UNIVERSITY OF CALIFORNIA, SAN DIEGO

A search for new physics using the M_{T2} variable in all-hadronic final states produced in proton-proton collisions with a center of mass energy of 13 TeV

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Physics

by

Mark Derdzinski

Committee in charge:

        Professor Frank Würthwein, Chair
        Professor Avraham Yagil
        Professor Benjamin Grinstein
        Professor Farhat Beg
        Professor Ian Galton

2018

The dissertation of Mark Derdzinski is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Chair

University of California, San Diego

2018

DEDICATION

FIXME: Dedication here.

EPIGRAPH

For knowledge comes slowly, and when it comes, it is often at great personal expense.

— Paul Auster, *Ghosts*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

FIXME: Acknowledgements

2011	B. A. in Physics and Mathematics, University of California, Berkeley
2014	M. S. in Physics, University of California, San Diego
2018	Ph. D. in Physics, University of California, San Diego

PUBLICATIONS

Search for new physics with the M_{T2} variable in all-jets final states produced in pp collisions at $\sqrt{s} = 13$ TeV, *CMS Collaboration*, J. High Energ. Phys. **10** (2016) 006, arXiv:1603.04053 [hep-ex]

Search for new physics in the one soft lepton final state using 2015 data at $\sqrt{s} = 13$ TeV, *CMS Collaboration*, Physics Analysis Summary (2016), CMS-PAS-SUS-16-011, cds.cern.ch/record/2161097

Search for new phenomena with the M_{T2} variable in the all-hadronic final state produced in proton-proton collisions at $\sqrt{s} = 13$ TeV, *CMS Collaboration*, Eur. Phys. J. C **77** (2017) no. 10, 710, arXiv:1705.04650 [hep-ex]

ABSTRACT OF THE DISSERTATION

A search for new physics using the M_{T2} variable in all-hadronic final states produced in proton-proton collisions with a center of mass energy of 13 TeV

by

Mark Derdzinski

Doctor of Philosophy in Physics

University of California, San Diego, 2018

Professor Frank Würthwein, Chair

A search for physics beyond the Standard Model (SM) is performed in events with final states including hadronic activity, missing energy, and significant momentum imbalance as measured with the M_{T2} variable. The results are based on data collected by the Compact Muon Solenoid detector at the Large Hadron Collider, and correspond to a total integrated luminosity of $35.9\,\mathrm{fb}^{-1}$ of proton-proton collisions at a center-of-mass energy of $13\,\mathrm{TeV}$. No significant excess above the predicted SM background is observed. The results are interpreted as 95% confidence-level exclusion limits on the masses of hypothesized

particles in a variety of simplified models of R-parity conserving supersymmetry (SUSY). Additional techniques for extending the search into final states with low-momentum leptons to target natural SUSY models in ewkino decays is discussed, and compared to the exclusion strength of other variations targeting such models.

Chapter 1

Introduction

1.1 The Standard Model of Particle Physics

1.2 Issues with the Standard Model

1.3 Beyond Standard Model Physics

Chapter 2

The Large Hadron Collider and the

CMS Detector

2.1 The Large Hadron Collider

FIXME: add paragraphs on LHC

2.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is a general-purpose physics detector at the LHC, situated at one of the five collision points along the main ring. The detector encapsulates the collision point with layers of various subsystems designed to interact with the outgoing particles of the proton-proton collisions, and measure the position and energies of the collision products. Because of the extremely high rate of interactions at the collision

point (on the order of one billion interactions per second), saving data from every bunch crossing would be unsustainable, and so the detector is also equipped with a system of hardware and software implemented "triggers" which identify events of interest for physics analyses to be saved to disk for further analysis.

The physical construction of the detector is motivated by the different interaction of particles with different types of materials, and consists of several subsystems layered as coaxial cylinders around the interaction point. Each subsystem consists of (sometimes different) components covering the fiducial area coaxial with the beamline (the "barrel") and also the ends of the cylinder (the "endcaps"). The innermost subsystem of CMS is the silicon tracker, which consists of many layered silicon pixels designed to pinpoint the locations of charged particles while minimally interacting with the particle's trajectory. The layer beyond the tracker is the electromagnetic calorimeter (ECAL), a grid of lead-tungstate crystals which scintillate to measure the energies of electromagnetic particles. Beyond the ECAL is the hadronic calorimeter (HCAL), a sampling calorimeter designed to measure the energies of hadronic particles (which deposit minimal energy in the ECAL). The final, outer layer is the CMS muon detector, where the muon detection stations are interweaved with the magnetic return yoke that generates the toroidal 3.8T magnetic field inside the detector volume. The total dimensions of the detector are 21.6m long and 14.6m in diameter, weighing over 12,500 tons.

2.2.1 Silicon Tracker

The silicon vertex tracker (SVT) is a series of silicon pixels and strips designed to measure the position of charged particles in the detector, while disturbing their path as little as possible. The position of particles in the interior is of particular importance in event reconstruction; charged particles traveling in a magnetic field will deflect in a curved path with radius proportional to the particles momentum as described in equation 2.1, and so the track reconstruct can be used to not only determine a particle's momentum with high precision, but also the sign of its charge based on the direction of curvature.

$$p = qrB \tag{2.1}$$

As the innermost detector subsystem, the SVT experiences the highest flux of particle radiation. In the barrel region, the tracker layers are oriented in 3 coaxial layers. Closest to the interaction point where particle flux is the greatest, very precise silicon pixels are used, measuring $100 \times 150 \mu\text{m}^2$, whereas in other layers the flux is low enough to use microstrip detectors, measuring $10\text{cm} \times 80\mu\text{m}$ and $25\text{cm} \times 180\mu\text{m}$ in the middle and outer layers respectively. In the endcaps, the pixel strips are arranged in a turbine-like pattern in two separate layers on each end. This configuration allows for the precise measurement of particle position for track reconstruction, while minimizing the amount of material which might deflect particles from their original trajectories. A partial geometry of the SVT layout can be seen in figure 2.1.
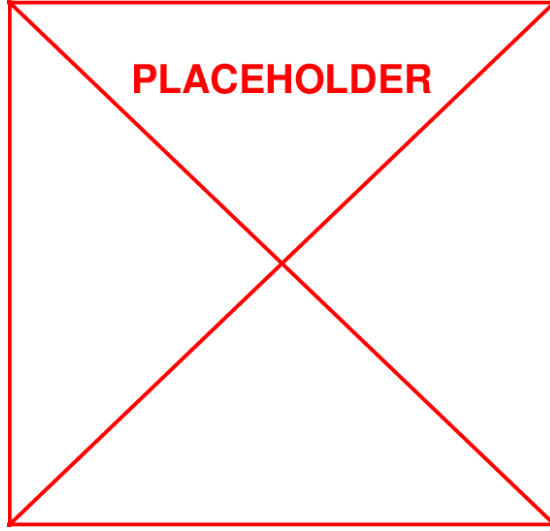
Figure 2.1: Geometry of silicon tracker inner layers in CMS.

2.2.2 Electromagnetic Calorimeter

The ECAL is used to measure the energies of particles which interact electromagnetically, both absorbing the incident particles and scintillating to provide an energy-readout to photodiodes attached to each crystal. Constructed of lead-tungstate ($PbWO_4$), electromagnetically interacting particles (such as electrons or photons) will interact with the crystal material, losing energy through a cascade of electromagnetic interactions including election-positron pair production and bremsstrahlung, as pictured in figure 2.2. This phenomenon — also referred to as "showering" — causes the crystals to scintillate proportional to the energy deposited in the crystal, which is then measured by various photodiodes to extract an accurate measurement of the particle energy, now fully absorbed by the calorimeter.

The fundamental principle of the calorimeter measurement relies on the energy loss
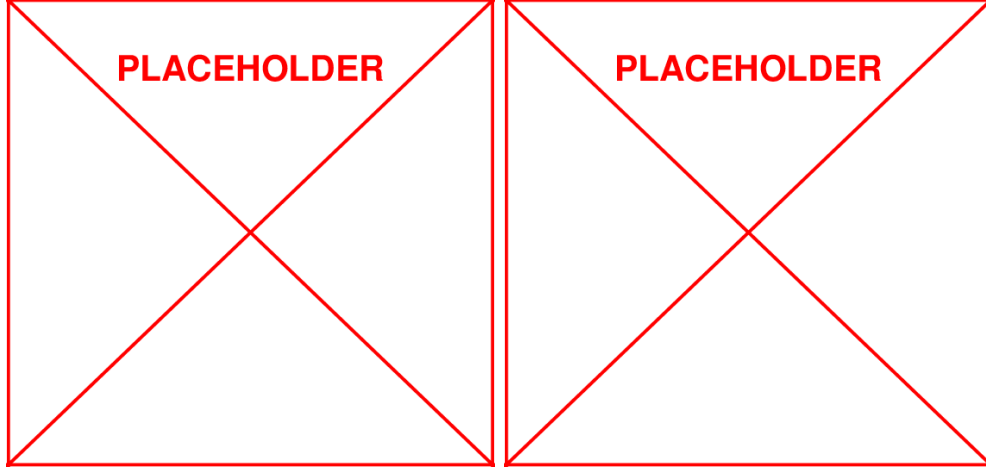
Figure 2.2: Feynman diagrams depicting the main processes by which particles shower in the ECAL. The left diagram depicts electron-positron pair production from a photon, and the right diagram depicts bremsstrahlung, where an electron radiates energy away through a photon.

of particles interacting with matter. In general, the energy of a particle traveling a distance X through some material is given by equation 2.2, where E_0 is the initial energy of the particle and X_0 is the material-dependent radiation length.

$$E(x) = E_0 e^{\frac{-x}{X_0}} \tag{2.2}$$

The design of the calorimeter is motivated by the choice of a scintillating, radiation-hard material with short X_0 such that incident electromagnetic particles deposit all their energy and are stopped by the ECAL. The resolution of the energy measurement is also dependent on the "stochastic term", which parametrizes the uncertainty due to statistical and measurement fluctuations in the calorimeter, and is given by equation 2.3, where S is
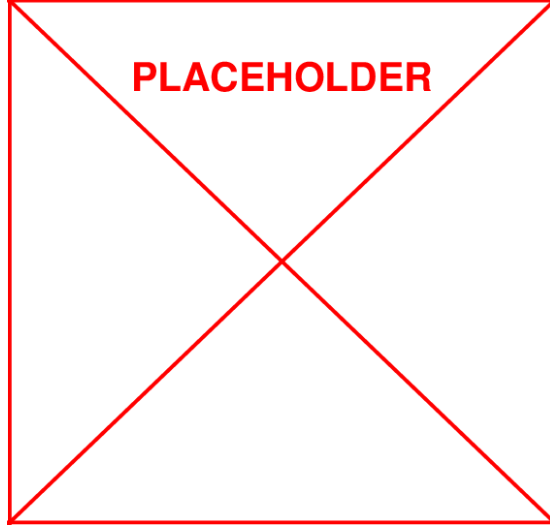
Figure 2.3: Energy resolution σ/E of the ECAL as a function of electron energy measured using a test beam. The energy was measured in a 3×3 crystal array with electrons incident on the center crystal, with electrons falling in a $4 \times 4\text{mm}^2$ region (lower points) and $20 \times 20\text{mm}^2$ region (upper points).

the stochastic term, N the noise, and C the constant term.

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{2.3}$$

The energy resolution can be measured by a test beam of known energy, as shown in figure 2.3.

The construction of the calorimeter is also divided into two sections by the cylindrical geometry, the ECAL barrel section (EB) and ECAL endcap sections (EE). The EB consists of 61,200 crystals arranged into 36 "supermodules", each spanning half the barrel length, and uses silicon avalanche photodiodes (APDs) as photodetectors. The individual crystals are tilted slightly (3°) in an $\eta - \phi$ grid with respect to the nominal interaction
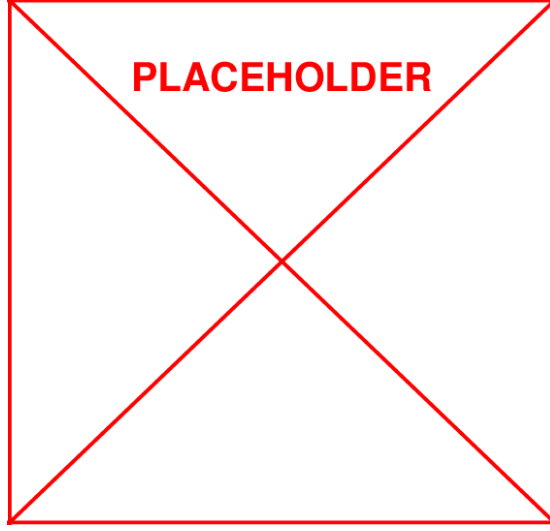
Figure 2.4: A cross section of the ECAL geometry, with the dashed lines marking the pseudorapidity values η covered by the various subsystems.

point, with a front-facing area of $22 \times 22\text{mm}^2$ and a length of 230mm. The EE instead uses vacuum phototriodes (VPTs) as photodetectors, and consists of approximately 15,000 crystals clustered in 5×5 units, also offset from the interaction point but arranged in an $x - y$ grid, with a cross section of $28.6 \times 28.6\text{mm}^2$ and a length of 220mm. The EE is also equipped with a "preshower" device placed in front of the crystal calorimeter, consisting of two strips of silicon strip detectors to enhance π^0 rejection. The layout of the ECAL can be seen in figure 2.4. Because of the depth of the ECAL crystals (which are $\sim 25 X_0$, and the confining properties of the crystals (which have a Moliere radius of 2.2cm, the radius of a cylinder containing 90% of a shower's energy on average), electrons and photons are typically well reconstructed in CMS, except in the transition region where EB and EE meet.

2.2.3   Hadronic Calorimeter

The CMS HCAL is a sampling calorimeter. Designed with alternating layers of scintillating and absorbing material, incident hadronic particles (such as charged pions, kaons, protons, etc.) interact with the absorber material and consequently shower into electromagnetic particles, whose energy can be read out by photodiodes connected to the scintillating material. Brass is used as the absorber material for both its interaction length and non-magnetic properties, and plastic scintillator tiles connected to embedded wavelength-shifting fibers carry the light to a readout system.

As with the ECAL, the energy loss of hadronic particles in the absorber is characterized by the (hadronic) interaction length and equation 2.2. However, unlike the ECAL, the HCAL contains both hadronic and electromagnetic showers. Electromagnetic particles generated in hadronic showers often fail to escape the absorber layers, and thus some electromagnetic energy is lost in the absorbers. The CMS HCAL is sometimes referred to as a *non-compensating calorimeter* because it is not constructed to actively compensate for the energy lost to these electromagnetic effects and the energy measurements must be corrected offline, known as "jet energy corrections" (JECs). JECs are typically calculated by examining data from collisions producing a boson recoiling against hadronic particles. By accurately measuring the boson energy in the ECAL, the sum of the recoiling hadronic energy in the HCAL is inferred (by momentum conservation) and compared to the detector response. Because the performance of the HCAL can fluctuate with time and run conditions, JECs are regularly recalculated and applied to the raw energy measurements
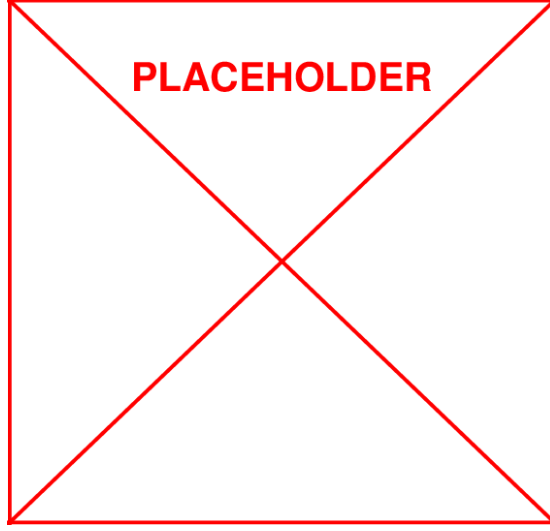
Figure 2.5: The jet transverse energy resolution as a function of the jet transverse energy, in different regions of pseudorapidity. For an explanation of jets, see section 2.3.5

taken by the HCAL to compensate for these effects. Additional information on the corrections contained in JECs is detailed in section 2.3.5. The energy resolution of the HCAL in different regions of pseudorapidity can be seen in figure 2.5.

