head of Department,	George R. Welch

May 2013

Major Subject: Physics

ABSTRACT

In this dissertation results are presented from a search for the pair production of heavy colored particles (gluinos, squarks) in R-parity conserving supersymmetric models, in which the lightest supersymmetric particle is a stable and neutral object. The search was performed for events with at least two tau leptons, two highly energetic jets and large missing transverse momentum in the final state on a data sample of proton-proton collisions at $\sqrt{s} = 7$ TeV. The data sample was collected by the Compact Muon Solenoid detector at the Large Hadron Collider in 2011, and it corresponds to an integrated luminosity of 5fb^{-1}. The tau isolation variable was optimized for this search. The number of events corresponding to standard model processes in the final selection was estimated to be 7.49 ± 0.74 using background estimation techniques based on data. Nine observed events are found to be in agreement with the standard model prediction, and exclusion limits on gluino mass are obtained in the context of supersymmetric models at the 95% confidence level.

Para Don Rafa

ACKNOWLEDGMENTS

I would like to thank all of my friends and collegues that made this analysis possible.

I want to express my sincere gratitude to my advisor, Teruki Kamon. Dr.Kamon constantly gave me valuable advice and guidance, and even though I argued almost every step of the way, he was always patient and understanding. Thanks to Dr. Kamon I also had the opportunity to work at Fermi National Accelerator Laboratory (Fermilab). Being at Fermilab by itself was a great learning experience thanks largely to Shuichi Kunori. Dr. Kunori was always available to talk and in many ocations went out of his way to help.

While working in Dr. Kamon's group I had the fortune to work closely with many great scientists. I'd like to acknowledge Alfredo (now Dr.) Gurrola who was a driving force and source of knowledge for this analysis, Dr. Ji Eun Kim for always being willing to travel far and work over time during critical times, and Will Flanagan who never hesitated to put in long hours of work to make many aspects of our search possible. To you and the other members of the CMS collaboration that had a role in making this work possible, thank you.

Finally, I want to thank my family. I am blessed to have an incredible family who has always supported and encouraged me. No one has influenced my life more than them. My father, Rafael, and my mother, Judith, made me the person I am today; they always showed interest and pride towards my work. They kept me going. I also had the continued support of my brothers, Rafael and Robert, their wives, Karla and Yoko, as well as my sister, Judith, and her husband, Andrew. All of them, along with their children, are always a source of inspiration.

TABLE OF CONTENTS

Page

ABSTRACT ii

DEDICATION iii

ACKNOWLEDGMENTS iv

TABLE OF CONTENTS v

LIST OF TABLES viii

LIST OF FIGURES ix

1 INTRODUCTION 1

2 THEORY AND MOTIVATION FOR SUPERSYMMETRY AT THE LHC . 3

2.1 The Standard Model of Particle Physics 3
2.2 Beyond The Standard Model 4
2.3 Supersymmetry 5
2.4 Motivation for SUSY Searches at the LHC 6

3 EXPERIMENTAL TOOLS 9

3.1 The Large Hadron Collider 9
3.2 The CMS Detector 10
3.2.1 CMS Coordinate System 13
3.2.2 CMS Tracker System 13
3.2.3 The Electromagnetic Calorimeter 15
3.2.4 The Hadron Calorimeter 16
3.2.5 Muon System 18
3.2.6 Triggering 19
3.2.7 Data Acquisition 23

4 TAU LEPTONS 25

5 ANALYSIS STRATEGY 27

6 OBJECT RECONSTRUCTION AND IDENTIFICATION 30

Page

6.1 Jet Reconstruction . . . 30
6.2 Tau Reconstruction and Identification . . . 30
6.3 Electron Reconstruction and Identification . . . 32
6.4 Muon Reconstruction and Identification . . . 32
6.5 $\not\!\!E_T$ and $\not\!\!H_T$ . . . 33

7 TRIGGER AND EVENT SELECTIONS . . . 34

7.1 Trigger Performance . . . 34
7.2 Event Selection . . . 36

8 BACKGROUND ESTIMATION . . . 41

8.1 Estimation of $\mathrm{t\bar{t}}$ Events in the Signal Region . . . 41
8.1.1 Upper Limit on the Real $\tau_h\tau_h$ $\mathrm{t\bar{t}}$ Estimate . . . 50
8.2 Estimation of Z ($\rightarrow \nu\nu$) + Jets Events in the Signal Region . . . 51
8.2.1 Upper Limit on the $Z(\rightarrow \nu\nu)$ + Jets Estimate . . . 53
8.3 Estimation of $Z(\rightarrow \tau\tau)$ + Jets Events in the Signal Region . . . 57
8.4 Estimation of QCD Events in the Signal Region . . . 59
8.4.1 Upper Limit on the QCD Background . . . 63
8.5 Estimation of W + Jets Events in the Signal Region . . . 65
8.6 Summary of Background Estimations . . . 73

9 STATISTICS . . . 76

9.1 CL_s Method . . . 76
9.2 Bayesian Method . . . 76

10 SYSTEMATIC UNCERTAINTIES . . . 78

10.1 Systematic Uncertainties on Backgrounds . . . 78
10.1.1 Systematic Uncertainties on the $\mathrm{t\bar{t}}$ Prediction . . . 78
10.1.2 Systematic Uncertainties on the W + Jets Prediction . . . 81
10.1.3 Systematic Uncertainties on the Z + Jets Prediction . . . 84
10.2 Systematic Uncertainties on Signal . . . 87

11 RESULTS . . . 93

11.1 Data in the Signal Region . . . 93
11.2 Limits to New Physics . . . 95
11.2.1 CMSSM . . . 95

Page

11.2.2 SMS 95
11.2.3 GMSB 98

12 CONCLUSIONS 99

REFERENCES 100

APPENDIX A. EFFECT OF B-TAGGING ON CONTROL REGIONS 102

APPENDIX B. VALIDATION OF THE B-TAGGING EFFICIENCY IN A HIGH MULTIPLICITY JET SAMPLE 106

LIST OF TABLES

TABLE Page

2.1 Fermions of the standard model of particle physics. 4

3.1 Number of sub-detector data sources [11]. 23

6.1 Reconstructed tau decay modes 31

6.2 μ identification 32

8.1 Events in the $t\bar{t}$ control region for data and MC. 49

8.2 $t\bar{t}$ extraction efficiencies. 49

8.3 Events in the $Z(\rightarrow \mu\mu)$ + Jets control region for data and MC. . . . 53

8.4 $Z(\rightarrow \nu\nu)$ + Jets extraction efficiencies 54

8.5 $Z \rightarrow \tau\tau$ + Jets extraction efficiencies 59

8.6 Events in the QCD control region for data and MC. 64

8.7 Events in the W + Jets enhanced region for data and MC. 66

8.8 W + Jets events extraction efficiencies 73

8.9 Summary of background predictions with statistical uncertainties. . . 75

10.1 List of systematic uncertainties for signal and background estimates. . 89

11.1 Number of data events and estimated background rates with statistical and systematic uncertainties, respectively. 93

LIST OF FIGURES

FIGURE Page

3.1 Aerial view of the LHC [8]. 10

3.2 CERN accelerator complex [10]. 11

3.3 Schematic drawing of the CMS detector. 13

3.4 Drawing of the CMS detector showing inner sub-systems. 14

3.5 CMS Pixel detector layout. 14

3.6 Electromagnetic shower created due to the interaction of an electron with the ECAL crystals. 16

3.7 The CMS Electromagnetic Calorimeter. 17

3.8 A barrel crystal and the APD capsule [11]. 17

3.9 An endcap crystal and VPT [11]. 18

3.10 CMS HCAL half-barrel in the assembly hall [11]. 20

3.11 View of the HB wedge showing scintillator sampling design [11]. . . . 20

3.12 Longitudinal view of the CMS detector [11]. 21

3.13 Depiction of muon going trough the CMS detector. 21

3.14 Quarter panel view of the CMS detector. 22

3.15 Architecture of the Level-1 Trigger [11]. 22

3.16 Architecture of the CMS DAQ System [11]. 24

4.1 Feynman diagram of leptonic tau decays. 26

4.2 Sample Feynman diagram of a hadronic tau decay mode. 26

FIGURE Page

5.1 Analysis strategy flow chart. 29

6.1 Depiction of τ_h decay modes. 31

7.1 Trigger turn-on curve. 36

7.2 $\mathrm{p_T}$ distribution for 1^{st} leading jet. 38

7.3 $\not\!\!E_T$ distribution. 39

7.4 $\not\!\!H_T$ distribution. 39

7.5 Simulated SUSY event: 3D Lego View. 40

7.6 Simulated SUSY event: 3D View. 40

8.1 $\not\!\!E_T$. 43

8.2 $\Delta\phi(j_1, \not\!\!H_T)$. 43

8.3 H_T. 44

8.4 $\not\!\!H_T$. 44

8.5 1^{st} Leading Jet p_T. 45

8.6 2^{nd} Leading Jet p_T. 45

8.7 τ p_T. 46

8.8 τ η. 46

8.9 Number of photons. 48

8.10 Number of tracks in the τ isolation cone. 48

8.11 H_T distribution in the $Z(\rightarrow \mu\mu)$+ Jets control region. 54

8.12 $\not\!\!H_\mathrm{T}$ distribution in the $Z(\rightarrow \mu\mu)$+ Jets control region. 55

FIGURE Page

8.13 1^{st} Leading Jet p_T distribution in the $Z(\rightarrow \mu\mu)$+ Jets control region. 55

8.14 2^{nd} Leading Jet p_T distribution in the $Z(\rightarrow \mu\mu)$+ Jets control region. 56

8.15 $\not{H}_T$ obtained from events in the $Z(\rightarrow \mu\mu)$ + Jets control sample and treating the muons as taus to model $Z(\rightarrow \tau\tau)$ + Jets events. 60

8.16 H_T obtained from events in the $Z(\rightarrow \mu\mu)$ + Jets control sample and treating the muons as taus to model $Z(\rightarrow \tau\tau)$ + Jets events. 60

8.17 Track isolation distribution for "taus"/jets in the $Z(\rightarrow \mu\mu)$ + Jets control region. 61

8.18 γ isolation distribution for "taus"/jets in the $Z(\rightarrow \mu\mu)$ + Jets control region. 61

8.19 Distribution for number of primary vertices obtained using events from a control sample enriched with $\sim$ 86% purity of QCD events according to MC. 64

8.20 Distribution for $\not{H}_{\rm T}$ obtained using events from a control sample enriched with $\sim$ 99% purity of QCD according to MC. 66

8.21 $\not{H}_T$. 67

8.22 $\not{E}_T$. 67

8.23 First leading jet p_T. 68

8.24 Second leading jet p_T. 68

8.25 $\not{H}_T$ distribution after subtracting contributions from other SM backgrounds. 71

8.26 $\not{E}_T$ distribution after subtracting contributions from other SM backgrounds. 71

8.27 First leading jet p_T distribution after subtracting contributions from other SM backgrounds. 72

FIGURE Page

8.28 Second leading jet p_T distribution after subtracting contributions from other SM backgrounds. 72

8.29 Number of photons. 74

8.30 Number of tracks in the τ isolation cone. 74

10.1 $j \rightarrow \tau_h$ mistag rate vs. p_T obtained from the $t\bar{t}$ dominated control sample. 82

10.2 $j \rightarrow \tau_h$ mistag rate vs. p_T obtained from the W + Jets dominated control sample. 84

10.3 Depiction of the systematic effect of tau energy scale on signal (τ seed track p_T). 90

10.4 Depiction of the systematic effect of tau energy scale on signal ($\not{H}_{\mathrm{T}}$). . 90

10.5 Depiction of the systematic effect of JEC on signal (jet p_T). 91

10.6 Depiction of the systematic effect of JEC on signal ($\not{H}_{\mathrm{T}}$). 91

10.7 Signal acceptance vs. Number of vertices. 92

11.1 Stacked distributions of H_{T}. 94

11.2 Stacked distributions of $\mathrm{M_{eff}}$ in the signal region. 94

11.3 Exclusion limit in the CMSSM plane at $\tan\beta = 40$. 96

11.4 Feynman diagram for the $\tau\tau$ SMS model. 96

11.5 95% CL cross section upper limits for the T3tauh model where the solid red line represents the limits on the mass of the gluino and the LSP. 97

11.6 95% CL cross section upper limits as a function of gluino mass in the GMSB scenario. 97

A.1 τ track isolation. 103

FIGURE Page

A.2 τ γ isolation. 103

A.3 τ p_T. 104

A.4 τ η. 104

A.5 τ track isolation. 105

A.6 τ γ isolation. 105

B.1 Number of jets tagged as b-jets using the track counting high efficiency "medium" working point. 107

1 INTRODUCTION

The concept of matter being composed of minute fundamental components has been around for a very long time. Over two thousand years ago the greek philosopher Democritus first developed the concept of an undividable substance he called the atom, which in greek means indivisible. It wasn't until the nineteenth century that the existence the atom could be indirectly established. The twentieth century would yield many physics theories and experiments that showed that the atom itself has a structure and thus making the term a misnomer. The later half of the twentieth century produced many results that pushed our physical understanding of the most elementary particles and the fundamental forces. Those results were integrated into a comprehensive theory of force and matter interaction called the Standard Model (SM) of particle physics. With the tools of the twenty first century, scientists hope to answer many of the remaining open questions of the last century.

The SM describes various phenomena of elementary particle interactions to a high degree of precision. However, there are a growing number of observations that suggest that the SM is an incomplete, low energy approximation of a more fundamental description of nature. These observations range from astronomical data suggesting the existence of weakly interacting but abundant form of matter called Dark Matter (DM) to other problems like: matter anti-matter asymmetry, neutrino oscillations, the Hierarchy Problem [1] and other phenomena that the SM, as it stands, cannot account for.

The Large Hadron Collider (LHC) is the largest experiment in human history. It is designed to collide two proton beams at center of mass energies just over seven times larger than its predecessor, the Tevatron at Fermi National Accelerator Laboratory, and is expected to probe physics beyond the SM (BSM), giving physicists a new edge to search for answers. One possible extension to the SM is Supersymmetry (SUSY). SUSY is favored by many physicists because it allows for the unification of gauge

couplings while also naturally providing a DM candidate in the form of the lightest SUSY particle (LSP). A description of SUSY is presented in subsection 2.4.

The dissertation is structured as follows: Chapter 2 contains a brief description of the SM and SUSY is presented as well as the motivation to search for SUSY. The experimental tools used for the analysis presented in this dissertation are described in some detail in Chapter 3. Since this dissertation deals with tau leptons as part of the final state signature, a short chapter is dedicated to its properties, 4. The analysis strategy and the description of how physics objects are identified and reconstructed in the CMS collaboration are presented in chapters 5 and 6, respectively. Chapter 7 describes the event selections used in this analysis while Chapter 8 deals with the estimation of standard model sources in the search region. Statistics as well as systematic uncertainties are presented in Chapters 9 and 10. Finally the results and final remarks are in Chapters 11 and 12.

2 THEORY AND MOTIVATION FOR SUPERSYMMETRY AT THE LHC

2.1 The Standard Model of Particle Physics

The Standard Model (SM) is a conceptually simple but very powerfull description of the known fundamental particles and their interactions [1]. A large amount of experimental observations can be explained within the framework of the SM as well as some aspects of cosmology in the early universe.

In the SM, matter is made of half-integer spin particles called fermions, while the electromagnetic, weak and strong fundamental forces are mediated by integer spin particles called bosons. The SM divides fermions into two categories: leptons and quarks. The lepton family includes the electron-type particles and the neutrino-type particles that are primarily involved in the electroweak interactions which are mediated by the $\gamma/\mathrm{W}^{\pm}/\mathrm{Z}^{0}$ bosons. The quark family includes the up-type quark (up, charm, and top) and the down-type quarks (down, strange, and bottom). Quarks and leptons are divided into three generations where the most significant distinction is made by the masses and not their quantum numbers.

The fermion field ψ can be decomposed into chirality components designated left-handed (L) and right-handed (R) chirality (ψ_L and ψ_R). This distinction is becomes importand in the SM because the left and right handed fermions have different properties and thus act differently under gauge transformations. The left-handed fermions are arranged in doublets in the SM while the right-handed fermions are arranged in singlets (see Table 2.1).

The forces governing the elementary particle interactions are mediated by bosons which are the quanta of the gauge fields and couple to the corresponding charges. The interactions are described by the symmetry transformations of the $SU(3)_C \times SU(2)_L \times U(1)_Y$ groups, where the meaning of S stands for special ($\det(U) = 1$) and U is unitary ($UU^{\dagger} = 1$). For instance, the symmetry group of quantum chromo dynamics (QCD) contains the underlying symmetries governing the strong force 'felt'

Table 2.1
Fermions of the standard model of particle physics.

	1^{st}	2^{nd}	3^{rd}	Interactions
Quarks	$\begin{pmatrix} u \\ d \end{pmatrix}_L$	$\begin{pmatrix} c \\ s \end{pmatrix}_L$	$\begin{pmatrix} t \\ b \end{pmatrix}_L$	Weak, electromagnetic, strong
	u,d,	c, s	b, t	Electromagnetic, strong
Leptons	$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$	Weak, electromagnetic
	e_R	μ_R	τR	Electromagnetic

by particles with color charge, such as quarks. The symmetries of the electroweak force are contained by the $SU(2)_L$ and the $U(1)_Y$ gauge groups.

2.2 Beyond The Standard Model

In spite of the success of the SM, there are an increasing number of observations that suggest that the SM is an incomplete approximation of a more fundamental description of nature. For instance, in the SM neutrinos are massless, however experiments show that a neutrino detected as one type (e.g. electron neutrino) can later be detected and identified as a different type of neutrino (e.g. muon neutrino). This phenomenon, called neutrino oscillations, suggests that neutrinos have a mass and thus are not fully consistent with the SM. Another shortcoming of the SM it that, even though the strength of gravitational interactions is much smaller than the other fundamental forces and thus does not play a role in collider experiments, gravity has not yet been incorporated to the theory. Astrophysical observations indicate a large discrepancy in the mass of large astronomical objects when calculated via their gravitational effects as opposed to their luminous effects. This suggests a type of matter that does not interact via the electromagnetic force and thus cannot be observed by optical means. This type of matter is known as dark matter (DM) and it is calculated to be about six times more abundant than common matter. The SM

does not provide a good candidate to account for the DM content in the universe. These are only a few of the questions that have not yet been explained within the framework of the SM.

In an attempt to answer some of these questions, many theories of physics beyond the SM (BSM) have been developed. One of such theories is SUSY which is, perhaps currently, the most appealing. In the next section a brief description of SUSY is presented.

2.3 Supersymmetry

Without a doubt, symmetry is one of the most important concepts in physics and a fundamental characteristic of modern particle physics. Finding an "ultimate" symmetry to describe nature has been a goal for many physicists. SUSY might prove to be that supreme symmetry since, under quite general assumptions, is the largest attainable symmetry of the S-Matrix [2]. It is a symmetry that relates the fermionic multiplets which are the basic constituents of matter with the force carriers which are bosons. SUSY is an extension of the Poincaré group that combines states of different angular momentum into a single irreducible multiplet. The generators of SUSY do not commute with the generators of the Poincaré group. The underlying property of these generators is that they carry angular momentum and thus relate fermions and bosons. The SUSY algebra is obtained by grading the Poincaré algebra by introducing four spinor anticommuting generators Q_α in addition to the ten generators of the Poincaré group. Thus, giving us a total of fourteen hermitian operators that obey the following rules:

$$[Q_\alpha, P_\mu] = 0 \tag{2.1}$$

$$[Q_\mu, J_{\nu\rho}] = 0 \tag{2.2}$$

$$\{Q_\mu, Q_\nu\} = 2P_{\mu\nu} \tag{2.3}$$

And the Poincaré generators obey the familiar rules:

$$[P_\mu, P_\nu] = 0 \tag{2.4}$$

$$[J_{\mu\nu}, P_\rho] = \eta_{\mu\rho}P_\nu - \eta_{\nu\rho}P_\mu \tag{2.5}$$

$$[J_{\mu\nu}, J_{\rho\sigma}] = \eta_{\mu\rho}M_{\nu\sigma} - \eta_{\mu\sigma}M_{\nu\rho} + \eta_{\nu\sigma}M_{\mu\rho} - \eta_{\nu\rho}M_{\mu\sigma} \tag{2.6}$$

Here the $P's$ are the translator generators, the $J's$ represent the lorentz transformations and η is the well known Minkowski metric. Looking at the rules for the superalgebara we see immediately that the $Q's$ are translation invariant, (Eq. 2.1). Equation 2.2 transforms as a Dirac spinor under the homogeneous Lorentz transformations and 2.3 implies that the Q's are not independent. Extensions to the SUSY can be found by adding an index $j = 2 \ldots N$ to the supergenerators Q_α thus having any number of SUSY generators.

2.4 Motivation for SUSY Searches at the LHC

Supersymmetry is one of the leading theories for BSM physics since it allows for the unification of gauge couplings while also providing a candidate for DM in the form of the LSP. SUSY models predict a spectrum of new particles similar to those of the SM but with spins that differ by one-half to their respective SM partners. In the case of gauge mediated breaking, the LSP is often the gravitino [3], while in supergravity inspired models the lightest neutralino ($\tilde{\chi}_1^0$) is often the LSP. Colored SUSY particles can be copiously pair-produced at the LHC and will very often decay into color singlet states that ultimately decay to an LSP. Since these SUSY particles have not been observed by previous experiments, they are expected to be heavy (TeV scale). Therefore, a typical SUSY signature at the LHC consists of a high multiplicity of highly energetic jets and a large momentum imbalance in the detector.

Additionally, the final state containing tau leptons is particularly important and interesting in many SUSY models. For example, in gauge mediated SUSY, the NLSP (often the slepton) can decay to a lepton and a gravitino. Additionally, the couplings to third generation leptons can be enhanced in many regions of parameter space, thus leading to final states that predominantly contain tau leptons. Minimal supergravity (mSUGRA) is another model in which the cascade decays of SUSY particles can consist of tau leptons. Minimal supergravity is one of the most widely studied models due to it's ability to resolve a number of problems with the SM [4,5] and also naturally providing a cold dark matter candidate with the correct annihilation cross sections in the early universe to create the amount of dark matter relic density observed today by the Wilkinson Microwave Anisotropy Probe (WMAP) [6]. Although the analysis presented is not limited to a particular BSM theory, mSUGRA is chosen as a benchmark due to the simplicity of the model as it contains only four parameters and one sign to determine all masses and couplings.

An interesting feature of mSUGRA is that certain regions of parameter space give rise to SUSY mass hierarchies where the lightest slepton, $\tilde{\tau}_1$, and the lightest neutralino, $\tilde{\chi}_1^0$, are nearly mass degenerate *i.e.*, $\Delta M = M_{\tilde{\tau}} - M_{\tilde{\chi}_1^0} \sim 5 - 20$ GeV/c^2. Thus, the $\tilde{\chi}_2^0 \to \tau\tilde{\tau} \to \tau\tau\tilde{\chi}_1^0$ decays are dominant and the branching ratio for $\tilde{\chi}_2^0 \to l\tilde{l} \to ll\tilde{\chi}_1^0$, where $l = e, \mu$, is effectively zero. Therefore, a search for new physics in final states with high energetic jets, large momentum imbalance, and tau leptons can provide a "golden" mode for discovery. This stau-neutralino coannihilation region has extremely significant cosmological implications as it provides the much needed $\tilde{\tau}_1$–$\tilde{\chi}_1^0$ "coannihilation" in the early universe to produce the correct DM relic abundance observed today. The stau-neutralino coannihilation region is particularly difficult to identify due to the production of very low energy taus from $\tilde{\tau}_1 \to \tau\tilde{\chi}_1^0$ ($p_T^\tau \sim$ 15 GeV before decaying). Because the analysis presented is sensitive to the stau-neutralino coannihilation region where the identification of low p_T taus is difficult and backgrounds are large, sensitivity to many other models and regions of SUSY

parameter space is naturally achieved. Therefore, the stau-neutralino coannihilation region provides a very good benchmark model for designing a search for BSM physics processes with highly energetic jets, large momentum imbalance, and tau leptons.

3 EXPERIMENTAL TOOLS

3.1 The Large Hadron Collider

The LHC is a two-ring superconducting particle accelerator located at the European Center for Nuclear Research (CERN), in the French-Swiss border near the city of Geneva, Switzerland. The LHC rings are located approximately 100 meters underground in a 26.7 km tunnel originally build for the Large Electron Collider [7]. The two LHC rings carry hadron beams traveling in opposite directions at speeds close to the speed of light. These beams are crossed at four interaction points distributed around the LHC circumference. Figures 3.1 and 3.2 show an aerial view of the LHC complex and a schematic view of the CERN accelerator complex, respectively.

The charged particle beams are guided around the LHC ring by over a thousand superconducting dipole magnets built from coils of electric cable capable of operating in a superconducting state, thus conducting electricity efficiently and without resistance or energy losses. To keep the magnets in a superconducting state they must be cooled to extremely low temperatures, approximately -270 degrees C. The probability that two subatomic particles will interact as the beams cross is small. In order to increase the probability of an interaction, the protons are arranged in bunches and squeezed. This is done so that the protons within the bunches will be situated within a smaller cross-section. To achieve the squeezing of the proton bunches quadrupole magnets are used. Each quadrupole magnet measures between 5 and 7 meters in length.

The LHC hosts multiple experiments. Located at one of the four interaction points, the Compact Muon Solenoid (CMS) is a multi-purpose detector aimed at exploring physics at the TeV scale. The analysis presented in this dissertation makes use of the 5 fb^{-1} of p-p collision data collected by CMS at a center of mass energy $\sqrt{s} = 7\text{TeV}$, delivered by the LHC in 2011. In the following section a description of CMS is afforded to the readers.

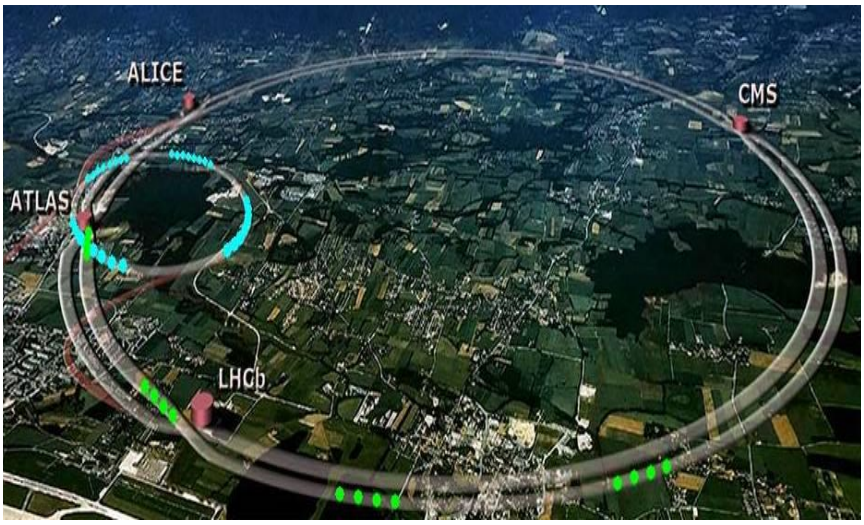

Fig. 3.1. Aerial view of the LHC [8].

3.2 The CMS Detector

Installed about 100 meters underground near the French village of Cessy, the CMS detector is one of six multi-purpose detectors along the LHC main ring that are designed to probe a wide variety of physics processes [7]. Located on the opposite side of the ring is the ATLAS detector. CMS and ATLAS are the two largest experiments at the LHC.

The CMS detector was designed to meet the goals of the LHC program summarized as follows [9]:

- Dimuon mass resolution of $\sim 1\%$ at 100 GeV as well as good muon resolution over a wide variety of angles and momenta and the ability to determine muon charge unambiguously for muons with momentum, $p < 1$ TeV.

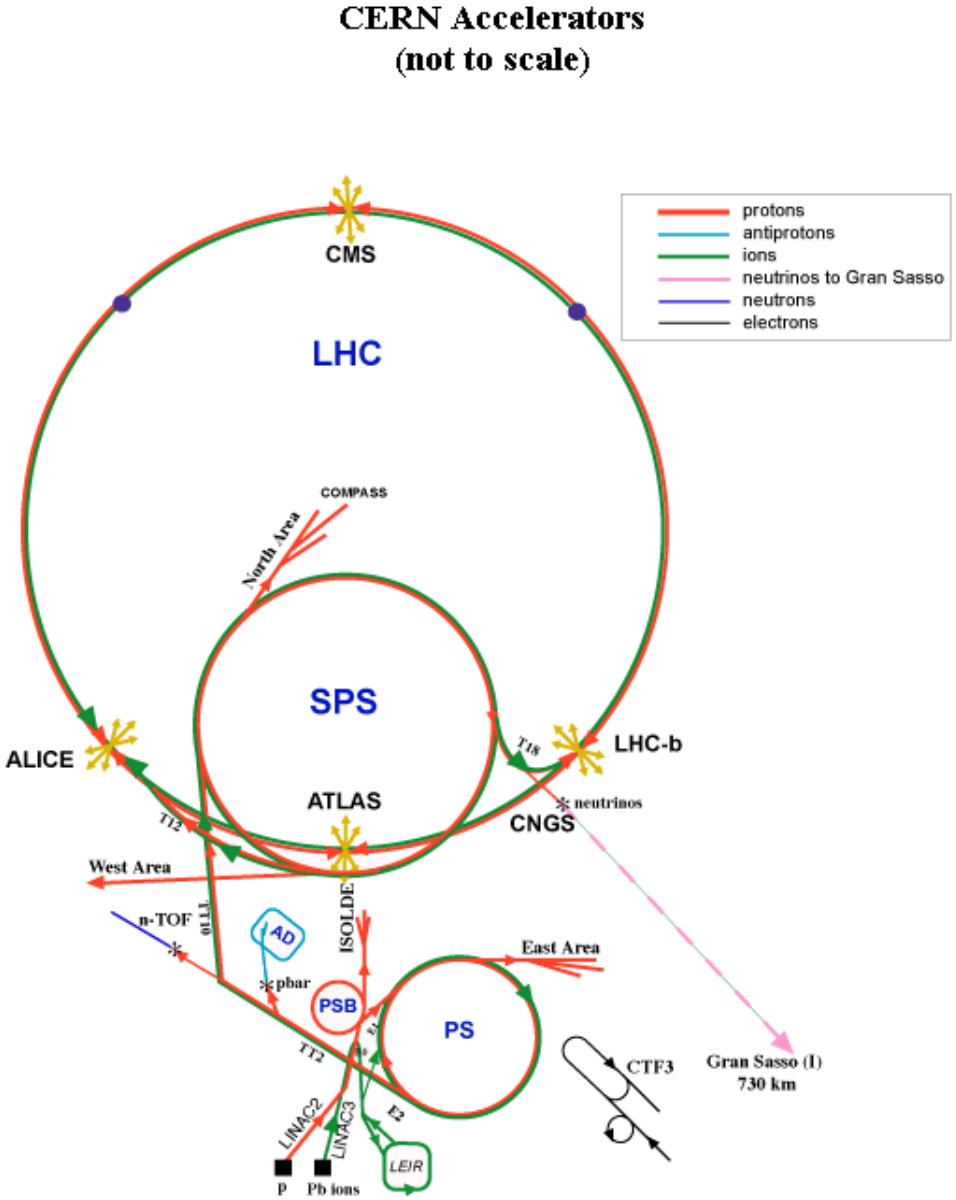

Fig. 3.2. CERN accelerator complex [10].

- A pixel detector close to the interaction point is required in order to ensure efficient triggering and tagging of $\tau's$ and b-jets.

- To obtain a diphoton and dielectron mass resolution of $\sim 1\%$ at 100 GeV, a good electromagnetic energy resolution is required.

- To achieve good missing-transverse-energy and dijet mass resolution, calorimeters with a large hermetic coverage and fine lateral segmentation are vital.

