ABSTRACT

Search for Supersymmetric Partners of the Top Quark with the CMS Detector and Novel Top Quark Tagging Algorithms

Kenneth Remington Call, Ph.D.

Advisor: Kenichi Hatakeyama, Ph.D.

The standard model of particle physics has been highly successful in explaining observed interactions of subatomic particles. The model was largely vindicated with the announced observation of the Higgs boson by the ATLAS and CMS experiments at the LHC. The astronomical observation of dark matter, among other reasons, motivates us to look for particles not part of the standard model. One extension of the model posits an additional symmetry which pairs every particle with a supersymmetric partner; fermions paired to bosons, and bosons paired with fermions. Using data collected with the CMS detector at a center of mass energy of 13 TeV during 2016, we have looked for evidence that the superpartner of the top quark was produced then decayed into an undetected particle and a top quark that decayed to hadronic jets. No statistically significant deviation from standard model expectations is observed. Limits are placed on the production cross section in the context of a simplified supersymmetric model. The search for the production of supersymmetric particles

will continue with larger datasets and more sophisticated analyses. An improved top quark tagging algorithm with enhanced sensitivity to top quarks that decay into three hadronic jets is presented.

Search for Supersymmetric Partners of the Top Quark with the CMS Detector and Novel Top Quark Tagging Algorithms

by

Kenneth Remington Call, B.Sc., M.A.

A Dissertation

Approved by the Department of Physics

____

Dwight P. Russell, Ph.D., Interim Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

Approved by the Dissertation Committee

____

Kenichi Hatakeyama, Ph.D., Chairperson

____

Jay Dittmann, Ph.D.

____

Gerald Cleaver, Ph.D.

____

Howard Lee, Ph.D.

____

Kevin Pinney, Ph.D.

Accepted by the Graduate School
August 2019

____

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF ACRONYMS

BDT Boosted Decision Tree

BSM Beyond the Standard Model

CSC Cathode Strip Chamber

CL Confidence Level

CR Control Region

CMS Compact Muon Solenoid

CTF Combinatorial Track Finder

DM Dark Matter

DNN Dense Neural Network

DT Drift Tube

ECAL Electromagnetic Calorimeter

FPR Fake Prediction Rate

GSF Gaussian Sum Filter

L1 Level-1 Trigger

LHC Large Hadron Collider

LPC LHC Physics Center

LSP Lightest Supersymmetric Particle

HCAL Hadron Calorimeter

HLT High Level Trigger

ISR Initial State Radiation

MSSM Minimal Supersymmetric Standard Model

NLL Next-to-Leading Logarithmic

NLO Next-to-Leading Order

PU Pileup

PF Particle Flow

QCD Quantum Chromodynamics

QED Quantum Electrodynamics

QIE Charge Integrator and Encoder

RNN Recurrent Neural Network

ROC Receiver Operating Characteristic

RPC Resistive Plate Chamber

SM Standard Model

SUSY Supersymmetry

TPR True Prediction Rate

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I wish to acknowledge the many individuals and organizations that have taught me, guided me, and been my partners in this process. Principally, I must recognize my advisor, Dr. Kenichi Hatakeyama. His expectations, guidance, encouragement was irreplaceable. He and Dr. Jay Dittmann stood ever ready to explain details of particle physics research and accelerator operations. The Baylor research group has been an amazing team to work with, and we have enjoyed a high degree of camaraderie. Working with the other graduate students, postdocs, professors, and undergraduates has been a pleasure and a great source of mutual support. I am particular grateful for the personal guidance and education provided by a Dr. Hongxuan Liu and Dr. Joe Pastika. I worked closely with them, and learned much from them. The CMS collaboration has been great to work with, and I appreciate the LPC at Fermilab for being at the center of CMS collaboration within the United States.

For my beloved Susie.

My partner in every adventure.

CHAPTER ONE

Introduction

Particle physics is the continued campaign to understand the physical structure of the universe at its most fundamental level. This means developing theories and designing experiments that look at physical phenomena that occur at the smallest scales and at the highest energies. Particle physics has its roots in the the discoveries of the X-rays in 1895 by Wilhem Roentgen, the electron in 1897 by J.J. Thomson, and the atomic nucleus in 1911 by Earnest Rutherford. With the dawning understanding that the universe is composed of smaller, more fundamental particles, the goal becomes to discover these particles, and to develop theories that both describe them and explain them.

Early discoveries were made with naturally occurring radioisotopes and cosmic rays. Progress has continued through the construction of increasingly powerful and intricate particle accelerators and particle detectors. The Standard Model (SM) is currently the best explanation for observed physical phenomena. It is not without itself weaknesses and deficiencies, but the theory has been highly successful. The predictions of the theory are some of the most accurate and precise of any type of physical experiment. With the discovery of the Higgs boson by the ATLAS [1] and CMS [2] experiments at the LHC in 2012, all particles of the standard model had been observed. Despite the success of the standard model, there is evidence that it is incomplete, and work to find particles and interactions that are not included within the standard model, or Beyond the Standard Model (BSM) physics, is an important and ongoing effort.