The geometry of the HCAL can be reduced to four sections. The HCAL barrel (HB) consist of 32 "towers" of alternating absorber/scintillator material spanning the pseudorapidity region $-1.4 < \eta < 1.4$. The HCAL outer (HO) detector lies outside the vacuum tank of the magnetic coil and measures energy from any hadronic particles "leaking" through the HB, covering the pseudorapidity region $-1.26 < \eta < 1.26$. The HCAL endcap (HE) consists of 14 η towers spanning the pseudorapidity range $1.3 < |\eta| < 3.0$. Finally, the HCAL forward (HF) is a different steel/quartz fibre calorimeter spanning the very-forward $3.0 < |\eta| < 5.0$ region. Instead utilizing Cherenkov radiation generated in the quartz fibers, the HF preferentially samples neutral hadronic energy and is ideally designed

for the hadronic-heavy radiation environment in the forward region.

2.2.4 Muon Detectors

The muon detector is the only subsystem which is constructed outside of the toroidal magnetic field. Interleaved with the magnetic return yoke, different muon detectors are used to aid in the identification and reconstruction of muon tracks. Because muons typically penetrate every other layer of the detector and have a large bending radius, additional measurements in the muon system — combined with measurements in the SVT — can lead to vastly improved resolution of muon momentum. The non-uniform magnetic field in the muon detector region (beyond the toroidal regime) causes an s-shaped trajectory and tighter bending radius (than in the SVT) for the muons, which improves the resolution for particles with transverse momentum above $\sim 200 \mathrm{GeV}$ as seen in figure 2.6.

Different types of muon detectors are deployed (sometimes in combination) in different sections of the full muon system. In the barrel region $(-1.2 < \eta < 1.2)$ where the muon flux, neutral background, and magnetic field are small, drift tube (DT) chambers are used. In the endcaps where the the muon flux, neutral background, and magnetic field are much greater, cathode strip chambers (CSC) are used to increase coverage up to $|\eta| < 2.4$. In addition to these technologies, both the barrel and endcap systems are supplemented with resistive plate chambers (RPC) to provide complementary information to the DT and CSC detectors. The layout of the different muon detector components in the barrel and endcap can be seen in figure 2.7.

The DT detectors are chambers filled with gas surrounding a wire, with a voltage

11
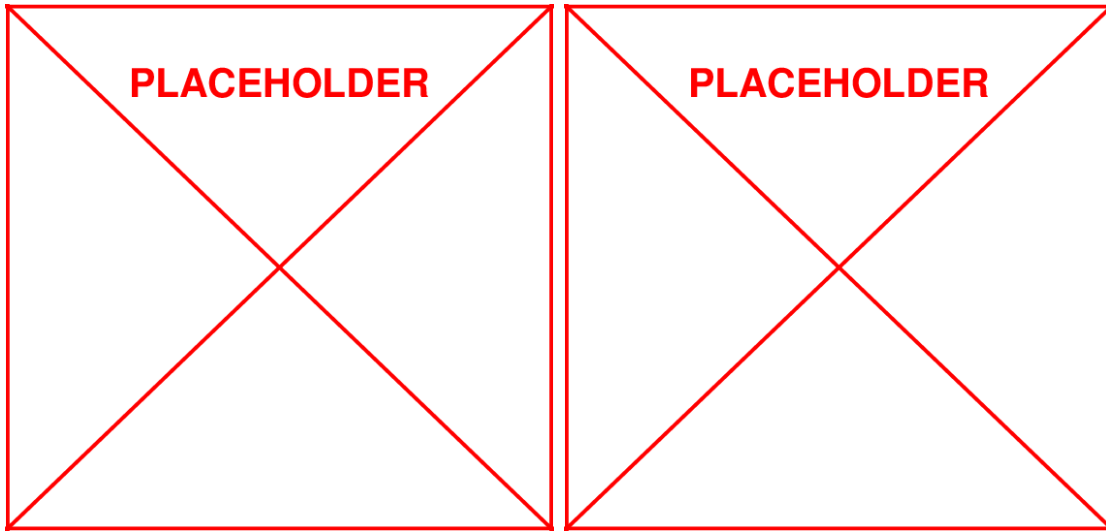
Figure 2.6: Muon momentum resolution as a function of muon momentum using only the inner tracking system, only the muon system, or both combined in the barrel region (left) or endcap region (right).
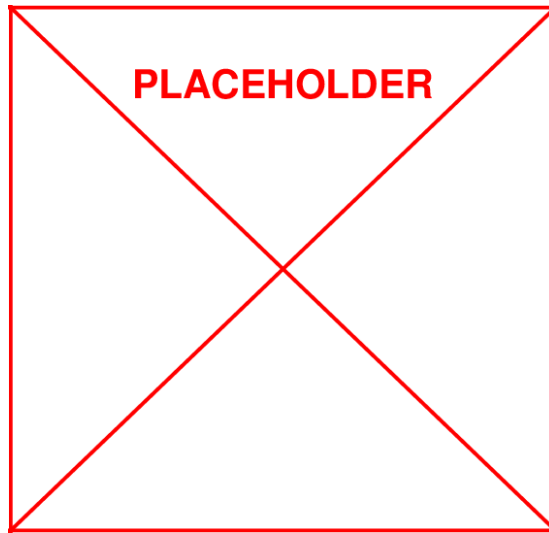
Figure 2.7: Layout of the CMS muon system for initial run configurations. The DT system is used only in the barrel region and CSCs in the endcap region, while RPCs are deployed in both the barrel and endcap.

difference between the wire and outside of the DT. When charged particles pass through the drift chamber, they ionize the gas, and the ionization products will drift across the voltage difference in the tube, resulting in a detectable voltage change in the DT. By measuring both the position along the DT wire where charge is deposited, as well as reconstructing the "drift time" it takes for the ionized particles to reach the wire, DTs provide a 2-dimensional measurement of a particle's position. In CMS, the DT chambers consist of a dozen layers arranged into 3 groups, each with up to 60 DTs. The layers are arranged in such a way that some measure the direction of the muon parallel to the proton beam and others along the perpendicular coordinate, such that the full muon trajectory can be reconstructed by using the DT station information.

FIXME: add paragraph on CSCs

FIXME: add paragraph on RPCs

2.2.5 Trigger Systems

When operating at design luminosity, the LHC can deliver proton bunches to the collision point at a rate of 40MHz (or 25ns between bunches), resulting in an average collision rate on the order of one billion collisions per second. In order to achieve reasonable rates of data collection for offline storage and processing, the detector must suppress the event rate by six orders of magnitude when selecting events of interest to be saved for physics analyses. This is accomplished through a combination of readout electronics and the trigger systems: the Level-1 trigger (L1) processors and online High-Level triggers (HLT).

The L1 trigger system is comprised of specialized hardware processors to rapidly pre-select events of interest based on the calorimeter and muon systems. Based on the beam crossing frequency, the L1 electronics have only a few microseconds to collect readout data from the front end electronics and execute logic to select events of interest, such that the total time allotted for L1 trigger calculations is $< 1\mu s$. During this time the bulk of detector data is held in a buffer while L1 trigger decisions are made based on data with reduced granularity and resolution rapidly collected from the calorimeter and muon systems, where triggers typically check for "trigger primitive" objects (such as photons, muons, electrons, etc.). Trigger primitives must meet certain momentum or energy thresholds, and L1 triggers may also check global data about the event such as the sum of transverse energy or the missing transverse energy (inferred from momentum conservation).

After an L1 trigger tags an event, high-resolution data is read out from buffers for additional data processing before reaching the HLT. Each event of $\sim 1.5 MB$ is transferred to front-end readout buffers which then pipes data to a processor containing the HLT code. HLT code is designed to discard events as soon as possible when making trigger decisions and only a subset of objects or partial reconstruction of events may occur before the final trigger decision is made, though HLT algorithms typically approach the quality of final reconstruction. The HLT reduces the L1 output rate to $\mathcal{O}(100\mathrm{Hz})$ for event storage and full reconstruction.

2.3   CMS Physics Objects

When physics events are fully reconstructed, detector data is used to identify *physics objects* representing real particles and event quantities for use in a physics analysis. The physics objects in an event — such as leptons, jets, or missing energy — and their properties are used to select events of interest for physics analyses targeting different final states. The properties of physics objects and global event data are also be used to make analysis level decisions of the quality of different objects. Here we describe some of the physics objects referred to in the M_{T2} analysis and how they are reconstructed, as well as some global event properties and quality variables used as discriminants for physics objects and events.

2.3.1   Particle Flow

Most of the physics objects described in the following sections are reconstructed and identified in CMS using the particle flow (PF) algorithm. The PF algorithm is a holistic, iterative algorithm which uses all the available data in the detector to classify "PF candidate" particles in an event. PF works iteratively by identifying tracks and calorimeter deposits into a PF candidate, removing all energy and hits associated with the candidate and repeating the algorithm until all the detector information has been associated to PF objects. First any muon tracks in the inner tracker associated with muon system hits are associated and removed. remaining tracks are extrapolated into the calorimeters, and any energy deposits on the path are associated with the track and removed from further consideration. Once all the tracks have been associated, the remaining energy clusters can
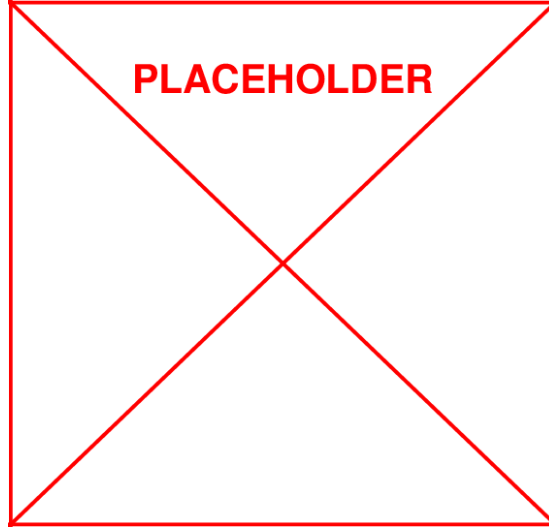
Figure 2.8: A graphic depiction of different particles leaving various signatures in the different CMS detector subsystems. Particles may be detected via tracks hits, energy deposits, or a combination of both.

be identified with photons and neutral hadrons (depending on their presence in the ECAL or HCAL, respectively). An example of the different tracks and energy deposits associated with various particles can be seen in figure 2.8.

2.3.2 Isolation

Isolation is an important distinguishing variable for physics objects measured in the detector. Simply put, isolation measures the total amount of energy in some proximity to a physics object. Because many physics processes results in particle production outside the primary interaction vertex (such as showering, decays, hadronization, pair production, etc.), the isolation of a particle parametrizes the amount of other "activity" surrounding the physics object, and consequently proves a useful discriminant when determining if

particles are produced in the primary interaction, or in subsequent physics processes. Particles which are produced in the primary physics process in a collision are sometimes referred to as *prompt*, and isolation is a powerful tool in identifying prompt leptons in particular.

2.3.3   Electrons, Muons, and Photons

The primary leptons considered in this analysis are electrons and muons (discussed in section 2.3.4), and photons are also used as a cross-check in some control regions for background estimates. Prompt electrons and photons are identified in CMS primarily through the use of the ECAL, where they are stopped and deposit all their energy. Prompt electrons are customarily distinguished by both the presence of a track originating at the primary vertex and a sufficiently low isolation.

Electrons in CMS are identified primarily through the existence of a charged particle track terminating in an energy cluster in the ECAL. As described in section 2.2.2, the ECAL crystals are ~ 25 radiation lengths deep, and will stop and contain nearly all of the energy of an electron (except perhaps miniscule losses to scattering in the tracker). The charged track leading to the energy deposit distinguishes the charged particle from a neutral particle, and the direction of curvature of the path in the tracker can be used to identify the charge of an electron (or positron).

Photons are distinguished from electrons by the lack of a track in the SVT. As a neutral particle, the only sign of a photon will be the energy deposit in the ECAL where the photon is stopped and deposits all its energy. Photons are typically distinguished

from other neutral electromagnetic particles (such as a π^0) by the shape of the shower in the ecal. Whereas photons typically shower by the processes in figure 2.2, other hadronic particles may cascade via different physics processes leading to measurable difference in the shape of the energy cluster.

2.3.4 Muons

Prompt muons produced in collisions at the LHC are generally minimum ionizing particles, and will penetrate the bulk of the detector. Without any matching calorimeter depisots, they are reconstructed using hits in the muon system matched to inner tracker hits. The muon system flags muon candidates, which are then matched to inner tracker hits for the best fit of the muon track. Given the track hits, the transverse momentum and energy of the muon can be calculated by a fit to the track parameters. As with electrons, muons are customarily distinguished by both the presence of a track originating at the primary vertex and a sufficiently low isolation.

2.3.5 Jets

Many particles produced at the LHC are created through the strong interaction, which may produce particles carrying color charge in the final state (i.e. quarks or gluons). Colored particles cannot exist individually due to the phenomenon known as *confinement*, and so via the process of *hadronization* (also refered to as *fragmentation*) they will proliferate in a series of interactions producing additional particle-antiparticle pairs, cascading in a parton shower to form hadronic bound states. When a "bare" quark or gluon is pro-

duced in the primary interaction, the ensuing hadronization results in a stream of tightly collimated hadrons aligned with the trajectory of the original bare particle, referred to as a *jet.*

Jets present themselves in CMS as collimated energy deposits in both the ECAL and HCAL, as well as tracks originating at the primary vertex. Between the two calorimeters, all the electromagnetic and hadronic components of the jet are stopped and measured by the calorimeters, and any leptons (including muons detected in the muon system) which originate in the cone of the jet are assigned to the total jet energy. The jets are reconstructed using the "anti-k_T" algorithm, which clusters jet activity in cones of a fixed radius. The jet is treated as a single physics object with a total energy equal to the sum of its electromagnetic, hadronic, and leptonic components, which by momentum conservation dictates the energy of the prompt quark or gluon that was produced in the primary interaction.

Furthermore, the substructure of jets and their content can be analyzed to determine the flavor of the parent parton, known as "tagging". Jets are typically tagged to distinguish between those originating from heavy-flavor quarks (bottom or charmed) and light-flavor quarks. Top quarks produced in primary interactions are unique because of their short lifetime. A top quark will decay before hadronization occurs, instead producing a W boson and down-type (down, bottom, or strange) quark which will subsequently hadronize. Searches involving top quarks in the final state typically employ "top-taggers" to search for these signatures of the top quark.

In practice, jets are complicated objects consisting of many constituent particles,

and must be clustered and calibrated so that the jet energy closely matches that of the parent parton which produced the jet. JECs are applied to raw jet energies to compensate for different experimental deviations from the parent parton energy. In particular, JECs include corrections to compensate for pile-up energy, detector effects (as a function of η), energy scale as a fundtion of jet p_{T}, and residual corrections to account for differences between data and simulation. The JECs are calculated by collecting data events with a Z boson (decaying to electrons or muons) or photon recoiling against a jet. By precisely measuring the bosons energy with the ECAL, scale factors are derived for jets as a function of pseudorapidity η and momentum p_{T}.

2.3.6   Missing Energy

Missing energy refers to the sum of all energy which has escaped the detector, and is inferred from a momentum-imbalance in the physics objects measured by the detector. Because the momentum of the incident protons is zero perpendicular to the direction of the beam, the momentum of physics objects produced in any collision must sum to zero in the direction transverse to the beamline. The missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$), is calculated by taking the negative vector sum of all PF candidates as in equation 2.4. $E_{\mathrm{T}}^{\mathrm{miss}}$ is of particular importance to analyses targeting BSM physics, as BSM signatures characteristically contain invisible particles in the final state which escape detection.

$$\vec{E}_T^{miss} = -\sum_i \vec{p}_T^{\,i} \tag{2.4}$$

The missing energy in an event is inferred from all other measured quantities in an

event, and there are many sources of $E_{\mathrm{T}}^{\mathrm{miss}}$ which are unphysical in nature but rather dependent on experimental effects. Particles from the primary interaction with a sufficiently large pseudorapidity may escape the fiducial region of the detector subsystems, and resolution effects or intrinsic noise in the detector may lead to fluctuations in measured energies. These experimental effects must be suppressed or distinguished from "real" $E_{\mathrm{T}}^{\mathrm{miss}}$ due to physics processes creating particles which escape the detector (e.g. neutrinos). Analyses sensitive to final states with $E_{\mathrm{T}}^{\mathrm{miss}}$ often employ robust data-driven methods to predict or suppress backgrounds which might generate experimental sources of $E_{\mathrm{T}}^{\mathrm{miss}}$ or physics processes which can contribute real $E_{\mathrm{T}}^{\mathrm{miss}}$ (such as $Z \to \nu\bar{\nu}$).