The apparatus itself is made of many layers of materials that use the properties of particles to detect and measure the charge and momentum of said particles to a high degree of precision. Closest to the interaction point is the tracking system which identifies the path (tracks) of passing particles and matches them to an originating vertex. The first three layers of the tracking system make up the silicon pixel and the next 10 make up the strip tracker. Surrounding the tracking system are the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL), respectively. The ECAL system also includes a silicon sensor preshower detector in front of its endcaps. Around the calorimeters at 7m in diameter is perhaps the central feature of this apparatus, the superconducting magnet, which provides a field of 3.8 T. Muons are measured in gas-ionization detectors embedded in the steel return yoke. In addition to the barrel and endcap detectors, CMS has extensive forward calorimetry. The relative luminosity is measured using the forward calorimeters. Collision events are selected by a first level trigger made of a system of fast electronics and a higher level trigger that consists of a farm of commercial CPUs running a version of the offline reconstruction optimized for fast processing. A detailed description of the CMS detector coordinate system and subdetector systems is presented in the following sections. The overall layout of the CMS detector and its substructure is shown in Figures 3.3 and 3.4.

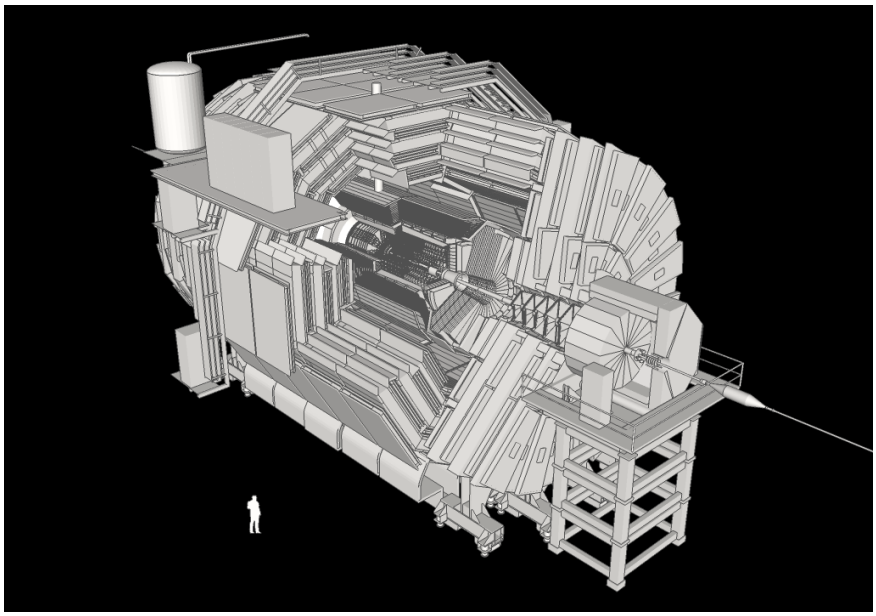

Fig. 3.3. Schematic drawing of the CMS detector.

3.2.1 CMS Coordinate System

In the CMS experiment, a right-handed coordinate system is used [7]. In this coordinate system the origin is placed at the nominal interaction point with the x-axis pointing towards the center of the LHC ring and the y-axis pointing perpendicularly to the LHC's ring plane. The z-axis is set along the counterclockwise-beam direction. The polar angle, θ, is measured from the positive z-axis and the azimuthal angle, ϕ, is measured in the x-y plane. The pseudorapidity is defined in the usual way (*i.e.* $\eta = -\ln[\tan(\theta/2)]$).

3.2.2 CMS Tracker System

Surrounding the interaction point, the CMS tracking system is designed to efficiently measure the trajectories of charged particles as well as to provide precise reconstruction of secondary vertices. Due to the LHC design specifications a detector

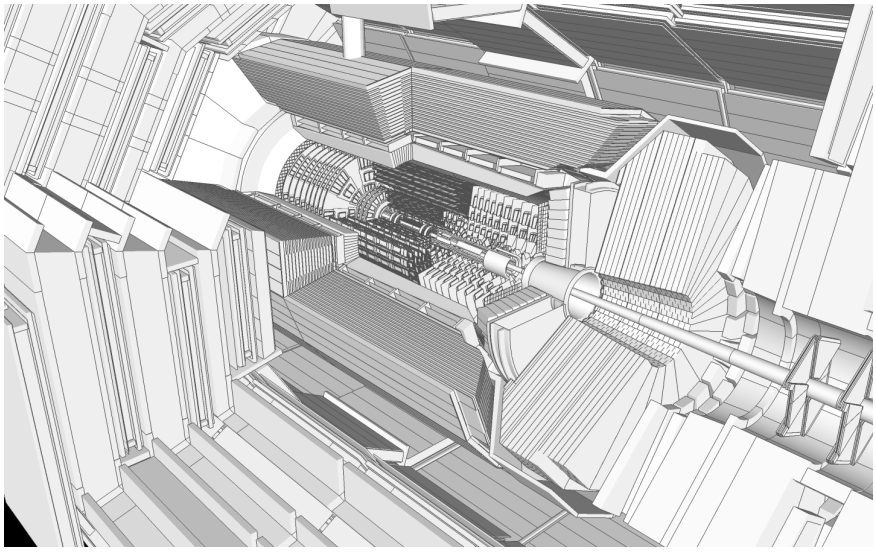

Fig. 3.4. Drawing of the CMS detector showing inner sub-systems.

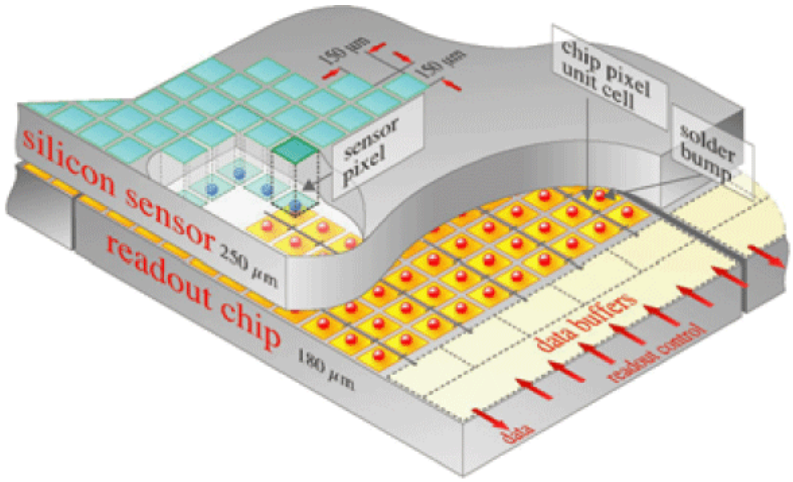

Fig. 3.5. CMS Pixel detector layout.

with high granularity and fast response is required to reliably identify the trajectories and attribute them to the correct bunch crossing [11]. The entire CMS tracker is composed of 1440 pixel and 15,148 strip detector modules.

Pixel Detector

The pixel detector is the CMS system closest to the interaction point. It contains 65 million pixels with a pixel cell size of 100 $\times$ 150 μm^2 covering a pseudorapidity of -2.5 $< \eta$ 2.5, which matches the acceptance of the central tracker. This detector is essential for the reconstruction of primary and secondary vertices of particles such as tau leptons and b quarks. It consists of three 53 cm long barrel layers located at mean radii of 4.4, 7.3, and 10.2 cm, and two endcap disks extending from 6 to 15 cm in radius. Figure 3.5 shows the pixel detector layout.

Silicon Strip Tracker

Immediately following the pixel detector, the silicon strip extends to a radius of 1.3 m. There are a total of 10 layers of silicon strip detectors, four tracker inner barrel (TIB) layers with two inner endcap disks (TID), and six outer barrel (TOB) layers. The TIB uses silicon sensors of 10 cm $\times$ 80 μm and a thickness of 320 μm. The TOB uses silicon sensors 25 cm $\times$ 180 μm and 500 μm of thickness [11]. On both sides, the tracker endcap (TEC) sensors are used with a thickness of 320 μm in the inner four rings and a thickness of 500 μm in the outer rings.

3.2.3 The Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) at CMS is designed to identify and reconstruct electrons and photons. Photodetectors are attached to the back of the crystals to detect scintillation light and convert it to electrical signal. Figure 3.6

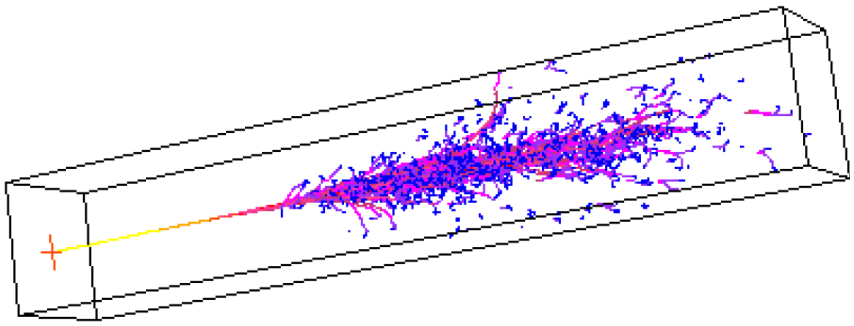

Fig. 3.6. Electromagnetic shower created due to the interaction of an electron with the ECAL crystals.

shows an illustration of an electromagnetic shower initiated by an electron interaction with the ECAL crystal.

The ECAL is a hermetic homogeneous calorimeter composed of over 61,000 lead tungstate ($PbWO_4$) crystals forming the barrel, and over 14,000 more in both endcaps (see Figure 3.7). The ECAL system also includes a silicon sensor preshower detector in front of its endcaps. Avalanche photodiodes (APDs) are used as photodetectors in the barrel while the endcaps use vacuum photodiodes (VPTs) [11]. Figures 3.8-3.9 show pictures of the barrel and endcap crystals with the photodetectors attached.

3.2.4 The Hadron Calorimeter

The hadron calorimeter (HCAL) at CMS is designed to measure the energy of hadrons and is of particular importance to the measurement of missing transverse momentum. The HCAL system is divided into four subsystems: the barrel (HB), endcaps (HE), outer (HO), and forward (HF). The HB and HE are sampling calorimeters made of layers of absorber materials (70% Cu, 30% Zn) and scintillator materials (Bicron BC408) [11]. The HB is itself divided into two half-barrel sections (HB+ and HB-) and extends to $|\eta|$ <1.3, the HE extends the η coverage to 3. A picture of

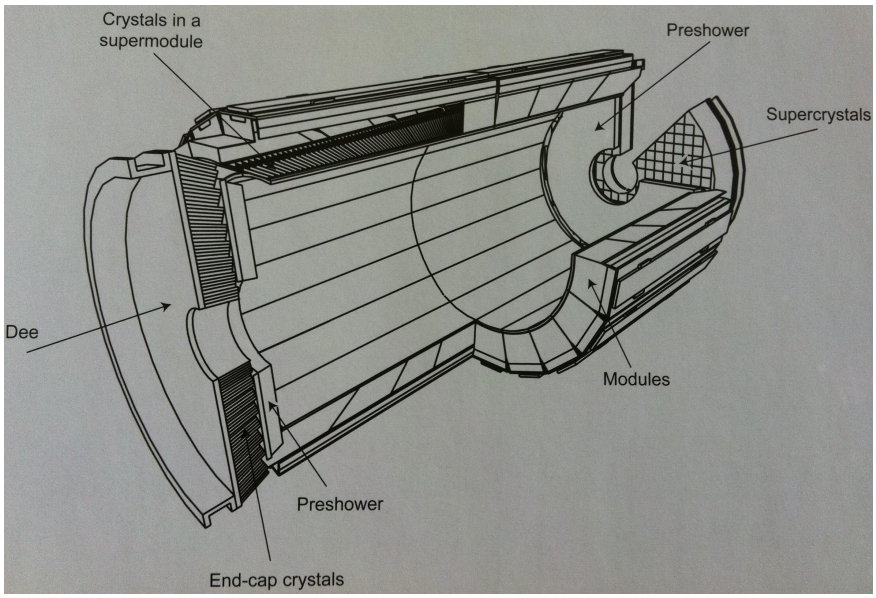

Fig. 3.7. The CMS Electromagnetic Calorimeter.

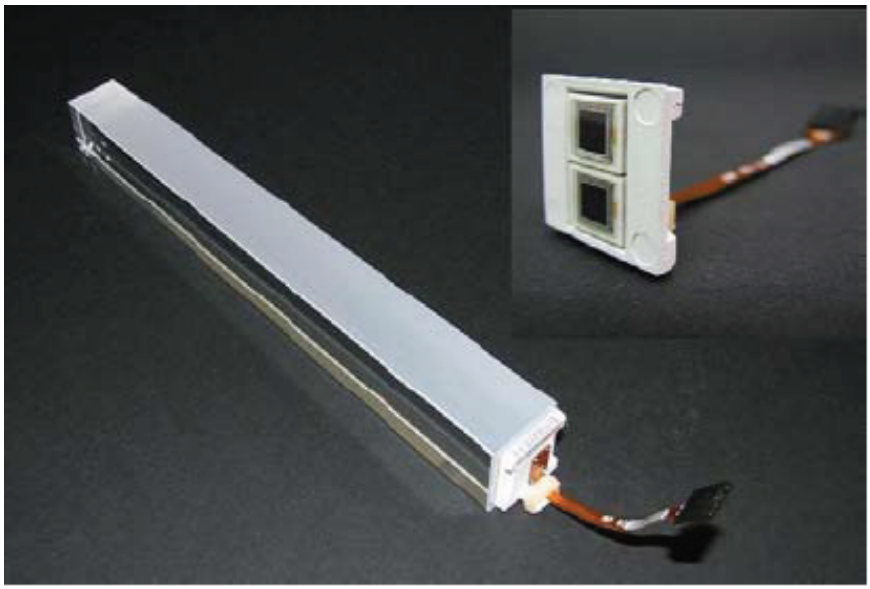

Fig. 3.8. A barrel crystal and the APD capsule [11].

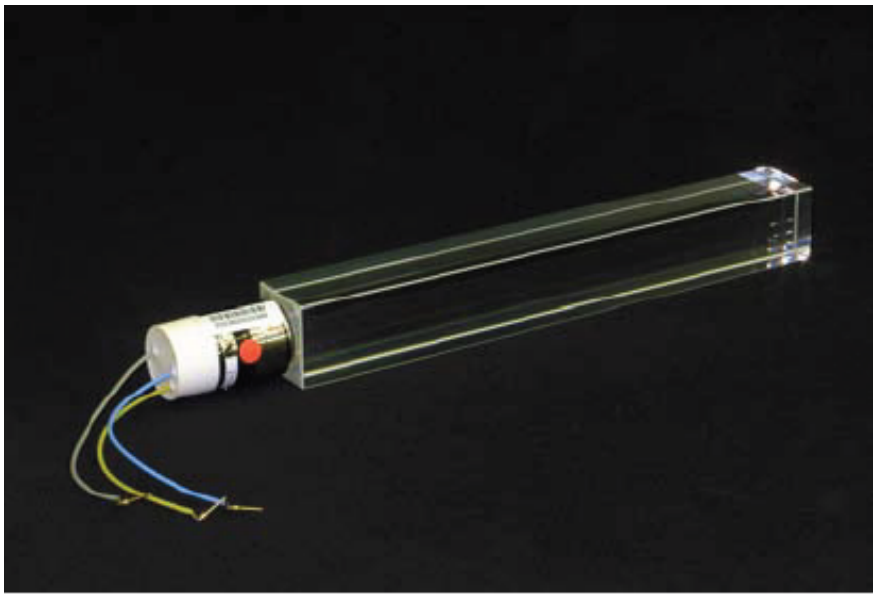

Fig. 3.9. An endcap crystal and VPT [11].

an HCAL half-barrel can be seen in Fig. 3.10. Figure 3.11 shows a drawing of the HB wedge design.

In addition to the HB and HE the HO is placed outside the solenoid in the central pseudorapidity region where ECAL and HB stopping power is not sufficient to contain hadron showers. It is utilized to detect late starting showers and showers penetrating past the HB. Finally, at 11.m from the interaction point the HF uses Cherenkov-based technology and extends the pseudorapitidy coverage to $|\eta|$ =5.2. Figure 3.12 shows a longitudinal view of the CMS with the locations of HB, HE, HO and HF and their respective η coverage.

3.2.5 Muon System

The CMS muon system is composed of about 25,000 m^2 of detection planes [11]. It was designed to be capable of reconstructing the charge and momentum of muons over the kinematic range of the LHC. Like other sub-systems the muon system has

a cylindrical barrel, and two endcaps on each end. It uses 3 types of gas particle detectors to adequately identify muons: drift tube chambers (DT), cathode strip chambers (CSC), and resistive plate chambers (RPC). Figure 3.13 shows a depiction of a muon traveling through the CMS barrel.

The barrel section of the muon system consists of 4 concentric stations and about 172,000 sensitive wires. The inner 3 cylinder stations have 60 DTs while the outer one has 70. A DT chamber is made of 2 or 3 superlayers (SL), where each SL is made of 4 layers of staggered drift cells. The endcap regions of the muon system uses CSCs to identify muons between $|\eta|$ values of 0.9 and 2.4. There are 4 stations of CSCs in each of the two endcaps, each CSC consists of six anode wire planes intervaled among 7 cathode panels. Figure 3.14 shows a quarter plane view of the CMS detector with the CSCs highlighted. In addition to the DTs and CSCs the muon system has a total of 6 layers of RPCs in the barrel region and a plane of RPCs in each of the first 3 station in the endcaps. The RPCs are gaseous parallel-plate detectors capable of tagging the time of an ionizing events in much less than 25 ns, which is the LHC's bunch crossing (BX) target at design luminosity (in 2011 the BX were 75 ns and 50 ns).

3.2.6 Triggering

At design luminosity the LHC will deliver forty million BXs per second averaging more than twenty interactions per BX. Since it is not possible to process and store the massive amounts of data associated with this number of events, a system has to be in place to select the relevant physics events. The trigger system performs this task and thus is the first step in the event selection process for physics analyzes. The trigger system does the selection in two steps callet Level-1 Trigger (L1) and High-Level Trigger (HLT). The L1 Trigger consists of high performance electronics custom-designed for CMS while the HLT consists of software implemented in a farm of about one thousand commercial processors. The combined rate reduction for the

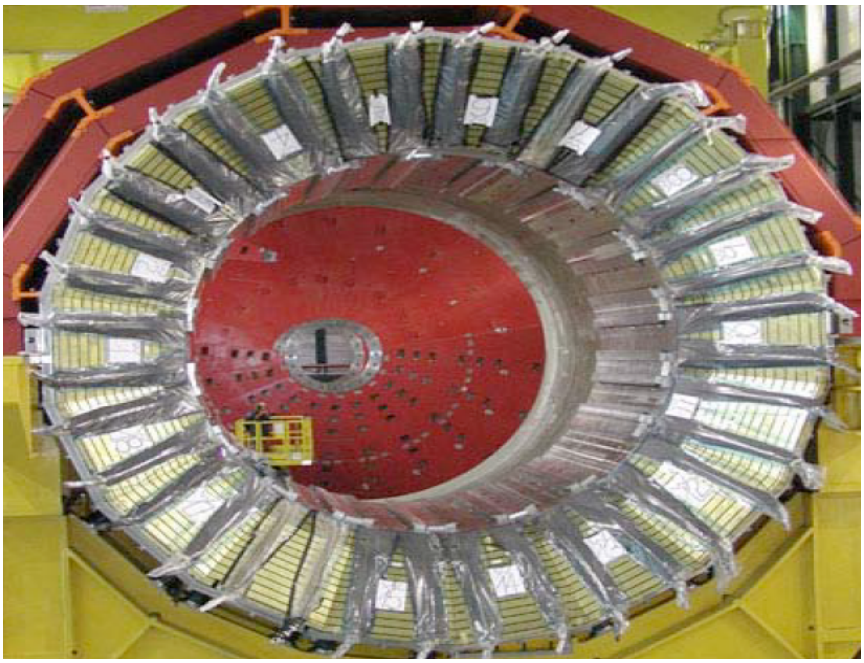

Fig. 3.10. CMS HCAL half-barrel in the assembly hall [11].

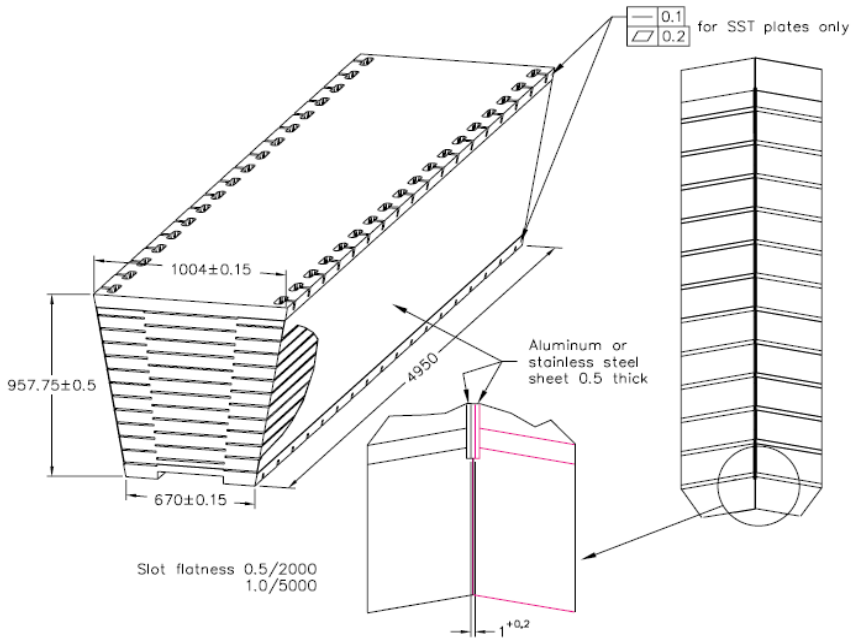

Fig. 3.11. View of the HB wedge showing scintillator sampling design [11].

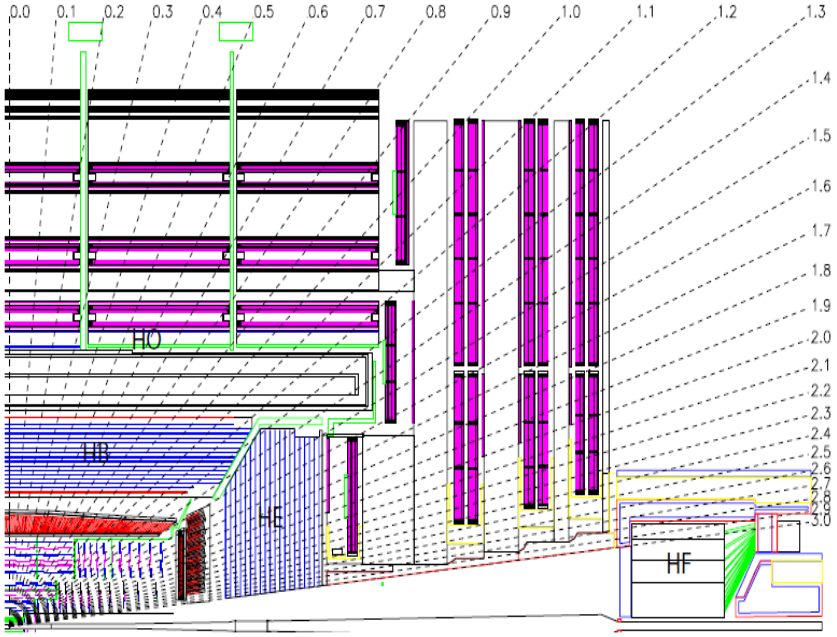

Fig. 3.12. Longitudinal view of the CMS detector [11].

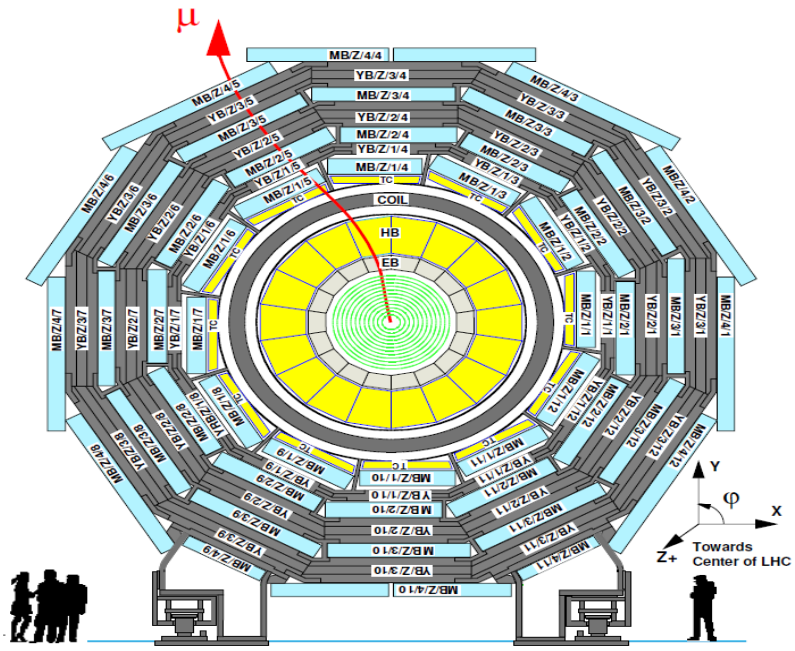

Fig. 3.13. Depiction of muon going trough the CMS detector.

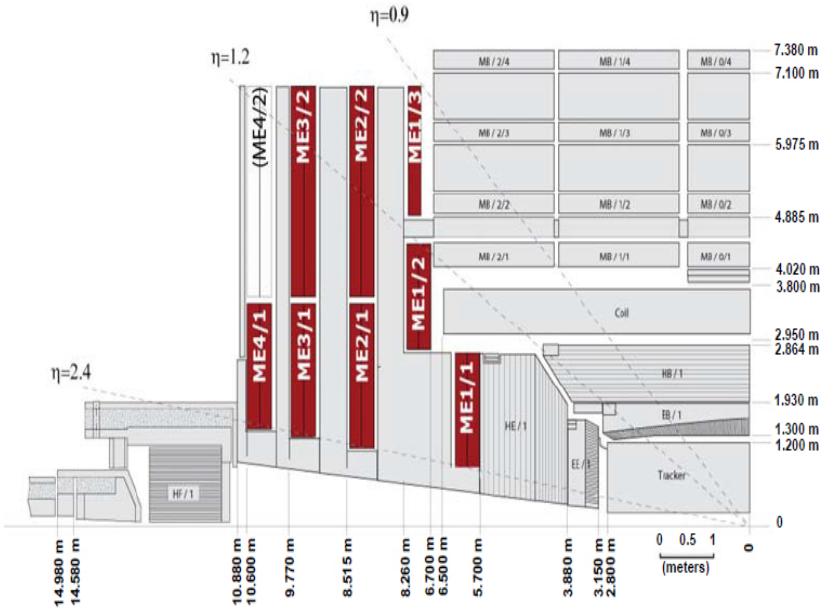

Fig. 3.14. Quarter panel view of the CMS detector.

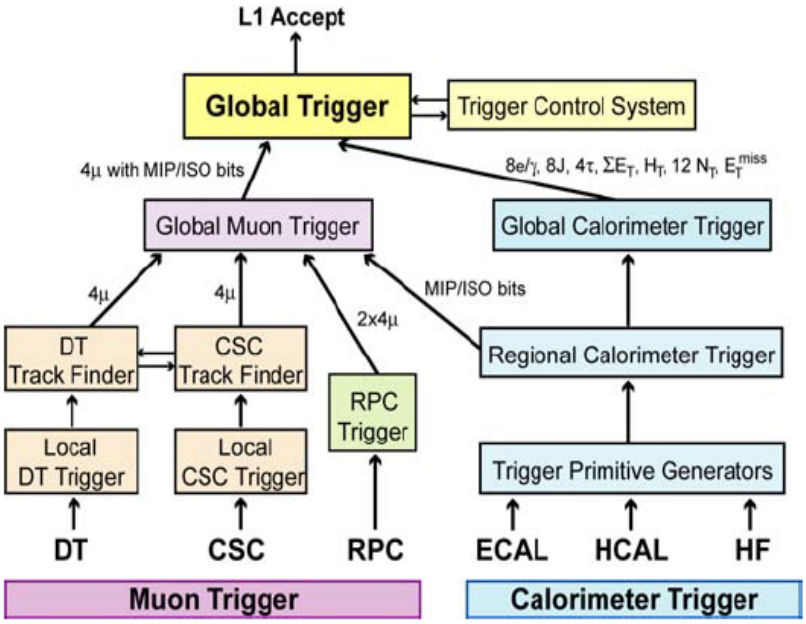

Fig. 3.15. Architecture of the Level-1 Trigger [11].

L1 and HLT combined is in the order of 10^6. Figure 3.15 shows a schematic of the L1 Trigger system.

3.2.7 Data Acquisition

The CMS data acquisition (DAQ) system is capable of recording nearly 100 GB of data per second from approximately 650 data sources while providing enough computing power for the HLT. Table 3.1 shows the number of data sources for every CMS sub-system. Figure 3.16 shows a schematic view of the CMS DAQ architecture.

Table 3.1
Number of sub-detector data sources [11].

Sub-detector	Number of data sources
Tracker pixel	40
Tracker strips	440
Preshower	56
ECAL	54
HCAL	32
Muons CSC	8
Muons RPC	3
Muons DT	10
Global Trigger	3
CSC, DT Track Finder	2
Total	626

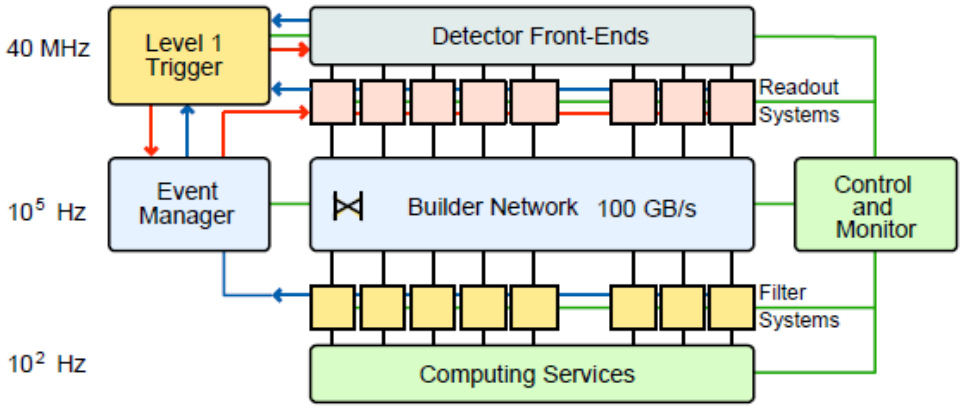

Fig. 3.16. Architecture of the CMS DAQ System [11].

4 TAU LEPTONS

Since the reconstruction and identification of taus is key to this analysis, it is appropriate to dedicate a chapter to give a general overview of the tau lepton and some of its physical properties. The present chapter will provide the necessary introduction to tau leptons and in Chapter 6, the discussion is followed by details on tau identification in the CMS experiment.

The tau discovery was made in at the Stanford Linear Accelerator Center (SLAC) in experiments carried out in 1975 using the SPEAR electron positron colliding ring [12]. Due to its large mass, it can decay leptonically as well as hadronically (see Figs. 4.1–4.2). One third of taus decay leptonically ($\tau \rightarrow \nu_\tau l \nu_l$ where l is a light lepton). The remainder decay into hadronic jets (in this dissertation, the notation τ_h is used to represent a hadronically decaying tau). In the latter case, tau jets consist of final states with one, three, or (rarely) five charged mesons usually accompanied by one or more neutral pions.

