UNIVERSITY OF CALIFORNIA, SAN DIEGO

Search for New Physics in a Final State with Same-Sign Dileptons, Jets, and Missing Transverse Energy at 7 TeV Center of Mass Energy

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

in

Physics

by

Frank Golf III

Committee in charge:

Professor Avraham Yagil, Chair
Professor Claudio Campagnari
Professor Clark Gibson
Professor Aneesh Manohar
Professor Frank Würthwein

2012

The dissertation of Frank Golf III is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2012

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

A thesis is not only a summary of years of research and study, but represents the efforts and friendship of all those without whom I would surely have faltered. I would like to express my gratitude to all who have been important to the realization of this thesis.

I would like to thank my advisor, Avi Yagil. I would not be on the verge of completing this work without the guidance and opportunities he provided. His patience and encouragement supported me along the way, as did his delicious cooking. Thanks also to Frank Würthwein and Claudio Campagnari. The insight and motivation you provided were invaluable. Throughout my thesis analysis, Frank's patience and friendship never wavered. And I am indebted to Aneesh Manohar for teaching me so much and for his understanding. His support has been steadfast, first as an advisor and later as a mentor.

The work done in this thesis, like most work in high energy physics, was a collaborative effort. The list of all those who contributed to my development is numerous and their offerings many. I have fond memories of the many discussions shared with Randy Kelley. I am thankful to everyone in my broader research group with whom I shared the past three years. The time would not have been so enjoyable without Puneeth Kalavase and Jake Ribnik. The importance of their support and friendship cannot be overstated.

My transition from being a theorist to an experimentalist would have faltered from the beginning without Johannes Mülmenstädt. And while much thanks is owed for all he taught me, to end there would be to understate how important his wisdom and support have been to me. Throughout the years our relationship has been different things at different times — colleagues, running partners, connoisseurs of fine fish tacos — but his friendship has always been constant.

One person refuses to fit into the timeline and categories of my life. I consider John McRaven to be a brother. Whatever successes I've enjoyed have been with his support; failures overcome because of his friendship.

I would not have made it here without the support of Ikaros Bigi. He mentored me throughout my undergraduate studies and was my strongest advocate for continuing with graduate studies. Thanks also to Vivek Sharma for opening the doors of this institution to me.

Of course, I wouldn't have had any of these opportunities without my parents. Their hard work and passion for learning continue to be a source of inspiration for me.

My siblings - Mary, Ann, Tony and Stephen - helped me become the person I am today. They've never stopped encouraging me, nor reminding me of how many years I've avoided having a real job. That I've had these opportunities is only because of my wife, Barbara. Without her love, support and sacrifices, none of this would have been possible.

VITA

2003	B. S. in Mathematics and Physics *cum laude*, University of Notre Dame
2007	M. S. in Physics, University of California, San Diego
2011	Ph. D. in Physics, University of California, San Diego

PUBLICATIONS

J. Chiu, F. Golf, R. Kelley and A. Manohar, "Electroweak Sudakov Corrections Using Effective Field Theory", *Phys. Rev. Lett.*, 100, 2008.

J. Chiu, F. Golf, R. Kelley and A. Manohar, "Electroweak Corrections in High Energy Processes Using Effective Field Theory", *Phys. Rev. D*, D77, 2008.

V. Khachatryan, *et al.* (CMS Collaboration) "First Measurement of the Cross Section for Top–Quark Pair Production in Proton–Proton Collisions at $\sqrt{s} = 7$ TeV", *Phys. Lett.*, B695, 2011.

S. Chatrchyan, *et al.* (CMS Collaboration) "Search for New Physics with Same–Sign Isolated Dilepton Events with Jets and Missing Transverse Energy at the LHC", *JHEP*, 1106, 2011.

ABSTRACT OF THE DISSERTATION

Search for New Physics in a Final State with Same-Sign Dileptons, Jets, and Missing Transverse Energy at 7 TeV Center of Mass Energy

by

Frank Golf III

Doctor of Philosophy in Physics

University of California, San Diego, 2012

Professor Avraham Yagil, Chair

We report on a search for new physics in a final state with two same sign leptons, missing transverse energy, and significant hadronic activity at a center of mass energy $\sqrt{s} = 7$ TeV. The data were collected with the CMS detector at the CERN LHC and correspond to an integrated luminosity of 0.98 fb^{-1}. Data–driven methods are developed to estimate the dominant Standard Model backgrounds. No evidence for new physics is observed.

The dominant background to the analysis comes from failures of lepton identification in Standard Model $t\bar{t}$ events. The $t\bar{t}$ production cross section in the dilepton final state is measured using 3.1 pb^{-1} of data. The cross section is measured to be 194 $\pm$ 72 (stat) $\pm$ 24 (syst) $\pm$ 21 (lumi) pb.

An algorithm is developed that uses tracking information to improve the reconstruction of missing transverse energy. The reconstruction of missing transverse energy is commissioned using the first collisions recorded at 0.9, 2.36 and 7 TeV data. Events with abnormally large values of missing transverse energy are identified as arising from anomalous signals in the calorimeters. Tools are developed to identify and remove these anomalous signals.

Chapter 1

Introduction

1.1 Overview of the LHC and CMS

The standard model (SM) of particle physics successfully describes all current high-energy experimental data. In the last few years, experimental evidence for neutrino masses suggests the existence of new interactions at energies higher than the electroweak scale. One of the key remaining questions at the electroweak scale is the origin of the masses of the W and Z bosons. In the SM in its simplest implementation, it is attributed to the spontaneous breaking of electroweak symmetry, accomplished by introducing a new scalar field. The existence of the associated field quantum, the Higgs boson, has yet to be experimentally confirmed.

The search for this new particle influenced the design of the Large Hadron Collider (LHC) and the Compact Muon Solenoid (CMS) experiment. The likely mass of the Higgs together with the small production cross section necessitated building a machine capable of delivering energetic collisions at a high rate. The design luminosity of $\mathcal{L} = 10^{34}$ cm^{-2} s^{-1} leads to approximately 1 billion proton–proton collisions per second.

Technical limitations in the production and storage of antiprotons led to the decision to build a pp collider. The nominal design energy of each proton beam was 7 TeV. First collisions occurred in September 2008, but no significant data was collected before an incident on 19 September 2008 shut down the accelerator complex. Collisions resumed more than a year later, but with 3.5 TeV beams in order to operate within the safe range of the accelerator magnets. Collisions will likely continue at this energy until the first repair and upgrade, tentatively scheduled to begin sometime during 2013.

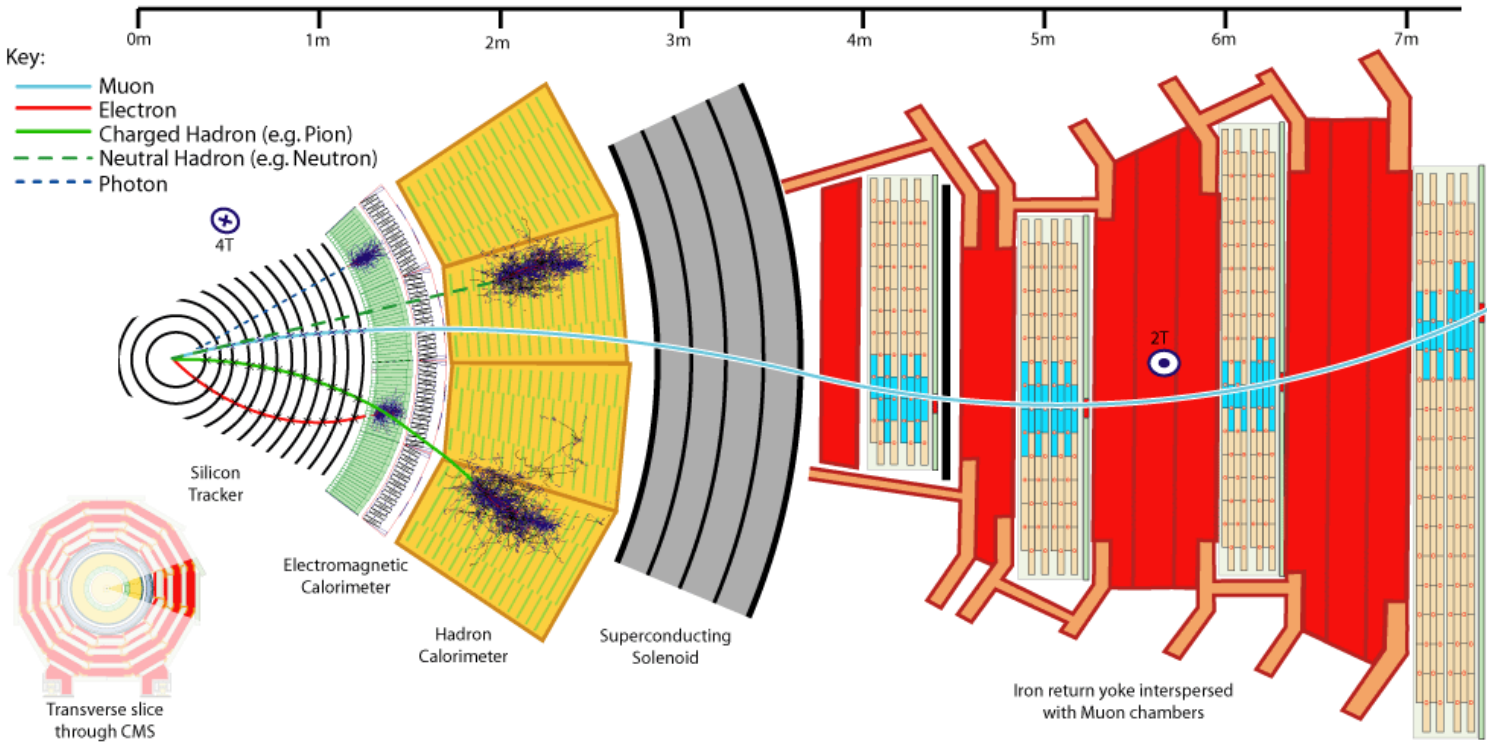

Figure 1.1: A slice of the CMS detector together with schematic trajectories for different types of particle produced in pp collisions. From [1].

The CMS detector was designed to operate in a high luminosity environment. Figure 1.1 shows a slice of the detector together with schematic trajectories for different types of particles produced in pp collisions. A tracker system composed of layers of silicon pixels and silicon strips surrounds the LHC beam pipe. The tracker is enclosed by a two-depth calorimeter — the inner layer contains the electromagnetic calorimeter (ECAL) with the hadron calorimeter (HCAL) positioned behind it. A magnet producing a 3.8 T field encompasses both the tracker and calorimeter system. Finally, a series of chambers for the reconstruction and identification of muons comprises the outer layer. The large magnetic field provides significant bending power to allow for the desired momentum resolution in the tracker and the muon spectrometer.

The silicon tracker is comprised of three pixel layers close to the beam pipe, enclosed by 10 silicon strip layers at larger radii, providing nearly complete coverage out to $|\eta| \sim 2.5$. As charged particles travel through silicon layers, they interact with the material and deposit energy. The resulting ionized charge is read out by the tracker electronics and a set of points is constructed that represents the path of the charged particle [2]. The reconstructed trajectory is used to determine how much the path of the particle was bent by the enclosing magnetic field, and its momentum is determined according to

$$p = \frac{0.3qBR}{\cos\theta}, \tag{1.1}$$

where q is the charge of the particle, B the strength of the magnetic field, R the radius of curvature of the reconstructed trajectory and θ the angle between the trajectory and the z-axis. The coordinate system adopted by CMS has the origin centered at the nominal collision point, the y–axis pointing vertically upward, the x-axis pointing radially inward toward the center of the LHC, and the z-axis pointing along the counterclockwise beam.

The tracker system has over 70 million readout channels — approximately 60 million pixel channels and 10 million strip channels — to allow for an average occupancy on the order of 1% in the strip layers at peak luminosity. The larger number of readout channels in the pixels results in an average occupancy much less than that of the strips. Low occupancy is important in a high luminosity environment to minimize the incorrect assignments of signals in the silicon during the reconstruction of particle trajectories [2]. The large material budget of the tracker presents interesting challenges and opportunities for the reconstruction of physics objects such as photons and missing transverse energy

($\not\!\!E_T$).

The tracker is surrounded by a two depth calorimeter. The front layer is occupied by a homogenous electromagnetic calorimeter, with a sampling hadron calorimeter positioned behind it. The electromagnetic calorimeter is designed to measure the energies of electrons and photons. When such a particle enters the ECAL, it interacts with the calorimeter material and an electromagnetic shower develops. Electrons scatter and emit bremsstrahlung radiation while sufficiently energetic photons convert into electron–positron pairs. The resulting light produced by the shower is read out by the ECAL electronics and converted into an energy measurement.

The electromagnetic calorimeter is comprised of nearly 75000 lead tungstate ($PbWO_4$) crystals with an energy resolution less than 1% for energies above 20 GeV. The crystals are quasi-projective and cover about 1° in $\Delta\phi$ and $\Delta\eta$, approximately equal to a Moliere radius — i.e., the radius of a cylinder that contains approximately 90% of the lateral shower of an electron [3]. The crystals are dozens of radiation lengths in depth, emit nearly 80% of the light from an electromagnetic interaction in a single 25 ns bunch crossing, and are radiation hard [2]. The requirement of a highly granular ECAL with good energy resolution was driven by searches for $H \to \gamma\gamma$ at low mass. Both the ECAL and HCAL provided coverage out to $|\eta| \sim 3$.

The HCAL serves two roles, measuring the energy of hadronic particles and providing containment to prevent the same particles from entering the muon system. The HCAL is composed of alternating layers of brass plates and scintillating material. Hadronic particles, bound states of colored particles, undergo elastic and inelastic interactions with the brass nuclei, producing additional hadronic particles. The subsequent particles interact forming a shower of charged and neutral hadrons. As the shower crosses a scintillating layer, light is produced by interactions between the layer and ionizing radiation. The light is readout by the HCAL electronics and converted into a reconstructed energy. The HCAL is organized into a series of towers whose size very with the angle from the beam pipe.

Coverage in the forward region, $3 < |\eta| < 5$, is provided by the forward hadron calorimeter (HF). The name is somewhat deceptive, as the calorimeter is responsible for measuring the energy of both hadronic and electromagnetic particles. The proximity of the HF to the beam pipe constrained the design due to the extremely high radiation levels present in this region. The towers of the HF are blocks of steel with embedded quartz

fibers. Energy is reconstructed from light signals due to Čerenkov radiation, emitted as charged particles pass through the fibers. Each tower contains a long and short fiber. The long fiber extends closer to the face of the tower and is important for reconstructing the energy of electromagnetic particles, for which the electromagnetic shower dissipates quickly in the material. The short fiber provides additional sensitivity for reconstructing the energy from hadronic showers.

The muon system comprises the outermost component of the CMS detector. Muons are minimum ionizing particles over much of the momentum range of interest at the LHC. On average, a muon deposits only a small fraction of its energy in the calorimeter. Additionally, muons are unlikely to decay within the detector, simplifying their reconstruction. The muon system is comprised of three subsystems — drift tubes (DTs), cathode strip chambers (CSCs), and resistive plate chambers (RPCs) — that all work on similar principles. Each device contains a significant amount of inactive material to increase the likelihood of differentiating between a muon and hadron, and chambers of gas provide a material that can be ionized by the passage of a charged particle, allowing for the presence and path of the particle to be reconstructed [2]. Good muon identification and momentum resolution were driving requirements of the detector design and enable sensitive searches for the Higgs boson in the $H \to ZZ \to 4\mu$ channel. This channel is particularly sensitive for a high mass Higgs.

Before moving on, it is worth making a brief aside to discuss units that will appear frequently in this thesis. Particle physics represents the probability for an interaction to take place — the interaction cross section (σ) — by the effective area presented by a target to an incident particle. The cross section is expressed in units of barns, where one barn is equal to 10^{-24} cm^{-2}, the approximate size of a uranium nucleus. The total amount of data collected is then traditionally reported as the total integrated luminosity ($\mathcal{L}$), a quantity with units of inverse area given by the relation

$$N = \mathcal{L}\sigma, \tag{1.2}$$

where N is the total number of collisions, or interactions, occurring in dataset with integrate luminosity $\mathcal{L}$ for a process with an interaction cross section σ. For example, the total pp interaction cross section from Figure 1.2 is approximately 100 mb. Thus, a 1 fb^{-1} dataset represents 10^{14} pp collisions. Similarly, with a cross section for inclusive $t\bar{t}$ production of 157.5 pb, over 150000 $t\bar{t}$ pairs would be produced in a dataset of 1 fb^{-1}.

1.2 Motivation

Although Higgs physics played an important role in detector design considerations, the strengths of CMS allow for sensitive searches for physics beyond the SM, while the weaknesses present opportunities for creative license. The work discussed in this thesis was motivated by both. The physics analysis discussed here attempts to take advantage of the strengths in lepton reconstruction to search for previously unobserved physics. Work on improving $\not\!\!E_T$ reconstruction uses the resolution of the tracker to improve upon shortcomings of the calorimeters.

The physics analysis is motivated by theoretical understanding and experimental observations and conditions. Figure 1.2 shows the production cross section for a number of SM processes, as well as the predicted cross section for a Higgs boson with varying hypothetical mass, as a function of the center of mass energy of the colliding beams. Comparing the relative size of the total inelastic cross section, dominated by QCD processes, with that for W production, the leading source of prompt, isolated leptons, it is apparent that leptons are rare in the SM. Dilepton final states are significantly more rare. The dominant source is $Z \to \ell^+\ell^-$. Additional sources include $t\bar{t}$, WW, WZ and ZZ production, all with smaller production cross sections. Most of the dilepton pairs from these have opposite electric charge. Experimentally, leptons leave much cleaner signatures in the detector that are simpler to reconstruct with smaller rates of mismeasurement. The combined effect is much smaller backgrounds for searches in dilepton final states, particularly same-sign dilepton final states, compared to single lepton and fully hadronic analyses.

The evidence for astrophysical dark matter intimates that the Standard Model is incomplete. Although direct detection experiments have yet to make an observation, indirect measurements suggest that whatever its nature, dark matter is weakly interacting. Thus, a significant $\not\!\!E_T$ signature should accompany other final state objects in direct dark matter production scenarios at the LHC [5, 6].

The relative rarity of dilepton final states presents a potential difficulty for early observation. With the first inverse femtobarn of data, such an analysis only has sensitivity to processes with sufficiently large production cross section times branching ratio. As the branching ratio to electrons and muons is approximately 20% total for W decays and 3% for Z decays, a large cross section is needed. Particles with large production rates tend to be produced via the strong interaction. Thus, final state leptons are expected to

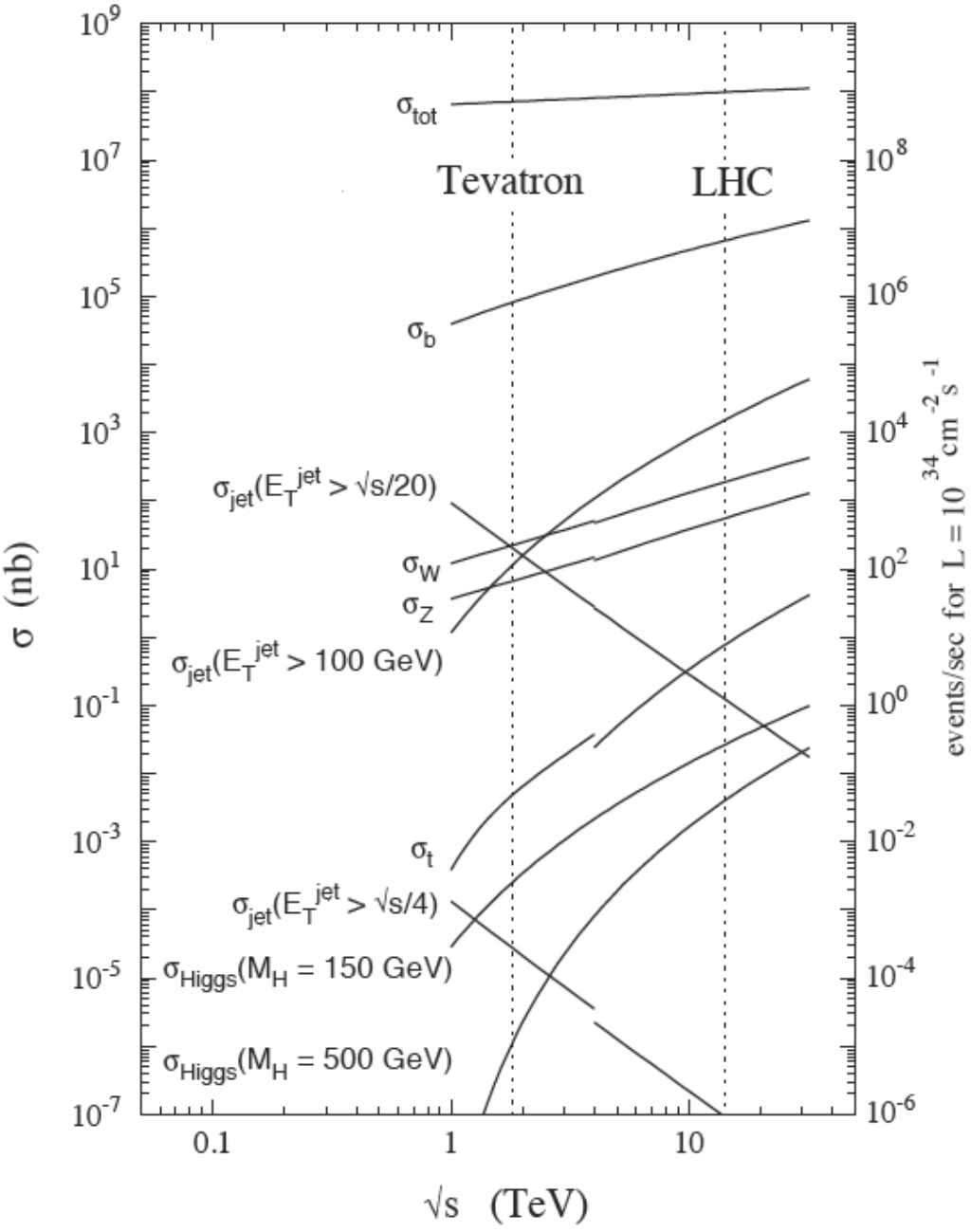

Figure 1.2: Cross sections for some Standard Model processes and for different potential Higgs masses. The production rate for new physics (i.e., Higgs production) is much smaller than that for SM electroweak and $t\bar{t}$ production, the main backgrounds to new physics searches in dilepton final states. The cross sections of SM processes giving final state leptons are in turn many orders of magnitude smaller than QCD cross sections. From [4].

be accompanied by significant hadronic activity in the form of high-p_T jets.

The theoretical and experimental considerations discussed above are used to define a search for new physics in a final state with same-sign dileptons, $\not\!\!E_T$ and significant hadronic activity. The analysis is designed to be as model independent as possible and to have sensitivity to deviations from SM expectations with the approximately 1 fb^{-1} of available data.

As discussed above, astrophysical evidence for dark matter suggests that $\not\!\!E_T$ may be an important signature in new physics searches at the LHC. The significant material budget of the tracker, non-linear response of the calorimeters that do not have depth segmentation, and the large encompassing magnetic field present challenges for $\not\!\!E_T$ reconstruction with the CMS detector. We developed a new $\not\!\!E_T$ algorithm that improves upon the traditional algorithm that uses only calorimeter information. The new track-corrected $\not\!\!E_T$ algorithm (tc$\not\!\!E_T$) makes use of the superior resolution of the silicon tracker compared to the calorimeter at energy scales $\mathcal{O}$(1 GeV). The tc$\not\!\!E_T$ reconstruction also provides an important independent cross-check of other $\not\!\!E_T$ reconstruction algorithms to help gain confidence in the interpretation of events with significant $\not\!\!E_T$ signatures.

1.3 Overview of the Thesis

This thesis describes a search for new physics in a final state with same-sign dileptons, $\not\!\!E_T$ and significant hadronic activity. The analysis was performed using 0.98 fb^{-1} of 7 TeV data. No excess with respect to Standard Model expectations is observed. Model independent limits are set on new physics. An efficiency model is developed that allows non-CMS collaborators to interpret the results in the context of a broad range of physics models.

Chapter 2 describes the $\not\!\!E_T$ algorithms used by the CMS experiment, with a detailed discussion of the tc$\not\!\!E_T$ algorithm developed as part of this thesis. The improved performance of the tc$\not\!\!E_T$ algorithm is demonstrated in simulation. This is followed by a discussion of work done to commission and understand the $\not\!\!E_T$ algorithms with the first collisions. This work uncovered several sources of anomalous signals in the calorimeter. Algorithms were developed to correct for the presence of the anomalous signals. Both the tc$\not\!\!E_T$ algorithm and the cleaning algorithms continue to be used by the CMS collaboration.

Chapter 3 discusses the measurement of the $t\bar{t}$ cross section in the dilepton final

state with early data. The measurement was the first top physics result produced by the LHC and one of the first results to be produced by the CMS collaboration. This analysis served two important purposes. First, the $t\bar{t}$ dilepton final state involves reconstructed leptons, jets and missing transverse energy. These physics objects are important to many new physics analyses, and thus it is necessary to first develop an understanding of them using a SM process that is well understood from previous experiments. Second, $t\bar{t}$ is a significant background to many searches for physics beyond the SM. Thus, it is necessary to develop a detailed understanding of top physics at the new energy scale.

The remainder of this thesis discusses a search for new physics with final states with same-sign dileptons, $\not\!\!E_T$ and hadronic activity. Chapter 4 describes the analysis selections. Physics motivation is provided whenever possible. Signal regions are defined based on considerations of expected SM backgrounds and general models of interest to the community. Control regions are defined that are enriched in SM backgrounds. These regions are used to gain confidence in the data-driven background estimates. Data-driven methods used to estimate backgrounds are discussed in Chapter 5. Selection efficiencies and systematic uncertainties necessary to set upper limits are described in Chapter 6. Results are discussed in Chapter 7. Here, observed event yields in data are provided together with background estimates. Model independent limits are set. Some models standard to the CMS experiment are considered as examples. Several points in model parameter space are excluded at the 95% confidence level.

Each chapter begins with a plain-English introduction that attempts to describe for the non–CMS member what follows in the remainder of the chapter. The introduction ends with a summary of the main results presented in the sections that follow. Attempts are made, whenever possible, to provide physics motivation for the choices made.

Chapter 2

Missing Transverse Energy

2.1 Introduction

Neutrinos and other, hypothetical, weakly interacting particles do not interact with the detector and thus cannot be directly measured. However, information on the presence of these particles can be inferred by exploiting the expected conservation of momentum in a plane transverse to the colliding beams. The two-dimensional vector representing this imbalance is referred to as the missing transverse momentum. The magnitude of this vector is referred to as the missing transverse energy, $\not\!\!E_T$ [7].

The missing transverse momentum vector is important in many physics analyses, such as the reconstruction of the mass of the W boson. This analysis requires an accurate reconstruction of both the magnitude and the direction of the $\not\!\!E_T$, and the reconstruction should aim for optimizing resolution. For many other measurements, though, the missing transverse energy is sufficient. This use case arises frequently in measurements of SM processes such as the $t\bar{t}$ cross section measurement and in studies of potential new physics, such as the search for dark matter candidates, where the $\not\!\!E_T$ is used as a discriminating variable, not as an input to the reconstruction of another quantity. New physics searches require $\not\!\!E_T$ reconstruction algorithms aimed at reducing tails due to mis-measurement rather than at optimizing resolution.

The reconstruction of $\not\!\!E_T$ is impacted by the performance of all detector components. Reconstruction of $\not\!\!E_T$ with the CMS detector is complicated by the large material budget of the tracker, the non-linear response of the calorimeter system, the presence of anomalous readout from the calorimeter electronics and the strong encompassing mag-

netic field. The importance and complexity of reconstructing $\not\!\!E_T$ has resulted in the development of a variety of algorithms that address different aspects of the problem.

This chapter provides an overview of the different algorithms used to reconstruct missing transverse momentum with the CMS detector. It begins with a discussion of calorimeter-based $\not\!\!E_T$ (calo$\not\!\!E_T$) and the reconstruction of calorimeter towers that are the building blocks for this object. It then discusses corrections to the basic calo$\not\!\!E_T$ objects for muons as well as for the jet energy scale (JES). The next section provides a brief overview of a particle flow-based algorithm for $\not\!\!E_T$ reconstruction. Section 2.6 provides a detailed description of the track-corrected $\not\!\!E_T$ algorithm developed as part of the work for this thesis. This section describes the selection and derivation of the inputs to the algorithm as well as the implementation. It is followed by a section demonstrating the expected performance of the tc$\not\!\!E_T$ algorithm in simulation for events with real and mismeasured missing transverse momentum. The following section discusses work done on the commissioning of $\not\!\!E_T$ with early data. The focus is on the identification of anomalous calorimeter energy and the development of algorithms to remove these signals. The chapter ends with a brief section discussing some aspects of the performance of the $\not\!\!E_T$ algorithms in data compared with that expected from simulation.

2.2 Calorimeter–based $\not\!\!E_T$

Missing transverse momentum is typically calculated from energy deposits in the calorimeter. A description of the relevant details of the reconstruction of calorimeter towers appears at the end of this section. From the collection of calorimeter towers, the calculation of the missing transverse momentum vector is straightforward:

$$\vec{\not\!\!E}_T = -\sum_k \left(E_k \sin\theta_k \cos\phi_k \hat{i} + E_k \sin\theta_k \sin\phi_k \hat{j} \right) = \not\!\!E_x \hat{i} + \not\!\!E_y \hat{j}, \tag{2.1}$$

where the sum runs over calorimeter towers and $\hat{i}, \hat{j}$ are unit vectors along the x and y axes, respectively. The overall minus sign accounts for the distinction between what is observed and what is missing.

In the absence of systematic effects the individual components $\not\!\!E_x$, $\not\!\!E_y$ are expected to follow Gaussian distributions with mean 0 and standard deviation σ. However, detector effects can introduce an offset and non-Gaussian tails that are challenging to model. The distribution of $\not\!\!E_T$ is positive-definite and tends to exhibit a Gaussian core

with an exponential tail, the width of the distribution increasing with the total transverse energy in the detector.

2.2.1 Reconstruction of Calorimeter Towers

Energy deposits reconstructed in each sub-detector of the calorimeter, referred to generically as rechits, are combined into a collection of calorimeter towers used to seed the reconstruction of higher-level quantities such as $\not\!\!E_T$. In the barrel and endcap, a calorimeter tower is built from a single HCAL rechit and a 5×5 grid of ECAL crystals. In the forward region, a calorimeter tower is the sum of the energy reconstructed in a long and a short fibers. Only rechits passing the so-called Scheme B thresholds are used in the reconstruction of calorimeter towers. Table 2.1 summarizes the thresholds applied to rechits in different sub-detectors. Rechits reconstructed in the hadronic outer calorimeter, HO, are not used by default.

Table 2.1: Scheme B thresholds are applied to rechits from the various calorimeter sub-detectors during the reconstruction of calorimeter towers. The HO threshold is 1.1 GeV in the central ring of the detector and 3.5 GeV elsewhere.

Sub-detector	Threshold (GeV)
EB	$E > 0.07$ $\sum_{5\times5} E > 0.2$
EE	$E > 0.3$ $\sum_{5\times5} E > 0.45$
HB	$E > 0.7$
HE	$E > 0.8$
HO	$E > 1.1$, 3.5
HF	$E > 0.5$, long $E > 0.85$, short
Tower	$E_T > 0.3$

2.3 Correcting $\not\!\!E_T$ for Muons

At the relativistic scales at which charged particles are typically produced at the LHC, ionization and atomic excitation are the dominant modes of energy loss. Figure 2.1 shows the average energy loss of positively charged muons in copper, as a function of the

muon momentum. The distribution evidences a broad minimum covering several decades of momentum that overlaps with the bulk of the momentum distribution of muons produced in electroweak interactions such as in the decay of W and Z bosons. A particle displaying this behavior is called "minimum-ionizing". Radiative effects only begin to become significant when the momentum reaches several hundred GeV. This contrasts sharply with electrons, nearly 200 times lighter than muons, for which bremsstrahlung is the principal mechanism of energy loss.

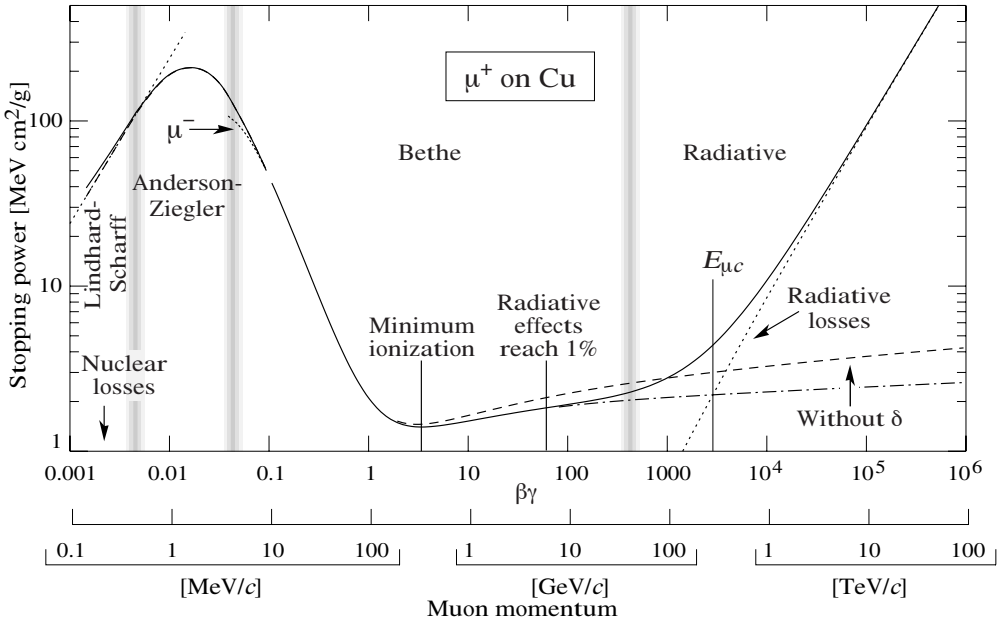

Figure 2.1: Stopping power of Cu for positively charged muons [3]. Muons from electroweak processes are typically produced with momenta of $\mathcal{O}(100)$ GeV. This momentum scale is still in the region where muons are approximately minimum ionizing, before radiative effects become significant.

Excepting electrons, most charged particles produced in the CMS detector are minimum ionizing. However, muons present a more significant challenge than do, for example, the similarly massive pions. A pion is a charged hadron whose constituent quarks participate in colored interactions. Although charged pions in vacuum decay predominantly to muons, it is a relatively slow process as the decay proceeds electromagnetically. At the relativistic energies at which they are produced in the CMS dectector, the lifetime of charged pions is sufficient that they usually do not decay before reaching the HCAL, where they interact and deposit their energy.

As a muon deposits only a few GeV of energy in the calorimeter, independent of

its momentum, a simple algorithm based upon (2.1) will fail to properly reconstruct $\not\!\!E_T$ in events containing muons. However, the momentum of a muon is well reconstructed in the tracker. Thus, the $\not\!\!E_T$ can be corrected by adding the energy a muon deposited in the calorimeter and subtracting the corresponding momentum measured in the tracker:

$$\vec{\not\!\!E}_T = \vec{\not\!\!E}_T + \sum_{\mu} \vec{E}^{\mu}_{T,\text{calo}} - \sum_{\mu} \vec{p}^{\mu}_{T}. \tag{2.2}$$

The muon collection over which the sum in (2.2) is performed contains significant contamination from fakes, mainly from hadrons that punch through the calorimeter into the muon system. Generator studies using simulated QCD events show that only approximately 10% of reconstructed muons with $p_T > 10$ GeV come from a b or c quark or a decay in flight.

To avoid generating fake $\not\!\!E_T$ by correcting for a charged hadron mistakenly reconstructed as a muon, the sum in (2.2) should be restricted to real muons. Absent generator information in data, the differentiation is made by restricting the sum to muons passing the selections in Table 2.2.

Table 2.2: Default selection for muons used to correct $\not\!\!E_T$. The selection is configurable.

	Selection
type	global and tracker
p_T	> 10 GeV
$\lvert\eta\rvert$	< 2.5
$\lvert d_0\rvert$	< 2 mm
$\chi^2_{\text{global}}/n(\text{dof})$	< 10
$N^{\text{Si}}_{\text{hits}}$	> 10
$N^{\text{SA}}_{\text{hits}}$	> 0

The p_T requirement on muons is designed to select a kinematic region where there is significant gain from correcting the $\not\!\!E_T$. Muons are reconstructed in the CMS detector using three different algorithms. Standalone muons are reconstructed using only hits in the muon system and have the poorest resolution. Global muons are seeded by a standalone muon and reconstructed by simultaneously fitting hits in the silicon tracker and the muon system. Tracker muons are seeded by a silicon track and use hits in the muon system only for identification. All three algorithms use the beam spot position as a constraint. Most good muons are reconstructed as all three types, although the efficiency of the standalone reconstruction falls below 10 GeV. Tracker muons provide the best

combination of efficiency and resolution at the typical momentum scales of interest in this analysis. The momentum resolution for muons at this scale is typically on the order of a few percent. At higher momenta, improved performance can be obtained by using the global fit.

Requiring that a muon is reconstructed by both the global and tracker algorithms helps to reduce contamination from fakes. The η selection comes from the extent of the tracking system itself. Farther forward there are fewer tracking layers and the reconstructed momentum is less reliable. The impact parameter (d_0), χ^2, and number of silicon hits requirements select tracks for which the reconstruction of the muon and matching track are reliable. The final requirement of at least 1 hit in the muon system reduces contamination from fakes due to punch through. Figure 2.2 shows that these selections increase the purity of the sample from about 10% to over 70%.

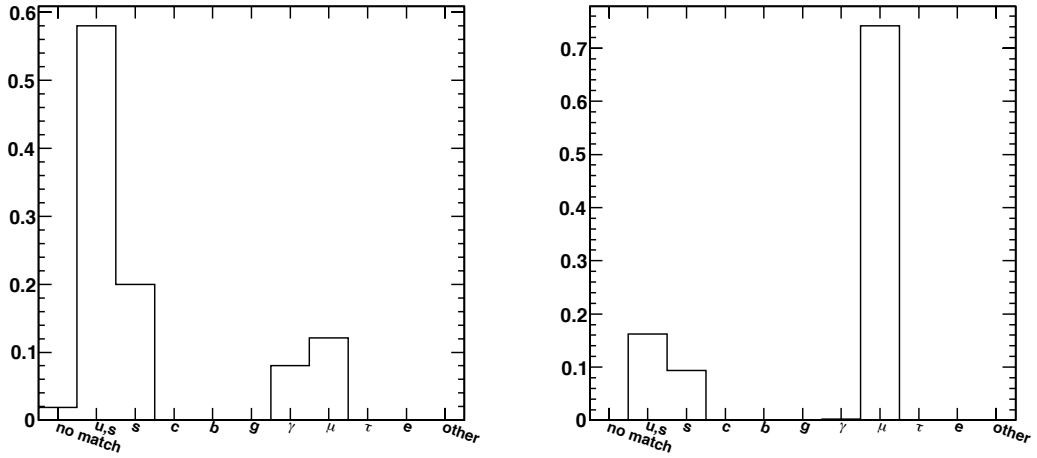

Figure 2.2: Left: Classification of reconstructed muons by ΔR matching to generator level particles. No additional requirements are placed on the muons. Right: Muon classification after selections in Table 2.2 are applied. The sample is QCD Monte Carlo with $\hat{p_T}$ = 80–120 GeV.

The final consideration in the $\not\!\!E_T$ correction for muons is isolation. If a muon is isolated, the $E^{\mu}_{T,\mathrm{calo}}$ contribution to the $\not\!\!E_T$ correction is given by a sum over towers traversed by the muon. Otherwise, the contribution is taken to be the median value of the sum expected from the muon were it isolated. Correcting a non-isolated muon as isolated would end up over-correcting the $\not\!\!E_T$ since the calorimeter deposits surrounding the muon would not be due solely to the muon itself. A muon is defined to be isolated if

$$\begin{aligned} \sum_{\text{tracks}} p_T &< 3 \text{ GeV} \\ \sum_{\text{ECAL+HCAL}} E_T &< 5 \text{ GeV}, \end{aligned} \tag{2.3}$$

where the first sum is over tracks in a cone of $\Delta R < 0.3$ with respect to the muon, and the second sum is over calorimeter towers in a cone of the same size. Only the ECAL and HCAL components of the tower are included in the latter sum; i.e., deposits in the HO and HF are excluded. The median value was derived from a high statistics single-muon MC sample and is parameterized as a function of the p_T and η of the muon [8].

2.4 Type–I $\not\!\!E_T$ Corrections

The energies of reconstructed jets tend to be under-measured, most significantly due to the non-linear response of the calorimeter to neutral and charged hadrons because of its non-compensating nature. Corrections have been devised to make up for this under-measurement and move the average energy scale of jets closer to unity. These corrections are derived from MC by comparing the response of a reconstructed jet to the associated jet formed by clustering particles at the generator level. The standard correction involves two components. The first, known as the L2 correction, flattens the jet response in η to account for the non-uniformity of the detector. An additional multiplicative factor, the so-called L3 correction, is then applied to account for the non-linear response of the detector and move the average energy scale closer to unity. Residual corrections are derived by comparing uncorrected jet response in data to simulation [9].

These jet energy scale corrections can be applied to correct $\not\!\!E_T$ in a manner analogous to the correction for muons. The uncorrected energies of jets are removed from the calorimeter and replaced with corrected quantities:

$$\vec{\not\!\!E}_T' = \vec{\not\!\!E}_T + \sum_{\text{jets}} \vec{p}_T^{\,\text{uncor}} - \sum_{\text{jets}} \vec{p}_T^{\,\text{cor}}. \tag{2.4}$$

Only jets with uncorrected $p_T < 20$ GeV for which at least 10% of the jet energy was reconstructed in the HCAL are considered in the sum. These requirement serves two purposes. They avoid correcting for jets arising from partons that fragmented predominantly into electromagnetic particles, for which the energy was already well–measured

in the ECAL, resulting in a jet whose energy scale is already close to 1. They also avoid correcting for fake jets arising principally from noise in the ECAL electronics, which would introduce additional fake energy into the $\not\!\!E_T$ sum [10].

2.5 Particle Flow $\not\!\!E_T$

A different approach to the standard calorimeter-based $\not\!\!E_T$ calculation can be taken by combining information from the various sub-detectors in an attempt to reconstruct the individual particles produced in a collision. Energy deposits in the ECAL and HCAL are clustered and links between clusters and tracks are made. A track and its linked clusters is identified as a charged particle while stand-alone clusters are reconstructed as neutrals [11].

$$\mathrm{pf}\vec{\not\!\!E}_T = -\sum_i \vec{p}_T^{\,i} \tag{2.5}$$

The missing transverse momentum is calculated by adding as vectors the transverse momenta of all reconstructed particles as shown in (2.5). Despite the better resolution of particle flow-based $\not\!\!E_T$ compared to calorimeter $\not\!\!E_T$, the method still suffers from under-measurement due to the non-linear response of the detector. Type-I jet energy scale corrections have been developed, analogous to those discussed in Section 2.4. Here the type-I corrections are applied to particle flow-based jets. These jets are clustered using the same algorithms as for calorimeter-based jet, but with reconstructed particles rather than calorimeter towers as input. The p_T threshold of jets to which the correction is applied is lowered to 10 GeV. The cut on the electromagnetic fraction remains the same.

2.6 Track Corrected $\not\!\!E_T$

The details of any new physics drive the requirement on how well the $\not\!\!E_T$ must be understood. If new physics presents with $\not\!\!E_T$ much larger than that expected in the standard model, the particular algorithm used is less important. However, if new physics exists close to the SM, understanding and reducing the tails of $\not\!\!E_T$ distributions will be vital. The combination of the non-linearity of the CMS calorimeter, the large material budget of the tracking system and the strong encompassing magnetic field result in a pion

with a p_T of 10 GeV depositing an average of approximately 6 GeV in the calorimeter, with the response degenerating further for softer tracks.

Our interest in new physics searches led us to develop a new track-corrected $\not\!\!E_T$ (tc$\not\!\!E_T$) reconstruction algorithm aimed at mitigating the above detector effects and reducing the tail of the $\not\!\!E_T$ distribution due to mis-measurement. At energy scales of a few GeV, the tracker resolution is superior to that of the calorimeter. Thus, the calorimeter-only measurement of $\not\!\!E_T$ can be improved by replacing the expected energy depositions of well reconstructed tracks, assumed to be pions, with the corresponding momenta measured in the tracker. The expected energy deposition is mapped in a response function in bins of the η and p_T of the track.

Mitigating the effects discussed above requires a detailed understanding of the interaction of charged particles with the detector. Leptons, such as electrons and muons, do not participate in colored interactions and thus have simpler, cleaner signatures in the detector, allowing for independent treatments that provide reliable corrections to the $\not\!\!E_T$. Hadrons, however, are strongly interacting and abundant and the properties of the CMS detector make devising a dependable correction more difficult.

The tc$\not\!\!E_T$ algorithm addresses these difficulties by relating a poorly measured quantity, the calorimeter deposition of low-energy charged hadrons, to something well known, the momentum obtained from the tracker system. The map between the two is the response function (RF). The RF is determined using a Monte Carlo (MC) sample of single, charged pions. Tracking and calorimeter information is extracted and the detector response, E/p, is calculated for each pion that passes a selection on kinematics and the quality of the reconstruction.

2.6.1 Selection of Pions

If the response function is to reliably represent the performance of the calorimeter, it is necessary to select particles that are well reconstructed. To this end, only pions with reconstructed tracks passing the selections in Table 2.3 are considered. The quality flag is a minimal categorization set during the reconstruction that reflects the reliability of the reconstruction [12]. The p_T was chosen as a threshold above which the response and momentum of a pion are sufficiently high that the potential to introduce fake $\not\!\!E_T$ by correcting for a pion whose momentum was poorly reconstructed exceeds the expected benefit to be gained by correcting the $\not\!\!E_T$ using the response function. The η require-

ment reflects the tracker geometry while the χ^2 and relative p_T error cuts reject a track whose measured momentum is unreliable. Selections on the impact parameter and the number of silicon hits also serve to reject poorly reconstructed tracks. The thresholds on these requirements differ depending on the iteration in which the track was constructed. Figure 2.3 shows N−1 distributions for tracks reconstructed from a single pion Monte Carlo sample, requiring that the tracks are flagged as high purity.

Table 2.3: Requirements used by the tc$\not\!\!E_T$ algorithm to select well-reconstructed tracks.