The standard model of particle physics has been highly successful in explaining the observed interactions of subatomic particles. The model was largely vindicated with the announced observation of the Higgs boson by the ATLAS and CMS experiments at the LHC 1.1. The astronomical observation of dark matter, among other reasons, motivates us to look for particles not part of the standard model. One extension of the model posits an additional symmetry which pairs every particle with a supersymmetric partner; fermions paired to bosons, and bosons paired with fermions. Using data collected with the CMS detector at a center of mass energy of 13 TeV during 2016, we have looked for evidence that the superpartner of the top quark was produced then decayed into an undetected particle and a top quark that decayed to hadronic jets. No statistically significant deviation from standard model expectations is observed. Limits are placed on the production cross section in the context of a simplified supersymmetric model. The search for the production of supersymmetric particles will continue with larger datasets and more sophisticated analyses. An improved top quark tagging algorithm with enhanced sensitivity to top quarks that decay into three hadronic jets is presented.

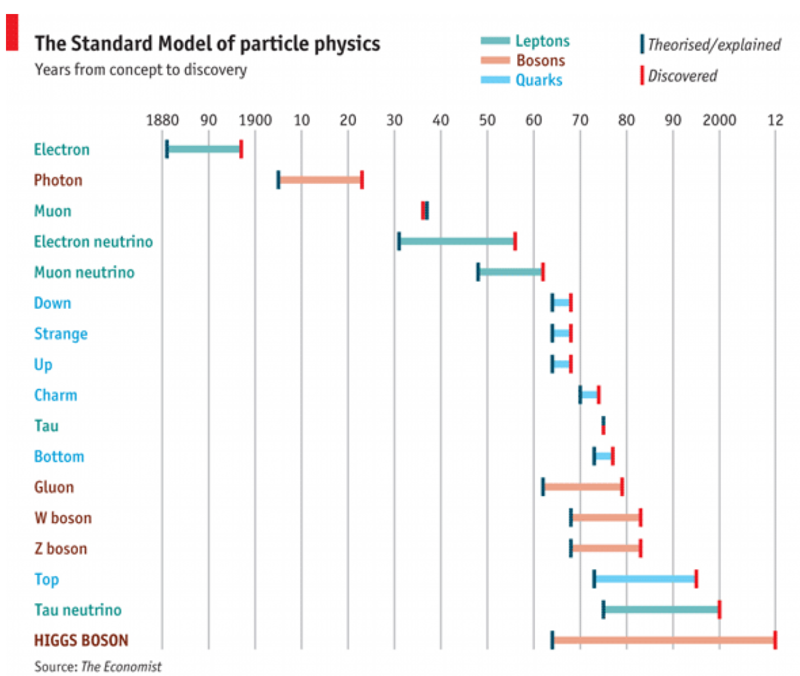


Figure 1.1: Timeline of standard model particle discovery. Image source [3]

The dissertation is organized as follows. In Chapter Two the standard model is introduced, evidence for physics beyond the standard model is presented, and the role of particle accelerators and particle detectors is discussed. Chapter Three gives an overview of the LHC and the CMS detector. In Chapter Four, more details are given about how the CMS detector is modeled in simulation, the production of simulated events, and the method used to reconstruct physics objects from the detector response. Chapter Five presents the method and results of the analysis which looks for BSM physics by looking at all-hadronic events with identified top quarks. Chapter Six delves deeper into the identification of top quarks, and presents a novel algorithm to identify top quarks using machine learning techniques. The summary is presented in Chapter Seven.

CHAPTER TWO

A Brief Overview of Particle Physics

2.1 The Standard Model

The standard model is the current generally accepted model of particle physics. The standard model is an example of a quantum field theory (QFT) in which all physical phenomena are described in terms of fields that exist everywhere and the interactions between these fields. The classical concept of a particle is an excitation of these fields to states above the vacuum state. It is built up from gauge fields, matter fields, and the Higgs field. It explains the electromagnetic, weak, and strong interactions. All of the fields have associated particles. Figure 2.1, shows the particles associated with the fields of the standard model and some of their qualities including mass, electric charge, and spin.

2.1.1 Symmetries in the Standard Model

The quantum fields of the standard model interact through gauge fields, which are in turn, defined by certain symmetries. Figure 2.2 shows which fields interact with one another in the standard model. During a collision in the accelerator, the quarks and gluons of the protons, which are collectively known as partons, interact with each other, and the excited states of the quark fields may induce excited states in other fields through the interactions depicted in Fig. 2.2. When calculating possible interactions and their cross sections, we can use the perturbative methods introduced by Feynman, which introduce the terms in the series in the forms of graphs, the edges of the graphs are excitations of the standard model fields, and the nodes are vertices where those fields interact according to the theory. Figure 2.3 shows the possible vertices.