Chapter 3

The M_{T2} Variable Search

3.1 Analysis Strategy

Searches for new physics targeting all-hadronic final states present unique challenges and opportunities at the LHC. While such searches typically implement stringent vetoes on lepton candidates and thus bypass the need to correctly identify "real" leptons, the high rate of QCD processes in proton-proton collisions generates large amounts of SM events with all-hadronic final states. Designing a search targeting signatures with all-hadronic final states requires a mechanism to distinguish and suppress the selection of multi-jet QCD events from new physics signatures, as well as robust background estimation methods to predict the yield of standard model events which may generate missing energy (such as $Z \to \nu\bar{\nu}$).

The M_{T2} analysis harnesses the discriminating power of the M_{T2} variable to distinguish standard model events from possible signatures including new physics. By first

requiring a significant amount of missing energy in the event, multi-jet QCD processes are greatly suppressed. Additional requirements on the topology of the event implemented using M_{T2} further suppress QCD-like processes and favor events with real missing energy anti-aligned with the hadronic energy deposits in the detector. After estimating the minimal QCD contribution remaining by extrapolating from a region orthogonal to the signal selection, the only remaining backgrounds are leptonic events where the lepton failed reconstruction or identification (or "lost-lepton" events), and SM events creating real $E_{\text{T}}^{\text{miss}}$, in the form of neutrinos from a decaying Z boson recoiling against jets (or "invisible Z" events).

3.2   The M_{T2} Variable

M_{T2} is a particularly useful kinematic mass variable for final states where two particles decay (possibly in a chain) to a final state containing an invisible particle X of mass m_X. The typical transverse mass M_T, is defined in equation 3.1 for particles $i = 1, 2$, where the mass $m^{\text{vis}(i)}$, transverse momentum $\vec{p}_t^{\,\text{vis}(i)}$, and transverse energy $E_T^{\text{vis}(i)}$ characterize the visible kinematics of the decay chain, and $\vec{p}_t^{\,\text{X}(i)}$ and $E_T^{\text{X}(i)}$ characterize the unknown kinematics of the invisible particle X.

$$(M_T^{(i)})^2 = (m^{\text{vis}(i)})^2 + m_{\text{X}}^2 + 2\left(E_T^{\text{vis}(i)} \cdot E_T^{\text{X}(i)} - \vec{p}_t^{\,\text{vis}(i)} \cdot \vec{p}_t^{\,\text{X}(i)} \right) \tag{3.1}$$

In principle, if the correct values of m_{X} and $\vec{p}_t^{\,\text{X}(i)}$ were accessible, then the transverse mass would have a kinematic endpoint and not exceed the mass of the parent particles. However, the individual momenta of the invisible particles in the two decay chains cannot

be measured; the only quantity experimentally accessible is the total missing momentum $\vec{p}_T^{\text{miss}}$. With this in mind, the generalized transverse mass variable M_{T2} is defined in equation 3.2, where the unknown mass m_{X} is a free parameter and a minimization is performed over the sum of invisible momenta $\vec{p}_t^{\text{X}(i)}$ that satisfy the measured $\vec{p}_T^{\text{miss}}$ constraint.

$$M_{\text{T2}}(m_{\text{X}}) = \min_{\vec{p}_t^{\text{X}(1)} + \vec{p}_t^{\text{X}(2)} = \vec{p}_T^{\text{miss}}} \left[\max \left(M_T^{(1)}, M_T^{(2)} \right) \right] \tag{3.2}$$

Because this analysis selects final states with (often several) jets in the final state, the calculation of M_{T2} first requires grouping the hadronic jet activity into two large *pseudojets* to act as the visible components in the M_{T2} equation. The jet activity in each event is divided into two hemispheres, and the jets in each hemisphere are summed together to created the pseudojets. The hemisphere algorithm (as defined in **??**) proceeds as follows:

- The direction of the two jets with largest invariant mass is chosen as the initial seed for the two axes.

- Jets are associated to one of the two axes according to the minimal Lund distance, such that jet k is associated to hemisphere i instead of j if the condition in equation 3.3 is true.

- After each jet is associated to one of the two axes, the axes are recalculated by summing the momenta of all jets associated to an axis.

- The association algorithm iterates using the new axes, and continues until no jets are associated to a different axis after an iteration.

Table 3.1: Summary of physics objects and preselection for events.

$$(E_i - p_i \cos\theta_{ik})\frac{E_i}{(E_i + E_k)^2} \leq (E_j - p_j \cos\theta_{jk})\frac{E_j}{(E_j + E_k)^2} \qquad (3.3)$$

QCD multijet events (when clustered using this pseudojet algorithm) may yield high M_{T2} events if the pseudojets have high jet masses, thus the visible masses $m^{\mathrm{vis}(i)}$ are set to zero to suppress such SM events. Since the kinetic components of M_{T2} will be large for signal events, this suppression does not isgnificantly impact sensitivity to many BSM signatures, thus M_{T2} is calculated in this analysis using only $E_{\mathrm{T}}^{\mathrm{miss}}$ and the two pseudojets.

3.3 Event Selection Criteria

The general strategy for the event selection is to first apply baseline cuts on mo-tivated by hardware and software-level triggers (discussed in section 2.2.5) and reducing the QCD multi-jet background to negligible levels. Events are further categorized using the scalar sum of the transverse momenta p_{T} of all selected jets (H_{T}), the total number of jets in the event (N_{jets}), the total number of b-tagged jets in the event ($N_{\mathrm{b\text{-}jets}}$), and M_{T2}. A summary of the event preselections can be found in table 3.1.
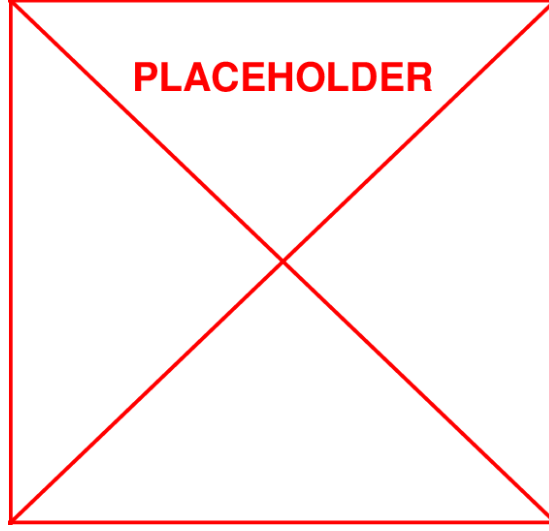
Figure 3.1: Topological regions in N_{jets} and $N_{\text{b-jets}}$ for the [575,1000] H_{T} region. Within each region, the relative fraction of background events from different SM processes is shown based on simulation.

3.4 Search Regions

The search regions are defined by categorizing events in bins of H_{T}, N_{jets}, $N_{\text{b-jets}}$, and M_{T2} (in addition to the baseline selection described in section 3.3). First events are categorized into "topological regions" according to H_{T}, N_{jets}, and $N_{\text{b-jets}}$:

- H_{T} (GeV): [250, 450] (Very Low), [450,575] (Low), [575,1000] (Medium), [1000, 1500] (High), [1500, ∞] (Extreme)

- N_{jets} & $N_{\text{b-jets}}$: 2-3j 0b, 2-3j 1b, 2-3j 2b, 4-6j 0b, 4-6j 1b, 4-6j 2b, $\geq$7j 0b, $\geq$7j 1b, $\geq$7j 2b, 2-6j $\geq$3b, and $\geq$7j $\geq$3b (except in the region with $250 < H_{\text{T}} < 450$ GeV, where bins $\geq$7j are merged with 4-6j bins due to lack of events).

The different topological regions for one H_{T} region and their background composition are

Table 3.2: M_{T2} binning in the Very Low, Low, and Medium H_{T} topological regions.

Table 3.3: M_{T2} binning in the High and Extreme H_{T} topological regions.

depicted in figure 3.1. Each topological region is further divided into bins based on M_{T2} . The M_{T2} binning is constructed such that the low edge of the first bin is 400 GeV in regions with $H_{\mathrm{T}} > 1500$ GeV and 200 GeV everywhere else, and the low edge of the final bin is constructed to contain approximately one background event based on simulation and not exceeding the maximum H_{T} value in that topological region (since an upper limit on H_{T} places an upper limit on M_{T2}). The detailed M_{T2} binning is as follows:

- Very Low H_{T}: [200,300], [300,400], [400,∞]

- Low H_{T}: [200,300], [300,400], [400,500], [500,∞]

- Medium H_{T}: [200,300], [300,400], [400,600], [600,800], [800,∞]

- High H_{T}: [200,400], [400,600], [600,800], [800, 1000], [1000, 1200], [1200,∞]

- Extreme H_{T}: [400,600], [600,800], [800,1000], [1000,1400], [1400,∞]

The various H_{T} bins and associated M_{T2} binning can be seen in figure 3.2, and the full breakdown of signal regions (including M_{T2} binning) is listed in tables 3.2 and 3.3 . In addition to multi-jet search regions, this anlysis also considers monojet events. Because
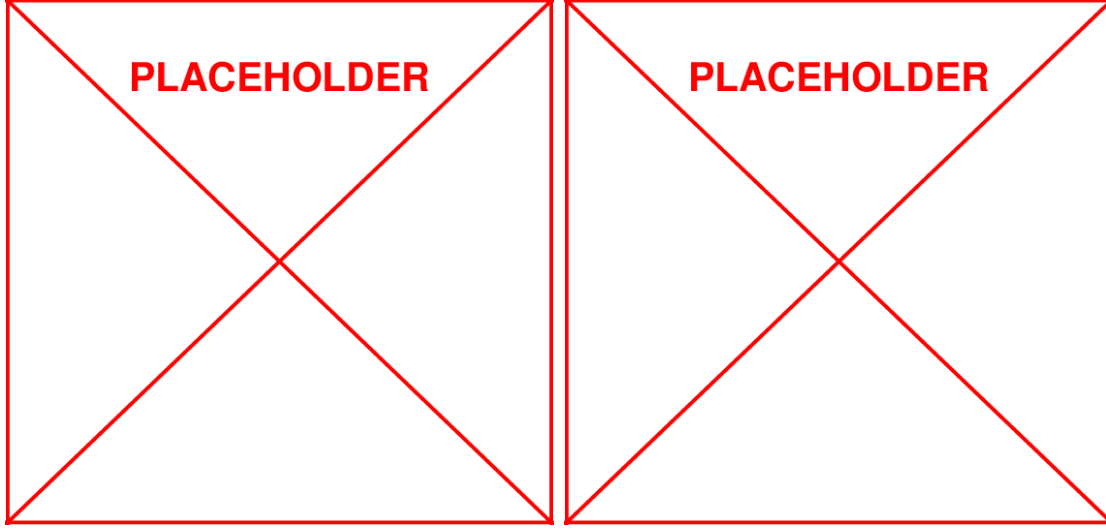
Figure 3.2: Signal region bins in H_T and E_T^{miss} (left) and M_{T2} binning within each H_T region (right). If simulation predicts less than one background event in the greatest M_{T2} bin within a region, it is merged with the previous bin.

there is only a single jet (and M_{T2} is ill-defined without multiple jets), binning in these regions is defined using $N_{\mathrm{b\text{-}jets}}$ and H_T as follows:

- $N_{\mathrm{b\text{-}jets}}$: 0b, $\geq$1b

- H_T: [250,350], [350,450], [450,575], [575,700], [700,1000], [1000,1200], [1200,∞]

As with the multi-jet regions, monojet H_T bins with less than one simulated background event in the final bin are merged with the penultimate bin.

In addition to these signal regions used to interpret results in the context of various BSM physics models, the analysis also provides results in "super signal regions" (SSRs) as defined in table 3.4. These regions provide a simpler set of selection than the nominal signal regions so that phenomenologists may easily reinterpret results in the context of

Table 3.4: Definition of "super signal regions" used in reinterpretations of the analysis.

different signal models (as in **FIXME: reference paper using reinterpretations**). Results obtained using the SSRs are not as sensitive as the nominal binning — finely binned regions have a higher signal-to-background ratio and the global background fit reduces the background uncertainties – but are much easier to use for reinterpretation than the many correlated bins of the nominal analysis.

3.5   Control Regions

In order to anchor the data-driven background estimates used in this analysis, *control regions* (CR) orthogonal to the signal region selection are defined for various processes. In particular, there are control regions corresponding to enriched samples of $\gamma + $ jet events, single lepton events, $Z \to l^+l^-$ events, and QCD multijet events.

3.5.1   Single Lepton Control Region

The single lepton CR is constructed to select a sample enriched with single lepton events, the most dominant contributions being from $t\bar{t}$ and W + jetsproduction. The same baseline selections described in section 3.3 are applied with the exception of the following:

- In lieu of the lepton veto, exactly one lepton candidate passing the reco or PF lepton selections is required. In order to avoid double counting (for leptons which

are reconstructed both as a reco lepton and PF candidate), PF leptons within ΔR < 0.1 of a reco lepton are not considered.

- The transverse mass M_{T2} between the lepton and E_T^{miss} must be less than $100\,\text{GeV}$ to reduce possible signal contamination.

Since non-isolated leptons in the fiducial region of the detector are usually successfully reconstructed, the closest jet within $\Delta R < 0.4$ of the lepton is removed and the lepton instead counted as a visible object for the purposes of computing H_T, H_T^{miss}, $\Delta\phi(\text{j}_{1234}, E_T^{\text{miss}})$, $|\vec{H}_T^{\text{miss}} - \vec{E}_T^{\text{miss}}|/E_T^{\text{miss}}$, and M_{T2} (as well as the hemispheres used to calculate M_{T2}). Events are further subdivided into the topological regions described in section 3.4 using the modified H_T and N_{jets} and $N_{\text{b-jets}}$, but not in M_{T2} to increase the statistical power of the CR. The signal regions with $\geq 7\text{j},\geq 1\text{b}$ are all predicted using CRs with identical H_T bins but $\geq 7\text{j},1\text{-}2\text{b}$ to increase the statistical power of those CRs (and to avoid signal contamination in regions with $\geq 7\text{j},\geq 3\text{b}$). In addition, for regions with $H_T > 1500\,\text{GeV}$, the minimum M_{T2} threshold is set to $200\,\text{GeV}$ to increase available statistics. The monojet CR is binned identically to the signal region.

3.5.2 Dilepton Control Region

Control regions corresponding to opposite-sign same-flavor leptons (OSSF) from $Z \to l^+l^-$ events are used to estimate the $Z \to \nu\bar{\nu}$ background, with corresponding sets of control regions requiring an opposite-sign opposite-flavor (OSOF) pair to estimate the flavor-symmetric background component in the former dilepton CR. The same baseline

30

selections as described in section 3.3 are applied with the exception of the follow:

- In lieu of the lepton veto, exactly 2 leptons (ee, $e\mu$, or $\mu\mu$) passing loose lepton selections are required.

- There is no requirement on E_T^{miss}. Instead, the dilepton system must have a transverse momentum $p_T(\ell\ell) > 200\,\text{GeV}$ to mimic the kinematics of the $Z \to \nu\bar{\nu}$ background and suppress the $t\bar{t}$ contribution.

- Without a missing energy requirement, events are selected in data using leptonic trigger paths. Dimuon events are selected using a combination of dimuon and high-p_T single muon triggers, dielectron events using a combination of dielectron and high-p_T single photon paths (which do not require isolation and recover efficiency for high-p_T electrons or those highly co-linear high-p_T Z bosons), and opposite-flavor events are selected using a combination of $e\mu$ triggers and higher threshold single-lepton paths (to again recover efficiency for some events).

- To improve trigger efficiency for these regions, the leading lepton is required to have a minimum momentum $p_T > 100\,\text{GeV}$ and the sub-leading lepton $p_T > 30\,\text{GeV}$.

- When selecting Z boson candidates for the $Z \to \nu\bar{\nu}$ estimate, the leptons are also required to be OSSF with an invariant mass $|m_{\ell\ell} - m_Z| < 20\,\text{GeV}$, where m_Z is the nominal Z boson mass.