The tau lepton is a heavy electron-like particle. It's the heaviest known lepton with a mass of 1.777 GeV/c^2 and a mean lifetime of 2.9×10^{-13} seconds. Because of the tau's short lifetime it can only be detected indirectly by looking at its decay products. Furthermore, when a tau decays leptonically the light leptons cannot be distinguished from electrons or muons produced by other processes. Therefore only hadronically decaying taus modes are considered in this analysis.

Since the tau lepton decays to $e\overline{\nu}_e\nu_\tau$ (17.8%); $\mu\overline{\nu}_\mu\nu_\tau$ (17.4%); and hadrons (τ_h) + ν_τ (64.8%), there are six distinct possible final states of ditau decays, namely ee (3.1%), $\mu\mu$ (3.1%), $e\mu$ (6.2%), $e\tau_h$ (23.1%), $\mu\tau_h$ (22.5%), and $\tau_h\tau_h$ (42%). In general ditau related searches, the $\mu\mu$, ee, and $e\mu$ final states have the lowest background contamination, while the double hadronic final state $\tau_h\tau_h$ has the largest branching ratio but suffers from large QCD jet backgrounds.

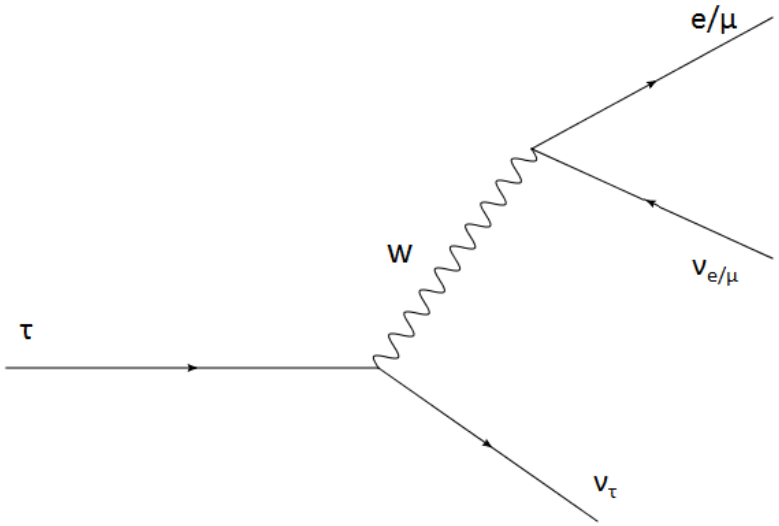

Fig. 4.1. Feynman diagram of leptonic tau decays.

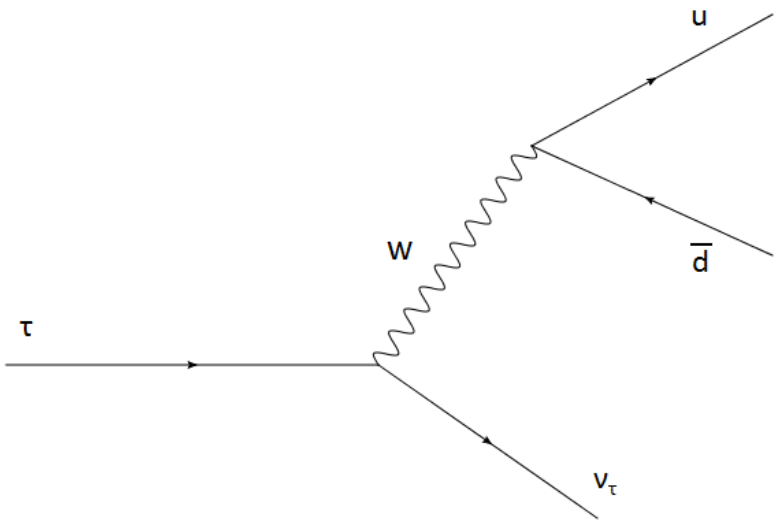

Fig. 4.2. Sample Feynman diagram of a hadronic tau decay mode.

5 ANALYSIS STRATEGY

In R-parity conserving SUSY the dominant production mechanisms at the LHC are $\tilde{q}\tilde{q}$, $\tilde{q}\tilde{g}$, and $\tilde{g}$ pair production. In the case of $\tilde{q}\tilde{g}$ pair production the presence of SUSY is likely to reveal itself in final states with high p_T jets. This is due to its heavy nature ($M_{\tilde{q},\tilde{g}}$ at TeV scale) and the frequency of decays to neutralinos plus quark/gluons (e.g. $\tilde{q} \rightarrow q\tilde{\chi}_2^0$). Additionally, because the LSP is heavy, weakly interacting and stable, events are expected to have a large momentum imbalance. In general, these SUSY events have at least two high p_T jets and large missing transverse energy (MET, $\not\!\!E_T$). Furthermore, some SUSY models and regions of parameter space (e.g. mSUGRA at high tanβ) can give rise to $\tilde{\chi}_2^0 \rightarrow \tau\tilde{\tau} \rightarrow \tau\tau\tilde{\chi}_1^0$ decays with a branching ratio close to 100%. It is in this region that the presence of SUSY can be discovered in final states containing a pair of tau leptons (see Chapter 4).

Due to the degeneracy between the $\tilde{\tau}$ and $\tilde{\chi}_1^0$, taus from $\tilde{\tau}_1 \rightarrow \tau\tilde{\chi}_1^0$ decays are expected to be low energy τ's. Therefore, these events will have a pair of tau leptons in the final state, one of which would be a low p_T tau. As explained in Chapter 4 a search using two τ_h's in the final state would suffer from a large QCD contamination. However, the analysis presented in this dissertation does not suffer from large QCD contribution because of the requirement of a large $\not\!\!E_T$. Then,it is preferable to make use of the double hadronic final state $\tau_h\ \tau_h$ since it has a larger branching fraction ($\sim$ 65%) as well as a smaller energy loss due to neutrinos ($p_T^{\tau_h} \sim \frac{1}{2}p_T^{\tau}$).This dissertation focuses on the search for SUSY in the double hadronic tau final state.

To select the trigger for this analysis one must consider the following. The efficiency for events passing a double hadronic tau trigger is typically a rising function that plateaus at $\sim$ 85–90% per tau leg in the τ_h p_T range $p_T > 50$ GeV/c. Then using a double tau trigger would diminish the sensitivity to many BSM processes with low p_T taus. For this reason a trigger based on $\not\!\!E_T$ or $\not\!\!H_T$ must be selected (see

Chapter 7). The use of such trigger confines the search to high $\not\!\!E_T$ or $\not\!\!H_T$ regions but has the advantage that the τ_h selections remain unbiased and allow the lowest possible p_T thresholds for tagging τ_h candidates.

Finally to validate tau identification and ensure the robustness of the tau selections it is important to obtain a clean sample of $Z \rightarrow \tau\tau$ events. However in the double hadronic tau final state it is not possible to obtain such a sample due to the contamination from QCD. Instead, one can obtain relatively clean samples of $Z \rightarrow \tau\tau$ events in the $\mu\tau_h$ final state where the presence of the μ significantly reduces the QCD background. If a control sample of $Z \rightarrow \tau\tau$ events is achieved in the $\mu\tau_h$ final state with consistency in shape and event rates between collision data and MC simulation, then this indicates that the tau tagging criteria is robust and that the proper scale factors were applied. In order to ensure the robustness of the analysis and confidence in the results, collision data is used whenever possible to understand and validate the efficiency of reconstruction methods as well as the estimation of the background contributions. For that purpose control regions (CRs) are defined with most of the selections similar to those used in the main search but enriched with events from background processes. Once a background enhanced region is created, selection efficiencies are measured in those regions and extrapolated to the region where the signal is expected. In cases where a method based in collision data (data-driven method) is not possible scale factors between collision data and MC simulation are applied to estimate the background contribution to the signal region. Such scale factors are calculated as the ratio between the observed collision data events and the expected MC events in the background enhanced region.

Figure 5.1 shows the overall analysis strategy ranging from the definition of the signal region to the determination of CRs.

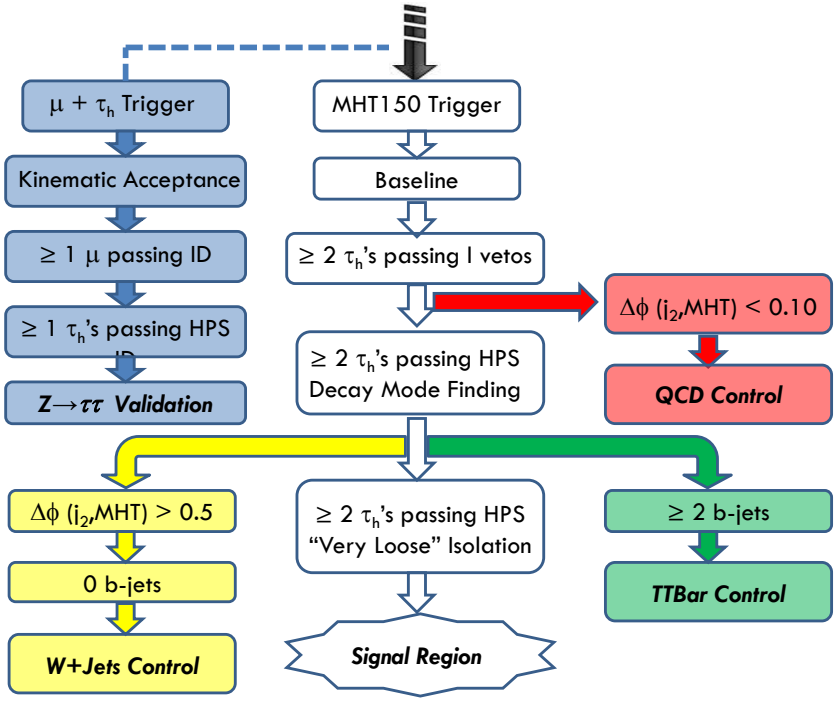

Fig. 5.1. Analysis strategy flow chart.

6 OBJECT RECONSTRUCTION AND IDENTIFICATION

6.1 Jet Reconstruction

In CMS there are several jet reconstruction algorithms available. In this analysis we use Particle Flow (PF) jets, which use information from all subdetectors to produce a mutually exclusive collection of particles which in turn are used as input to the jet clustering algorithms. PF jets are reconstructed using the Anti-KT jet clustering algorithm [13] with a reconstruction cone R=$\sqrt{\eta^2 + \phi^2}$. In this analysis a reconstruction cone with R = 0.5 is used. The PF jet objects are corrected using three levels of a factorized jet correction system: L1FastJet, L2Relative, and L3Absolute. L1 corrections are based on event energy densities and jet areas and remove additional contributions to the measured jet energies due to pile-up particles. The L2 and L3 corrections are based on Monte Carlo (MC) truth information and seek to remove any p_T and η dependencies.

6.2 Tau Reconstruction and Identification

Identifying taus at hadron colliders is particularly challenging because the physics signature of a τ_h will closely resemble that of a quark/gluon jet. Then, a major challenge in any τ_h identification algorithm is in discriminating against quark/gluon jets, which are produced with a cross-section several orders of magnitude larger than τ_h. To do this the CMS collaboration has developed several algorithms to reconstruct and identify τ_h leptons. In this analysis the method used is called the Hadrons Plus Strips (HPS) algorithm, which makes use of PF jets as inputs.

For any PF jet to be considered as a possible τ_h candidate it must first pass an isolation requirement that was optimized for this search. This requirement is referred to as "very loose" (VLoose) isolation and it has a value of ΔR = 0.3. This isolation

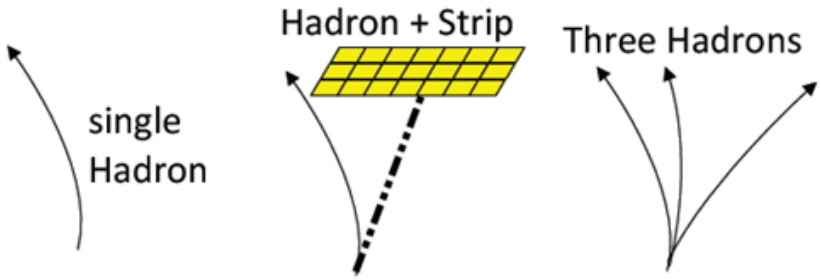

Fig. 6.1. Depiction of τ_h decay modes.

requirement allows for an increased signal acceptance while maintaining a low rate of fake τ_h leptons.

In HPS the PF jet information is passed to an algorithm that uses strips of clustered electromagnetic particles to reconstruct neutral pions. These strips ("neutral pions") are then combined with the charged hadrons within the PF jets to reconstruct the main tau decay modes (PF taus) outlined in Table 6.1. Figure 6.1 shows a depiction of a single hadron, one charge hadron plus one strip, and a three charged hadron decay mode.

Table 6.1
Reconstructed tau decay modes

HPS Tau Decay Modes
Single Hadron + Zero Strips
Single Hadron + One Strip
Single Hadron + Two Strips
Three Hadrons

6.3 Electron Reconstruction and Identification

Electrons are reconstructed using information from the tracker and ECAL detectors. Electron tracks are reconstructed by matching trajectories in the silicon strip tracker and matching to hits in the pixel detector. A Gaussian Sum Filter is used for the reconstruction of trajectories in the silicon strips. The track that best matches the energy supercluster in ECAL is chosen to be reconstructed. Two energy clustering algorithms are used to measure the energy of electrons and photons: one for the barrel and one for the endcaps. In addition, electrons are required to be isolated in the calorimeter and tracker. Calorimeter isolation requires recorded hits (RecHits) to have an energy threshold of > 0.1 GeV inside a cone of ΔR$= 0.4$ in the endcaps and > 0.08 GeV in the barrel.

6.4 Muon Reconstruction and Identification

In this analysis we use "global" muons, which are isolated by having minimal energy in a cone of ΔR$= 0.5$ around the lepton trajectory. Table 6.2 shows a complete list of the μ identification criteria.

Table 6.2
μ identification

"Global" μ
Tracker hits ≥ 10
Pixel hits ≥ 1
≥ 2 chambers with matching segments
Global fit $\chi^2/NDOF < 10$
≥ 1 hit in muon system
$\lvert d_0 \rvert < 0.2$ cm
Σp_T of iso. tracks /ECAL RecHits < 1

6.5 $\not\!\!E_T$ and $\not\!\!H_T$

The CMS detector is capable of detecting all known particles produced in pp collisions (except neutrinos) up to $|\eta| \sim 5$. This allows a precise measurement of the momentum-imbalance in the transverse direction relative to the beam also known as missing transverse energy or MET. The MET vector, $\vec{\not\!\!E}_T$, when calculated based on the PF algorithm is referred to as PFMET. PFMET makes use of the PF algorithms to reconstruct the momenta of individual particles:

$$\vec{\not\!\!E}_T = -\Sigma_i \vec{p}_T^{\,i}, \tag{6.1}$$

where the index i runs over all PF object candidates. In the proposed analysis we also define the missing momentum quantity $\not\!\!H_T =\mid \vec{\not\!\!H}_T \mid= \mid -\Sigma_i \vec{p}_T^{\,i} \mid$, with the sum running over all PF jets with transverse momenta $p_T > 30$ GeV. We favor $\not\!\!H_T$ over $\not\!\!E_T$ in order to reduce the PU dependence in our search.

7 TRIGGER AND EVENT SELECTIONS

7.1 Trigger Performance

The signal region of this analysis is selected by requiring events with at least two jets which pass a fully hadronic HLT trigger based on $\not{H}_T > 150$ GeV (HLT_PFMHT150). The HLT_PFMHT150 trigger is validated by obtaining a CR of pure $t\bar{t}$ events. This type of events is selected because their topology is very similar to that of the signal region in this analysis, with two high p_T jets from the top quark decay as well as real or fake τ_h's from the hadronic decay of the W bosons. To avoid a potential bias in the high $\not{H}_T$ region the events are selected using $\mu + \tau_h$ and $e + \tau_h$ cross-triggers. Additonally, the following selection criteria are applied to the sample:

Acceptance Selections:

- ≥ 1 l (l = Global muon or electron) with $p_T > 10$ GeV/c and $|\eta| < 2.1$
- ≥ 1 PFTau with $p_T > 15$ GeV/c and $|\eta| < 2.1$
- ≥ 1 $e/\mu\tau_h$ pair with $\Delta R(e/\mu, \tau_h) > 0.7$

Muon Selections:

- ≥ 1 μ with tracker hits ≥ 10
- ≥ 1 μ with pixel hits ≥ 1
- ≥ 1 μ with ≥ 2 matching segments
- ≥ 1 μ with χ^2/ndof < 10
- ≥ 1 μ $|d_0| < 0.2$ cm
- ≥ 1 μ passing isolation $\sum \mathrm{p_T}$ of tracks and Ecal RecHits < 1

Electron Selections:

- Electron must be a Ecal driven electron
- $|\Delta\eta_{in}| < 0.005$ in EB; $|\Delta\eta_{in}| < 0.007$ in EE
- $|\Delta\phi_{in}| < 0.09$ in EB and EE
- $H/E_m < 0.05$ in EB and EE
- $\sigma_{i\eta i\eta} < 0.03$ in EE
- $E^{2\times5}/E^{5\times5} > 0.94$ $||E^{1\times5}/E^{5\times5} > 0.83$ in EB
- Ecal Isolation : $\sum E_T^{ecal} < 4.5$ GeV
  ($E_{barrel}^{EcalRecHit} > 0.08$ GeV, $E_{T\,endcap}^{EcalRecHit} > 0.1$ GeV, $\Delta R_{iso} = 0.4$)
- Track Isolation: $\sum p_T^{trk} < 3.5$ GeV ($p_T^{trk} > 0.7$ GeV, $\Delta R_{iso} = 0.4$)

Tau Selections:

- ≥ 1 τ_h passing the HPS "tight" muon veto
- ≥ 1 τ_h passing the HPS "tight" electron veto
- ≥ 1 τ_h passing the HPS decay mode finding
- ≥ 1 τ_h passing the HPS "very loose" isolation

Topological Selections:

- $Q(e/\mu) \times Q(\tau_h) < 0$
- $\not\!\!E_{\mathrm{T}} > 30$ GeV
- ≥ 1 jet tagged as a b-jet using the TCHE working point

The HLT_PFMHT150 trigger efficiency is defined as follows:

$$\varepsilon = \frac{\text{Number of Events Passing Selections Above AND HLT\_PFMHT150}}{\text{Number of Events Passing Selections Above}} \tag{7.1}$$

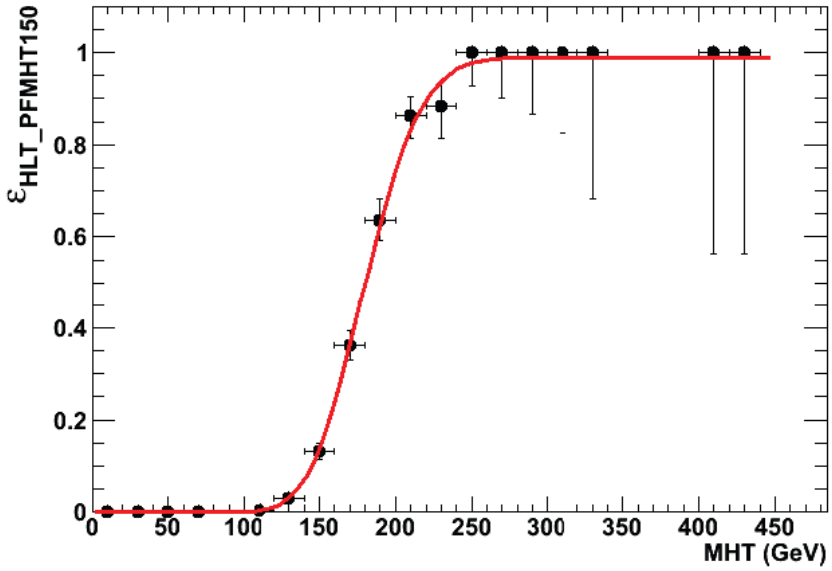

Fig. 7.1. Trigger turn-on curve.

Figure 7.1 shows the trigger HLT_PFMHT150 turn-on-curve extracted directly from data using the methodology described in this section. One can see the plateau is reached at $\not{H}_T > 230$ GeV, therefore validating the use of $\not{H}_T > 250$ GeV in this analysis. The HLT_PFMHT150 trigger efficiency can be directly measured from this validation sample, by fitting the trigger turn-on curve to an error function. The efficiency at the plateau is 98.9 $\pm$ 2.5 %.

7.2 Event Selection

The signal selections are outlined below:

Baseline selections:

- 1^{st} Leading Jet $p_T > 100$ GeV and $|\eta| < 3$
- 2^{nd} Leading Jet $p_T > 100$ GeV and $|\eta| < 3$
- $\not{H}_T > 250$ GeV

Tau selections:

- ≥ 2 τ_h's with $\mathrm{p_T} > 15$ GeV and $|\eta| < 2.1$
- ≥ 2 τ_h's passing the HPS muon veto
- ≥ 2 τ_h's passing the HPS electron veto
- ≥ 2 τ_h's passing the HPS decay mode finding
- ≥ 2 τ_h's passing the HPS isolation

Topological selections:

- 1^{st} Leading Jet separated from τ_h's ($\Delta R(j_1, \tau_h) > 0.3$)
- 2^{nd} Leading Jet separated from τ_h's ($\Delta R(j_2, \tau_h) > 0.3$)
- $\Delta\phi(j_2, \not{H}_T) > 0.5$
- ≥ 1 τ_h pair with $\Delta R(\tau_{h,i}, \tau_{h,j}) > 0.3$

Figure 7.2 shows p_T distributions for the highest p_T jets for an mSUGRA benchmark point ($m_0 = 360, m_{1/2} = 560, \tan\beta = 40, \mathrm{A_0} = -500$ GeV) as well as the major SM background contributions. This distribution shows that the requirement of a first and second leading jets to have a $p_T > 100$ GeV achieves a very good discrimination against backgrounds. Even though a higher threshold could be applied, the lower threshold of 100 GeV is kept to maintain a high signal acceptance and allow the highest possible sensitivity to BSM processes.

Figures 7.3-7.4 show the $\not{E}_T$ and $\not{H}_T$ distributions, respectively, after the requirements of the leading and next-to-leading jets. It is observed that the $\not{H}_T > 250$ GeV requirement achieves very good discrimination against SM backgrounds while maintaining high signal acceptance. Also the $\not{H}_T > 250$ GeV requirements is robust in terms of systematics because it allows the signal region to stay away from the trigger turn-on curve (Fig. 7.1) where systematic effects can be large.

In QCD events, a non-zero measurement of $\not\!\!H_T$ or $\not\!\!E_T$ is due to mismeasurements in jet energies so a back-to-back correlation between $\not\!\!E_T$ or $\not\!\!H_T$ and leading jets is expected. Thus, applying the requirement of $|\Delta\phi(j_2, \not\!\!H_T)| > 0.5$ is very effective in removing the QCD contamination in the signal region. A simulated event passing all selection criteria is shown in Figures 7.5–7.6.

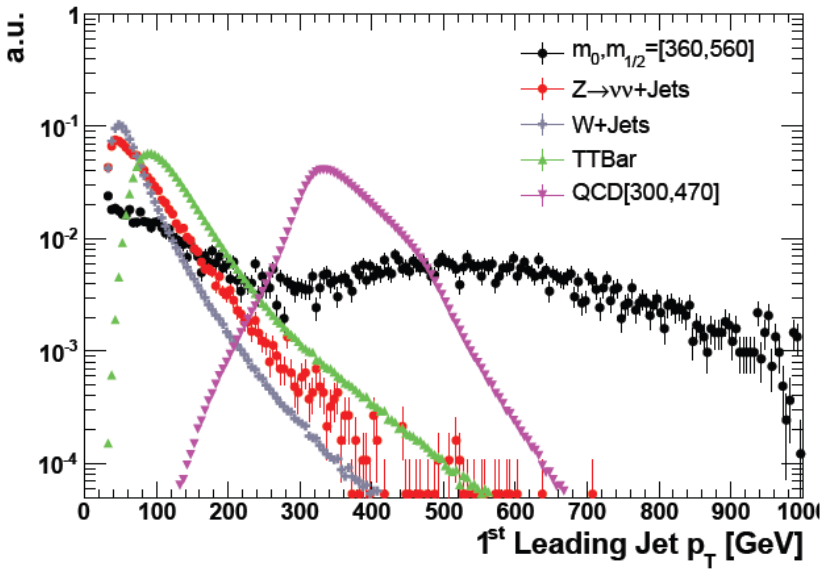

Fig. 7.2. p_T distribution for 1^{st} leading jet.

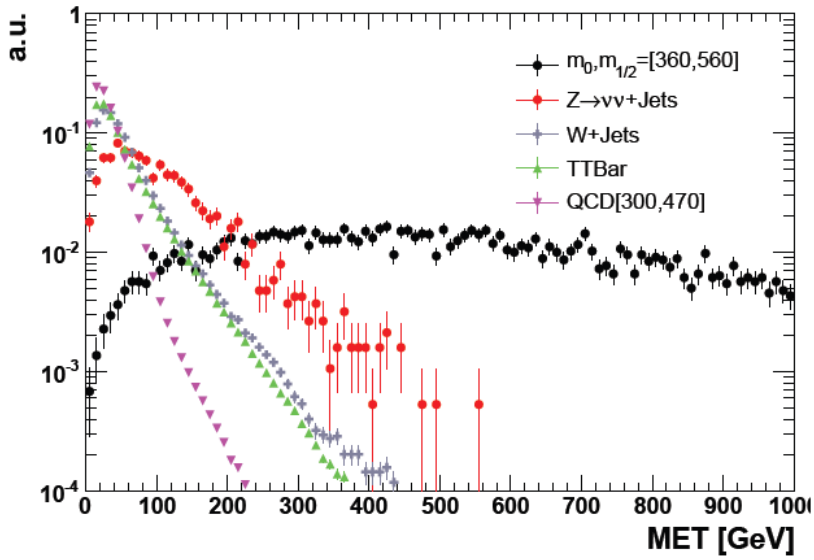

Fig. 7.3. $\not\!\!E_T$ distribution.

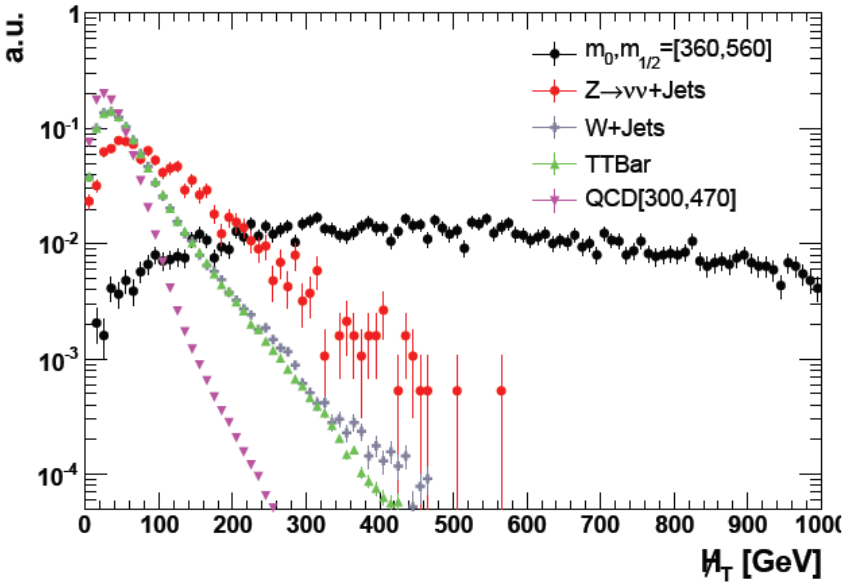

Fig. 7.4. $\not\!\!H_T$ distribution.

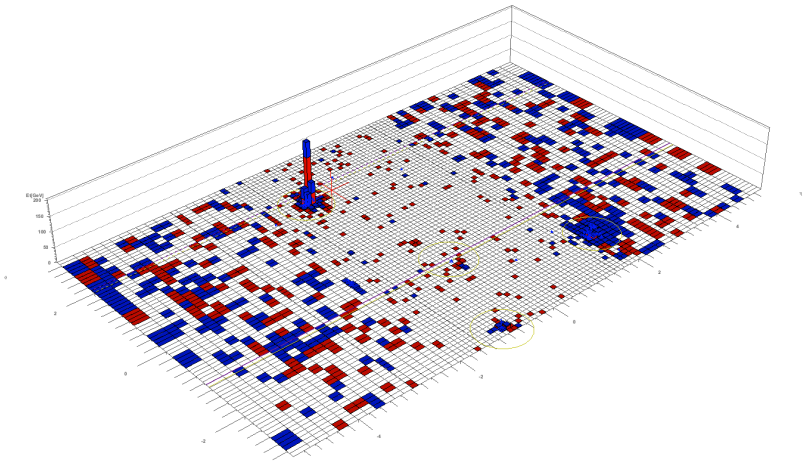

Fig. 7.5. Simulated SUSY event: 3D Lego View.

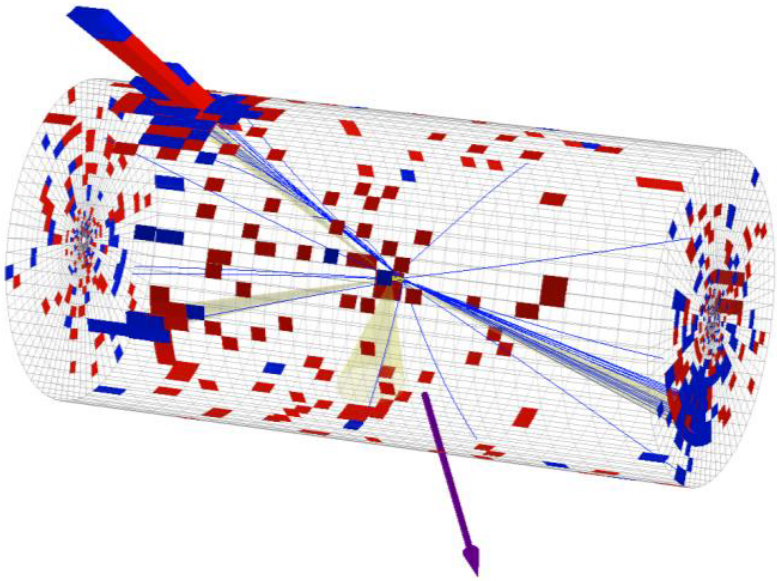

Fig. 7.6. Simulated SUSY event: 3D View.

8 BACKGROUND ESTIMATION

In order to ensure robustness of the analysis and our confidence in the results, whenever possible we rely on the data itself to understand and validate the efficiency of reconstruction methods as well as the estimation of the background contributions. For that purpose we define control regions with most of the selections similar to what we use in our main search but enriched with events from background processes. Once a background enhanced region is created, we measure selection efficiencies in those regions and extrapolate to the region where we expect to observe our signal. In cases where a complete data–driven method is not possible we make use of scale factors, ratio between observed data events and expected MC events in the background enhanced region to estimate the background contribution in the signal region. Figure 5.1 displays the general methodology for obtaining the various control regions. In general, a few key discriminating variables are used to define control regions with most of the selections similar to what we use in our main search. This ensures confidence that our selections do not bias our ability to measure correct efficiencies. In cases where this is not immediately clear, cross-checks are made and discussed in detail.

8.1 Estimation of $t\bar{t}$ Events in the Signal Region

The fact that b-jets are contained in 99% of $t\bar{t}$ events can be exploited to create a $t\bar{t}$ enhanced region. To obtain such a control sample one can require the presence of one or more tagged b-jets. In the case of this analysis, to reduce a contribution from W + jets events the presence of two jets tagged as b-jets is required when selecting the $t\bar{t}$ control region. Modifications to the final signal selections outlined in section 7 can be made to enhance the statistics of the $t\bar{t}$ enhanced region, allowing for a more precise measurement of the expected $t\bar{t}$ contribution in the SUSY signal region.