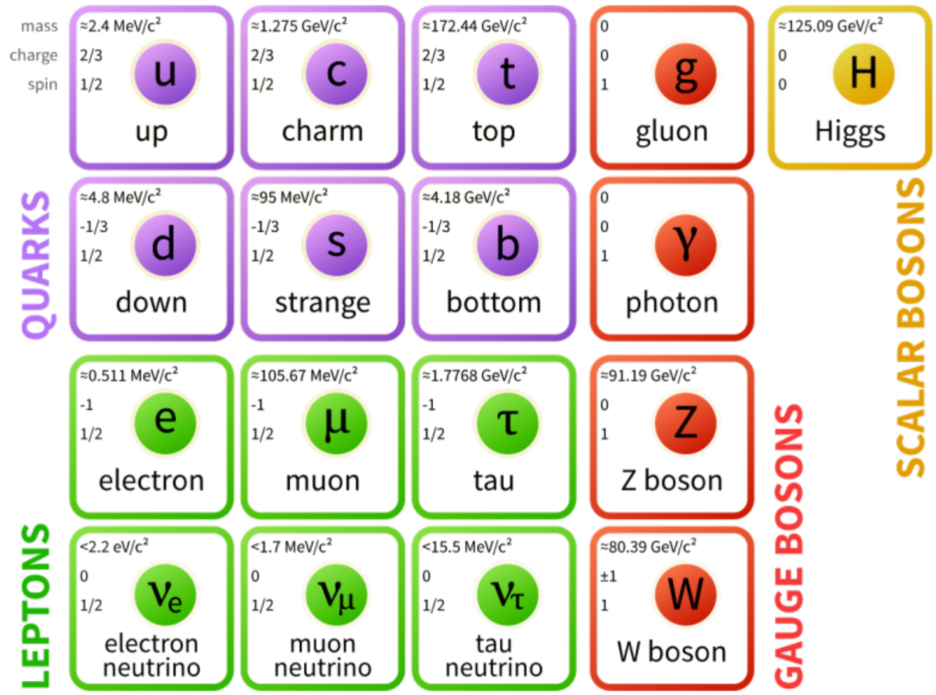


Figure 2.1: The particles of the standard model. The standard model includes three generations of quarks and leptons, the gauge fields that give rise to the electromagnetic, weak, and strong forces, and the Higgs field which gives mass to W and Z bosons through the Higgs mechanism. Image Source [4].

2.1.2 Open Questions Beyond the Standard Model

Despite the success of the standard model to describe subatomic particle interactions and its spectacular predictions of the Higgs boson, there are open questions about the SM itself, as well as phenomena that it is unable to explain.

- Why are there three generations of matter particles?
- Why are there the same number of quarks as leptons?
- Are there right-handed neutrinos?
- Is there a mechanism that explains the observed hierarchy of the electroweak and Planck scales?
- Is there a scale at which the coupling constants unify?
- Is dark matter a particle? If so how does it fit into the model?
- Can the model be extended to include gravity?

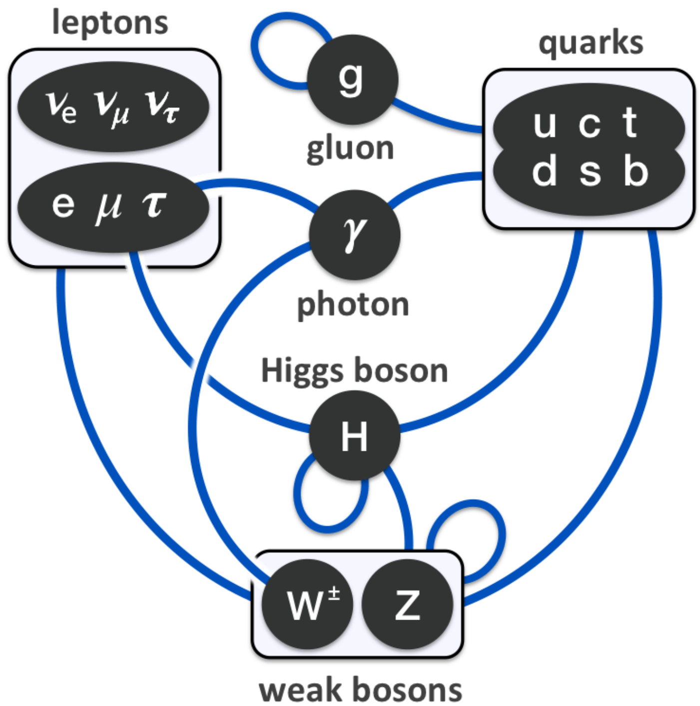


Figure 2.2: Standard model field interactions. The particles of the standard model interact through the electroweak and strong forces. Many also couple to the Higgs boson. This figure shows the particles that have vertices together. Image Source [4].