Similar to the single lepton control region, the closest jet within $\Delta R < 0.4$ of each lepton is removed and the leptons added to the E_T^{miss} vector for the purposes of computing

N_{jets}, $N_{\text{b-jets}}$, H_{T}, $H_{\text{T}}^{\text{miss}}$, $\Delta\phi(\text{j}_{1234}, E_{\text{T}}^{\text{miss}})$, $|\vec{H}_{\text{T}}^{\text{miss}} - \vec{E}_{\text{T}}^{\text{miss}}|/E_{\text{T}}^{\text{miss}}$, and M_{T2} (as well as the hemispheres used to calculate M_{T2}). Additional information on the M_{T2} binning for these regions is detailed in section 4.4.

3.5.3   Multijet Control Region

The multijet control region is designed to select a sample enriched in QCD events to estimate the multijet background. The same baseline selections described in section 3.3 are applied with the exception of the $\Delta\phi(\text{j}_{1234}, E_{\text{T}}^{\text{miss}})$ requirement, which is inverted to select a sample dominated by QCD events with large fake $E_{\text{T}}^{\text{miss}}$ due to jet energy mismeasurements.

The transfer factor which is used to extrapolate the control region yield to the signal region is measured in a separate QCD-dominated region with $M_{\text{T2}} < 200\,\text{GeV}$, described in detail in section 4.2. Because of the lower $E_{\text{T}}^{\text{miss}}$ requirement, different trigger paths must be used. for regions with $H_{\text{T}} > 1000\,\text{GeV}$ trigger paths seeded by a single high-p_{T} jet are used. For other H_{T} regions, similar trigger paths with lower jet p_{T} thresholds are used, but due to the rate of QCD events creating low-p_{T} jets these paths are prescaled to suppress the data acquisition rate.

Chapter 4

Background Estimates

4.1   Types of Backgrounds

The M_{T2} all-hadronic search is primarily a search for missing energy, and the backgrounds present in the signal region can be roughly categorized into two groups: events with "fake" missing energy due to mismeasurement of energies or other experimental effects, and those with "real" missing energy due to the underlying physics process which escapes the detector. In this analysis, the primary backgrounds from SM processes with $E_{\mathrm{T}}^{\mathrm{miss}}$ and hadronic activity in the final state can be divided into three categories:

- *Multijet events*, with no true missing energy generating fake $E_{\mathrm{T}}^{\mathrm{miss}}$. Such all-hadronic events usually enter analysis signal regions due to mismeasurement of jet energy or sporadic noise.

- *Lost lepton events*, with a lepton in the final state that is either not identified, reconstructed, in the fiducial region of the detector, or sufficiently isolated to pass the lepton selections. The bulk of these events are due to leptonic W boson or top quark decays (which may also produce significant real E_T^{miss} with an associated neutrino) recoiling against jets.

- *Invisible Z events*, where a Z boson produced in association with jets decays into neutrinos. The neutrinos are extremely weakly interacting particles which escape the detector without leaving any signature, and so this background is fundamentally very similar to many of the BSM signals which may appear in each search region. It is a dominant background in many signal regions, though the relative fraction decreases in regions with more b-tagged jets ($N_{\text{b-jets}}$).

In order to provide a robust prediction of each background that is not heavily dependent on simulation alone, each background is estimated using data-driven techniques as described in the following sections. The control regions for each background can be found in chapter 3.

4.2   Multijet Estimate

The estimate of the SM multijet background utilizes two different techniques depending on the number of jets in a given signal region. For regions with two or more jets, the background is estimated by extrapolating to regions with high $\Delta\phi_{\min}$ after inverting the $\Delta\phi_{\min}$ requirement. In the monojet signal regions, a sample enriched in unbalanced

dijet events is used to extrapolate to regions with low sub-leading jet momentum.

4.2.1 Multijet Signal Region Prediction

As outlined in section 3.5.3, the multijet background in control regions with two or more jets is estimated using a QCD-enriched sample where H_{T} triggers are used to select events with an inverted $\Delta\phi_{\mathrm{min}}$ cut. The ratio of events with high to low $\Delta\phi_{\mathrm{min}}$ (r_ϕ) is modeled as a power law function in M_{T2}, as shown in equation 4.1.

$$r_\phi(M_{\mathrm{T2}}) = \frac{N(\Delta\phi_{\mathrm{min}} > 0.3)}{N(\Delta\phi_{\mathrm{min}} < 0.3)} = a \cdot M_{\mathrm{T2}}^b \tag{4.1}$$

The power-law dependence of r_ϕ on M_{T2} is verified in simulation for events with $M_{\mathrm{T2}} > 60\,\mathrm{GeV}$,, as illustrated in figure 4.1. Because the dominant source of $E_{\mathrm{T}}^{\mathrm{miss}}$ in low M_{T2} events is not necessarily due to jet mismeasurement, the fit is performed in the window $60 < M_{\mathrm{T2}} < 100\,\mathrm{GeV}$ except in extreme H_{T} regions where a lower bound of $70\,\mathrm{GeV}$ is used (as a conservative measure since these regions have high statistics). The upper bound of 100 GeV is chosen such that the contamination of electroweak and top-quark events is small compared to the QCD multijet yield. A systematic uncertainty in the r_ϕ value is assigned based on variations of the fit window, by shifting the lower boundary of the window in either direction while preserving the fit statistics by shifting the upper boundary of the window. The systematic uncertainty is then taken as the maximal deviation of all such variation with respect to the nominal window.

Due to the the total integrated luminosity available and the deliberate suppression rate with which some H_{T} triggers save events (known as the trigger *prescale*), the statistics
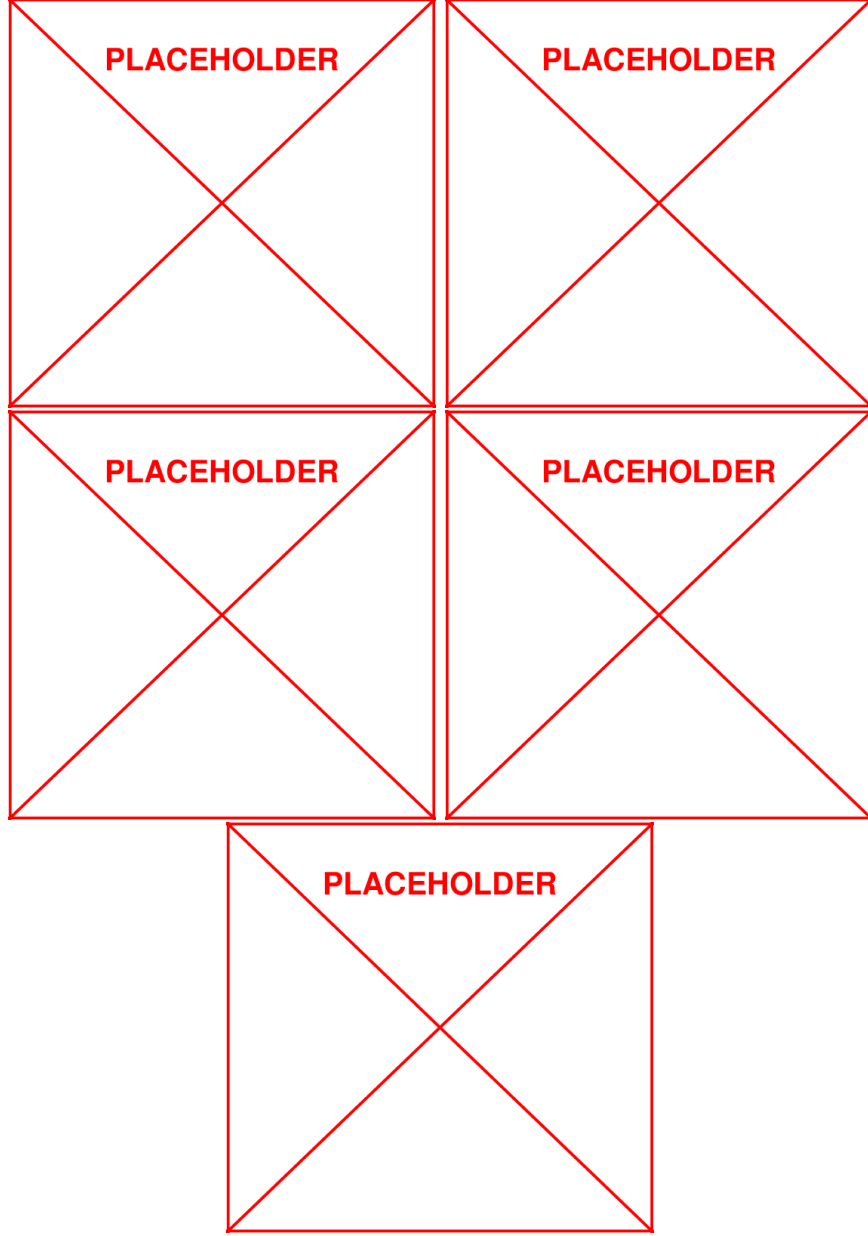
Figure 4.1: Simulated distribution of the r_ϕ ratio as a function of M_{T2} for the low (top left), medium (top right), high (medium left), extreme (medium right), and very low (bottom) H_T signal regions. Solid points indicate the total background based on simulation which the hollow points show the QCD multijet contribution only. The red line (and grey error band) illustrates the best fit (and associated uncertainty) to a power law function performed in the fit window, which is indicated by dashed lines.

Table 4.1: Relative uncertainty of f_j and r_b associated with the assumed invariance with respect to M_{T2} and $\Delta\phi$ (and H_{T} for r_b).

are sufficient to perform fits in H_{T} regions inclusive in N_{jets} and N_{jets}. The inclusive multijet background estimate can be determined using r_ϕ as a function of M_{T2} $N_{\mathrm{inc}}^{SR} = r_\phi(M_{\mathrm{T2}}) \cdot N_{\mathrm{inc}}^{CR}(M_{\mathrm{T2}})$. The final estimate in a given N_{jets}-$N_{\mathrm{b\text{-}jets}}$ bin can be determined as in equation 4.2, where f_j is the fraction of multijet events in bin N_{jets} in a given H_{T} bin, and r_b is the ratio of events with $N_{\mathrm{b\text{-}jets}}$ b-tags in a given N_{jets} bin.

$$N_{j,b}^{SR}(M_{\mathrm{T2}}) = r_\phi(M_{\mathrm{T2}}) \cdot N_{\mathrm{inc}}^{CR}(M_{\mathrm{T2}}) \cdot f_j(H_{\mathrm{T}}) \cdot r_b(N_{\mathrm{jets}}) \tag{4.2}$$

The values of f_j and r_b are measured in data using QCD-enriched regions with an inverted $\Delta\phi$ requirement and $100 < M_{\mathrm{T2}} < 200$. Based on simulation, f_j and r_b do not significantly depend on M_{T2} , and r_b is independent of H_{T}. Since f_j is dependent on H_{T} and r_b on N_{jets}, different values of f_j are measured in each inclusive H_{T} region, and different values of r_b are measured in in inclusive N_{jets} regions. Figures 4.2 and 4.3 illustrates the data-driven values of f_j and r_b respectively, along with a comparison to simulation. The robustness of invariance with respect to M_{T2} and $\Delta\phi$ (and H_{T} for r_b) is calculated by making several variations of the aforementioned variables and measuring consequent variations in f_j and r_b , and is summarized in table 4.1.

Figure 4.2: The values of f_j as measured in data in different H_T regions, compared to simulation. The uncertainties include both the statistical error and the systematic sources as listed in table 4.1
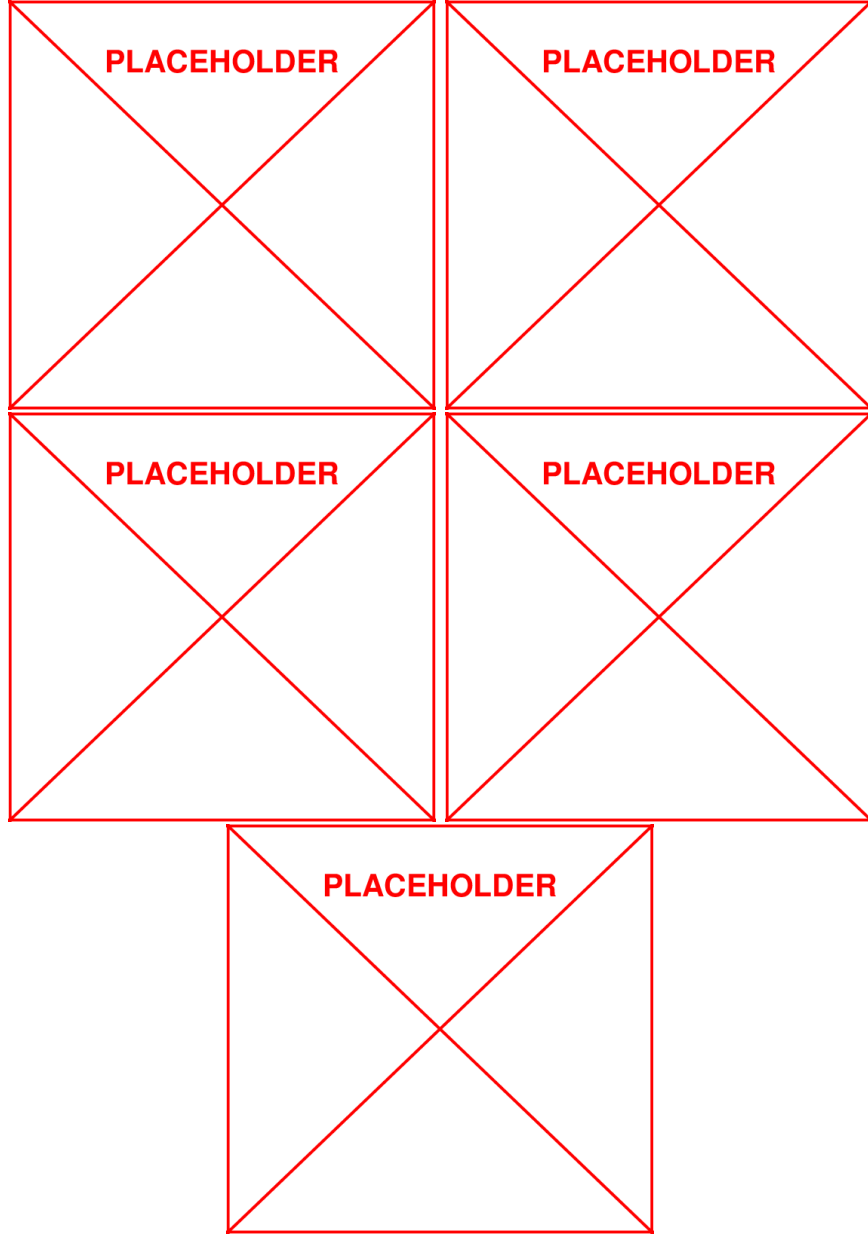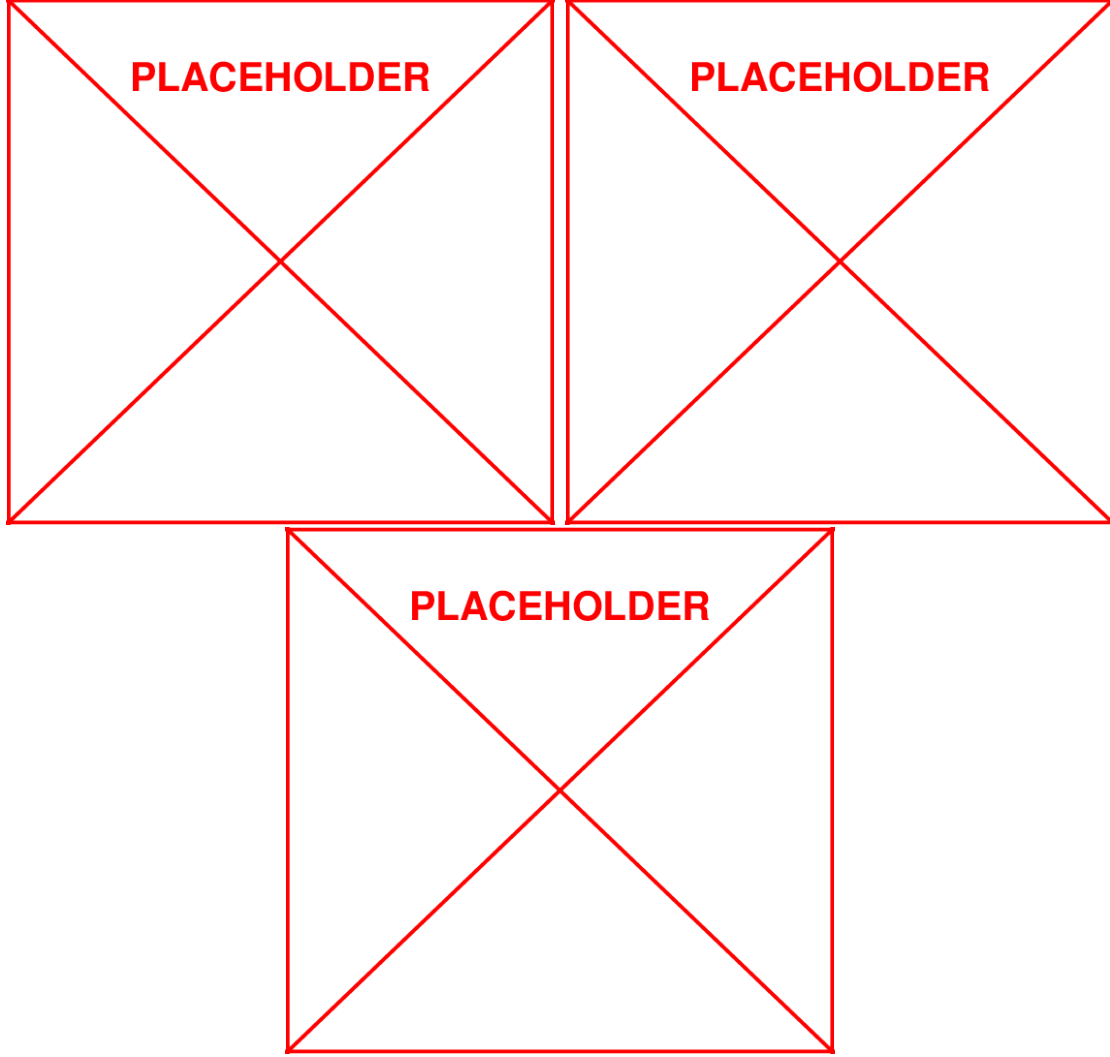
Figure 4.3: The values of r_b as measured in data in different N_{jets} regions, compared to simulation. The uncertainties include both the statistical error and the systematic sources as listed in table 4.1