Therefore, the $t\bar{t}$ enhanced region is defined by making the following modifications to the final signal selection criteria:

- Remove the isolation requirement on τ's
- Require ≥ 2 jets tagged as a b-jets using the track counting high efficiency "medium" discriminator (TCHEM)

Figures 8.1– 8.8 show various relevant distributions obtained using events in the $t\bar{t}$ control region. The figures show a high purity of $t\bar{t}$ events in the control region. A good agreement between collision data and MC simulation is observed as well, even though this is not a requirement since the techniques employed to extract the background contributions are based on collision data. Additionally, agreement is not expected due to the difficulties related to the correct modeling of jet fragmentation, pile-up, and material budget. However, the fact that agreement is indeed observed points to the robustness of the analysis and signal selections.

Table 8.1 shows the number of observed events in collision data as well as the expected number of MC simulation events in the $t\bar{t}$ control region. The statistical uncertainty on the predicted rate in MC simulation is due to the statistics of the simulated samples, while the systematic uncertainty comes from the uncertainty on the $t\bar{t}$ cross-section. According to MC simulation, the fraction of the events in the $t\bar{t}$ control sample which contain one real τ_h and one jet mistagged as the second τ_h is $A_{\tau+j} = 0.166 \pm 0.011(stat) \pm 0.005(syst)$, while the fraction of the events in the $t\bar{t}$ control sample where two jets are mistagged as the two tau legs is $A_{j+j} = 0.834 \pm 0.025(stat) \pm 0.005(syst)$. Therefore, the number of $t\bar{t}$ events in the signal region can be calculated as follows:

$$\begin{aligned} N_{t\bar{t}}^{Signal} &= A_{\tau+j} \frac{N_{t\bar{t}}^{CR}}{P(2\ b\text{-jets})} \varepsilon^{\tau \text{ iso}} \sum_{N=1}^{\infty} P(N) \sum_{n=1}^{N} C(N,n) f^n (1-f)^{N-n} \\ &+ A_{j+j} \frac{N_{t\bar{t}}^{CR}}{P(2\ b\text{-jets})} \sum_{M=2}^{\infty} P(M) \sum_{m=2}^{M} C(M,m) f^m (1-f)^{M-m} \end{aligned} \quad (8.1)$$

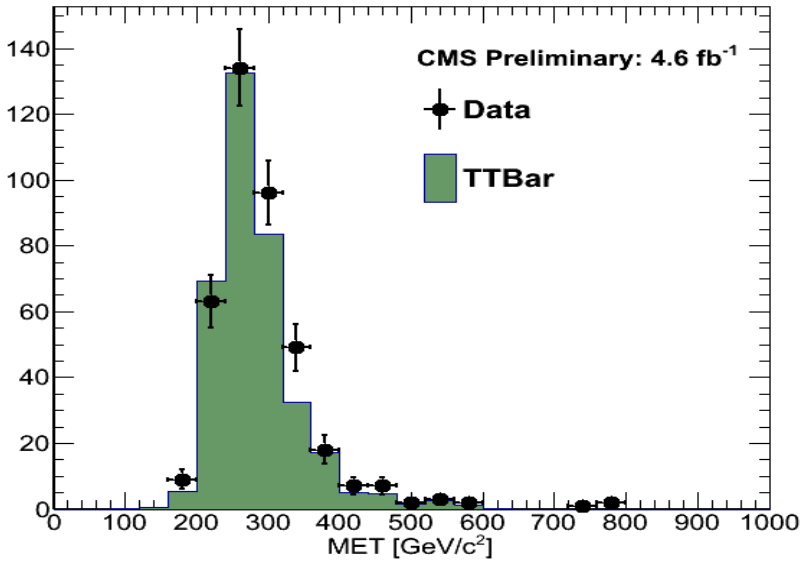

Fig. 8.1. $\not\!\!E_T$.

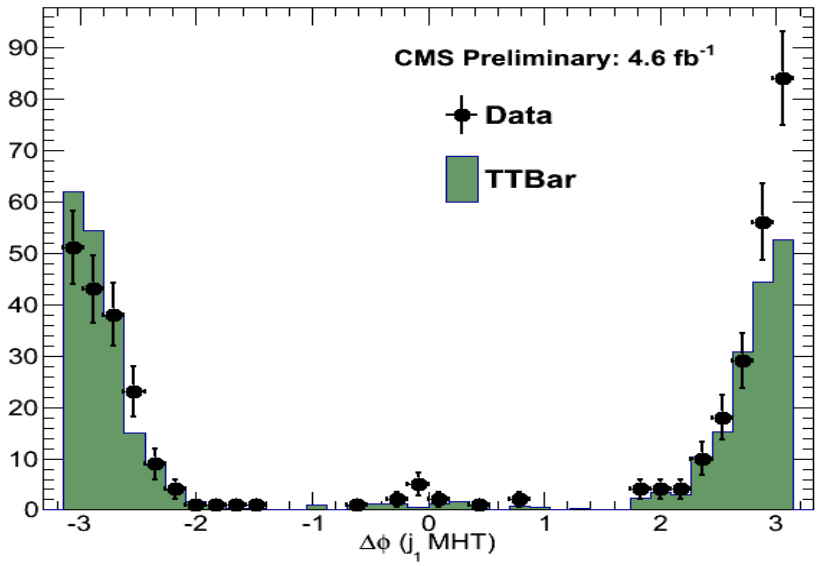

Fig. 8.2. $\Delta\phi(j_1, \not\!\!H_T)$.

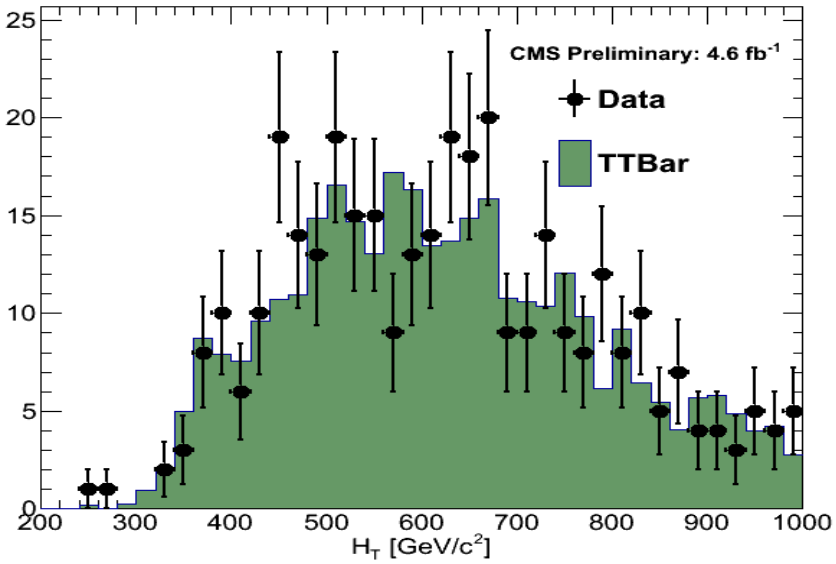

Fig. 8.3. H_T.

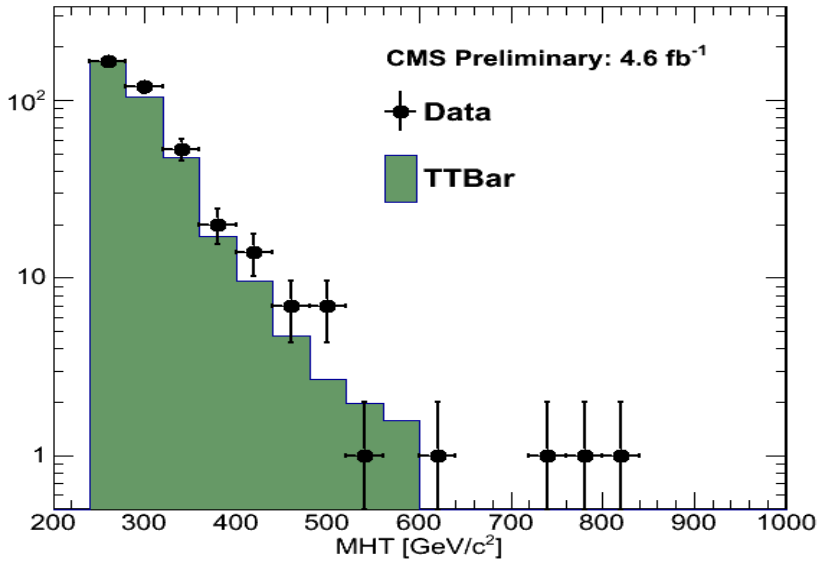

Fig. 8.4. $\not{H}_T$.

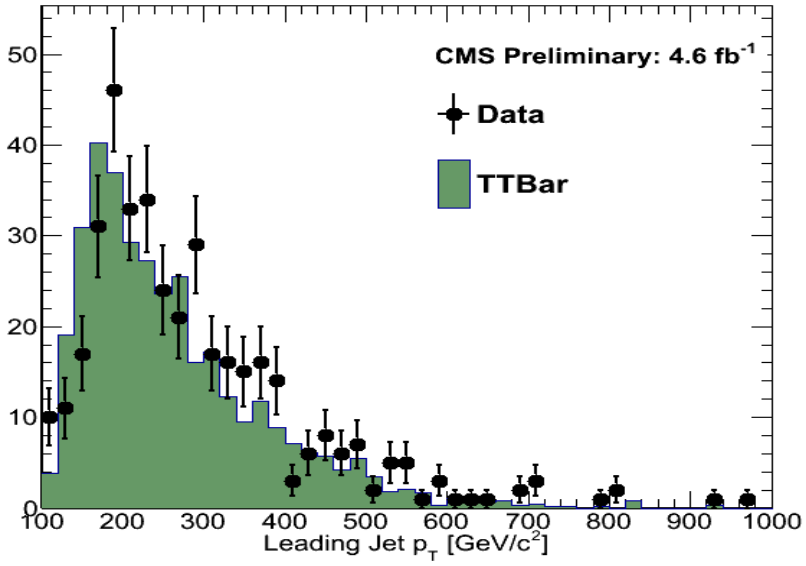

Fig. 8.5. 1^{st} Leading Jet p_T.

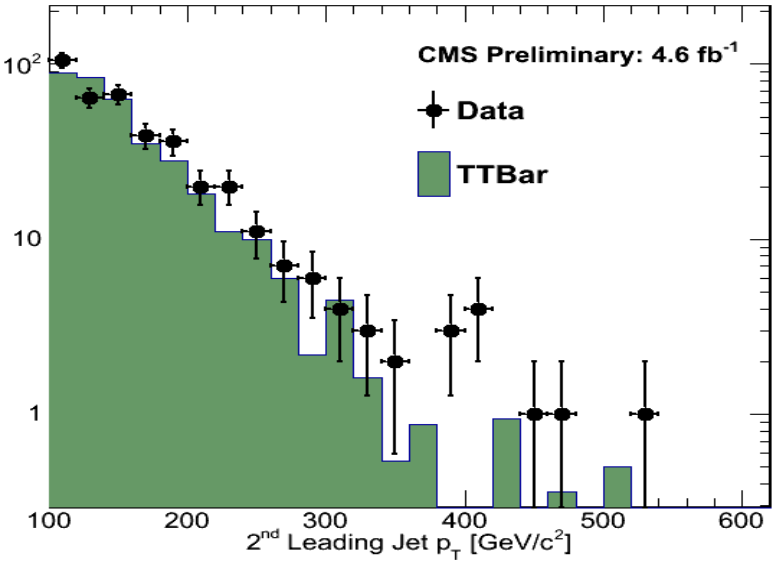

Fig. 8.6. 2^{nd} Leading Jet p_T.

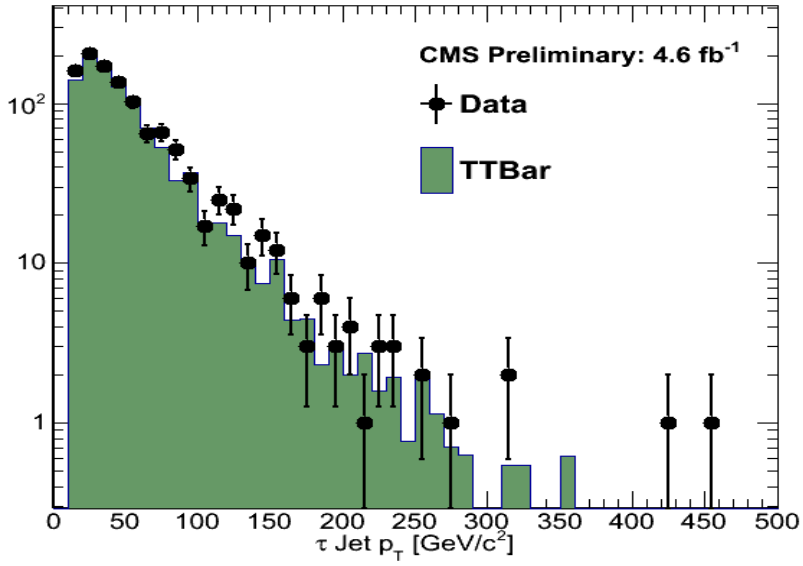

Fig. 8.7. τ p_T.

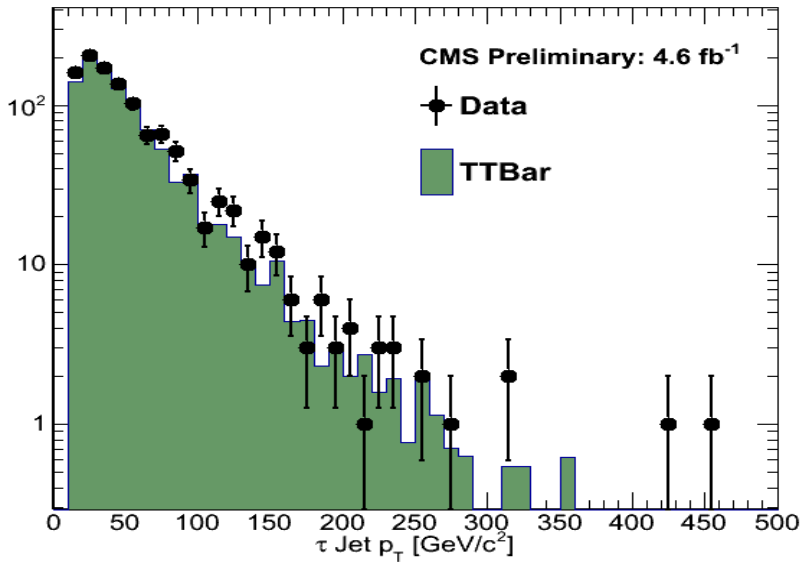

Fig. 8.8. τ η.

where $C(N,n)$ is the combinatoric of (N,n) (i.e.the number of possible ways for n jets to pass isolation given N possible jets), f is the "fake rate" for jets to get tagged as τ_h (probability for a jet to pass the "very loose" HPS isolation requirement), $\varepsilon^{\tau\ \mathrm{iso}}$ is the probability for a τ_h to pass the "very loose" HPS isolation requirement, and $P(N)$ ($P(M)$) are the probabilities for an event to have N (M) jets ("fake" τ_h candidates) considered as possible τ_h candidates. The "fake rate" f and jet multiplicities $P(N)$ are measured/extracted directly from the $\mathrm{t\bar{t}}$ control sample.

The fake rate f has the following definition:

$$f = \frac{\text{Jets Passing Decay Mode Finding, Lepton Vetoes, and VLoose Isolation}}{\text{Jets Passing Decay Mode Finding and Lepton Vetoes}} \quad (8.2)$$

The probability to tag both b-jets, $P(2\ b\text{-jets})$, is given by $\varepsilon^2_{b\text{-tag}}$ where $\varepsilon_{b\text{-tag}}$ is the b-tagging efficiency as measured in [14]. Figures 8.9–8.10 show the isolation distributions obtained from the $\mathrm{t\bar{t}}$ control sample which are used to measure the "fake rate" f. One can see from Fig. 8.9 that the probability for a jet to be mistagged as a tau in collision data is different from the fake rate as determined by MC simulation. However, it is important to note that the jet$\rightarrow \tau$ fake rate is not expected to be well modeled by MC simulation due to the difficulty of modeling rare fluctuations in jet fragmentation and other effects such as pile-up. For this reason we measure the probability for a jet to be misidentified as a tau directly from the control sample and use it to estimate the predicted rate in the signal region. Therefore, agreement between MC and data is not required.

Table 8.2 lists the relevant values used to estimate the $\mathrm{t\bar{t}}$ contribution in the signal region. The isolation efficiency has a measured value of $\varepsilon^{\tau\ \mathrm{iso}} = 0.55 \pm 0.006(stat) \pm 0.04(syst)$ [15]. The probability for a jet to be mistagged as a tau has a measured value of $f = 0.022 \pm 0.003(stat) \pm 0.002(syst)$. The probability to tag ≥ 2 b-jets using the track counting high efficiency "medium" working point is 0.468 ± 0.02 and is determined using the b-tagging efficiency (0.684 ± 0.021) as measured in [14], ε^2_b.

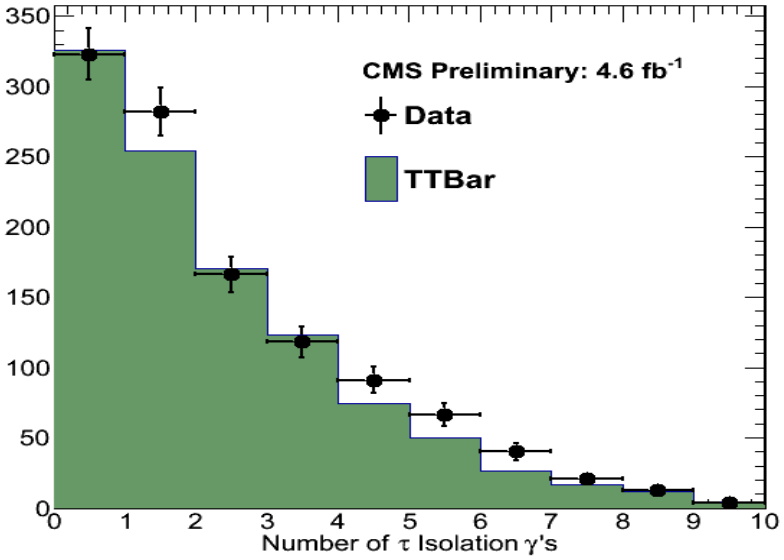

Fig. 8.9. Number of photons.

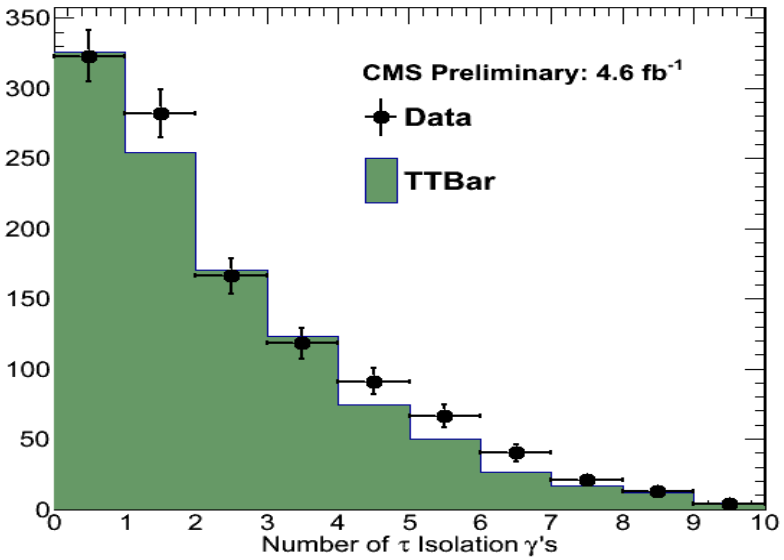

Fig. 8.10. Number of tracks in the τ isolation cone.

Table 8.1
Events in the $t\bar{t}$ control region for data and MC.

Sample	Events
Data	394
QCD	—
$W + Jets$	—
$t\bar{t}$	$357.5 \pm (55.4)_{syst} \pm (9.8)_{stat}$
$Z \to \nu\nu$ + Jets	—

Therefore, the estimated $t\bar{t}$ contribution in the signal region is $N_{t\bar{t}}^{Signal} = 2.03 \pm 0.46$. The uncertainty in the predicted rate is determined by propagating the uncertainties in the parameters listed in Table 8.2 and assuming uncorrelated uncertainties. A more detailed discussion of the systematic uncertainties used in the estimation of $t\bar{t}$ is presented in Section 10.1.1.

Table 8.2
$t\bar{t}$ extraction efficiencies.

Cut	Data
$A_{\tau+j}$	$0.166 \pm 0.011(stat) \pm 0.005(syst)$
A_{j+j}	$0.834 \pm 0.025(stat) \pm 0.005(syst)$
ε^{τ} iso	$0.55 \pm 0.006(stat) \pm 0.04(syst)$
f	$0.022 \pm 0.003(stat) \pm 0.002(syst)$
$\varepsilon^{b-Tagging}$(TCHEM) [14]	$0.684 \pm 0.021(stat + syst)$
Probability to tag ≥ 2 b-jets (TCHEM)	$0.468 \pm 0.02(stat + syst)$
Expected Number of Events	2.03 ± 0.46

Since measuring the $t\bar{t}$ contribution to the signal region makes use of b-tagging to obtain the control sample a natural concern is whether the use of b-tagging will produce a bias on the measured efficiencies. To answer this question cross-checks were carried out to determine if any possible bias was introduced in the b-tagging requirement. It was found that no bias was introduced due to the requirement of at least 2 b-tagged jets. The details of this study are presented in Appendix A. A second

possible concern is whether the use of the b-tagging efficiency as measured in [14] is applicable to the present analysis and selections. Thus, the use of the b-tagging efficiency has been validated by selecting a sample of events with a high multiplicity of jets, this allows a semi-clean $t\bar{t}$ control sample to be obtained in which the given efficiency can be validated. Further details on this study are presented in Appendix B, in which is shown that the use of the b-tagging efficiency as measured in [14] is indeed applicable to the present analysis.

8.1.1 Upper Limit on the Real $\tau_h\tau_h$ $t\bar{t}$ Estimate

In the discussions above and in equation 10.1, the contribution from $t\bar{t}$ events with two real hadronically decaying taus was ignored. A more accurate equation for the estimation of $t\bar{t}$ should contain a term which corresponds to the real $\tau_h\tau_h$ $t\bar{t}$ contribution:

$$A_{\tau+\tau}\frac{N_{t\bar{t}}^{CR}}{P(2\ b\text{-jets})}\varepsilon^{\tau}\ \text{iso}_{\varepsilon^{\tau}\ \text{iso}} \tag{8.3}$$

The reason this term was ignored is because it gives a negligible contribution to the signal region since $A_{\tau+\tau} = 0$ according to simulation. From simulation it was established that 0 out of the 1352 MC events are real $\tau_h\tau_h$ events. Although a negligible contribution it is important to set an upper limit on the real $\tau_h\tau_h$ contribution. To do this, an uncertainty of 1 to the 0 out of 1352 events is assigned to the above equation. Then, the upper limit on the real $\tau_h\tau_h$ contribution to the signal region is given by $N_{\tau_h\tau_h\ t\bar{t}}^{Signal} = \frac{1}{1352}\cdot\frac{394}{0.468}\cdot 0.55^2 = 0.19$. This is then assigned as an additional systematic uncertainty to the $t\bar{t}$ estimation of 2.03 ± 0.46. Therefore, the final estimate of the $t\bar{t}$ contribution in the signal region is $N_{t\bar{t}}^{Signal} = 2.03 \pm 0.46 \pm 0.19 = 2.03 \pm 0.50$.

8.2 Estimation of Z ($\rightarrow \nu\nu$) + Jets Events in the Signal Region

$Z(\rightarrow \nu\nu)$+ Jets events become a background when jets in the events fake tau leptons. Since the probability for $Z(\rightarrow \nu\nu)$+ Jets events to contain four jets, two of which pass all tau tagging criteria, is very small, these type of events can only have a very small contribution to the signal region. However, the contribution of $Z(\rightarrow \nu\nu)$+ Jets events in other background enhanced control regions is not negligible. As events with jets faking taus are difficult to model in MC, a data-driven method is used to extract the number of $Z(\rightarrow \nu\nu)$+ Jets events in the signal region.

Since there is no method to obtain a clean sample of $Z(\rightarrow \nu\nu)$+ Jets events in collision data, from which efficiencies can be measured and used to extrapolate to the signal region. Instead, the contribution from this background is determined by applying a selection criteria similar to those used in the final analysis to obtain a sample of $Z(\rightarrow \mu\mu)$ + Jets events with two clean muons. These muons are then treated as neutrinos in order to properly model the large $\not{H}_T$ values associated with $Z(\rightarrow \nu\nu)$ + Jets events. It is easier to obtain a clean sample of $Z(\rightarrow \mu\mu)$ + Jets events due to the much lower probability for a jet to fake a muon as compared to that for a jet to fake a hadronically decaying tau. Once a clean control sample of $Z(\rightarrow \mu\mu)$ + Jets events is obtained, efficiencies for the $Z(\rightarrow \nu\nu)$ + Jets events can be measured. Correcting the number of observed background events for the muon efficiencies and the ratio of the branching fraction of neutrinos to muons, the expected contribution of $Z(\rightarrow \nu\nu)$ + Jets in the signal region is extracted.

To obtain a clean sample of $Z(\rightarrow \mu\mu)$ + Jets, the final signal selections are modified as follows:

- Select events using the $\mu\tau_h$ cross-triggers
- Remove the requirement on $\not{H}_T$ in order to recalculate $\not{H}_T$ by treating the muons as neutrinos.
- Remove the tau identification criteria.

- Require exactly 0 jets tagged as a b-jets using the track counting high efficiency "medium" discriminator (TCHEM).

- Require ≥ 2 global muons with $p_T > 20$ GeV/c, $|\eta| < 2.1$, and passing the μ identification criteria outlined in section 6.4

Since the branching ratio of muons $B(Z \rightarrow \mu\mu)$ is ~ 6 times smaller than the branching ratio to neutrinos $B(Z \rightarrow \nu\nu)$, then the tau tagging criteria is removed to obtain a statistically significant sample. The requirement of zero b-jets is used in this case in order to minimize the contribution from $t\bar{t}$ events. Also, to properly model the topology of $Z(\rightarrow \nu\nu)$ + Jets events, jets considered for the 1^{st} and 2^{nd} leading jet requirements must not be identified as muons. Finally, because the $\not{H}_T$ must be recalculated by treating the muons as neutrinos, events are selected using the $\mu\tau_h$ cross-triggers. Figures 8.11–8.14 show the distributions for relevant variables obtained with the above selections. A very pure sample of $Z(\rightarrow \mu\mu)$ + Jets events is obtained ($\sim 99\%$ purity) with a very good agreement in both shapes and event rates. Table 8.3 shows the number of observed events in data as well as the expected number of MC simulated events in the $Z(\rightarrow \mu\mu)$ + Jets control region. The $Z(\rightarrow \nu\nu)$ + Jets contribution in the signal region can be calculated as follows:

$$N^{Signal}_{Z(\rightarrow\nu\nu)+Jets} = \frac{N^{CR}_{Z(\rightarrow\mu\mu)+Jets}}{A^2_\mu \varepsilon^2_\mu} \frac{B(Z \rightarrow \nu\nu)}{B(Z \rightarrow \mu\mu)} \frac{\varepsilon^{Trigger}_{\not{H}_T}}{\varepsilon^{Trigger}_{\mu\tau}} \varepsilon^{\not{H}_T} \sum_{N=2}^{\infty} P(N) \sum_{n=2}^{N} C(N,n) f^n (1-f)^{N-n} \tag{8.4}$$

where A_μ is the muon acceptance efficiency, ε_μ the muon identification efficiency, $B(Z \rightarrow \nu\nu)$ the branching ratio for $Z \rightarrow \nu\nu$, $B(Z \rightarrow \mu\mu)$ the branching ratio for $Z \rightarrow \mu\mu$, $\varepsilon^{Trigger}_{\not{H}_T}$ the HLT_PFMHT150 trigger efficiency at the plateau, $\varepsilon^{Trigger}_{\mu\tau}$ the $\mu\tau$ cross-trigger efficiency, $\varepsilon^{\not{H}_T}$ the efficiency for $\not{H}_T > 250$ GeV, $C(N,n)$ the combinatoric of (N,n) (i.e.the number of possible ways to n jets pass the tau selections given M possible jets), f the "fake rate" for jets to get tagged as taus, and $P(N)$ ($P(M)$) the probabilities for an event to have $N(M)$ jets ("fake" tau candidates)

Table 8.3
Events in the $Z(\rightarrow \mu\mu)$ + Jets control region for data and MC.

Sample	Events
Data	738
QCD	—
$W + Jets$	—
$t\bar{t}$	$6.81 \pm 1.20_{stat} \pm 1.06_{syst}$
$Z(\rightarrow \mu\mu)$ + Jets	$709.5 \pm 15.43_{stat} \pm 38.61_{syst}$

considered as possible tau candidates. The acceptance efficiency A_μ is taken from the simulated samples and has a value of $70.07 \pm 0.4 \pm 2.9\%$. The muon identification efficiency ε_μ is measured using standard tag-and-probe methods and has a value of 86.78 ± 0.14 % per muon [16]. The "fake rate" f, jet multiplicities $P(N)$, and $\varepsilon^{\not{H}_T}$ are measured/extracted directly from the $Z(\rightarrow \mu\mu)$ + Jets control sample. Table 8.4 lists the relevant values used to estimate the $Z(\rightarrow \nu\nu)$ + Jets contribution in the signal region. Figures 8.17–8.18 show the isolation distributions obtained from the $Z(\rightarrow \mu\mu)$ + Jets control sample and which are used to measure the "fake rate" f. The probability for a jet to be mistagged as a tau has a measured value of $f = 0.0164 \pm 0.00193(stat) \pm 0.001(syst)$. Therefore, the estimated $Z(\rightarrow \nu\nu)$ + Jets contribution in the signal region is $N^{Signal}_{Z(\rightarrow\nu\nu)+Jets} = 0.03 \pm 0.03$.

8.2.1 Upper Limit on the $Z(\rightarrow \nu\nu)$ + Jets Estimate

The predicted rate for $Z(\rightarrow \nu\nu)$ + Jets is negligible in simulation (0.0 ± 0.07). Additionally, this has been further validated by employing a data driven estimation of $Z(\rightarrow \nu\nu)$ + Jets where the probability for a jet to be mistagged as a τ_h is measured. The main reason this background is expected to increase in collision data compared to the estimate from MC simulation is because of the high probability that the MC simulation does not accurately model the $j \rightarrow \tau_h$ mistag rate. Since the mistag rates are measured directly from data and since the mistag rates measured in the $t\bar{t}$ and

Table 8.4
$Z(\to \nu\nu)$ + Jets extraction efficiencies

Cut	Data
$N^{pure}_{Z(\to\mu\mu)+\text{Jets}}$	738
$B(Z \to \nu\nu)$	0.20 ± 0.0006
$B(Z \to \mu\mu)$	0.03366 ± 0.00007
$\varepsilon^{Trigger}_{\mu\tau}$	$0.87 \pm 0.04(stat + syst)$
$\varepsilon^{Trigger}_{\not{H}_T}$	$0.989 \pm 0.025(stat + syst)$
$\varepsilon^{\not{H}_T}$	$0.0081 \pm 0.0033(stat) \pm 0.004(syst)$
A_μ	$0.7007 \pm 0.004(stat) \pm 0.029(syst)$
ε_μ	$0.8678 \pm 0.0014(stat + syst)$
f	$0.0164 \pm 0.00193(stat) \pm 0.001(syst)$
Expected Number of Events	0.03 ± 0.03

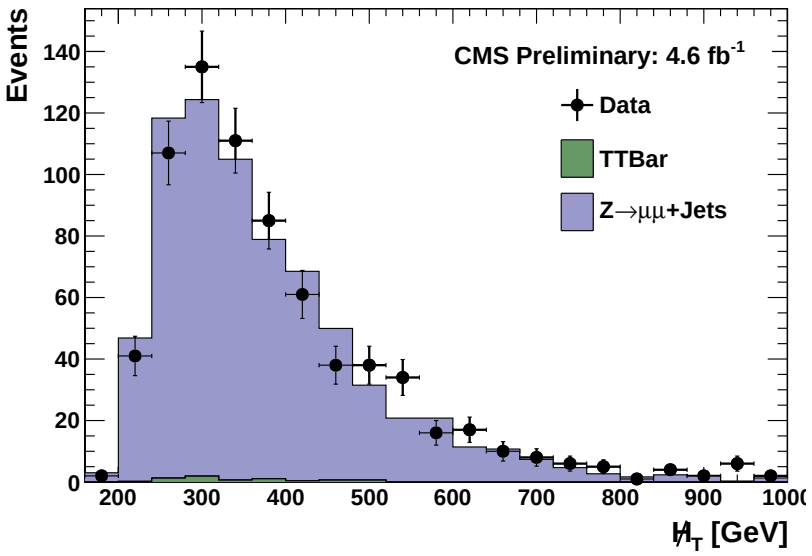

Fig. 8.11. H_T distribution in the $Z(\to \mu\mu)$+ Jets control region.