Selection	Value
quality	high purity
p_T	< 100 GeV
$\|\eta\|$	< 2.5
χ^2/n(dof)	< 5
$\delta p_T/p_T$	< 0.2
N^{Si}_{hits}	> 8, iterations 0-3
	> 5, iterations 4,5
$\|d_0\|$	$< 2.$ mm, iterations 0-3
	$< \infty$, iterations 4,5

2.6.2 Track Reconstruction

CMS employs an inside–out fitting procedure to exploit the superior resolution and low occupancy of the inner pixel layers while reducing inefficiency arising from pion–nucleon inelastic collisions due to the significant material budget of the tracker [13]. Seeding is hybrid, preferring pixel triplets but permitting the use of pairs of pixels, pixel–strip pairs and even allowing strip-only seeding. Pixel triplet seeding accounts for approximately 90% of reconstructed tracks, with the remainder using one of the other options together with either a primary vertex or beam spot constraint.

Track candidates are seeded as described and a trajectory is built using a Kalman filter to navigate outward and select additional hits based on χ^2 compatibility between the measured and expected hit position [14]. Candidates are then refit inside–out using the full set of hit parameters and smoothed by running the filter in reverse, outside–in. If a primary vertex or beam spot constraint was used at the seeding stage it is removed before performing the final fit to avoid introducing a bias.

The criteria for hit compatibility are relatively loose in an effort to maximize

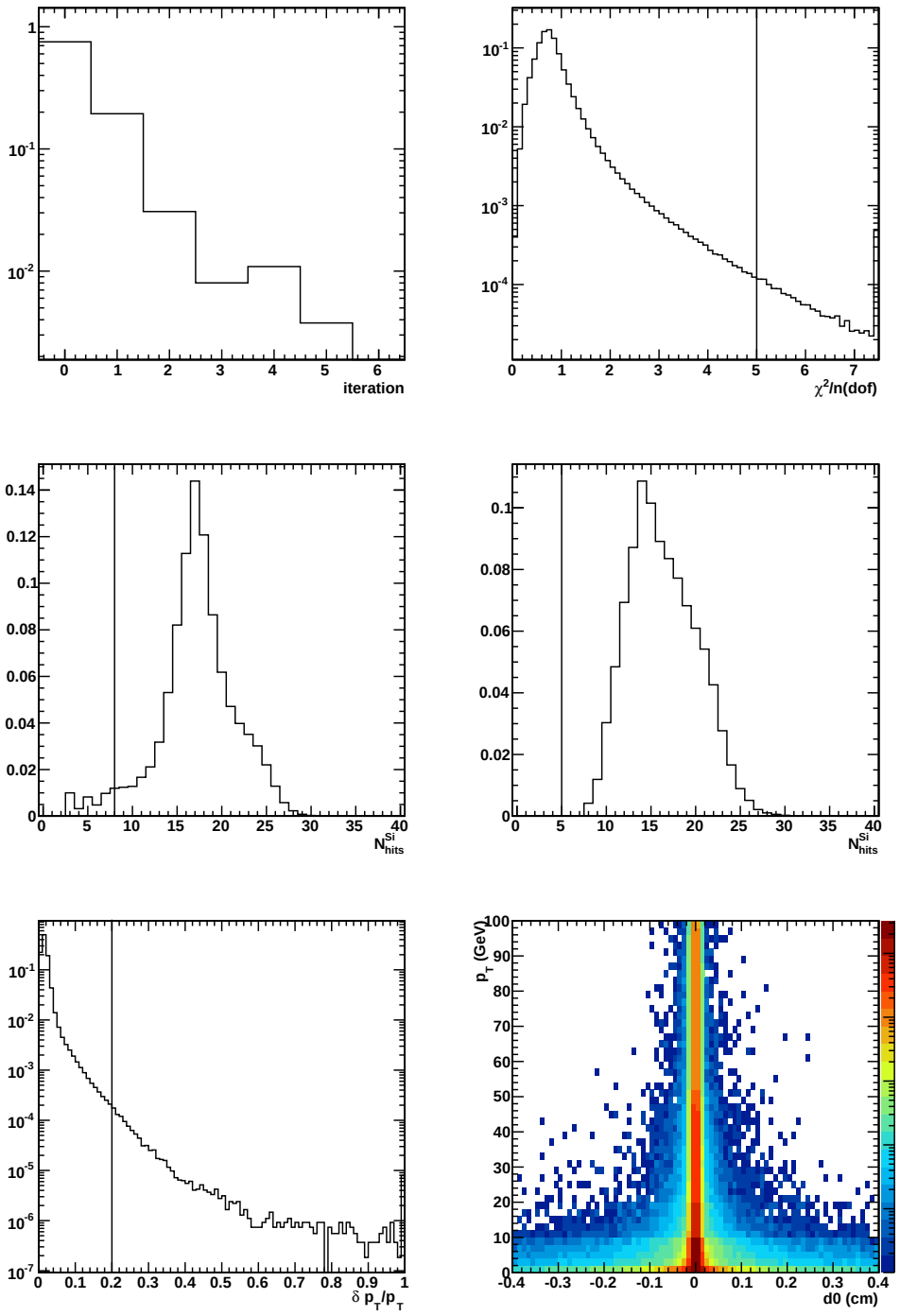

Figure 2.3: Tracking distributions made from a single pion Monte Carlo sample. Vertical lines indicate the selections used by the tc$\not\!\!E_T$ algorithm. The distribution in each panel is made after applying all selections in Table 2.3 but omitting selection on the plotted quantity.

the efficiency, causing significant track duplication. After all candidates are fit, ambiguities are resolved by comparing the fraction of shared hits between pairs of tracks. Equation (2.6) defines the hit-sharing criteria used.

$$f_{\text{shared}} = \frac{N_{\text{shared}}^{\text{hits}}}{\min\left(N_1^{\text{hits}},\ N_2^{\text{hits}}\right)} \tag{2.6}$$

If $f_{\text{shared}} > 0.5$, the track with the fewer hits is discarded. If the two tracks have the same numbers of hits, the one with the larger χ^2 is removed. This procedure is repeated until $f_{\text{shared}} < 0.5$ for all tracks.

An iterative approach is used to achieve high efficiency while minimizing the fake rate. The tracking routine described is performed and at the end the set of hits used to build good tracks is removed from the total collection available. The seeding and building criteria are then loosened and the procedure is performed again. Six passes are made in total with changes in the seeding being the most significant difference. The output of each iteration is merged to form the full track collection. Table 2.4 details the seeding configuration used for each iteration. Table 2.5 provides additional selections applied during the track building stage of each iteration [13, 12].

Table 2.4: Seeding configuration for each tracking iteration. Here, d_0 and z_0 are calculated with respect to the center of CMS, expect for the z_0 cut of iteration 1 which is calculated with respect to the production vertex [13].

Iteration	Seeding Layers	p_T cut (GeV)	d_0 cut (cm)	z_0 cut (cm)
0	pixel triplets	0.5	0.2	15.9
1	pixel pairs	0.9	0.2	0.2
2	pixel triplets	0.2	0.2	17.5
3	pixel pairs	0.35	1.2	7.0
4	TIB 1+2 TID/TEC ring 1+2	0.5	2.0	10.0
5	TOB 1+2 & TEC ring 5	0.8	5.0	10.0

2.6.3 Definition of the Response

A reconstructed track needs to be associated with energy deposits in the calorimeter in order to calculate a response. For each pion that passes the selections in Table 2.3, the momentum p measured at the vertex serves as a seed to an analytical propagator that extrapolates the trajectory of the pion from the vertex to the face of the ECAL.

Table 2.5: Cuts applied during track building for each iteration. In addition to a minimum p_T cut, N_{hit} is the minimum number of valid hits, N_{lost} the maximum number of allowed invalid hits and $N_{\text{hit}}^{\text{rebuild}}$ the minimum number of hits needed in the inside–out step to trigger an outside–in fit [13].

Iteration	p_T cut (GeV)	N_{hit}	N_{lost}	$N_{\text{hit}}^{\text{rebuild}}$
0	0.3	3	1	5
1	0.3	3	1	5
2	0.1	3	1	5
3	0.1	4	0	5
4	0.1	7	0	5
5	0.1	7	0	4

The propagator uses the measured CMS magnetic field, but does not account for multiple scattering in the tracker. The energy deposited in the calorimeter is obtained by summing towers in a cone of $\Delta R = 0.5$, centered on the track–calorimeter intersection point.

$$\text{response} = \left(\sum_{\Delta R < 0.5} E_{\text{tower}} \right) / p \tag{2.7}$$

The response is the ratio of the energy summed in the cone and the magnitude of the track momentum measured at the vertex. Figure 2.4 shows the detector response for central pions with p_T between 20 and 30 GeV for several different cone sizes. The response increases asymptotically with the cone size. The use of single particle Monte Carlo provides the flexibility to collect energy over a large number of towers without concern of contamination from other sources.

Choices made in defining the response introduce three potential sources of uncertainty. If the cone size is too small, the summed energy will underestimate the true amount of energy deposited in the calorimeter, and the response will be artificially low; given the small difference in response observed between cones of size $\Delta R = 0.3$ and 0.5 in Figure 2.4, this effect isn't important. The cone itself is centered around a point determined using the extrapolated trajectory. The analytical propagator used to perform the extrapolation accounts for the magnetic field but not the material budget. For high-p_T tracks this should have little effect. For low p_T tracks that undergo significant multiple scattering, however, the use of a simple propagator could result in the cone being off–center. Given the large cone size and the significant variation in response for

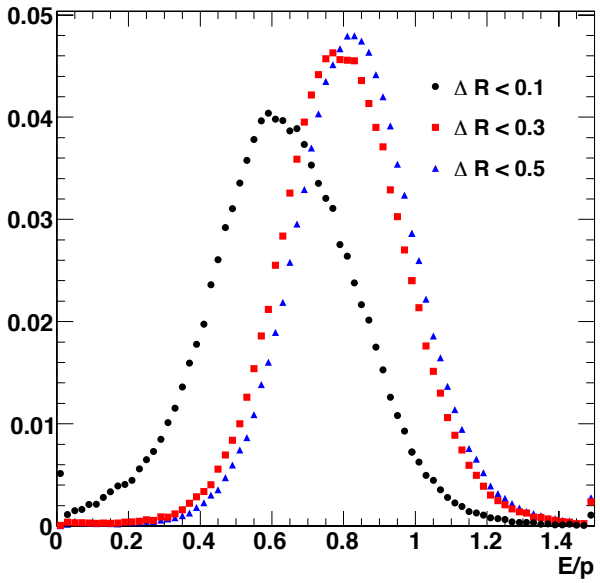

Figure 2.4: Response of simulated charged pions with $|\eta| < 1$ and p_T between 20 and 30 GeV using cones of size $\Delta R = 0.1$, 0.3 and 0.5.

low-p_T tracks due to other effects, we do not think this is a concern. Finally, the energy deposited by the pion is determined by summing calorimeter towers. This introduces a sensitivity to thresholds applied in the construction of calorimeter towers and to zero suppression applied at readout. However, thresholds are necessary to suppress noise and limit the number of channels that need to be readout and reconstructed. While studies show that the $\not\!\!E_T$ does have some sensitivity to readout thresholds on calorimeter deposits, this sensitivity is expected to be sub-dominant to other detector effects.

Figure 2.5 shows response distributions of pions in the barrel and endcap for several different p_T ranges. Two general features are apparent:

- The response improves with increasing p_T.
- The response is better in the endcap than in the barrel for a constant p_T interval.

Better response in the endcap is likely due to the larger material budget of the detector in that region and the relatively higher p for a given p_T at higher η. Constructing the response functions in bins of p rather than p_T was considered, but no significant difference was observed.

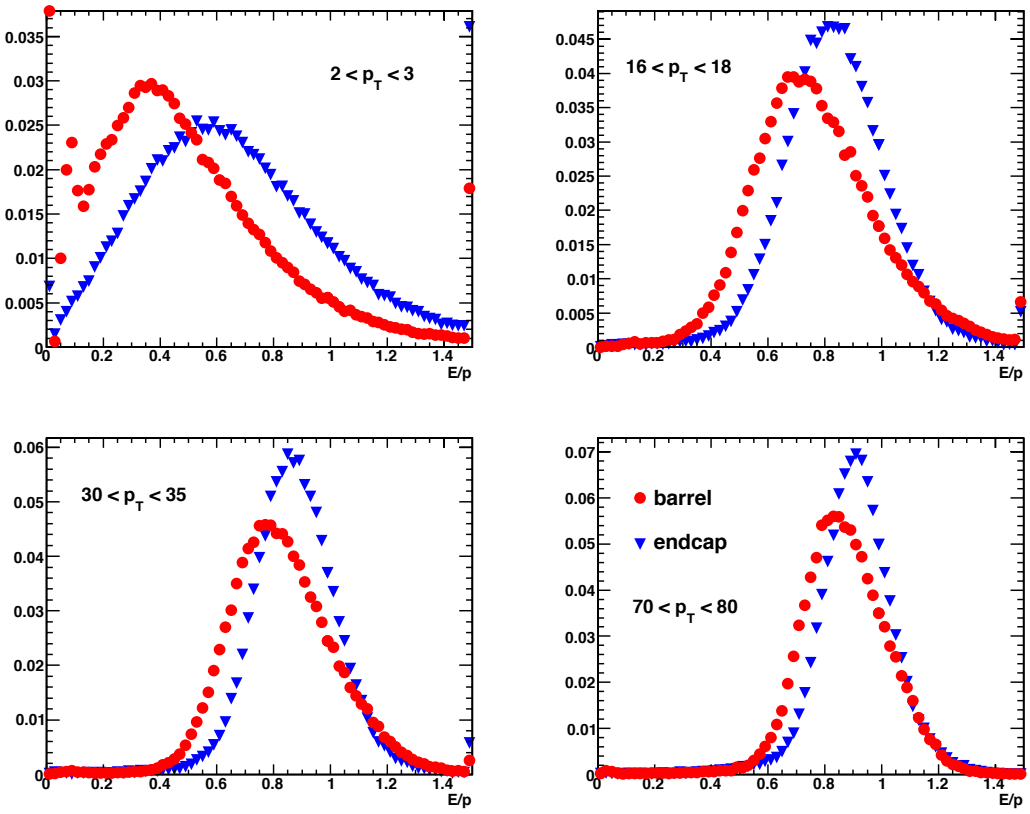

Figure 2.5: Calorimeter response for charged pions, separately in the barrel and endcap, for several different p_T intervals. The response is systematically better in the endcap than in the barrel and improves with increasing pion p_T.

2.6.4 Properties of the Response Function

The response, as defined in (2.7), is calculated for each pion passing the selections in Table 2.3. A distribution is constructed in each bin of η, p_T and a Gaussian fit is performed on the interval $(\mu - 1.5 \cdot \sigma,\ \mu + 1.5 \cdot \sigma)$, where μ is the mean and σ the RMS of the distribution. The response is taken to be the mean of the Gaussian fit. We investigated other methods of extracting the response function including using the mode and mean of the distribution. The mode was found to be unstable at low-p_T and led to a response that fluctuated unnaturally between bins of η, p_T. The average and fit are comparable, as expected, since the fit acts approximately like a truncated mean. However, we opt to use the fit, as an average over the full interval can be unduly influenced by outliers in the tail. Table 2.6 compares the response found using the different methods in several bins of η, p_T.

Table 2.6: Comparison of response extracted using the mode, mean and Gaussian fit methods.

Interval		Mean	Fit	Mode
$p_T = 2.5$,	$\eta = 0.05$	0.57	0.51	0.42
$p_T = 2.5$,	$\eta = 2.25$	0.67	0.61	0.54
$p_T = 17$,	$\eta = 0.05$	0.80	0.76	0.70
$p_T = 17$,	$\eta = 2.25$	0.84	0.83	0.82
$p_T = 32.5$,	$\eta = 0.05$	0.86	0.84	0.82
$p_T = 32.5$,	$\eta = 2.25$	0.88	0.88	0.82
$p_T = 75$,	$\eta = 0.05$	0.89	0.89	0.86
$p_T = 75$,	$\eta = 2.25$	0.93	0.91	0.90

The η and p_T bins used for the response function are specified in Table 2.7. Only positive η bins are listed in the table, but identical bins are used for negative η between -2.5 and 0. The η bins are symmetric, but the full range is used as differences were observed between the positive and negative halves of the detector during the 2010 run. Otherwise, bins in η are chosen to coincide with the physical boundaries of the calorimeter towers. Bins in p_T are variable, ranging in width from 500 MeV at low-p_T to 10 GeV at the upper end of the spectrum. The use of variable bin size is important as a larger variation in detector response is expected at low-p_T, requiring finer bins, than at higher p_T where coarser separation is sufficient.

Table 2.7: Definition of η, p_T bins used for the response function.

$\eta =$	{0.000, 0.087, 0.174, 0.261, 0.348, 0.435, 0.522, 0.609, 0.696, 0.783, 0.879, 0.957, 1.044, 1.131, 1.218, 1.305, 1.392, 1.479, 1.566, 1.653, 1.740, 1.830, 1.930, 2.043, 2.172, 2.322, 2.500}
$p_T =$	{0.0, 0.5, 1, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100}

2.6.5 Implementation of the tc$\not\!\!E_T$ Algorithm

Track-corrected $\not\!\!E_T$ is calculated for an event using reconstructed muons, electrons and tracks, the response function and the calorimeter-based $\not\!\!E_T$. Muons are corrected at the outset using the standard algorithm discussed in Section 2.3. Tracks matched to reconstructed electrons with $h/e < 0.1$ are identified and no correction is applied. An object reconstructed as an electron with a small h/e deposited a large fraction of its energy in the ECAL with a response close to unity. The remaining reconstructed tracks are treated as pions and are considered for correction using the response function.

The $\not\!\!E_T$ is corrected for those remaining tracks that additionally have $p_T > 1$ GeV and pass the selections in Table 2.3 by removing for each good track the expected energy, $\langle \vec{E}_T \rangle$, deposited in the calorimeter, determined using the response function, and replacing it with the momentum reconstructed at the vertex. Tracks with $p_T < 1$ GeV that pass the selections in Table 2.3 are fully compensated for assuming no calorimeter response $\Big($i.e. $\langle \vec{E}_T \rangle = 0\Big)$. These corrections are summarized in (2.8).

$$\begin{aligned} tc\vec{\not\!\!E}_T &= \vec{\not\!\!E}_T + \delta\vec{\not\!\!E}_T^{\,\mu} + \delta\vec{\not\!\!E}_T^{\,\mathrm{tc}} \\ &= \vec{\not\!\!E}_T^{\,\mu} + \sum_{\substack{\text{good} \\ \text{tracks}}} \langle \vec{E}_T \rangle - \sum_{\substack{\text{good} \\ \text{tracks}}} \vec{p}_T \end{aligned} \tag{2.8}$$

where $\vec{\not\!\!E}_T^{\,\mu}$ is the result of applying the correction in (2.2) and the sum is over all tracks not matched to muons or electrons and passing the selections in Table 2.3.

It is important to note that the correction for each charged pion involves two sets of coordinates. The expected energy is removed using a position at the face of the ECAL, determined by extrapolating the track from the vertex using an analytical propagator. The track momentum is taken at the vertex. Explicitly, for the x-component of the $\not\!\!E_T$,

$$\left(tc\not{E}_T\right)_x = \left(\not{E}_T^{\mu}\right)_x + \sum_{\substack{\text{good}\\\text{tracks}}} \langle E\rangle \sin\theta_c \cos\phi_c - \sum_{\substack{\text{good}\\\text{tracks}}} p_T \cos\phi_v \tag{2.9}$$

where θ_c, ϕ_c are the polar and azimuthal coordinates of the extrapolated trajectory at the face of the ECAL and ϕ_v is the azimuthal angle of the track at the vertex.

Design Considerations

The tc$\not{E}_T$ algorithm was designed from the outset not only to provide a needed physics tool, but also to maximize its operational utility. The unique needs of a particular analysis demand a tool that is flexible. Lepton selections, in particular, vary significantly depending on the degree of trade-off between efficiency and fakes a particular analysis can permit. Varied electron selections are not relevant in the context of tc$\not{E}_T$, as the detector in which the primary interaction occurs is more germane than is the correct identification of the particle. Muons, on the other hand, are minimum ionizing and thus the interpretation of a muon candidate has significant implications on the reconstructed $\not{E}_T$. To this end, we created a tool that provides the necessary information for modifying the reconstructed $\not{E}_T$, depending upon how a muon candidate is interpreted in the context of a specific analysis. One map contains the x and y components of the calorimeter deposits associated with the track, assuming it is a muon, as well as a flag that indicates whether or not the $\not{E}_T$ was corrected for the muon, and which fit — global, tracker or standalone — was used for the correction. A second map contains the x and y components of the expected energy deposit, assuming the object is a pion, and a flag indicating how the object was treated by the tc$\not{E}_T$ algorithm — that is, corrected as a muon, corrected as a pion or no correction applied. Together, the maps allow the $\not{E}_T$ to be modified to accommodate any desired muon selection.

The creation of the maps described above also contributes to a second design consideration. The volume of data collected by CMS when the LHC operates with design parameters necessitates that only the subset of each reconstructed event that is essential to performing an analysis is disseminated to users. One implication of this is that it is not possible for an individual user to redo large parts of the reconstruction. The tc$\not{E}_T$ algorithm was designed from the beginning to be reproducible using the response function and information available to users at any data tier. As a result, users can modify and recalculate tc$\not{E}_T$ on the fly as required by an analysis rather than wait for a new

reconstruction to be performed centrally by the experiment.

2.7 Performance of tc$\not\!\!E_T$ in Simulation

The goal at the outset was to reduce the sizable tails of the Drell-Yan $\not\!\!E_T$ distribution. The correction is tested using simulated Drell-Yan events with the metric being the number of events with $\not\!\!E_T > 30$ (50) GeV. Additionally, the tc$\not\!\!E_T$ algorithm is tested on simulated $t\bar{t}$ events in the dileptonic final state. The use of a sample with real $\not\!\!E_T$ is an important cross check. Correcting the Drell-Yan tails with a naive additive or multiplicative factor will eliminate events with high $\not\!\!E_T$, but will also remove events with real $\not\!\!E_T$. Thus, it is necessary check that any attempt to correct for mis-measured $\not\!\!E_T$ does not significantly impact the ability to identify those events for which $\not\!\!E_T$ is an important and real signature. The baseline $\not\!\!E_T$ is defined to be calo$\not\!\!E_T$ plus muon corrections, $\not\!\!E_T^{\mu}$. Final states are selected by requiring two oppositely charged, well identified and isolated leptons. Muons are required to pass the default selections used by the tc$\not\!\!E_T$ algorithm as discussed in Table 2.2. Electrons are selected using a requirement similar to that for the $t\bar{t}$ analysis, discussed in Section 3.2.4.

2.7.1 Final States with Fake $\not\!\!E_T$

Figure 2.6 shows the $\not\!\!E_T$ distribution for Drell-Yan events with two good muons or electrons in the final state. Here, we only consider events with electron or muon final states at the generator level. Drell-Yan events decaying to τ's that subsequently decay to electrons or muons have real $\not\!\!E_T$ and may be suppressed by cuts at the analysis level.

The distributions in Figure 2.6 evidence non-Gaussian tails, characteristic of fake $\not\!\!E_T$ in Drell-Yan events. Correcting for tracks reduces the number of events above 50 GeV by a factor of 4.3 compared to the present best caloMET (corrected for muons and the JES). Table 2.8 shows the breakdown of events by the number of jets in the event. Here, we count uncorrected calorimeter-based jets with $p_T > 15$ GeV and $|\eta| < 3$.

The majority of Drell-Yan events fall into the 0-jet bin, where the tail is reduced by a factor of 2.1. Although this correction was conceived with the 0-jet bin in mind, the large number of tracks in events with jets gives reason to expect similar reductions. The improvement in the 2-jet bin is more than 50% that in the 0-jet bin. The 1-jet bin shows even better performance, although this is not entirely unexpected, as these events contain one sizable jet which provides additional tracks for correction. The presence

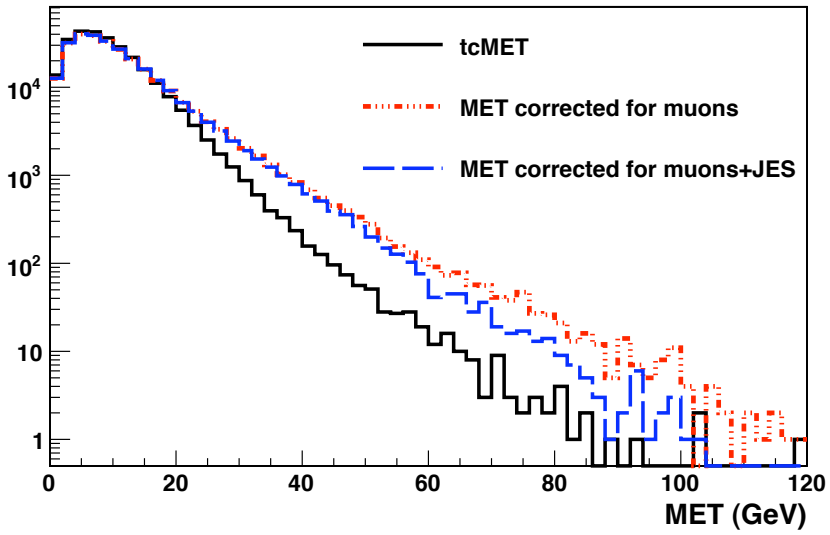

Figure 2.6: Distribution for tc$\not\!\!E_T$ and corrected calo$\not\!\!E_T$ objects in $Z \to \ell^+\ell^-$ events, where the leptons are required to have $p_T > 20$ GeV. The tc$\not\!\!E_T$ algorithm reduces the number of events with $\not\!\!E_T > 30$ (50) GeV by a factor of 3.4 (6.8) compared to caloMET corrected for muons only. The reduction is a factor of 3.0 (4.3) when compared to caloMET corrected for both muons and the JES.

Table 2.8: Performance of the tc$\not\!\!E_T$ algorithm on Drell-Yan events in the dilepton final state as a function of the number of reconstructed jets. The numbers reported are for a requirement of $\not\!\!E_T > 30$ GeV.

Case	0 jets	1 jet	2 jets	jets
$\not\!\!E_T^{\mu}$	915/201863 0.5%	4860/53607 9%	852/2978 22%	2044/5201 39%
tc$\not\!\!E_T$	435/201863 0.2%	1085/53607 2%	852/2978 6%	797/2044 15%
factor of improvement	2.1	4.5	3.5	2.6

of a single large jet that is typically under-measured produces significant asymmetry, providing excellent conditions for correction. Even in events with more than 2 jets, the improvement is by a factor of 2.6.

2.7.2 Final States with Real $\not\!\!E_T$

Although it was not the intent at the outset to improve resolution, it is nevertheless important to consider the effect of a track-based correction on processes with real $\not\!\!E_T$ for reasons discussed in the introduction to this section. Here we consider the effect of the track–based correction on the resolution of the magnitude and direction of the missing transverse momentum for $t\bar{t} \to \ell^+\ell^- + \mathrm{X}$. Again, we require two good, oppositely charged electrons or muons, although we relax the same flavor constraint. We also allow the leptons to be the daughter's of $\tau's$ as we are interesting in how well we measure real $\not\!\!E_T$.

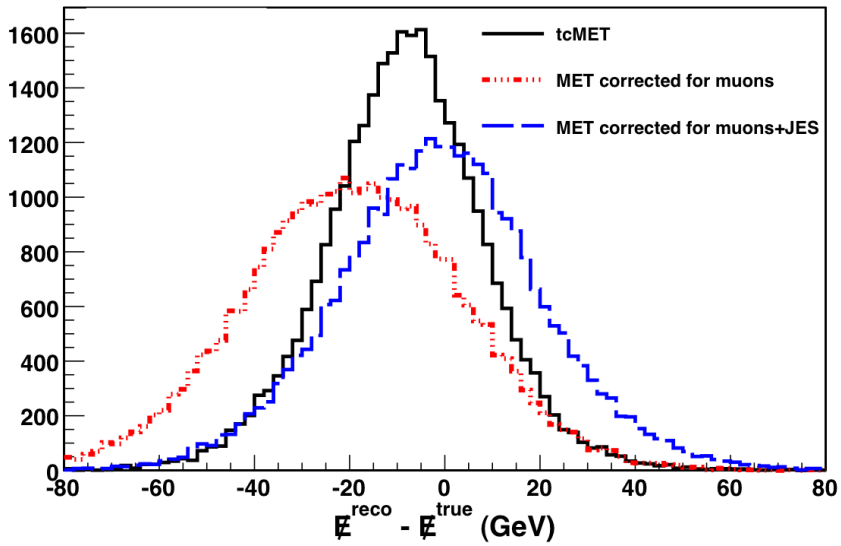

Figure 2.7: Distribution of $\Delta\not\!\!E_T = \not\!\!E_T^{\mathrm{reco}} - \not\!\!E_T^{\mathrm{true}}$ for $t\bar{t} \to \ell^+\ell^- + \mathrm{X}$ events with $\not\!\!E_T^{\mathrm{true}} > 50$ GeV. The $\not\!\!E_T$ resolution is improved nearly 25% (20%) for tc$\not\!\!E_T$ compared to calo$\not\!\!E_T$ corrected for muons (calo$\not\!\!E_T$ corrected for muons + JES).

Figure 2.7 shows the $\not\!\!E_T$ resolution for events with $\not\!\!E_T^{\mathrm{true}} > 50$ GeV. An improvement of approximately 25% is observed. There exists an offset at the level of 3 GeV. This is expected as the tc$\not\!\!E_T$ algorithm does not include any treatment for neutral objects. Figure 2.8 shows the resolution of the the azimuthal $\not\!\!E_T$ direction for dilepton $t\bar{t}$ events. Correcting with tracks improves the determination of the $\not\!\!E_T$ direction by nearly

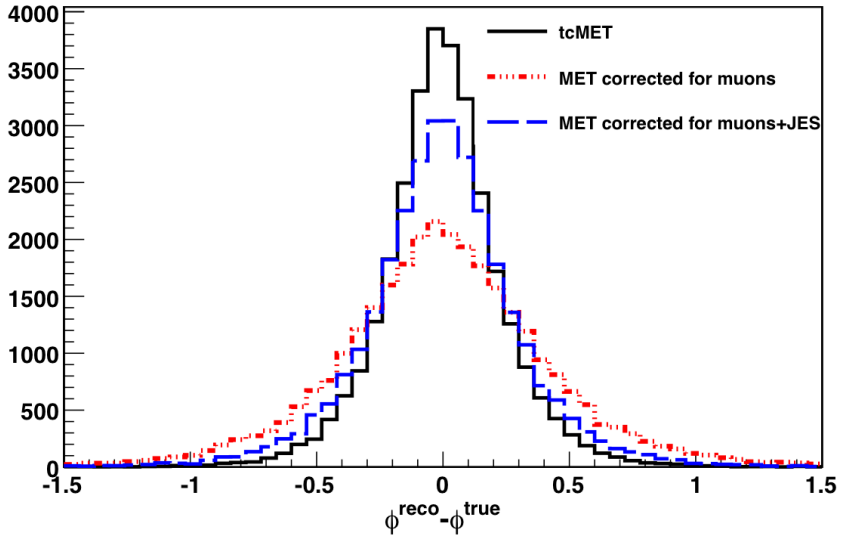

Figure 2.8:]
Distribution of $\Delta\phi = \phi^{\text{reco}} - \phi^{\text{true}}$ for $t\bar{t} \to \ell^+\ell^- + \text{X}$ events with $\not\!\!E_T^{\text{true}} > 50$ GeV. The resolution of the $\not\!\!E_T$ direction is improved approximately 40% (20%) for tc$\not\!\!E_T$ compared to calo$\not\!\!E_T$ corrected for muons (calo$\not\!\!E_T$ corrected for muons + JES).

40% compared to correcting for muons only and by approximately 20% compared to correcting for correcting for muons and the JES.

2.7.3 Fake $\not\!\!E_T$ in QCD Processes

Before moving on it is worthwhile to consider the effect a track–based correction has on the $\not\!\!E_T$ for QCD processes. These events are expected to contain little real $\not\!\!E_T$ while having significant hadronic activity, and thus serve as a good platform for comparison. Here, inclusive QCD samples over several $\hat{p_T}$ ranges were tested. In all cases, the tc$\not\!\!E_T$ algorithm is seen to improve both the tails and the resolution of the the $\not\!\!E_T$ distribution. This improvement increases with increasing $\hat{p_T}$ up to several hundred GeV, as is visible in Figure 2.9.

2.8 Commissioning $\not\!\!E_T$ with Early Data

Collisions in late 2009 presented the first opportunity to commission $\not\!\!E_T$ using data. Events were collected at two center–of–mass energies, 900 and 2360 GeV, with a small number of bunches per fill and relatively few protons per bunch. These conditions

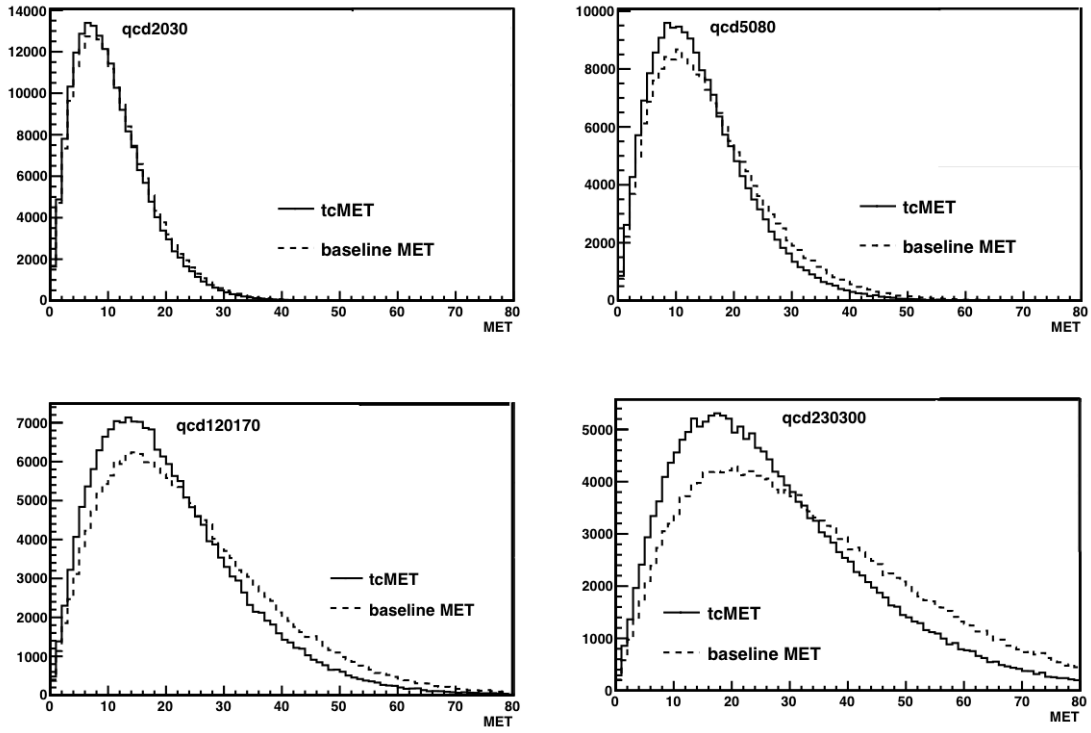

Figure 2.9: Comparison of tc$\not\!\!E_T$ and corrected calo$\not\!\!E_T$. Four different $\hat{p_T}$ ranges are shown, where qcdXY refers to the lower value, X, and upper value, Y, of the $\hat{p_T}$ interval. In all cases, tc$\not\!\!E_T$ improves both the tail and resolution of the $\not\!\!E_T$ distribution with the degree of improvement increasing with $\hat{p_T}$.

were very useful for commissioning as the low intensity beams ensured only one interaction per crossing and the small number of bunches reduced machine-related effects such as beam halo, where protons in the beam interact with residual gas in the vacuum pipe, producing hadrons that decay to muons and pass longitudinally through the detector. Halo muons can deposit energy via bremsstrahlung, resulting in significant, fake $\not\!\!E_T$. Collisions were predominantly minimum bias with no real $\not\!\!E_T$, simplifying the search for fake $\not\!\!E_T$ arising from detector effects.

Collisions were selected by requiring a signal in both Beam Scintillation Counters (BSC) coincident with either of the Beam Pick–up Timing eXperiment (BPTX) devices and at least one well–reconstructed primary vertex. To suppress machine–induced background, an event was vetoed if any of the BSC–based beam halo triggers fired. It was observed that some bunch crossings were associated with high levels of activity in the pixel layers resulting in the reconstruction of an exceptionally large number of fake tracks. This pathology was rejected by requiring that, in events with at least 10 tracks, at least 25% were high purity. Figure 2.10 shows the $r-\phi$ and $r-z$ views of the tracker in such an event.

Early commissioning activities revealed an excess of events in the tails of the $\not\!\!E_T$ distributions. A large fraction of these events evidenced one of three characteristics:

- The $\vec{\not\!\!E}_T$ was anti–aligned with a single high-energy tower in the forward hadron calorimeter with little surrounding activity.
- The $\vec{\not\!\!E}_T$ was anti–aligned with a single high-energy ECAL crystal in the barrel with little surrounding activity.
- The hemisphere opposite the $\vec{\not\!\!E}_T$ contained either a high energy readout in a small number of towers or 1–4 parallel rows of approximately 18 adjacent high energy towers.

Figure 2.11 shows the tc$\not\!\!E_T$ distribution at the start of commissioning. A large excess of high $\not\!\!E_T$ events are observed in data that are not present in the simulation. The data distribution has nearly 300 events with $\not\!\!E_T > 14$ GeV. Table 2.9 shows a classification of these high-$\not\!\!E_T$ events determined by a visual inspection.

We developed tools to identify and correct for anomalous deposits in the ECAL and HF. Methods based upon this early work are used to clean the rechit collections during reconstruction. Fake $\not\!\!E_T$ due to noise in the HCAL was expected prior to first

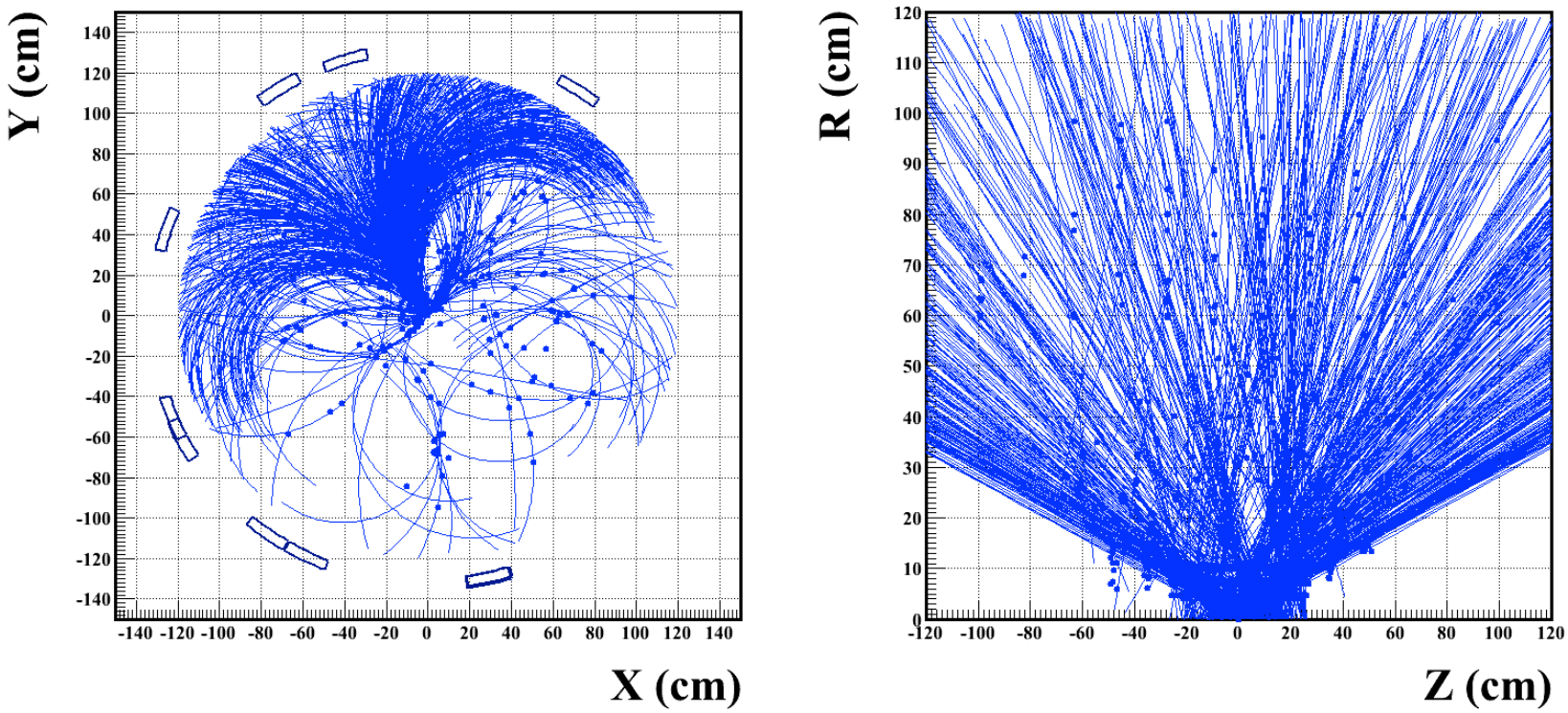

Figure 2.10: $r-\phi$ and $r-z$ views of the tracker for an event with anomalous pixel activity resulting in a large number of reconstructed tracks. This detector failure is rejected by requiring that, in events with 10 or more tracks, at least 25% are high purity.

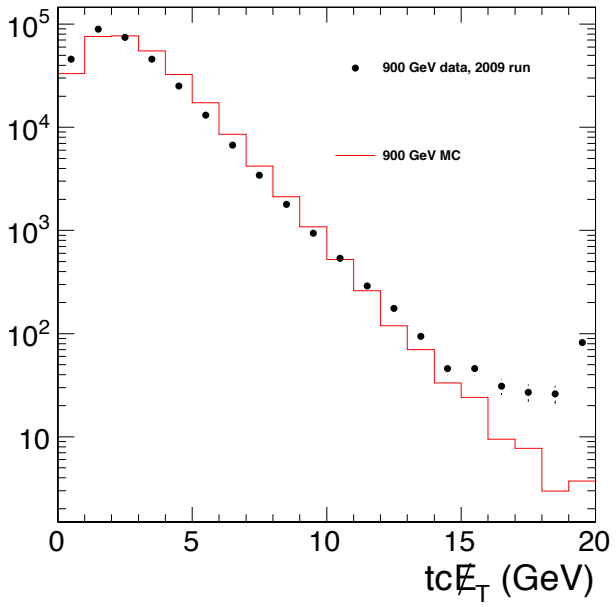

Figure 2.11: Distribution of tc$\not\!\!E_T$ in data and MC at the start of the commissioning of 2009 data.

collisions and, due to the details of the HCAL readout, no reliable method is available to correct for this source of noise. The current CMS recommendation is to reject events identified as containing significant HCAL noise [15].

Table 2.9: Classification of events with $\not\!\!E_T > 14$ GeV in the 2009 dataset. The classification is based on visual inspection.

Source	% of Events
HF	43%
ECAL	24%
HCAL	5%
other detector effects	16%
cause not easily identifiable	12%

2.8.1 Anomalous HF Deposits

The primary source of fake $\not\!\!E_T$ tails was a single high-energy, isolated tower in the forward hadron calorimeter. The tower, in these cases, had large energy reconstructed in one channel with little energy measured in the other fiber. Figures 2.12– 2.14 show a visualization of a characteristic anomalous HF signal seen in the 2009 data. Anomalous

deposits were distributed uniformly through the HF.

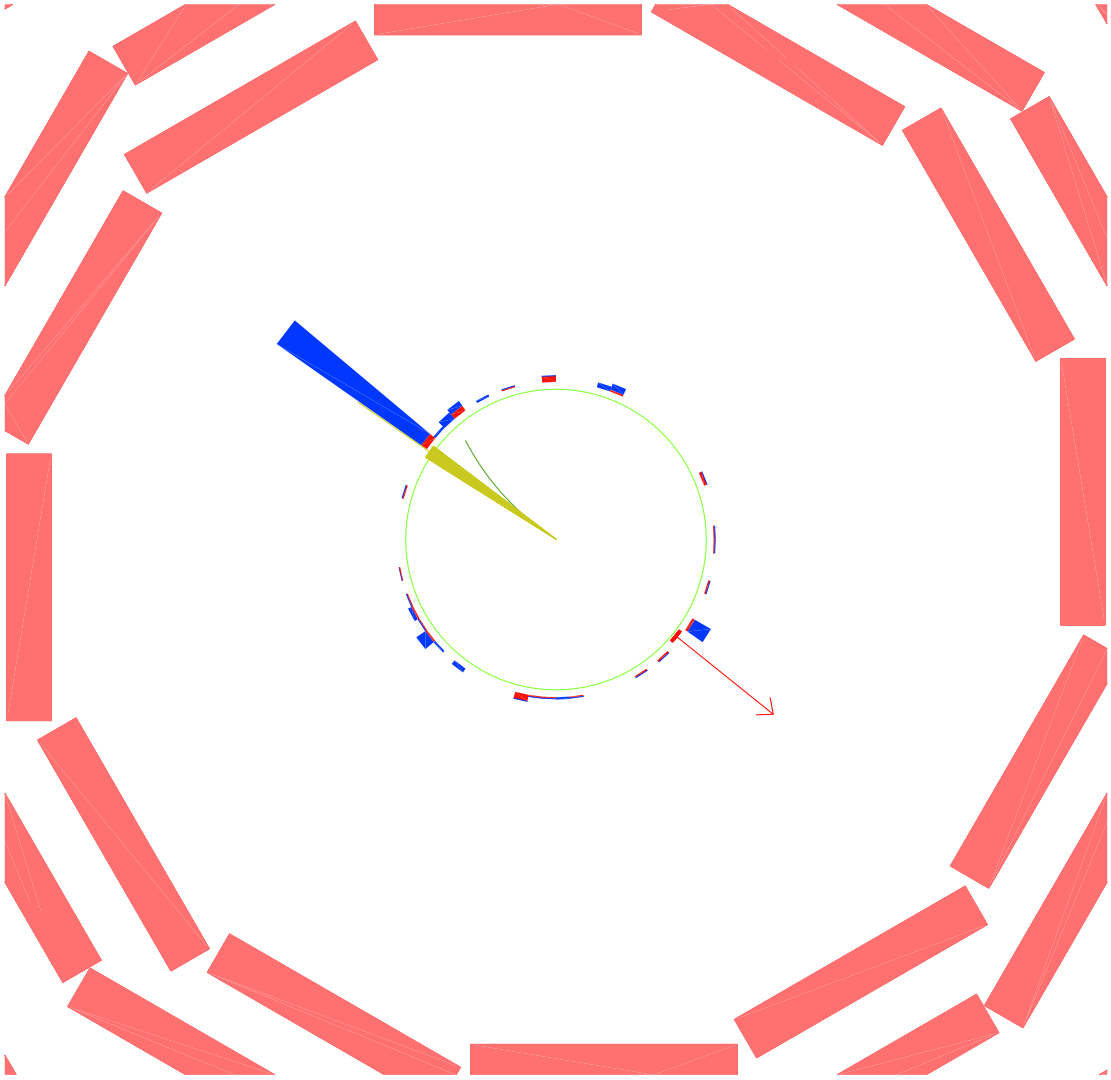

Figure 2.12: Visualization in the $r-\phi$ plane of an event from 2009 data showing clear evidence of a high energy calorimeter deposit that is anti-aligned with the $\vec{\not{E}}_T$.