4.2.2 Monojet Signal Region Prediction

The multijet background in control regions with a single jet cannot be estimated using the $\Delta\phi$ technique since $E_{\mathrm{T}}^{\mathrm{miss}}$ is usually very similar to the H_{T} in these events and typically anti-aligned with the jet. As outlined in section 3.5.3, a separate control region is devised which instead selects dijet events that are orthogonal to the multijet signal regions because of an inverted $\Delta\phi(\mathrm{j}_{1234}, E_{\mathrm{T}}^{\mathrm{miss}})$ cut (and orthogonal to the the monojet SR with the presence of multiple jets).

The sub-leading jet momentum for in this CR can be seen in figure 4.4. Because jets with p_{T} below $30\,\mathrm{GeV}$ are not considered, monojet events can be classified as those with $p_{\mathrm{T}}^{\mathrm{jet2}} < 30\,\mathrm{GeV}$, and the CR is used to extrapolate into the regime where $p_{\mathrm{T}}^{\mathrm{jet2}}$ is small. With decreasing $p_{\mathrm{T}}^{\mathrm{jet2}}$, events appear more imbalanced and approximate the topology of a true monojet event, as depicted in figure 4.5.

The predicted yield of multijet background in a monojet $p_{\mathrm{T}}^{\mathrm{jet1}}$ bin is determined according to equation 4.3, where f_{QCD} is the fraction of QCD events as measured in the region with $30 < p_{\mathrm{T}}^{\mathrm{jet2}} < 60\,\mathrm{GeV}$ and N_{data} is the yield in data of dijet events with $30 < p_{\mathrm{T}}^{\mathrm{jet2}} < 60\,\mathrm{GeV}$. Assuming $N_{\mathrm{data}}(0-30) < N_{\mathrm{data}}(30-60)$, this estimate provides an upper bound on the total multijet background contribution in each monojet CR. A systematic uncertainty of 50% on the QCD fraction f_{QCD} is assigned as a conservative estimate.

$$N_{\mathrm{QCD}}(p_{\mathrm{T}}^{\mathrm{jet1}}) = N_{\mathrm{data}}(30-60, p_{\mathrm{T}}^{\mathrm{jet1}}) \cdot f_{\mathrm{QCD}}(30-60, p_{\mathrm{T}}^{\mathrm{jet1}}) \tag{4.3}$$

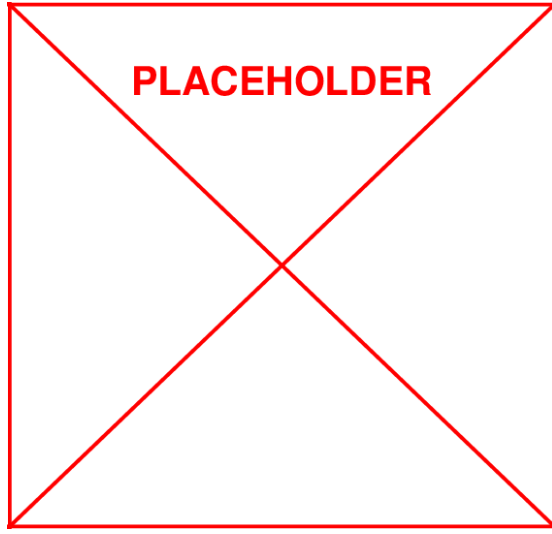**Figure 4.4**: The transverse momentum of the sub-leading jet for dijet events in the monojet QCD background control region. The total yield of the simulation is normalized to the overall yield in data.
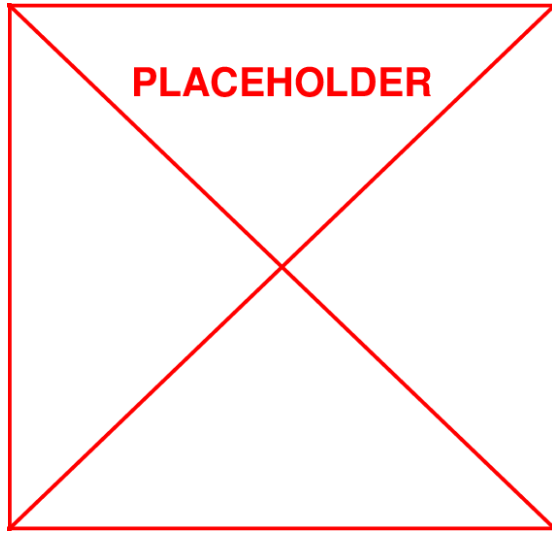
Figure 4.5: An illustration of "unbalanced" dijet events. As the momentum of the sub-leading jet decreases, $E_{\mathrm{T}}^{\mathrm{miss}}$ is more anti-aligned with the primary jet and approaches the topology of a monojet event.

4.3   Lost Lepton Estimate

The lost lepton background is predicted by taking the yield of single lepton events in similar kinematic regions, measuring the ratio of events where the lepton is lost to those where it is found, and extrapolating into the lost lepton regime. The control region for the lost lepton estimate is described in detail in section 3.5.1, and is comprised of events in data with the same signal triggers and preselection, with the exception of an inverted lepton veto and an additional requirement on lepton M_T (to reduce signal contamination).

4.3.1   Background Prediction

The estimate in a given signal region is obtained from a control region using transfer factors as described in equation 4.4, where N_{1L}^{CR} is the number of events in the corresponding CR, $R_{MC}^{0l/1l}$ is the ratio of zero-lepton to one-lepton events derived from simulation, and $k(M_{T2})$ is the transfer factor into bins of M_{T2} in a given topological region. The ratio $R_{MC}^{0l/1l}$ is measured in Monte Carlo after normalizing the yield to data in each CR and correcting for differences in lepton efficiency between data and simulation, and also factors in lepton acceptance, efficiency of reconstruction and identification, as well as contributions from W bosons decaying through τ leptons to a hadronic final state. The M_{T2} transfer factor $k(M_{T2})$ is taken from data where statistics permit, and uses information from simulation to project events into bins of M_{T2} in the low-statistics regime.

$$N_{LL}^{SR}(H_T, N_j, N_b, M_{T2}) = N_{1L}^{CR}(H_T, N_j, N_b, M_{T2}) \cdot R_{MC}^{0l/1l}(H_T, N_j, N_b, M_{T2}) \cdot k(M_{T2}) \quad (4.4)$$

To reduce the dependence of the estimate on the M_{T2} shape modeling in simulation,

Table 4.2: The last M_{T2} bin and the M_{T2} extrapolation point for each topological region, beyond which shape data from simulation is used to extrapolate the lost lepton estimate into M_{T2} bins.

the transfer factor $k(M_{\text{T2}})$ uses a combination of data and simulation information. In each topological region, the greatest M_{T2} bin is iteratively combined with the next-greatest bin until the total expected SM background yield in simulation is at least 50 events. These combined bins together form the CR for a range of M_{T2} values, where the fraction of events falling in a particular M_{T2} bin, $k(M_{\text{T2}})$, is determined from simulation. In all the other M_{T2} bins in the topological region, statistics are sufficient for a direct measurement in data and $k(M_{\text{T2}}) = 1$. The extrapolation point for each topological region can be found in table 4.2. The shape modeling in simulation is verified in data by selecting an inclusive sample enriched in either W boson or $t\bar{t}$ production (using the number of b-tags in the event), and predicting the M_{T2} distribution using simulation after normalizing simulation yield to data and summing all the control regions together, as seen in figure 4.6.

4.3.2   Systematic Uncertainties

Several sources of uncertainty are assessed for the lost lepton estimate, including those associated with the topological transfer factor and the M_{T2} shape modeling in simulation. The full list of systematic uncertainties is as follows:
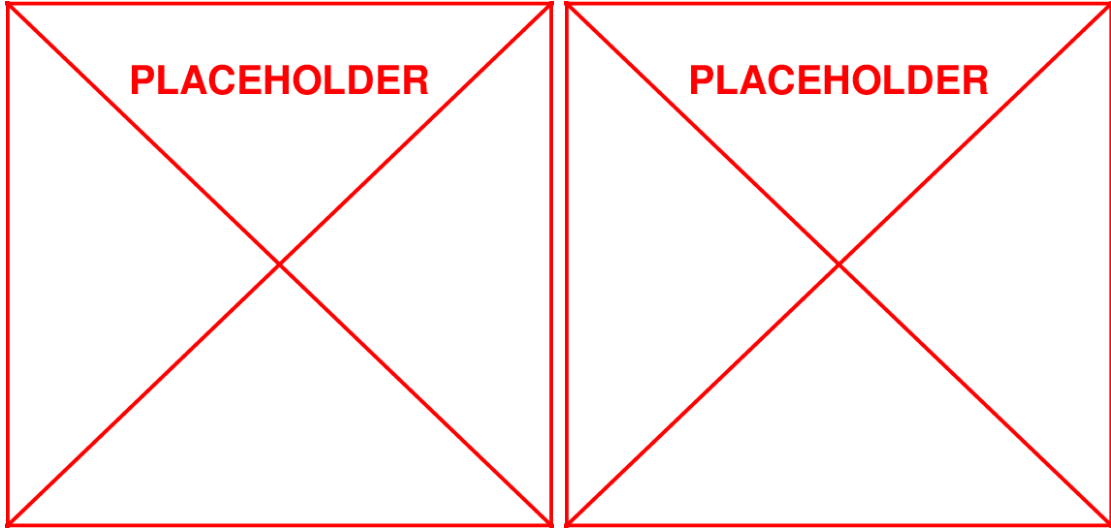
Figure 4.6: The M_{T2} shape in data and simulation using the single lepton CR selection, for events with zero b-tagged jets (left) or at least one b-tagged jet (right). The simulation is normalized to data in each topological region before summing to create the inclusive region. The hatched bands in each upper plot show the MC statistical uncertainty, while the shaded bands in each lower plot represent the systematic shape uncertainty.

- *Control region statistical error:* the Poisson error on the number of observed events in data is taken as a correlated uncertainty in each signal region using the same control region. The error is uncorrelated amongst M_{T2} bins, except in regions sharing merged M_{T2} bins for the estimate.

- *Monte Carlo statistical error:* where the simulation is used to compute the transfer factor (and in some cases, the M_{T2} shape), the MC statistical uncertainty ranges from 1-50%, taken as uncorrelated amongst all bins.

- *Electron and muon selection efficiency:* the reconstruction and identification efficiency of electrons and muons is computed using a procedure known as *tag and probe*, where leptons from $Z \to \ell\ell$ decays are used to evaluate the lepton id efficiencies as a function of lepton p_{T} and η and lepton isolation efficiencies as a function of p_{T} and nearby activity. The scale factors applied to simulation to compensate for such effects approach unity with uncertainties on the order of a few percent, and are correlated across all bins. The maximum effect of this uncertainty is up to 7% in some signal regions.

- *Tau selection efficiency:* the efficiency for hadronically decaying taus is classified according to the number of charged particles in the τ decay, whether a *1-prong* tau leaving a single charged track, or a *3-prong* tau leaving three charged tracks in the final state. This efficiency is measured in simulation by measuring the isolation efficiency as a function of candidate p_{T} for electrons, muons, and taus in various decay modes. Based on half the difference in efficiency between 1-prong taus and muons, an

45

uncertainty of 10% is taken for 1-prong taus which also covers and differences in veto efficiency as a function of the primary kinematic variables. For 3-prong taus, the PF hadron veto is very inefficient since most fail the isolation requirement (the typical selection efficiency is 8%), and a 100% relative uncertainty is taken to cover any differences as a function of kinematics. These uncertainties are correlated amongst all bins.

- M_T *cut efficiency:* the use of simulation to compute the CR-to-SR transfer factor also relies on the satisfactory modeling of the M_T cut in Monte Carlo. A sample of $Z \to \ell\ell$ events is selected in data with one of the leptons "deleted" from the event to mimic a leptonic W boson decay, and compared with simulation. Based on data-MC agreement, a correlated error of 3% is taken across all bins.

- *b-tagging efficiency:* the effect of varying the b-tag scale factor efficiency is calculated in each bin, and taken as a correlated error amongst all bins. The maximum effect of this variation is about 4% in some bins.

- *Jet energy corrections:* by varying the jet energy scales across all bins, a maximum deviation of about 5% is observed in regions with sufficient statistics, and a correlated error of 5% is taken across al bins.

- *MC renormalization and factorization scales:* the overall effect of varying the simulation renormalization and factorization scales of the underlying physics processes (and subsequent effect on event kinematics) are computed separately in each bin.

Taken as correlated across all bins, they are typically on the order of a few percent but range up to 10% in some regions.

- M_{T2} *shape uncertainty:* in regions where the simulation is used to model the M_{T2} distribution, additional variations of the renormalization and factorization scales, parton distribution functions, b-tagging scale factor uncertainties, and jet energy scale uncertainties are performed to measure their effect on the M_{T2} shape modeling. The most significant impact is seen in the highest M_{T2} bins from theoretical uncertainties ($\sim$15%), and up to 40% in low statistics bins due to jet energy scale variations. With this in mind, the shape uncertainty (in regions where MC M_{T2} shape modeling is used) is assigned as a linear morphing of the M_{T2} shape starting in the first bin from which MC extrapolation is used, growing to 40% in the final bin. The shape morphing in every distinct topological region is taken as an uncorrelated error.

4.3.3   Signal Contamination

Nearly every control region in this analysis (including those for multijet and invisible Z backgrounds) is not only orthogonal to the signal region selection, but also crafted such that any potential BSM signal contribution to any CR is negligible and will not bias the background estimate. However, certain SUSY simplified models yield final states with prompt lepton decays — in some cases kinematically similar to SM $t\bar{t}$ decays — which may be non-negligible in the lost lepton CR. To account for this effect when calculating

limits on such models (described in detail in section refsec:interpretations), the amount by which the lost lepton background is overestimated is modeled as a loss in signal efficiency. The modified signal yield $N_{\text{sig}}^{SR\prime}$ is defined in equation 4.5, where N_{sig}^{SR} and N_{sig}^{CR} represent the simulated signal yield in the signal and control regions respectively, and TF the total transfer factor used in the lost lepton estimate for a given signal region.

$$N_{\text{sig}}^{SR\prime} = N_{\text{sig}}^{SR} - TF \cdot N_{\text{sig}}^{CR} \tag{4.5}$$

The correction is most significant for simplified models with stop decays, where the stop-neutralino mass splitting is close to the SM top quark mass. In such cases, the signal contamination is maximally 5% of the expected background yields in each CR for mass points near the expected exclusion limits at high mass.