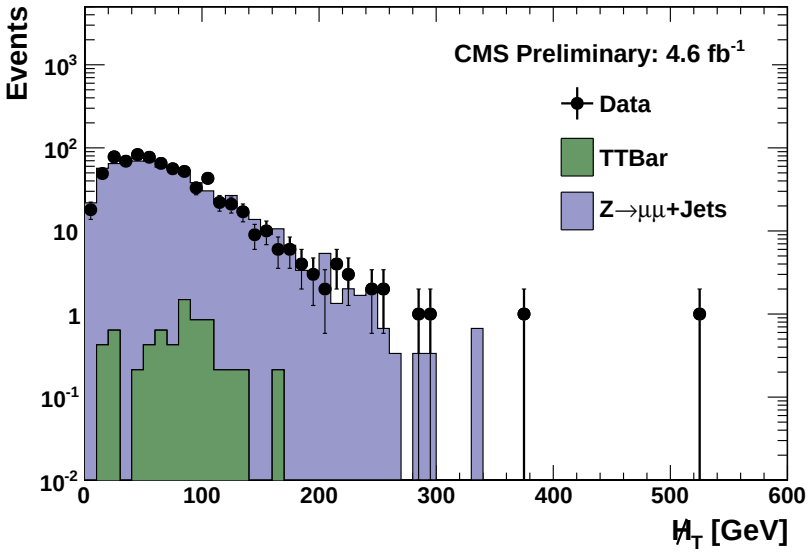

Fig. 8.12. $\not\!\!H_{\mathrm{T}}$ distribution in the $Z(\rightarrow \mu\mu)$+ Jets control region.

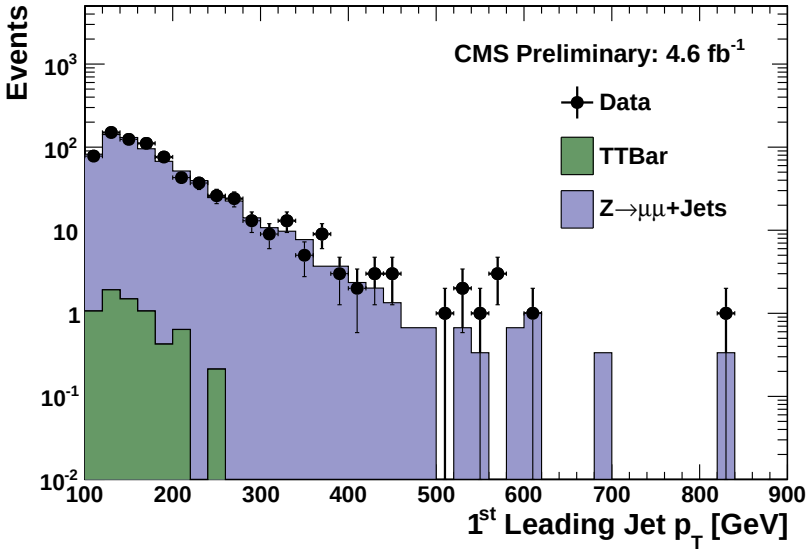

Fig. 8.13. 1^{st} Leading Jet p_T distribution in the $Z(\rightarrow \mu\mu)$+ Jets control region.

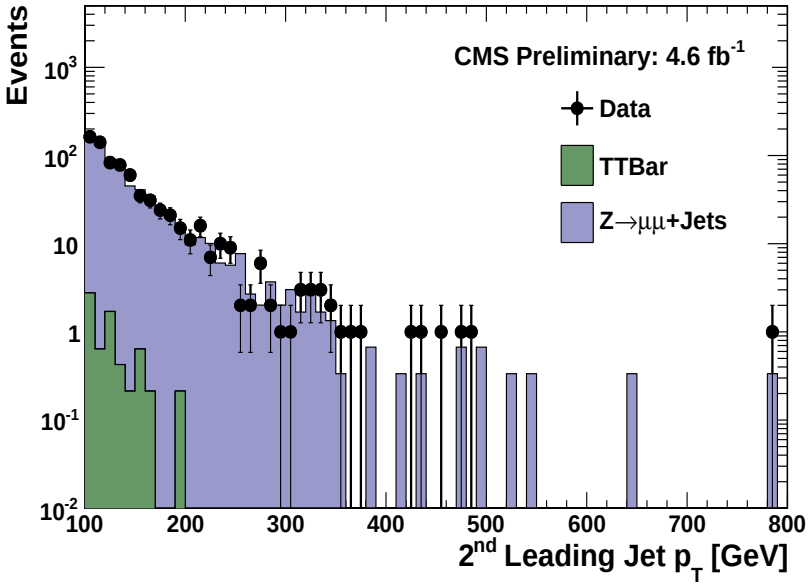

Fig. 8.14. 2^{nd} Leading Jet p_T distribution in the $Z(\rightarrow \mu\mu)$+ Jets control region.

W + Jets control regions agree with those measured in the control sample of Z + Jets, this is a strong validation that the contribution is indeed small. However, to provide with additional confidence in the estimate, in this section an upper limit on the $Z(\rightarrow \nu\nu)$ + Jets contribution in the signal region is determined. The equation used to estimate the $Z(\rightarrow \nu\nu)$ + Jets contribution is listed below:

$$N^{Signal}_{Z(\rightarrow\nu\nu)+\text{Jets}} = \frac{N^{CR}_{Z(\rightarrow\mu\mu)+\text{Jets}}}{A^2_\mu \varepsilon^2_\mu} \frac{B(Z\rightarrow\nu\nu)}{B(Z\rightarrow\mu\mu)} \frac{\varepsilon^{Trigger}_{\not{H}_T}}{\varepsilon^{Trigger}_{\mu\tau}} \varepsilon^{\not{H}_T} \sum_{N=2}^{\infty} P(N) \sum_{n=2}^{N} C(N,n) f^n (1-f)^{N-n} \tag{8.5}$$

The probability for a $Z \rightarrow \nu\nu$ event to have $\not{H}_T > 250$ GeV, $\varepsilon^{\not{H}_T}$, is small ($<<$ 1%). Additionally, $\varepsilon^{Trigger}_{\not{H}_T}$ is approximately equal to $\varepsilon^{Trigger}_{\mu\tau}$. Therefore, an upper limit on the estimation of $Z(\rightarrow \nu\nu)$ + Jets can be determined by only considering the

branching ratios for Z (which are known to good accuracy) and the $j \rightarrow \tau_h$ mistag rate, f:

$$N^{Signal}_{Z(\rightarrow\nu\nu)+\text{Jets}} << N^{CR}_{Z(\rightarrow\mu\mu)+\text{Jets}} \frac{B(Z \rightarrow \nu\nu)}{B(Z \rightarrow \mu\mu)} f^2 \tag{8.6}$$

Using the above equation, an upper limit for $Z \rightarrow \nu\nu$ of $N^{Signal}_{Z(\rightarrow\nu\nu)+Jets} << 738 \cdot 6 \cdot 0.01^2 \sim 0.5$ is obtained. This is an extremely conservative upper limit on the $Z(\rightarrow \nu\nu)$ + Jets contribution in the signal region since the fact that $\varepsilon^{\not{H}_T}$ is a very small quantity ($< 1\%$ in both MC simulation and collision data) has been ignored. This upper limit is assigned on the $Z(\rightarrow \nu\nu)$ + Jets prediction as a systematic uncertainty on the predicted rate.

8.3 Estimation of $Z(\rightarrow \tau\tau)$ + Jets Events in the Signal Region

$Z(\rightarrow \tau\tau)$+ Jets becomes a background mostly when: (1) both taus from the decay of the Z boson pass the tau tagging criteria, provide large $\not{H}_T$ from their decays to neutrinos, and two additional jets pass the jet criteria; (2) one tau from the decay of the Z boson passes the tau tagging criteria and an additional jet is mistagged as a tau.

Like in the case of $Z(\rightarrow \nu\nu)$ + Jets, there is no method to obtain a clean sample of $Z(\rightarrow \tau\tau)$+ Jets where efficiencies can be measured and used to extrapolate to the signal region. Therefore, the contribution from this background is determined using the same sample of $Z(\rightarrow \mu\mu)$ + Jets events used to estimate $Z(\rightarrow \nu\nu)$+ Jets. However, in this case the muons are treated taus in order to properly model the large $\not{H}_T$ values associated with $Z(\rightarrow \tau\tau)$ + Jets events. Figures 8.15–8.16 shows the $\not{H}_T$ and H_T distributions obtained using events from the $Z(\rightarrow \mu\mu)$ + Jets control sample and applying this method of treating the muons as taus. Efficiencies for $Z(\rightarrow \tau\tau)$ + Jets events can be measured from this sample and the number of observed events can be corrected for the muon efficiencies and branching ratios to muons in order

to determine the expected contribution of $Z(\rightarrow \tau\tau)$ + Jets in the signal region. Therefore, $Z(\rightarrow \tau\tau)$ + Jets contribution in the signal region can be calculated as follows:

$$
\begin{aligned}
N_{Z\rightarrow\tau\tau}^{Signal} &= N_{Z\rightarrow\mu\mu}^{CR} \frac{B(Z\rightarrow\tau\tau)B(\tau\rightarrow\tau_h)}{B(Z\rightarrow\mu\mu)} \frac{\varepsilon_{\not{H}_T}^{Trig}}{\varepsilon_{\mu\tau}^{Trig}} \varepsilon^{\not{H}_T} [\frac{A_\tau^2\varepsilon_\tau^2}{A_\mu^2\varepsilon_\mu^2}] \\
&+ N_{Z\rightarrow\mu\mu}^{CR} \frac{B(Z\rightarrow\tau\tau)B(\tau\rightarrow\tau_h)}{B(Z\rightarrow\mu\mu)} \frac{\varepsilon_{\not{H}_T}^{Trig}}{\varepsilon_{\mu\tau}^{Trig}} \varepsilon^{\not{H}_T} [\frac{A_\tau^2(2\varepsilon_\tau(1-\varepsilon_\tau))}{A_\mu^2\varepsilon_\mu^2} \\
&\times \sum_{N=1}^{\infty} P(N) \sum_{n=1}^{N} C(N,n) f^n (1-f)^{N-n}] \\
&+ N_{Z\rightarrow\mu\mu}^{CR} \frac{B(Z\rightarrow\tau\tau)B(\tau\rightarrow\tau_h)}{B(Z\rightarrow\mu\mu)} \frac{\varepsilon_{\not{H}_T}^{Trig}}{\varepsilon_{\mu\tau}^{Trig}} \varepsilon^{\not{H}_T} [\frac{2A_\tau(1-A_\tau)\varepsilon_\tau}{A_\mu^2\varepsilon_\mu^2} \\
&\times \sum_{M=1}^{\infty} P(M) \sum_{m=1}^{M} C(M,m) f^m (1-f)^{M-m}] \\
&+ N_{Z\rightarrow\mu\mu}^{CR} \frac{B(Z\rightarrow\tau\tau)B(\tau\rightarrow\tau_h)}{B(Z\rightarrow\mu\mu)} \frac{\varepsilon_{\not{H}_T}^{Trig}}{\varepsilon_{\mu\tau}^{Trig}} \varepsilon^{\not{H}_T} [\frac{(1-A_\tau)^2}{A_\mu^2\varepsilon_\mu^2} \\
&\times \sum_{K=1}^{\infty} P(K) \sum_{k=1}^{K} C(K,k) f^k (1-f)^{K-k}] \qquad (8.7)
\end{aligned}
$$

where A_μ is the muon acceptance efficiency, ε_μ the muon identification efficiency, A_τ the tau acceptance efficiency, ε_τ the tau identification efficiency, $B(Z \rightarrow \tau\tau)$ the branching ratio for $Z \rightarrow \tau\tau$, $B(\tau \rightarrow \tau_h)$ the branching ratio for tau leptons to hadrons, $B(Z \rightarrow \mu\mu)$ the branching ratio for $Z \rightarrow \mu\mu$, $\varepsilon_{\not{H}_T}^{Trigger}$ the HLT_PFMHT150 trigger efficiency at the plateau, $\varepsilon_{\mu\tau}^{Trigger}$ the $\mu\tau$ cross-trigger efficiency, $\varepsilon^{\not{H}_T}$ the efficiency for $\not{H}_T > 250$ GeV, $C(N,n)$ the combinatoric of (N,n) (i.e.the number of possible ways to n jets pass the tau selections given M possible jets), f the "fake rate" for jets to get tagged as taus, and $P(N)$ ($P(M)$) the probabilities for an event to have $N(M)$ jets ("fake" tau candidates) considered as possible tau candidates. The "fake rate" f and jet multiplicities $P(N)$ are measured/extracted directly from the $Z(\rightarrow \mu\mu)$ + Jets control sample. Table 8.5 lists the relevant values used to

Table 8.5
$Z \rightarrow \tau\tau$ + Jets extraction efficiencies

Cut	Data
$N^{pure}_{Z(\rightarrow\mu\mu)+\text{Jets}}$	738
$B(Z \rightarrow \tau\tau)$	0.03367 ± 0.0008
$B(\tau \rightarrow \tau_h)$	0.6479 ± 0.0008
$B(Z \rightarrow \mu\mu)$	0.03366 ± 0.00007
$\varepsilon^{Trigger}_{\mu\tau}$	$0.87 \pm 0.04(stat + syst)$
$\varepsilon^{Trigger}_{\not{H}_T}$	$0.989 \pm 0.025(stat + syst)$
$\varepsilon^{\not{H}_T}$	$0.00271 \pm 0.00192(stat)$
A_τ	$0.503 \pm 0.001(stat) \pm 0.014(syst)$
ε_τ	$0.649 \pm 0.007(stat) \pm 0.045(syst)$
A_μ	$0.7007 \pm 0.004(stat) \pm 0.029(syst)$
ε_μ	$0.8678 \pm 0.0014(stat + syst)$
f	$0.0164 \pm 0.00193(stat) \pm 0.001(syst)$
Expected Number of Events	0.21 ± 0.20

estimate the $Z(\rightarrow \tau\tau)$ + Jets contribution in the signal region. The tau isolation efficiency has a measured value of $\varepsilon_\tau = 0.55 \pm 0.04$ [15]. Figures 8.17–8.18 show the isolation distributions obtained from the $Z(\rightarrow \mu\mu)$ + Jets control sample and which are used to measure the "fake rate" f. The probability for a jet to be mistagged as a tau has a measured value of $f = 0.0164 \pm 0.00193$. Therefore, the estimated $Z \rightarrow \tau\tau$ + Jets contribution in the signal region is $N^{Signal}_{Z(\rightarrow\tau\tau)+\text{Jets}} = 0.21 \pm 0.19$.

8.4 Estimation of QCD Events in the Signal Region

Though QCD has the largest cross section of all the backgrounds affecting this analysis, it plays only a minor role after tau identification. Nonetheless, an understanding of QCD is important for other background estimations and minimizing our total uncertainty. Additionally, because the QCD contribution in the signal region depends on the probability for a jet to fake a τ_h, it is important to validate and/or

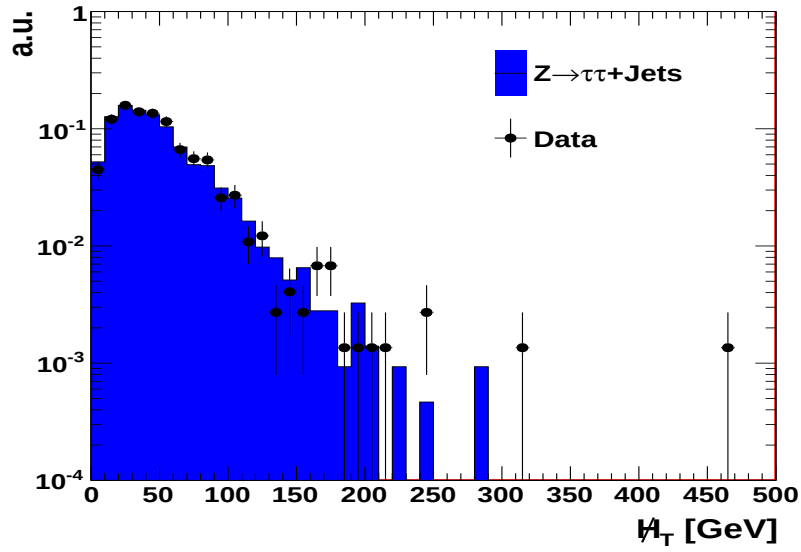

Fig. 8.15. $\not{H}_T$ obtained from events in the $Z(\rightarrow \mu\mu)$ + Jets control sample and treating the muons as taus to model $Z(\rightarrow \tau\tau)$ + Jets events.

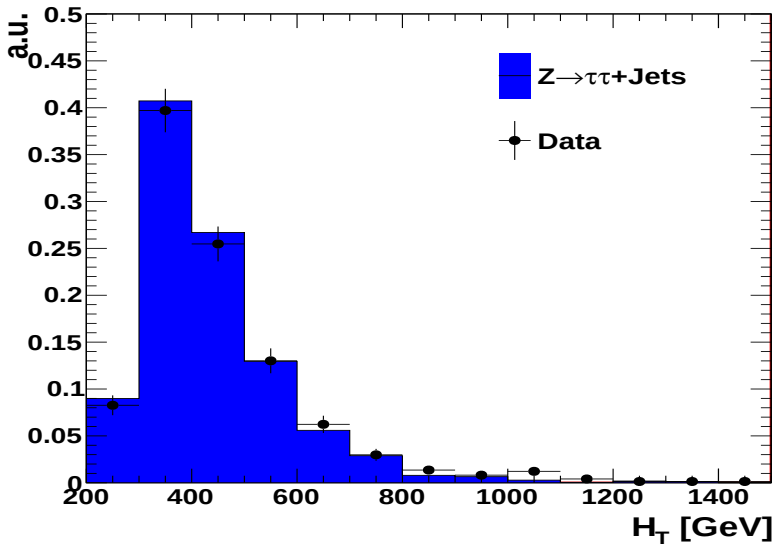

Fig. 8.16. H_T obtained from events in the $Z(\rightarrow \mu\mu)$ + Jets control sample and treating the muons as taus to model $Z(\rightarrow \tau\tau)$ + Jets events.

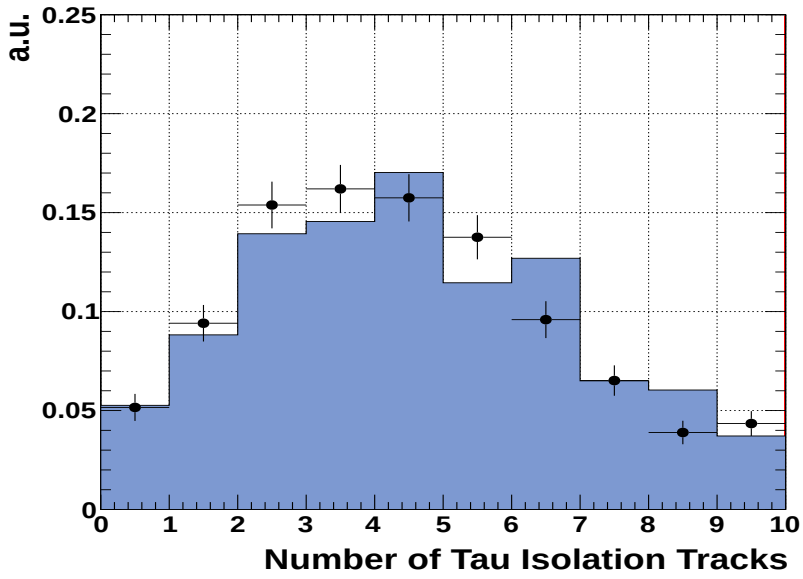

Fig. 8.17. Track isolation distribution for "taus"/jets in the $Z(\rightarrow \mu\mu)$ + Jets control region.

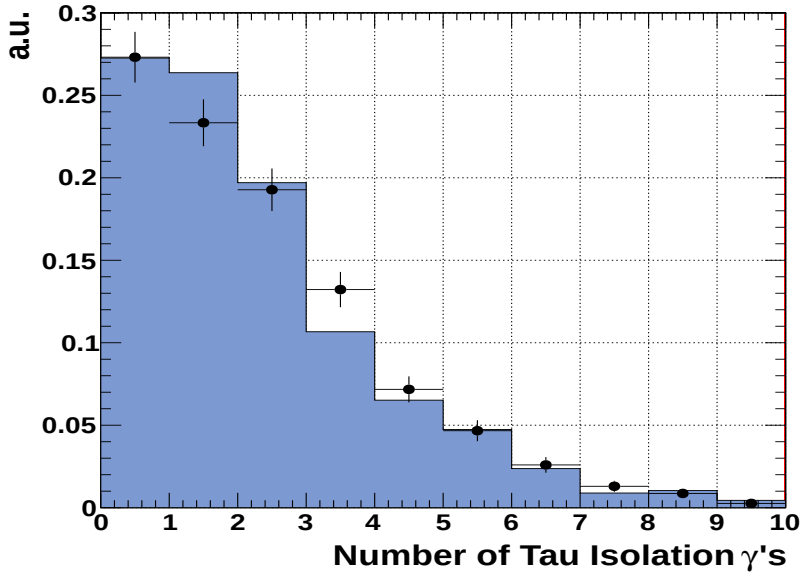

Fig. 8.18. γ isolation distribution for "taus"/jets in the $Z(\rightarrow \mu\mu)$ + Jets control region.

measure that the signal contribution is indeed small. To obtain a good purity QCD control region the following cuts are applied:

- $p_T > 100$ GeV and $|\eta| < 3.0$ for 2 leading jets
- $\not{H}_{\rm T} > 250$ GeV (To eliminate trigger bias)
- ≥ 2 "τ_h's" passing the light lepton vetoes (To minimize W+Jets and $t\bar{t}$ backgrounds)
- $\Delta\phi(\not{H}_{\rm T}, jet_2) < 0.10$

Since $\not{H}_T$ tends to be quite small for QCD events, $\not{H}_T$ is greater than 250 GeV almost exclusively in QCD events in which a jet is mismeasured. As a result, $\not{H}_T$ tends to align with the second leading jet of most QCD events surviving the $\not{H}_T > 250$ GeV cut. After applying the event selections above, a data sample is obtained with 1678 events and $\sim$ 86% purity according to MC simulation. Table 8.6 shows the number of observed events in data and the expectations based on the MC simulated samples. In order to estimate the QCD contribution to the signal region, a data-MC scale factor (SF_{QCD}) can be applied to the QCD expectation in MC simulation. This results in a scale factor of 0.74 ± 0.02. Since the QCD simulation samples are limited by low statistics, the predicted QCD rate in MC is obtained by slightly relaxing the $\not{H}_T$ criteria such that it remains unbiased to the remaining selection efficiencies and provides enough increase in statistics to obtain a MC simulated prediction to the signal region. The estimated QCD contribution to the signal region is then calculated as follows:

$$N_{QCD}^{Signal} = SF_{QCD} \times N_{QCD}^{MC} \tag{8.8}$$

This method gives a QCD estimation of $N_{QCD}^{Signal} = (0.74 \pm 0.02) \times (0.024 \pm 0.024) = 0.018 \pm 0.018$. As expected, the QCD contribution to the signal region is small. The

$\sim$ 86% purity of this QCD dominated control sample is shown in Figures 8.19–8.20, which show the distributions for N_{vtx} and $\not{H}_T$. Since a full data-driven method is not employed and since the control sample is obtained with selections similar to the signal region, the figures provide a degree of confidence in the usage of the data-MC correction factor by showing that the MC accurately describes the bulk of the distributions in collision data. In the following section, an upper limit on the predicted rate is calculated and used to assign an additional systematic uncertainty.

8.4.1 Upper Limit on the QCD Background

The final event selection requires a $\Delta\phi(\not{H}_T, jet_2) > 0.50$. To obtain a control region of high QCD purity, the requirement applied is $\Delta\phi(\not{H}_{\rm T}, jet_2) < 0.10$. A full data-driven QCD estimate would employ the following:

$$N_{QCD}^{Signal} = N_{QCD}^{ControlRegion} \times \frac{\epsilon(\Delta\phi(\not{H}_{\rm T}, jet_2) > 0.50)}{\epsilon(\Delta\phi(\not{H}_{\rm T}, jet_2) < 0.10)} \times {\epsilon_{Fake}^{\tau}}^2 \tag{8.9}$$

where ϵ_{Fake}^{τ} is the probability of a jet faking a tau (approximately 1%). (QCD events in the signal region require two fake taus.) Since $\epsilon(\Delta\phi(\not{H}_T, jet_2) > 0.50$ is approximately 10% and $\epsilon(\Delta\phi(\not{H}_{\rm T}, jet_2) < 0.10)$ is approximately 80%, a very conservative upper limit of the QCD contribution in the signal region can be calculated.

$$N_{QCD}^{Signal} << N_{QCD}^{ControlRegion} \times {\epsilon_{Fake}^{\tau}}^2 \tag{8.10}$$

This bounds the QCD contamination in our signal region to be less than 0.1678 events. Since a full data-driven estimate for QCD is not performed, this upper limit is applied on the QCD prediction as a systematic uncertainty.

Table 8.6
Events in the QCD control region for data and MC.

Sample	Events
Data	1678
QCD	$1847.5 \pm 37.6_{stat} \pm 286.22_{syst}$
$W \rightarrow l\nu$ + Jets	$242.0 \pm 19.7_{stat} \pm 12.04_{syst}$
$t\bar{t}$	$66.2 \pm 3.6_{stat} \pm 10.26_{syst}$
$Z \rightarrow \nu\nu$ + Jets	—

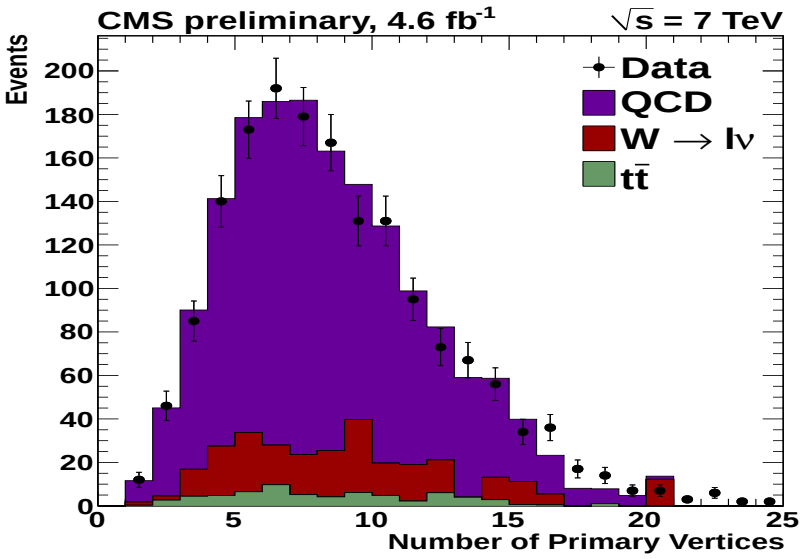

Fig. 8.19. Distribution for number of primary vertices obtained using events from a control sample enriched with $\sim$ 86% purity of QCD events according to MC.

8.5 Estimation of W + Jets Events in the Signal Region

W + Jets events become a background due to recoil jets or jets from initial and final state radiation faking one or both of the tau legs. In order to enhance the contribution of W + Jets events, the tau isolation criteria, which discriminates τ_h's from other jets, is removed. On the other hand, the lack of a tau isolation requirement causes the presence of other backgrounds to increase in the W + Jets control region since most of the backgrounds arise due to jets faking τ_h's. To reduce the $t\bar{t}$ event contribution to the W + Jets control sample, events are required to have zero jets tagged as b-jets via the TCHEM discriminator. This requirement will minimize the contamination from $t\bar{t}$ events to $\sim 5\%$, however the presence of other backgorunds is not negligible. Furthermore, no aditional requirements can be imposed to reduce the remaining backgrounds while allowing an unbiased extraction of efficiencies. In this case a subtraction technique is employed where each background contribution to the W + Jets control region is calculated from MC samples:

$$N^{control} = N_{t\bar{t}}^{\text{W+Jets enhanced}} + N_{QCD}^{\text{W+Jets enhanced}} + N_{Z(\rightarrow\nu\nu)+\text{Jets}}^{\text{W+Jets enhanced}} + N_{\text{W+Jets}}^{\text{W+Jets enhanced}} \tag{8.11}$$

$$N_{\text{W+Jets}}^{\text{W+Jets enhanced}} = N^{control} - N_{t\bar{t}}^{\text{W+Jets enhanced}} - N_{\text{QCD}}^{\text{W+Jets enhanced}} - N_{Z(\rightarrow\nu\nu)+\text{Jets}}^{\text{W+Jets enhanced}} \tag{8.12}$$

where $N_{t\bar{t}}^{\text{W+Jets enhanced}}$ is the $t\bar{t}$ contribution, $N_{\text{QCD}}^{\text{W+Jets enhanced}}$ the QCD contribution, and $N_{Z(\rightarrow\nu\nu)+\text{Jets}}^{\text{W+Jets enhanced}}$ the $Z(\rightarrow \nu\nu)$ + jets contribution in the W + jets enhanced region.

Table 8.7 lists the observed and expected number of MC simulated events in this W + Jets enhanced region while Figs. 8.21–8.24 show the relevant distributions obtained from these events. The number of W + Jets events in collision data would

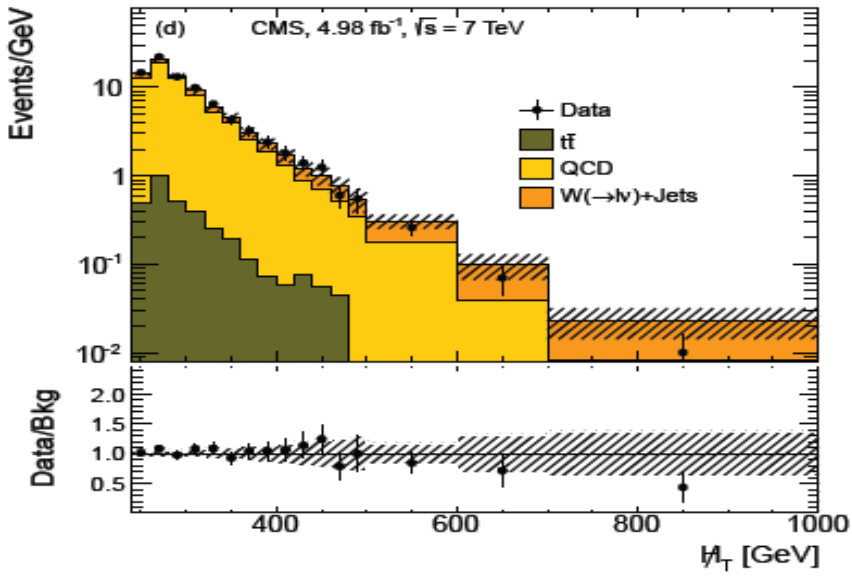

Fig. 8.20. Distribution for $\not\!\!H_T$ obtained using events from a control sample enriched with $\sim 99\%$ purity of QCD according to MC.