Consultation with HF experts [15] led us to characterize towers in the HF using the quantity $\alpha = \frac{L-S}{L+S}$, where L, S are the reconstructed energies in the long and short readout fibers, respectively. Figure 2.15 shows the distribution of α for simulation and for data. Only towers with $E_T > 5$ GeV are shown, as towers with smaller transverse energy do not contribute significantly to the $\not{E}_T$. The Monte Carlo is normalized to the number of events in data.

A tower in the forward hadron calorimeter is composed of a long and short fiber.

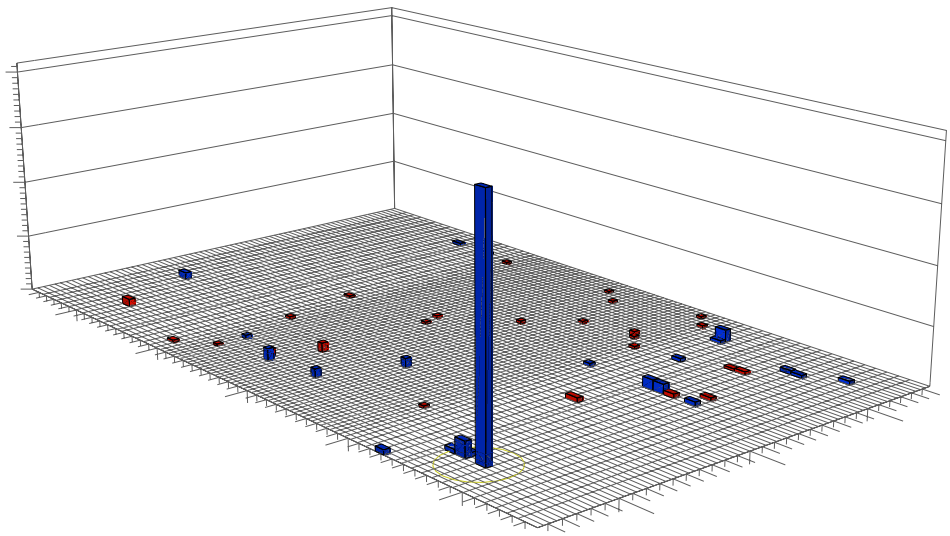

Figure 2.13: Visualization of the CMS calorimeter unrolled to form a plane in $\eta - \phi$ space. There is clear evidence for a high energy tower in the forward region with little surrounding activity.

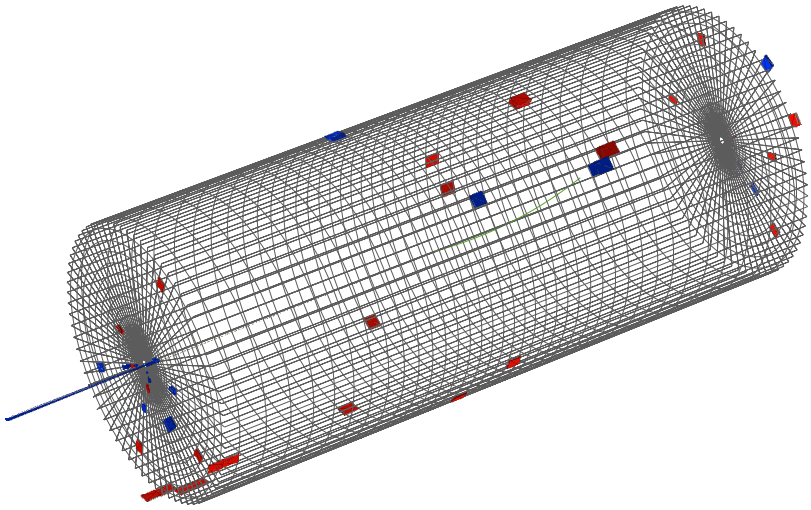

Figure 2.14: Visualization of a characteristic anomalous signal in the HF. The image shows a three–dimensional representation of the CMS detector.

The long fiber extends nearly to the face of the tower while the short fiber is recessed approximately 22 cm. Hadronic particles shower throughout the tower resulting in a distribution for α that peaks at 0 with a tail that extends to 1. The shower of an electron or photon develops more quickly and so the distribution is shifted towards 1, although α rarely is larger than 0.98. These features are apparent in Figure 2.15 where large peaks are observed near $\alpha = \pm 1$ in data that are not seen in the simulation.

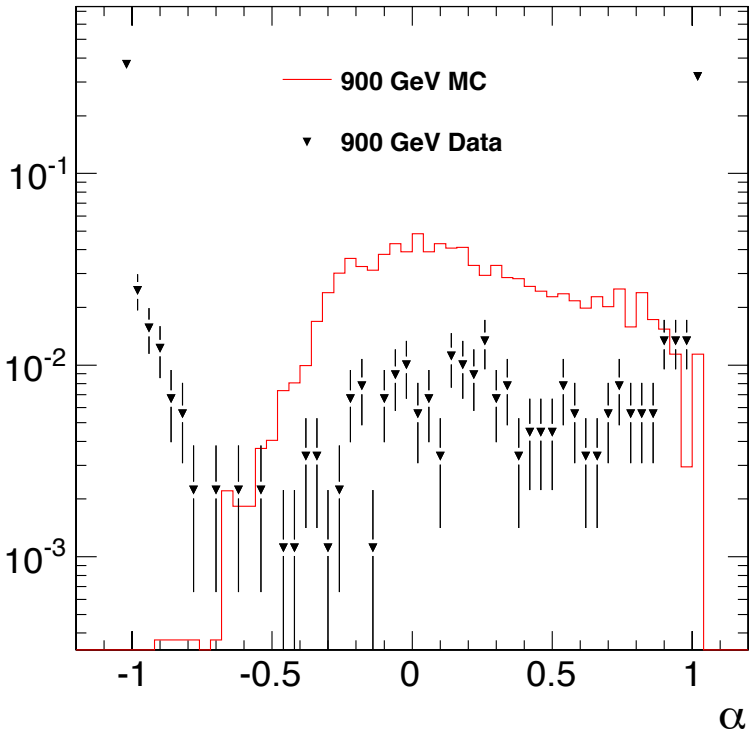

Figure 2.15: Distribution of $\alpha = \frac{L-S}{L+S}$ used to characterize energy deposits in the forward hadron calorimeter. There is an excess of towers in data near $\alpha = \pm 1$, which is indicative of anomalous signals in the HF. Only towers with $E_T > 5$ GeV are shown as towers with smaller E_T do not contribute significantly to the $\not\!\!E_T$ tails.

Based on Figure 2.15, we identify towers in the forward hadron calorimeter with $\alpha < -0.8$ or $\alpha > 0.99$ as anomalous, hereafter referred to as HF spikes. The unphysical nature of towers with values of α in the far negative tail permits a looser cut than is possible in the positive tail where it is necessary to cut tight to avoid introducing an inefficiency for reconstructed electrons, photons and neutral pions. Only towers with $E_T > 5$ GeV are classified as spikes. In minimum bias events, approximately 80% of HF deposits with $E_T > 5$ GeV were found to be anomalous while less than 1% of HF towers

with $E_T > 5$ GeV are classified as spikes in Monte Carlo. HF spikes occur in data at a rate of approximately 1 per 500 events. [16].

Evidence that HF spikes are producing fake $\not\!\!E_T$ tails can be seen in Figure 2.16. The ordinate in the bottom panel is misleadingly referred to as $\not\!\!E_T$ ϕ, by which is meant the direction opposite to the $\vec{\not\!\!E}_T$. If an event contained multiple deposits identified as spikes, the tower E_T and ϕ were calculated by a vector sum of the spiking towers. In the 2009 data, six events with two HF spikes were found. One event with four HF spikes was observed. The two distributions show a clear correlation, confirming that HF spikes are producing fake $\not\!\!E_T$ tails.

We proposed a method to clean the $\not\!\!E_T$ for HF spikes by adding the components of an anomalous tower to the corresponding components of the $\not\!\!E_T$. Figure 2.17 shows the $\not\!\!E_T$ distribution for the subset of events in data that contain at least one HF spike. The cleaning procedure significantly reduces the tail of the $\not\!\!E_T$ distribution.

2.8.2 Anomalous ECAL Deposits

A secondary source of fake $\not\!\!E_T$ tails was a single high-energy, isolated crystal in the ECAL. Anomalous crystals appear almost exclusively in the barrel. Figures 2.18–2.20 show a visualization of a characteristic anomalous ECAL signal seen in the 2009 data.

The procedure for identifying anomalous ECAL signals is similar to that described in 2.8.1. Although the crystals have only a single readout channel, the physical design of the ECAL provides a handle, as the width of a crystal is approximately the same as the Moliere radius of an electron or photon interacting in the material. As a result, a typical shower from an electromagnetic object is expected to be shared across a group of neighboring crystals. Thus, instead of α, we define a topological measure to exploit the isolated nature of these crystals. The variable in this case is called $R4$ and is defined as the ratio of the energy summed in the neighbors of a seed crystal divided by the energy of the seed crystal. An electron or photon that deposits energy in the ECAL will have its shower spread to neighboring crystals with the $R4$ distribution peaking near 0.2 and falling off quickly. Figure 2.21 shows the distribution of $R4$ versus the E_T of the seed crystal for both data and minimum bias Monte Carlo.

Based on Figure 2.21, we identify crystals in the ECAL barrel with $R4 < 0.05$ and $E_T > 5$ GeV as anomalous, hereafter referred to as ECAL spikes. ECAL spikes occur

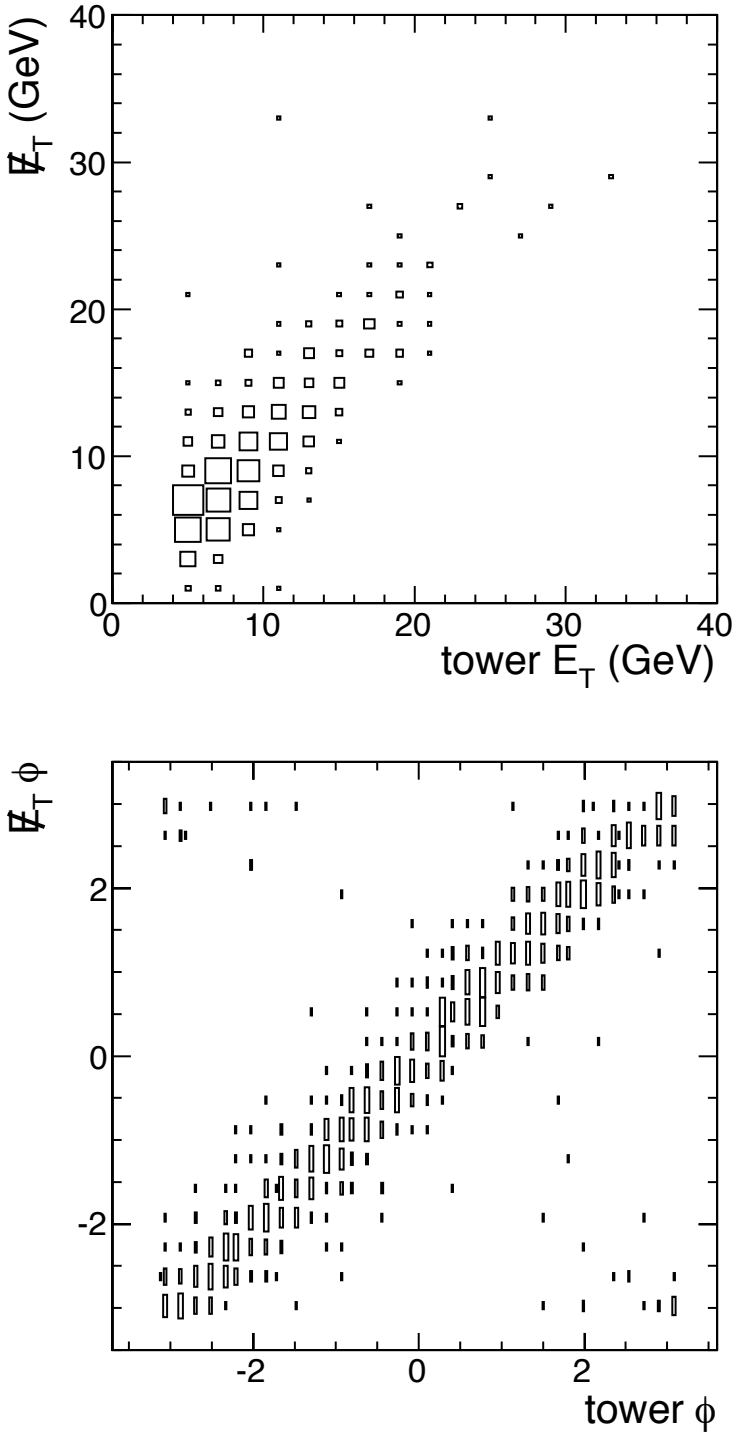

Figure 2.16: Correlation of the magnitude (top) and direction (bottom) between HF spikes and the $\vec{\not\!\!E}_T$.

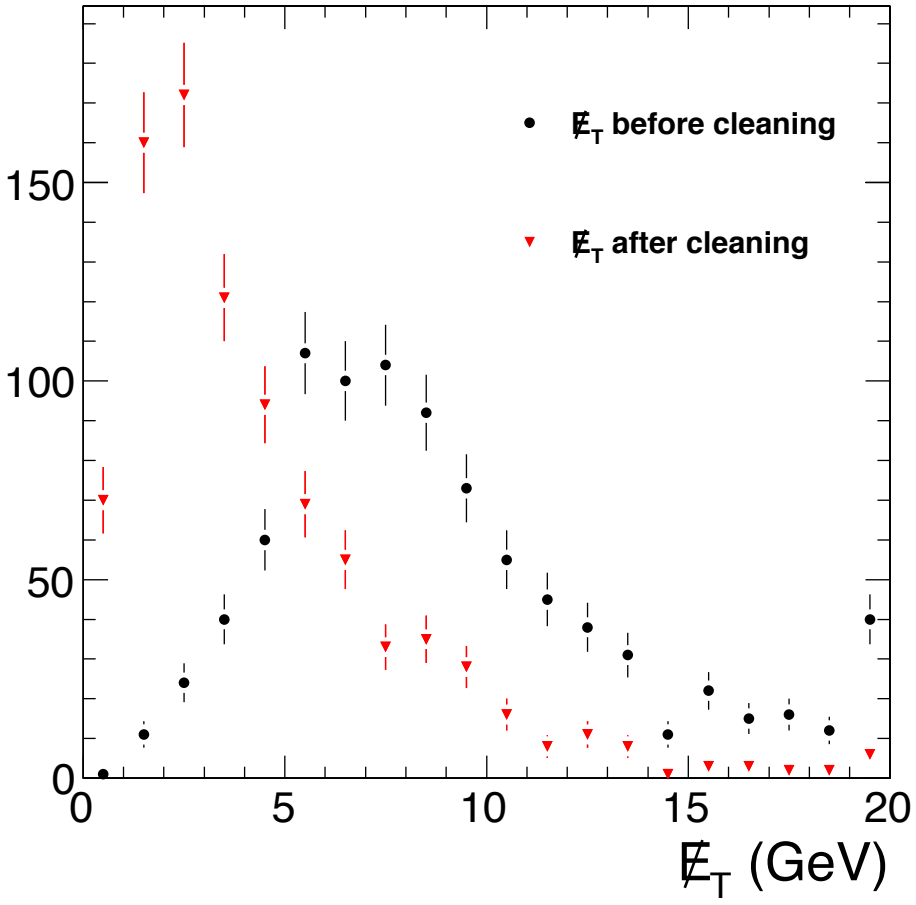

Figure 2.17: Distribution of $\not\!\!E_T$ before and after cleaning HF spikes. Only the subset of events in data containing a HF tower identified as a spike is shown.

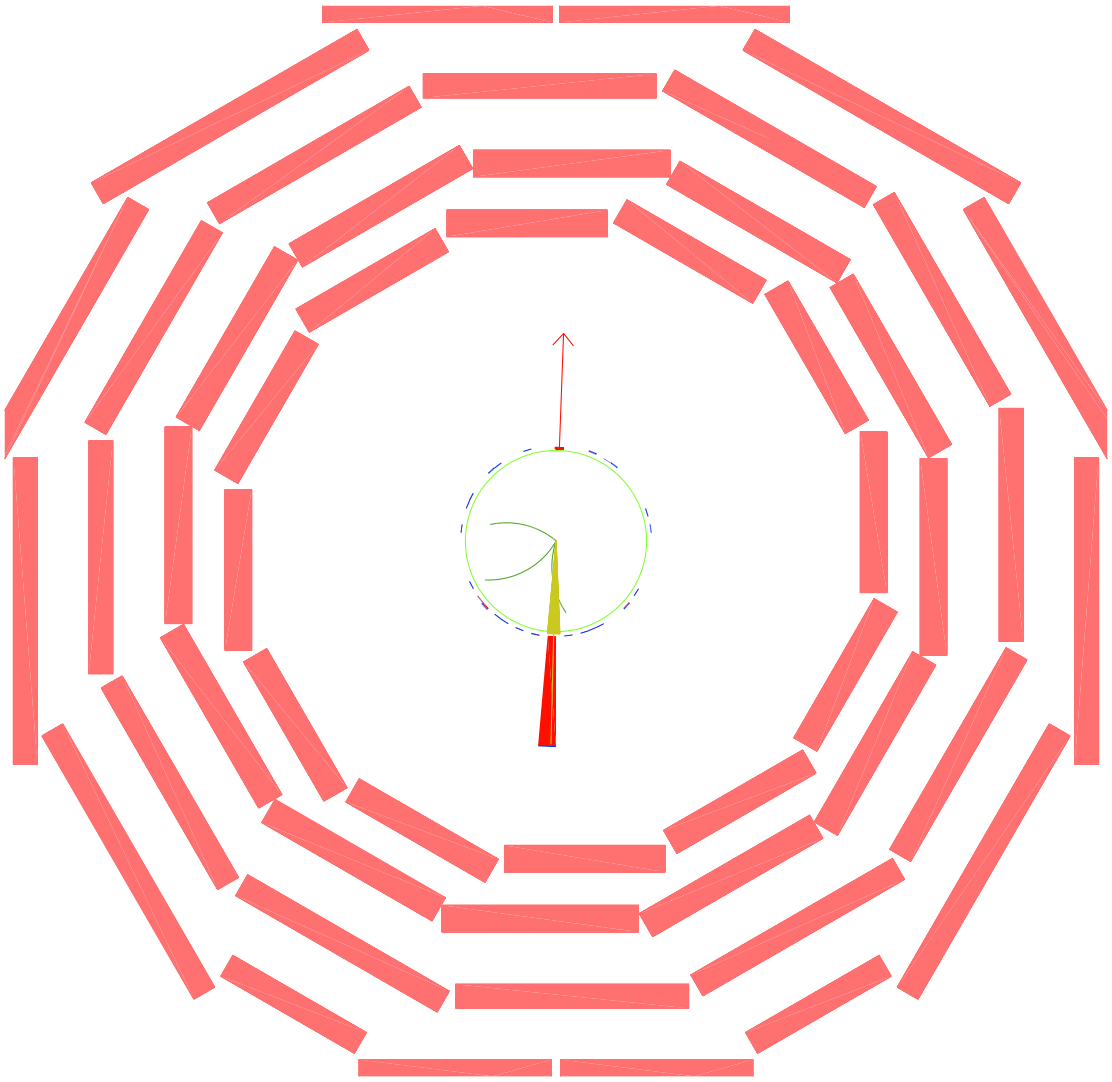

Figure 2.18: Visualization in the $r - \phi$ plane of an event from 2009 data showing clear evidence of a high energy calorimeter deposit that is anti-aligned with the $\vec{\not{E}}_T$.

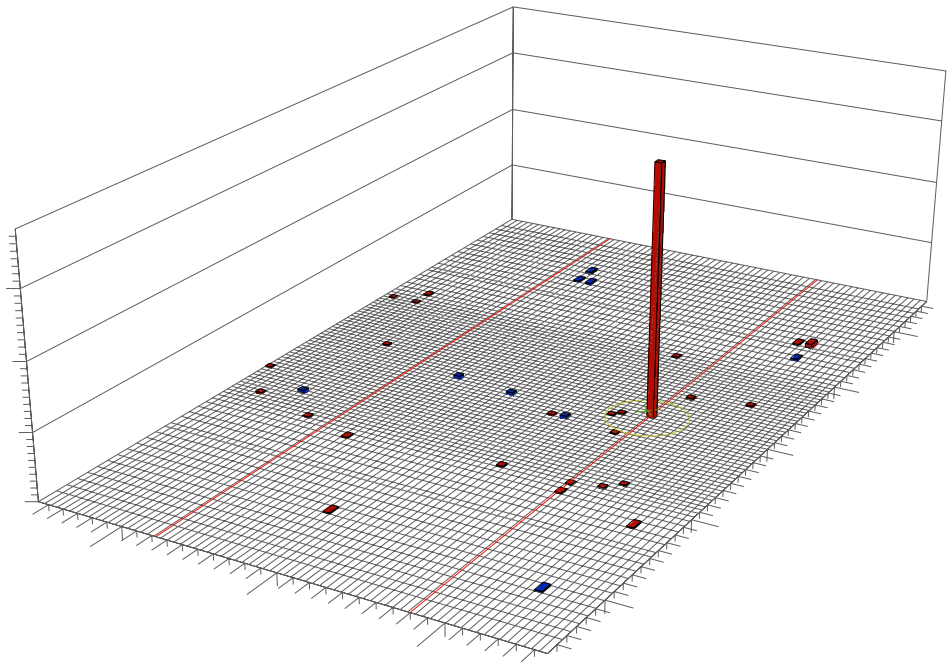

Figure 2.19: Visualization of the CMS calorimeter unrolled to form a plane in $\eta - \phi$ space. There is clear evidence for a single high energy tower in the barrel ECAL with little surrounding activity.

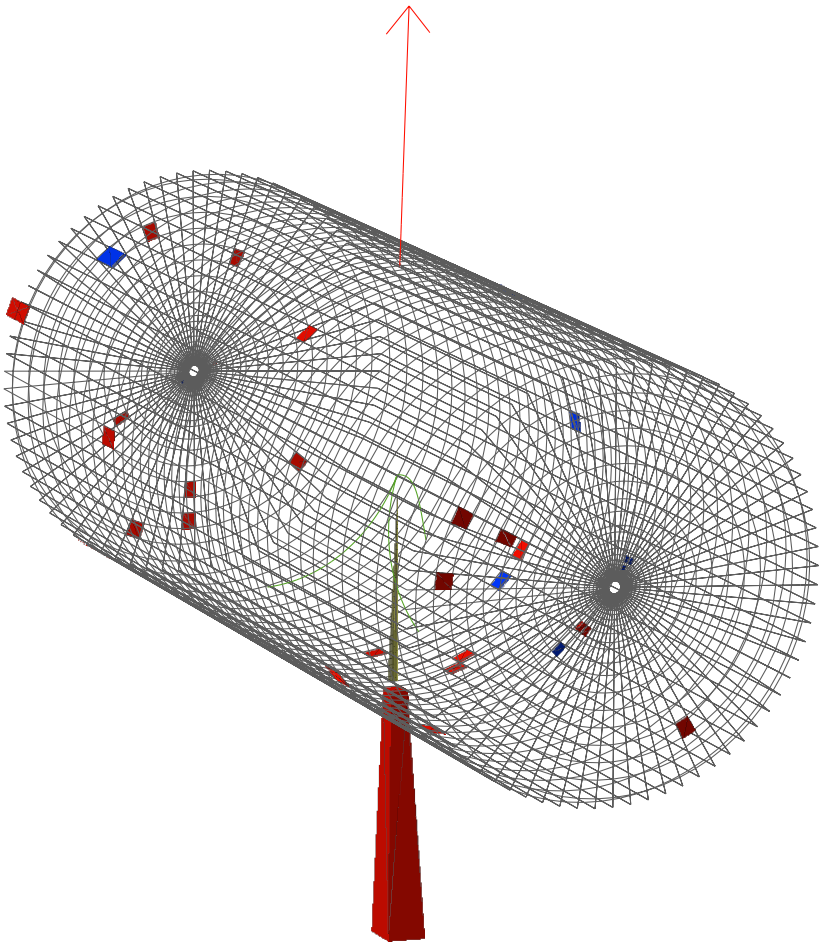

Figure 2.20: Visualization of a characteristic anomalous signal in the EB. The image shows a three–dimensional representation of the CMS detector.

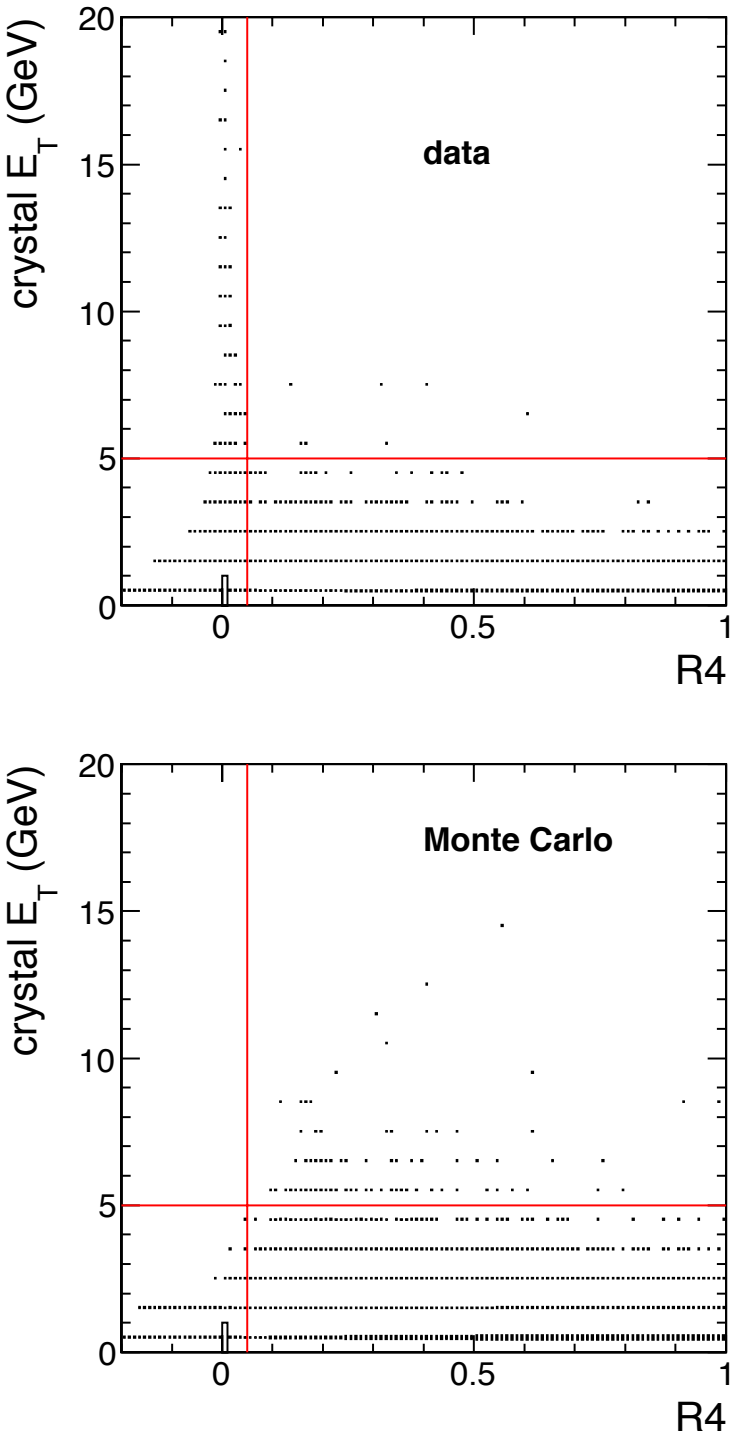

Figure 2.21: Distribution of $R4$ used to characterize energy deposits in the ECAL barrel. There is an excess of towers in data with $E_T > 5$ GeV and $R4 < 0.05$ which is indicative of anomalous signals in the ECAL.

in data at a rate of approximately 1 per 1000 events. The rate was relatively constant across data taking periods and was not associated with a nearby reconstructed track or energy in the HCAL of the containing tower. The crystals were observed uniformly in η, ϕ in the barrel as evident in Figure 2.22.

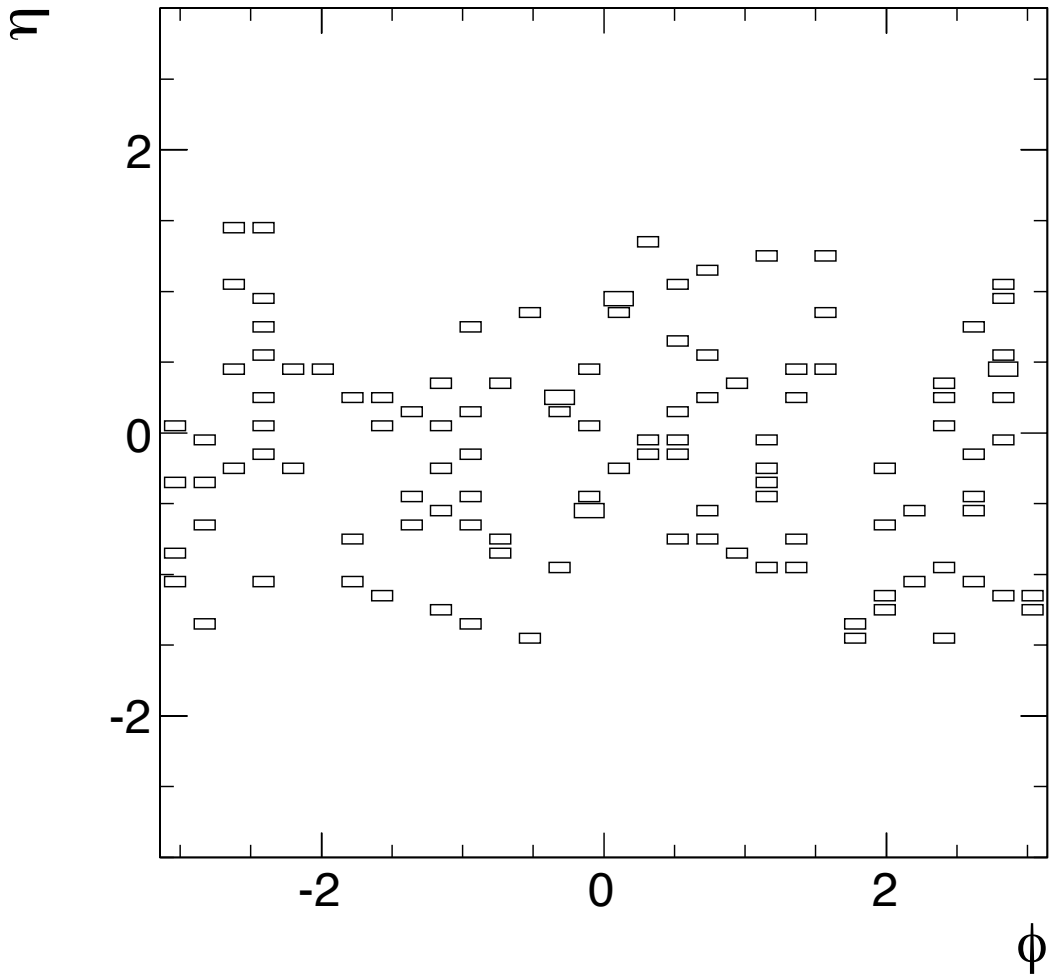

Figure 2.22: ECAL spikes are uniformly distributed throughout the barrel.

Evidence that ECAL spikes are producing fake $\not\!\!E_T$ tails can be seen in Figure 2.23. The ordinate in the bottom panel is misleadingly referred to $\not\!\!E_T$ ϕ, by which is meant the direction opposite to the $\vec{\not\!\!E}_T$. No events were observed with more than one ECAL spike in the 2009 data. The two distributions show a clear correlation, confirming that ECAL spikes are producing fake $\not\!\!E_T$ tails.

We proposed a method to clean the $\not\!\!E_T$ for ECAL spikes by adding the compo-

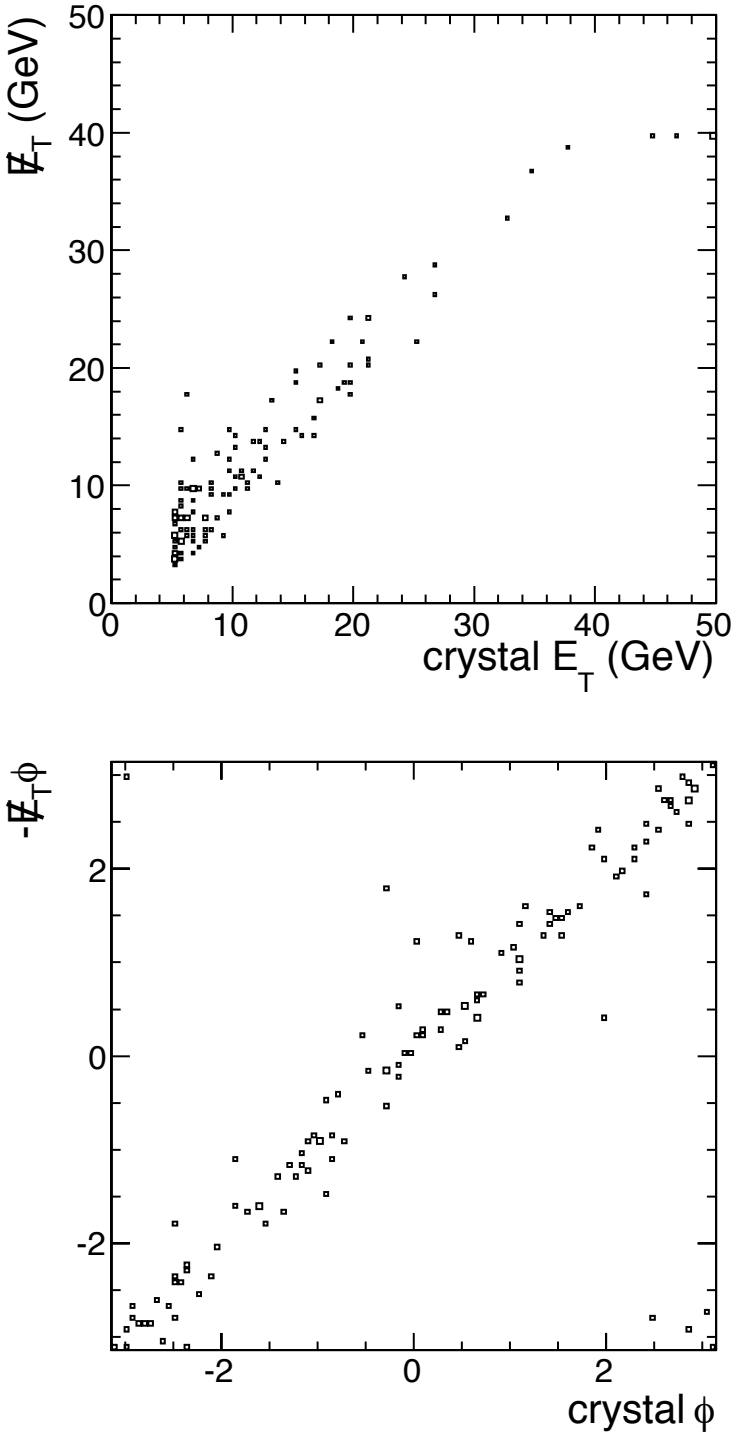

Figure 2.23: Correlation of the magnitude (top) and direction (bottom) between ECAL spikes and the $\vec{\not\!\!E}_T$.

nents of an anomalous crystal to the corresponding components of the $\not\!\!E_T$. Figure 2.24 shows the $\not\!\!E_T$ distribution for the subset of events in data that contain an ECAL spike. The cleaning procedure significantly reduces the tail of the $\not\!\!E_T$ distribution.

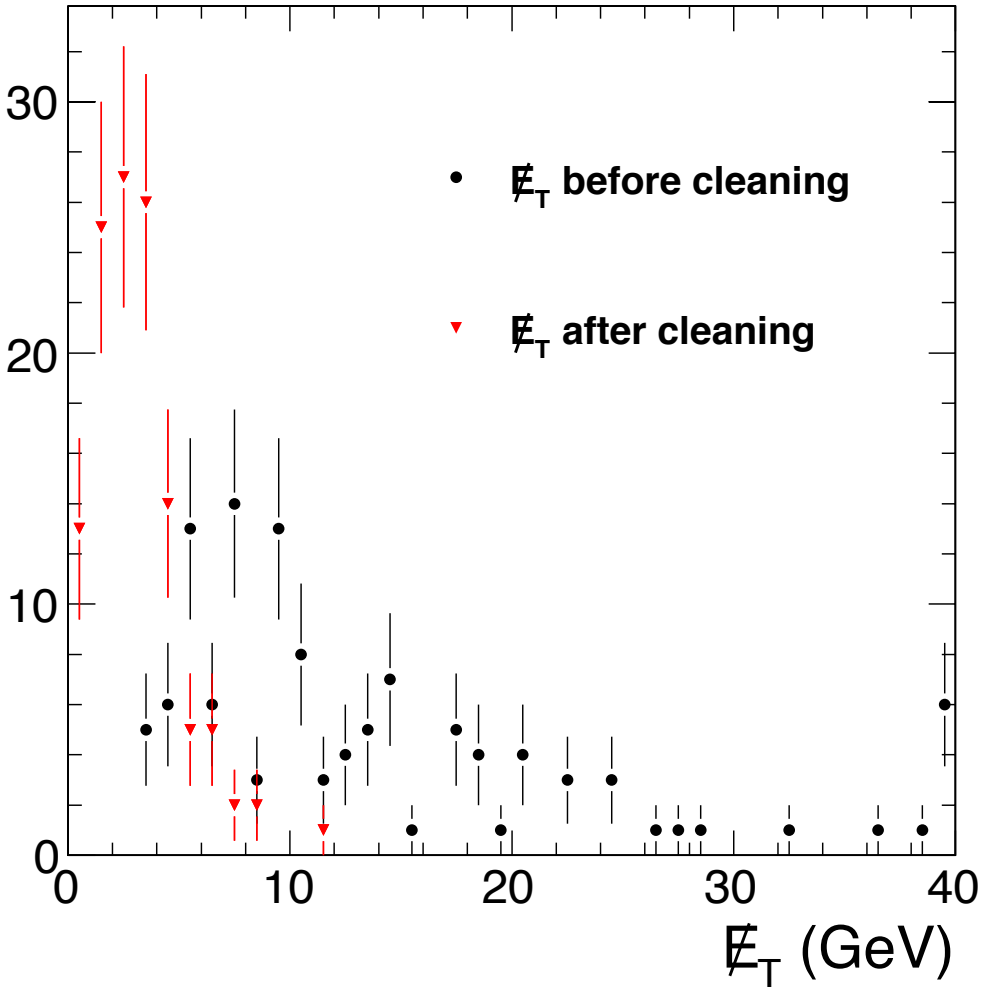

Figure 2.24: Distribution of $\not\!\!E_T$ before and after cleaning ECAL spikes. Only the subset of events in data containing an ECAL crystal identified as a spike is shown.

2.9 Performance of $\not\!\!E_T$ in Data

This section presents a few brief results from studies of the performance of $\not\!\!E_T$ algorithms in data. Figure 2.25 shows tc$\not\!\!E_T$ distributions from data and simulation for minimum bias and di-jet events using approximately 11.7 nb^{-1} of $\sqrt{s} = 7$ TeV data.

These events are expected to have no real $\not\!\!E_T$. Data distributions are after applying the cleaning algorithms discussed in Sections 2.8.1 and 2.8.2. Overall, the data and simulation agree well. There remain some residual tials in the $\not\!\!E_T$ distribution. Some fraction of these events are understood to still be due to noise in the ECAL, HF and HCAL that the $\not\!\!E_T$ noise algorithms and filters were unable to correct for at this stage in data taking. The distribution on the right is for events selected by requiring two JPT jets with $p_T > 25$ GeV and $|\eta| < 3$. Good agreement is observed between simulation and data, although a few outliers exist in the data at high $\not\!\!E_T$ values.

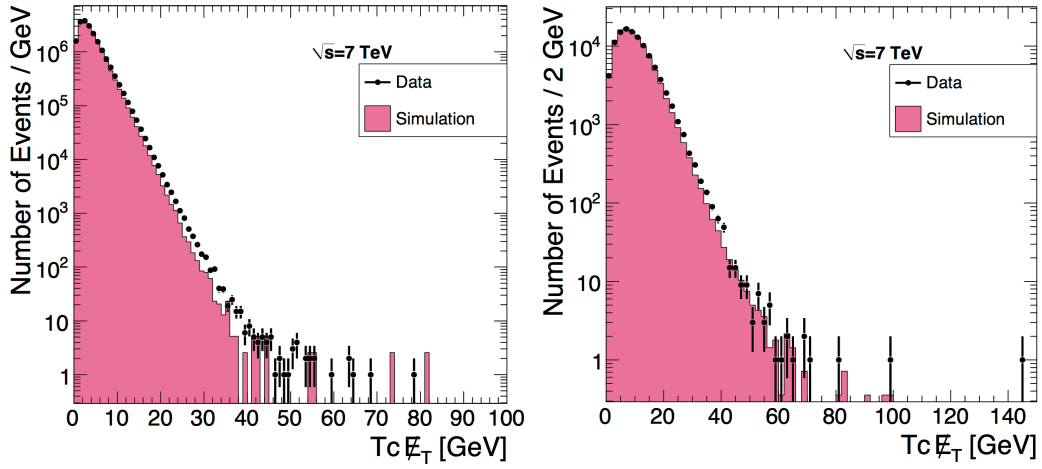

Figure 2.25: Comparison of tc$\not\!\!E_T$ distributions in data and simulation for minimum bias events (left) and di-jet events (right). Distributions were made using approximately 11.7 nb^{-1} of $\sqrt{s} = 7$ TeV data [17].

Figure 2.26 shows $\not\!\!E_T$ distributions for calo$\not\!\!E_T$, tc$\not\!\!E_T$ and pf$\not\!\!E_T$ using 36 pb^{-1} of $\sqrt{s} = 7$ TeV data. Events were selected by requiring an electron or muon from a single lepton trigger with $p_T > 25$ GeV. Electrons were required to have $|\eta| < 2.5$, while muons were accepted if $|\eta| < 2.1$. Additionally, the event was required to have at least 25 GeV of $\not\!\!E_T$ and the transverse mass formed from the lepton and the missing transverse momentum was required to be greater than 50 GeV. The latter two requirements were imposed to suppress background from QCD processes. These requirements select a sample enriched in $W \to \ell\nu$ events where real $\not\!\!E_T$ is expected from the neutrino that accompanies the lepton.

The distributions evidence good agreement between data and simulation for all three algorithms. Both tc$\not\!\!E_T$ and pf$\not\!\!E_T$ demonstrate better separation between $W \to \ell\nu$ events and those from background. The muon channel, shown in the lower row of plots,

has significantly less background than the electron channel, where mis-identified leptons from QCD processes contribute significantly to the final selection.

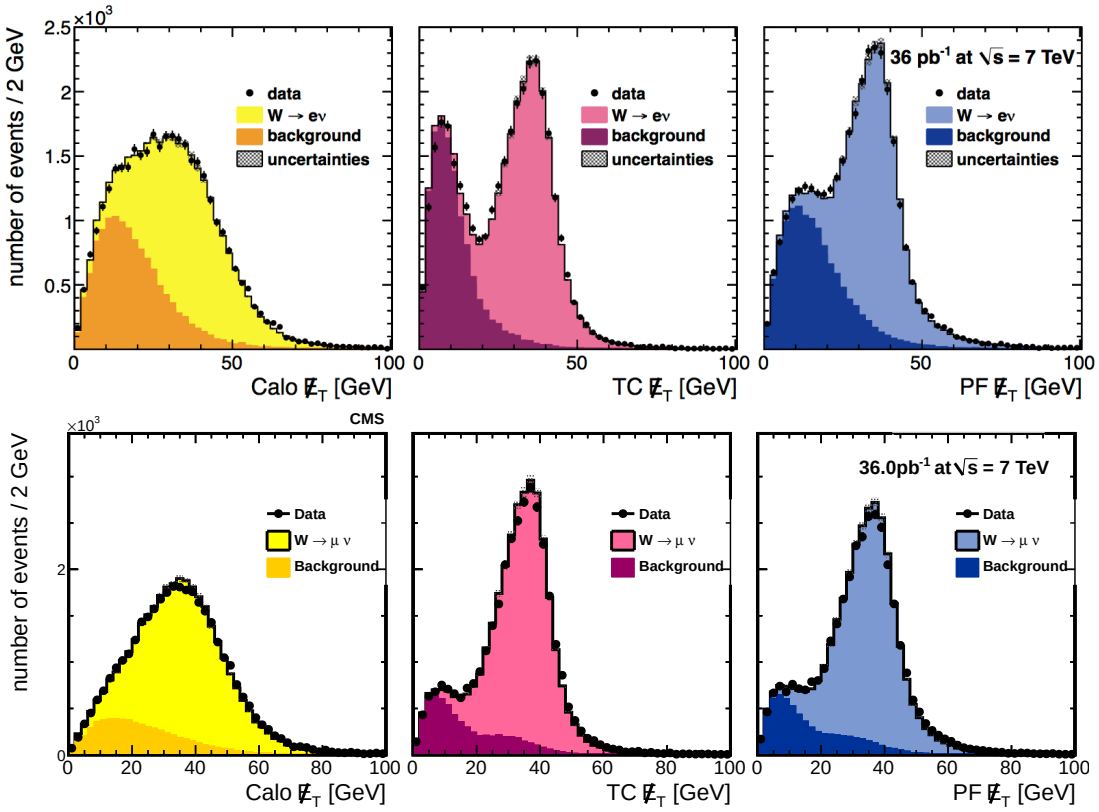

Figure 2.26: Comparison of $\not\!\!E_T$ distributions for different algorithms in $W \to e\nu$ (top) and $W \to \mu\nu$ (bottom) candidate events. The distributions correspond to the corrected calo$\not\!\!E_T$, tc$\not\!\!E_T$ and pf$\not\!\!E_T$ algorithms (left to right). Good agreement is observed between data and simulation for all algorithms. Both tc$\not\!\!E_T$ and pf$\not\!\!E_T$ show significant improvement compared to the baseline calorimeter-based $\not\!\!E_T$ [18].