4.4   Invisible Z Estimate

The invisible Z background is the dominant SM contribution in many signal regions due to the "irreducible" nature of the underlying physics. Because the primary interaction produces a massive particle decaying to an invisible final state ($Z \to \nu\bar{\nu}$) recoiling against hadronic activity, the signature is fundamentally similar to that of the BSM physics the search is designed to target and difficult to reduce by conventional cuts removing SM background contributions. A robust method leveraging the well-understood Drell-Yan process ($Z \to l^+l^-$) — kinematically similar to the invisible Z background — is used to predict the expected SM contribution based on data while minimizing the reliance on Monte Carlo modeling of the kinematics.

4.4.1 Background Prediction

The invisible Z background is estimated using dilepton events selected in data. The control region consists of events selected with dilepton triggers, with addition requirements that the leptons are of the same flavor and opposite sign. The momentum of the leading and trailing lepton must also be at least $100\,\mathrm{GeV}$ and $30\,\mathrm{GeV}$, respectively, and the invariant mass of the lepton system $m_{\ell\ell}$ must be withing $20\,\mathrm{GeV}$ of the Z boson mass. The individual CRs are then constructed by removing the dilepton system from the event and applying the baseline preselection requirements as for the signal regions. A detailed description of the CR can be found in section 3.5.2.

The invisible Z prediction is computed as described by equation 4.6, where $N_{\ell\ell}^{\mathrm{CR(SF)}}$ is the number of events in the dilepton same-flavor (SF) region, $N_{\ell\ell}^{\mathrm{CR(OF)}}$ the number of events in the dilepton opposite-flavor (OF) region, $R^{\mathrm{SF/OF}}$ the SF-OF transfer factor, $R_{\mathrm{MC}}^{Z\to\nu\bar{\nu}/Z\to l^+l^-}$ the transfer factor from $Z \to l^+l^-$ to $Z \to \nu\bar{\nu}$ events, and $k(M_{\mathrm{T2}})$ the transfer factor into bins of M_{T2} . Each factor is explained in detail below.

$$N_{Z\to\nu\bar{\nu}}^{\mathrm{SR}}(H_{\mathrm{T}}, N_{\mathrm{j}}, N_{\mathrm{b}}, M_{\mathrm{T2}}) = \left[N_{\ell\ell}^{\mathrm{CR(SF)}}(H_{\mathrm{T}}, N_{\mathrm{j}}, N_{\mathrm{b}}) - N_{\ell\ell}^{\mathrm{CR(OF)}}(H_{\mathrm{T}}, N_{\mathrm{j}}, N_{\mathrm{b}}) \cdot R^{\mathrm{SF/OF}} \right]$$
$$\times R_{\mathrm{MC}}^{Z\to\nu\bar{\nu}/Z\to l^+l^-}(H_{\mathrm{T}}, N_{\mathrm{j}}, N_{\mathrm{b}}) \cdot k(M_{\mathrm{T2}}) \tag{4.6}$$

The second term in equation 4.6, $N_{\ell\ell}^{\mathrm{CR(OF)}} \cdot R^{\mathrm{SF/OF}}$, is a correction factor applied to the control region yield to correct for the contribution from processes producing SF and OF event, or *flavor-symmetric processes* (such as $t\bar{t}$). To compensate for this contribution, a separate control region enriched in $t\bar{t}$ events is selected from data by using the same selections as the invisible Z CR, except for an inverted selection on the dilepton p_{T} and

Table 4.3: The control region predicted Drell-Yan (DY) yield, SF yield, OF yield, purity (the rfraction of $Z \to l^+l^-$ events), and ratio $R_{\mathrm{MC}}^{Z \to \nu\bar{\nu}/Z \to l^+l^-}$ for various topological regions. Note that the 7+ jet regions with b-tags (marked with an asterisk) share the same CR, and the fraction of events with different numbers of b-tags is folded into the ratio.

$m_{\ell\ell}$ requirements. The ratio of $t\bar{t}$ events is then measured directly from data by counting the yield of SF (ee or $\mu\mu$) and OF ($e\mu$ or μe) events in this CR. The ratio $R^{\mathrm{SF/OF}}$ is expected to be close to unity based on the underlying physics process, but due to varying acceptance and efficiency effects for different flavor leptons is measured as $R^{\mathrm{SF/OF}} = 1.13 \pm 0.15$, and is stable with respect to event kinematics as shown in figure 4.7. The Drell-Yan yield in each control region, as well as the SF yield, OF yield, and transfer factors can be found in table 4.3.

The ratio of invisible Z events to Drell-Yan events, $R_{\mathrm{MC}}^{Z \to \nu\bar{\nu}/Z \to l^+l^-}$, is measured in simulation for each CR. The ratio accounts for the branching fraction differences between $Z \to l^+l^-$ and $Z \to \nu\bar{\nu}$ decays, as well as differences in lepton acceptance and efficiency for dilepton pairs in the CR (including corrections for differences in lepton efficiency between data and simulation).

The transfer factor $k(M_{\mathrm{T2}})$ uses a combination of data and simulation information to reduce the dependence of the prediction on the M_{T2} shape modeling in simulation. Based on simulation, the M_{T2} shape in ever H_{T} region is independent of $N_{\mathrm{b\text{-}jets}}$. In addition, in the extreme H_{T} region $H_{\mathrm{T}} > 1500\,\mathrm{GeV}$, the shape is also independent of N_{jets}.
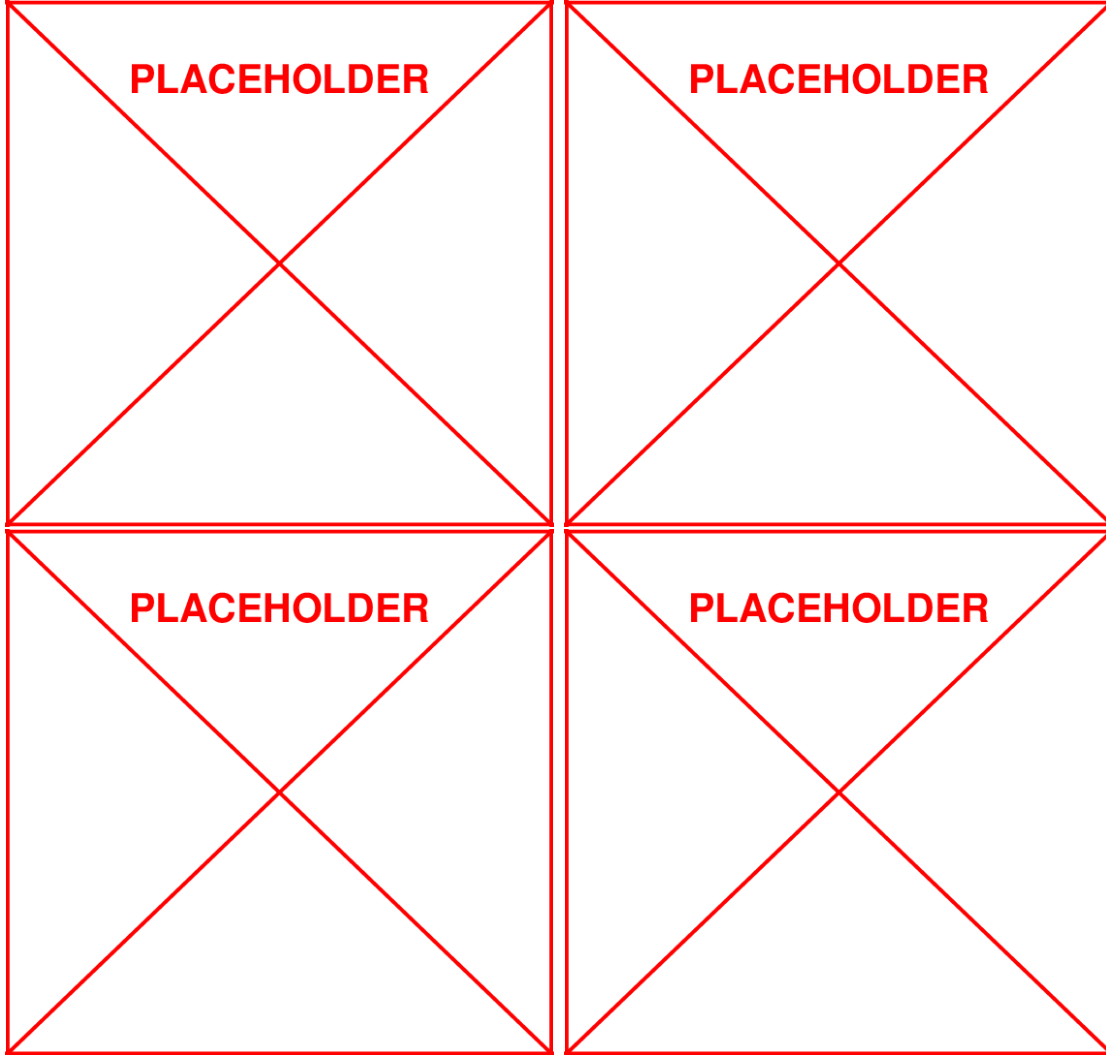
Figure 4.7: The ratio of same-flavor to opposite-flavor events in a $t\bar{t}$ enriched control region, as a function of N_{jets} (top left), $N_{\text{b-jets}}$ (top right), H_{T} (bottom left), and M_{T2} (bottom left). The solid black line corresponds to a constant value of 1.13 ± 0.15, while the dashed black lines correspond to the statistical uncertainty and the dashed red lines the total systematic uncertainty.

Table 4.4: The M_{T2} extrapolation point for each topological region, beyond which shape data from simulation is used to extrapolate the invisible Z estimate into M_{T2} bins. "NA" indicates regions where the simulation shape is not used at all since dilepton statistics in data are sufficiently large to perform the estimate bin-by-bin.

Based on these findings, M_{T2} shape templates are constructed for each H_{T} and N_{jets} region (inclusive in $N_{\text{b-jets}}$[1]), with one single template for the extreme H_{T} region (which is also inclusive in N_{jets}).

The M_{T2} distribution, $k(M_{\text{T2}})$, for each topological region is constructed from dilepton events in data where statistics allows, and $Z \to \nu\bar{\nu}$ simulation where statistics are sparse. In each template region, the greatest M_{T2} bin is iteratively combined with the next-greatest bin until the total expected SM background yield in simulation is at least 50 events. For uncombined bins (where statistics in data is sufficient), the M_{T2} shape is taken directly from dilepton data, corrected for the ratio $R_{\text{MC}}^{Z \to \nu\bar{\nu}/Z \to l^+l^-}$. In the low-statistics combined regime, the M_{T2} shape in $Z \to \nu\bar{\nu}$ simulation is used to determine the fraction of events in each M_{T2} bin after normalizing the simulation yield to data in the combined bins. The extrapolation point after which the M_{T2} shape is based on simulation in each signal region can be found in table 4.4.

The accuracy of the M_{T2} shape modeling in simulation is verified using other control

---

[1]The only exception is for regions with more than 2 b-tags (2j3b and 2-6j3b), where at least 3 jets are required so as to avoid biasing the N_{jets} distribution when requiring more b-tags than jets.

samples enriched in γ, $W \to \ell\nu$, and $Z \to l^+l^-$ events in each H_T bin, as shown in figure 4.8. The γ-enriched sample is selected using photon triggers and requiring $p_T^\gamma > 180\,\text{GeV}$, with corrections applied for multijet background contributions and the ratio of M_{T2} distributions for photon to Z boson events, $R_{MC}^{Z/\gamma}$. The W and Z boson samples are selected in data using leptonic triggers, corrected to compensate for the contribution from top quark production, as well as the ratio of M_{T2} distributions, $R_{MC}^{Z/W}$ and $R_{MC}^{Z \to \nu\bar{\nu}/Z \to l^+l^-}$ respectively.

4.4.2   Systematic Uncertainties

Several sources of uncertainty are assessed for the invisible Z estimate, including those associated with the various transfer factors and the M_{T2} shape modeling in simulation. The dominant uncertainty in regions using an M_{T2} template constructed from data is the statistics of the template. The full list of systematic uncertainties is as follows:

- *Control region statistical error:* the Poisson error on the number of observed events in $Z \to l^+l^-$ data is taken as an uncorrelated uncertainty across all signal regions.

- $R_{MC}^{Z \to \nu\bar{\nu}/Z \to l^+l^-}$ *statistical error:* the statistical error associated with the number of MC events generated factors into the transfer factor.

- $R_{MC}^{Z \to \nu\bar{\nu}/Z \to l^+l^-}$ *systematic error:* a 5.5% uncertainty based on variations of lepton efficiency uncertainties as well as other modeling parameters (jet energy scales, factorization and renormalization scales, etc.) is applied as a correlated error in each topological region.
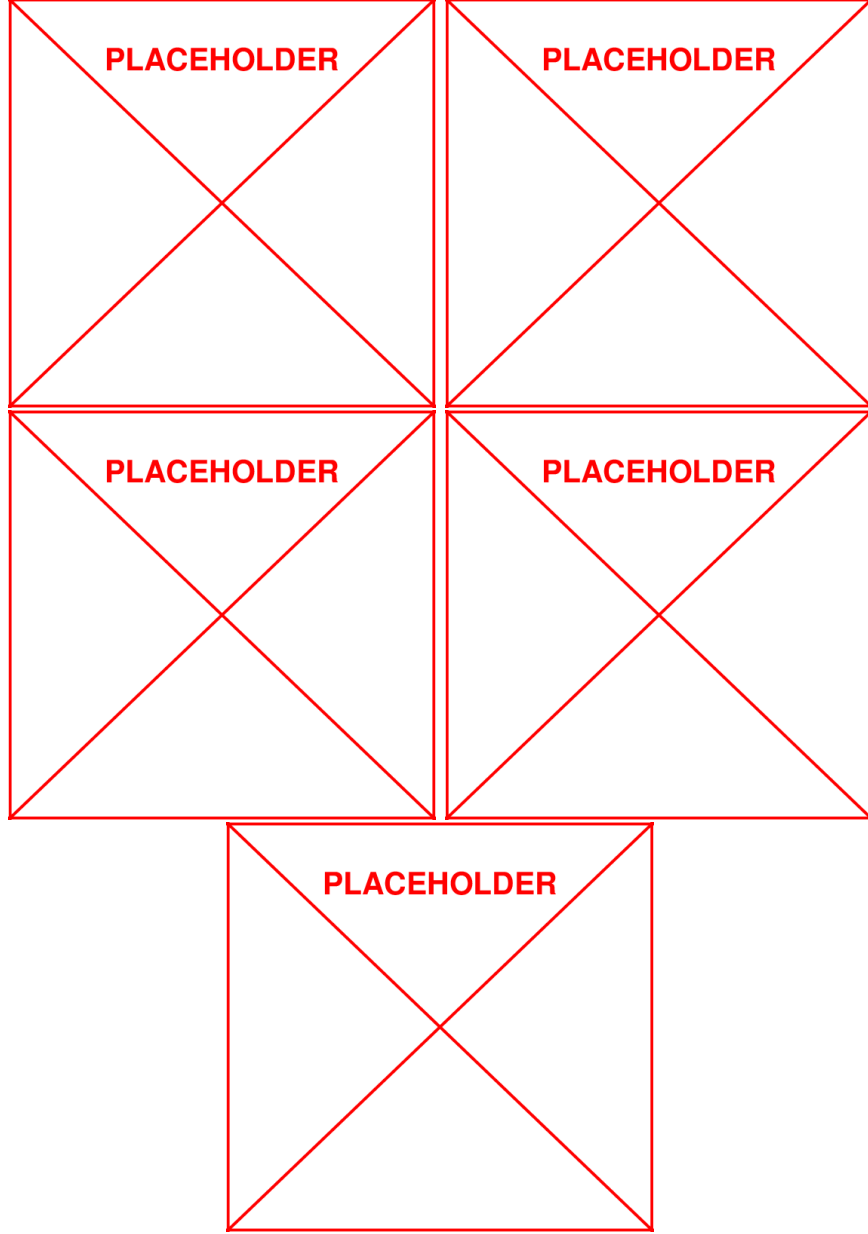
Figure 4.8: The M_{T2} shape distribution in $Z \to \nu\bar{\nu}$ simulation compared to γ, $W \to \ell\nu$, and $Z \to l^+l^-$ enriched samples in data, for each H_{T} region. The solid grey band indicates the systematic uncertainty associated with the M_{T2} shape modeling.