Table 8.7
Events in the W + Jets enhanced region for data and MC.

Sample	Events
Data	2874
QCD	$194.3 \pm 32.1_{stat} \pm 30.1_{syst}$
$W + Jets$	$1734.3 \pm 87.8_{stat} \pm 86.3_{syst}$
$t\bar{t}$	$116.2 \pm 7.2_{stat} \pm 18.0_{syst}$
$Z \rightarrow \nu\nu$ + Jets	$549.9 \pm 73.8_{stat} \pm 29.9_{syst}$

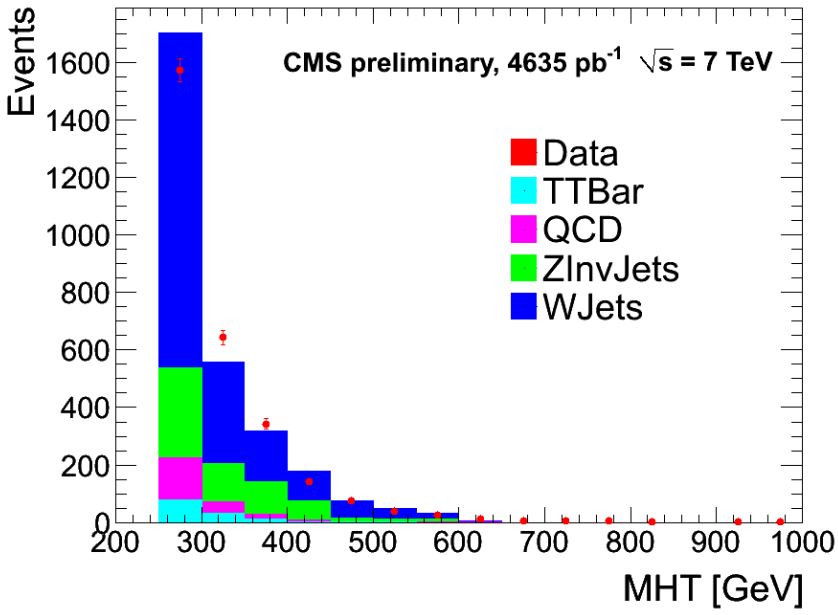

Fig. 8.21. $\not{H}_T$.

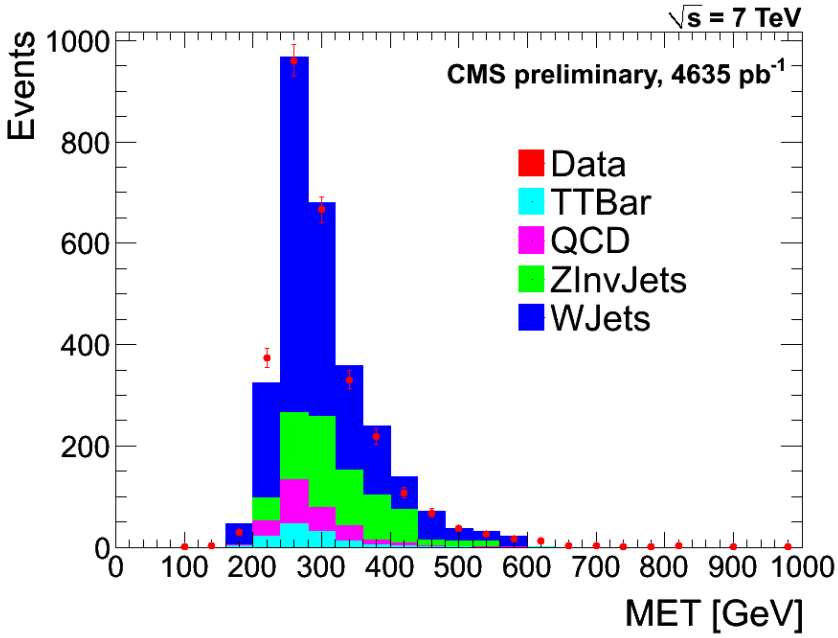

Fig. 8.22. $\not{E}_T$.

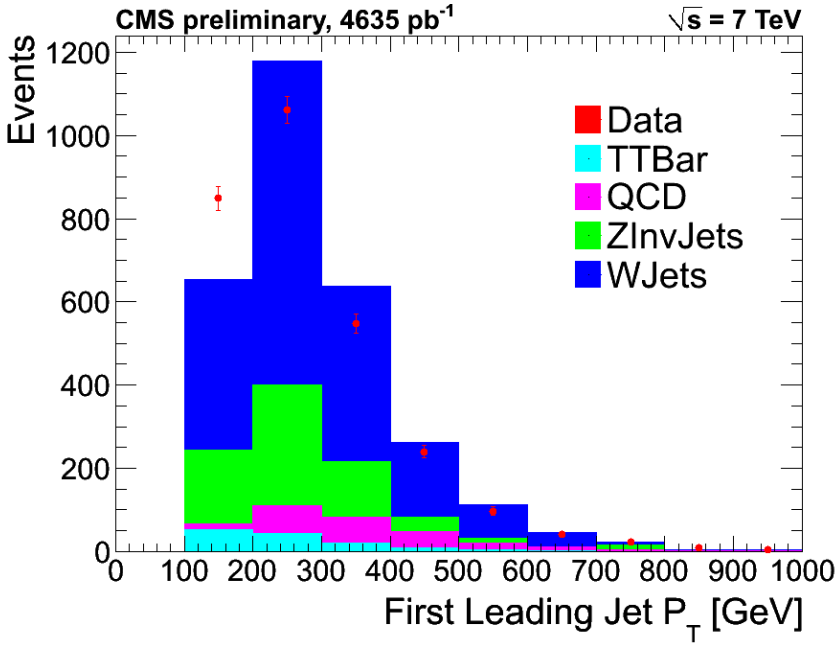

Fig. 8.23. First leading jet p_T.

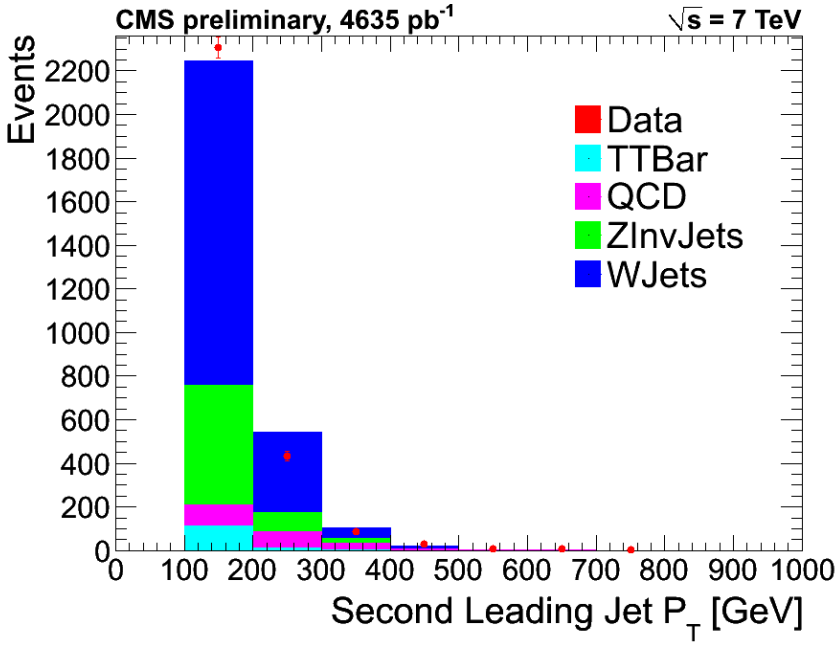

Fig. 8.24. Second leading jet p_T.

then be data − non W + Jets background which is 2013.6 events. Hence, the data-MC scale factor is $2013.6/1734.3 = 1.16 \pm 0.10$. Figures 8.25–8.28 show the resultant distributions after performing the subtraction of BGs in the W + Jets events CR as outlined on Equation 8.12. In order to extrapolate from the W + Jets control region to the signal region we can use the following equation:

$$N_{signal}^{\text{W+Jets}} = \frac{N_{\text{W+Jets}}^{\text{W+Jets enhanced}} \times \varepsilon^{\tau_{iso}}}{P_{b-Tagging}(0)} \tag{8.13}$$

where $N_{\text{W+Jets}}^{\text{W+Jets enhanced}}$ is the number of events in data after the subtraction (equation 8.12), $\varepsilon^{\tau_{iso}}$ is the efficiency to find ≥ 2 jets/taus passing isolation, and $P_{b-Tagging}(0)$ is the probability to get zero jets tagged as b-jets. We find that the measured efficiency $\varepsilon^{\tau_{iso}}$ to find ≥ 2 jets/taus passing isolation is $\varepsilon^{\tau_{iso}} = 6/2874 = 0.21 \pm 0.09\%$. The probability to mistag zero jets as b-jets is calculated by determining the following quantity:

$$P(N,n) = C(N,n) * f_{0b}^{n} * (1 - f_{0b})^{N-n} \tag{8.14}$$

where $P(N,n)$ is the probability to tag n fake b-jets given N jets, $C(N,n)$ the combinatoric of (N,n) (i.e. the number of possible ways to obtain n given N), and f_{0b} the fake rate for non-b jets to get tagged as b's. The value of $N = 4$ chosen for the above calculation represents the mean number of jets in the control region. Therefore, the probability to tag zero jets as b-jets is given by:

$$P(0) = 1 - P(4,1) - P(4,2) - P(4,3) - P(4,4) \tag{8.15}$$

Using the fake rate $f_{0b} = 2.8\%$ as measured in [14], the calculated probability to tag zero jets as b-jets is $83.0 \pm 8.0\%$. The quoted uncertainty is determined by varying N in the above equations by the RMS of the jet multiplicity distribution, $RMS(N) = 1$. Therefore, the total number of expected W+Jets in the signal region is 4.7 ± 1.9. This number is consistent with the prediction in the signal region using

the data-MC scale factor of 1.16 ± 0.10. We find 7.4 ± 3.7 events using the data-MC scale factor.

The W + jet events in the signal region is also calculated using the following equation ("fake rate method"):

$$\begin{aligned} N^{Signal}_{\text{W+Jets}} &= A_{\tau+j}\frac{N^{After\ subtraction}_{\text{W+jets}}}{P(0\ b\text{-jets})}\varepsilon^{\tau\ \text{iso}}\sum_{N=1}^{\infty}P(N)\sum_{n=1}^{N}C(N,n)f^{n}(1-f)^{N-n} \\ &+ A_{j+j}\frac{N^{After\ subtraction}_{\text{W+Jets}}}{P(0\ b\text{-jets})}\sum_{M=2}^{\infty}P(M)\sum_{m=2}^{M}C(M,m)f^{m}(1-f)^{M-m} \quad (8.16) \end{aligned}$$

where, $A_{\tau+j}$ is the fraction of events in the W + jets control sample which contain one real tau and one jet mistagged as the second tau, A_{j+j} is the fraction of events in the W + jets control sample where two jets are mistagged as the two tau legs, $C(N,n)$ is the combinatoric of (N,n) (i.e.the number of possible ways for n jets to pass the isolation condition given M possible jets), f is the "fake rate" for jets to get tagged as taus (probability for a jet to pass the "very loose" HPS isolation requirement), $\varepsilon^{\tau\ \text{iso}}$ is the probability for a tau to pass the "very loose" HPS isolation requirement, and $P(N)$, $(P(M))$ are the probabilities for an event to have N (M) jets ("fake" tau candidates) considered as possible tau candidates. Table 8.8 shows the values obtained for each variable in the above equation. Figures 8.29–8.30 show the isolation distributions obtained from the W + jets events control sample which are used to measure the "fake rate" f. The fake rate f has the following definition:

$$f = \frac{\text{Jets Passing Decay Mode Finding, Lepton Vetoes, and VLoose Isolation}}{\text{Jets Passing Decay Mode Finding and Lepton Vetoes}} \quad (8.17)$$

The number of expected W + Jets events in the signal region is cross checked using the tau isolation fake rate, as was done in section 8.1 for the $t\bar{t}$ estimation in order to reduce the systematic uncertainty. The fake rate is measured to be

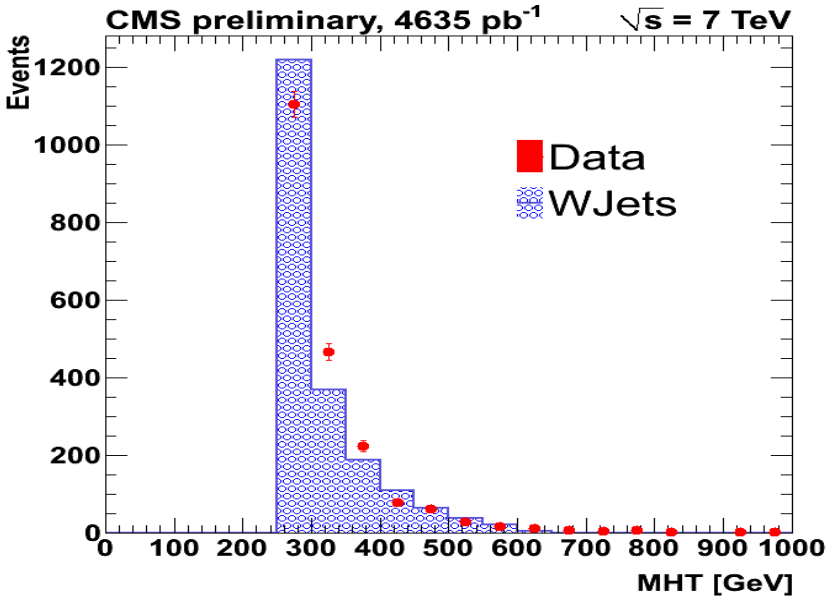

Fig. 8.25. $\not{H}_T$ distribution after subtracting contributions from other SM backgrounds.

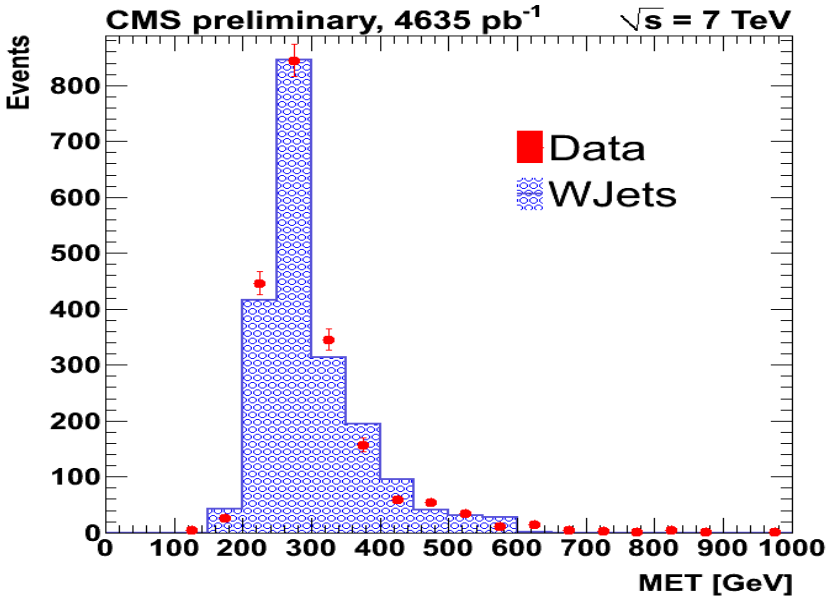

Fig. 8.26. $\not{E}_T$ distribution after subtracting contributions from other SM backgrounds.

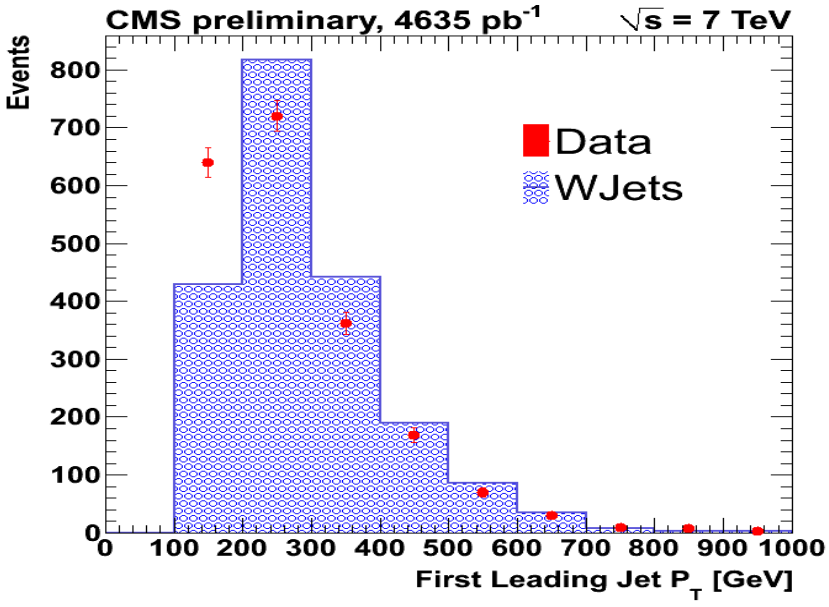

Fig. 8.27. First leading jet p_T distribution after subtracting contributions from other SM backgrounds.

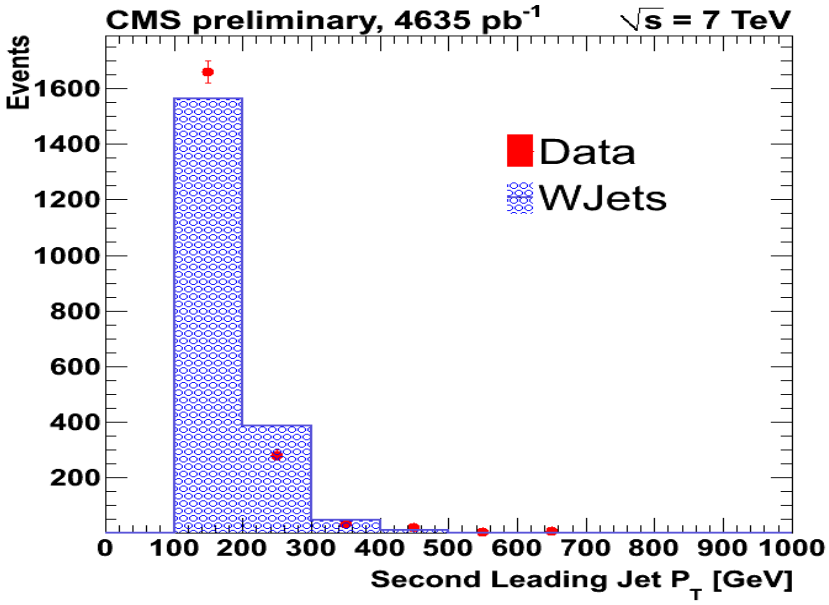

Fig. 8.28. Second leading jet p_T distribution after subtracting contributions from other SM backgrounds.

Table 8.8
W + Jets events extraction efficiencies

Cut	Data
$A_{\tau+j}$	$0.149 \pm 0.016(stat) \pm 0.004(syst)$
A_{j+j}	$0.851 \pm 0.038(stat) \pm 0.004(syst)$
ε^{τ} iso	$0.649 \pm 0.007(stat) \pm 0.045(syst)$
f	$0.019 \pm 0.001(stat) \pm 0.001(syst)$
Probability to tag 0 $b-jets$	$0.83 \pm 0.08(stat+syst)$
Expected Number of Events	5.20 ± 0.89

$0.019 \pm 0.001(\text{stat}) \pm 0.001(\text{syst})$. Combining the two types of events in the W + Jets events control region which consist of events with one real hadronic tau and one jet faking a tau or events with two jets faking two taus, the contribution from W + Jet events is estimated to be 5.20 ± 0.89. This number is used for the W + Jets events estimation in the signal region.

8.6 Summary of Background Estimations

As presented in the previous sections, a careful study was done on the major SM processes that could impact the signal region. However no study was done for WW, WZ and ZZ processes because they provide negligible contributions to both the signal region and the control regions. To validate this, the process with the largest cross section of the three is considered, WW events. Events with direct WW production are expected to be negligible because of the much smaller cross-section [17] than e.g. $t\bar{t}$ events where the topology is similar to that of the signal region. In addition, unlike the $t\bar{t}$ case, WW is not expected to have a large jet multiplicity, thus significantly reducing the chance of jets being tagged as τ_h leptons. Table 8.9 shows a summary of the background estimations with their respective statistical uncertainties.

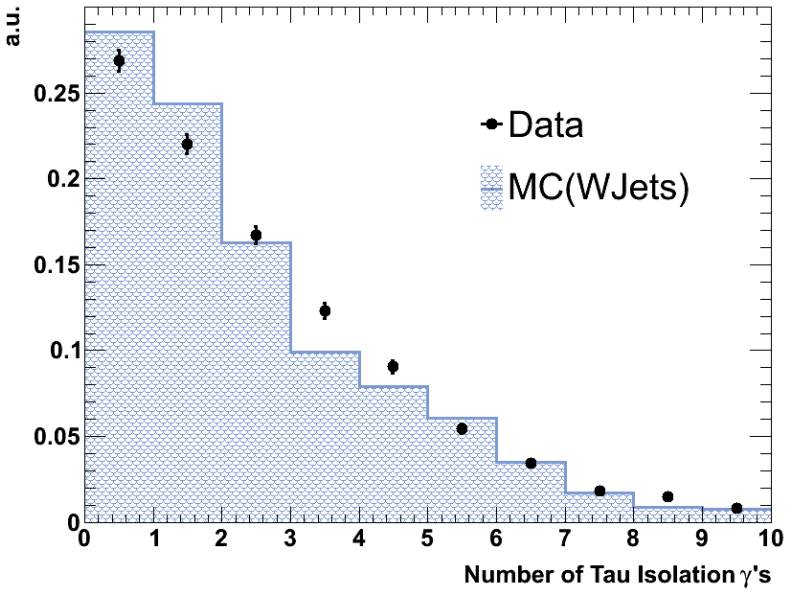

Fig. 8.29. Number of photons.

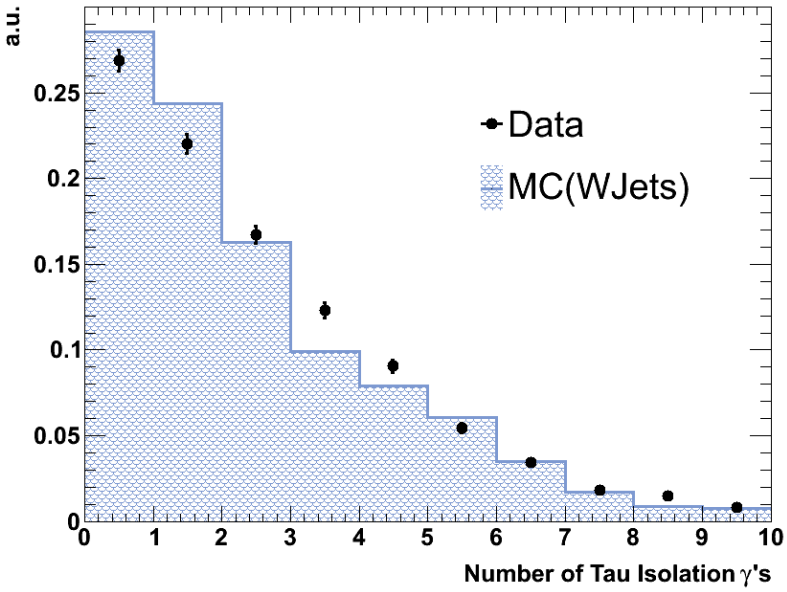

Fig. 8.30. Number of tracks in the τ isolation cone.

Table 8.9
Summary of background predictions with statistical uncertainties.

Process	Signal Region
$t\bar{t}$	2.03 ± 0.36
$Z(\rightarrow \nu\nu)$ + Jets	0.03 ± 0.02
$Z(\rightarrow \tau\tau)$ + Jets	0.21 ± 0.13
QCD	0.02 ± 0.02
W + Jets	5.20 ± 0.63

9 STATISTICS

9.1 CL_s Method

The CL_s is the CMS recommended method of calculating the 95% Confidence Level (CL) upper limit on the cross section. The method uses a log-likelihood ratio (LLR) as a test statistic:

$$LLR = -2ln\frac{p(data|H_1)}{p(data|H_0)} \tag{9.1}$$

where H_1 denotes the test hypothesis under the assumption that both signal and background are present, H_0 is the background only test hypothesis, and p the Poisson probability of observing events ν_i in data for bin i, given an expectation of μ_i:

$$p = \frac{\mu_i^{\nu_i} e^{-\mu_i}}{\nu_i!} \tag{9.2}$$

9.2 Bayesian Method

For the statistical interpretation of data and evaluation of signal significance a Bayesian fit of data against the expected mass distribution is performed to calculate the binned likelihood:

$$\mathcal{L}(\epsilon_1, \epsilon_2, .., \epsilon_n) = \prod_{i=1}^{Nbins} \mathcal{L}_i(\mu_i, \nu_i) \tag{9.3}$$

where $\mathcal{L}_i$ is the Poisson probability of observing events ν_i in data for bin i, given an expectation of $\mu_i(\sigma) = [\text{Background + Signal}(\sigma)]_i$:

$$\mathcal{L}_i = \frac{\mu_i^{\nu_i} e^{-\mu_i}}{\nu_i!} \tag{9.4}$$

The likelihood distribution is used to obtain the 95% CL limit for the signal cross section. For sensitivity studies, pseudo-data samples are generated from background

only distributions, using a Poisson based random event generator. Systematic uncertainties can affect the global normalization of the event rate and also create an uncertainty in the knowledge of the mass shape. To incorporate systematic uncertainties into the likelihood calculation, MC numerical integration methods are used to integrate over nuisance parameters. The nuisance parameters, α_k, are generated according to a log normal probability density function for normalizations and Gaussian for mass spectrum uncertainties. If $\epsilon_{\rm n}$ is an efficiency with systematic error $\delta\epsilon$, the likelihood integral becomes:

$$\int \mathcal{L}(\epsilon_1, \epsilon_2, .., \epsilon_n) \mathrm{d}^n\epsilon = N^{-1} \sum_{j=1}^{N} \mathcal{L}(\epsilon_1 + \alpha_1^j \delta\epsilon_1, \epsilon_2 + \alpha_2^j \delta\epsilon_2, .., \epsilon_n + \alpha_n^j \delta\epsilon_n) \tag{9.5}$$

To incorporate the effects of possible variations in shape, a "morphing" procedure is applied on default, unsmeared mass templates, D_i^{def}, to generate variated templates, D_i^j . Taking into account the smeared templates, the likelihood integral is modified to:

$$N^{-1} \sum_{j=1}^{N} \mathcal{L}(\epsilon_1 + \alpha_1^j \delta\epsilon_1, .., \epsilon_n + {\alpha_n}^j \delta\epsilon_n, D_1^{def} + \alpha_1^j \delta D_1^j, ..., D_n^{def} + \alpha_n^j \delta D_n^j) \tag{9.6}$$

where $\delta\mathcal{D}_i^k = \mathcal{D}_i^k - \mathcal{D}_i^{def}$ is the difference between the default and the deviated shape for the k^{th} systematic effect. The fitting framework also takes into account correlations between systematic uncertainties considered in each channel and across channels. For example, to incorporate correlations across channels, the nuisance parameters can be modified as such:

$$\alpha_k = f * \alpha_f + g * \alpha_g \tag{9.7}$$

where f and g represent the correlated and uncorrelated terms respectively.

10 SYSTEMATIC UNCERTAINTIES

Both signal and background systematic uncertainties must be taken into account for this analysis and are described separately. The systematic uncertainty in the backgrounds is described in section 10.1 while the systematic uncertainty in the signal is described in section 10.2 Both signal and background are affected by the uncertainty in the identification of the τ_h candidate. The systematic uncertainty for τ_h identification is obtained by using a fit of data in a $Z \rightarrow \tau\tau$ enhanced region and fixing the cross section to that measured using $ee/\mu\mu$. This is uncertainty is validated on a control sample of $Z \rightarrow \tau\tau$ events which is selected using a criteria similar to that used for the signal region described in section 5. The difference between collision data and Monte Carlo simulation is at the level of 7%.

10.1 Systematic Uncertainties on Backgrounds

The dominant source of systematic uncertainties on the background estimations are due to uncertainties in the correction factors, the statistical uncertainty on the number of observed events in the CR's ($\delta n \neq \sqrt{N}$), uncertainties in the measured jet $\rightarrow \tau_h$ misidentification rates, and the level of agreement betwen predicted rates and observed number of events in the CR's.

10.1.1 Systematic Uncertainties on the $t\bar{t}$ Prediction

In section 8.1, the $t\bar{t}$ contribution to the signal region was estimated by obtaining a control sample enriched with a high purity of $t\bar{t}$ events where the jet multiplicity and $j \rightarrow \tau_h$ mistaging rate could be measured directly from data. The following equation was utilized to obtain our $t\bar{t}$ signal prediction:

$$
\begin{aligned}
N_{t\bar{t}}^{Signal} &= A_{\tau+j}\frac{N_{t\bar{t}}^{CR}}{P(2\ b\text{-jets})}\varepsilon^{\tau\ \text{iso}}\sum_{N=1}^{\infty}P(N)\sum_{n=1}^{N}C(N,n)f^{n}(1-f)^{N-n} \\
&+ A_{j+j}\frac{N_{t\bar{t}}^{CR}}{P(2\ b\text{-jets})}\sum_{M=2}^{\infty}P(M)\sum_{m=2}^{M}C(M,m)f^{m}(1-f)^{M-m} \qquad (10.1)
\end{aligned}
$$

The fraction of events with one real tau, $A_{\tau+j}$, and the fraction of events with two tau fakes, A_{j+j}, were taken directly from MC simulation and have values of $A_{\tau+j} = 0.166 \pm 0.011$ and $A_{j+j} = 0.834 \pm 0.025$. These values correspond to 224 out of 1352 events with one real tau and 1128 out of 1352 events with no real taus. The uncertainty of 0.011 on $A_{\tau+j} = 0.166$ and of 0.025 on $A_{j+j} = 0.834$ are both purely statistical ($\frac{\sqrt{224}}{1352} = 0.011$ and $\frac{\sqrt{1128}}{1352} = 0.025$, respectively). Since these values represent the fraction of events with real taus in a control sample where the tau identification selections have been removed, they are driven by branching ratios ($W \to \tau\nu$ vs. $W \to q\bar{q}$) and biases from the $\not\!\!H_T$ and jet kinematic selections. Therefore, collision data vs. MC discrepancies are not used to assign a systematic uncertainty to these values. Instead jet energy correction uncertainties are considered as well as tau energy scale systematics in order to quantify how the changes in $\not\!\!H_T$ and jet kinematics results in systematic shifts in the measured values of $A_{\tau+j}$ and A_{j+j}. An uncertainty of $2-5\%$ due to jet energy corrections ($2-5\%$ depending on jet p_T/η) results in an uncertainty 2.7% on $A_{\tau+j}$ and 0.5% on A_{j+j}. Similarly, an uncertainty of 3% due to tau energy scale results in an uncertainty 1.3% on $A_{\tau+j}$ and 0.3% on A_{j+j}. The final values for $A_{\tau+j}$ and A_{j+j} are $A_{\tau+j} = 0.166 \pm 0.011(stat) \pm 0.005(syst)$, $A_{\tau+j} = 0.834 \pm 0.025(stat) \pm 0.005(syst)$.