Figure 2.27 shows the $\not\!\!E_T$ resolution for the three primary algorithms. The distribution was made using 36 pb^{-1} of $\sqrt{s} = 7$ TeV data. Di-jet events were selected by requring two jets with $p_T > 25$ GeV and $|\eta| < 3$. The resolution was determined from the width of a Gaussian fit to the $\not\!\!E_x$,$\not\!\!E_y$ distributions. The resolution is shown as a function of the total sum E_T in the event. Here, the sum E_T calculated using particle-flow candidates is used for all three algorithms. The sum E_T is calibrated to the particle-level sum E_T, on average, based on the relationship between the measured sum E_T and the particle-level sum E_T observed in Monte Carlo simulation. Again, both tc$\not\!\!E_T$ and pf$\not\!\!E_T$ show a significant improvement over the corrected calo$\not\!\!E_T$. The resolution of

the pf$\not\!\!E_T$ algorithm is slightly better than that for tc$\not\!\!E_T$, although the difference is not as large as one might expect given the differences in the complexity of the algorithms.

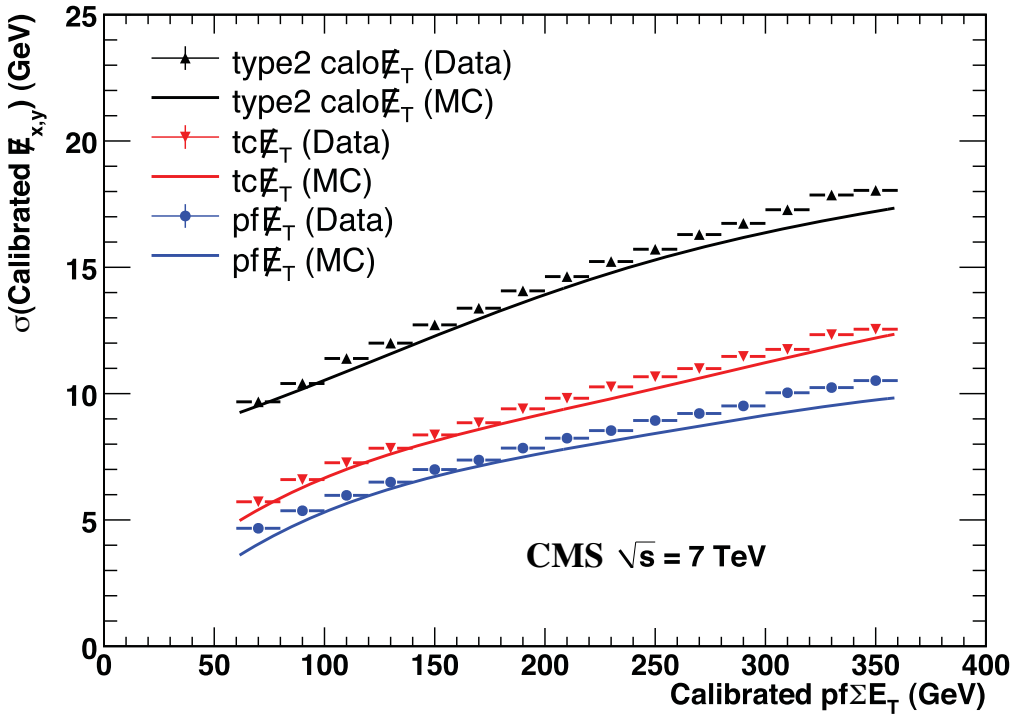

Figure 2.27: Resolution of the $\not\!\!E_T$ components for calo$\not\!\!E_T$, tc$\not\!\!E_T$ and pf$\not\!\!E_T$ as a function of the calibrated sum E_T of particle–flow candidates [17].

Chapter 3

Measurement of the $t\bar{t}$ Cross Section

3.1 Introduction

The measurement of the SM $t\bar{t}$ production cross section is one of the most important early analyses performed at the LHC. The analysis is important from a commissioning perspective, as it measures a process that is well understood from decades of study at the Tevatron. Additionally, it is vital to CMS as $t\bar{t}$ production in the dilepton final state involves the reconstruction of electrons, muons, jets and $\not\!\!E_T$. This complex signature requires the proper performance of all major sub–detectors. A proper determination of SM $t\bar{t}$ production cross section then is a requirement to convincing ourselves and others that the performance of the LHC and the CMS detector is understood at a level sufficient to have confidence in any results that may follow [19].

Many searches for new physics involve complex signatures such as those present in $t\bar{t}$ events. Importantly, SM $t\bar{t}$ is a major background for many new physics searches in final states involving leptons. Thus, understanding $t\bar{t}$ production and accurately modeling it in simulation is a necessary precursor to any new physics search of this type. Neither the CMS collaboration nor the community at large will have confidence in any signal of new physics if the important SM backgrounds are not well understood [20, 21].

We measured the SM $t\bar{t}$ production cross section using approximately 3.1 pb^{-1} of $\sqrt{s} = 7$ TeV data. The cross section was measured to be 194 $\pm$ 72 (stat) $\pm$ 24 (syst) $\pm$ 21(lum) pb. This value agrees with the NLO theory cross section of 157.5 pb.

This chapter starts with a description of the analysis selections, beginning with the trigger selection and moving on to lepton, jet and met selections. This is followed by a discussion of the trigger and lepton selection efficiencies. The following section describes data–driven methods used to estimate the primary backgrounds from off–peak Z/γ^* and from failures of lepton identification. The next section discusses the systematic uncertainties from experimental and theoretical sources. The final section summarizes the yields and background estimates and the resulting measured $t\bar{t}$ production cross section.

3.2 Analysis Selection

3.2.1 Datasets

Collision Data

This analysis uses data collected during the 2010 run at $\sqrt{s} = 7$ TeV. Dilepton events are selected from datasets collected using electron and muon trigger paths. These datasets are used for both the selection of a signal–like sample as well as for data–driven background estimates. Only runs and luminosity sections from good data taking periods are used, where such periods are defined using flags delivered by CMS data quality monitoring and detector and physics validation teams. The datasets correspond to an integrated luminosity of 3.1 pb^{-1}. A detailed list of the samples used and the associated run ranges can be found in Table 3.1.

Table 3.1: A summary of the collision datasets used for the $t\bar{t}$ analysis.

Dataset	Run Range
`/MinimumBias/Commissioning10-SD_EG-Jun14thSkim_v1`	135059 – 135802
`/MinimumBias/Commissioning10-SD_Mu-Jun14thSkim_v1`	135059 – 135802
`/EG/Run2010A-Jun14thReReco_v1`	135821 – 137028
`/Mu/Run2010A-Jun14thReReco_v1`	135821 – 137028
`/EG/Run2010A-PromptReco_v4`	138564 – 139459
`/Mu/Run2010A-PromptReco_v4`	138564 – 139459
`/EG/Run2010A-Jul16thReReco-v1`	139799 – 140159
`/Mu/Run2010A-Jul16thReReco-v1`	139799 – 140159
`/EG/Run2010A-PromptReco_v4`	140160 – 144114
`/Mu/Run2010A-PromptReco_v4`	140160 – 144114

Simulated Data

This analysis uses simulated data samples for the $t\bar{t}$ signal and relevant backgrounds. Simulated samples are used to study the systematics of the data–driven background estimation methods. The estimation of the Drell–Yan background uses the ratio of dilepton events in and out of a Z mass window determined from simulation. Studies in simulation show that the $V\gamma$ background is negligibly small and is ignored, in agreement with [22]. Other backgrounds, such as s/t–channel single top, are found top to be small and are considered to be negligible for the purpose of this analysis. Available QCD samples correspond to an integrated luminosity much smaller than what is used in data. The background from QCD multijet events is estimated using a data–driven method, while simulated QCD samples are used to evaluate some dependencies of the estimation method. Events in the simulated data samples are normalized to an integrated luminosity of 3.1 pb^{-1}.

Table 3.2 lists the samples and cross sections used. Samples are not listed for those processes considered to be negligible backgrounds. The Monte Carlo (MC) samples were generated using Pythia and MadGraph generators [23, 24]. Those samples with MadGraph in the name used the MadGraph generator to calculate the matrix element and Pythia for the hadronization. The remaining datasets were generated entirely with Pythia. All cross sections are NLO unless otherwise specified.

Table 3.2: A summary of the simulated datasets used for the $t\bar{t}$ analysis. Those samples with MadGraph in the name used the MadGraph generator for the matrix element and Pythia for the hadronization. The remainder of the samples were generated entirely with Pythia. Cross sections listed are NLO unless otherwise specified.

Sample	Cross Section (pb)	Comment
$t\bar{t}$ MadGraph	157.5	Inclusive
tW MadGraph	10.6	Inclusive
VV+jets MadGraph	4.8	LO, at least one lepton in final state
W+jets MadGraph	31314	$W \to \ell\nu$
Z/γ^*+jets MadGraph	3048	$M_{ll} > 50$ GeV
$Z/\gamma^* \to ee$ Pythia	5123	$10 < M_{ll} < 50$ GeV
$Z/\gamma^* \to \mu\mu$ Pythia	5123	$10 < M_{ll} < 50$ GeV
$Z/\gamma^* \to \tau\tau$ Pythia	1666	$20 < M_{ll} < 50$ GeV

3.2.2 Event Selection

To reject a known machine background resulting in high activity in the pixel layers, we require that no fewer than 25% of tracks are high purity in events having 10 or more tracks. Additionally, we reject events with significant noise in the hadronic calorimeters. Finally, to ensure that there was a reconstructed collision, we require at least one good vertex, as specified in Table 3.3. A fake vertex is defined to be one without any associated tracks. The number of degrees of freedom is determined by a sum of weighted tracks, while the ρ, z requirements are loose, requiring only that the vertex is within the luminous region [25].

Table 3.3: Good vertex selections.

Selections
vertex is not fake
n(dof) > 4
$\|\rho\| < 2$ cm
$\|z\| < 24$ cm

3.2.3 Trigger Selection

Events are selected using electron and muon triggers. Those with muons are selected from a sample collected by a trigger requiring an online muon object with $p_T >$ 9 GeV. Events with electrons are selected from a sample collected by single and double electron triggers as listed in Table 3.4. The electron trigger choice is more complicated as it followed changes in the trigger menu that evolved with the rapidly increasing luminosity. These triggers are chosen to have online requirements much looser than those used in the offline analysis selections. Only one electron trigger definition is used for all simulated data. Events in the $e\mu$ final state are required to pass either an electron or a muon trigger. Care is taken to exclude duplicate events.

3.2.4 Muon Selection

Muons are required to have $p_T > 20$ GeV and $|\eta| < 2.5$ and to be reconstructed by both the global and tracker algorithms. To further reject fake and poorly reconstructed muons:

Table 3.4: Electron triggers used for the $t\bar{t}$ analysis. The first trigger is used in both data and simulation.

Trigger Requirement	Run Range	Comment
$E_T > 10$ GeV	135059 – 137999	match online object, $p_T > 15$ GeV
$E_T > 15$ GeV	138000 – 143999	
$E_T > 15$ GeV, loose ID	144000 – 144114	
$E_T > 20$ GeV	144000 – 144114	
$E_T > 10$ GeV	144000 – 144114	2 electrons above E_T cut

- We require that the global fit of the muon have $\chi^2/\text{n(dof)} < 10$, to reject poorly reconstructed muons.

- We require that the muon have more than 10 silicon hits, to reject poorly reconstructed muons as well as muons originating late in the tracker from decays in flight.

- We require that the muon have $|d_0| < 200$ μm, where the impact parameter has been calculated with respect to the beam spot. A tight selection on the impact parameter reduces background muons from the decays of b-mesons, as well as decays in flight.

- We require that at least one muon sub-detector hit is used in the global fit.

- We require that the relative uncertainty of the muon p_T be less than 0.1. This requirement suppresses fake $\not\!\!E_T$ coming from mis-measured muon momentum.

- We require that relIso $\equiv E_T^{\text{iso}}$ / max $(p_T, 20) < 0.15$, where E_T^{iso} is the sum of transverse energy and momentum deposits in the ECAL, HCAL and tracker in a cone of $\Delta R = 0.3$, excluding deposits in the veto regions. Veto cones of $\Delta R = 0.01$, 0.07 and 0.1 are used in the tracker, ECAL and HCAL, respectively.

3.2.5 Electron Selection

Electrons, like muons, are required to have $p_T > 20$ GeV and $|\eta| < 2.5$. In addition:

- We require that the electron have $|d_0| < 400$ μm, where the impact parameter has been calculated with respect to the beam spot. The electron impact parameter

requirement is looser than that for muons as the uncertainty of the Gaussian sum filter (GSF) track is larger due to electron bremsstrahlung.

- We require that there is no global or tracker muon within a cone of $\Delta R = 0.1$ about the electron. This veto rejects a particular pathology where a muon is misreconstructed as an electron.

- We require that the electron pass the VBTF90 identification (see Table 3.5).

- To reject conversions, we require that there are no missing, expected inner hits. We further reject conversions via partner track finding, by requiring that there is no additional general track such that the dist < 0.02 and $\Delta \cot\theta < 0.02$, where the dist is the distance of closest approach between the electron and partner general tracks [26].

- We require that relIso $\equiv E_T^{\text{iso}}$ / max $(p_T,\ 20) < 0.15$, where E_T^{iso} is the sum of transverse energy and momentum deposits in the ECAL, HCAL and tracker in a cone of $\Delta R = 0.3$, excluding deposits in the veto regions. We subtract 1 GeV from the isolation sum in the barrel of the ECAL to account for a pedestal. A veto cone of $\Delta R = 0.15$ is used in the HCAL. The tracker and ECAL use the so-called Jurassic algorithm, which employs a strip 1.5 crystals in width, covering the full extent of the cone, to better account for the electron footprint [27].

The VBTF90 identification uses the selections in Table 3.5. Electrons are classified using the η of the supercluster. The selection values were chosen to have approximately 90% efficiency for electrons from W and Z decays.

Table 3.5: Details of the VBTF90 electron identification.

Selection Variable	Selection Value (barrel)	Selection Value (endcap)
$\sigma_{i\eta i\eta}$	< 0.01	< 0.03
$\Delta\phi_{\text{in}}$	< 0.8	< 0.7
$\Delta\eta_{\text{in}}$	< 0.007	< 0.009
H/E	< 0.12	< 0.05

The $\sigma_{i\eta i\eta}$ variable cuts on the shower shape in the ECAL. Electrons have a characteristic shower shape that is narrow in η but wide in ϕ, the latter due to bremsstrahlung. The $\Delta\phi_{\text{in}}$ and $\Delta\eta_{\text{in}}$ variables cut on the track-to-cluster matching. The final variable,

H/E, is calculated as a ratio of energies in the ECAL and HCAL veto cones. An electron is expected to deposit nearly all of its energy in the ECAL, while a hadron usually deposits significant energy in the HCAL, as well.

3.2.6 Dilepton Selection

Events are required to have at least one oppositely charged dilepton pair. We require a lepton trigger path to have fired and the leptons to pass the identification and isolation requirements discussed above. If more than one dilepton pair satisfies these criteria, we select the pair with the highest scalar sum of the lepton p_T. The rare effect of having multiple candidates is ignored in the estimate of the background from mis-identified leptons. To reject the contribution to the background from $Z/\gamma^* \rightarrow \ell\ell$, we exclude events with dilepton invariant mass between 76 and 106 GeV. Note that events with dilepton mass within 15 GeV of the Z peak are still used in the data-driven prediction of the off-peak Z/γ^* contribution to the background. Events with a dilepton mass below 10 GeV are rejected to avoid low-mass hadronic resonances. The low mass cutoff is expected to have almost no inefficiency for the signal.

3.2.7 Jet Selection

The signal events are expected to have jets from b quarks, with significant p_T, and less energetic jets from initial and final state radiation. As the backgrounds are not expected to have as much hadronic activity, we require the presence of at least two jets in the final state in order to suppress the background with little loss in signal efficiency. We use jets reconstructed with the anti-k_T algorithm with a distance parameter of R = 0.5 [28]. We count jets with corrected $p_T > 30$ GeVand $|\eta| < 2.5$. Additionally, the selected jets are required to not overlap within a cone of $\Delta R = 0.4$ with the selected leptons at each stage of the event selections.

For this analysis, we choose to use jets reconstructed in the calorimeter and corrected by replacing the expected deposition of charged particles with the momentum reconstructed in the tracker, in a manner similar to the tc$\not\!\!E_T$ algorithm. These so-called jet-plus-tracks (JPT) jets have better response and resolution than the progenitor calorimeter jets, although the under-response of the calorimeter together with the averaging effect of the correction still requires the application of the L2 and L3 jet energy corrections discussed in Section 2.4. Additionally, to suppress detector noise, we require

jets to pass loose jet identification criteria.

The jet energy scale uncertainty is one of the largest contributors to the $t\bar{t}$ cross section measurement. The expected scale uncertainty for JPT jets is 5% for p_T as low as 20 GeV, while the uncertainty of calorimeter-only jets is 10%. The smaller scale uncertainty is one of the driving factors in deciding to use JPT jets for this analysis.

3.2.8 $\not\!\!E_T$ Selection

Missing transverse energy is natural in dilepton $t\bar{t}$ events due to the presence of neutrinos from W decays. We impose a $\not\!\!E_T$ requirement to suppress background from $Z/\gamma^* \rightarrow \ell\ell$ as well as QCD multijet events. For this analysis, we choose to use tc$\not\!\!E_T$ as it is better at rejecting fake $\not\!\!E_T$ tails in Z/γ^* events (see Section 2.7). Dilepton events are required to have tc$\not\!\!E_T > 30$ GeV in same flavor final states. In $e\mu$ final states, where the background from Drell-Yan processes comes only via leptonic tau decays, the requirement is lowered to tc$\not\!\!E_T > 20$ GeV.

3.3 Selection Efficiencies

We estimate the selection efficiencies in data using a tag-and-probe method on Z events, described below, and compare them to the efficiencies obtained from simulation. No jet or $\not\!\!E_T$ requirements are imposed in the efficiency measurements in order to have enough events to make meaningful comparisons. Additionally, we compare the lepton efficiencies from Z events with those obtained from $t\bar{t}$ simulation after all analysis selections. The comparison is used to assess the systematic uncertainty on the signal event selections.

3.3.1 Trigger Efficiencies

Efficiencies are measured using a tag–and–probe based method. We select $Z \rightarrow ll$ events from a single lepton triggered sample by requiring opposite sign, same flavor leptons in a window around the Z mass, $|M_Z - M_{ll}| < 15$ GeV. Events outside of the Z mass window are not used to extract efficiencies, but are retained as sidebands to monitor the contamination from backgrounds. Contributions from W and QCD processes are found to be small. An event is tagged by requiring one lepton to pass the tight selections described in Section 3.2 and be matched to an online trigger object. Tagged events are

then classified into mutually exclusive categories.

- TT: Both leptons are matched to trigger objects and pass the tight criteria.
- TP: The probe fails the tight criteria, but passes a looser set of criteria chosen to not bias the selection criteria under study.
- TF: The probe fails the looser selection criteria.

Those events classified as TT should enter the efficiency calculation twice, as either leg could be used as the probe. Efficiencies are extracted following 3.1.

$$\epsilon = \frac{2TT + TP}{2TT + TP + TF} \tag{3.1}$$

The trigger efficiency for muons does not have a significant dependence on the muon momentum. There is, however, a significant dependence on muon rapidity. A summary of the single muon trigger efficiency is given in Table 3.6. Assuming the dimuon efficiency factorizes, the average trigger efficiency per dimuon event is estimated at 0.980 ± 0.005 in simulation and 0.977 ± 0.003 in data. Top dilepton events are expected to have a slightly higher efficiency as the muons tend to be more central. A conservative estimate of the difference is 1%.

Table 3.6: Single muon trigger efficiency in data and simulation in two pseudorapidity ranges. No background subtraction is necessary. The statistical uncertainty on the efficiency in simulation is negligible.

Source	$\|\eta\| < 2.1$	$\|\eta\| > 2.1$
Simulation	0.926	0.230
Data	0.882 ± 0.008	0.50 ± 0.04

Electron trigger efficiencies are consistent throughout the full data sample, even though different trigger are employed for different run ranges to select events. There is no significant dependence on electron momentum or pseudorapidity. Good agreement between data and simulation is observed, with the single electron efficiency measured to be 0.995 in simulation and 0.985 ± 0.003 in data. The trigger inefficiency on dielectron events is less that 0.1%. In combination with the single muon trigger above, the inefficiency of the trigger on $e\mu$ events is less than 0.3%, although it is somewhat different between data and simulation.

3.3.2 Electron Efficiencies

Electron identification and isolation efficiencies are measured using a tag-and-probe method similar to that used for the measurement of single lepton trigger efficiencies. We measure the the identification efficiency with respect to isolated electrons and the isolation efficiency with respect to those passing identification. This method is preferable, as the background contribution is large in dilepton pairs where the probe is required to only be a reconstructed electron. Table 3.7 shows the efficiencies measured in data and simulation. For the purposes of estimating a systematic uncertainty, we estimate an average difference in the identification requirement of 2% and about 1% for the isolation.

Table 3.7: Electron identification and isolation efficiencies in data and simulation measured using a tag-and-probe method on selected $Z \to e^+e^-$ events. The probe is required to pass identification (isolation) requirements for an estimate of the isolation (identification) efficiency. The uncertainty on the efficiency in simulation includes a 100% uncertainty on the background, when present. The uncertainty in data is purely statistical.

Source	$p_T < 40$ GeV	$p_T > 40$ GeV	$\lvert\eta\rvert < 1.5$	$\lvert\eta\rvert > 1.5$
Identification				
DY MC	0.924	0.937	0.947	0.871
DY/W+jets MC	0.906 ± 0.007	0.933 ± 0.004	0.942 ± 0.005	0.863 ± 0.008
Data	0.918 ± 0.008	0.935 ± 0.008	0.934 ± 0.006	0.886 ± 0.013
Isolation				
DY MC	0.961	0.984	0.970	0.972
DY/W+jets MC	0.948 ± 0.013	0.980 ± 0.004	0.961 ± 0.009	0.962 ± 0.010
Data	0.966 ± 0.005	0.988 ± 0.004	0.974 ± 0.004	0.972 ± 0.007

3.3.3 Muon Efficiencies

Dimuon events near the Z-mass peak are essentially background-free, and the efficiencies can be measured relative to a probe without identification or isolation requirements. We measure the identification efficiency with respect to such a probe. The isolation efficiency is measured with respect to an identified muon.

No dependence on muon momentum or pseudorapidity is observed. The muon identification efficiency is found to be 0.992 in simulation and 0.992 ± 0.002 in data. The isolation efficiency is found to be 0.981 ± 0.003 (0.970 ± 0.005) in data for $p_T <$ 40 GeV ($p_T >$ 40 GeV) and 0.980 (0.971) in simulation. After the estimates of the

identification and isolation efficiency are combined, a total systematic uncertainty of 0.5% is conservatively taken from the difference between data and simulation.

3.4 Background Estimation Methods

This analysis uses data-driven methods to estimate the primary backgrounds arising from failures of lepton identification and contributions from off-peak Z/γ^* in same flavor final states. The remaining backgrounds, including single-top (tW), diboson production and Drell-Yan in $\tau\tau$ final states are small and are taken directly from simulation. This section describes the off-peak Z/γ^* background estimate followed by a brief discussion of the estimate of the background from failures of lepton identification. A detailed discussion of the latter method can be found in the context of the search for new physics in final states with same sign dileptons in Section 5.2.

3.4.1 Estimating the Background from Off–Peak Z/γ^*

Drell-Yan events contribute significantly to the background in same flavor final states. A $\not\!\!E_T$ requirement is imposed to suppress this background, as discussed in Sections 3.2.8. Additionally, we veto same flavor dilepton pairs with an invariant mass between 76 and 106 GeV. These vetoed events are used as an independent control region. Despite these selections, Drell-Yan events remain a significant background because of the much larger production cross section for this process.

The total number of Drell-Yan events, N_{DY}, is the sum of events inside and outside the Z-mass peak, $N_{DY}^{\text{in}} + N_{DY}^{\text{out}}$. A scale factor between data and simulation is obtained by normalizing the event count in the control region in simulation to that observed in data. This scale factor can then be used to normalize the out–of–peak contribution from simulation, as shown in (3.2).

$$N_{DY}^{\text{out}(est)} = \frac{N_{DY,DATA}^{\text{in}}}{N_{DY,MC}^{\text{in}}} \cdot N_{DY,MC}^{\text{out}} \tag{3.2}$$

Grouping the two terms from simulation into a single term, $R_{\text{out/in}}$, the number of Drell-Yan events outside the control region can be estimated using the observed control region events according to (3.3).

$$N_{DY}^{\text{out}(est)} = R_{\text{out/in}} \cdot N_{DY,DATA}^{\text{in}} \tag{3.3}$$

The simple expression above does not account for contamination from signal and other background processes in the control region. Processes other than Drell–Yan that contribute to the near-Z control region come in two types: peaking and non-peaking. Peaking refers to contributions from the WZ and ZZ diboson processes that have a natural Z-mass peak. Simulation shows that the ZZ contribution is dominated by $ZZ \to \ell\ell\nu\nu$, which is covered by the above procedure. A significant fraction of contributing WZ events involve one lepton from the Z with the other coming from the W, together with a real source of $\not\!\!E_T$ from the neutrino that accompanies the lepton in the W decay. This contribution is not predicted by the above method, but studies in simulation show that the overall contribution from WZ is a small fraction of the total background and so it is neglected here.

The remaining non-peaking background comes from signal $t\bar{t}$ events as well as WW, tW and W+jets background. These processes contribute equally to the same and opposite flavor final states, and their contribution to the control region can be estimated directly using the opposite flavor yield in the near-Z region.

Thus, the estimate of the number of off–peak events coming from Drell-Yan and ZZ can be estimated according to (3.4).

$$N_{DY/ZZ}^{\mathrm{out}(est)} = R_{\mathrm{out/in}} \cdot \left(N_{ll,DATA}^{\mathrm{in}} - k \cdot N_{e\mu,DATA}^{\mathrm{in}}\right) \tag{3.4}$$

The constant, k, is equal to 0.5 times a correction due to differences in the reconstruction, identification and isolation efficiencies for electrons and muons. The factor of 0.5 accounts for combinatorics between the $e\mu$ and $\ell\ell$ final states. The efficiency correction factor is determined from the number of ee and $\mu\mu$ events in the control region after relaxing the $\not\!\!E_T$ requirement. This is expressed in (3.5), where it is assumed that the branching ratio to and acceptance for electrons and muons are the same. The correction factor takes the form of (3.5) when estimating the off-peak background in the $\mu\mu$ final state. For the ee final state, the ratio under the square root should be inverted.

$$\begin{aligned} \frac{n_{\mu\mu}^{\mathrm{obs}}}{n_{ee}^{\mathrm{obs}}} &= \frac{N_{\mu\mu}^{\mathrm{true}} \cdot A_{\mu\mu} \cdot \epsilon_{\mu}^{2}}{N_{ee}^{\mathrm{true}} \cdot A_{ee} \cdot \epsilon_{e}^{2}} = \frac{\epsilon_{\mu}^{2}}{\epsilon_{e}^{2}} \\ k &= \frac{1}{2}\sqrt{\frac{n_{\mu\mu}^{\mathrm{obs}}}{n_{ee}^{\mathrm{obs}}}} \end{aligned} \tag{3.5}$$

The correction factor is determined without applying a $\not\!\!E_T$ requirement to remove any difference due to $\not\!\!E_T$ mis-measurement in events with electron compared to muons. Such

mis-measurement is taken into account directly in the determination of the number of events observed in the control region after the $\not\!\!E_T$ cut is applied.

3.4.2 Estimating the Background From Lepton Misidentification

The data-driven method used to estimate the background arising from failures of lepton identification is discussed extensively in Section 5.2. Here, we restrict the discussion to the denominator definitions used in this analysis, the implementation of the correction and the resulting systematic uncertainties.

The muon denominator is based on the selections discussed in Section 3.2.4 with the following requirements relaxed:

- χ^2/n(dof) $< 50.$ (numerator cut is $< 10.$);
- $|d_0| < 2$ mm, taken with respect to the beam spot (numerator cut is 200 μm);
- relIso < 0.4 (numerator cut is 0.15).

The electron denominator is based on the selections discussed in Section 3.2.5 with the following requirements relaxed:

- remove the VBTF90 requirement;
- remove the impact parameter requirement (numerator cut is 400 μm).

The QCD background is estimated by counting the number of dilepton pairs passing the denominator selections, but failing the numerator selections. Each events is weighted by a factor of $\epsilon_{\rm fr}/(1-\epsilon_{\rm fr})$ for each lepton and the sum over all such events is the estimate of the QCD background, as expressed in (3.6)

$$N_{nn}^{\rm QCD} = \sum_{i,j} \frac{\epsilon_{\rm fr}^i \epsilon_{\rm fr}^j}{\left(1-\epsilon_{\rm fr}^i\right)\left(1-\epsilon_{\rm fr}^j\right)} N_{\bar{n}\bar{n}}^{ij} \tag{3.6}$$

The background from W+jets is estimated in a similar manner, selecting events with one lepton passing the numerator selection, while the other lepton fails the numerator selection, but passes the denominator selection. Each such event is weighted by an appropriate factor of $\epsilon_{\rm fr}/(1-\epsilon_{\rm fr})$ with the total W+jets background given by the sum over all such events (3.7).

$$N_{nn}^{W+\text{jets}} = \sum_{i,j} \frac{\epsilon_{\text{fr}}^{j}}{\left(1-\epsilon_{\text{fr}}^{j}\right)} N_{n\bar{n}}^{ij} \tag{3.7}$$

The total background arising from failures of lepton identification is expressed in (3.8)

$$N_{nn}^{\text{fakes}} = N_{nn}^{W+\text{jets}} - N_{nn}^{\text{QCD}} - \Delta_{\text{signal}}, \tag{3.8}$$

where Δ_{signal} corresponds to a contribution of real dilepton events which enter into $N_{n\bar{n}}^{ij}$. The correction for signal spillage is derived from the expected number of signal events from simulation together with a spillage rate, $\text{SR}_{\ell\ell}$, determined from data using Z events without a jet or $\not\!\!E_T$ requirement. The spillage rate is determined separately for electrons and muons. For muons, the rate comes directly from counting the number of dimuon events near the Z-mass peak for which one muon passes the numerator selection, while the other muon passes the denominator selection, but fails the numerator selection. Determining the spillage rate for electrons requires more care as contamination from jets that are reconstructed as lepton can effect SR_{ee}. Equation (3.9) defines the spillage rate for electrons,

$$SR_{ee} = \frac{1}{N_{nn}^{ee}} \sum_{i,j} \frac{\epsilon_{\text{fr}}^{j}}{\left(1-\epsilon_{\text{fr}}^{j}\right)} N_{n\bar{n}}^{ij,ee}, \tag{3.9}$$

where we have neglected second order effects coming from leptons that are jets in N_{nn} and contribute to Δ_{signal}. The spillage rate in $e\mu$ events is simply $SR_{e\mu} = 0.5(SR_{ee} + SR_{\mu\mu})$. The spillage rate for electrons is expected to extrapolate well to the signal region, as lepton identification efficiency is relatively independent of the number of jets in the event. The muon spillage rate is expected to be an underestimate, as the isolation efficiency decreases with increasing hadronic activity.

The systematic uncertainty on the data-driven estimate of the background coming from failures of lepton identification is primarily from differences between the momentum spectra of the progenitor partons in QCD and W+jets events and differences in the flavor composition of the QCD-enriched sample from which the fake rate is measured and the sample to which it is applied. Smaller sources of systematic uncertainty may come from contamination of the fake rate by leptons from electroweak processes and possible biases arising from differences in triggers used to select the fake-enriched and signal samples. For the purposes of this analysis, the systematic uncertainty on the

electron fake rate is taken to be 50% for the W+jets estimate, where a single factor of the fake rate enters, and 100% for the QCD estimate, where two factors enter. The systematic uncertainty on the background from mis-identified muons is 100% for the QCD background and +50%/-100% for the W+jets background estimate.

3.5 Systematic Uncertainties

Systematic uncertainties arise from uncertainties on event selections expected in simulation compared to the actual performance of the detector, from uncertainties on the fraction of events passing all selections due to uncertainties on the signal production and from the absolute normalization of the total number of expected events due to an uncertainty on the total integrated luminosity. The latter has an uncertainty of 11%. Uncertainties on the background estimates contribute to the systematic uncertainty of the cross section measurement. A summary of all systematic uncertainties is provided at the end of this section.

3.5.1 Systematic Uncertainty of the Lepton Selection

The systematic uncertainty of the lepton selection is estimated from tag-and-probe efficiency measurements in data and simulation. Based upon the combined identification and isolation efficiencies, a scale factor, SF, is derived such that

$$\epsilon_{\text{data}} = \epsilon_{\text{MC}} \cdot SF$$

Estimates of the lepton selection efficiencies in data and simulation were discussed in Section 3.3. Efficiencies between data and simulation were sufficiently similar that we take the scale factor to be unity with an uncertainty, δSF, matching the residual difference between data and simulation. An additional uncertainty of 2% per lepton is assigned on the isolation uncertainty to account for differences observed between simulated Z and $t\bar{t}$ events. Contributions to the systematic uncertainty on the lepton selections are summarized in Table 3.8. The uncertainties on electron and muon reconstruction efficiencies are from [29].

Table 3.8: Summary of contributions to the systematic uncertainty on the lepton selections, estimated for dilepton pairs from $t\bar{t}$ events. The values displayed are fractional uncertainties.

Source	ee	$\mu\mu$	$e\mu$	all
Trigger	0.1%	1.0%	0.3%	0.3%
Reconstruction	3.0%	3.0%	2.1%	2.1%
Identification and Isolation	5.0%	1.0%	2.5%	2.5%
Simulated $t\bar{t}$ vs Z	4.0%	4.0%	2.8%	2.8%
Combined	7.1%	5.2%	4.4%	4.4%

3.5.2 Systematic Uncertainty of the Jet and $\not\!\!E_T$ Selection

We vary the jet energy in simulation by ± 5% simultaneously with a ± 5% variation in the hadronic part of the missing transverse energy to test the sensitivity of the jet and missing transverse energy selections to the uncertainty in the energy scale. The variations in the number of events passing the full event selections are used as an estimate of the jet and missing transverse energy scale systematic uncertainties. The results are included in the summary table at the end of this section.

3.5.3 Theoretical Uncertainties on the Signal Production

We assess the systematic uncertainties due to settings used at the event production/generation step by comparing event yields in simulated $t\bar{t}$ samples produced with different configurations. Comparisons are made between samples with different ISR/FSR configuration and between samples simulated with different decay models (EVTGEN and Tauola).

The ratio of events passing full selections with larger (smaller) ISR/FSR compared to the nominal is 0.98 ± 0.01 (0.99 ± 0.01), where the uncertainties are statistical only. We assign a 1% systematic uncertainty due to ISR/FSR.

Hadronic and tau decays in simulation are expected to be better modeled with EVTGEN for heavy flavor and Tauola for tau decays. We compare samples with and without these decay models included. Each indicate an expected difference of about 2%, going in different directions. We assign a systematic uncertainty of 2% due to the hadron and lepton decay modeling based on these differences.

The branching ratio of $W \to \ell\nu$ decays used in the MadGraph samples is set to its leading order value of 1 / 9. The current world average is 0.1080 ± 0.0009 [3]. We

apply a scale factor equal to $(0.108 \cdot 9)^2$ with a fractional uncertainty of 1.7%, which is added to the total systematic uncertainty in each final state.

The effect on the acceptance of the PDF uncertainty was assessed in the $t\bar{t}$ signal Monte Carlo sample. The intrinsic uncertainty of the acceptance with the default CTEQ6.1 PDF was calculated using the weights method after selecting two leptons with $p_T > 20$ GeV and $|\eta| < 2.5$ at the generator level. Additionally, we required two generator level jets within the same pseudorapidity having $p_T > 30$ GeV and separated from each lepton by $\Delta R > 0.4$. The relative uncertainty on the acceptance was found to be approximately 0.5%. Uncertainties were also assessed using the MSTW 2008 NLO and CTEQ66 PDFs and the change was found to be less than 0.1%. Based on the negligible variation in acceptance with variation in the PDF parameters, no additional systematic uncertainty is assigned.

3.5.4 Systematic Uncertainties on the Background Estimates

Backgrounds considered in this analysis are estimated using data-driven techniques as well as simulation. The backgrounds from off-peak Drell-Yan in same flavor final states and from mis-identified leptons are estimated using data-driven methods discussed in Section 3.4. Simulation is used to estimate the backgrounds from single-top (tW) and diboson production and from Drell-Yan production in the ditau final state, where both taus decay to electrons or muons. The uncertainties on these backgrounds arise from the same sources as for the $t\bar{t}$ signal. In addition, there is an uncertainty on the total rate from the uncertainty on the production cross section and the luminosity uncertainty. Each background is assigned a conservative 50% fractional systematic uncertainty. This systematic uncertainty is reported in Table 3.9, relative to the expected cross section.

3.5.5 Summary of Systematic Uncertainties

Table 3.9 summarizes the systematic uncertainties from all sources.

3.6 Results

A summary of the number of events observed in data and the estimated contribution from backgrounds is given in Table 3.10. We attribute the excess of events above

Table 3.9: Summary of contributions to the systematic uncertainty on the signal selection and background estimation. Reported values are fractional, relative to the total cross section. The systematic uncertainty due to background is normalized to the expected standard model signal yield.

Source	ee	$\mu\mu$	$e\mu$	all
Lepton Selection	7.1%	5.2%	4.4%	4.4%
Energy Scale	3.8%	4.0%	3.4%	3.7%
ISR/FSR	1.0%	1.0%	1.0%	1.0%
Decay Model	2.0%	2.0%	2.0%	2.0%
Branching Ratio	1.7%	1.7%	1.7%	1.7%
Backgrounds	50%	40%	+ 10% -5%	15%
Total without Luminosity	50%	40%	+ 12% -8%	16%
Integrated Luminosity	11%	11%	11%	11%

the background expectation to events from top pair production in dilepton final states.

Table 3.10: Expected signal and background contributions compared to the number of events observed in data passing the full signal selection. Contributions from off-peak $Z/\gamma^*/ZZ$ and events with fake leptons are estimated from data. All other contributions are estimated from data. All other contributions are estimated from simulation.

Source	ee	$\mu\mu$	$e\mu$	all
Dilepton $t\bar{t}$	1.5 ± 0.1	1.7 ± 0.1	4.5 ± 0.3	7.7 ± 0.5
Diboson	0.03 ± 0.02	0.03 ± 0.02	0.08 ± 0.04	0.13 ± 0.07
Single–top (tW)	0.05 ± 0.03	0.05 ± 0.03	0.15 ± 0.08	0.25 ± 0.13
$Z/\gamma^* \to \tau\tau$	0.04 ± 0.02	0.07 ± 0.03	0.07 ± 0.04	0.18 ± 0.09
Off–peak Z/γ^*	$0.8 \pm 0.4 \pm 0.4$	$0.6 \pm 0.4 \pm 0.3$	–	$1.4 \pm 0.5 \pm 0.5$
Fake Leptons	$0.3 \pm 0.4 \pm 0.2$	$0.1 \pm 0.2 \pm 0.2$	$-0.3\ ^{+0.4}_{-0.1}\ ^{+0.2}_{-0.1}$	$0.1 \pm 0.5 \pm 0.3$
Total Background	1.2 ± 0.7	0.8 ± 0.6	$0.3\ ^{+0.4}_{-0.1}$	2.1 ± 1.0
Data	3	3	5	11

The signal production cross section can be measured according to:

$$\sigma^{t\bar{t}}_{\rm data} = \frac{N_{\rm obs} - B_{\rm est}}{\mathcal{L} \cdot \epsilon \cdot A \cdot {\rm BR}} = \sigma^{t\bar{t}}_{\rm theory} \frac{N_{\rm obs} - B_{\rm est}}{SF \cdot S_{\rm exp}}, \tag{3.10}$$

where $\sigma^{t\bar{t}}_{\rm theory} = 157.5$ pb is the cross section used to normalize the expected number of signal events in the simulation, $S_{\rm exp}$. The estimate of the number of observed signal events

is the difference between the number of observed events, $N_{\rm obs}$, and the estimated number of background events, $B_{\rm est}$. Differences between data and simulation are absorbed into the scale factor, SF. The scale factor accounts for differences in the expected normalization due to the estimated integrated luminosity, differences in the event selection and differences in theoretical modeling of the fraction of events passing event selections. Each difference corresponds to a single scale factor of which SF is the product.

We use the observed events and expected backgrounds in Table 3.10 together with (3.10) to derive the cross section, which we find to be

$$\begin{aligned} \sigma^{t\bar{t}}_{\rm data} &= 194 \pm 72\ ({\rm stat}) \pm 24\ ({\rm syst}) \pm 21\ ({\rm lum})\ {\rm pb} \\ &= 194 \pm 76\ ({\rm syst} \oplus {\rm stat}) \pm 21\ ({\rm lum})\ {\rm pb}. \end{aligned} \tag{3.11}$$

The result is less than one standard deviation away from the standard model cross section of 157.5 pb. The uncertainty on the measurement is dominated by the statistical uncertainty.

Chapter 4

Same Sign Analysis Selections

4.1 Introduction

The commissioning of the $\not\!\!E_T$ algorithms and the measurement of the $t\bar{t}$ production cross section provide confidence in the inputs requisite for many new physics searches. We build upon our work in these areas by conducting a search for physics beyond the SM in a final state with two same-sign leptons, missing transverse energy and significant hadronic activity. Many of the selections and background estimation methods used to measure the $t\bar{t}$ cross section are carried over, with some modification, to this new physics search.

LHC operation in 2011 produced nearly 100 million collision per second in the CMS detector, of which only approximately 100 are written to disk for later analysis. Most collision events come from well understood processes that are not of high interest, as can be seen in Figure 1.2, where the cross section for top production, the dominant background to this analysis, is already many orders of magnitude smaller than the total pp cross section. The CMS trigger system is responsible for selecting the most interesting events for processing and storage. Data is acquired using a trigger menu that represents the interests and compromises made amongst the members of the experiment and defines the physics program that the experiment can carry out. A trigger menu is composed of several hundred individual paths. Each path represents a series of decisions made using characteristic event signatures that are measured using specially designed electronics and software. Reconstructed data based on the set of trigger paths is organized by physics signature into a small number of primary datasets.

Despite the drastic reduction by the trigger system, nearly six orders of magnitude, in the number of events written to disk, only a small fraction are of interest to a particular analysis. Our analysis looks for events with two same sign leptons in primary datasets defined by dielectron, dimuon and muon–electron trigger paths. Leptons constituting our final state are expected to come from W/Z decays and thus we require the leptons to be prompt and isolated. In the context of this analysis, the word lepton is used to refer to an electron or a muon, including those secondary from tau decays. The word prompt is used to distinguish leptons from W and Z bosons as opposed to those coming from heavy flavor decays (b or c quarks) and decays in flight.

Reconstructed leptons are composite objects built from simpler elements measured in each sub-detector — i.e. the silicon tracker, calorimeters and muon system. Lepton identification selections require the individual elements be of high quality and well matched. Identification criteria help to differentiate between actual leptons, regardless of the source, and non-leptonic objects that leave similar signatures in the detector. Isolation concerns the amount of ambient energy — charged particles reconstructed in the tracker as well as energy deposits from charged and neutral particles in the calorimeters — surrounding the electron or muon. Leptons from electroweak bosons are solitary objects without significant surrounding activity, whereas those from the decay of colored particles tend to be surrounded by other colored particles. Thus, lepton isolation is a powerful handle for differentiation between prompt leptons from W and Z decays and those from other sources.

In addition to selecting events with two leptons with the same electric charge, we require $\not\!\!E_T$ and hadronic activity. Hadronic activity refers to the amount of energy in the event that is carried by colored particles. Quarks and gluons are frequently produced and subsequently shower and hadronize as they pass through the detector. Although we cannot measure the quarks themselves, the charged and neutral particles produced in the shower leave dense, characteristic signals in the silicon tracker and calorimeters that allow us to identify that an energetic colored particle was produced in the collision. A reconstructed jet is a cluster of associated tracks and calorimeter deposits that are the remnants of the originating quark or gluon.

Neutrinos and other, hypothetical, weakly interacting particles pass through the detector without interacting. The presence of such particles can be inferred from the imbalance of the total reconstructed momentum in a plane transverse to the beams. This

imbalance is referred to as the missing transverse momentum, $\vec{\not{E}}_T$, and its magnitude as the missing transverse energy, $\not{E}_T$.

We consider separately searches with high-p_T leptons, where both leptons are required to have $p_T > 10$ GeV and at least one must have $p_T > 20$ GeV and inclusive dilepton searches that lower the p_T thresholds to 5 GeV (10 GeV) for muons (electrons). Control regions are defined by selecting events with same-sign isolated dileptons, at least 2 jets and a moderate amount of $\not{E}_T$. These regions are chosen to be enriched in the SM processes that are the backgrounds to potential new physics, allowing us to test the data-driven background estimation methods. Signal regions are defined by tightening the requirements on the $\not{E}_T$ and the scalar sum of the p_T of selected jets (H_T).

This chapter discusses the selections and search regions. The first section lists the data sets used to conduct the analysis as well as simulated samples used for studies. The following section provides a detailed description of the various object selections. The section begins with a description of the trigger paths used and follows with a discussion of the identification and isolation requirements imposed on muons and electrons. A list of criteria for handling events with multiple candidates is provided. This is followed by a discussion of the selection of jets and an explanation of the $\not{E}_T$ requirement. The final section describes the pre-selection used to define control regions for the analysis and a definition of the $\not{E}_T$ and H_T selections used to define the signal regions.

4.2 Data Samples

4.2.1 Collision Data

This analysis uses data collected during the 2011 run at $\sqrt{s} = 7$ TeV. Dilepton events are selected from primary datasets collected using electron and muon trigger paths. These datasets are used for both the selection of a signal-like sample as well as for the data-driven background estimates. Only runs and luminosity sections from good data taking periods are used, where such periods are defined using flags delivered by CMS data quality monitoring and detector and physics validation teams. The datasets used correspond to an integrated luminosity of 0.98 fb^{-1}.

The details of the trigger paths used to define the primary datasets are complicated by the frequent changes to the CMS trigger menu necessary to accommodate increases in the instantaneous luminosity delivered by the accelerator. A precise expla-

Table 4.1: Primary datasets used by the high-p_T and inclusive analyses. Events are considered in the order of the datasets shown. An event passing the analysis selections is skipped if it has already appeared in a previous dataset. The last run in the `PromptReco-v4` datasets corresponds to the last good run used.

Dataset	Run Range
`/DoubleMu/Run2011A-May10ReReco-v1/AOD`	160329-163869
`/DoubleMu/Run2011A-PromptReco-v4/AOD`	165071-167784
`/MuEG/Run2011A-May10ReReco-v1/AOD`	160329-163869
`/MuEG/Run2011A-PromptReco-v4/AOD`	165071-167784
`/DoubleElectron/Run2011A-May10ReReco-v1/AOD`	160329-163869
`/DoubleElectron/Run2011A-PromptReco-v4/AOD`	165071-167784
`/MuHad/Run2011A-May10ReReco-v1/AOD`	160329-163869
`/MuHad/Run2011A-PromptReco-v4/AOD`	165071-167784
`/ElectronHad/Run2011A-May10ReReco-v1/AOD`	160329-163869
`/ElectronHad/Run2011A-PromptReco-v4/AOD`	165071-167784

nation of the lepton trigger paths used and the associated efficiencies can be found in Sections 4.3.1 and 6.3, respectively.