One of the most tantalizing clues for the existence of physics beyond the standard model is the existence of dark matter. Dark matter was first proposed to explain the light curves of rotating galaxies. Measurements showed that rotation of stars around galaxies was faster at distances far from the galactic core than was expected based on the visible matter in the galaxy and the application of the current known theories of gravity. The expected rotation curve and a representative observed rotation curve is depicted in Fig. 2.4. Such rotation curves could be explained by a large amount of matter in the galaxy that is not observable. Other explanations could include a modification of the current theories of gravity. Tantalizing evidence that dark matter is in fact a particle emerged with the observation of the Bullet Cluster. As seen in Fig. 2.5, the visible matter as observed by the Chandra X-ray observatory does not

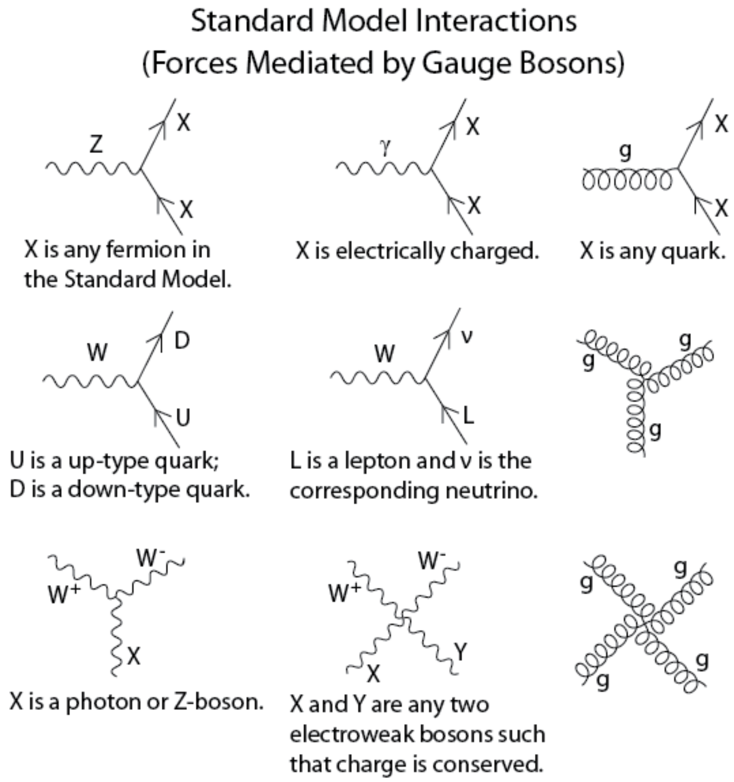


Figure 2.3: Gauge field interactions. These are the vertices that show the interaction of the standard model particles through gauge bosons. Image Source [4].

match the distribution of matter as measured through gravitational lensing. This is inconsistent with the hypothesis that dark matter phenomena are due to a modified theory of gravity.

2.2 Supersymmetry

A supersymmetric theory gives every field a supersymmetric partner. Every bosonic field has a fermionic supersymmetric partner. Every fermionic field has a bosonic supersymmetric partner. There are any number of ways that the standard model can be extended to become a supersymmetric theory. There are several motivations to believe that there is a supersymmetric extension of the standard model. One of these motivations is that it would lead to a natural explanation of the Higgs mass. Calculations of the Higgs mass just based on the standard model would indicate that

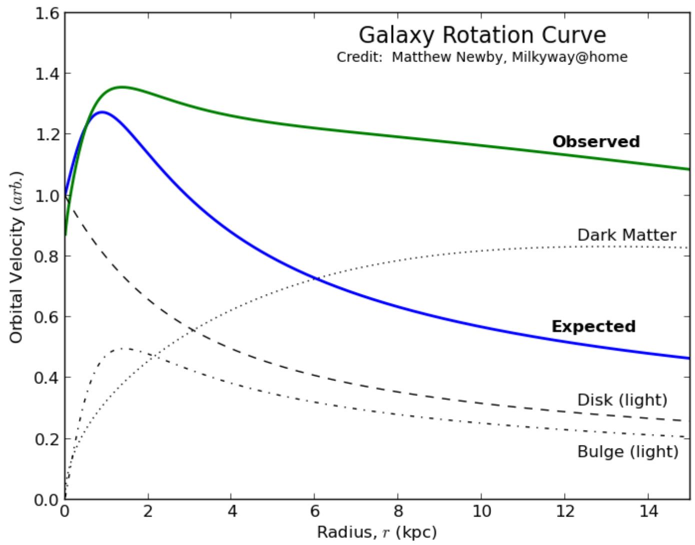


Figure 2.4: Galactic rotation curve. Given the distribution of visible matter in galaxies we expect the the velocities of the stars rotating around the galaxy to go as $1/\sqrt{r}$; however, the observed velocities have a flatter distribution indicating that there is matter present that is not visible. Image source [5].