The probability of tagging at least 2 jets as b-jets is calculated by utilizing the b-tagging efficiency as measured in [14], $\varepsilon_b = 0.684 \pm 0.021$. The uncertainty of $\sim 3\%$ on the b-tagging efficiency is considered as a systematic uncertainty in the $t\bar{t}$ estimate. Therefore, the systematic uncertainty on the probability to tag $\geq 2\ b-jets$ using the track counting high efficiency "medium" working point is determined by

propagating the uncertainty of $\sim 3\%$ on the b-tagging efficiency to the determination of $P(2\ b\text{-jets}) = \varepsilon_b^2$. This results in a systematic uncertainty of 4.3% on $P(2b\text{-jets})$.

The tau isolation efficiency, $\varepsilon^{\tau\ \mathrm{iso}}$, is determined by taking the efficiency in MC simulation and applying a data/MC correction factor. In this case, comparisons between collision data and MC in $Z \to \tau\tau$ and $W \to \tau\nu$ control samples are used to assign systematic uncertainties on $\varepsilon^{\tau\ \mathrm{iso}} = \varepsilon_{MC}^{\tau\mathrm{iso}} \cdot SF_{\tau_h}$. $Z \to \tau\tau$ events were used to show that the level of agreement between data and MC was $\sim 6 - 7\%$. Additionally, the τ_h ID scale factor was extracted directly from data by fixing the cross section to the value measured by CMS in the electron and muon channels, $\sigma(pp \to ZX) \cdot B(Z \to ee, \mu\mu) = 0.931 \pm 0.026(stat.) \pm 0.023(syst.) \pm 0.102(lumi.)$ nb. The extracted value of the τ_h ID scale factor is 1.04 ± 0.068. Therefore, the systematic uncertainty on the tau isolation efficiency, $\varepsilon^{\tau\ \mathrm{iso}}$, is $\sim 7\%$. Since the tau isolation efficiency in MC is $\varepsilon_{MC}^{\tau\ \mathrm{iso}} = 0.55$, a systematic uncertainty of $0.07 \cdot 0.55 = 0.04$ $\to \varepsilon^{\tau\ \mathrm{iso}} = 0.55 \pm 0.04$ is assigned.

The $j \to \tau_h$ mistag rate is determined directly from data by determining the percentage of "τ_h" candidates (which are $\sim 100\%$ jets) within the $\mathrm{t\bar{t}}$ control sample which pass the tau isolation criteria. The measured mistag rate f is 0.022 ± 0.003. The uncertainty of 0.003 is entirely statistical in nature, driven by the number of jets used to calculate the mistag rate. It is important to note the mistag rate has been measured "on average." The standard strategy employed by the CMS collaboration when "fake rate" methods are employed is to utilize a QCD enriched sample of jets to measure the mistag rates vs. p_T and η. The reason the mistag rates are determined in bins of p_T and η is because the topology and kinematics of the QCD dominated control sample are not necessarily similar to the kinematics of the background being estimated. In fact, in most cases, they are entirely different. However, the approach taken in this analysis is to measure the mistag rates directly from background enriched control samples where the proper kinematics are naturally considered. Therefore, the measurement of the mistag rates can be measured "on average" and do

not need to be measured in bins of p_T and η. Since fake rates can be measured "on average", the strategy employed in this analysis has the added advantage that systematic uncertainties on the measured mistag rates are minimized since the number of degrees of freedom have themselves been minimized. Additionally,the correct percentage of gluon jets, light quark jets, etc. is ensured in the sample used to measure the mistag rates. This is an added advantage because mistag rates vary depending on the jet type/structure. This is a complication that often results in large systematic uncertainties in "standard" CMS fake rate techniques due to differences in jet type/structure between the QCD dominated control sample and the background being estimated. Having mentioned the advantages of the strategy pursued in this analysis, a very conservative approach is taken to assign a systematic uncertainty on the mistag rates by studying the dependence with p_T and η. Figure 10.1 shows the $j \rightarrow \tau_h$ mistag rate as a function of p_T. No strong dependence is observed with p_T. A fit of the data points to a "flat line" results in a fit value of 0.02 ± 0.002 with $\chi^2/NDOF = 2.6/6 = 0.4$. The uncertainty on the fit value is used as the systematic uncertainty on the mistag rate ($\sim 10\%$). Therefore, the final value for the $j \rightarrow \tau_h$ mistag rate is $f = 0.022 \pm 0.003(stat) \pm 0.002(syst)$.

10.1.2 Systematic Uncertainties on the W + Jets Prediction

In section 8.5, the W + Jets contribution to the signal region was estimated by obtaining a control sample enriched with a high purity of W + Jets events where the jet multiplicity and $j \rightarrow \tau_h$ mistag rate could be measured directly from data. The following equation was utilized to obtain our W + Jets signal prediction:

$$\begin{aligned} N_{W+jets}^{Signal} &= A_{\tau+j} \frac{N_{W+jets}^{After\ subtraction}}{P(0\ b\text{-jets})} \varepsilon^{\tau} \text{ iso} \sum_{N=1}^{\infty} P(N) \sum_{n=1}^{N} C(N,n) f^n (1-f)^{N-n} \\ &+ A_{j+j} \frac{N_{W+jets}^{After\ subtraction}}{P(0\ b\text{-jets})} \sum_{M=2}^{\infty} P(M) \sum_{m=2}^{M} C(M,m) f^m (1-f)^{M-m} \quad (10.2) \end{aligned}$$

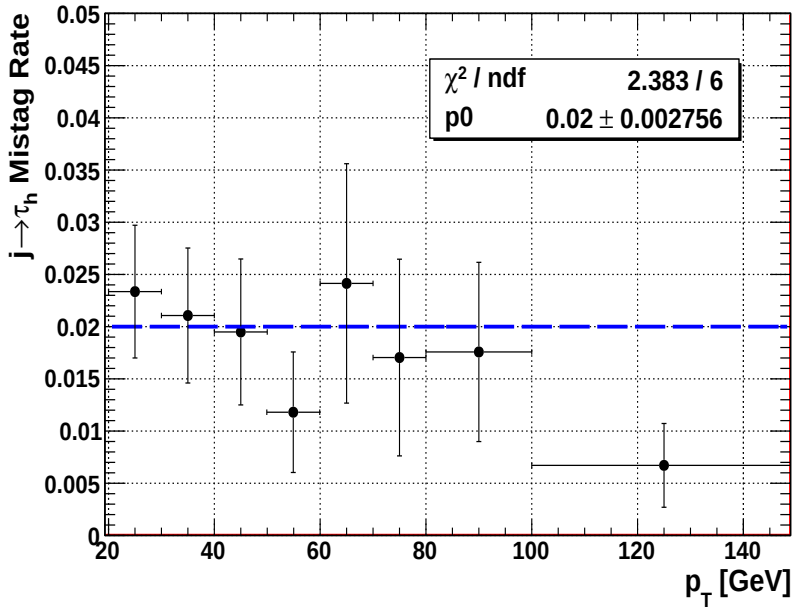

Fig. 10.1. $j \rightarrow \tau_h$ mistag rate vs. p_T obtained from the $t\bar{t}$ dominated control sample.

The fraction of events with one real tau, $A_{\tau+j}$, and the fraction of events with two tau fakes, A_{j+j}, were taken directly from simulation and have values of $A_{\tau+j} = 0.149 \pm 0.016$ and $A_{j+j} = 0.851 \pm 0.038$. These values correspond to 90 out of 605 events with one real tau and 515 out of 605 events with no real taus. The uncertainty of 0.016 on $A_{\tau+j} = 0.149$ and of 0.038 on $A_{j+j} = 0.851$ are both purely statistical ($\frac{\sqrt{90}}{605} = 0.016$ and $\frac{\sqrt{515}}{605} = 0.038$, respectively). Since these values represent the fraction of events with real taus in a control sample where the tau identification selections have been removed, they are driven by branching ratios ($W \rightarrow \tau\nu$ vs. $W \rightarrow q\bar{q}$) and biases from the Hsl and jet kinematic selections. Therefore, collision data vs. MC discrepancies are not used to assign a systematic uncertainties to these values. Instead the same strategy discussed in section 10.1.1 is used while taking into consideration jet energy correction uncertainties, and tau energy scale systematics to determine the systematic uncertainties on $A_{\tau+j}$ and A_{j+j}. An uncertainty of

$2-5\%$ due to jet energy corrections results in an uncertainty 2.2% on $A_{\tau+j}$ and 0.3% on A_{j+j}. Similarly, an uncertainty of 3% due to tau energy scale results in an uncertainty 1.1% on $A_{\tau+j}$ and 0.2% on A_{j+j}. The final values for $A_{\tau+j}$ and A_{j+j} are $A_{\tau+j} = 0.149 \pm 0.016(stat) \pm 0.004(syst)$, $A_{\tau+j} = 0.851 \pm 0.038(stat) \pm 0.004(syst)$.

The probability to mistag zero jets as b-jets, $P(0\ b\text{-jets})$, is calculated by utilizing the $j \rightarrow$b mistag rate as measured in [14], $f_b = 0.028$, and determining the probability to tag n fake b-jets given N jets as described in section 8.5. A central value of $N = 4$ jets (as determined by the observed jet multiplicity distribution in data) is used to determine the central value of the probability to mistag exactly zero jets as b-jets, which has a calculated value of $P(0\ b\text{-jets}) = 0.83$. To assign a systematic uncertainty to this value, a variation to the mean of jet multiplicity distribution of ± 1 is applied to then quantify how much the calculation of $P(0\ b\text{-jets})$ varies from the central value obtained with $N = 4$. A systematic uncertainty of 9.6% is calculated on the measured value for $P(0\ b\text{-jets})$.

The systematic uncertainty on the tau isolation efficiency, $\varepsilon^{\tau\ \mathrm{iso}}$, is determined using the same argument/method described in section 10.1.1. Therefore, since the tau isolation efficiency in MC is $\varepsilon^{\tau\ \mathrm{iso}}_{MC} = 0.649$, a systematic uncertainty of $0.07 \cdot 0.649 = 0.045 \rightarrow \varepsilon^{\tau\ \mathrm{iso}} = 0.649 \pm 0.045$ is assigned.

The systematic uncertainty on the $j \rightarrow \tau_h$ mistag rate is determined using the same argument/method described in section 10.1.1 by studying the dependence with p_T and η. Figure 10.2 shows the $j \rightarrow \tau_h$ mistag rate as a function of p_T. No strong dependence is observed with p_T. A fit of the data points to a "flat line" result in a fit value of 0.018 ± 0.001 with $\chi^2/NDOF = 6.2/7 = 0.9$. We use the uncertainty on the fit value as the systematic uncertainty on the mistag rate ($\sim 5.6\%$). Therefore, the final value for the $j \rightarrow \tau_h$ mistag rate is $f = 0.019 \pm 0.001(stat) \pm 0.001(syst)$.

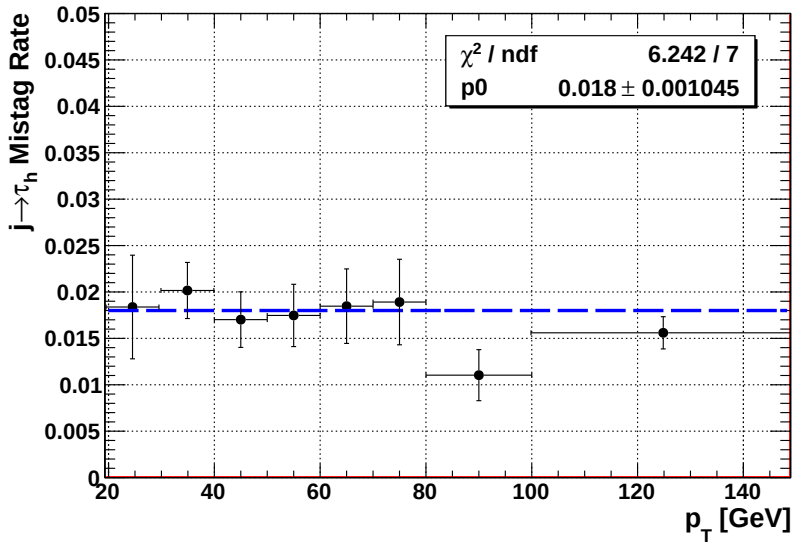

Fig. 10.2. $j \rightarrow \tau_h$ mistag rate vs. p_T obtained from the W + Jets dominated control sample.

10.1.3 Systematic Uncertainties on the Z + Jets Prediction

In sections 8.2 and 8.3, the Z + Jets contribution to the signal region was estimated by obtaining a control sample enriched with a high purity of $Z(\rightarrow \mu\mu)$ + Jets events, where efficiencies for Z + Jets events are measured and used to extrapolate to the signal region after proper corrections for the muon efficiencies and branching fraction to muons.

The branching fractions for Z are taken from the CMS collaboration Particle Data Group (PDG). Therefore, the uncertainty on the branching fractions as referenced by the PDG serves as a systematic uncertainty in the estimation of Z + Jets.

The determination of the trigger efficiency for HLT_PFMHT150, $\varepsilon_{\not{H}_T}^{Trigger}$, is determined by utilizing a $t\bar{t}$ control sample (see section 7). The systematic uncertainty on $\varepsilon_{\not{H}_T}^{Trigger}$ is determined by fitting the trigger turn-on curve to an error function and extracting the uncertainty on the fit parameter corresponding to the plateau. A

systematic uncertainty of $\sim 2.5\%$ is calculated. Additionally, since the uncertainties on the fit parameters are correlated and are often constrained by the values and uncertainties in the other parameters (constrained by the high statistics of the rising slope) through the error matrix, an additional validation to the $\sim 2.5\%$ systematic uncertainty is done by fitting the region $\not{H}_{\mathrm{T}} > 250$ GeV to a "flat line" and extracting the uncertainty on the constant fit parameter. This approach also validates the use of $\sim 2.5\%$ as the systematic uncertainty on the trigger efficiency for HLT_PFMHT150, $\varepsilon_{\not{H}_{\mathrm{T}}}^{Trigger}$.

The determination of the trigger efficiency for the $\mu\tau$ cross-triggers was performed as part of the $H/Z' \to \tau\tau$ analyzes [18]. The strategy consists of using standard tag and probe methods in $Z \to \mu\mu$ to determine the efficiency on the HLT_Mu and utilizing a control sample of $Z \to \tau\tau \to \mu\tau_h$ events selected with single muon triggers to measure the efficiency for HLT_PFTau. The trigger efficiency $\varepsilon_{\mu\tau}^{Trigger}$ has a measured value of 0.87 ± 0.04. The uncertainty of $\sim 4.6\%$ contains a systematic uncertainty due to HLT_Mu ($\sim 1.15\%$) and a systematic uncertainty due to HLT_PFTau ($\sim 4\%$). Therefore, the systematic uncertainty of $\sim 4.6\%$ on $\varepsilon_{\mu\tau}^{Trigger}$ is treated as a systematic uncertainty in the background estimation.

The muon acceptance A_μ is determined from MC simulation. In this case, the level of agreement between collision data and MC is used in the $Z(\to \mu\mu)$ + Jets control sample to assign a systematic uncertainty on A_μ. The observed number of events in collision data is 738, while the predicted rate for $Z(\to \mu\mu)$ + Jets and $t\bar{t}$ from MC simulation is 709.5 and 6.81 respectively. Therefore, the level of agreement between collision data and MC is $\frac{(738.-6.81-709.5)}{709.5} \sim 3\%$. The level of agreement between collision data and MC in this control sample is driven by much more than just the muon acceptance. Therefore, a systematic uncertainty of $\sim 3\%$ is very conservative. Also, to remain on the conservative end, the systematic uncertainty on the muon momentum scale (MMC $\sim 1\%$) and jet energy corrections (JEC) is used to assign an additional systematic uncertainty on A_μ. It was found

that a $\sim$ 1% systematic uncertainty on the muon momentum scale results in << 1% uncertainty on the muon acceptance. Similarly, a $2-5\%$ uncertainty due to jet energy corrections results in an uncertainty of $\sim$ 1.5% on A_μ. Therefore, the quoted value for A_μ is $0.701 \pm 0.004(stat) \pm 0.027(data-MC) \pm 0.01(JEC/MMS) = 0.701 \pm 0.004(stat) \pm 0.029(syst)$. These uncertainties are propagated through the estimation of the Z + Jets contribution in the signal region. A similar approach is taken for A_τ: the systematic uncertainty on tau energy scale ($\sim$ 3%) and jet energy corrections (JEC) are used to assign a systematic uncertainty on A_τ. A $\sim$ 3% systematic uncertainty on the tau energy scale results in 2% uncertainty on the tau acceptance. Similarly, a $2-5\%$ uncertainty due to jet energy corrections results in an uncertainty of $\sim$ 1.8% on A_τ. Therefore, the quoted value for A_τ is $0.503 \pm 0.001(stat) \pm 0.01(TES) \pm 0.009(JEC) = 0.503 \pm 0.001(stat) \pm 0.014(syst)$.

The muon identification efficiency, ε_μ, is measured using standard tag and probe methods in $Z \rightarrow \mu\mu$. They validated as part of the $H/Z' \rightarrow \tau\tau$ analyzes [18]. The efficiency ε_μ is known to very good accuracy as muons are fairly well measured objects and control samples of $Z \rightarrow \mu\mu$ have high purity high statistics. The collision data to MC agreement in the $Z(\rightarrow \mu\mu)$ + Jets control sample provides a very conservative upper limit on the systematic uncertainty on ε_μ. However, as this has already been included in the systematic uncertainty for A_μ, it is not included in ε_μ.

The efficiency for $\not\!\!H_{\rm T} > 250$ GeV, $\varepsilon^{\not\!\!H_{\rm T}}$, is determined by calculating the percentage of events from the $\not\!\!H_{\rm T}$ distributions in the region $\not\!\!H_{\rm T} > 250$ GeV. Therefore, for $Z(\rightarrow \nu\nu)$ + Jets the efficiency has a measured value of $\varepsilon^{\not\!\!H_{\rm T}} = \frac{6}{738} = 0.0081 \pm 0.0033$. The uncertainty of 0.0033 represents the purely statistical uncertainty $\frac{\sqrt{6}}{738} = 0.0033$. To be conservative, the level of agreement between collision data and MC is used in the region $\not\!\!H_{\rm T} > 250$ GeV to assign a systematic uncertainty on $\varepsilon^{\not\!\!H_{\rm T}}$. Since $\varepsilon^{\not\!\!H_{\rm T}}_{MC}$ in MC is 0.004, while $\varepsilon^{\not\!\!H_{\rm T}}_{Data}$ in collision data is 0.008, a systematic uncertainty of $\frac{0.008-0.004}{0.008} = 50\%$ to $\varepsilon^{\not\!\!H_{\rm T}}$ is assigned. Therefore, the quoted value for $\varepsilon^{\not\!\!H_{\rm T}}$ in $Z(\rightarrow \nu\nu)$

+ Jets is $0.0081 \pm 0.0033(stat) \pm 0.004(syst)$. This uncertainty is propagated through the estimation of the $Z(\rightarrow \nu\nu)$ + Jets contribution in the signal region.

The systematic uncertainty on the $j \rightarrow \tau_h$ mistag rate is determined using the same argument/method described in section 10.1.1 by studying the dependence with p_T and η. Figure 10.2 shows the $j \rightarrow \tau_h$ mistag rate as a function of p_T. No strong dependence is observed with p_T. A fit of the data points to a "flat line" results in a systematic uncertainty on the mistag rate of $\sim 5.6\%$). Therefore, the final value for the $j \rightarrow \tau_h$ mistag rate is $f = 0.0164 \pm 0.0019(stat) \pm 0.001(syst)$.

10.2 Systematic Uncertainties on Signal

One of the main sources of systematics is due to the uncertainty in the calculation of the backgrounds and the imprecise knowledge of the luminosity (2.2%). Sources of systematics such as trigger efficiencies, identification efficiencies, energy and momentum scale, parton distribution functions, and initial and final state radiation have been included. In all cases, the values used are the ones recommended by the CMS collaboration and the corresponding physics objects group (POG). Scale factors for τ_h identification are taken from the tau POG and obtained using a fit of data in a $Z \rightarrow \tau\tau$ enhanced region and fixing the cross section to that measured using $ee/\mu\mu$. Systematic uncertainties on the triggers are measured using a $t\bar{t}$ enhanced region obtained by selecting data with the $\mu\tau_h$ and $e\tau_h$ cross-triggers as described in section 7. Tau and jet energy scale systematics also affect the knowledge of the signal acceptance. The $\not\!\!E_T$ scale uncertainties contribute via the jet energy scale (2-5% depending on η and p_T) and unclustered energy scale (10%), where unclustered energy is defined as the energy found "outside" the reconstructed leptons and jets with $p_T > 10$ GeV.The unclustered energy scale uncertainty has a negligible systematic effect on the signal acceptance and M_{eff} shape. Uncertainties that contribute to the M_{eff} shapes include tau energy scale, jet energy scale, and the $\not\!\!E_T$ scale. The systematic effect due to the imprecise knowledge of the parton distribution functions (PDFs)

is determined by comparing CTEQ6.6, MRST2006, and NNPDF10 PDFs with the default PDF and variations within the family of parameterizations. The PDF4LHC recommendation is employed to determine the overall systematic uncertainty due to the parton distribution functions. The systematic effect due to imprecise modeling of initial and final state radiation is determined by re-weighting events to account for effects such as missing α terms in the soft-collinear approach [19] and missing NLO terms in the parton shower approach [20]. Table 10.1 summarizes the sources of systematics considered and their effect on the acceptance. Figures 10.3–10.6 show the effect of τ_h/jet energy scale on signal for the mSUGRA benchmark point. In order to quantify the systematic effect of pile-up on the signal acceptance, the acceptance efficiency is parametrized as a function of the number of reconstructed vertices. Figure 10.7 shows that the signal acceptance is constant as a function of the number of vertices. To determine the systematic effect of PU on the signal acceptance, the dependence of $\varepsilon^{signal}(N_{vtx})$ is fitted to a constant factor. The resultant fit value is denoted as X_{avg}. To determine whether there is a systematic deviation of the signal acceptance from a "flat" dependence, the spread of the points in Figure 10.7 are calculated from the central fit value:

$$\sigma_A^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i^2 - X_{avg}^2) \tag{10.3}$$

where N is the number of points used in the fit and X_i is the signal acceptance for point i. To untangle the statistical uncertainty from the systematic uncertainty due to a possible dependence of the acceptance on PU, σ_A^2 is compared to the spread due to the known statistical uncertainties δX^2 on the points:

$$\sigma_B^2 = \frac{1}{\sum_{i=1}^{N}\frac{1}{\delta X_i^2}} \tag{10.4}$$

If $\sigma_A > \sigma_B$,a systematic uncertainty $\sigma_{sys} = \sqrt{\sigma_A^2 - \sigma_B^2}$ is assigned. The entire parameter space was scanned for the benchmark SUSY models and it was found that the systematic uncertainty due to PU is negligible in all cases.

Table 10.1
List of systematic uncertainties for signal and background estimates.

Source of Systematic	Systematic Uncertainty
Luminosity	2.2%
$\not{H}_{\mathrm{T}}$ Trigger	2.5%
Tau ID	6.8%
Parton Distribution Functions	11.0%
Initial State Radiation	*Negligible*
Final State Radiation	*Negligible*
Tau Energy Scale (3.0%)	2.3%
Jet Energy Corrections (2-5%)	4.6%
Pile-up	*Negligible*
Background Estimation	15.6%

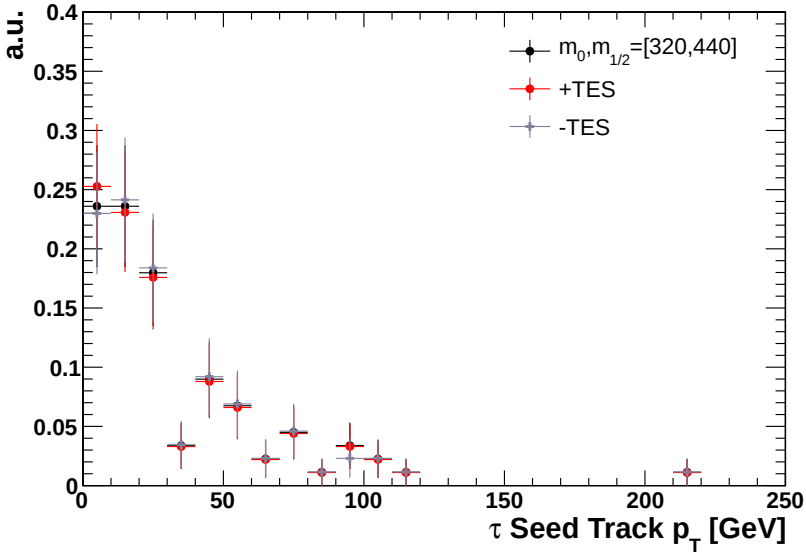

Fig. 10.3. Depiction of the systematic effect of tau energy scale on signal (τ seed track p_T).

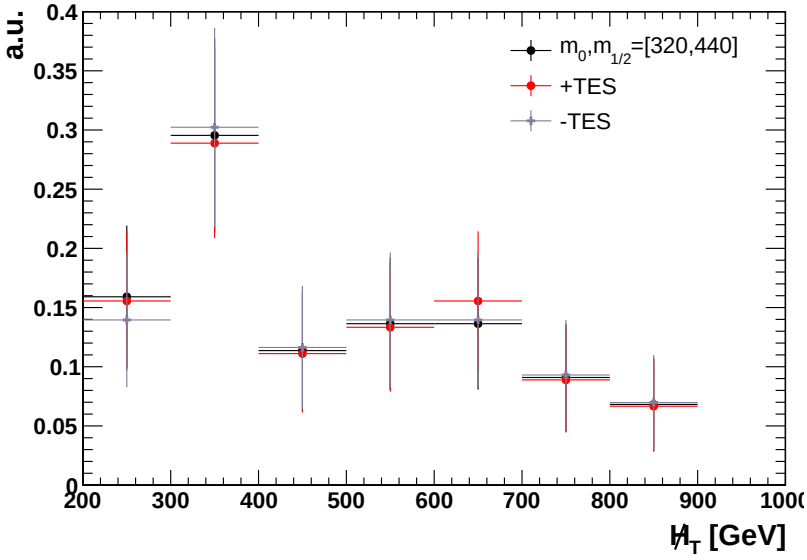

Fig. 10.4. Depiction of the systematic effect of tau energy scale on signal ($\not{H}_{\mathrm{T}}$).

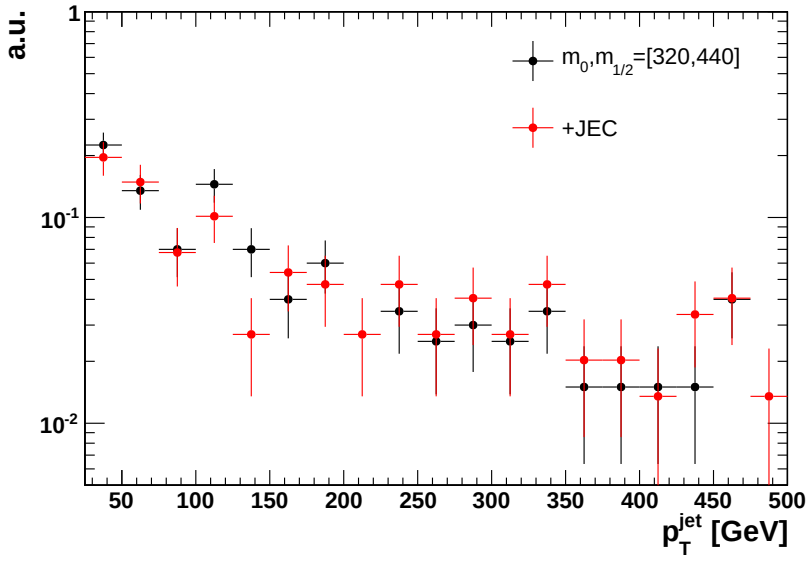

Fig. 10.5. Depiction of the systematic effect of JEC on signal (jet p_T).

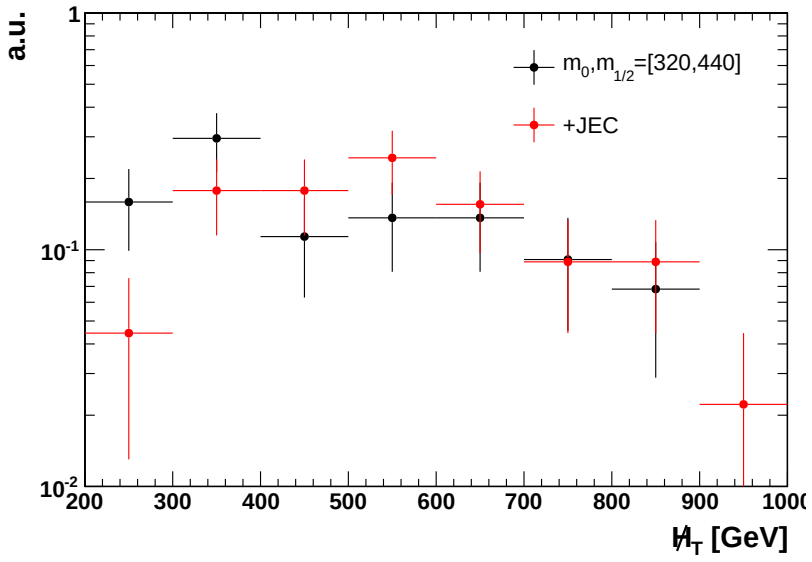

Fig. 10.6. Depiction of the systematic effect of JEC on signal ($\not{H}_T$).

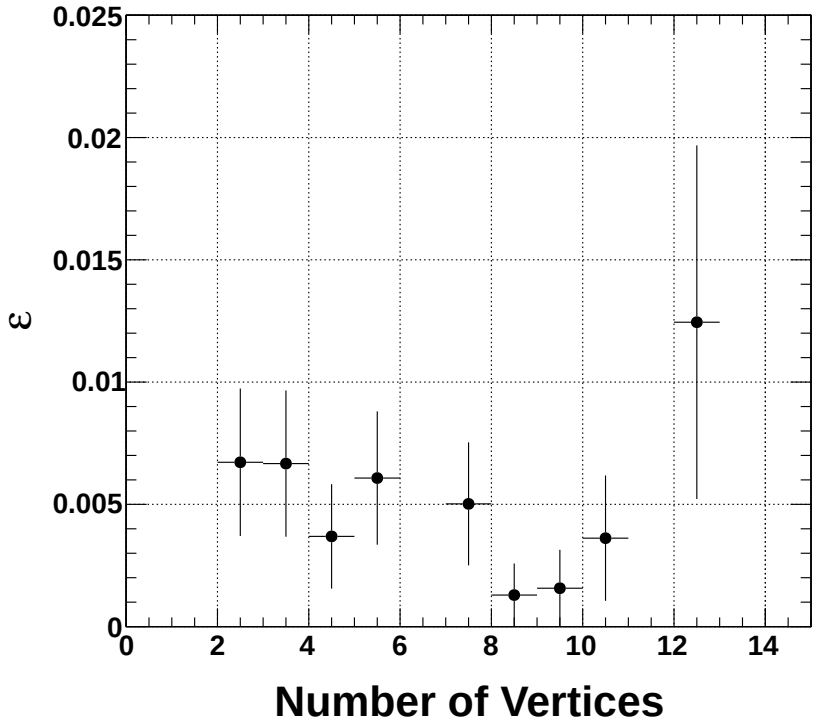

Fig. 10.7. Signal acceptance vs. Number of vertices.