4.2.2 Simulated Data

This analysis uses simulated data samples for candidate signal models and the relevant backgrounds. The contribution of real same sign dileptons from diboson processes and $t\bar{t}W$ production is estimated directly from simulation. Simulated samples of Z/γ^* are used to cross-check the partially data-driven estimate of the background contribution from charge mis-reconstruction, as well as to study the systematic uncertainties of the lepton selection. The data-driven technique developed to estimate the background from fake leptons is applied to simulated $t\bar{t}$ and W+jets samples as one part of the study of systematics. Simulated SUSY samples are used for the study of systematic uncertainties as well as to develop a model of the selection efficiencies. These samples are also used in the extraction of limits. The contribution of double fakes from QCD processes is determined using a data-driven method, but the available simulated samples are used to study some dependencies of the fake lepton prediction method. Simulated data samples are normalized to an integrated luminosity of 0.98 fb^{-1}, unless otherwise stated. Details of the samples are summarized in Table 4.2.

Table 4.2: Cross sections used for simulated signal and background samples. The same sign W samples and the $t\bar{t}W$ sample do not use the full event simulation and reconstruction and do not simulate pile-up.

Sample	σ (pb)	Equivalent Luminosity (fb^{-1})	Comment	Generator
$t\bar{t}$	157.5	6.8	Inclusive	MadGraph, Tauola
tW	10.6	4.6	$t \to b\ell\nu$	MadGraph
t(s-channel)	1.4	354	$t \to b\ell\nu$	MadGraph
t(t-channel)	20.9	23	$t \to b\ell\nu$	MadGraph
W+jets	31314	0.48	$W \to \ell\nu$	MadGraph
Z/γ^*	15369		$Z/\gamma^* \to \ell\ell$, $M_{ll} > 10$ GeV	MadGraph, Pythia, Tauola
WW	43	2.6	$WW \to 2\ell 2\nu$	Pythia
WZ	18.2	116	Inclusive	Pythia, Tauola
ZZ	5.9	349	Inclusive	Pythia, Tauola
Vγ	173	6.4	$V \to 2\ell$, $p_T^\gamma > 20$ GeV	MadGraph
$t\bar{t}W$	0.079	2532	non-t W is $W \to \ell\nu$	Pythia
qqW^+W^+	0.165	811		Pythia
qqW^-W^-	0.055	3183		Pythia
DPS $W^\pm W^\pm$	0.378	529		Pythia
LM1	6.55	34	Inclusive	Pythia
LM6	0.40	550	Inclusive	Pythia

4.3 Event Selection

The event selection consists of the following general requirements:

- we require a lepton trigger;
- we select events with two good same sign isolated leptons; the high-p_T analysis requires both leptons to have $p_T > 10$ GeV and at least one to have $p_T > 20$ GeV; the inclusive analysis selects muons with $p_T > 5$ GeV or electrons with $p_T > 10$ GeV.
- we require $m_{ll} > 5$ GeV to reject low mass resonances;
- we require a significant amount of $\not\!\!E_T$ to reduce standard model backgrounds, particularly that from charge mis-measurement in $Z \to \ell^+\ell^-$ events;
- we require a significant amount of hadronic energy to reduce standard model backgrounds, particularly mis-identified leptons in W+jets events;
- we veto events if the invariant mass of either hypothesis lepton and a third good lepton in the event is consistent with the Z mass. This requirement is chosen to suppress the irreducible background from WZ and ZZ events.

To reject a known machine background resulting in high activity in the pixel layers, we require that no fewer than 25% of tracks are high purity in events having 10 or more tracks. Also, to ensure that there was a reconstructed collision, we require at least one good deterministic annealing vertex, as specified in Table 3.3.

4.3.1 Trigger Selection

We use a cocktail of double lepton trigger paths to select events in data. An event in the ee final state is required to pass at least 1 double electron trigger, a $\mu\mu$ event at least 1 double muon trigger and an $e\mu$ event is required to pass one of the electron-muon cross triggers. Because of the rapidly changing trigger menu, many trigger paths were not implemented in the simulation and thus no trigger requirement is made here. Instead, as discussed in Section 6.3, a trigger efficiency weight is applied to each event, based on the trigger efficiencies measured from data. The triggers are close to fully efficient except for the dimuon trigger at low p_T. A list of trigger paths used to select signal–like events can be found in Appendix D.

The high-p_T analysis selects events with two leptons having $p_T > 10$ GeV and additionally requires at least one lepton to have $p_T > 20$ GeV. Dielectron events in the high-p_T analysis are selected using triggers with E_T requirements of 17 and 8 GeV on the two legs. Additionally, the triggers impose online calorimeter identification and isolation requirements. The online selections are sufficiently loose compared to the analysis selections that little inefficiency is introduced (see Section 6.3 for details). However, care must be taken in the data-driven background estimate so that a bias is not introduced. This issue was studied in detail and found to not be a significant concern. The electron–muon cross triggers have no online isolation requirement on the electron leg. The cross triggers have muon p_T requirements of 17 GeV (8 GeV) and corresponding electron E_T requirements of 8 GeV (17 GeV). The dimuon triggers have only p_T requirements. The primary dimuon trigger in the `May10ReReco` dataset had a p_T requirement of 7 GeV on both legs. The p_T requirement increased to 13 GeV on one leg and 8 GeV on the other in the `PromptReco-v4` primary dataset.

The trigger paths used for the inclusive dilepton analysis have lower p_T requirements but impose an additional H_T requirement to keep the rate within the allotted bandwidth. The online H_T requirement was 160 GeV in the `May10ReReco` datasets and 150 GeV for later runs. Additionally, a dimuon trigger with a 200 GeV online H_T requirement was used throughout the data taking period. The dielectron triggers have an E_T requirement of 8 GeV on both legs. Additionally, the triggers impose online identification requirements on the shape of the energy deposit in the calorimeter and the matching between the calorimeter deposit and the reconstructed track. No online isolation requirement is made. The electron–muon cross triggers make the same identification requirements and select muons with $p_T > 3$ GeV and electrons with $E_T > 8$ GeV. The p_T requirement in the dimuon trigger was 3 GeV on both legs. Table 4.3 lists the online requirements placed on various trigger paths.

4.3.2 Muon Selection

Muon reconstruction was discussed previously in Section 2.3. To reject fakes with relatively little loss in efficiency, we require muons to be reconstructed as both global muons and tracker muons. To further reject fake and poorly reconstructed muons:

- We require that the muon have $p_T > 5$ GeV.
- We require that the muon have $|\eta| < 2.4$, which overlaps with the muon detector

Table 4.3: Online identification and isolation requirements used in the electron trigger paths.

Requirement	Barrel	Endcap
CaloIdL	$H/E < 0.15$	$H/E < 0.10$
	$\sigma_{i\eta i\eta} < 0.014$	$\sigma_{i\eta i\eta} < 0.035$
CaloIdT	$H/E < 0.10$	$H/E < 0.075$
	$\sigma_{i\eta i\eta} < 0.011$	$\sigma_{i\eta i\eta} < 0.031$
CaloIsoVL	ecalIso/$E_T < 0.2$	ecalIso/$E_T < 0.2$
	hcalIso/$E_T < 0.2$	hcalIso/$E_T < 0.2$
TrkIdVL	dη lt 0.01	d$\eta < 0.01$
	d$\phi < 0.15$	d$\phi < 0.10$
TrkIsoVL	trkIso/$E_T < 0.2$	trkIso/$E_T < 0.2$

coverage. Due to the finite luminous region, some muons can have a track with $|\eta|$ > 2.4, but we do not try to recover these as the inefficiency is small.

- We require that the global fit of the muon have χ^2/n(dof) < 10, to reject poorly reconstructed muons.

- We require that the muon have more than 10 silicon hits, to reject poorly reconstructed muons as well as muons originating late in the tracker from decays in flight.

- We require that the muon have $|d_0| < 200$ μm, where the impact parameter has been calculated with respect to the primary vertex. A tight selection on the impact parameter reduces background muons from the decays of b-mesons, as well as decays in flight.

- We require that at least one muon sub-detector hit is used in the global fit.

- We require that the energy deposits in the veto cones are less than 4 GeV in the ECAL and 6 GeV in the HCAL. Veto deposits are calculated in cones of size ΔR $= 0.07$, 0.1 for the ECAL and HCAL, respectively. Vetoing on the energy in the cones rejects fake muons produced when hadrons punch through the calorimeter into the muon system. This requirement can introduce an inefficiency for muons with high-p_T or in a busy event environment such as $t\bar{t}$–like events[30].

- We require that the inner track z be within 1 cm of the first good vertex, which

we take to be the event vertex. This requirement helps reject mis-reconstructed muons as well as those originating from pile-up interactions.

- We require that the relative uncertainty of the muon p_T be less than 0.1. This requirement suppresses fake $\not{E}_T$ coming from mis-measured muon momentum.

- We require that relIso $\equiv E_T^{\text{iso}}$ / $p_T < 0.15$, where E_T^{iso} is the sum of transverse energy and momentum deposits in the ECAL, HCAL and tracker in a cone of $\Delta R = 0.3$, excluding deposits in the veto regions. A veto cone of $\Delta R = 0.01$ is used in the tracker.

4.3.3 Electron Selection

Electrons are reconstructed in the CMS detector using 2 different algorithms. Tracker-driven electrons are seeded using silicon tracks and are primarily useful for reconstructing low-p_T electrons common from the decays of low mass resonances. ECAL–driven electrons are seeded by superclusters reconstructed in the calorimeter. The supercluster seed is matched to pixel hits in the inner layers of the tracker, from which an electron track is reconstructed. This tracking reconstruction uses a special Gaussian Sum Filter (GSF) algorithm that is designed to account for the energy loss an electron undergoes due to bremsstrahlung. The latter algorithm is optimized for electron p_T scales typical of those observed in electroweak processes. The reconstruction efficiency for ECAL–seeded barrel(endcap) electrons is measured to be approximately 99%(97%) using a tag-and-probe method on Z decays[31]. For this analysis, we require electrons to be ECAL seeded. In addition:

- We require that the electron have $p_T > 10$ GeV.

- We require that the electron have $|\eta| < 2.4$.

- We require that the electron have $|d_0| < 200$ μm, where the impact parameter has been calculated with respect to the primary vertex. A tight selection on the impact parameter reduces background from fake electrons and conversions.

- We require that there is no muon within a cone of $\Delta R = 0.1$ about the electron. Only muons passing the selections of Section 4.3.2 are considered. This veto rejects a particular pathology where a muon is mis-reconstructed as an electron.

- We require that the electron pass the VBTF80 identification (see Table 4.4), where the requirement on H/E is removed in the endcap. Removing the H/E requirement represents a departure from the published 2010 analysis [32] in order to reduce the loss of efficiency that occurs when applying this selection in a higher pile-up environment[33].

- We require that electrons with $p_T < 20$ GeV have $f_{brem} > 0.15$, or have $|\eta_{\rm SC}| < 1$ with $E/P > 0.95$. Nearly all electrons in the forward tracker radiate due to the larger material budget, and thus we reject all candidates in this region with $f_{brem} < 0.15$. The material budget in the central tracker is smaller, and thus we permit electrons with minimal bremsstrahlung provided that the calorimeter and tracker measurements are consistent[33].

- To reject conversions, we require that there are no missing, expected inner hits. We further reject conversions via partner track finding, by requiring that there is no additional general track such that the dist < 0.02 and $\Delta \cot\theta < 0.02$, where the dist is the distance of closest approach between the electron and partner general tracks [26].

- We require that the GSF track z be within 1 cm of the first good vertex, which we take as the event vertex. This requirement helps reject mis–reconstructed electrons as well as those originating from pile–up interactions.

- We require that all three charge measurements for an electron agree. One charge comes from the curvature of the GSF track. A second comes from the curvature of the associated CTF track. We require all electrons to have an associated CTF track. The last charge, the so-called supercluster charge, is determined using the relative position of the supercluster with respect to the projected track from the pixel seed.

- We require that relIso $\equiv E_T^{\rm iso}$ / $p_T < 0.15$, where $E_T^{\rm iso}$ is the sum of transverse energy and momentum deposits in the ECAL, HCAL and tracker in a cone of $\Delta R = 0.3$, excluding deposits in the veto regions. We subtract 1 GeV from the isolation sum in the barrel of the ECAL to account for a pedestal. A veto cone of $\Delta R = 0.15$ is used in the HCAL. The tracker and ECAL use the so-called Jurassic algorithm, which employs a strip 1.5 crystals in width covering the full extent of the cone, to better account for the electron footprint [27].

The VBTF80 identification uses the selections in Table 4.4, where the requirement on H/E in the endcap has been removed. The reported H/E value of 0.15 is implicit in the electron reconstruction algorithm. CMS defines the barrel/endcap transition to be at $|\eta| = 1.479$. Electrons are classified using the η of the supercluster. The selection values were chosen to have approximately 80% efficiency for electrons from W and Z decays.

Table 4.4: Details of the modified VBTF80 electron identification.

Selection Variable	Selection Value (barrel)	Selection Value (endcap)
$\sigma_{i\eta i\eta}$	< 0.01	< 0.03
$\Delta\phi_{\text{in}}$	< 0.06	< 0.03
$\Delta\eta_{\text{in}}$	< 0.004	< 0.007
H/E	< 0.04	< 0.15

The $\sigma_{i\eta i\eta}$ variable cuts on the shower shape in the ECAL. Electrons have a characteristic shower shape that is narrow in η but wide in ϕ, the latter due to bremsstrahlung in the tracker. The $\Delta\phi_{\text{in}}$ and $\Delta\eta_{\text{in}}$ variables cut on the track-to-cluster matching. The final variable, H/E, is calculated as a ratio of energies in the ECAL and HCAL veto cones. An electron is expected to deposit nearly all of its energy in the ECAL, while a hadron usually deposits significant energy in the HCAL, as well.

4.3.4 Lepton Pair Disambiguation

In events with multiple same sign dilepton candidates passing the selections described above, only one pair is selected according to the following prescription:

- We give preference to $\mu\mu$ pairs over $e\mu$ pairs, which are chosen over ee candidates.
- If multiple candidates remain, the pair with the highest scalar sum p_T is chosen.

4.3.5 Jet Selection

We require the presence of energetic hadronic activity in the event as new physics with a large cross section is expected to be produced via the strong interaction. We choose to use jets built from particle-flow candidates, as they provide the best scale and resolution for jets typical of those produced in SM processes such as $t\bar{t}$ events. Charged particle-flow candidates are reconstructed by matching silicon tracks to energy

deposits in the calorimeters and segments in the muon system. Neutral candidates are reconstructed from the remaining, unaccounted energy deposits in the calorimeter. The reconstructed candidates are clustered to form jets. Clustering is performed using the the anti-k_t algorithm [28]. We correct the p_T of all jets using a three tiered scheme. The first level of corrections subtracts the contribution from pile–up interactions and the underlying event. The second correction level is relative and accounts for differences in response across the detector. The third correction is absolute and scales the average jet response to unity.

The first correction level uses the L1FastJet correction. This method is an event-by-event, data-driven correction for the diffuse energy deposited in the detector by pile-up interactions. It accounts for both in-time and out-of-time pile-up as well as the underlying event. Using an algorithm that tends to cluster noise into soft jets, a distribution of jet p_T over jet area is constructed and the median of this distribution is the average diffuse energy per unit area, ρ. The p_T of each jet is then corrected by subtracting a term $\rho \times A$, where A is the jet area. This correction is valid provided that the pile-up energy is sufficiently dense, that the distribution is uniform in rapidity and azimuth, and that the number of pile-up jets is much larger than the number of jets from the hard interaction[34].

Jet momenta are further required to have L2 and L3 corrections applied. L2 corrections flatten the jet response in η while L3 corrections scale the average response to unity [35]. The corrections applied here were derived on simulation samples. Studies of the jet response in data show that the corrections derived from simulation are correct to within a few percent out to $|\eta| \sim 2$, beyond which the size of the residual correction increases [36]. This discrepancy is covered by a systematic uncertainty of 7.5% per jet.

Table 4.5: Selections for the loose particle flow jet ID.

Selection Variable	Selection Value	Comment
fraction of energy from neutral hadrons	< 0.99	
fraction of energy from neutral EM particles	< 0.99	
number of particle flow candidates	> 1	
fraction of energy from charged hadrons	> 0	$\|\eta\| < 2.4$
fraction of energy from charged EM particles	< 0.99	$\|\eta\| < 2.4$
number of charged particle flow candidates	> 0	$\|\eta\| < 2.4$

Jets used in this analysis are required to pass the loose particle flow jet ID

described in Table 4.5. These selections are very loose and are primarily intended to reject jets that are unambiguously due to detector noise. The inefficiency of this selection on simulation is less than 1% [37]. Additionally, jets are required to be separated by at least $\Delta R = 0.4$ from all hypothesis leptons and other leptons in the event.

For this analysis we require at least two particle flow jets passing the loose ID with $p_T > 40$ GeV and $|\eta| < 2.5$. The fiducial cut coincides with the extent of the tracker. In addition to counting jets, this analysis also makes a selection based on the total hadronic activity in the event, H_T. The H_T is calculated as the scalar sum of the p_T of all jets passing the jet counting criteria discussed above. Even a modest H_T requirement suppresses the background from failures of lepton identification in W+jets events. The two jet requirement made in all search regions imposes a minimal H_T requirement of 80 GeV.

4.3.6 $\not\!\!E_T$ Selection

Missing transverse energy is a natural requirement for new physics searches. Many models contain weakly interacting particles. For example, some of the SUSY models considered produce leptons via decay chains ending in a lightest, non-interacting particle, often generically called the LSP. The LSP escapes undetected, resulting in considerable missing energy. CMS uses 3 different $\not\!\!E_T$ reconstruction algorithms, discussed in detail in Chapter 2. For this analysis, we use $\not\!\!E_T$ reconstructed from a vector sum of particle flow candidates. Using tc$\not\!\!E_T$ gives similar performance, but we use pf$\not\!\!E_T$ as a compromise with other groups in CMS performing same sign searches. We impose a $\not\!\!E_T$ requirement of at least 30 GeV in all search regions. Even a modest $\not\!\!E_T$ requirement is effective at suppressing background from Drell-Yan processes in ee and $e\mu$ final states, where the charge of one of the final state electrons is mis–reconstructed.

4.3.7 Additional Z Veto

One of the primary irreducible backgrounds to the same sign dilepton search comes from WZ and ZZ production, where the bosons both decay to leptons. A natural same sign hypothesis is formed using a lepton from each of the two bosons. In the case of WZ, the lepton from the W comes together with a neutrino of the same flavor providing a natural source of $\not\!\!E_T$. To reduce this background, we reject events for which one of the hypothesis leptons and a third lepton in the event have an invariant mass consistent

with the Z, defined to be between 76 and 106 GeV. We require the third lepton to be of the same flavor and opposite sign and to pass the identification and isolation criteria described in Sections 4.3.2, 4.3.3.

4.3.8 Pile–up

The rapidly increasing luminosity resulted in a difference in pile-up conditions between simulation and data. This difference is accounted for by re-weighting events in simulation by the data-to-simulation ratio of the number of reconstructed vertices. Figure 4.1 shows the ratio used to re-weight the simulation. The distributions in data and simulation were obtained by counting the number of reconstructed deterministic annealing vertices in events with two leptons passing the selections of reference [38].

4.4 Definition of the Search Regions

We define a pre–selection with the following requirements:

- A dilepton trigger to have fired.
- Two same sign leptons passing the selections defined in Sections 4.3.2 and 4.3.3.
- At least 2 particle–flow jets passing loose identification with $p_T > 40$ GeV and $|\eta| < 2.5$
- $H_T > 80$ GeV
- pf$\not{E}_T > 30$ GeV
- Additional Z veto.

We start by defining control regions that are enhanced in the SM backgrounds. These regions provide sufficient statistics to compare observed yields with estimated backgrounds and gain confidence in the methods. In addition to the selections above, the high-p_T control region requires both leptons to have $p_T > 10$ GeV and at least one lepton to have $p_T > 20$ GeV. The inclusive search control region has lower p_T requirements, 5 GeV for muons and 10 GeV for electrons, and a higher H_T requirement of 200 GeV. The higher H_T requirement for the inclusive search results from the online H_T thresholds in the trigger.

Signal regions are defined in the H_T, $\not{E}_T$ plane as follows:

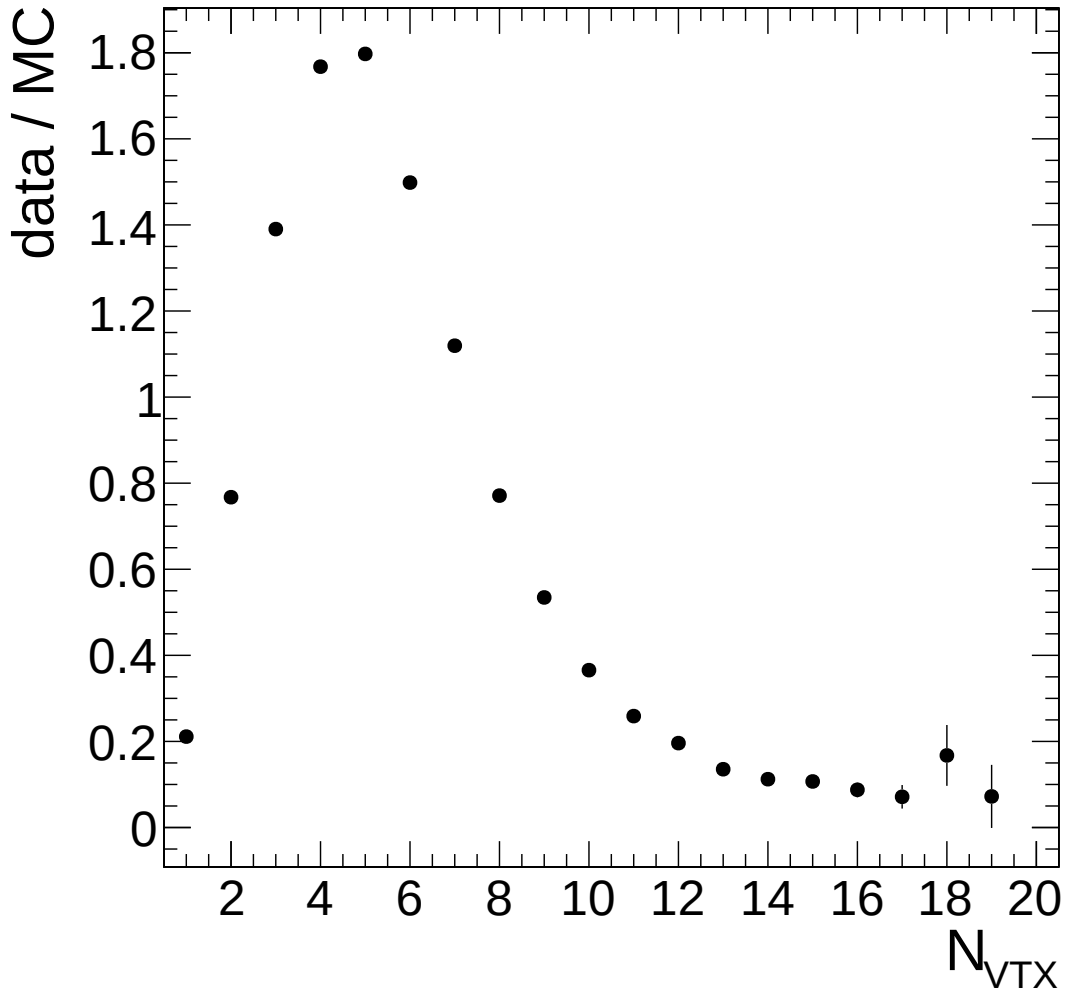

Figure 4.1: Data-to-simulation ratio of the number of reconstructed DA vertices. The simulation is re-weighted according to the above distribution to account for differences between the simulated pile–up and that observed in the data.

- $H_T > 400$ GeV, $\not\!\!E_T > 120$ GeV, inspired by SUSY models with low m_0;
- $H_T > 400$ GeV, $\not\!\!E_T > 50$ GeV, inspired by SUSY models with high m_0;
- $H_T > 200$ GeV, $\not\!\!E_T > 120$ GeV, inspired by a simplified model of squark–gluino production;
- $H_T > 80$ GeV, $\not\!\!E_T > 100$ GeV, inspired by the pMSSM model with sneutrino LSP (meaningful only for the high-p_T dilepton search).

Chapter 5

Background Estimation Methods

5.1 Introduction

At the level of theory, leptons are uncommon in the standard model. Prompt, isolated same sign dileptons are very rare. However, complications exist at the experimental level such that things aren't always what they seem. Not all reconstructed leptons are from W/Z decays. Some leptons are from heavy flavor decays, while other leptons aren't even leptons, but arise from failures of lepton reconstruction and identification. These aren't the leptons that we're looking for.

Chapter 4 described the selections used to differentiate interesting leptons from background. Even after these selections, background leptons remain. Same sign leptons primarily come from two sources: failures of lepton identification, often referred to as fake leptons or simply fakes, and failures of electron charge reconstruction. The latter are sometimes colloquially referred to as charge flips.

Fake leptons are the dominant background to this analysis. The term fake leptons refers to leptons from heavy flavor decays and decays in flight as well as jets that leave a lepton–like signature in the detector. In either case, fake leptons come from jets. For the fake lepton to pass identification and isolation selections, the underlying parton must have showered, or fragmented, in an unlikely manner. These occurrencesare not well modeled by the simulation. Thus, data-driven methods to estimate these backgrounds are needed.

A typical fake muon comes from the decay of a heavy flavor parton. For example, a bound state containing a b quark decays via a virtual W to a new bound state with

a c quark. A bit more than 20% of the time the W decays to an electron or muon. As this is a real lepton, the muon will pass the identification criteria with high probability. However, because of the low q^2 of the decay, the lepton from the virtual W will frequently have small separation from hadron and thus isolation provides a handle to differentiate this muon from a muon from a W or Z decay.

Fake electrons come from heavy flavor decays, in a manner similar to muons, as well as from light flavor jets. Details of the electron reconstruction algorithms make it less likely to reconstruct an electron embedded in a high-p_T jet as compared to a muon. Despite this, fake electrons are relatively common. Consider the case of a light flavor jet (i.e. a jet originating from a parton such as a u or a d quark) that fragments to a leading π^0 and a charged pion. The π^0 will deposit its energy in the ECAL, and if the track reconstructed from the charged pion is matched to the supercluster reconstructed from the π^0, an electron candidate can be formed. Electron identification is a powerful tool for rejecting this scenario, but fakes remain. If however, an electron comes from a heavy flavor decay, isolation remains the best handle for rejection.

Leptons from heavy flavor decays, for which isolation is the best differentiator from those from electroweak processes, pass the full analysis selection with a probability that depends on the momentum of the progenitor parton. If the lepton carries a large fraction of the momentum of the original parton or is emitted at a large relative angle, the lepton has a non-negligible chance of being isolated. The larger the difference between the momentum of the originating parton and that of the daughter lepton, the more ambient energy there is potentially surrounding the lepton and the less likely the lepton is to pass the isolation requirement. Thus, the probability for a lepton of a given momentum to pass the analysis selection is dependent upon the underlying parton kinematics.

We attempt to estimate the probability for a reconstructed lepton passing some loose criteria to also pass the full analysis selections by measuring the ratio in a statistically independent sample. We assume that the fake rate is universal. That is, a jet is sufficiently similar in all samples and the likelihood for a jet to fake a lepton is largely independent of the underlying process and kinematics. The validity of this assumption of universality is tested in both data and simulation and the measured dependence on the sample composition and the underlying kinematics is taken as a systematic uncertainty on the method.

Lepton charge is typically determined from the curvature of the reconstructed

track. The mis-measurement rate for muons is negligibly small at the W/Z momentum scale. The charge mis-reconstruction rate for electrons, however, is significant. Bremsstrahlung is emitted as electrons are deflected by the silicon atoms of the tracker. A radiated photon has a significant probability to convert into an e^+e^- pair due to the large material budget of the tracking system. The combination of the radiation of a hard photon and a subsequent asymmetric conversion (i.e., when one of the two resulting electrons carries a significant fraction of the momentum of the original photon) can result in the reconstruction of a track that combines hits from the original electron with those from the converted photon. As a result, the reconstructed track may have a curvature opposite to that of the prompt electron, resulting in an incorrect charge assignment. We estimate the background from this source by determining the electron charge mis–reconstruction rate using a simulated sample of electrons and applying that rate in data to events passing the full analysis selection described in Chapter 4, but with the same sign electric charge requirement replaced by an opposite sign one.

Standard model sources of prompt, same-sign dileptons are rare and thus are expected to make a small contribution to our data sample. The simulation-based prediction of the irreducible background coming from the processes in Table 5.4 comprises about 10% in the control regions, increasing to approximately 40% in the tightest signal region. The remaining 60–90% of the background is estimated using two data-driven techniques. This chapter is comprised of three sections. The first describes a data-driven method to estimate the number of events with fake leptons, the dominant source of background for this analysis. The second describes a method that uses information from both data and simulation to estimate the number of prompt electrons reconstructed with an incorrect charge sign. Finally, the chapter concludes with a brief discussion of the irreducible backgrounds estimated using simulated samples.

5.2 Data-driven Fake Rate Method

This section describes the method used to estimate the background from fake leptons. An overview of the method is provided, followed by a description of the determination of the fake rate, as a function of the lepton η and p_T, in a sample selected to be enriched in fakes. The fake rates themselves are given, followed by a detailed study of the systematics of the method using both Monte Caro and data. Closure tests in simulation are performed to assess the expected performance of the method in a signal–like region.

5.2.1 The Fake Rate Method

The idea of the fake rate (FR) method is that QCD data is used to measure a lepton FR as a function of the lepton p_T and η. The fake rate is defined as the probability for a lepton candidate passing a loose lepton selection to also pass the tight, analysis selection. Leptons passing the loose selection are called "fakeable objects" (FO). The fake rate is used to estimate the background from fake leptons as follows:

- Events are selected using all analysis cuts, except for the lepton selection. Events with one lepton passing the tight selection and one failing it, but passing the loose selection, are used to estimate the background to dilepton final states from one fake lepton. Backgrounds with two fake leptons are estimated by requiring both leptons to pass the loose selection and fail the full analysis selection.

- The background from a single fake lepton is estimated by weighting each tight-FO pair by a factor of $\epsilon_{\text{fr}}/(1-\epsilon_{\text{fr}})$, where ϵ_{fr} is the fake rate for the chosen definition of the FO.

- The background from two fake leptons is estimated by weighting each FO–FO pair by the product of two factors of $\epsilon_{\text{fr}}/(1-\epsilon_{\text{fr}})$, where the value of ϵ_{fr} is determined separately for each of the two FOs.

- The sum of the weights over the selected events is the background estimate.

Note that the estimate of single fakes for a pure sample of two fake leptons overestimates the background by a factor of two. Here, a fake lepton refers to a jet mis-reconstructed as a lepton as well as leptons from heavy flavor decays. Implicit in the above prescription are two assumptions upon which the method relies.

- We assume universality — i.e. that the lepton fake rate measured in an inclusive QCD sample represents the lepton fake rate in the signal sample.

- We assume locality — i.e. that the fake rate per lepton is independent for the two leptons, e.g. to predict the fake contribution to an $e\mu$ final state, we consider the electron and muon fakes separately, assuming no correlations between the two estimates. (Just like efficiency is physics dependent, the fake rate may be too, depending on the extrapolation.)

The validity of these assumptions is tested as part of the study of the systematic uncertainties of the method.

5.2.2 Fakeable Object Definitions

The fake rate method uses an extrapolation from the analysis lepton requirements to a looser lepton selection. A lepton passing the full analysis selection (or tight selection) is referred to as a "numerator" lepton. Similarly, a lepton passing the loose selection is defined as a "denominator" object. The fake rate itself is determined by the ratio of numerator to denominator leptons in bins of the p_T and η of the leptons. As described in Section 5.2.1, the estimation of the background counts leptons that pass the denominator selection, but fail the numerator. We will refer to these as non-numerator leptons.

The numerator selections were described in Chapter 4. The denominator selections are described below, specifying only looser selections. The muon denominator definition relaxes the following requirements with respect to Section 4.3.2:

- global fit χ^2/n(dof) < 50 (numerator cut is < 10);
- $|d_0| < 2$ mm, taken with respect to the primary vertex (numerator cut is < 200 μm);
- relIso < 0.4 (numerator cut is < 0.15).

The electron denominator definition relaxes the following requirements with respect to Section 4.3.3:

- the impact parameter cut is removed (numerator cut is < 200 μm);
- relIso < 0.6 (numerator cut is < 0.15).

We thus use an extrapolation in isolation and impact parameter to estimate the fake lepton backgrounds for electrons and muons. Based on simulation, we expect that the background is dominated by $t\bar{t}$ events, where the fake lepton is predominantly a real lepton from the semileptonic decay of a b quark. In this scenario, relaxing the isolation and impact parameter requirements, while keeping the identification tight, provides sufficient lever arm while keeping roughly the same composition in events with denominator leptons.

5.2.3 Definition of Fake Rate Datasets

The fake rate is measured in a multi-jet sample. The original idea was to use single jet triggers to obtain the sample. However, the low-p_T jet triggers are so heavily prescaled that this is no longer feasible. Instead, we select a sample using single lepton triggers. The triggers are chosen to coincide as closely as possible with the dilepton triggers used to obtain the signal sample. A list of primary datasets is given in Table 5.1 and for simulation in Table 5.2.

Table 5.1: Datasets used for the measurement of the fake rate. The measurement is performed using pre-scaled single lepton triggers.

Dataset	Run Range
`/DoubleElectron/Run2011A-May10ReReco-v1/AOD`	160329-163869
`/DoubleElectron/Run2011A-PromptReco-v4/AOD`	165071-167151
`/DoubleMu/Run2011A-May10ReReco-v1/AOD`	160329-163869
`/DoubleMu/Run2011A-PromptReco-v4/AOD`	165071-167151
`/SingleMu/Run2011A-May10ReReco-v1/AOD`	160329-163869
`/SingleMu/Run2011A-PromptReco-v4/AOD`	165071-167151

Events are required to have a lepton passing the denominator requirements discussed in Section 5.2.2. To enrich the sample in fakes, we require an away jet separated from the lepton by $\Delta R > 1$. The jet is a particle-flow jet that is required to pass the loose particle-flow jet ID listed in Table 4.5. The jet is further required to have $p_T >$ 40 GeV. The p_T is corrected using the L2L3 corrections together with the L1FastJet pile-up subtraction. To reject a known machine background resulting in high activity in the pixel layers, we require that no fewer than 25% of tracks are high purity in events having 10 or more tracks. Also, to ensure that there was a reconstructed collision, we require at least one good deterministic annealing vertex, as specified in Table 3.3. Only runs and luminosity sections certified as good are used in the determination of the fake rate.

The muon fake rate was measured using a cocktail of the `Mu5`, `Mu8`, `Mu12`, `Mu30` and `Mu8_Jet40` triggers. The number in the trigger names following the `Mu` refers to the online p_T requirement of the selected muon. The `Mu8_Jet40` triggers additionally have a requirement of a 40 GeV calorimeter–jet. The `Mu8_Jet40` triggers reside in the `DoubleMu` primary dataset. The remainder of the triggers are found in the `SingleMu` PD. The prescales of these triggers increase significantly over the selected run range.

Differences in the high-p_T and inclusive dilepton analyses necessitate different

Table 5.2: Monte Carlo datasets. All datasets are in the AODSIM data tier. The common part of each dataset name, `Spring11-PU_S1_START311_V1G1`, is replaced with the shorthand `Spring11`. The `MuEnriched` samples are used to extract the muon fake rates (`MuPt5Entriched` samples are used up to 15GeV, after which the `MuEnrichedPt15` takes over); the electron fake rates are extracted from the QCD samples.

Name	Cross section, pb	Luminosity, pb^{-1}
`/QCD_Pt_30to50_TuneZ2_7TeV_pythia6_Spring11-v1`	5.31×10^7	0.12
`/QCD_Pt_50to80_TuneZ2_7TeV_pythia6_Spring11-v1`	6.36×10^6	0.68
`/QCD_Pt_80to120_TuneZ2_7TeV_pythia6_Spring11-v1`	7.8×10^5	6.6
`/QCD_TuneD6T_HT-100To250_7TeV-madgraph_Spring11-v1`	7×10^6	0.17
`/QCD_Pt-20to30_MuPt5Enriched_TuneZ2_7TeV-pythia6_Spring11-v1`	2.38×10^8	9.1
`/QCD_Pt-30to50_MuPt5Enriched_TuneZ2_7TeV-pythia6_Spring11-v1`	5.31×10^7	14
`/QCD_Pt-50to80_MuPt5Enriched_TuneZ2_7TeV-pythia6_Spring11-v1`	6.35×10^6	74
`/QCD_Pt-80to120_MuPt5Enriched_TuneZ2_7TeV-pythia6_Spring11-v1`	7.85×10^5	109
`/QCD_Pt-20_MuEnrichedPt15_TuneZ2_7TeV-pythia6_Spring11-v1`	2.97×10^8	302
`TTToLNu2Q2B_7TeV-powheg-pythia6_Spring11-v1`	65.8	73×10^3
`WJetsToLNu_TuneZ2_7TeV-madgraph-tauola_Spring11-v1`	31314	483

electron fake rates. The high-p_T analysis uses dielectron triggers with an online isolation requirement, as listed in Table D.1. For consistency, we measure the fake rate using the `Ele8_CaloIdL_CaloIsoVL` and `Ele17_CaloIdL_CaloIsoVL` triggers, where the number refers to the online E_T requirement on the electron. We also use the `Ele8_CaloIdL_CaloIsoVL_Jet40` triggers which have an additional requirement of a 40 GeV calorimeter–jet. These triggers are found in the `DoubleElectron` PD. The prescales of these triggers change significantly over the acquired data.

Table D.4 lists the dielectron triggers used for the inclusive dilepton analysis. These triggers do not have an online isolation requirement. To maintain consistency, we use the `Ele8` and `Ele8_CaloIdL_TrkIdVL` triggers to measure the fake rate for the inclusive dilepton searches. The same fake rate is used for $e\mu$ events in the high-p_T dilepton analysis, as these triggers do not have an online isolation requirement on the electron leg (Table D.2).

5.2.4 Fake Rate Contamination from Prompt Leptons

Even after the away jet requirement, there is contamination from leptons coming from the decays of W and Z bosons. The contamination from $W \to \ell\nu$ is suppressed by requiring that the pf$\not\!\!E_T < 20$ GeV and the transverse mass $M_T < 25$ GeV. Contamination from $Z \to \ell\ell$ is reduced by rejecting events with a second FO for which the pair has an invariant mass $|M_Z - M_{\ell\ell}| < 20$ GeV. The Z veto requires that both fakeable objects have $p_T > 20$ GeV, although the second FO is only required to pass a looser selection used in our $t\bar{t}$ analysis (see Section 3.4.2).

Contamination from electroweak processes remains a problem at high lepton p_T, even after selecting against W/Z events. The contamination of the electron fake rate from these processes is small up to a p_T of approximately 55 GeV. This is evident in Figure 5.1, where here is no significant contribution from $W \to e\nu$ events observed. Evidence for the contamination being small is further supported by the observation of only a marginal increase in the fake rate in data after removing the W suppression requirements in data, as seen in Figure 5.2. Based on this evidence, we measure the electron fake rate in data for electrons with $p_T < 55$ GeV, and assume a constant value of the fake rate at higher electron p_T. The situation is similar for muons, but we only measure the fake rate for muons with $p_T < 35$ GeV, taking a constant value above, as the impact of W contamination is more pronounced for muons than for electrons.

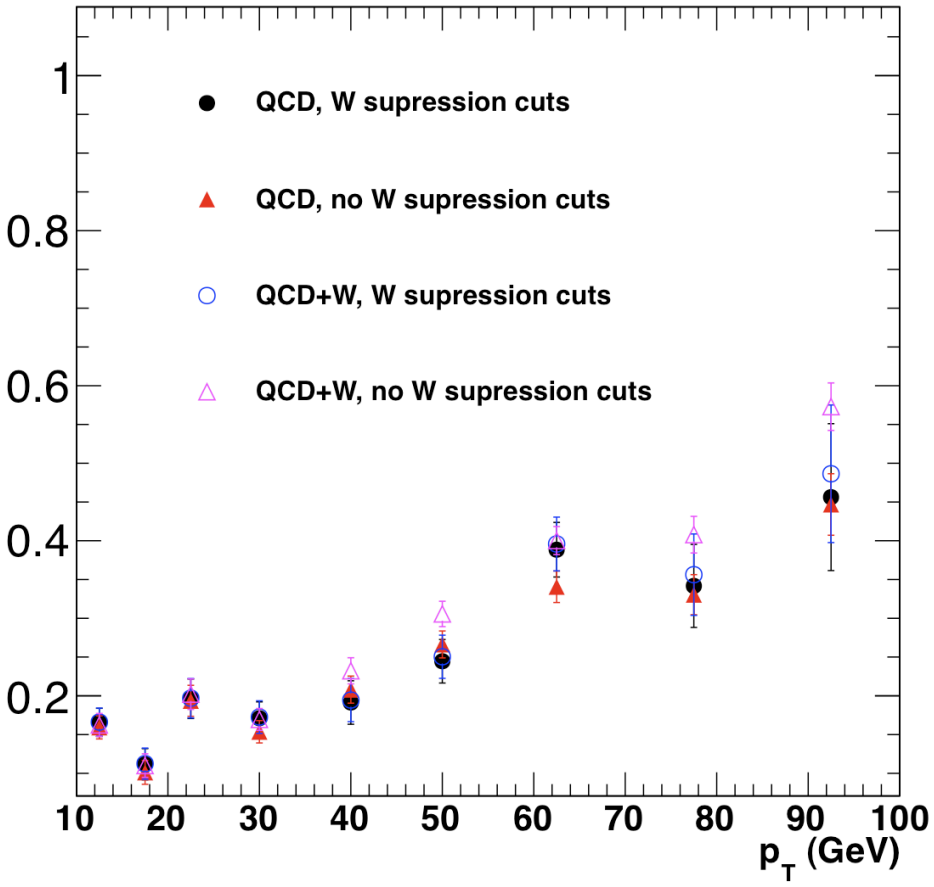

Figure 5.1: Electron fake rate as a function of the electron p_T measured over a range of selections enhancing the contribution from W/Z events. The fake rate measured in simulation is shown for the nominal measurement using only QCD samples (filled circles), and that with the W sample included (open circles), as well as for the selection without the W suppression measured in the QCD sample alone (filled triangles) and that with the W sample included (open triangles). Removing the W suppression cuts results in little increase in the value of the fake rate.

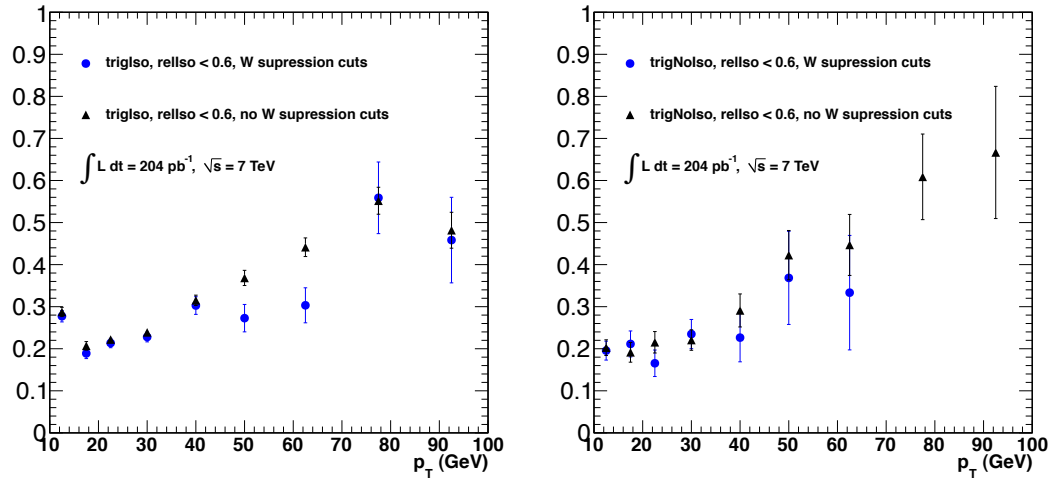

Figure 5.2: Electron fake rate as a function of the electron p_T measured with (filled circles) and without (filled triangles) W suppression cuts in data. Removing the W suppression cuts results in a small change in the fake rate for triggers with an online isolation cut. The fake rate measured in events selected using triggers with an online isolation cut shows more sensitivity to W contamination, as expected, although the value of the fake rate remains stable with p_T after applying rejection cuts. Based on these observations and those from simulation, the electron fake rate is measured out to a p_T of 55 GeV, and take to be a constant value at higher p_T.

5.2.5 Fake Rates for Electrons and Muons

The nominal fake rates are measured using the selections described in Section 5.2.2. As discussed in Section 5.2.3, the electron fake rates are measured separately for triggers with and without an online isolation requirement. The results of the measurement for electrons are summarized in Tables A.1 and A.2. The muon fake rates are measured using the single–muon triggers described in Section 5.2.3. The results are summarized in Table A.3.

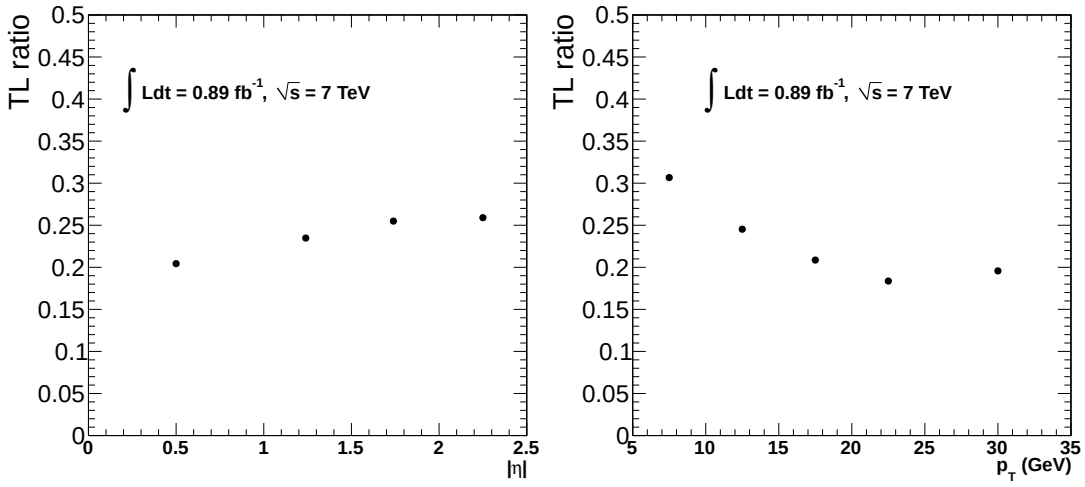

Figure 5.3: Projection in η and p_T of the muon fake rate measured in data. This fake rate is used for the high-p_T and inclusive dilepton analyses.