- *Flavor-symmetric subtraction statistical error:* a Poisson error based on the number of observed opposite-flavor events.

- *Flavor-symmetric subtraction systematic error:* a 15% uncertainty on the $t\bar{t}$ contamination based on the $R^{\mathrm{SF/OF}}$ uncertainty.

- M_{T2} *shape uncertainty:* in regions where the simulation is used to model the M_{T2} distribution, additional variations of the renormalization and factorization scales, parton distribution functions, b-tagging scale factor uncertainties, and jet energy scale uncertainties are performed to measure their effect on the M_{T2} shape modeling. The most significant impact is seen in the highest M_{T2} bins of up to 20%. In order to cover the uncertainty from additional electro-weak corrections not present in simulation (and possibly not covered by the above variations), a conservative upper threshold of 40% is used, and the shape uncertainty (in regions where MC M_{T2} shape modeling is used) is assigned as a linear morphing of the M_{T2} shape starting in the first bin from which MC extrapolation is used, growing to 40% in the final bin. The shape morphing in every distinct topological region is taken as an uncorrelated error.

Chapter 5

Results

The techniques described in this analysis are applied to $35.9\,\mathrm{fb}^{-1}$ of proton-proton collision data gathered at the LHC and recorded by the CMS detector. Observed yields in each signal region are further interpreted in the context of simplified SUSY models to establish new limits on the masses of hypothesized BSM particles.

5.1 Yields and Significances

The observed yield in the search regions is statistically compatible with the predicted background from SM processes. A summary of the total yield in each signal region and predicted background contribution relying only on the techniques described in chapter 4, referred to as *pre-fit* results, is illustrated for each topological region in figure 5.1, and the individual yield in each M_{T2} bin can be found in figures 5.2, 5.3, and 5.4.

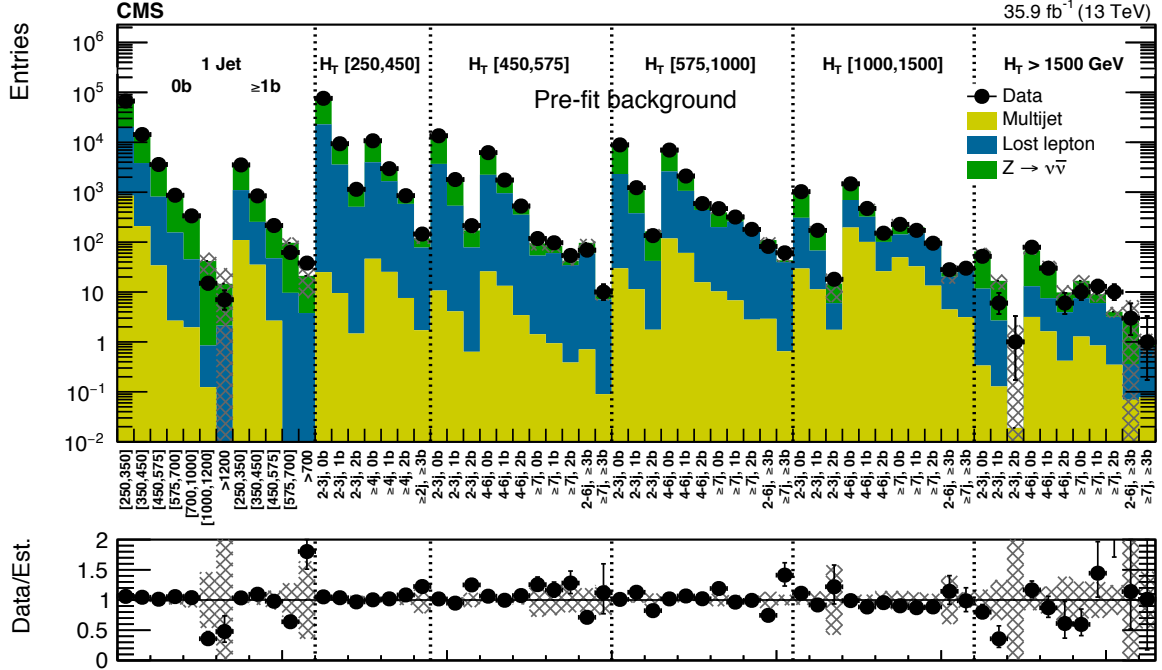The background estimate is further refined by performing a maximum-likelihood fit

Figure 5.1: The data yield in each topological region compared to the pre-fit background prediction. The hatched bands illustrate the total uncertainty in the background estimate. Results in the monojet regions are binned in jet p_{T}, while those in the multijet regions are labeled according to N_{jets} and $N_{\mathrm{b\text{-}jets}}$.
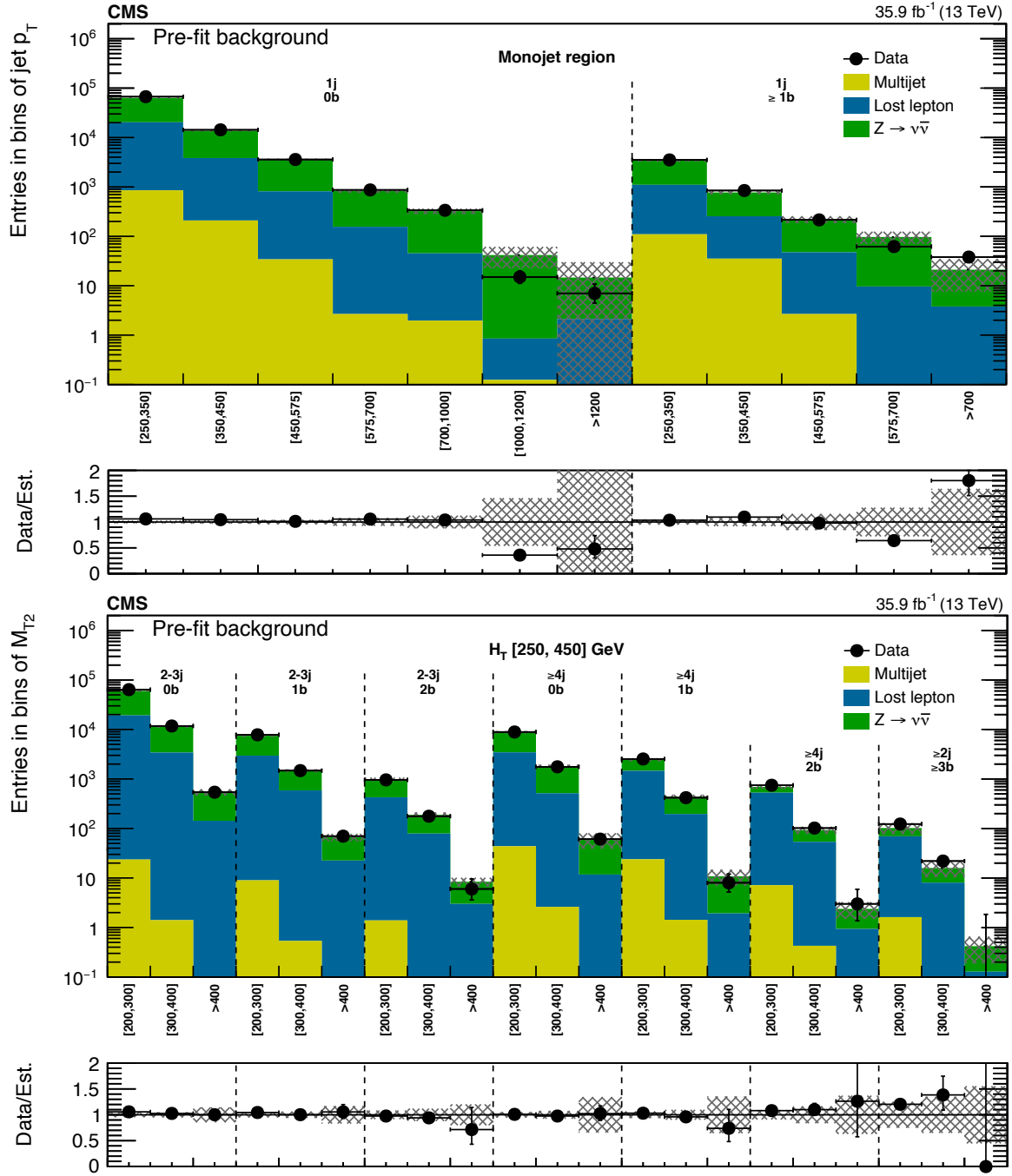
Figure 5.2: The data yield in the monojet and very-low H_T regions compared to the pre-fit background prediction. The hatched bands illustrate the total uncertainty in the background estimate. Results in the monojet regions are binned in jet p_T in units of GeV, while those in the multijet regions are labeled according to M_{T2} bin in units of GeV.
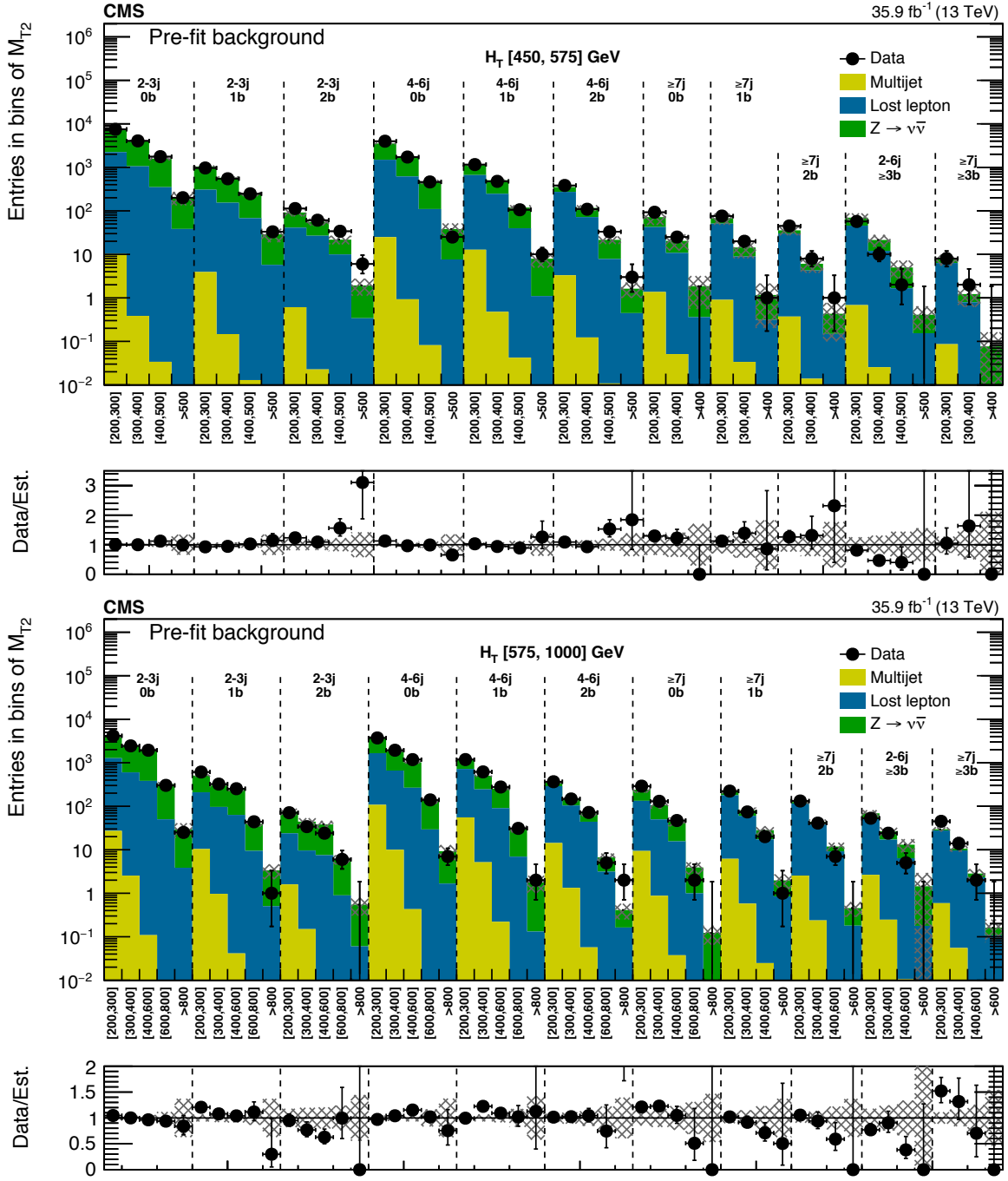
Figure 5.3: The data yield in the low H_T and medium H_T regions compared to the pre-fit background prediction. The hatched bands illustrate the total uncertainty in the background estimate. Results are labeled according to M_{T2} bin in units of GeV.
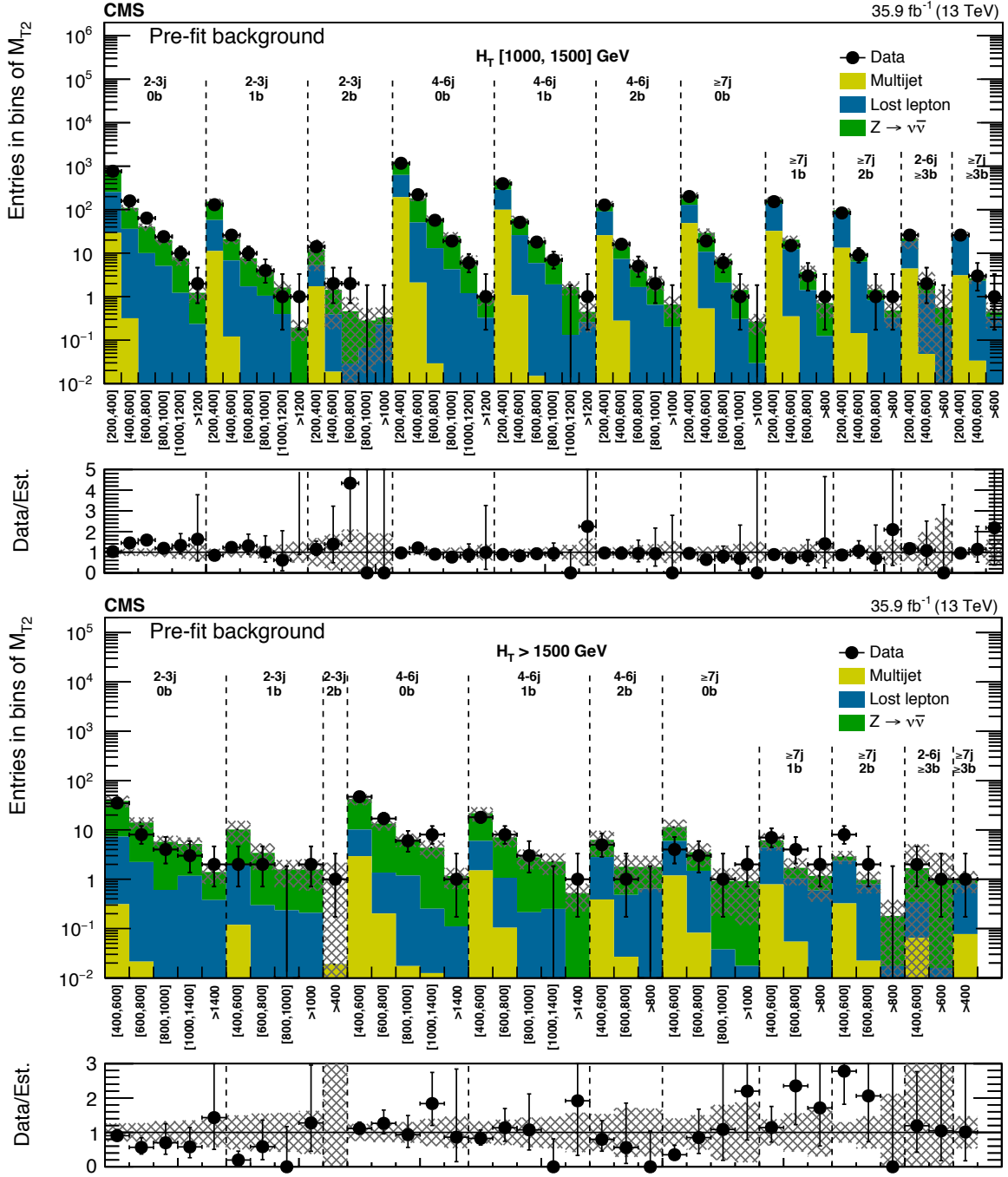
Figure 5.4: The data yield in the high H_T and extreme H_T regions compared to the pre-fit background prediction. The hatched bands illustrate the total uncertainty in the background estimate. Results are labeled according to M_{T2} bin in units of GeV.