11 RESULTS

11.1 Data in the Signal Region

The largest sources of background in this analysis come from SM top-pair production events and from W + jets events. A counting experiment was performed to compare the SM background predictions from the collision data collected at the CMS detector in 2011. Table 11.1 lists these SM background predictions and the number of observed events in the signal region. Figure 11.1 shows the H_{T} distribution in the signal region while Fig. 11.2 shows the $M_{\mathrm{eff}}(= \not{H}_T + H_T)$ distribution in the signal region. The background distributions in Figs. 11.1–11.2 are taken from MC simulation and are normalized to the predictions based on data over the full spectrum. The estimated number of events due to the SM background processes are in agreement with the number of observed events in the signal region. Thus, no deviation from SM physics was observed in this analysis.

Table 11.1
Number of data events and estimated background rates with statistical and systematic uncertainties, respectively.

Process	Signal Region
QCD multijet events	$0.02 \pm 0.02 \pm 0.17$
W+jets	$5.20 \pm 0.63 \pm 0.62$
$t\bar{t}$	$2.03 \pm 0.36 \pm 0.34$
$Z(\rightarrow \tau\bar{\tau})$+ jets	$0.21 \pm 0.13 \pm 0.17$
$Z(\rightarrow \nu\bar{\nu})$+ jets	$0.03 \pm 0.02 \pm 0.50$
Estimated $\sum SM$	$7.49 \pm 0.74 \pm 0.90$
Data	9

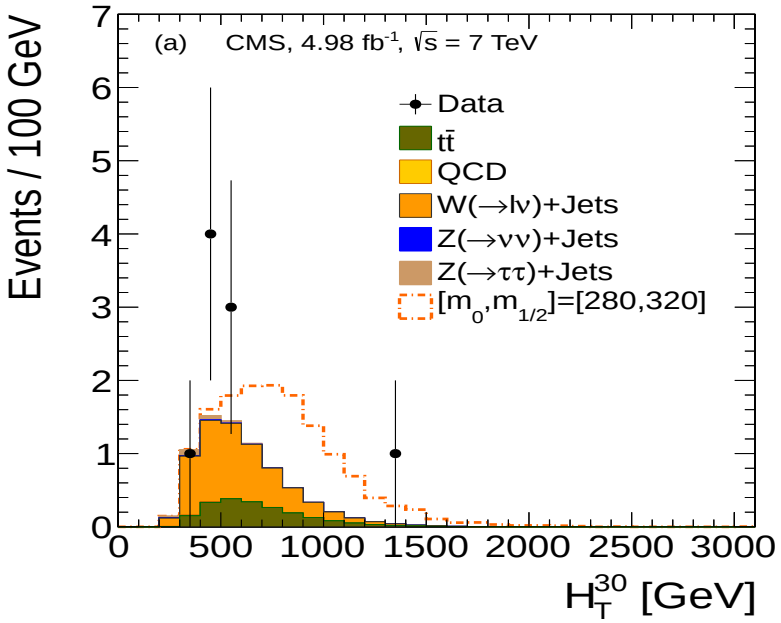

Fig. 11.1. Stacked distributions of H_{T}.

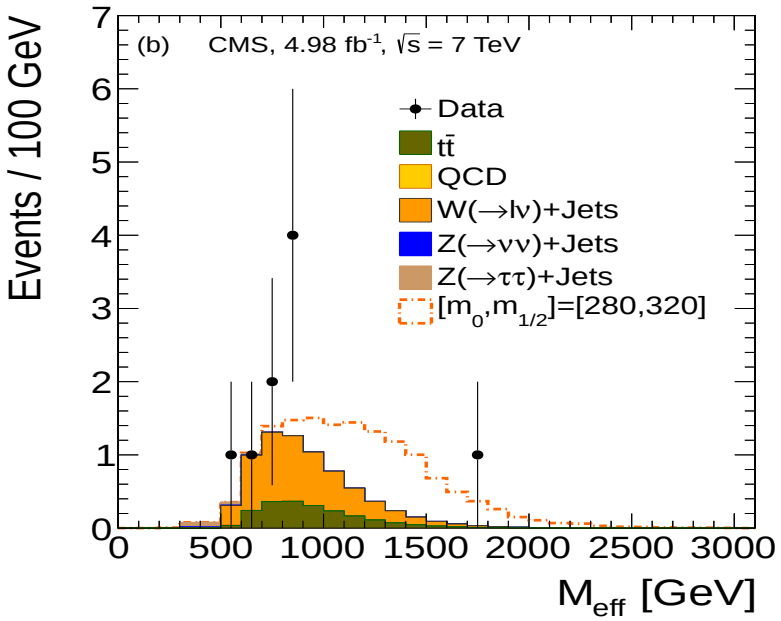

Fig. 11.2. Stacked distributions of $\mathrm{M_{eff}}$ in the signal region.

11.2 Limits to New Physics

11.2.1 CMSSM

Limits are set using the CL_S criterion in the context of CMSSM. The parameter space with $\tan\beta = 40$, A_0 = -500 GeV, $\mu > 0$ and M_t = 173.2 GeV is chosen as a possible scenario with a light $\tilde{\tau}$ and a value of $\Delta M \leq 20$ GeV. The limits are set using a counting experiment where systematic uncertainties are treated as nuisance parameters and marginalized. The signal contamination is taken into account. In this analysis a common gaugino mass ($m_{1/2}$) of < 495 GeV is excluded at 95% CL for a common scalar mass (m_0) of 400 GeV. A gluino with mass < 1.15 TeV is excluded at 95% CL in this region. Figure 11.3 shows the excluded regions, the solid red line denotes the experimental limit and the dotted red lines represent the uncertainty on the experimental limit due to scale variations by a factor of two, and PDF effects on the theoretical cross sections. The blue band represents the expected uncertainties.

11.2.2 SMS

The results are also interpreted in the context of a simplified scenario (SMS) [21]. The $\tau\tau$ SMS scenario (T3tauh) is studied where gluinos are produced in pairs and subsequently decay to τ lepton pairs and a LSP via neutralino ($\tilde{g} \rightarrow q\bar{q}\tilde{\chi}_2^0 \rightarrow \tau + \tilde{\tau} + \text{LSP}$) and the mass of $\tilde{\chi}_2^0$ is the average of the masses of the gluino and the LSP. A Feynman diagram for this scenario is shown in Figure 11.4.

In the SMS scenario a gluino mass of < 740 GeV is excluded at 95% CL for LSP masses up to 290 GeV. This results are shown in Figure 11.5 where the limits of the gluino and LSP are shown with a solid black line.

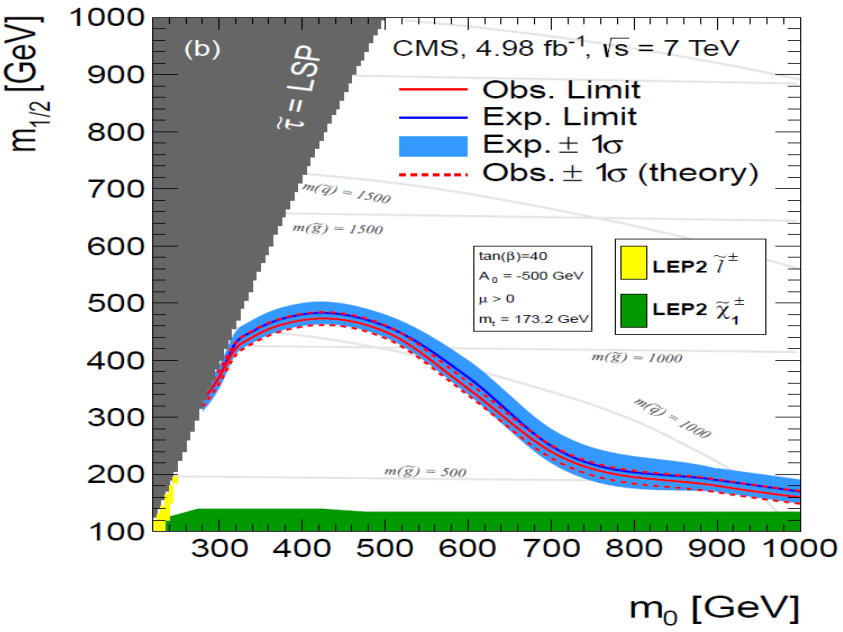

Fig. 11.3. Exclusion limit in the CMSSM plane at $\tan\beta = 40$.

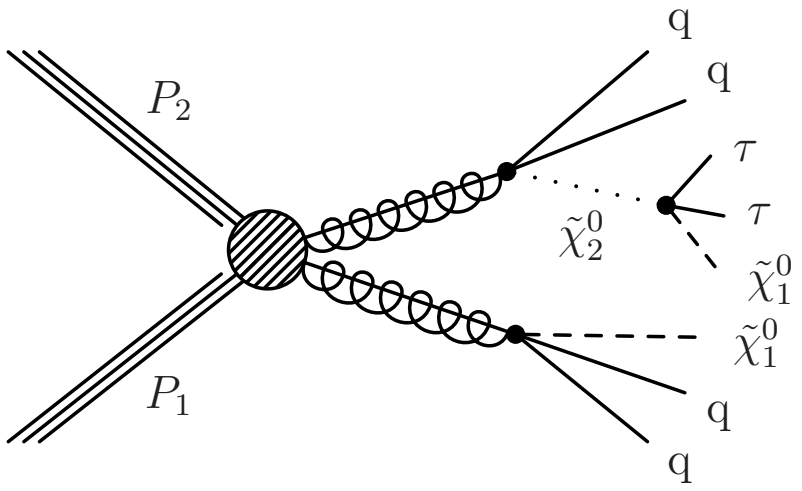

Fig. 11.4. Feynman diagram for the $\tau\tau$ SMS model.

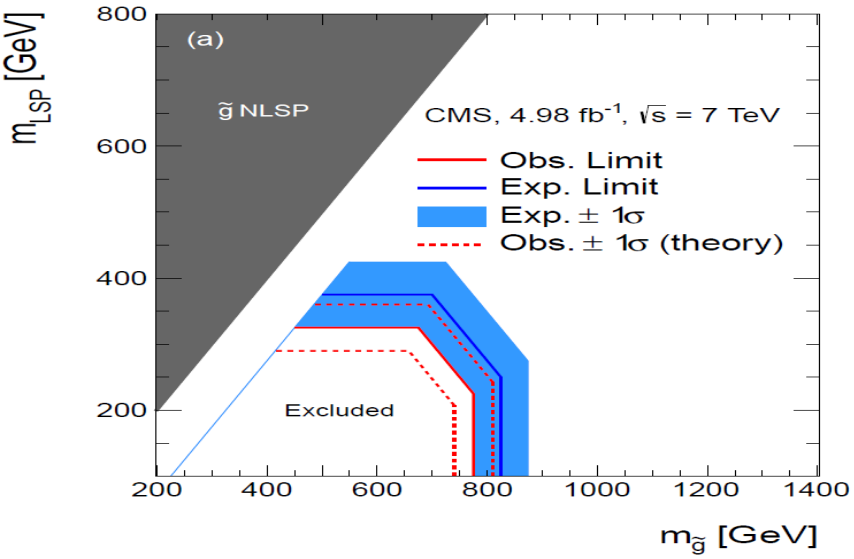

Fig. 11.5. 95% CL cross section upper limits for the T3tauh model where the solid red line represents the limits on the mass of the gluino and the LSP.

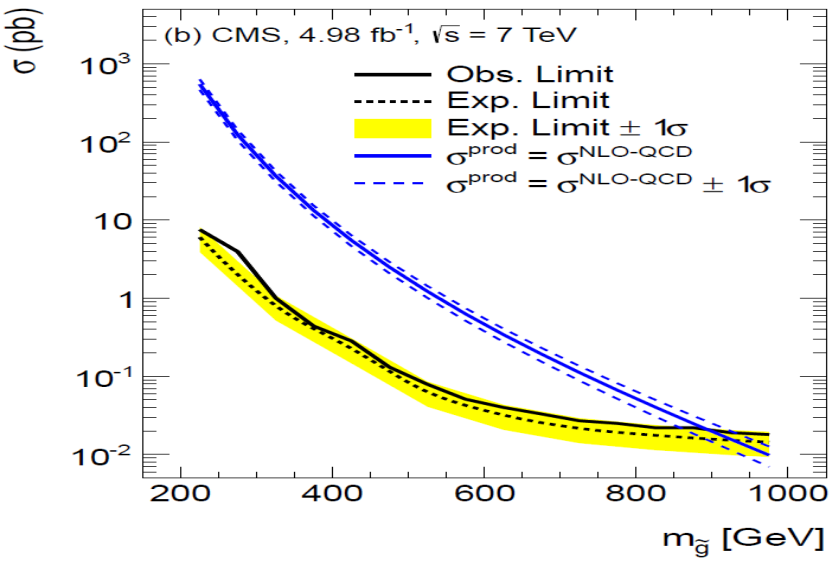

Fig. 11.6. 95% CL cross section upper limits as a function of gluino mass in the GMSB scenario.

11.2.3 GMSB

In the simplified gauge mediated symmetry breaking (GMSB) scenario, the $\tilde{\tau}$ is the NLSP and decays to a τ-lepton and a gravitino ($\tilde{G}$) with a mass of the order of keV [22–24] ($\tilde{\chi}_2^0 \to \tau\tilde{\tau} \to \tau\tau\tilde{G}$). The topology for this simplified GMSB is similar to that of the T3tauh except for the assumption that both gluinos decay to τ-lepton pairs with a branching fraction of 100%. Then the results can also be interpreted in the GMSB scenario using the T3tauh scenario by correcting the signal acceptance to account for the final state containing up to four τs. In this scenario a gluino mass $<$ 860 GeV is excluded with 95% CL, this is shown in Figure 11.6 as a function of the gluino mass.

12 CONCLUSIONS

A search for SUSY in events with two or more highly energetic jets, two or more hadronically decaying taus and a large momentum imbalance was performed. The data used corresponds to 5 fb^{-1} of CMS data from 7 TeV proton-proton collisions at the LHC. The observed number of events is consistent with the SM background contributions from $t\bar{t}$, W + jets, Z ($\rightarrow \nu\nu$) + jets, Z ($\rightarrow \tau\tau$) + jets and QCD events. Thus, there is no evidence of physics beyond the standard model.

To reduce the reliance on simulation the SM backgrounds were estimated using robust data driven methods. The largest uncertainties are of a statistical nature. Upper limits on the cross sections for CMSSM, GMSB and SMS scenarios were set to 95% CL. Within the CMSSM framework at $\tan\beta = 40$, a gaugino mass $m_{1/2} <$ 495 GeV is excluded at 95% CL for scalar masses m_0 <400 GeV. This results places a lower limit in the mass of the gluino at 1.15 TeV with 95% CL at the peak. For the T3tauh simplified scenario a gluino mass of less than 740 GeV is excluded while for the GMSB scenario a gluino mass < 860 GeV is excluded at 95% CL.

REFERENCES

[1] H. Spiesberger, M. Spira, and P. Zerwas, "The Standard Model: Physicsl Basis and Scattering Experiments", http://arxiv.org/abs/hep-ph/0011255, (2000).

[2] S.J. Gates Jr, M.T. Grisaru, M. Rocek, and W. Siegel, "Superspace, or One Thousand and One Lessons in Supersymmetry", http://arxiv.org/abs/hep-th/0108200, (2001).

[3] S. P. Martin, "A Supersymmetry Primer", http://arxiv.org/abs/hepph/9709356, (1997).

[4] A. H. Chamseddine, R. Arnowitt, and Pran Nath, "Locally Supersymmetric Grand Unification", Phys. Rev. 49, 970-974 (1982).

[5] L. Hall, J. Lykken and S. Weinberg, "Supergravity as the Messenger of Supersymmetry Breaking", Phys. Rev. D 27, 2359-2378 (1983).

[6] E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold et al., "Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation", Astrophys. J. Suppl. 192, 18-75 (2011).

[7] A. Breskin, "The CERN Large Hadron Collider: Accelerator and Experiments", Geneva, CERN Publication Vol 1, (2009).

[8] Science blogs of physical sciences, "Aerial View of the LHC", http://www.webelements.com/nexus/aggregator/sources/23?page=4, Retreived 6 June, 2011.

[9] CMS Collaboration, "The CMS Experiment at the CERN LHC", J. Instrum. 3, S08004 (2008).

[10] CERN, "CERN Accelerator Complex", http://atlas.kek.jp/sub/photos/Accel erator/PhotoAccelerator.html, 2011.

[11] A. Breskin, "The CERN Large Hadron Collider: Accelerator and Experiments", Geneva, CERN Publication Vol 2, (2009).

[12] M. L. Perl, G. S. Abrams, A. M. Boyasski, M. Breidenbach, D. D. Briggs et al.,"Evidence for Anomalous Lepton Production in e^+ - e^- Annihilation", Phys. Rev . Lett. 35, 1489-1492, (1975).

[13] M. Cacciari, G. P. Salam and G. Soyez, "The Anti-k_t Jet Clustering Alrorithm", http://arxiv.org/abs/0802.1189v2, (2008).

[14] CMS Collaboration, "Performance of the b-jet Identification in CMS", CMS PAS BTV-11-001, http://cds.cern.ch/record/1366061, (2011).

[15] CMS Collaboration, "Performance of τ-lepton Reconstruction and Identification in CMS", JINST 7, 1001, (2012).

[16] CMS Collaboration, "Search for Higgs Bosons in Di-tau Final States With CMS Run2011 data", CMS Note (2011).

[17] CMS Collaboration, "Measurement of the WW, WZ and ZZ Cross Sections at CMS", CMS PAS EWK-11-010, http://cds.cern.ch/record/1370067, (2011).

[18] CMS Collaboration, "Search for Heavy Resonances Decaying to Di Tau States", CMS Note (2011).

[19] G. Nanava, and Z. Was, "How to Use SANC to Improve the PHOTOS Monte Carlo Simulation of Bremsstrahlung in Leptonic W-Boson Decays", Acta Phys.Polon. B34, 4561-4570, (2003).

[20] G. Miu, and T. Sj *o*strand, "W Production in an Improved Parton-Shower Approach", Phys. Let. B 449, 313-320, (1999).

[21] D. Alves, N. Arkani-Hamed, S. Arora, Y. Bai, M. Baumgart et al., "Simplified Models for LHC New Physics Searches", J. Phys. G: Nucl. Part. Phys. 39, (2012).

[22] M. Dine and W. Fishler, "A Phenomenological Model of Particle Physics Based on Supersymmetry", Phys. Lett. B 110, 227-231, (1982).

[23] C. R. Nappi and B. A. Ovrut, "Supersymmetric Extension of the $SU(3) \times SU(2) \times U(1)$ Model", Phys. Lett. B113, 175-179, (1982).

[24] L. Alvarez-Gaume, M. Claudson, and M. B.Wise, "Low-Energy Supersymmetry", Nucl. Phys. B207, 96-110, (1982).

APPENDIX A
EFFECT OF B-TAGGING ON CONTROL REGIONS

The use of b-tagging plays a major role in the creation of a high purity $t\bar{t}$ events control region. To show that the b-tagging requirement does not introduce a bias in the measured efficiencies it is necessary to determine the effect of b-tagging in the tau isolation. The study presented here uses MC simulated events to show that the requirement of two jets tagged as b-jets does not bias the isolation and kinematic distributions. To show this the relevant distributions are made with no b-tagging requirement as well as with the requirement of more than 1 jet tagged as a b-jet. The Figures A.1–A.4 below display the tau isolation distributions as well as the tau kinematic distributions for $t\bar{t}$ events within the constraints of this study. These distributions show that the b-tagging requirement does not produce a bias on the measurement of the efficiencies in the $t\bar{t}$ events control region.

Similar cross-checks were made for all control regions to validate the use of b-tagging. In all cases it was found that the use of b-tagging produces negligible effects. For example, Figures A.5–A.6 show the the tau isolation distributions and kinematic distributions for QCD events with and without the b-tagging requirement. As expected, the b-tagging requirement does not produce any bias on the measurement of efficiencies in the QCD control region.

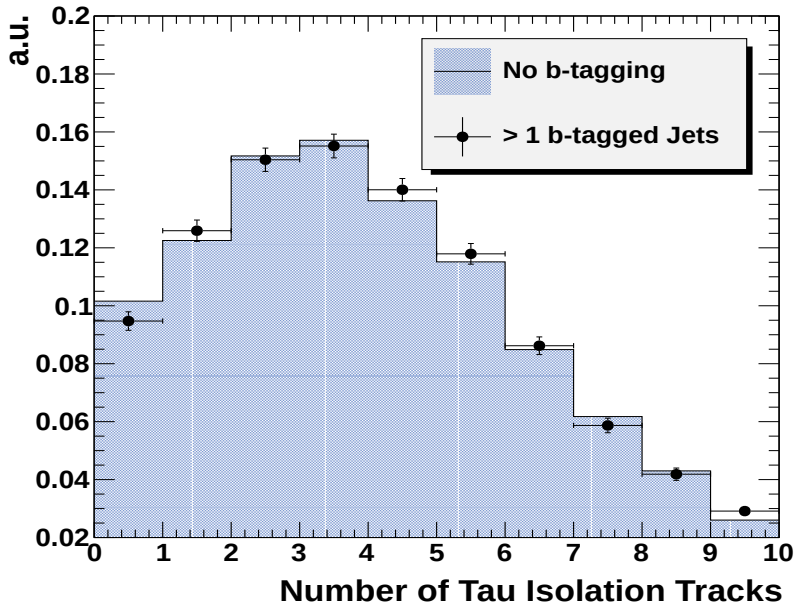

Fig. A.1. τ track isolation.

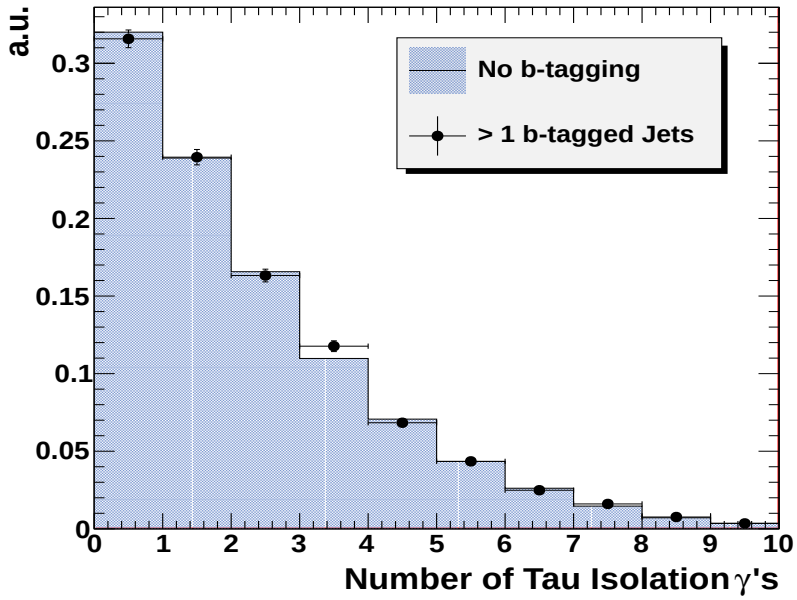

Fig. A.2. τ γ isolation.

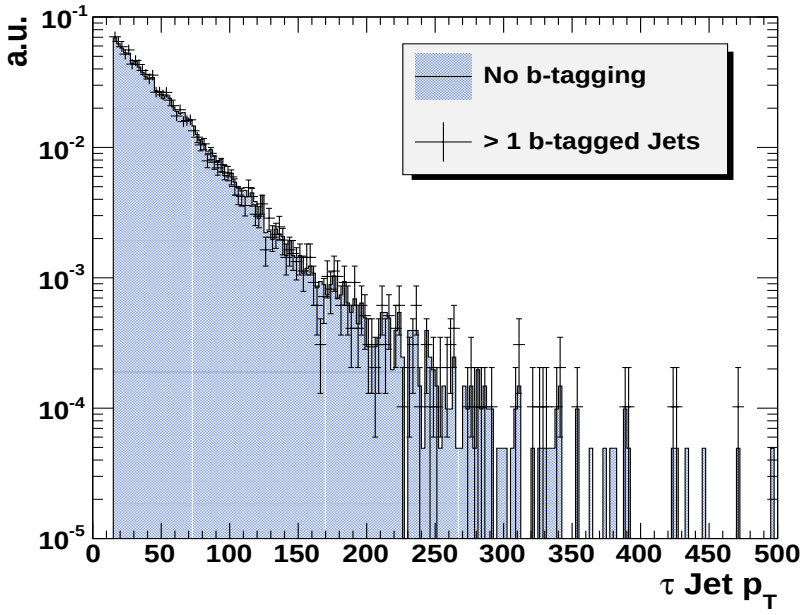

Fig. A.3. τ p_T.

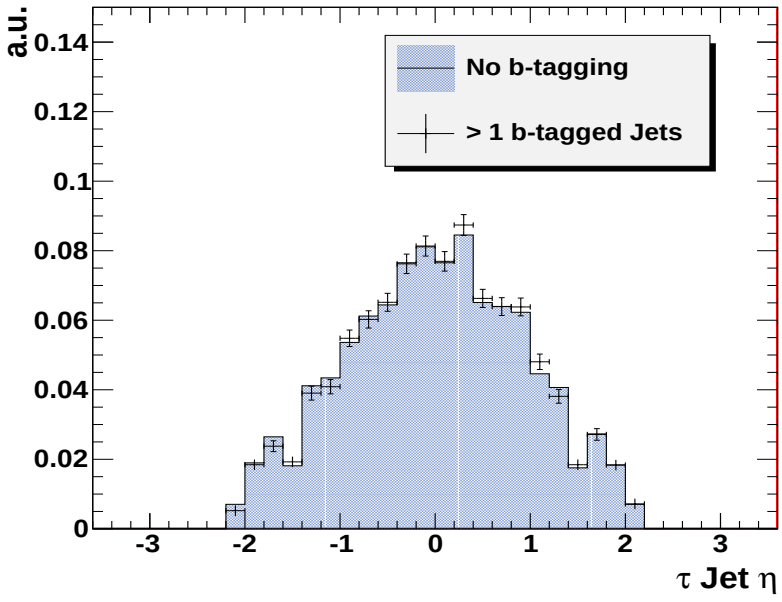

Fig. A.4. τ η.

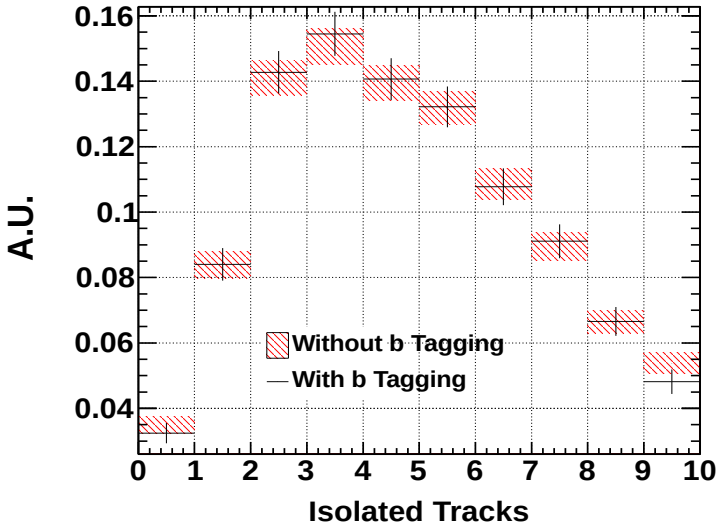

Fig. A.5. τ track isolation.

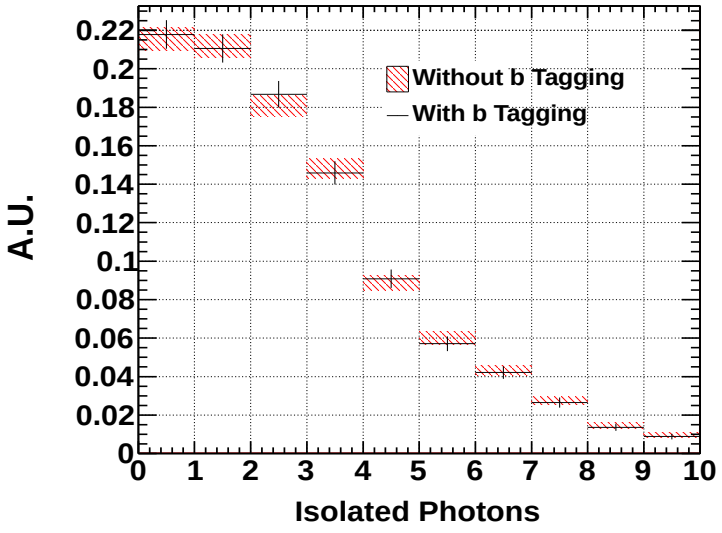

Fig. A.6. τ γ isolation.

APPENDIX B
VALIDATION OF THE B-TAGGING EFFICIENCY IN A HIGH MULTIPLICITY JET SAMPLE

Since the b-tagging efficiency as measured by [14] is used to estimate the signal contribution from $t\bar{t}$, it is necessary to validate the use of said measurement in the present analysis. To do this a semi-clean $t\bar{t}$ sample is selected using a selection for a high multiplicity of jets. Once the sample is created, the probability to tag > 1 jet as a b-jet as measured in this region can be compared to the one in [14]. Figure B.1 shows the number of jets tagged as b-jets using the TCHE "medium" working point. The selections used to obtain the $t\bar{t}$ sample for this study are similar to those used in the calculation of the $t\bar{t}$ events contribution to the signal region 8.1:

- Begin with the signal selections
- Remove the isolation requirement on τ_h's
- Require ≥ 7 jets with $p_T > 30$ GeV/c and $|\eta| < 3$

The expected number of MC $t\bar{t}$ simulated events is 69.18 ± 1.59. The expected number of MC simulated W + Jets events is 15.12 ± 1.83. The number of observed events in data is 76. Therefore, the probability to tag > 1 jets as a b-jet can be measured as follows:

$$P(2 \text{ b-jets}) = \frac{N^{control}_{>1 \text{ b-jet}} - N^{W+jets}_{>1 \text{ b-jet}}}{N^{control} - N^{W+jets}} \tag{B.1}$$

The measured probability to tag > 1 jet as a b-jet is $40.15 \pm (8.17)_{stat} \pm (1.86)_{stat}$. This is to be compared to the value obtained if the b-tagging efficiency measured in [14] is used: 45.83 ± 7.76. This validates that the use of the b-tagging efficiency, as measured in [14] is valid for this analysis.

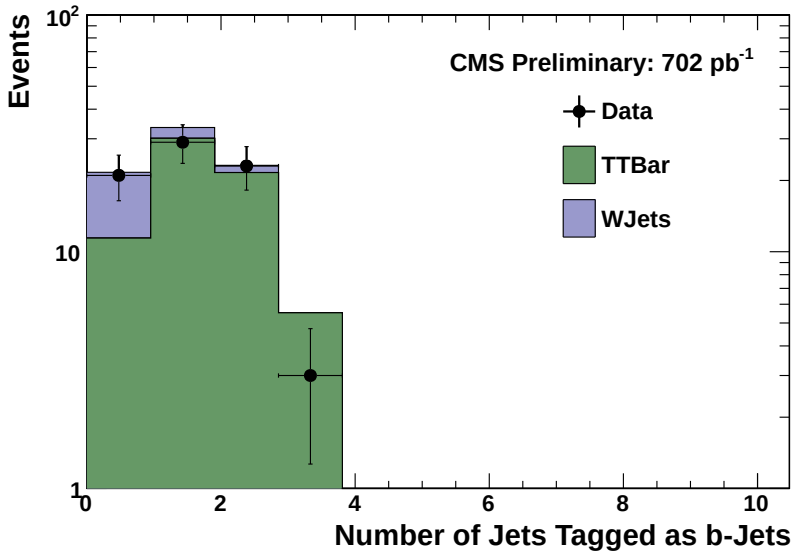

Fig. B.1. Number of jets tagged as b-jets using the track counting high efficiency "medium" working point.