Figure 5.3 shows the η and p_T projection of the muon fake rate as measured in data. The fake rate is approximately flat in $|\eta|$. The muon fake rate decreases with increasing p_T, from a value of approximately 0.3 at 5 GeV to about 0.2 at high-p_T. The highest p_T bin shows a slight increase in the fake rate that may be an indication that W contamination is present, but the increase is not significant enough to warrant considerable attention at this time.

Figure 5.4 shows the same projections for the electron fake rate measured with and without an online isolation requirement. The fake rate is relatively stable, although it does show an upward trend at high-p_T. The fake rate measured using triggers without online isolation shows more variability, although the sample used is smaller due to increasing trigger pre-scales. Overall, the differences between the electron fake rates measured using triggers with and without an online isolation requirement are not large.

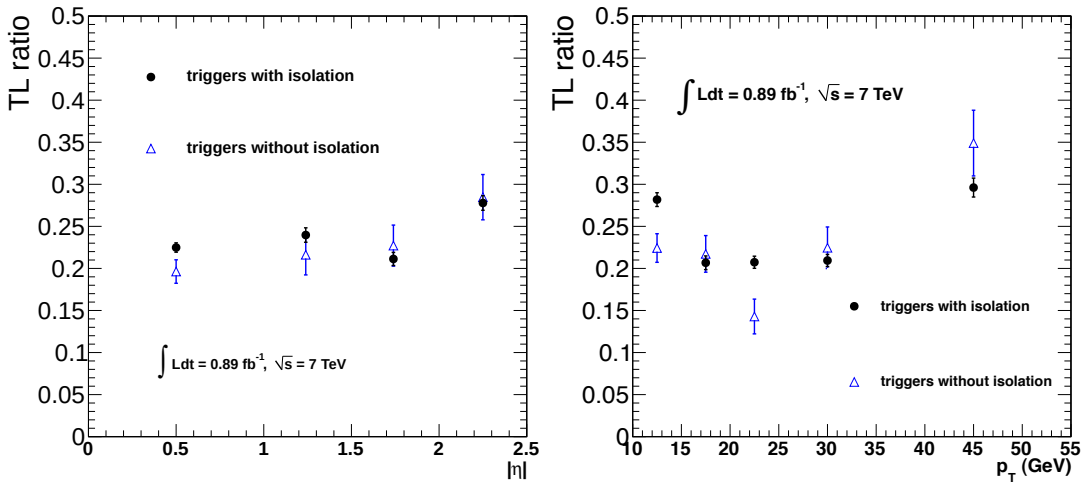

Figure 5.4: Projection in η and p_T of the electron fake rate measured in data using triggers with (filled circle) and without (open triangle) an online isolation requirement. The fake rate measured using triggers with an online isolation requirement is used for the high-p_T analysis in the dielectron final state. Triggers without an online isolation requirement are used to measure the fake rate for the inclusive dilepton analysis, as well as the $e\mu$ final state of the high-p_T analysis.

Similar fake rates were derived in simulation, but without a trigger requirement. The results are given in Tables A.4 and A.5 for muons and electrons, respectively. The corresponding projections in η and p_T appear in Figures 5.5 and 5.6. On average, the muon fake rate is slightly larger in simulation than in data. It is more interesting that the value of the fake rate falls much further in simulation than in data, decreasing from approximately 0.3 to 0.1 for muon p_T ranging from 5 to 35 GeV. The data begins to exhibit a flattening out beginning in the p_T 15–20 GeV bin and continues as one goes to higher p_T. This may be an indication of contamination from electroweak processes. The electron fake rate is quite stable in simulation with an overall lower value than that observed in data. Overall, the comparison of the trends is more interesting than a comparison of the absolute values, but simulation is not expected to model the fake rate well as it depends on details of the tails of the fragmentation functions as well as pile-up and the underlying event.

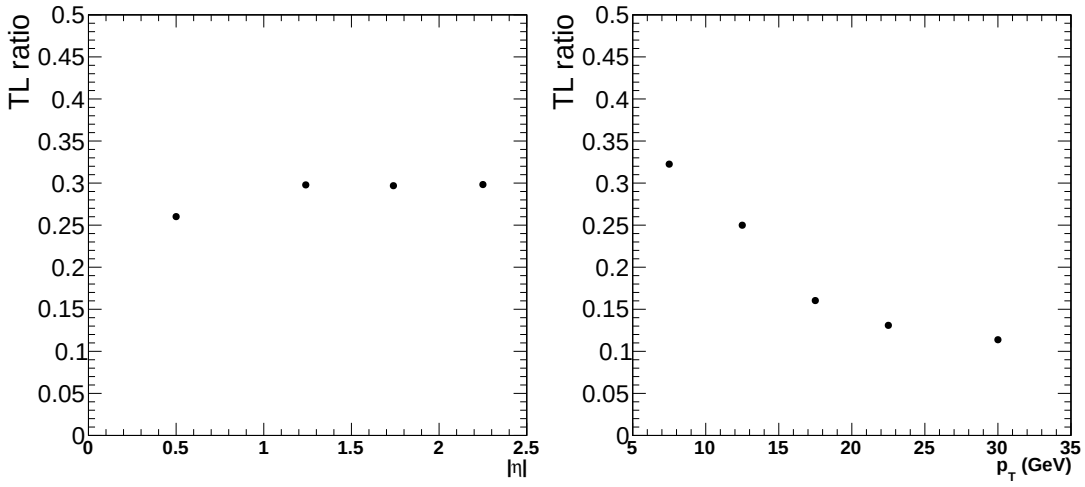

Figure 5.5: Projection in η and p_T of the muon fake rate measured in simulation.

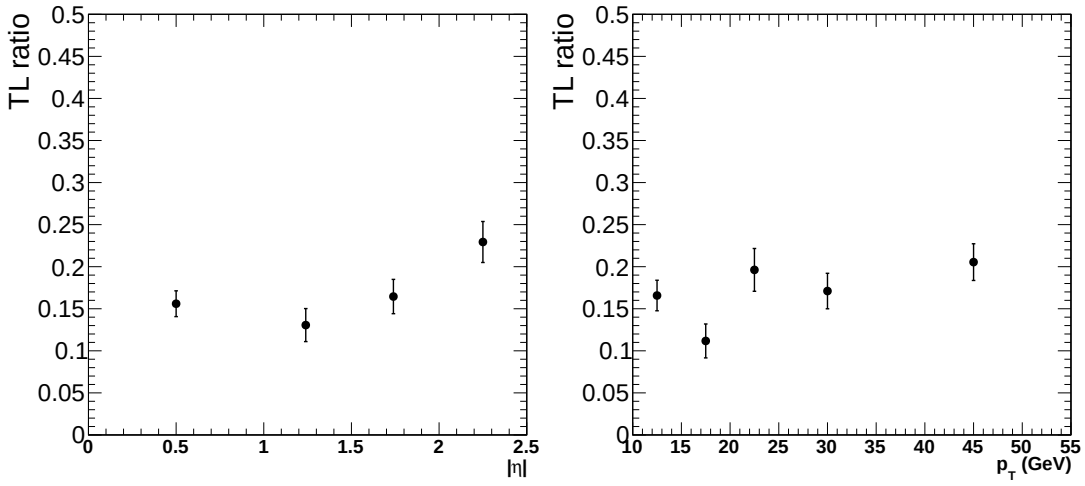

Figure 5.6: Projection in η and p_T of the electron fake rate measured in simulation.

5.2.6 Study of Fake Rate Dependences

Several studies are performed in data and simulation to assess the dependence of the fake rate on the sample composition and kinematics. Studies of the contamination of the fake rate in data due to prompt leptons was already considered in Section 5.2.4. Here we perform the following additional studies:

- Measurement of the fake rate dependence on the away-side jet p_T, as a measure of the dependence on the originating parton momentum.
- Measurement of the fake rate dependence on the sample composition. This is performed by requiring that the away-side jet is b-tagged, to enhance the heavy flavor contribution.
- Closure tests are performed using simulated W+jets and $t\bar{t}$ samples.

Fake Rate Dependence on Away Jet p_T

The lepton fake rates are dependent upon the away-jet p_T requirement because of the extrapolation in isolation. Lepton candidates in high-p_T jets have a smaller probability to pass an isolation requirement than a similar p_T lepton candidate in a low-p_T jet. The difference can be understood by considering the relative size of the lepton p_T and the $|p_T^{\text{parton}}$-$p_T^{\text{lepton}}|$. The latter term is a measure of the maximum transverse energy that can be deposited in the isolation cone. Thus, for a given lepton p_T, the higher the p_T of the originating parton, the less likely it is for the lepton to pass the numerator isolation cut. The probability flattens out when the momentum of the parton becomes much larger than that of the lepton candidate.

Figures 5.7– 5.9 show the η and p_T projections of the muon and electron fake rates measured in data with $p_T > 20$, 40 and 60 GeV away–jet requirements. The range of jet p_T considered was determined by looking at the typical p_T spectrum of b–quarks with FO daughters in simulated $t\bar{t}$ events. The dependence of the muon fake rate on the away-jet p_T requirement is on the order of 30%. The dependence is flat in $|\eta|$, but increases with the p_T of the muon candidate. The dependence of the electron fake rates on the away-jet p_T is similiarly on the order of 30%. There is not a strong trend in either $|\eta|$ or p_T. The dependence appears similar for electron fake rates measured with and without an isolation requirement.

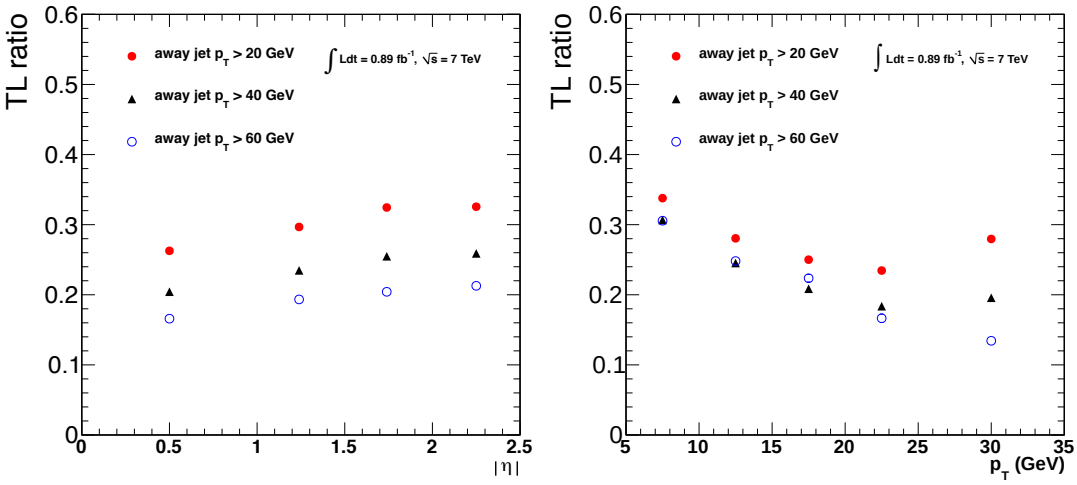

Figure 5.7: Projection in η and p_T of the muon fake rate measured in data for different away-jet p_T requirements. The away-jet is a corrected particle-flow jet and is required to be at least $\Delta R > 1.0$ away from the muon candidate.

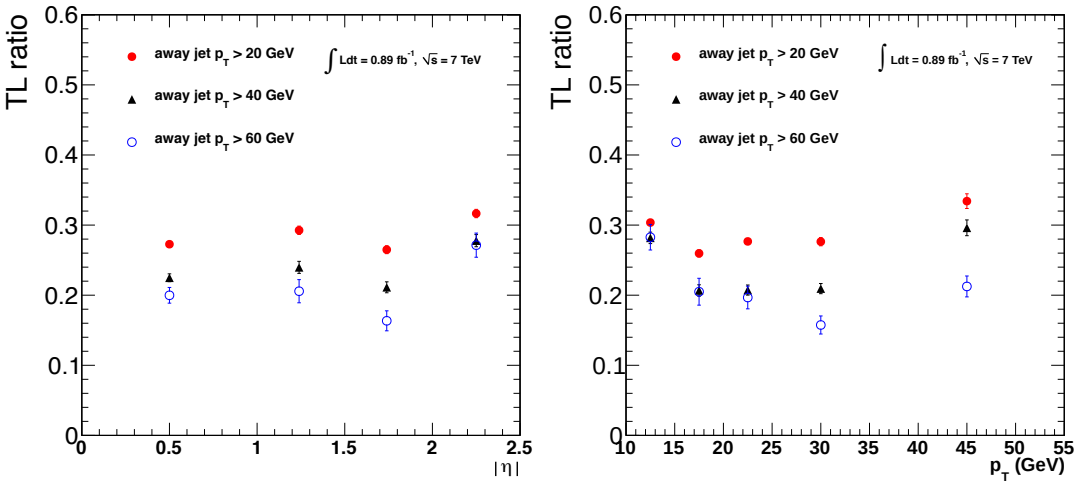

Figure 5.8: Projection in η and p_T of the electron fake rate measured in data for different away-jet p_T requirements. The away-jet is a corrected particle-flow jet and is required to be at least $\Delta R > 1.0$ away from the electron candidate. The electron fake rate is measured using triggers with an online isolation requirement.

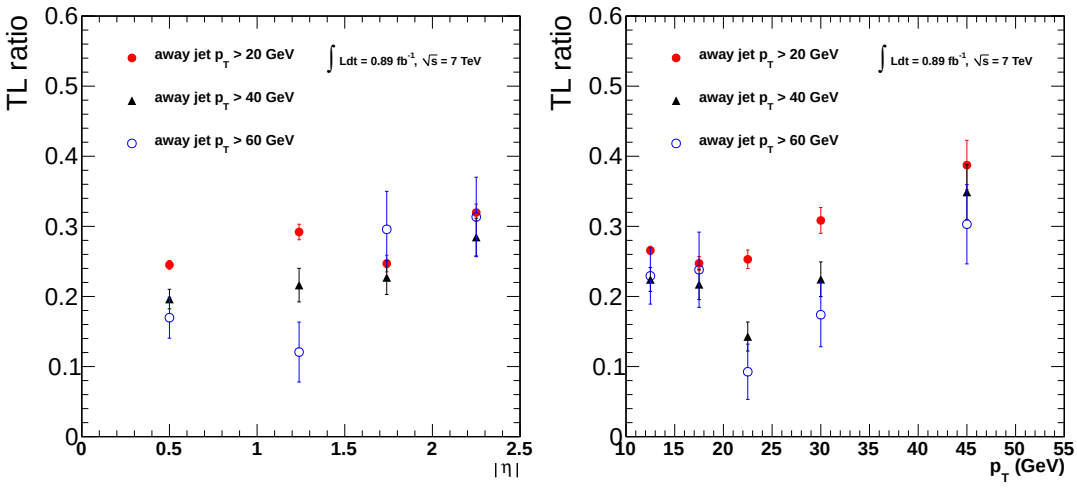

Figure 5.9: Projection in η and p_T of the electron fake rate measured in data for different away-jet p_T requirements. The away-jet is a corrected particle-flow jet and is required to be at least $\Delta R > 1.0$ away from the electron candidate. The electron fake rate is measured using triggers without an online isolation requirement.

Fake Rate Dependence on Sample Composition

Fake leptons come from different sources. Studies in simulation indicate that the primary background for the same sign analysis is $t\bar{t}$ events, where one of the leptons is prompt from the decay of a W and the other is a fake from a b decay. The fake rate is measured in a sample that is predominantly QCD. Similar studies in simulation indicate that the muons in this sample are also from the decay of a b quark, but that electron fakes come primarily from light flavor jets. To test the effect of sample composition on the fake rate and to obtain a sample closer in composition to the expected background, we measure the lepton fake rates using samples enriched in heavy flavor. This enrichment is accomplished by requiring that the away-jet is b-tagged.

Figure 5.10 compares the standard lepton fake rates with those measured requiring that the away-jet is b-tagged. The b-tagging algorithm used here is the simple secondary vertex high efficiency (SSVHE) algorithm with a requirement that the discriminator value is greater than 1.74. The top row shows the electron fake rate for triggers with (left) and without (right) an online isolation requirement. The bottom row shows the muon fake rate. None of the lepton fake rates show a strong dependence on the sample composition. There is, however, a drop in the muon fake rate at high muon

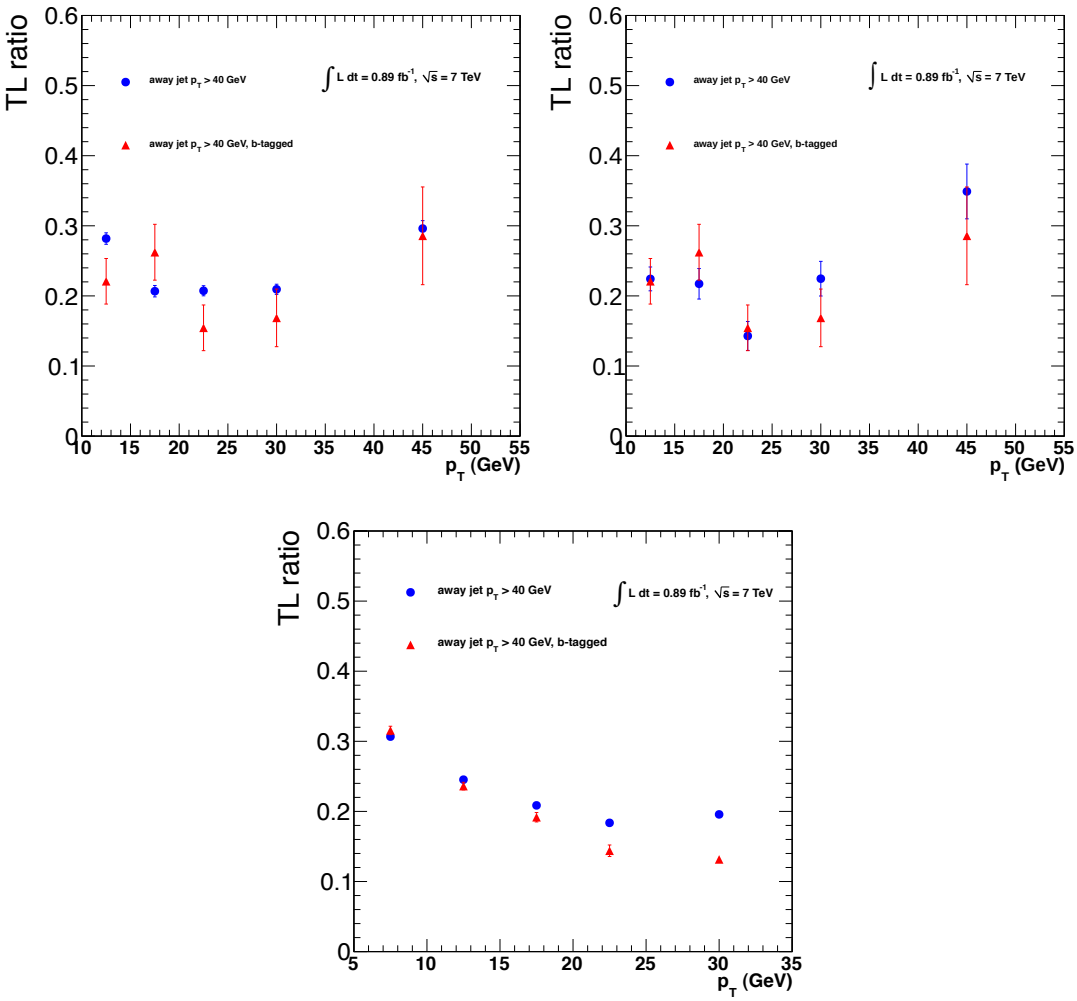

Figure 5.10: Projection in p_T of the lepton fake rates measured in data for different away-jet requirements. The away-jet is a corrected particle-flow jet with $p_T > 40$ GeV, is required to be at least $\Delta R > 1.0$ away from the lepton, with (filled triangle) and without (filled circle) a b-tagging requirement. We use the simple secondary vertex high efficiency (SSVHE) algorithm and require the discriminant have a value of at least 1.74. The top row shows the electron fake rate for triggers with (left) and without (right) an online isolation requirement. The bottom row shows the muon fake rate. Neither the fake rate for electrons nor muons shows a strong dependence on the heavy flavor content of the sample.

p_T when requiring that the away-jet is b-tagged that provides additional evidence that there may some contamination from $W \to \mu\nu$ for muon $p_T > 20$ GeV.

Fake Rate Closure Tests

Closure tests are performed by applying the lepton fake rates extracted from QCD Monte Carlo to simulated W+jets and $t\bar{t}$ events. We select the subset of these samples that contain one prompt lepton from the decay of a W and one fake lepton. Matching reconstructed leptons to those at generator level requires care as a small failure of the truth matching can significantly skew the outcome of the closure test. The closure tests are performed as follows:

- Select events passing the baseline selections.
- Require that one lepton is matched to a leptonic W decay and that it passes the numerator selection. Require that the other lepton is not matched to a leptonic W decay and that it passes at least the denominator selection.
- Scale the number of non-numerator leptons by $\epsilon_{\text{fr}}/(1-\epsilon_{\text{fr}})$. This is the prediction of the number of fake leptons passing the full lepton selections.
- Compare the predicted and observed numbers of fake leptons.

Table 5.3 shows the results of the closure test. The number of events with fake leptons in $t\bar{t}$ events is consistently overestimated for both electrons and muons by approximately 70%. These observations are consistent with studies performed during the 2010 analysis [39]. We attribute this overestimate to differences in the underlying parton momenta in $t\bar{t}$ and inclusive QCD events. The average parton momentum is higher in $t\bar{t}$ events resulting in a smaller effective fake rate. A marginally significant underestimate is observed for the prediction of the number of events with fake leptons in W+jets events. The statistical uncertainty of this test is much larger for muons than for electrons. An underestimate is expected if the average parton momentum in W+jets events is lower than in QCD. Based upon the results of the closure test, the systematic uncertainty for the fake rate method is taken to be 50%.

Table 5.3: Fake rate closure test on $t\bar{t}$ and W+jets events for the high-p_T dilepton selections. The muon FR test in $e\mu$ is done with $\not\!\!E_T > 20$ GeV. The number of events is scaled to 1 fb^{-1}. Except for the test in $t\bar{t}$ with electrons (done with jet $p_T > 40$ GeV), the results are reported for events with at least two jets with $p_T > 30$ GeV, used to increase the number of events passing the selections.

Sample	result	Electron FR		Muon FR	
		ee	$e\mu$	$\mu\mu$	$e\mu$
$t\bar{t}$	observed	2.8 ± 0.2	4.2 ± 0.2	3.9 ± 0.2	4.0 ± 0.2
	predicted	4.9 ± 0.4	6.8 ± 0.5	7.2 ± 0.3	6.5 ± 0.2
	ratio	1.8 ± 0.2	1.6 ± 0.2	1.8 ± 0.1	1.6 ± 0.1
W+jets	observed	< 2.1	8.4 ± 4.2	2.1 ± 2.1	
	predicted	1.5 ± 0.8	3.4 ± 1.4	2.1 ± 1.2	
	ratio	< 1.4	0.4 ± 0.3	1.0 ± 1.2	

5.3 Estimating the Charge Mis-measurement Background

This section describes the method used to estimate the background from leptons for which the sign of the charge is mis-reconstructed. The charge mis-measurement rate is determined as a function of η and p_T using a simulated sample of single electrons. The method is tested by selecting a sample in data enriched in $Z \to ee$ decays and comparing the observed same-sign yield with that expected by applying the charge mis-measurement rate to the opposite-sign sample. The charge mis-measurement rate is scaled to the data in the control region and the simulation is used to extend the coverage to high and low p_T.

As discussed in Section 4.3.3, we require agreement between all three charge measurements for an electron. This requirement reduces the charge mis-measurement rate for electrons to the point where it is almost a negligible background, accounting for less than 10% of the total background, as was shown in the 2010 analysis [32]. Even though this background is small, it is not necessarily well modeled in simulation. Thus, we use a quasi data–driven method.

Figure 5.11 shows the electron charge mis-measurement rate projected in p_T and η. The mis-measurement rate is approximately flat in the barrel, increasing sharply through the transition region and then falling slightly again in the endcap. Overall, the flip rate in the barrel is several times smaller than in the endcap. This is expected as the material budget is larger in the endcap, increasing the likelihood that a photon from bremsstrahlung converts. The mis-measurement increases monotonically with the p_T of

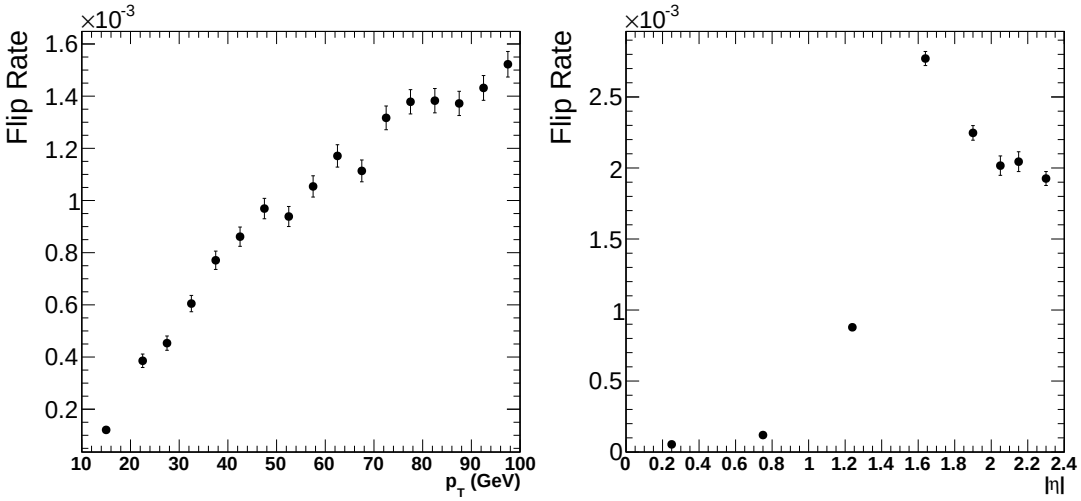

Figure 5.11: Charge mis–reconstruction rate projected in p_T (left) and η (right) measured using a single electron Monte Carlo sample. The rate is measured after requiring consistency amongst the three reconstructed electron charges.

the electron. This is expected as the curvature of the reconstructed track decreases with increasing momentum.

Studies in simulation indicate that the charge mis-measurement does not affect the energy scale of the reconstructed electron. Using this observation, we test the mis-measurement rate obtained from simulation is by applying it to an opposite sign $Z \to ee$ sample in data. Here, we select dielectron events passing the selections of Section 4.3.3 and having an invariant mass between 76 and 106 GeV. Additionally, to suppress contamination from W+jets, where one of the leptons is fake, we require pf$\not\!\!E_T < 20$ GeV and $M_T < 25$ GeV. The transverse mass is calculated using the lepton with the higher p_T. Each opposite sign event is scaled by

$$k = \frac{\epsilon}{1-\epsilon} + \frac{\epsilon^{'}}{1-\epsilon^{'}}, \tag{5.1}$$

where $\epsilon, \epsilon^{'}$ are the probabilities that the charge of each the electrons was mis-measured.

The sum over opposite sign events is the prediction of the number of same sign Z events coming from charge mis-measurement. The number of events expected from simulation is 94 $\pm$ 10. The predicted same sign invariant mass distribution is shown in Figure 5.12 together with that expected from simulation. The two distributions are

consistent within statistics. We find 129 events with same sign electron pairs in data in the Z control region, compared to an expectation of 100.0 $\pm$ 0.3 events from the opposite sign dielectron sample and 8 $\pm$ 4 events from fake electrons. Based on these observations, we assign a scale factor of 1.2 $\pm$ 0.2 to the expected number of same sign dielectron obtained using the opposite sign dielectron sample. The correction factor corresponds to the relative difference between 121 $\pm$ 11 (stat.) $\pm$ 4 (syst) and 100 $\pm$ 0.3. The systematic uncertainty is taken to be 20% to account for potential effects not covered by this test arising from differences in lepton kinematics between Z and $t\bar{t}$–like events.

5.4 Backgrounds Estimated from Simulation

Backgrounds arising from failures of lepton identification and electron charge mis-reconstruction together account for 60–90% of the estimated background in a given search region. The remaining 10–40% of the background is irreducible and is taken from simulation. A list of these processes and the associated cross sections is provided in Table 5.4. We neglect the contributions from WWW and WWZ as they are expected to be negligible.

Table 5.4: Sources of true same sign dileptons from Standard Model processes. Cross sections are leading order. The contribution of these processes to the background is taken directly from simulation.

Process	Cross Section (pb)
WZ	18.2
ZZ	5.9
Vγ	173
qqW^+W^+	0.165
qqW^-W^-	0.055
DPS $W^\pm W^\pm$	0.378
$t\bar{t}W$(non–t W is $W \to \ell\nu$)	0.079

The WZ, ZZ and $V\gamma$ backgrounds have been measured with the CMS detector [40, 41, 42]. The first two processes have prompt, isolated same sign dileptons in the final state. The background from $V\gamma$ involves a final state with one real lepton from the vector boson and a fake electron when the photon converts. However, this fake background is not estimated by the fake rate method as the fake rate is derived for fake leptons from jets in a QCD sample. The remaining backgrounds taken from simulation

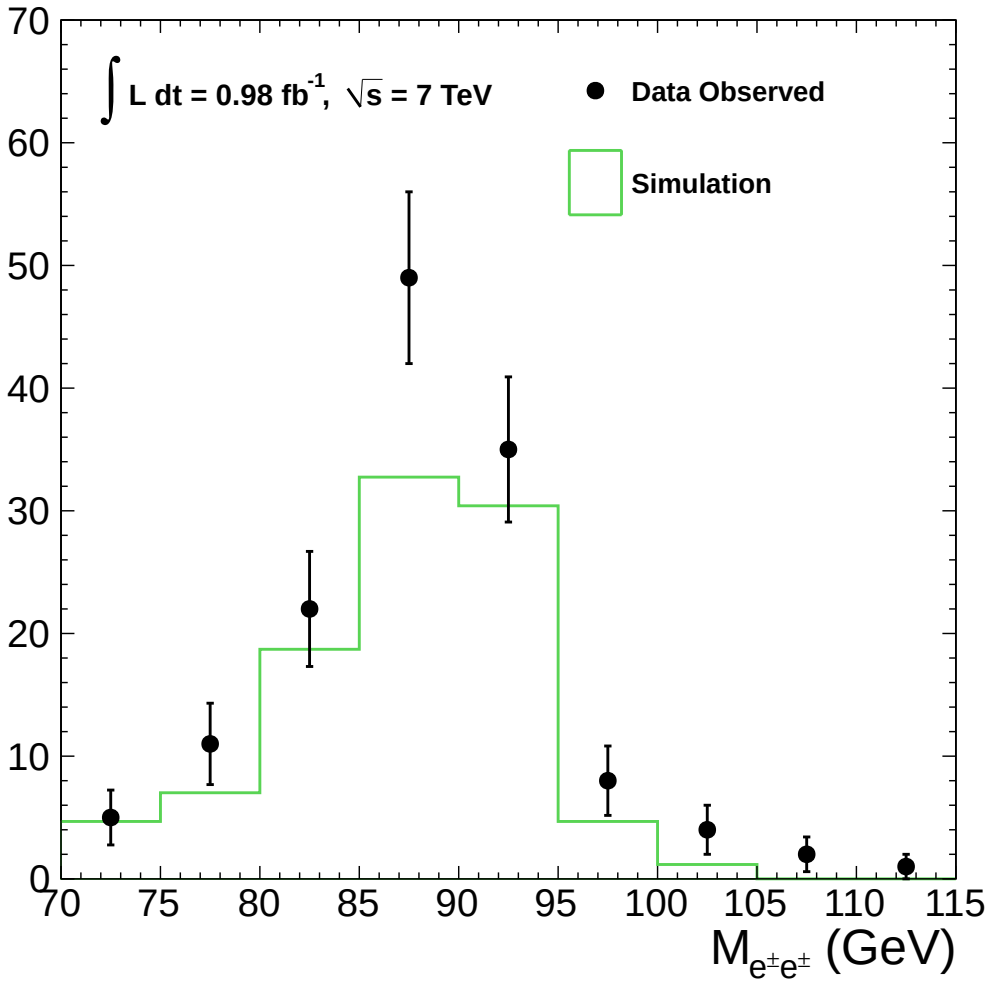

Figure 5.12: Invariant mass distribution of $e^{\pm}e^{\pm}$ pairs in data compared with $Z \to ee$ Monte Carlo expectations. To reduce backgrounds from W+jets, we require pf$\not{E}_T$ < 20 GeV and M_T < 25 GeV. The highest p_T electron is used in the calculation of the transverse mass.

involve rare SM processes that have never been measured before. The cross sections used for these processes are leading order. The resulting systematic uncertainty of 50% on the background estimate from simulation reflects the uncertainty in the cross sections and the kinematic distributions.

The same sign dilepton search described here will likely become limited by these processes as their contribution to the background becomes more significant as the H_T and $\not\!\!E_T$ requirements are increased. Continuing the same sign search may require suppressing these backgrounds, for example, by applying a b-tagging requirement. These backgrounds also offer the possibility of establishing complementary analyses to make first observations and to measure the cross sections of these processes.

Chapter 6

Selection Efficiencies

6.1 Introduction

Selection efficiencies and acceptance are needed to measure a cross section or to set an exclusion limit. The term acceptance refers to the presence of generated physics objects within the coverage of the detector. The combined selection efficiency represents the fraction of events within acceptance that also pass the analysis selections at the reconstructed level. Consider the lepton efficiency, for example. This efficiency represents the fraction of leptons within acceptance that pass the identification and isolation selections of the analysis. The total efficiency is a combination of efficiencies measured independently for each of the analysis selections.

Consider the analysis as follows. We apply selections to the data and count the number of observed events. Backgrounds are estimated using data-driven methods together with simulation. Based upon the number of observed events, the estimated backgrounds and the total uncertainties, a limit is set on the maximum number of events consistent with observation at the 95% confidence level.

To determine whether a specific signal model is consistent with the data, the same analysis selections are applied to a sample of generated events that have been run through the detector simulation and offline reconstruction. No trigger requirement is made in the simulation. Instead, we scale events in simulation by the measured efficiency of the relevant trigger paths. Similarly, differences exist in the efficiency of the lepton identification and isolation criteria between data and simulation. Scale factors are applied to the simulation to account for this difference. Only after applying scale

factors for all relevant efficiencies can the expected yield from simulation be compared with the observation from the data.

This chapter describes the determination of the selection efficiencies and acceptance. The first section describes the measurement of lepton efficiencies using a tag-and-probe method and the associated uncertainties. The next section discusses the efficiencies and associated uncertainties of the trigger requirements. Finally, to maximize the utility of our results, we present an efficiency framework that is as model independent as possible, to allow those so inclined to compare the predictions of their favorite model with our results.

6.2 Lepton Efficiencies

The lepton efficiencies are measured using a tag-and-probe based method, as descrbed in Section 3.3. We select $Z \to \ell^+\ell^-$ events from a lepton triggered sample by requiring opposite-sign, same flavor leptons in a window around the Z mass, $|M_Z - M_{ll}| < 15$ GeV. Out of window events are not used to extract efficiencies but are retained as sidebands to monitor the contamination from backgrounds. Contributions from W and QCD processes are found to be small. An event is tagged by requiring one lepton to pass the tight selections described in Section 4.3 and be matched to an online trigger object by requiring $\Delta R < 0.1$. Tagged events are then classified in the following mutually exclusive categories:

- TT: Both leptons are matched to trigger objects and pass the tight criteria.
- TP: The probe fails the tight criteria, but passes a looser set of criteria chosen to not bias the selection criteria under study.
- TF: The probe fails the looser selection criteria.

Those events classified as TT should enter the efficiency calculation twice, as either leg could be used as the probe. We extract the efficiency defined in (3.1) by plotting distributions of the probes for each category in bins of p_T and η. Efficiencies are extracted for lepton identification and isolation with respect to one another. As an example, consider the measurement of the isolation efficiency. Here, the tag criteria is the full analysis lepton selection plus trigger matching, while the probe criteria is the

analysis selection without isolation. The isolation efficiency is then extracted according to (3.1). The lepton identification efficiency was extracted in an analogous way.

It should be noted that the isolation efficiency for low-p_T leptons measured using tag-and-probe on $Z \to \ell^+\ell^-$ events is not reliable. Low p_T leptons are rare in this sample and the isolation selection at low-p_T is sensitive to the event topology as well as to details of the underlying event and pile-up.

Table 6.1: Electron isolation and identification efficiencies measured with the tag-and-probe method. The uncertainties are statistical only.

Type	source	Electron p_T range			
		10–15 GeV	15–20 GeV	20–40 GeV	$>$ 40 GeV
iso	mc	0.914 ± 0.013	0.930 ± 0.007	0.976 ± 0.001	0.995 ± 0.001
	data	0.870 ± 0.016	0.908 ± 0.008	0.972 ± 0.001	0.994 ± 0.001
	data/mc	0.952 ± 0.022	0.977 ± 0.011	0.997 ± 0.001	0.998 ± 0.001
id	mc	0.519 ± 0.018	0.645 ± 0.010	0.808 ± 0.002	0.861 ± 0.002
	data	0.429 ± 0.016	0.596 ± 0.010	0.789 ± 0.002	0.839 ± 0.002
	data/mc	0.827 ± 0.042	0.924 ± 0.022	0.976 ± 0.003	0.974 ± 0.002
id x iso	mc	0.474 ± 0.018	0.599 ± 0.011	0.788 ± 0.002	0.857 ± 0.002
	data	0.373 ± 0.016	0.541 ± 0.010	0.767 ± 0.002	0.834 ± 0.002
	data/mc	0.787 ± 0.044	0.903 ± 0.023	0.973 ± 0.003	0.972 ± 0.002

Tables 6.1 and 6.2 contain the identification and isolation efficiencies for electrons and muons, respectively. Efficiencies in simulation are extracted using Z samples as well as W and QCD events to simulate the dominant backgrounds. A total lepton efficiency is constructed by multiplying the individual efficiencies in each bin, as the efficiencies for identification and isolation are largely uncorrelated. A scale factor is derived by taking the ratio of efficiencies in data to simulation. Events yields from Monte Carlo are re–weighted by the appropriate scale factor. A resulting systematic uncertainty of half the difference between the scale factor and unity is taken on the data-to-simulation ratio.

6.3 Trigger Efficiencies

As described in Section 4.3.1, this analysis relies on two types of trigger paths: a) dilepton trigger paths without an additional jet requirement; b) dilepton trigger paths with an additional requirement on H_T. The trigger efficiencies are only relevant for the estimates of the signal selection efficiency.

Table 6.2: Muon isolation and identification efficiencies measured with the tag-and-probe method. The uncertainties are statistical only.

Type	source	Muon p_T range				
		5–10 GeV	10–15 GeV	15–20 GeV	20–40 GeV	$>$ 40 GeV
iso	mc	0.753 ± 0.086	0.858 ± 0.009	0.916 ± 0.005	0.966 ± 0.001	0.9945 ± 0.0003
	data	0.594 ± 0.087	0.772 ± 0.010	0.864 ± 0.006	0.963 ± 0.001	0.9927 ± 0.0003
	data/mc	0.788 ± 0.146	0.899 ± 0.015	0.943 ± 0.008	0.997 ± 0.001	0.9982 ± 0.0004
id	mc	0.974 ± 0.036	0.982 ± 0.004	0.973 ± 0.003	0.962 ± 0.001	0.958 ± 0.001
	data	0.950 ± 0.049	0.972 ± 0.005	0.974 ± 0.003	0.960 ± 0.001	0.955 ± 0.001
	data/mc	0.975 ± 0.062	0.990 ± 0.006	1.001 ± 0.004	0.998 ± 0.001	0.997 ± 0.001
id x iso	mc	0.733 ± 0.088	0.843 ± 0.009	0.891 ± 0.006	0.929 ± 0.001	0.953 ± 0.001
	data	0.564 ± 0.088	0.750 ± 0.011	0.842 ± 0.006	0.924 ± 0.001	0.948 ± 0.001
	data/mc	0.769 ± 0.151	0.890 ± 0.017	0.945 ± 0.010	0.995 ± 0.002	0.995 ± 0.001

The efficiencies of the high-p_T dilepton trigger paths was measured using tag-and-probe [33]. The dielectron trigger paths are measured to be approximately 99% efficient per electron over the entire η and p_T ranges relevant for this analysis. As the inefficiency increases slightly at lower p_T, we assign an efficiency of 99% $\pm$ 1% (99% $\pm$ 2%) for electrons with $p_T > 20$ GeV ($p_T < 20$ GeV). We consider efficiencies measured for the high-p_T dielectron triggers to be applicable to the dielectron+H_T triggers, where an online selection is made on track identification variables instead of isolation, and assign an efficiency of 98% $\pm$ 2% per electron for this trigger. This is confirmed directly in Z events, where we find 96.7% $\pm$ 0.7% of dielectrons passing the high-p_T dilectron trigger with an offline H_T requirement to also pass the dielectron+H_T trigger. As there is not sufficient statistics to confirm this efficiency for low-p_T electrons, we use a value with a larger uncertainty, 98% $\pm$ 3%, for electrons with $p_T < 20$ GeV.

The muon trigger path efficiencies measured in reference [33] are provided for muons with $p_T > 10$ GeV. The average efficiency for a trigger with a HLT requirement of $p_T > 7$ GeV and a level-1 requirement of a 3 GeV muon is approximately 96%. This measurement is expected to underestimate the efficiency of muons collected with an HLT requirement of 3 GeV and a level-1 requirement of a 0 GeV muon, as is the case for the dimuon+H_T trigger, by approximately 1–2%. Note, the electron–muon cross triggers without an H_T requirement use an even less restrictive level–1 muon seed, the so–called open-muon seed, which has a 1–2% higher efficiency. We simplify these results and assign an efficiency of 96% $\pm$ 2% (90% $\pm$ 5%) per muon for $p_T > 10$ GeV ($p_T < 10$ GeV), where for the low-p_T region we rely on results reported in reference [43].

Trigger paths with an H_T requirement have an additional inefficiency. The H_T trigger turn-on curves are shown in Figures 6.1 and 6.2, both made with the first 150 pb^{-1} of data. For each of these trigger turn-on curves, we choose a muon trigger for the denominator, and a corresponding muon plus H_T trigger for the numerator. Muons are chosen to show the worst case for the H_T-based triggers. The trigger requires the sum of the calorimeter-jets momenta with $p_T > 40$ GeV to be above a given threshold. While high-p_T electrons will naturally contribute to the H_T computed at the HLT-level, muons, being minimum-ionizing particles, will not. Figure 6.1 shows dimuon events with $p_T > 20$, 10 GeV with a dimuon mass of at least 40 GeV. The muons are required to pass the selections of Section 4.3.2. The denominator trigger is `DoubleMu7` while the numerator is the logical `OR` of the `DoubleMu3_HT150` and `DoubleMu3_HT160` triggers.

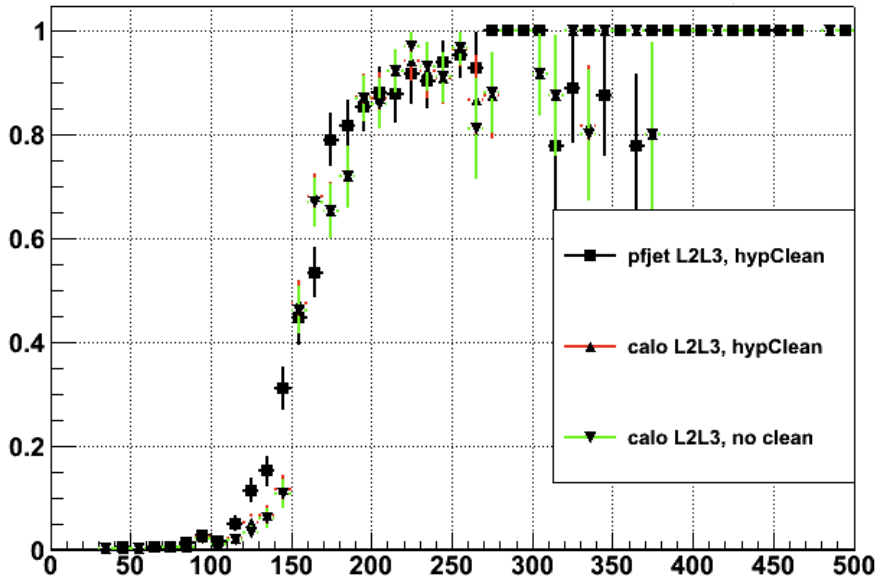

Figure 6.1: Turn-on curve for the H_T requirement of the DoubleMu3_HT150 and DoubleMu3_HT160 triggers measured using dimuon events near the Z mass. The distributions are shown as a function of H_T computed using L2L3-corrected particle-flow jets not overlapping with either of the muons (box), L2L3-corrected calorimeter jets not overlapping with either of the muons (up triangle) and L2L3-corrected calorimeter jets allowed to overlap with the two muons. In all cases, the jets are required to have $p_T > 30$ GeV and $|\eta| < 2.5$.

The precision of this measurement is limited by the number of dimuon events with large values of H_T. A higher precision measurement is performed using single–muon events, albeit for the only available single-muon trigger with a requirement of $H_T > 200$ GeV. Figure 6.2 compares `HT200` turn–on curves for single and double muon selections. For the single muon selection, we require $p_T > 30$ GeV, deliberately veto events with a second muon with $p_T > 15$ GeV, and require a logical `OR` of the `Mu8`, `Mu12` and `Mu30` triggers in the denominator. The numerator has the same selection except requiring as trigger `HLT_Mu8_HT200`. The double muon selection for Figure 6.2 is the same as in Figure 6.1. Only jets having $p_T > 40$ GeV, $|\eta| < 2.5$ and away from the muons by $\Delta R > 0.4$ are included in the H_T sum. The agreement between the curves increases confidence in the less precise measurement shown in Figure 6.1.

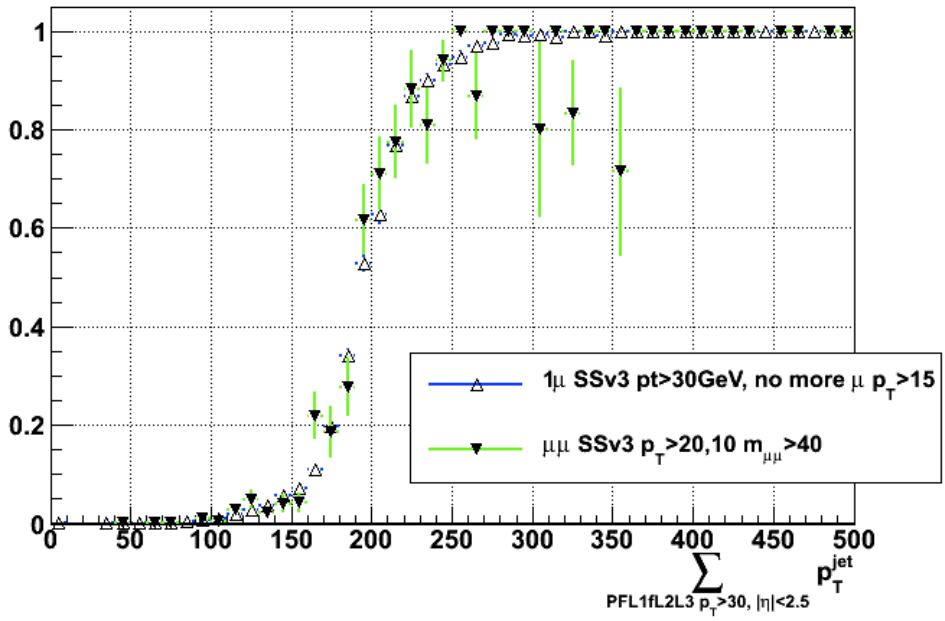

Figure 6.2: H_T trigger turn-on curves for the HT150/HT160 triggers as a function of the L1FastL2L3-corrected particle-flow jets. The jets are required to pass a selection of $p_T > 30$ GeV and $|\eta| < 2.5$. The curve measured for the single-muon+H_T trigger ine event with only one muon (open triangles) represent the efficiency in W+jets events. The curve measured for the dimuon H_T trigger in events with two muons (filled triangles) represents the efficiency in Z+jets events.