it should have an extremely high mass due to contributions from its interactions with the other fields. That is unless the parameters of the standard model are fine tuned just so in order to give the mass that is observed. Supersymmetry would eliminate the need for this fine tuning. The additional graphs provided by the supersymmetric partners would cancel out the contributions from the standard model interactions. Figure 2.7 shows how a supersymmetric partner of a top quarks would cancel the contribution from a top quark loop.

2.3 Top Squark Production

In natural supersymmetric models, it is expected that the supersymmetric top quark partner, also known as the top squark or stop, will be among the lightest supersymmetric particles, and possibly one of the most abundantly directly produced supersymmetric partners generated by the strong interactions of proton-proton collisions at the LHC. Depending on the superpartner spectra, stop decay may have

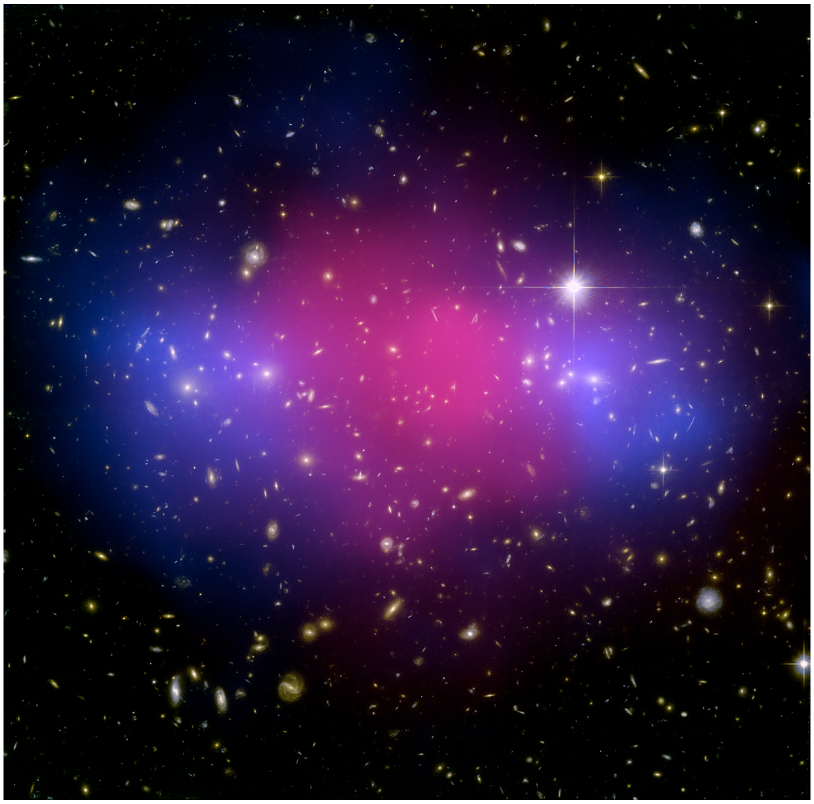


Figure 2.5: MACS J0025 as imaged by the Chandra X-ray observatory and visible light telescopes. Like the Bullet Cluster, MACS J0025.4-1222 is a galactic cluster that provides evidence that dark matter is a particle. When the cluster is observed by the Chandra X-ray observatory, the bulk of normal baryonic matter seems to be concentrated at the collision of the two galaxies. When the gravitational lensing of the cluster is calculated, non-visible matter within the galaxies appears to have moved past each other with relatively little interaction. Image source [6].

a easily observed signature; the decay of the stop into a light neutralino and a top quark.

2.3.1 The LSP

The LSP (lightest supersymmetric particle) plays a special role in the search for particles consistent with supersymmetric extensions of the standard model. In these extensions, baryon number and lepton number may no longer be conserved, and it is proposed that a number called R-parity be defined. It can be written as R-parity $= (-1)^{3(\mathrm{B}-\mathrm{L})+2\mathrm{s}}$ where B is the bayron number, L is the lepton number, and