60

to data in each signal region, referred to as *post-fit* results. The fit is performed using both background-only or background-plus-signal hypotheses to set limits on simplified physics models described in section 5.2. The estimates and uncertainties on each background as described in chapter 4 are used as inputs to the fitting procedure, where the likelihood is constructed as a product of Poisson probability density functions for each signal region with constraints set according to the background uncertainties and signal uncertainties (described in **FIXME: ref subsection on signal systematics**). The post-fit yield for each topological region is illustrated in figure 5.5, and the individual yield in each M_{T2} bin can be found in figures 5.6, 5.7, and 5.8. The post-fit procedure has the effect of constraining background and its associated uncertainties when the fitting procedure is applied to data consistent with predictions modeling uncertainties appropriately.
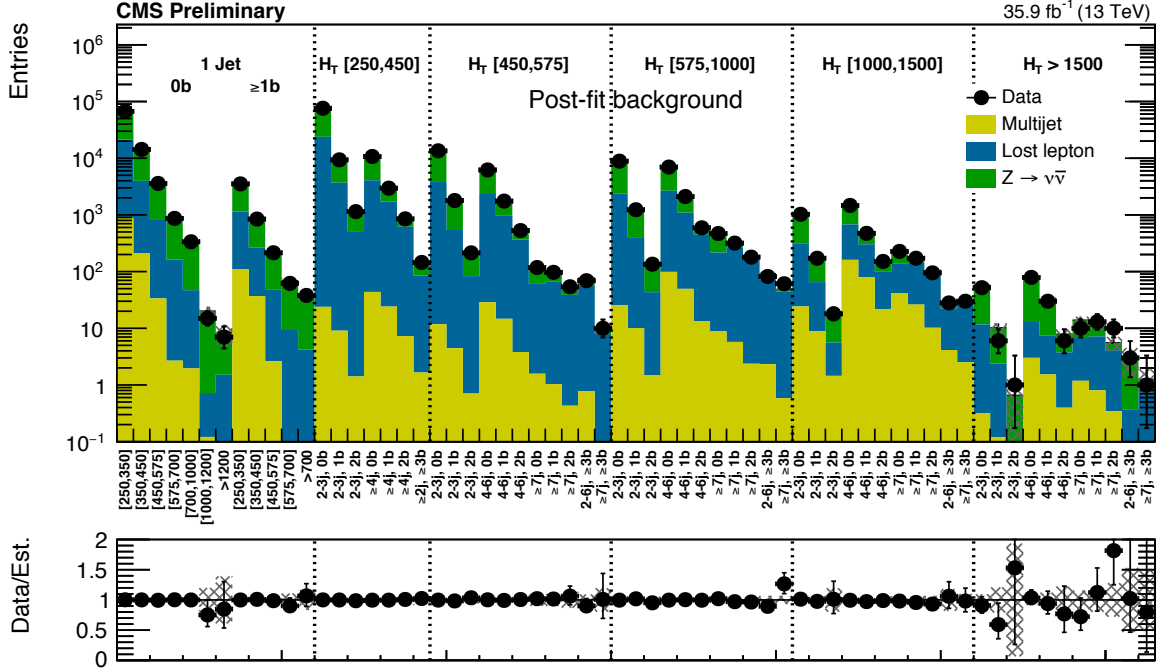
5.2   Signal Interpretations

Figure 5.5: The data yield in each topological region compared to the post-fit background prediction. The hatched bands illustrate the total uncertainty in the background estimate. Results in the monojet regions are binned in jet p_T, while those in the multijet regions are labeled according to N_{jets} and $N_{\text{b-jets}}$.
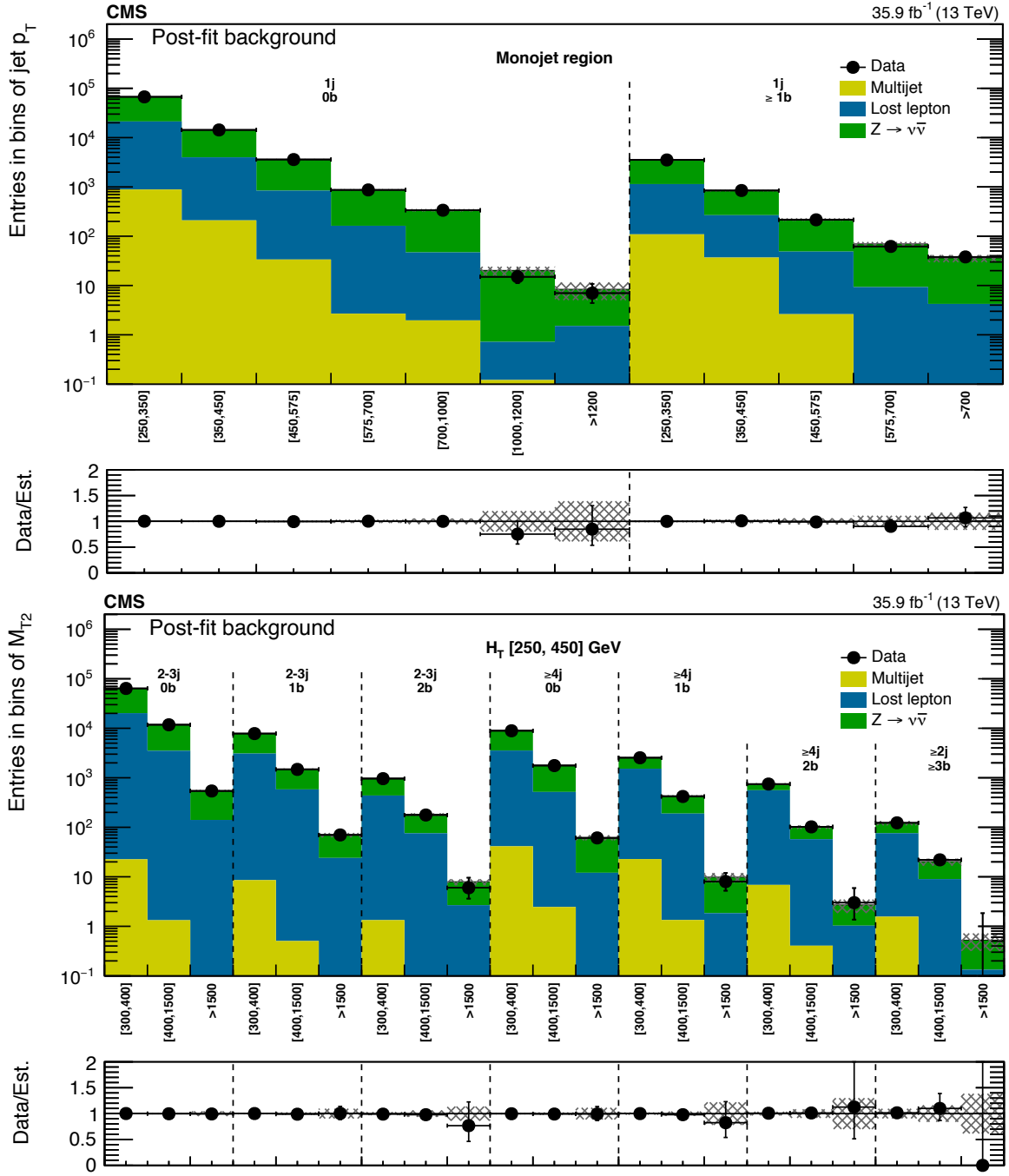
Figure 5.6: The data yield in the monojet and very-low H_T regions compared to the post-fit background prediction. The hatched bands illustrate the total uncertainty in the background estimate. Results in the monojet regions are binned in jet p_T in units of GeV, while those in the multijet regions are labeled according to M_{T2} bin in units of GeV.
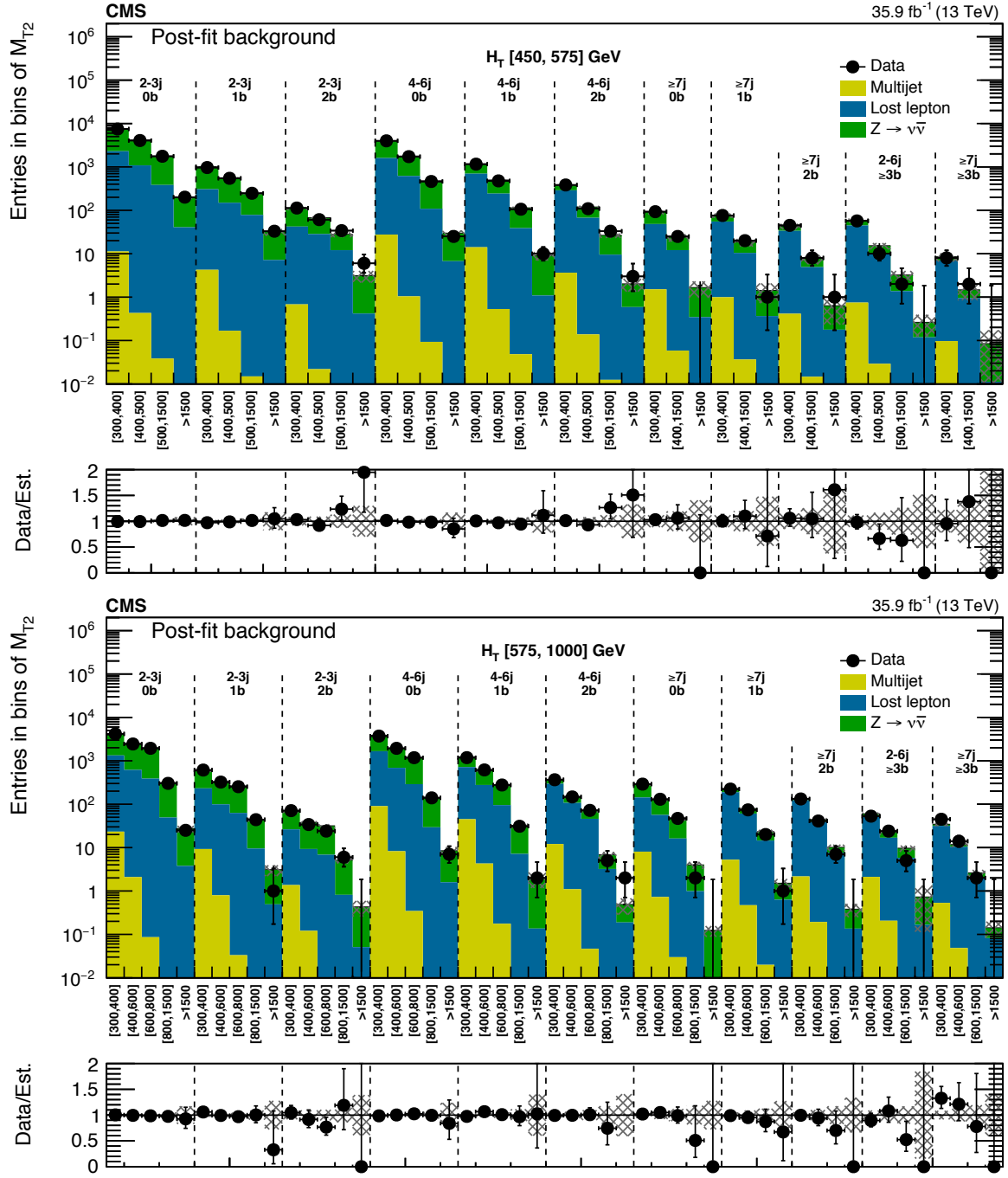
Figure 5.7: The data yield in the low H_T and medium H_T regions compared to the post-fit background prediction. The hatched bands illustrate the total uncertainty in the background estimate. Results are labeled according to M_{T2} bin in units of GeV.

Figure 5.8: The data yield in the high H_T and extreme H_T regions compared to the post-fit background prediction. The hatched bands illustrate the total uncertainty in the background estimate. Results are labeled according to M_{T2} bin in units of GeV.
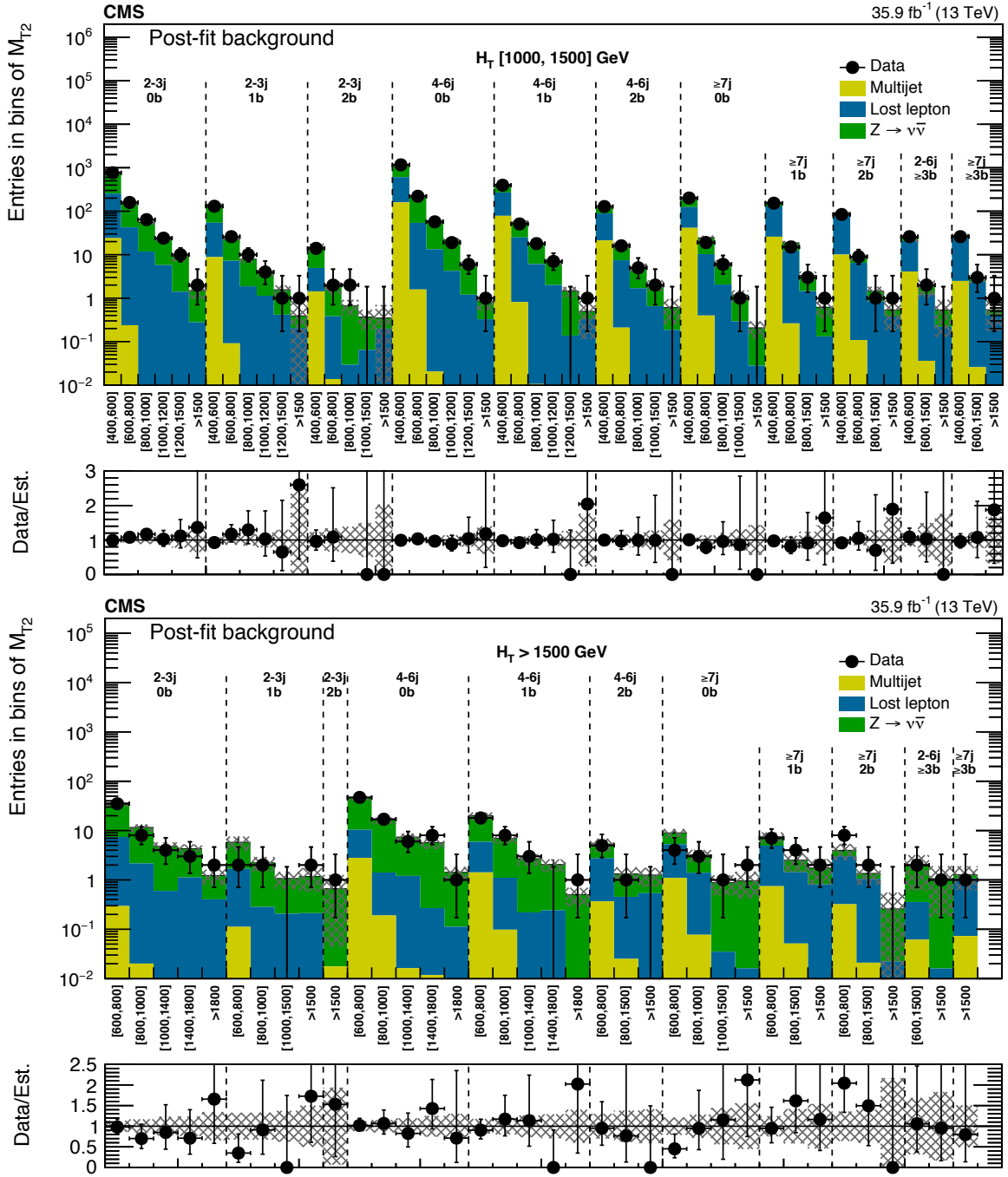
Chapter 6

Naturalness and Soft Leptons

6.1 Natural Supersymmetry

6.2 Extending the M_{T2} Search

6.3 Searching for Natural Susy with Soft Leptons

Bibliography