6.4 Efficiency Model

We would like to quote our results as a cross section, or cross section limit, in a manner that is as model-independent as possible, so that non-CMS theorists can interpret our results in the context of their favorite model. What we mean by this is that we carefully define the acceptance, and provide enough details about the selection efficiency within that acceptance, that anyone interested can use a Monte Carlo generator to define the acceptance at the hard scatter level and estimate the efficiency of the analysis selections in the context of this model to within the theoretical uncertainties with which the process is understood.

Here, we provide the scale factors needed to estimate the efficiency. The lepton selection efficiencies vary significantly as a function of η and p_T, especially for electrons. Second, both the $\not\!\!E_T$ and H_T have turn-on curves due to finite-resolution effects. Additional effects exist due to trigger efficiencies, pile-up distributions, and dependencies on p_T between data and simulation. Trigger efficiencies were discussed in Section 6.3 and lepton efficiency scale factors, as obtained from a tag-and-probe method applied to $Z \to \ell^+\ell^-$ events, can be found in Section 6.2. To simplify the efficiency model, these effects are included in the lepton efficiency parameterization described in Section 6.4.2. The efficiencies were determined using simulated events for the mSUGRA point LM6.

6.4.1 Acceptance

The acceptance is determined by counting the fraction of events that pass the analysis selection at generator level. The lepton acceptance is defined as having one lepton with $p_T > 20$ GeV and another lepton with $p_T > 10$ GeV for the high-p_T selection, or $p_T > 5$ GeV(10 GeV) for muons (electrons) in the inclusive dilepton selections. The leptons are required to have the same charge. The n-jet acceptance is determined using jets clustered at the generator level. The $H_T^{\rm gen}$ is calculated as the scalar sum p_T of all colored particles at the hard scattering level with $p_T > 40$ GeV and $|\eta| < 2.5$. The $\not\!\!E_T$ is defined as the absolute vector sum of the transverse momentum of all non-interacting particles — e.g. neutrinos and stable supersymmetric particles such as the LSP.

6.4.2 Lepton Efficiencies

These curves are taken directly from simulation; a similar set of curves was provided in the published 2010 analysis based on 36 $\text{pb}^{-1}$of 7 TeV data [32]. Lepton selection efficiencies, including data-to-simulation scale factors, are illustrated in Figure 6.3. The efficiency dependence can be parameterized as a function of p_T as

$$\epsilon = \epsilon_\infty \text{ erf } \left(\frac{p_T - C}{\sigma}\right) + \epsilon_C \left(1 - \text{erf } \left(\frac{p_T - C}{\sigma}\right)\right), \tag{6.1}$$

where ϵ_∞ gives the value of the efficiency plateau at high momentum, C is equal to 5 GeV (10 GeV) for muons (electrons), ϵ_C gives the value of the efficiency at $p_T = C$ and σ describes the size of the transition region. The results of the fits for electrons and muons are described in Table 6.3.

Table 6.3: Results of the fit of the p_T dependence in Figure 6.3 to the function specified in Eq. 6.1.

Parameter	Electrons	Muons
C	10	5
ϵ_∞	0.683 ± 0.010	0.736 ± 0.008
ϵ_C	0.186 ± 0.024	0.242 ± 0.029
σ	19.1 ± 1.8	14.7 ± 1.4

6.4.3 $\not\!\!E_T$ and H_T efficiency

We require at least two jets with $p_T > 40$ GeV and $|\eta| < 2.5$. Two such jets are present in approximately 95% of the events in LM1 and LM6 with $H_T^{\text{gen}} > 200$ GeV prior to any additional requirement on colored partons at the generator level. This represents the acceptance to two jets. In the following, we determine the H_T and $\not\!\!E_T$ efficiencies after generator-level requirements on the leptons and colored particles as described in Section 6.4.1.

The efficiency to pass a given reconstructed $\not\!\!E_T$ (H_T) threshold is shown in Figure 6.4 as a function of the $\not\!\!E_T^{\text{gen}}$ (H_T^{gen}) in events passing $H_T^{\text{gen}} > 200$ GeV ($\not\!\!E_T^{\text{gen}} >$ 30 GeV). Due to the small fraction of events in the LM6 simulation having low H_T, the H_T efficiency curves are made using a simulation of the mSUGRA point LM1. Results of the fits of these curves to the function

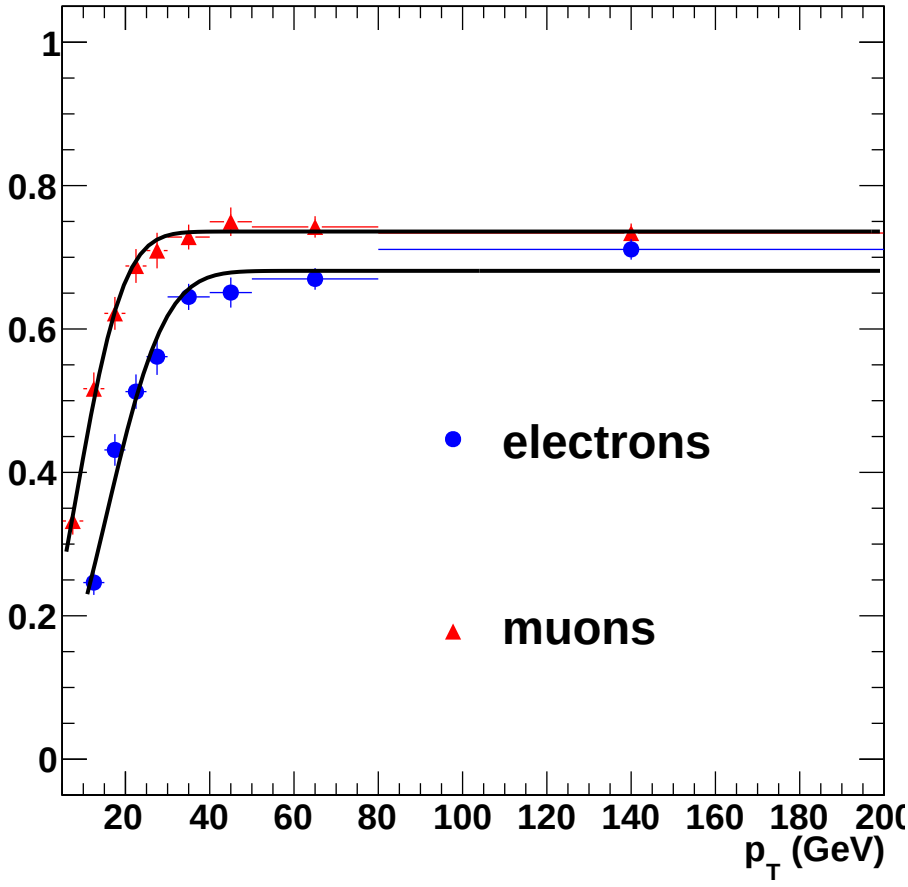

Figure 6.3: Lepton selection efficiency for electrons and muons as a function of p_T. The efficiencies were determined at the hard scatter level using simulated events for the mSUGRA point LM6. The displayed efficiencies include data-to-simulation scale factors for lepton identification and isolation determined using tag-and-probe in Z events as well as trigger efficiencies. The simulation has been re-weighted to reflect the pile-up distribution observed in the data.

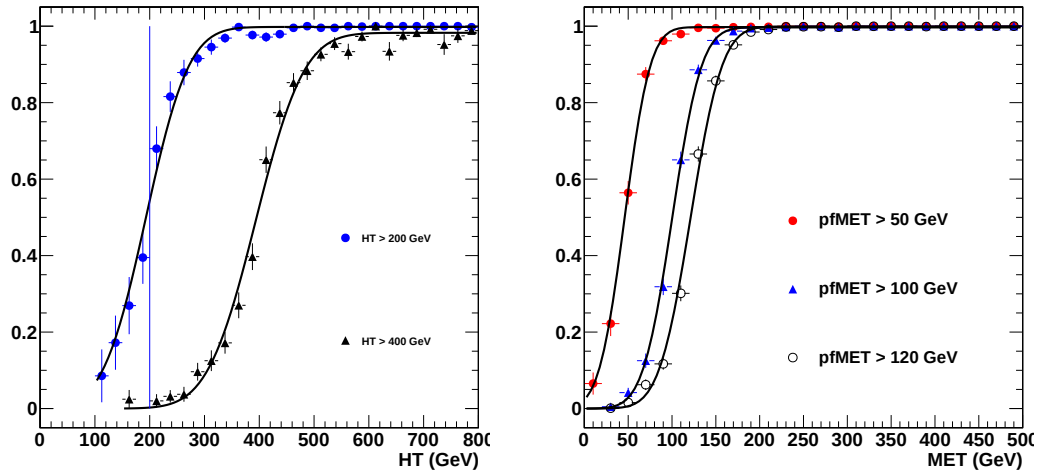

Figure 6.4: Efficiency for an event to pass a given reconstructed $\not\!\!E_T$ (H_T) threshold as a function of the $\not\!\!E_T^{\rm gen}$ ($H_T^{\rm gen}$). The curves are shown for $\not\!\!E_T$ thresholds of 50, 100 and 120 GeV. The thresholds for H_T are 200 and 400 GeV.

$$\epsilon = 0.5\ \epsilon_\infty \left[\mathrm{erf}\left(\frac{x - x_{1/2}}{\sigma} \right) + 1 \right] \tag{6.2}$$

are summarized in Table 6.4. Neither the $\not\!\!E_T$ nor the H_T curves show a significant bias in the position of the point with half the plateau efficiency, $x_{1/2}$. The inefficiency at the plateau, ϵ_∞, is negligible. The width of the turn-on region, σ, increases with the value of the cut.

Table 6.4: Results of the fit of the dependence in Figure 6.4 to the function specified in Eq. 6.2

	H_T		$\not\!\!E_T$		
	> 200 GeV	> 400 GeV	> 50 GeV	> 100 GeV	> 120 GeV
ϵ_∞	0.998 ± 0.001	0.987 ± 0.002	0.998 ± 0.001	0.997 ± 0.001	0.999 ± 0.001
$x_{1/2}$	193.0 ± 4.5	378.6 ± 3.1	45.9 ± 1.2	100.2 ± 0.8	121.2 ± 0.8
σ	87.4 ± 5.9	113.2 ± 4.9	32.6 ± 1.9	37.3 ± 1.3	40.2 ± 1.3

Chapter 7

Results

As described in Chapter 1, we have conducted a search for new physics in a final state with two same-sign leptons, missing transverse energy, and significant hadronic activity. Final state objects were selected according to Chapter 4, providing regions sensitive to new physics, while rejecting a large fraction of the SM background. Observed events yields are compared with background estimates in the control regions. Good agreement is observed, which provides confidence in the data-driven background estimation methods described in Chapter 5.

The event yields in the signal regions defined in Section 4.4 are reported together with the estimated background yields. No new physics is observed. Model independent limits are set. The observed limits are compared to the expected yields for benchmark points in mSUGRA parameter space.

The selections chosen for this analysis were motivated by considerations of Standard Model backgrounds and generic new physics signatures. Although opposite-sign dileptons are significantly more prevalent in the SM than same-sign, any NP model with a Majorana particle will produce the two in equal abundance. Astrophysical evidence for dark matter suggests the existence of a weakly interacting, massive particle; a final state with significant missing transverse energy is motivated by this evidence. Finally, particles produced via strong interactions tend to have larger cross sections than those produced via electroweak interactions; final states with significant hadronic activity are thus likely to require a smaller integrated luminosity for observation. These motivations are realized in a variety of models such as SUSY, models with extra dimensions, Majorana neutrinos, and same-sign top resonances.

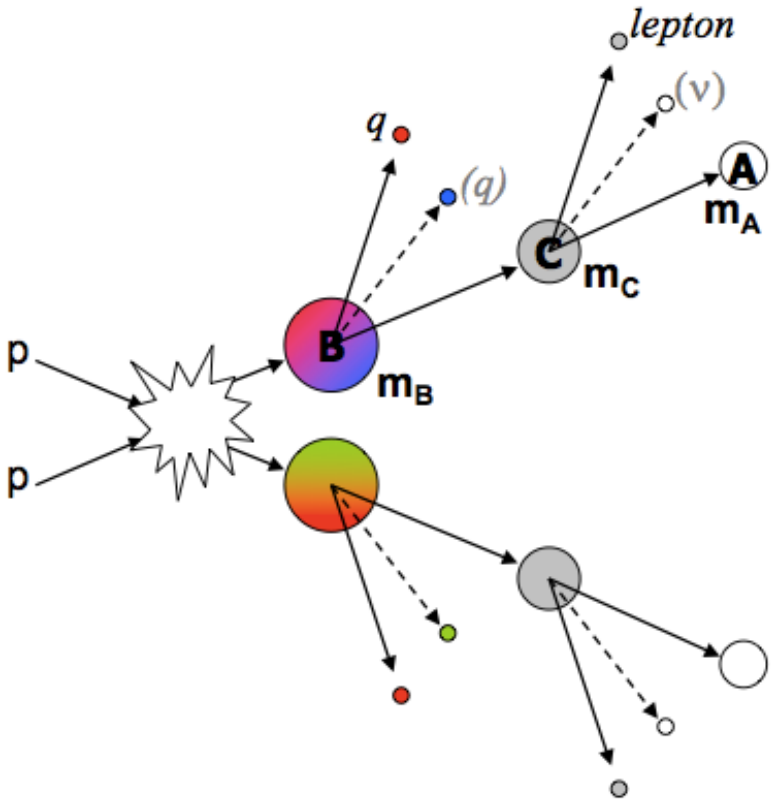

Figure 7.1: A depiction of a decay chain typically observed in SUSY models.

This chapter consists of three sections. The first section discuss the results of the high-p_T dilepton searches starting with a discussion of control region events having leptons with $p_T > 20$, 10 GeV, $\not\!\!E_T > 30$ GeV, and $H_T > 80$ GeV. Results for each signal region follow with particular attention paid to the variation in the background composition with changed H_T and $\not\!\!E_T$ requirements. The following section provides an analogous discussion of the inclusive dilepton search results, beginning with a control region that increases the H_T requirement to 200 GeV and lowers the lepton p_T requirements to 5 (10) GeV for muons (electrons). The final section discusses model independent limits. These limits are used to rule out several mSUGRA points and to exclude regions of the CMSSM model in the m_0–$m_{1/2}$ plane [44]. For clarity, the signal region definitions are repeated here.

- Region 1: $H_T > 400$ GeV, $\not\!\!E_T > 120$ GeV, inspired by mSUGRA models with low m_0;
- Region 2: $H_T > 400$ GeV, $\not\!\!E_T > 50$ GeV, inspired by mSUGRA models with high m_0;
- Region 3: $H_T > 200$ GeV, $\not\!\!E_T > 120$ GeV, inspired by a simplified model of squark–gluino production;
- Region 4: $H_T > 80$ GeV, $\not\!\!E_T > 100$ GeV, inspired by the pMSSM model with sneutrino LSP (meaningful only for the high-p_T dilepton search).

The selections and signal regions summarized allow for a broad area of the mSUGRA parameter space to be probed. Figure 7.1 shows a depiction of a cascade decay typically observed in the SUSY models considered. As a toy model, consider a scenario where m_B represents a gluino ($\tilde{\mathrm{g}}$), m_C a chargino ($\chi^{\pm}$), and m_A the lightest supersymmetric particle (LSP). If the mass of the gluino is larger than the squark mass, $m_{\tilde{g}} > m_{\tilde{q}}$, then gluino production leads to $\tilde{g} \to \tilde{q}\bar{q}$ followed by $\tilde{q} \to q\chi^{\pm}$. Thus, the p_T of the jets is determined by ΔM_{BC} and this mass difference drives the H_T. The cascade continues with $\chi^{\pm} \to \ell^{\pm}\nu\chi^0$, where the χ^0 is the LSP. The p_T of the final state leptons are thus driven by the mass difference between the chargino and the LSP, ΔM_{BA}. Thus, performing searches in regions of phase space with different H_T, $\not\!\!E_T$, and lepton p_T requirements provides added sensitivity as the masses of the supersymmetric particles vary in mSUGRA parameter space.

7.1 The High p_T Dilepton Analysis

7.1.1 High p_T Control Region

We define a control region enriched in background by requiring a same sign dilepton pair with $p_T > 20$, 10 GeV and at least 30 GeV of $\not\!\!E_T$ with no additional requirement on the hadronic activity beyond requiring at least 2 jets. This region is sensitive mostly to fake leptons in SM multi-jet processes: predominantly $t\bar{t}$ and W+jets. Figure 7.2 shows the distribution of events in the H_T–$\not\!\!E_T$ plane. The events cluster in the corner of low $\not\!\!E_T$ and low H_T, as expected for background. Horizontal and vertical lines denote the signal region definitions in this plane. The shaded area indicates the regions not accessible to this analysis.

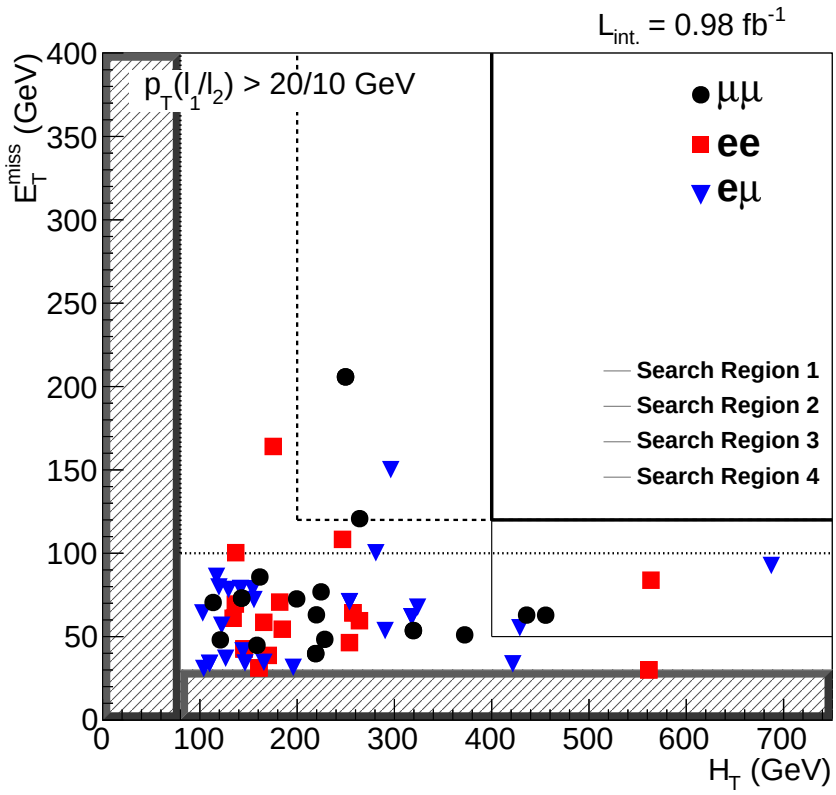

Figure 7.2: Distribution of events in the H_T–$\not\!\!E_T$ plane passing the analysis selections for the high-p_T control region. From [45]

The observed event yields and estimated backgrounds in the high-p_T control region can be seen in Table 7.1. The yield reported as *simulated backgrounds* includes

contributions from genuine same-sign lepton pairs, as well as electrons from converted photons in $V\gamma$ production. Entries with zero contributing events are reported with an uncertainty corresponding to one event. This uncertainty is not added to the total MC contribution. Systematic uncertainties (the second uncertainty if present) are displayed only for the final combined type of background, no systematic uncertainty is added for estimates with zero entries. Systematic uncertainties are 100% correlated among the channels.

Simulation indicates an expected background composition of approximately 65% fake leptons, 5% leptons with incorrect charge assignment and 30% from SM sources of prompt, isolated same sign dileptons. The background from fake leptons is approximately 75% top, 20% W+jets and 5% $V\gamma$. The latter contribution is not covered by the fake rate method and instead is taken directly from simulation. The W+jets contribution is based on 2 unweighted events. The background from electrons with mis-reconstructed charge comes from fully leptonic $t\bar{t}$ decays. Simulation predicts no contribution from Drell-Yan processes. The remaining irreducible background comes from several different processes, although predominantly $t\bar{t}W$, WZ and $qqW^{\pm}W^{\pm}$. The backgrounds from rare processes such as $qqW^{\pm}W^{\pm}$ and $t\bar{t}W$ are particularly interesting. The latter process has an expected yield of nearly 4.5 events in the control region in 1 fb^{-1} based on the LO cross section. This and other rare processes may become observable with the next year's data.

We observe 56 events in the control region with a total estimate of all backgrounds of approximately 75 events. The two agree within uncertainties. The total background estimate of 75 events is comprised of an estimated 62 events from fake leptons, 2.5 events from leptons with mis–reconstructed charge and a bit over 11 events from SM sources of prompt isolated same sign dileptons. The estimated background from fake leptons in data is nearly a factor of 3 larger than that expected from pure Monte Carlo. Even accounting for a 6% underestimate of the luminosity and as much as a 70% overestimate by the fake rate method, the simulation under predicts the background from fake leptons by about 50%. If we were to assign a scale factor of 0.6 to the background estimated using the fake rate method to account for the over prediction observed in the closure test, the background contribution from single and double fakes would fall from 61.9 to 36.4 while the total background from all sources would become about 50 events, in agreement with the 56 observed. No scale factor is applied to the fake lepton background estimate

at this time.

7.1.2 Signal Regions

We consider four signal regions as shown in Figure 7.2. The observed yields and estimated backgrounds for each signal region are shown in Table 7.2. The uncertainty reported is the total statistical and systematic uncertainty. We see no evidence of an event yield in excess of the background estimations and set 95% confidence level (CL) upper limits (UL) on the number of observed events using a hybrid frequentist-bayesian method with a flat prior on the signal strength. Log-normal distributions are used for the efficiency and background uncertainties. The limits include uncertainties on the signal efficiency of 14%. A detailed breakdown of results can be found in Appendix B.

Table 7.3 shows a breakdown of the total backgrounds in each of the high-p_T signal regions. Backgrounds taken from data are classified as coming from fake leptons or charge mis-reconstruction. The backgrounds classified as coming from simulation include SM sources of prompt, isolated same sign dileptons, as well as fakes from photon conversion in $V\gamma$ processes. Fake leptons are the dominant background, accounting for approximately 55–70% of the background depending on the signal region. Irreducible backgrounds, primarily from $t\bar{t}W$ and $qqW^{\pm}W^{\pm}$ account for 30–35% of the background, except for the tightest signal region where nearly 45% of the background comes from simulation. The background from charge mis-reconstruction is small, contributing 1–3%.

The background composition observed is different than that predicted from pure simulation, where fakes make a smaller contribution to the background, which is dominated by rare SM processes (except for Region 4, where fakes are expected to be the dominant background). As discussed in Section 7.1.1, some of this may be due to an overestimate of the background from fake leptons, although there is evidence from the control region that the simulation under represents the size of this background. Table 7.4 shows a breakdown of the backgrounds estimated from simulation for each high-p_T signal region. The uncertainties reported are purely statistical. The contribution from $V\gamma$ is omitted as there are no events passing the analysis selections in any signal region. The largest contributions come from $t\bar{t}W$ followed by $qqW^{\pm}W^{\pm}$, with the contribution from WZ and ZZ increasing as the H_T requirement is relaxed. It is interesting to observe how the $t\bar{t}W$ and $qqW^{\pm}W^{\pm}$backgrounds change with the $\not{E}_T$ and H_T requirements. Combined with b-tagging and reductions in the fake rate method systematics,

Table 7.1: Observed yields in the high-p_T control region together with background estimates and predictions from simulation.

Source	ee	$\mu\mu$	eμ	all
$t\bar{t} \to \ell\ell X$	0.996 ± 0.448	0.000 ± 0.144	0.746 ± 0.417	1.742 ± 0.612
$t\bar{t} \to \ell(b \to \ell)X$	3.545 ± 0.850	3.514 ± 0.830	5.634 ± 1.067	12.692 ± 1.597
$t\bar{t} \to \ell(\not{b} \to \ell)X$	1.832 ± 0.571	0.000 ± 0.144	0.821 ± 0.407	2.653 ± 0.701
$t\bar{t}$ other	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
tW	0.234 ± 0.084	0.121 ± 0.053	0.348 ± 0.098	0.703 ± 0.139
t, t-channel	0.173 ± 0.088	0.117 ± 0.076	0.458 ± 0.171	0.749 ± 0.206
t, s-channel	0.023 ± 0.010	0.007 ± 0.005	0.020 ± 0.009	0.049 ± 0.014
W+jets	0.000 ± 2.039	0.000 ± 2.039	4.117 ± 2.636	4.117 ± 2.636
$Z \to ee$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \mu\mu$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \tau\tau$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
WW	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040
WZ	0.536 ± 0.076	0.689 ± 0.091	1.132 ± 0.112	2.357 ± 0.164
ZZ	0.055 ± 0.015	0.051 ± 0.014	0.112 ± 0.020	0.218 ± 0.029
Vγ	1.066 ± 0.542	0.000 ± 0.210	0.428 ± 0.335	1.494 ± 0.637
qqW^+W^+	0.409 ± 0.022	0.598 ± 0.027	0.997 ± 0.035	2.005 ± 0.049
qqW^-W^-	0.131 ± 0.006	0.202 ± 0.008	0.330 ± 0.010	0.662 ± 0.014
dp$W^\pm W^\pm$	0.004 ± 0.003	0.011 ± 0.005	0.013 ± 0.005	0.028 ± 0.007
$t\bar{t}W$	0.839 ± 0.018	1.301 ± 0.022	2.178 ± 0.029	4.317 ± 0.041
Total MC	9.841 ± 1.251	6.611 ± 0.841	17.333 ± 2.932	33.786 ± 3.296
LM6	1.018 ± 0.043	1.334 ± 0.050	2.223 ± 0.064	4.576 ± 0.092
Prompt-fake	13.67 ± 2.32	14.18 ± 2.21	30.08 ± 3.99	57.93 ± 5.12
Double-fake	0.62 ± 0.21	1.11 ± 0.27	2.24 ± 0.50	3.97 ± 0.60
Total with fakes	14.29 ± 2.29 ± 7.15	15.29 ± 2.16 ± 7.65	32.32 ± 3.90 ± 16.16	61.90 ± 5.01 ± 30.95
Charge misreconstruction	1.77 ± 0.08 ± 0.35	- ± -	0.71 ± 0.04 ± 0.14	2.48 ± 0.09 ± 0.50
Simulated backgrounds	3.04 ± 0.55 ± 1.52	2.85 ± 0.10 ± 1.43	5.19 ± 0.36 ± 2.59	11.08 ± 0.66 ± 5.54
All backgrounds	19.10 ± 2.36 ± 7.31	18.15 ± 2.16 ± 7.78	38.21 ± 3.92 ± 16.37	75.46 ± 5.06 ± 31.45
Data	16	16	24	56

Table 7.2: Summary of results for the high-p_T signal regions. The observed yields and estimated backgrounds agree within uncertainties. A 95% CL upper limit is set on the yield.

Signal Region (minimum H_T/ $\not\!\!E_T$)	ee	$\mu\mu$	$e\mu$	all	95% CL UL yield
(400 / 120)					
Estimated background	0.4 ± 0.3	0.4 ± 0.3	0.7 ± 0.4	1.4 ± 0.7	3.0
Observed	0	0	0	0	
(400 / 50)					
Estimated background	1.4 ± 0.8	1.3 ± 0.8	1.3 ± 0.6	4.0 ± 1.7	7.5
Observed	1	2	2	5	
(200 / 120)					
Estimated background	1.2 ± 0.7	1.5 ± 0.8	1.8 ± 0.8	4.5 ± 1.9	5.2
Observed	0	2	1	3	
(80 / 100)					
Estimated background	2.5 ± 1.2	2.6 ± 1.2	4.9 ± 2.2	10 ± 4	6.0
Observed	3	2	2	7	

Table 7.3: Breakdown of the background estimates in the high-p_T signal regions.

Signal Region (minimum H_T/ $\not\!\!E_T$)	Fake Leptons	Incorrect Charge	Simulation	Total
(400 / 120) Background Estimates	0.8 ± 0.6	0.03 ± 0.01	0.6 ± 0.3	1.4 ± 0.7
(400 / 50) Background Estimates	2.6 ± 1.6	0.09 ± 0.02	1.3 ± 0.7	4.0 ± 1.7
(200 / 120) Background Estimates	2.9 ± 1.7	0.08 ± 0.02	1.6 ± 0.8	4.5 ± 1.9
(80 / 100) Background Estimates	6.8 ± 3.8	0.28 ± 0.06	2.9 ± 1.4	10 ± 4

these trends suggest potential search strategies for making first observations of these rare SM processes.

Table 7.4: Breakdown of the backgrounds estimated using simulation.

Region $H_T/\not\!\!E_T$	WZ/ZZ	$t\bar{t}W$	DPS $W^{\pm}W^{\pm}$	$qqW^{\pm}W^{\pm}$	Total
400/120	0.06 ± 0.03	0.26 ± 0.01	0.000 ± 0.002	0.31 ± 0.02	0.63 ± 0.04
400/50	0.10 ± 0.03	0.59 ± 0.02	0.002 ± 0.002	0.63 ± 0.03	1.32 ± 0.05
200/120	0.20 ± 0.05	0.79 ± 0.02	0.000 ± 0.002	0.57 ± 0.02	1.56 ± 0.06
80/100	0.43 ± 0.07	1.51 ± 0.02	0.004 ± 0.003	0.93 ± 0.03	2.87 ± 0.08

7.2 The Inclusive Dilepton Analysis

7.2.1 Control Region

We define a control region enriched in background by requiring a same sign dilepton pair. We select muons with $p_T > 5$ GeV and electrons with $p_T > 10$ GeV. We require at least 30 GeV of $\not\!\!E_T$ and at least 200 GeV of H_T in addition to requiring at least 2 jets. Figure 7.3 shows the distribution of events in the H_T–$\not\!\!E_T$ plane. The events cluster in the corner of low $\not\!\!E_T$ and low H_T, as expected for background. Horizontal and vertical lines denote the signal region definitions in this plane. The shaded area indicates the regions not accessible to this analysis.

The observed event yields and estimated backgrounds in the inclusive dilepton control region can be seen in Table 7.5. The yield reported as *simulated backgrounds* includes contributions from genuine same-sign lepton pairs, as well as electrons from converted photons in $V\gamma$ production. Entries with zero contributing events are reported with an uncertainty corresponding to one event. This uncertainty is not added to the total MC contribution. Systematic uncertainties (the second uncertainty if present) are displayed only for the final combined type of background, no systematic uncertainty is added for estimates with zero entries. Systematic uncertainties are 100% correlated among the channels.

Simulation indicates an expected background composition of approximately 80% fake leptons, 3% leptons with incorrect charge assignment and 17% from SM sources of prompt, isolated same sign dileptons. The background from fake leptons is approximately 87% top, 12% W+jets and 1 % $V\gamma$. The latter contribution is not covered by the fake rate

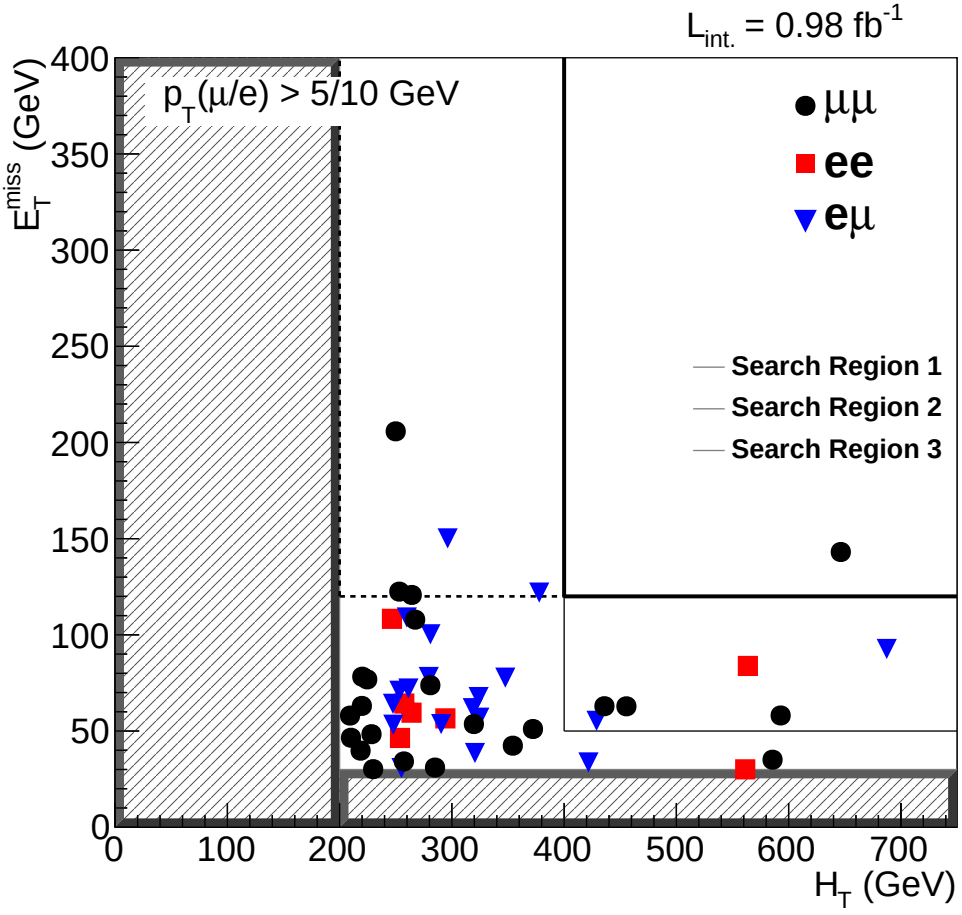

Figure 7.3: Distribution of events in the H_T–$\not\!\!E_T$ plane passing the analysis selections for the inclusive dilepton control region. From [45]

method and instead is taken directly from simulation. The W+jets contribution is based on 2 unweighted events. The background from electrons with mis-reconstructed charge comes from fully leptonic $t\bar{t}$ decays. Simulation predicts no contribution from Drell-Yan. The remaining irreducible background comes from several different processes, although predominantly $t\bar{t}W$, $qqW^{\pm}W^{\pm}$ and WZ. The backgrounds from rare processes such as $qqW^{\pm}W^{\pm}$ and $t\bar{t}W$ are particularly interesting. The latter process has an expected yield of nearly 3 events in the control region in 1 fb^{-1} based on the LO cross section. This and other rare processes may become observable with the next year's data.

We observe 49 events in the control region with a total estimate of all backgrounds of approximately 58 events. The two agree within uncertainties. The total background estimate of 58 events is comprised of an estimated 51 events from fake leptons, 1 event from leptons with mis-reconstructed charge and a bit over 6 events from SM sources of prompt isolated same sign dileptons. The estimated background from fake leptons in data is about a factor of 2 larger than that expected from pure Monte Carlo. Accounting for a 6% underestimate of the luminosity and as much as a 70% overestimate by the fake rate method would nearly bring the two into agreement. No scale factor is applied to the fake lepton background estimate at this time. The asymmetric p_T cuts on the different lepton generations results in a significant difference in observed yields and estimated backgrounds in the ee and $\mu\mu$ final states.

Table 7.5: Observed yields in the inclusive dilepton control region together with background estimates and predictions from simulation.

Source	ee	$\mu\mu$	eμ	all
$t\bar{t} \to \ell\ell X$	0.595 ± 0.346	0.000 ± 0.144	0.231 ± 0.203	0.826 ± 0.401
$t\bar{t} \to \ell(b \to \ell)X$	2.223 ± 0.679	8.062 ± 1.289	7.331 ± 1.197	17.616 ± 1.885
$t\bar{t} \to \ell(\not{b} \to \ell)X$	1.405 ± 0.526	0.216 ± 0.216	0.763 ± 0.372	2.384 ± 0.679
$t\bar{t}$ other	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
tW	0.038 ± 0.038	0.387 ± 0.100	0.353 ± 0.103	0.778 ± 0.149
t, t-channel	0.078 ± 0.052	0.657 ± 0.183	0.789 ± 0.218	1.524 ± 0.290
t, s-channel	0.005 ± 0.005	0.014 ± 0.008	0.012 ± 0.007	0.031 ± 0.011
W+jets	0.000 ± 2.039	0.000 ± 2.039	3.027 ± 2.401	3.027 ± 2.401
$Z \to ee$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \mu\mu$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \tau\tau$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
WW	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040
WZ	0.131 ± 0.037	0.231 ± 0.053	0.409 ± 0.068	0.771 ± 0.093
ZZ	0.018 ± 0.009	0.021 ± 0.008	0.038 ± 0.012	0.076 ± 0.017
Vγ	0.372 ± 0.372	0.000 ± 0.210	0.000 ± 0.210	0.372 ± 0.372
qqW^+W^+	0.299 ± 0.019	0.458 ± 0.023	0.761 ± 0.030	1.517 ± 0.043
qqW^-W^-	0.086 ± 0.005	0.146 ± 0.007	0.223 ± 0.008	0.455 ± 0.012
dp$W^\pm W^\pm$	0.000 ± 0.002	0.002 ± 0.002	0.004 ± 0.003	0.006 ± 0.003
$t\bar{t}W$	0.544 ± 0.014	0.898 ± 0.019	1.466 ± 0.024	2.908 ± 0.033
Total MC	5.792 ± 1.001	11.091 ± 1.325	15.408 ± 2.728	32.291 ± 3.193
LM6	1.008 ± 0.043	1.509 ± 0.053	2.332 ± 0.065	4.849 ± 0.095
Prompt-fake	8.68 ± 2.22	11.84 ± 3.91	19.83 ± 3.48	40.35 ± 5.68
Double-fake	0.25 ± 0.23	7.31 ± 1.10	2.63 ± 0.63	10.19 ± 1.29
Total with fakes	8.93 ± 2.18 ± 4.46	19.15 ± 3.41 ± 9.57	22.46 ± 3.30 ± 11.23	50.54 ± 5.22 ± 25.27
Charge misreconstruction	0.657 ± 0.049 ± 0.131	- ± -	0.273 ± 0.025 ± 0.055	0.931 ± 0.055 ± 0.186
Simulated backgrounds	1.449 ± 0.375 ± 0.724	1.755 ± 0.062 ± 0.878	2.901 ± 0.079 ± 1.450	6.105 ± 0.388 ± 3.052
All backgrounds	11.04 ± 2.22 ± 4.52	20.90 ± 3.41 ± 9.61	25.63 ± 3.30 ± 11.32	57.57 ± 5.24 ± 25.45
Data	7	23	19	49

7.2.2 Signal Regions

We consider three signal regions as shown in Figure 7.3. The observed yields and estimated backgrounds for each signal region are shown in Table 7.6. The uncertainty reported is the total statistical and systematic uncertainty. We see no evidence of an event yield in excess of the background estimations and set 95% confidence level (CL) upper limits (UL) on the number of observed events using a hybrid frequentist-bayesian method with a flat prior on the signal strength. Log-normal distributions are used for the efficiency and background uncertainties. The limits include uncertainties on the signal efficiency of 17%. A detailed breakdown of results can be found in Appendix B.

Table 7.6: Summary of results for the inclusive dilepton signal regions. The observed yields and estimated backgrounds agree within uncertainties. A 95% CL upper limit is set on the yield.

Signal Region (minimum H_T/ $\not{E}_T$)	ee	$\mu\mu$	$e\mu$	all	95% CL UL yield
(400 / 120)					
Estimated background	0.4 ± 0.4	1.2 ± 0.8	0.7 ± 0.4	2.3 ± 1.2	3.7
Observed	0	1	0	1	
(400 / 50)					
Estimated background	1.3 ± 0.7	2.5 ± 1.5	1.4 ± 0.7	5.3 ± 2.4	8.9
Observed	1	2	2	5	
(200 / 120)					
Estimated background	1.5 ± 0.9	3.0 ± 1.6	2.1 ± 1.0	6.6 ± 2.9	7.3
Observed	0	2	1	3	

Table 7.7 shows a breakdown of the total backgrounds in each of the inclusive dilepton signal regions. Backgrounds taken from data are classified as coming from fake leptons or charge mis-reconstruction. The backgrounds classified as coming from simulation include SM sources of prompt, isolated same sign dileptons, as well as fakes from photon conversion in $V\gamma$ processes. Fake leptons are the dominant background, accounting for approximately 70–75% of the background depending on the signal region. Irreducible backgrounds, primarily from $t\bar{t}W$ and $qqW^{\pm}W^{\pm}$ account for 25–30% of the background. The remaining background from charge mis–reconstruction is small, contributing 1–2%.

The background composition observed is similar to that predicted from pure simulation, except for Region 1 where fakes are predicted to be subdominant to the

Table 7.7: Breakdown of the background estimates in the inclusive dilepton signal regions.

Signal Region (minimum H_T/ $\not\!\!E_T$)	Fake Leptons	Incorrect Charge	Simulation	Total
(400 / 120) Background Estimates	1.6 ± 1.1	0.03 ± 0.01	0.7 ± 0.3	2.3 ± 1.2
(400 / 50) Background Estimates	3.9 ± 2.3	0.09 ± 0.02	1.4 ± 0.7	5.3 ± 2.4
(200 / 120) Background Estimates	4.8 ± 2.8	0.08 ± 0.02	1.7 ± 0.8	6.6 ± 2.9

contributions from rare SM processes. Table 7.8 shows a breakdown of the backgrounds estimated from simulation for each high-p_T signal region. The uncertainties reported are purely statistical. The contribution from $V\gamma$ is omitted as there are no events passing the analysis selections in any signal region. The largest contributions come from $t\bar{t}W$ followed by $qqW^{\pm}W^{\pm}$, with the contribution from WZ and ZZ increasing as the H_T requirement is relaxed.

Table 7.8: Breakdown of the backgrounds estimated using simulation.

Region $H_T/\not\!\!E_T$	WZ/ZZ	$t\bar{t}W$	DPS $W^{\pm}W^{\pm}$	qq$W^{\pm}W^{\pm}$	Total
400/120	0.06 ± 0.03	0.28 ± 0.01	0.000 ± 0.002	0.32 ± 0.02	0.64 ± 0.04
400/50	0.10 ± 0.03	0.62 ± 0.02	0.002 ± 0.002	0.65 ± 0.03	1.37 ± 0.05
200/120	0.20 ± 0.05	0.86 ± 0.02	0.002 ± 0.002	0.62 ± 0.02	1.64 ± 0.06

7.3 Limits on New Physics

The model-independent limits in Table 7.2 are compared with the expected yields from simulation for a number of mSUGRA points [46]. These points are standard benchmarks used by the CMS experiment. The parameters defining the benchmarks can be found in Appendix C. We consider here only the high-p_T signal region with $H_T > 400$ GeV and $\not\!\!E_T > 120$ GeV, for which the best limit is set. Based on the 95% CL upper limit yield of 3.0 events, we exclude all points except for LM2, LM5, LM7 and LM12. The point LM0 was excluded by the 2010 analysis [32]. We may be able to exclude the remaining points once the full 2011 dataset is collected.

Table 7.9: Expected yields from simulation for a variety of mSUGRA points for the high-p_T analysis with $H_T > 400$ GeV and $\not{E}_T > 120$ GeV.

Expected	LM1	LM2	LM3	LM4	LM5	LM6
	19.3 ± 0.8	1.9 ± 0.1	13.8 ± 0.6	5.5 ± 0.3	1.5 ± 0.1	4.2 ± 0.1
Yield	LM7	LM8	LM9	LM11	LM12	LM13
	1.2 ± 0.1	7.4 ± 0.2	4.0 ± 0.5	5.7 ± 0.2	1.6 ± 0.2	17.6 ± 0.9

The cross sections used for the mSUGRA points are NLO. Events from simulation have been re-weighted to account for trigger efficiencies as well as data-to-simulation scale factors based on differences in the lepton selection efficiencies as measured using a tag-and-probe method in Section 6.2. Events are additionally scaled according to Figure 4.1 to account for differences in pile-up distributions between data and Monte Carlo.

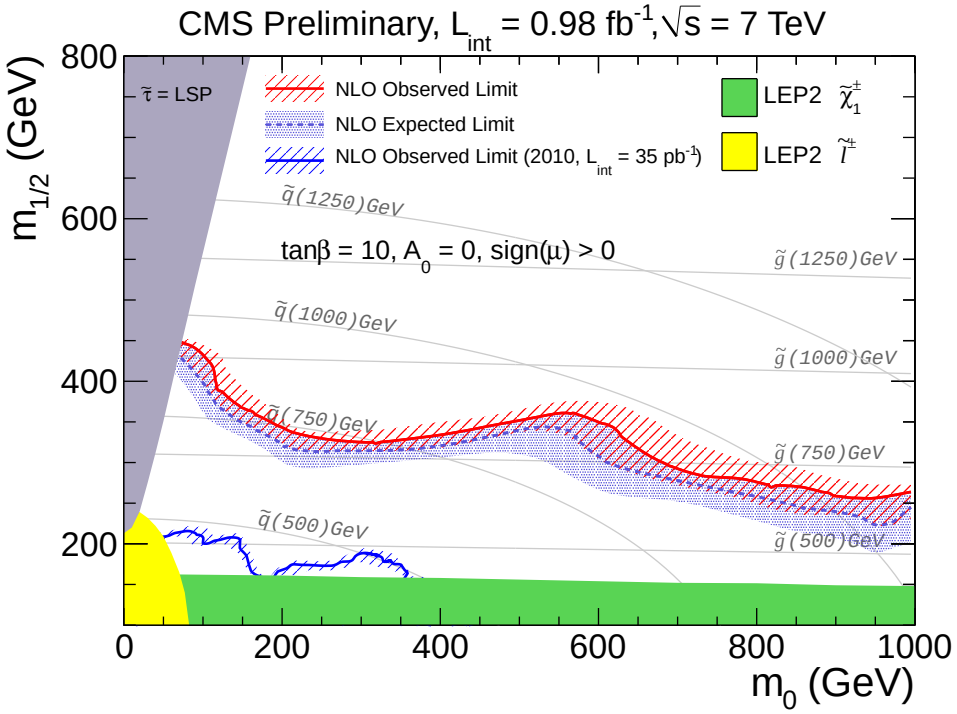

Figure 7.4: Exclusion region in the CMSSM corresponding to the observed upper limit of 3.0 events in the high-p_T signal region defined by $H_T > 400$ GeV and $\not{E}_T > 120$ GeV. The result of the previous analysis [32] is shown to illustrate the improvement since.