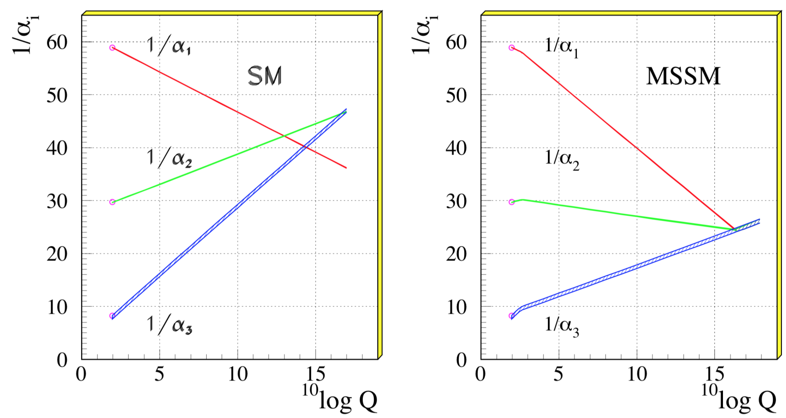


Figure 2.6: Running of coupling constants. At higher energy scales the coupling constants that describe the strength of the electromagnetic, weak, and strong forces approach a common value. Using only standard model calculations, the constants come close to, but fail to, meet at a common value. Introducing supersymmetry may mean that they meet at a common value at some energy scale indicating that there may be a theory that describes all three forces as a unified force. Image source [7].

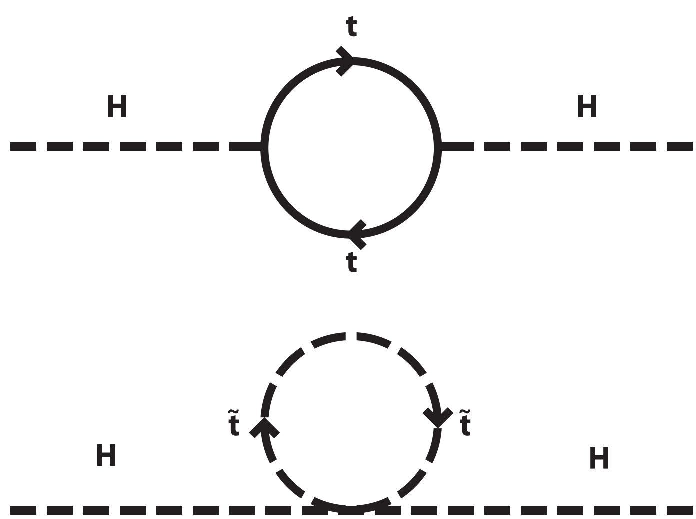


Figure 2.7: Diagrams showing the quantum loop corrections to the Higgs boson mass due to the top quark and its supersymmetric partner, top squark. A motivation to look for supersymmetric particles is that it may explain the relatively small mass of the Higgs boson.

s is the spin. R-parity = 1 for standard model particles and -1 for supersymmetric particles. If it is the case that R-parity is conserved, then it is expected that the LSP is stable since vertices with one supersymmetric partner and two standard model particles are forbidden. Up to this time there has been no evidence of electrically or color charged stable massive particles, so searches which assume a stable LSP look for evidence of an at most weakly interacting massive particle. Such a particle will show up in the detector as missing momentum.

A stable LSP is one of the motivations to continue looking at supersymmetric extensions of the standard model as it is a candidate as a dark matter particle.

2.4 The Role of Accelerators

Early observations in particles physics was done looking at particles produced in natural processes. Radioactive decay produces helium nuclei (alpha particles), electrons and positrons (beta particles), and gamma rays. Beta decays can also give support to the existence of neutrinos. Protons and neutrons are seen in fission events. Photographic plates can capture the tracks of muons and pions produced in the atmosphere from the showering of cosmic rays. Although the collisions that naturally occur in the atmosphere can be many orders of magnitude greater than any collision that has been produced in the laboratory, such collisions happen infrequently and far from detectors. In modern particle physics, accelerators are used to produce large numbers of high energy collisions very close to instrumentation.

Particle accelerators vary both in the particles that are collided, and in geometry.

2.4.1 Linear and Circular Colliders

All major colliders use high electric fields to accelerate electrically charged particles. The modern approach is to use radio frequency (RF) cavities and precise timing to accelerate particles with the oscillating electric field of a standing electromagnetic wave.

Some experiments collide accelerated particles with a stationary target. Such systems are not efficient in producing high mass products since to conserve momentum, much of the energy of the accelerated particles must go into boosting the products. A technically more challenging, but much more efficient, practice is to collide two beams of accelerated particles moving in opposite directions. In such collisions, the total momentum averages at 0, and much higher mass particles can be produced.

2.4.2 Hadron and Lepton Colliders

Any charged particle can be used in a particle accelerator.

Leptons are fundamental particles and interact through the relatively easy to compute electroweak force.

Hadrons are composite particles and in addition to electroweak interactions, interact through the computational difficult strong force. Any given inelastic collision can involve a variety of quarks and gluons with varying amounts of momentum. Figure 2.8 shows an example of an inelastic collision in a hadron collider.