As a reference to other SUSY searches, we interpret the results in high-p_T search region 1 in the context of a CMSSM model [47]. The observed upper limit on the number

of signal events of 3.0 is compared to the expected number of events in the CMSSM model in a plane of $(m_0, m_{1/2})$ for $\tan\beta = 10$, $A_0 = 0$ and $\mu > 0$. All points with mean expected values above this upper limit are interpreted as excluded at the 95% CL. The observed exclusion region for the high-p_T dilepton selection is shown in Figure 7.4. The shaded region represents the uncertainty on the position of the limit due to an uncertainty on the production cross section of CMSSM resulting from PDF uncertainties and the NLO cross section uncertainty estimated from varying the renormalization scale by a factor of two. The expected exclusion region is approximately the same as the observed one. An exclusion region based on the 2010 analysis [32] is also shown for comparison. The new result extends to gluino masses of 825 GeV in the region with similar values of squark masses and extends to gluino masses of 675 GeV for higher squark masses. This can be compared to the exclusion of just around 500 GeV in the 2010 analysis.

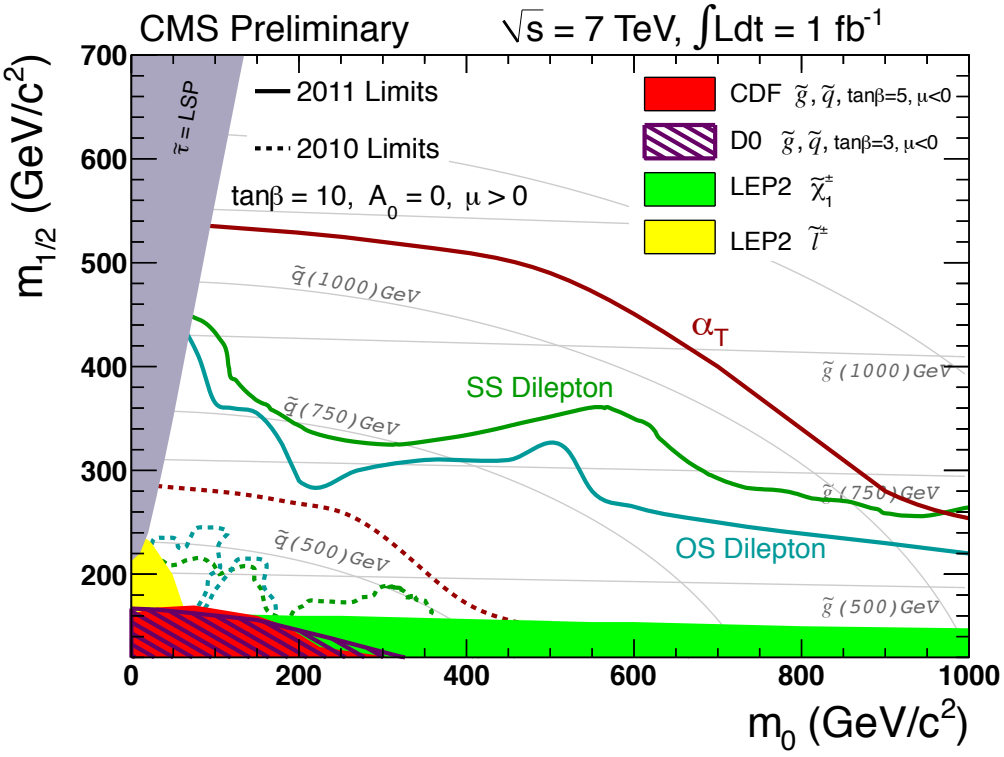

Figure 7.5: Observed limits for several 2011 CMS SUSY searches plotted in the CMSSM $(m_0, m_{1/2})$ plane. From [48].

Figure 7.5 presents the observed limits for the same-sign dilepton, opposite-sign dilepton [38], and all-hadronic (α_T) [49] new physics searches, separately for the 2010

and 2011 datasets. The same-sign dilepton analysis has slightly greater sensitivity than the opposite-sign search through the scanned parameter space. The all-hadronic search sets considerably better limits than either dilepton analysis, except at large values of m_0, where the same-sign dilepton search beings to overtake it. Some interesting physics of the CMSSM model unerlies the relative sensitivity of the all-hadronic and dilepton analyses.

The exclusion curves are shown in the m_0–$m_{1/2}$ plane. These are two of five parameters that define the CMSSM Lagrangian; the others are A_0, μ, and $\tan\beta$. The functional dependence of the supersymmetric particle masses on the Lagrangian parameters is generically given by [44]:

$$\begin{aligned} m_{\tilde{g}} &= f\left(m_{1/2}\right) \\ m_{\chi} &= g\left(m_{1/2}, \mu\right) \\ m_{\tilde{q}} &= h\left(m_0, m_{1/2}, \beta\right) \end{aligned}$$

If $m_{1/2} \gg m_0$, then $m_{\tilde{g}} \approx m_{\tilde{q}}$ and gluino production leads to $\tilde{g} \to \tilde{q}\bar{q}$ followed by $\tilde{q} \to q\chi$, while direct squark production gives just the latter. The final state will thus contain high-p_T jets and $\not\!\!E_T$ for which the all-hadronic analysis has better sensitivity. If instead, $m_{1/2} \ll m_0$, then $m_{\tilde{q}} \gg m_{\tilde{g}}$ and the favored gluino production can only proceed by a three body decay mediated by a virtual squark: $\tilde{g} \to \bar{q}\tilde{q}^* \to \bar{q}q\chi$. In the limit $q^2 \ll m_{\tilde{q}}^2$, the propagator of the virtual squark becomes constant, and so for fixed $m_{1/2}$, as m_0 increases the three body decay results in a final state with lower H_T and $\not\!\!E_T$. The χ, however, still offers the possibility for final state leptons with appreciable momentum. Thus, the sensitivity for the dilepton analyses begins to overtake that of the all-hadronic in the $m_0 \gg m_{1/2}$ region of parameter space.

Chapter 8

Summary and Conclusions

We have reported on a search for new physics in a final state with two same-sign leptons, missing transverse energy and significant hadronic activity. No evidence for new physics was observed. We set 95% CL upper limits based upon the observed yields, estimated backgrounds and reported uncertainties. We report exclusion limits for an mSUGRA model. The results extend significantly those reported previously from the Tevatron experiments.

The selections developed provide sensitivity to a broad range of potential new physics models. Data-driven methods were developed to estimate the dominant backgrounds from Standard Model processes. The background estimation methods were validated using control samples in data. To support these efforts, the SM $t\bar{t}$ cross section was measured using 3.1 pb^{-1} and found to be 194 $\pm$ 72 (stat) $\pm$ 24 (syst) $\pm$ 21(lum) pb. The tc$\not\!\!E_T$ algorithm was developed to improve the reconstruction of missing transverse energy by using tracking information to supplement energy deposits reconstructed in the calorimeter.

Rare Standard Model processes that have yet to observed were found to be significant backgrounds — in particular, $t\bar{t}W$ and $qqW^{\pm}W^{\pm}$. This observation has important implications for future analyses. The increasing importance of these backgrounds in regions with large H_T and $\not\!\!E_T$ may limit the future reach of this channel or, at a minimum, require modifications to the selections such as the introduction of a b-tagging requirement. The increasing contributions from these processes will also provide an opportunity to make first observations as the total integrated luminosity of the datasets increases. We intend to pursue both analysis paths with future data.

Appendix A

Table of Fake Rates

This appendix contains tables of fake rates values determined for electrons and muons using both data and simulation.

A.1 Fake Rates from Data

Fake rates for muons and electrons determined using 0.89 fb^{-1} of 7 TeV data.

A.2 Fake Rates from Simulation

Fake rates for muons and electrons using QCD events simulated at $\sqrt{s} = 7$ TeV.

Table A.1: Electron fake rate measured using triggers with an online isolation requirement. This fake rate is used in the dielectron final state of the high-p_T analysis. The fake rate was measured using 0.89 fb^{-1} of data.

$\lvert\eta\rvert$ \ p_T	10.000 – 15.000	15.000 – 20.000	20.000 – 25.000	25.000 – 35.000	35.000 – 55.000
0.000 – 1.000	0.2820 ± 0.0110	0.1903 ± 0.0110	0.1862 ± 0.0114	0.1901 ± 0.0122	0.2905 ± 0.0215
1.000 – 1.479	0.2929 ± 0.0189	0.2147 ± 0.0188	0.2128 ± 0.0170	0.2003 ± 0.0167	0.3111 ± 0.0282
1.479 – 2.000	0.2651 ± 0.0242	0.1922 ± 0.0208	0.1969 ± 0.0142	0.1902 ± 0.0139	0.2478 ± 0.0200
2.000 – 2.500	0.2790 ± 0.0212	0.2687 ± 0.0233	0.2531 ± 0.0171	0.2634 ± 0.0161	0.3404 ± 0.0219

Table A.2: Electron fake rate measured using triggers without an online isolation requirement. This fake rate is used in the inclusive dilepton analysis, as well as the $e\mu$ final state of the high-p_T analysis. The fake rate was measured using 0.89 fb^{-1} of data.

$\lvert\eta\rvert$ \ p_T	10.000 – 15.000	15.000 – 20.000	20.000 – 25.000	25.000 – 35.000	35.000 – 55.000
0.000 – 1.000	0.2234 ± 0.0217	0.1746 ± 0.0276	0.1228 ± 0.0307	0.1444 ± 0.0371	0.3400 ± 0.0670
1.000 – 1.479	0.2136 ± 0.0404	0.1867 ± 0.0450	0.1739 ± 0.0559	0.2353 ± 0.0594	0.3810 ± 0.1060
1.479 – 2.000	0.2241 ± 0.0548	0.2326 ± 0.0644	0.1528 ± 0.0424	0.2308 ± 0.0477	0.3409 ± 0.0715
2.000 – 2.500	0.2432 ± 0.0499	0.4038 ± 0.0680	0.1455 ± 0.0475	0.3182 ± 0.0573	0.3529 ± 0.0820

Table A.3: Muon fake rate measured in data. This fake rate is used for both the high-p_T and inclusive dilepton analyses. The fake rate was measured using 0.89 fb^{-1} of data.

$\lvert\eta\rvert$ \ p_T	5.000 – 10.000	10.000 – 15.000	15.000 – 20.000	20.000 – 25.000	25.000 – 35.000
0.000 – 1.000	0.2810 ± 0.0028	0.2205 ± 0.0025	0.1968 ± 0.0033	0.1725 ± 0.0045	0.1735 ± 0.0015
1.000 – 1.479	0.3106 ± 0.0039	0.2536 ± 0.0037	0.2143 ± 0.0047	0.1780 ± 0.0061	0.2033 ± 0.0023
1.479 – 2.000	0.3370 ± 0.0041	0.2730 ± 0.0039	0.2287 ± 0.0050	0.2065 ± 0.0068	0.2217 ± 0.0025
2.000 – 2.500	0.3304 ± 0.0052	0.2703 ± 0.0051	0.2039 ± 0.0062	0.1932 ± 0.0088	0.2351 ± 0.0038

Table A.4: Muon fake rate measured in simulation. No trigger requirement is made.

$\lvert\eta\rvert$ \ p_T	5.000 – 10.000	10.000 – 15.000	15.000 – 20.000	20.000 – 25.000	25.000 – 35.000
0.000 – 1.000	0.3016 ± 0.0018	0.2282 ± 0.0036	0.1390 ± 0.0008	0.1069 ± 0.0010	0.0895 ± 0.0012
1.000 – 1.479	0.3438 ± 0.0026	0.2577 ± 0.0053	0.1686 ± 0.0012	0.1421 ± 0.0016	0.1193 ± 0.0018
1.479 – 2.000	0.3362 ± 0.0026	0.2707 ± 0.0054	0.1818 ± 0.0012	0.1557 ± 0.0017	0.1432 ± 0.0020
2.000 – 2.500	0.3348 ± 0.0034	0.2749 ± 0.0073	0.1821 ± 0.0016	0.1498 ± 0.0022	0.1398 ± 0.0028

Table A.5: Electron fake rate measured in simulation. No trigger requirement is made.

$\lvert\eta\rvert$ \ p_T	10.000 – 15.000	15.000 – 20.000	20.000 – 25.000	25.000 – 35.000	35.000 – 55.000
0.000 – 1.000	0.1729 ± 0.0246	0.1355 ± 0.0325	0.1741 ± 0.0420	0.1327 ± 0.0354	0.1102 ± 0.0238
1.000 – 1.479	0.1536 ± 0.0369	0.0602 ± 0.0326	0.1215 ± 0.0431	0.1557 ± 0.0559	0.1623 ± 0.0388
1.479 – 2.000	0.1126 ± 0.0460	0.0904 ± 0.0392	0.2038 ± 0.0513	0.1600 ± 0.0381	0.2261 ± 0.0418
2.000 – 2.500	0.2224 ± 0.0631	0.1298 ± 0.0568	0.2734 ± 0.0601	0.2232 ± 0.0427	0.2572 ± 0.0425

Appendix B

Tables of Results

B.1 The High p_T Dilepton Analysis

B.1.1 High H_T, High $\not\!\!E_T$ Signal Region

B.1.2 High H_T, High $\not\!\!E_T$ Signal Region

B.1.3 Moderate H_T, High $\not\!\!E_T$ Signal Region

B.1.4 Low H_T, High $\not\!\!E_T$ Signal Region

B.2 The Inclusive Dilepton Analysis

B.2.1 High H_T, High $\not\!\!E_T$ Signal Region

B.2.2 High H_T, High $\not\!\!E_T$ Signal Region

B.2.3 Moderate H_T, High $\not\!\!E_T$ Signal Region

Table B.1: Observed yields in the high H_T, high $\not\!\!E_T$ signal region for the high-p_T analysis. The observed event count in data and the estimated background agree within uncertainties.

Source	ee	$\mu\mu$	eμ	all
$t\bar{t} \to \ell\ell X$	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
$t\bar{t} \to \ell(b \to \ell)X$	0.000 ± 0.144	0.000 ± 0.144	0.216 ± 0.216	0.216 ± 0.216
$t\bar{t} \to \ell(\not\!b \to \ell)X$	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
$t\bar{t}$ other	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
tW	0.000 ± 0.021	0.000 ± 0.021	0.000 ± 0.021	0.000 ± 0.021
t, t-channel	0.015 ± 0.042	0.000 ± 0.042	0.000 ± 0.042	0.015 ± 0.042
t, s-channel	0.000 ± 0.003	0.000 ± 0.003	0.000 ± 0.003	0.000 ± 0.003
W+jets	0.000 ± 2.039	0.000 ± 2.039	0.000 ± 2.039	0.000 ± 2.039
$Z \to ee$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \mu\mu$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \tau\tau$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
WW	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040
WZ	0.006 ± 0.006	0.000 ± 0.008	0.056 ± 0.026	0.062 ± 0.027
ZZ	0.000 ± 0.003	0.000 ± 0.003	0.000 ± 0.003	0.000 ± 0.003
Vγ	0.000 ± 0.210	0.000 ± 0.210	0.000 ± 0.210	0.000 ± 0.210
spW^+W^+	0.047 ± 0.008	0.069 ± 0.009	0.141 ± 0.013	0.256 ± 0.018
spW^-W^-	0.011 ± 0.002	0.016 ± 0.002	0.024 ± 0.003	0.052 ± 0.004
dp$W^\pm W^\pm$	0.000 ± 0.002	0.000 ± 0.002	0.000 ± 0.002	0.000 ± 0.002
$t\bar{t}W$	0.045 ± 0.004	0.082 ± 0.006	0.137 ± 0.007	0.264 ± 0.010
Total MC	0.126 ± 0.019	0.167 ± 0.011	0.574 ± 0.219	0.867 ± 0.220
LM6	0.804 ± 0.039	1.043 ± 0.044	1.712 ± 0.056	3.559 ± 0.081
Prompt-fake	0.25 ± 0.59	0.22 ± 0.56	0.31 ± 0.76	0.77 ± 0.83
Double-fake	0.00 ± 0.27	0.00 ± 0.26	0.00 ± 0.34	0.00 ± 0.34
Total with fakes	0.25 ± 0.25 ± 0.13	0.22 ± 0.22 ± 0.11	0.31 ± 0.32 ± 0.15	0.77 ± 0.46 ± 0.39
Charge misreconstruction	0.012 ± 0.006 ± 0.002	- ± -	0.014 ± 0.006 ± 0.003	0.026 ± 0.008 ± 0.005
Simulated backgrounds	0.110 ± 0.011 ± 0.055	0.167 ± 0.011 ± 0.083	0.358 ± 0.030 ± 0.179	0.635 ± 0.034 ± 0.317
All backgrounds	0.37 ± 0.25 ± 0.14	0.38 ± 0.22 ± 0.14	0.68 ± 0.33 ± 0.24	1.44 ± 0.47 ± 0.50
Data	0	0	0	0

Table B.2: Observed yields in the high H_T, low $\not\!\!E_T$ signal region for the high-p_T analysis. The observed event count in data and the estimated background agree within uncertainties.

Source	ee	$\mu\mu$	eμ	all
$t\bar{t} \to \ell\ell X$	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
$t\bar{t} \to \ell(b \to \ell)X$	0.255 ± 0.255	0.111 ± 0.111	0.216 ± 0.216	0.583 ± 0.353
$t\bar{t} \to \ell(\not{b} \to \ell)X$	0.255 ± 0.255	0.000 ± 0.144	0.247 ± 0.219	0.502 ± 0.336
$t\bar{t}$ other	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
tW	0.000 ± 0.021	0.000 ± 0.021	0.000 ± 0.021	0.000 ± 0.021
t, t-channel	0.015 ± 0.042	0.000 ± 0.042	0.000 ± 0.042	0.015 ± 0.042
t, s-channel	0.000 ± 0.003	0.000 ± 0.003	0.000 ± 0.003	0.000 ± 0.003
W+jets	0.000 ± 2.039	0.000 ± 2.039	0.000 ± 2.039	0.000 ± 2.039
$Z \to ee$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \mu\mu$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \tau\tau$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
WW	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040
WZ	0.013 ± 0.009	0.021 ± 0.016	0.062 ± 0.027	0.096 ± 0.032
ZZ	0.000 ± 0.003	0.001 ± 0.001	0.000 ± 0.003	0.001 ± 0.001
Vγ	0.000 ± 0.210	0.000 ± 0.210	0.000 ± 0.210	0.000 ± 0.210
spW^+W^+	0.102 ± 0.011	0.132 ± 0.013	0.271 ± 0.018	0.506 ± 0.025
spW^-W^-	0.026 ± 0.003	0.034 ± 0.003	0.059 ± 0.004	0.119 ± 0.006
dp$W^\pm W^\pm$	0.000 ± 0.002	0.000 ± 0.002	0.002 ± 0.002	0.002 ± 0.002
$t\bar{t}W$	0.113 ± 0.007	0.172 ± 0.008	0.300 ± 0.011	0.586 ± 0.015
Total MC	0.780 ± 0.362	0.473 ± 0.113	1.157 ± 0.310	2.410 ± 0.489
LM6	0.864 ± 0.040	1.122 ± 0.046	1.872 ± 0.058	3.858 ± 0.084
Prompt-fake	1.06 ± 0.76	0.97 ± 0.77	0.59 ± 0.81	2.62 ± 1.13
Double-fake	0.00 ± 0.27	0.00 ± 0.26	0.00 ± 0.34	0.00 ± 0.34
Total with fakes	1.06 ± 0.54 ± 0.53	0.97 ± 0.57 ± 0.48	0.59 ± 0.43 ± 0.30	2.62 ± 0.90 ± 1.31
Charge misreconstruction	0.064 ± 0.016 ± 0.013	- ± -	0.025 ± 0.007 ± 0.005	0.089 ± 0.017 ± 0.018
Simulated backgrounds	0.254 ± 0.016 ± 0.127	0.362 ± 0.022 ± 0.181	0.694 ± 0.034 ± 0.347	1.310 ± 0.044 ± 0.655
All backgrounds	1.38 ± 0.54 ± 0.54	1.33 ± 0.57 ± 0.52	1.31 ± 0.43 ± 0.46	4.02 ± 0.90 ± 1.46
Data	1	2	2	5

Table B.3: Observed yields in the moderate H_T, high $\not\!\!E_T$ signal region for the high-p_T analysis. The observed event count in data and the estimated background agree within uncertainties.

Source	ee	$\mu\mu$	eμ	all
$t\bar{t} \to \ell\ell X$	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
$t\bar{t} \to \ell(b \to \ell)X$	0.000 ± 0.144	0.255 ± 0.255	0.563 ± 0.326	0.819 ± 0.414
$t\bar{t} \to \ell(\not\!b \to \ell)X$	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
$t\bar{t}$ other	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
tW	0.000 ± 0.021	0.028 ± 0.020	0.000 ± 0.021	0.028 ± 0.020
t, t-channel	0.031 ± 0.022	0.000 ± 0.042	0.009 ± 0.042	0.040 ± 0.024
t, s-channel	0.000 ± 0.003	0.000 ± 0.003	0.000 ± 0.003	0.000 ± 0.003
W+jets	0.000 ± 2.039	0.000 ± 2.039	0.745 ± 2.039	0.745 ± 2.039
$Z \to ee$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \mu\mu$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \tau\tau$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
WW	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040
WZ	0.035 ± 0.017	0.033 ± 0.020	0.129 ± 0.039	0.197 ± 0.047
ZZ	0.005 ± 0.005	0.000 ± 0.003	0.000 ± 0.003	0.005 ± 0.005
Vγ	0.000 ± 0.210	0.000 ± 0.210	0.000 ± 0.210	0.000 ± 0.210
spW^+W^+	0.090 ± 0.010	0.132 ± 0.013	0.231 ± 0.017	0.454 ± 0.023
spW^-W^-	0.024 ± 0.003	0.037 ± 0.003	0.055 ± 0.004	0.116 ± 0.006
dp$W^\pm W^\pm$	0.000 ± 0.002	0.000 ± 0.002	0.000 ± 0.002	0.000 ± 0.002
$t\bar{t}W$	0.148 ± 0.008	0.238 ± 0.010	0.404 ± 0.012	0.790 ± 0.017
Total MC	0.333 ± 0.031	0.723 ± 0.257	2.136 ± 0.815	3.193 ± 0.855
LM6	0.899 ± 0.041	1.160 ± 0.047	1.914 ± 0.059	3.974 ± 0.086
Prompt-fake	0.90 ± 0.76	1.08 ± 0.76	0.80 ± 0.54	2.78 ± 0.95
Double-fake	0.00 ± 0.27	0.00 ± 0.26	0.09 ± 0.09	0.09 ± 0.09
Total with fakes	0.90 ± 0.54 ± 0.45	1.08 ± 0.56 ± 0.54	0.88 ± 0.52 ± 0.44	2.86 ± 0.93 ± 1.43
Charge misreconstruction	0.032 ± 0.009 ± 0.006	- ± -	0.046 ± 0.010 ± 0.009	0.078 ± 0.014 ± 0.016
Simulated backgrounds	0.302 ± 0.022 ± 0.151	0.440 ± 0.026 ± 0.220	0.819 ± 0.044 ± 0.409	1.562 ± 0.056 ± 0.781
All backgrounds	1.23 ± 0.54 ± 0.47	1.52 ± 0.56 ± 0.58	1.75 ± 0.52 ± 0.60	4.50 ± 0.93 ± 1.63
Data	0	2	1	3

Table B.4: Observed yields in the low H_T, high $\not\!\!E_T$ signal region for the high-p_T analysis. The observed event count in data and the estimated background agree within uncertainties.

Source	ee	$\mu\mu$	eμ	all
$t\bar{t} \to \ell\ell X$	0.162 ± 0.162	0.000 ± 0.144	0.000 ± 0.144	0.162 ± 0.162
$t\bar{t} \to \ell(b \to \ell)X$	0.733 ± 0.382	1.005 ± 0.473	1.117 ± 0.476	2.855 ± 0.772
$t\bar{t} \to \ell(\not{b} \to \ell)X$	0.333 ± 0.267	0.000 ± 0.144	0.000 ± 0.144	0.333 ± 0.267
$t\bar{t}$ other	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
tW	0.024 ± 0.024	0.028 ± 0.020	0.004 ± 0.021	0.056 ± 0.031
t, t-channel	0.031 ± 0.022	0.047 ± 0.047	0.009 ± 0.042	0.087 ± 0.053
t, s-channel	0.007 ± 0.005	0.004 ± 0.004	0.004 ± 0.004	0.014 ± 0.007
W+jets	0.000 ± 2.039	0.000 ± 2.039	0.745 ± 2.039	0.745 ± 2.039
$Z \to ee$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \mu\mu$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \tau\tau$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
WW	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040
WZ	0.051 ± 0.023	0.081 ± 0.030	0.287 ± 0.057	0.418 ± 0.068
ZZ	0.006 ± 0.005	0.001 ± 0.001	0.005 ± 0.005	0.012 ± 0.007
Vγ	0.000 ± 0.210	0.000 ± 0.210	0.000 ± 0.210	0.000 ± 0.210
spW^+W^+	0.147 ± 0.013	0.219 ± 0.016	0.359 ± 0.021	0.725 ± 0.030
spW^-W^-	0.041 ± 0.004	0.059 ± 0.004	0.103 ± 0.006	0.204 ± 0.008
dp$W^\pm W^\pm$	0.000 ± 0.002	0.002 ± 0.002	0.002 ± 0.002	0.004 ± 0.003
$t\bar{t}W$	0.280 ± 0.010	0.453 ± 0.013	0.776 ± 0.017	1.509 ± 0.024
Total MC	1.813 ± 0.495	1.899 ± 0.477	3.410 ± 0.887	7.122 ± 1.122
LM6	0.952 ± 0.042	1.217 ± 0.048	2.037 ± 0.061	4.206 ± 0.088
Prompt-fake	1.86 ± 0.95	1.81 ± 0.87	3.05 ± 1.33	6.71 ± 1.70
Double-fake	0.00 ± 0.27	0.00 ± 0.26	0.13 ± 0.10	0.13 ± 0.10
Total with fakes	1.86 ± 0.78 ± 0.93	1.81 ± 0.70 ± 0.90	3.18 ± 1.32 ± 1.59	6.84 ± 1.69 ± 3.42
Charge misreconstruction	0.122 ± 0.019 ± 0.024	- ± -	0.153 ± 0.018 ± 0.031	0.275 ± 0.026 ± 0.055
Simulated backgrounds	0.525 ± 0.029 ± 0.262	0.815 ± 0.037 ± 0.408	1.531 ± 0.063 ± 0.766	2.871 ± 0.079 ± 1.436
All backgrounds	2.51 ± 0.78 ± 0.97	2.62 ± 0.70 ± 0.99	4.86 ± 1.32 ± 1.76	9.99 ± 1.69 ± 3.71
Data	3	2	2	7

Table B.5: Observed yields in the high H_T, high $\not\!\!E_T$ signal region for the inclusive dilepton analysis. The observed event count in data and the estimated background agree within uncertainties.

Source	ee	$\mu\mu$	eμ	all
$t\bar{t} \to \ell\ell X$	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
$t\bar{t} \to \ell(b \to \ell)X$	0.000 ± 0.144	0.031 ± 0.144	0.328 ± 0.243	0.358 ± 0.245
$t\bar{t} \to \ell(\not\!b \to \ell)X$	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
$t\bar{t}$ other	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
tW	0.000 ± 0.021	0.000 ± 0.021	0.000 ± 0.021	0.000 ± 0.021
t, t-channel	0.015 ± 0.042	0.000 ± 0.042	0.000 ± 0.042	0.015 ± 0.042
t, s-channel	0.000 ± 0.003	0.000 ± 0.003	0.000 ± 0.003	0.000 ± 0.003
W+jets	0.000 ± 2.039	0.000 ± 2.039	0.000 ± 2.039	0.000 ± 2.039
$Z \to ee$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \mu\mu$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \tau\tau$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
WW	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040
WZ	0.006 ± 0.006	0.000 ± 0.008	0.056 ± 0.026	0.062 ± 0.027
ZZ	0.000 ± 0.003	0.000 ± 0.003	0.000 ± 0.003	0.000 ± 0.003
Vγ	0.000 ± 0.210	0.000 ± 0.210	0.000 ± 0.210	0.000 ± 0.210
spW^+W^+	0.047 ± 0.008	0.076 ± 0.010	0.146 ± 0.013	0.269 ± 0.018
spW^-W^-	0.011 ± 0.002	0.018 ± 0.002	0.025 ± 0.003	0.055 ± 0.004
dp$W^\pm W^\pm$	0.000 ± 0.002	0.000 ± 0.002	0.000 ± 0.002	0.000 ± 0.002
$t\bar{t}W$	0.046 ± 0.004	0.089 ± 0.006	0.148 ± 0.008	0.283 ± 0.010
Total MC	0.126 ± 0.019	0.213 ± 0.033	0.703 ± 0.245	1.042 ± 0.248
LM6	0.819 ± 0.039	1.230 ± 0.048	1.861 ± 0.058	3.909 ± 0.085
Prompt-fake	0.31 ± 0.97	0.90 ± 0.65	0.31 ± 0.76	1.52 ± 0.80
Double-fake	0.00 ± 0.46	0.08 ± 0.08	0.00 ± 0.34	0.08 ± 0.08
Total with fakes	0.31 ± 0.32 ± 0.15	0.98 ± 0.64 ± 0.49	0.31 ± 0.32 ± 0.15	1.60 ± 0.79 ± 0.80
Charge misreconstruction	0.012 ± 0.006 ± 0.002	- ± -	0.014 ± 0.006 ± 0.003	0.026 ± 0.008 ± 0.005
Simulated backgrounds	0.111 ± 0.011 ± 0.055	0.183 ± 0.011 ± 0.091	0.375 ± 0.031 ± 0.187	0.669 ± 0.034 ± 0.334
All backgrounds	0.43 ± 0.32 ± 0.16	1.16 ± 0.64 ± 0.50	0.70 ± 0.33 ± 0.24	2.29 ± 0.79 ± 0.87
Data	0	1	0	1

Table B.6: Observed yields in the high H_T, low $\not\!\!E_T$ signal region for the inclusive dilepton analysis. The observed event count in data and the estimated background agree within uncertainties.

Source	ee	$\mu\mu$	eμ	all
$t\bar{t} \to \ell\ell X$	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
$t\bar{t} \to \ell(b \to \ell)X$	0.255 ± 0.255	0.759 ± 0.381	1.098 ± 0.507	2.113 ± 0.683
$t\bar{t} \to \ell(\not{b} \to \ell)X$	0.255 ± 0.255	0.000 ± 0.144	0.247 ± 0.219	0.502 ± 0.336
$t\bar{t}$ other	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
tW	0.000 ± 0.021	0.024 ± 0.018	0.038 ± 0.038	0.062 ± 0.042
t, t-channel	0.015 ± 0.042	0.000 ± 0.042	0.032 ± 0.032	0.048 ± 0.036
t, s-channel	0.000 ± 0.003	0.000 ± 0.003	0.000 ± 0.003	0.000 ± 0.003
W+jets	0.000 ± 2.039	0.000 ± 2.039	0.000 ± 2.039	0.000 ± 2.039
$Z \to ee$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \mu\mu$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \tau\tau$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
WW	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040
WZ	0.013 ± 0.009	0.021 ± 0.016	0.062 ± 0.027	0.096 ± 0.032
ZZ	0.000 ± 0.003	0.001 ± 0.001	0.000 ± 0.003	0.001 ± 0.001
Vγ	0.000 ± 0.210	0.000 ± 0.210	0.000 ± 0.210	0.000 ± 0.210
spW^+W^+	0.102 ± 0.011	0.146 ± 0.013	0.279 ± 0.018	0.527 ± 0.025
spW^-W^-	0.026 ± 0.003	0.038 ± 0.003	0.061 ± 0.004	0.125 ± 0.006
dp$W^\pm W^\pm$	0.000 ± 0.002	0.000 ± 0.002	0.002 ± 0.002	0.002 ± 0.002
$t\bar{t}W$	0.114 ± 0.007	0.186 ± 0.008	0.317 ± 0.011	0.617 ± 0.015
Total MC	0.782 ± 0.362	1.175 ± 0.382	2.137 ± 0.555	4.094 ± 0.765
LM6	0.878 ± 0.040	1.315 ± 0.050	2.036 ± 0.061	4.229 ± 0.088
Prompt-fake	1.03 ± 1.05	1.97 ± 1.00	0.61 ± 0.49	3.61 ± 1.22
Double-fake	0.00 ± 0.46	0.17 ± 0.12	0.08 ± 0.08	0.25 ± 0.15
Total with fakes	1.03 ± 0.50 ± 0.51	2.15 ± 0.98 ± 1.07	0.68 ± 0.47 ± 0.34	3.86 ± 1.20 ± 1.93
Charge misreconstruction	0.057 ± 0.015 ± 0.011	- ± -	0.030 ± 0.008 ± 0.006	0.086 ± 0.017 ± 0.017
Simulated backgrounds	0.255 ± 0.016 ± 0.128	0.392 ± 0.023 ± 0.196	0.721 ± 0.035 ± 0.361	1.369 ± 0.044 ± 0.685
All backgrounds	1.34 ± 0.50 ± 0.53	2.54 ± 0.98 ± 1.09	1.43 ± 0.47 ± 0.50	5.31 ± 1.20 ± 2.05
Data	1	4	2	7

Table B.7: Observed yields in the moderate H_T, high $\not\!\!E_T$ signal region for the inclusive dilepton analysis. The observed event count in data and the estimated background agree within uncertainties.

Source	ee	$\mu\mu$	eμ	all
$t\bar{t} \to \ell\ell X$	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
$t\bar{t} \to \ell(b \to \ell)X$	0.000 ± 0.144	1.345 ± 0.573	1.228 ± 0.489	2.572 ± 0.753
$t\bar{t} \to \ell(\not\ell \to \ell)X$	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
$t\bar{t}$ other	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144	0.000 ± 0.144
tW	0.000 ± 0.021	0.032 ± 0.020	0.000 ± 0.021	0.032 ± 0.020
t, t-channel	0.031 ± 0.022	0.047 ± 0.047	0.041 ± 0.034	0.119 ± 0.062
t, s-channel	0.000 ± 0.003	0.000 ± 0.003	0.000 ± 0.003	0.000 ± 0.003
W+jets	0.000 ± 2.039	0.000 ± 2.039	0.745 ± 2.039	0.745 ± 2.039
$Z \to ee$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \mu\mu$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
$Z \to \tau\tau$	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169	0.000 ± 1.169
WW	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040	0.000 ± 0.040
WZ	0.035 ± 0.017	0.035 ± 0.020	0.129 ± 0.039	0.199 ± 0.047
ZZ	0.005 ± 0.005	0.000 ± 0.003	0.000 ± 0.003	0.005 ± 0.005
Vγ	0.000 ± 0.210	0.000 ± 0.210	0.000 ± 0.210	0.000 ± 0.210
spW^+W^+	0.092 ± 0.010	0.148 ± 0.013	0.248 ± 0.017	0.488 ± 0.024
spW^-W^-	0.024 ± 0.003	0.044 ± 0.004	0.060 ± 0.004	0.128 ± 0.006
dp$W^\pm W^\pm$	0.000 ± 0.002	0.002 ± 0.002	0.000 ± 0.002	0.002 ± 0.002
$t\bar{t}W$	0.150 ± 0.008	0.269 ± 0.010	0.437 ± 0.013	0.856 ± 0.018
Total MC	0.336 ± 0.031	1.922 ± 0.576	2.889 ± 0.893	5.146 ± 1.063
LM6	0.915 ± 0.041	1.367 ± 0.051	2.077 ± 0.062	4.359 ± 0.090
Prompt-fake	1.15 ± 1.10	2.17 ± 1.05	1.01 ± 0.75	4.33 ± 1.42
Double-fake	0.00 ± 0.46	0.28 ± 0.17	0.20 ± 0.15	0.49 ± 0.23
Total with fakes	1.15 ± 0.61 ± 0.58	2.45 ± 1.01 ± 1.23	1.22 ± 0.70 ± 0.61	4.82 ± 1.37 ± 2.41
Charge misreconstruction	0.032 ± 0.009 ± 0.006	- ± -	0.050 ± 0.010 ± 0.010	0.083 ± 0.014 ± 0.017
Simulated backgrounds	0.305 ± 0.022 ± 0.153	0.498 ± 0.027 ± 0.249	0.874 ± 0.045 ± 0.437	1.677 ± 0.057 ± 0.839
All backgrounds	1.49 ± 0.61 ± 0.60	2.95 ± 1.01 ± 1.25	2.14 ± 0.70 ± 0.75	6.58 ± 1.37 ± 2.55
Data	0	4	2	6

Appendix C

Description of mSUGRA Points

Table C.1: Parameters defining the various mSUGRA benchmarks.

Benchmark	m_0	$m_{1/2}$	A_0	$\tan\beta$	sign(μ)
LM0	200	160	-400	10	+
LM1	60	250	0	10	+
LM2	185	350	0	35	+
LM3	330	240	0	20	+
LM4	210	285	0	10	+
LM5	230	360	0	10	+
LM6	85	400	0	10	+
LM7	3000	230	0	10	+
LM8	500	300	-300	10	+
LM9	1450	175	0	50	+
LM11	250	325	0	35	+
LM12	2544.58	246.564	-865.752	47.5897	+
LM13	270	218	-553	40	+

Appendix D

Triggers

This appendix contains a list of the triggers used to select signal–like events and the run ranges for which those triggers were active. Only unprescaled triggers were used. Tables D.1– D.3 provide a list of triggers used for the high-p_T analysis. Triggers used to select events for the inclusive dilepton analysis can be found in Tables D.4– D.6

Table D.1: Double electron signal triggers used in the high-p_T analysis.

Trigger	Run Range
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL_v1	160329 – 161176
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL_v2	161210 – 163262
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL_v3	163265 – 164236
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL_v4	165088 – 165887
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL_v5	165922 – 166967
HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL_v6	166979 – 167784
HLT_Ele17_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL_Ele8_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL_v2	161200 – 163262
HLT_Ele17_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL_Ele8_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL_v3	163265 – 164236
HLT_Ele17_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL_Ele8_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL_v4	165088 – 165887
HLT_Ele17_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL_Ele8_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL_v5	165922 – 166967
HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v5	166979 – 999999

Table D.2: Electron-muon signal cross triggers used in the high-p_T analysis.

Trigger	Run Range
HLT_Mu17_Ele8_CaloIdL_v1	160329 – 161176
HLT_Mu17_Ele8_CaloIdL_v2	161210 – 163262
HLT_Mu17_Ele8_CaloIdL_v3	163265 – 164236
HLT_Mu17_Ele8_CaloIdL_v4	165088 – 165887
HLT_Mu17_Ele8_CaloIdL_v5	165922 – 166967
HLT_Mu17_Ele8_CaloIdL_v6	166979 – 167784
HLT_Mu8_Ele17_CaloIdL_v1	160329 – 161176
HLT_Mu8_Ele17_CaloIdL_v2	161210 – 163262
HLT_Mu8_Ele17_CaloIdL_v3	163265 – 164236
HLT_Mu8_Ele17_CaloIdL_v4	165088 – 165887
HLT_Mu8_Ele17_CaloIdL_v5	165922 – 166967
HLT_Mu8_Ele17_CaloIdL_v6	166979 – 167784

Table D.3: Double muon signal triggers used in the high-p_T analysis.

Trigger	Run Range
HLT_DoubleMu7_v1	160325 – 163262
HLT_DoubleMu7_v2	163265 – 164236
HLT_Mu13_Mu8_v2	165088 – 166345
HLT_Mu13_Mu8_v2	166347 – 167043
HLT_Mu13_Mu8_v3	166346 – 166346
HLT_Mu13_Mu8_v4	167078 – 167784

Table D.4: Double electron signal triggers used in the low-p_T analysis.

Trigger	Run Range
HLT_DoubleEle8_CaloIdL_TrkIdVL_HT160_v2	160329 – 161176
HLT_DoubleEle8_CaloIdL_TrkIdVL_HT160_v3	161210 – 163262
HLT_DoubleEle8_CaloIdT_TrkIdVL_HT160_v2	160329 – 161176
HLT_DoubleEle8_CaloIdT_TrkIdVL_HT160_v3	161210 – 163262
HLT_DoubleEle8_CaloIdL_TrkIdVL_HT150_v1	163265 – 164236
HLT_DoubleEle8_CaloIdL_TrkIdVL_HT150_v2	165088 – 165887
HLT_DoubleEle8_CaloIdT_TrkIdVL_HT150_v1	163265 – 164236
HLT_DoubleEle8_CaloIdT_TrkIdVL_HT150_v2	165088 – 165887
HLT_DoubleEle8_CaloIdT_TrkIdVL_HT150_v3	165922 – 166967
HLT_DoubleEle8_CaloIdT_TrkIdVL_HT150_v4	166979 – 167784

Table D.5: Electron-muon signal cross triggers used in the low-p_T analysis.

Trigger	Run Range
HLT_Mu3_Ele8_CaloIdL_TrkIdVL_HT160_v2	160329 – 161176
HLT_Mu3_Ele8_CaloIdL_TrkIdVL_HT160_v3	161210 – 163262
HLT_Mu3_Ele8_CaloIdT_TrkIdVL_HT160_v2	160329 – 161176
HLT_Mu3_Ele8_CaloIdT_TrkIdVL_HT160_v3	161210 – 163262
HLT_Mu3_Ele8_CaloIdL_TrkIdVL_HT150_v1	163265 – 164236
HLT_Mu3_Ele8_CaloIdL_TrkIdVL_HT150_v2	165088 – 165887
HLT_Mu3_Ele8_CaloIdL_TrkIdVL_HT150_v3	165922 – 166967
HLT_Mu3_Ele8_CaloIdT_TrkIdVL_HT150_v1	163265 – 164236
HLT_Mu3_Ele8_CaloIdT_TrkIdVL_HT150_v2	165088 – 165887
HLT_Mu3_Ele8_CaloIdT_TrkIdVL_HT150_v3	165922 – 166967
HLT_Mu3_Ele8_CaloIdT_TrkIdVL_HT150_v4	166979 – 999999

Table D.6: Double muon signal triggers used in the low-p_T analysis.

Trigger	Run Range
HLT_DoubleMu3_HT160_v2	160329 – 161176
HLT_DoubleMu3_HT160_v3	161210 – 163262
HLT_DoubleMu3_HT200_v2	160329 – 161176
HLT_DoubleMu3_HT200_v3	161210 – 163262
HLT_DoubleMu3_HT200_v4	163269 – 164236
HLT_DoubleMu3_HT200_v5	165088 – 165887
HLT_DoubleMu3_HT200_v6	165922 – 166345
HLT_DoubleMu3_HT200_v6	166347 – 167043
HLT_DoubleMu3_HT200_v7	166346 – 166346
HLT_DoubleMu3_HT200_v8	167078 – 999999
HLT_DoubleMu3_HT150_v1	163265 – 164236
HLT_DoubleMu3_HT150_v2	165088 – 165887
HLT_DoubleMu3_HT150_v3	165922 – 166345
HLT_DoubleMu3_HT150_v3	166347 – 167043
HLT_DoubleMu3_HT150_v4	166346 – 166346
HLT_DoubleMu3_HT150_v5	167078 – 999999

Bibliography

[1] The CMS, Cms - media, http://cms.web.cern.ch/cms/Welcome.html/CMSdocuments/DetectorDrawings/Slice/CMS_Slice.gif.

[2] The CMS, Cms technical design report, 1998.

[3] W.-M. Yao *et al.*, J. Phys. G **33**, 1+ (2006).

[4] The PDG, Pdg, http://pdg2.lbl.gov/atlasblog/wp-content/uploads/2010/04/Picture-1.png.

[5] G. Bertone, D. Hooper and J. Silk, Phys.Rept. **405**, 279 (2005), hep-ph/0404175.

[6] H. Baer, p. 211 (2009), 0901.4732.

[7] CMS PAS JME-09-010.

[8] G. Landsberg and F. Moortgat, CMS AN-2008/089.

[9] CMS PAS JME-10-010.

[10] S. Esen *et al.*, CMS AN-2007/041.

[11] CMS PAS PFT-09-001.

[12] P. Azzurri and B. Mangano, CMS IN-2008/017.

[13] CMS PAS TRK-09-001.

[14] P. Billoir and S. Qian, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **294**, 219 (1990).

[15] (2010), 0911.4881.

[16] W. Andrews *et al.*, CMS AN-2010/028.

[17] CMS PAS JME-10-004.

[18] CMS PAS JME-10-009.

[19] W. Andrews *et al.*, CMS AN-2009/050.

[20] W. Andrews *et al.*, CMS AN-2009/130.

[21] W. Andrews *et al.*, CMS AN-2009/138.

[22] CMS PAS TOP-09-002.

[23] T. Sjostrand, S. Mrenna and P. Z. Skands, JHEP **0605**, 026 (2006), hep-ph/0603175.

[24] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, JHEP **1106**, 128 (2011), 1106.0522.

[25] The CMS, Collisions2010recipes, `https://twiki.cern.ch/twiki/bin/viewauth/CMS/Collisions2010Recipes`.

[26] D. Barge *et al.*, CMS AN-2009/159.

[27] A. P. Vartak, M. LeBourgeois and V. Sharma, CMS AN-2010/106.

[28] M. Cacciari, G. P. Salam and G. Soyez, JHEP **04**, 063 (2008), 0802.1189.

[29] CMS PAS EWK-10-002.

[30] M. Mulders *et al.*, CMS AN-2008/098.

[31] CMS PAS EGM-10-004.

[32] JHEP **1106**, 077 (2011), 1104.3168.

[33] L. Bauerdick *et al.*, CMS AN-2011/155.

[34] M. Cacciari and G. P. Salam, (2007), 0707.1378v2.

[35] CMS PAS JME-07-002.

[36] CMS PAPER JME-10-011.

[37] CMS AN-2010/003.

[38] L. Bauerdick *et al.*, CMS AN-2011/269.

[39] W. Andrews *et al.*, CMS AN-2010/257.

[40] S. Chatrchyan *et al.*, Phys.Lett. **B699**, 25 (2011), 1102.5429.

[41] CMS PAS EWK-11-009.

[42] CMS PAS EWK-11-010.

[43] D. Acosta *et al.*, CMS AN-2011/184.

[44] The CMS, Cms technical design report, volume ii: Physics performance, 2006.

[45] L. Bauerdick *et al.*, CMS AN-2011/258.

[46] S. Sekmen, M. T. Zeyrek, S. Kraml and M. Spiropolu, CMS IN-2008/002.

[47] G. L. Kane, C. F. Kolda, L. Roszkowski and J. D. Wells, Phys.Rev. **D49**, 6173 (1994), hep-ph/9312272.

[48] The CMS, Physicsresultssus, `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS`.

[49] CMS PAS SUS-11-003.