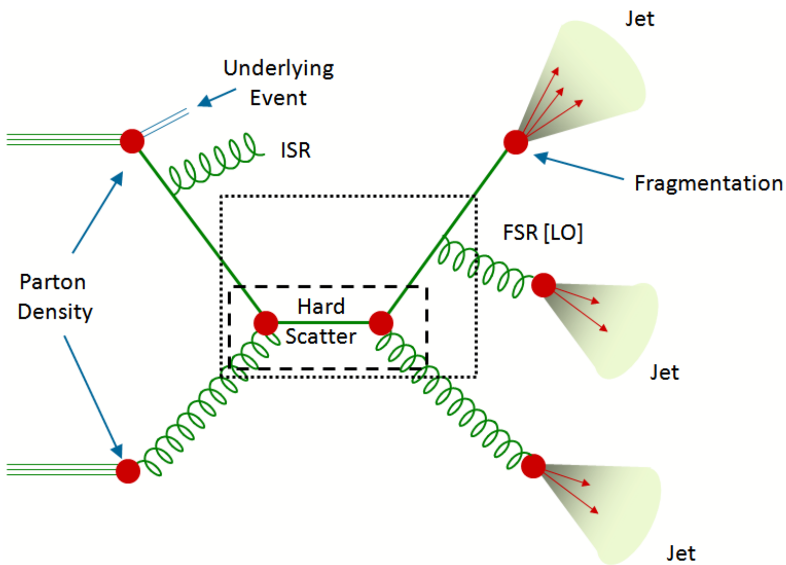


Figure 2.8: The partons are the parties involved in the collision. Before a hard scatter, there may be initial state radiation (ISR), after the hard scatter, the final products are produced and may decay further. There may be final state radiation (FSR) from these products.

CHAPTER THREE

Experimental Apparatus

The results presented in Chapter Five are based on data collected in 2016 by the CMS experiment. The complete experimental apparatus is properly understood to consist of the CERN accelerator complex that accelerates beams of protons or heavy ions through a series accelerators culminating with the unparalleled energies produced by Large Hadron Collider (LHC) [8] and the CMS detector [9] in which the beams are directed to collide and which collects information about the collisions using various subdetectors.

3.1 The Large Hadron Collider (LHC) at CERN

The LHC is located at CERN near Geneva, Switzerland. It has a circular shape 26.7 kilometers in circumference, and it is about 100 meters underground, as shown in Fig. 3.1. It occupies the tunnel created for the retired Large Electron-Positron collider (LEP) which operated from 1989 to 2000.

3.1.1 CERN Accelerator Complex

The LHC collides protons at the highest energy ever achieved by a particle accelerator. In 2012, the ATLAS and CMS experiments announced the discovery of the Higgs boson using data from proton-proton collision with center of mass energies of 7 TeV and 8 TeV. Figure 3.2 is a plot from the CMS announcement of the Higgs discovery. After the 2013–2014 long shutdown, from 2015 to 2018 (Run 2) the LHC collided protons with a center of mass energy of 13 TeV, meaning the individual protons each have an energy of 6.5 TeV.

Protons travel through a number of steps that start with a bottle of hydrogen and ends with a collision in the LHC. In proton-proton collisions, the accelerator stages

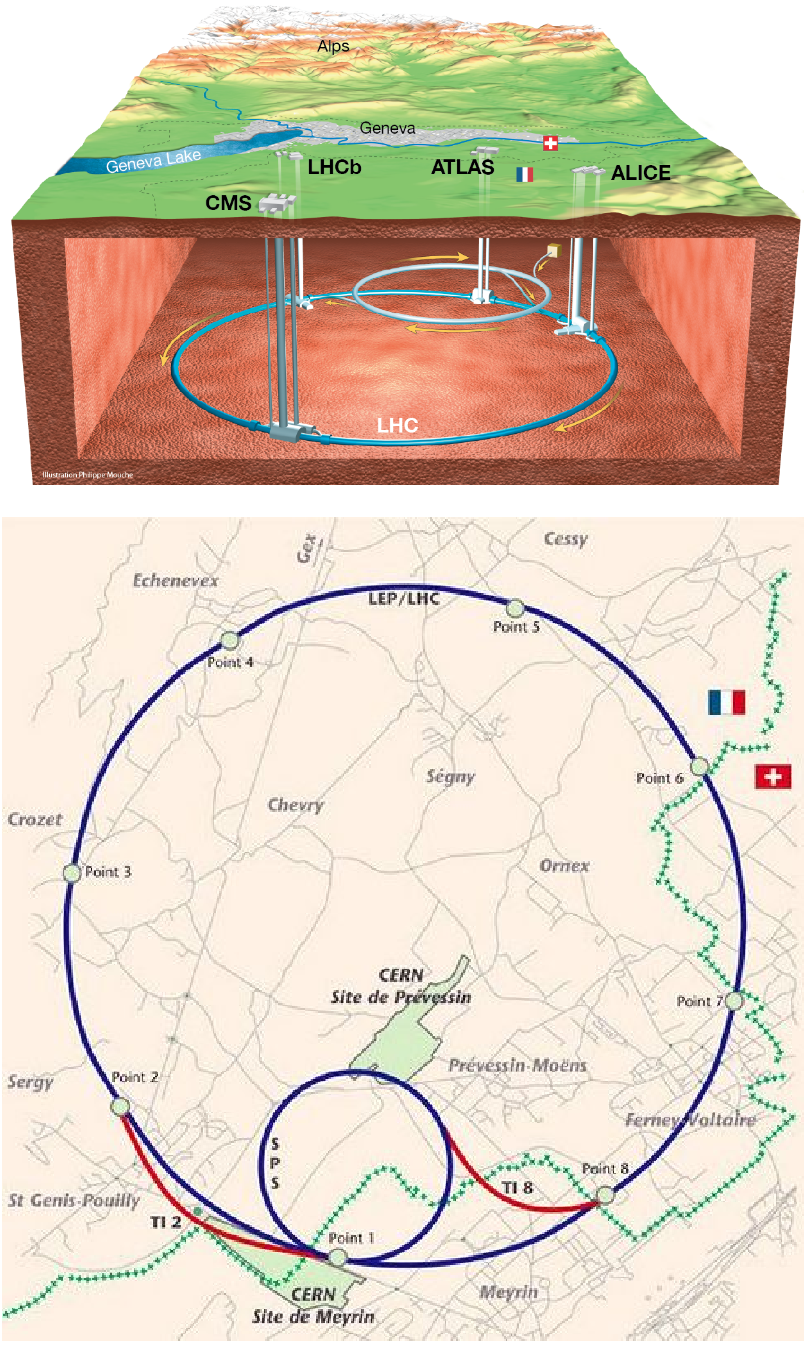


Figure 3.1: The layout of the LHC. The LHC is buried beneath the border of Switzerland and France near Geneva.

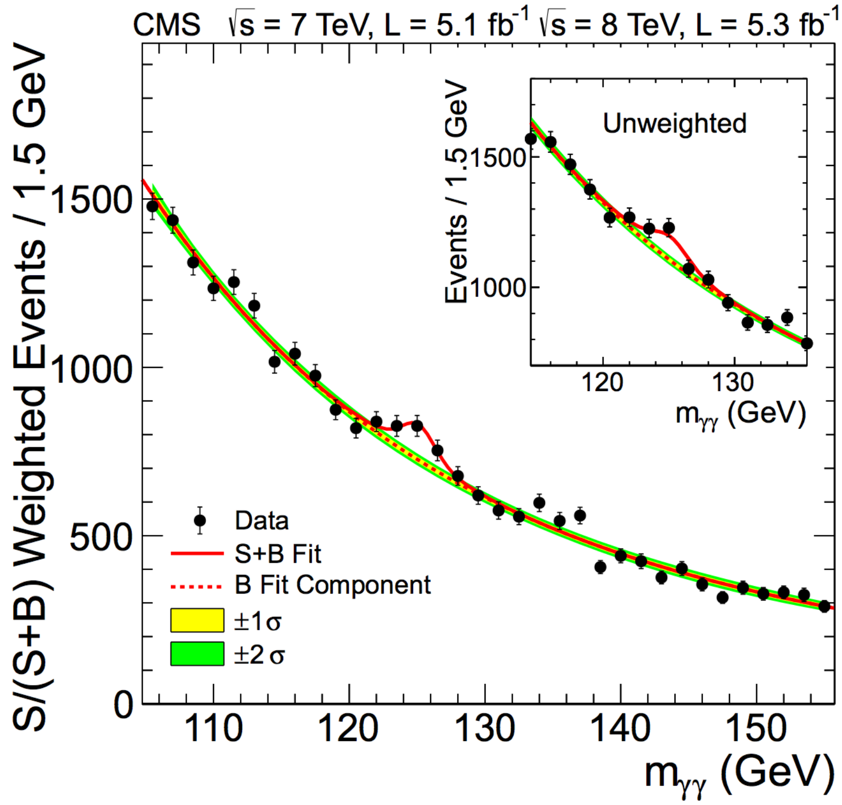


Figure 3.2: Discovery of the Higgs boson at CERN. Using early data from the LHC, the CMS and ATLAS experiments reported the discovery of a new boson in 2012 [2]. This figure shows a weighted distribution of the diphoton invariant mass. The signature of the Higgs boson is the bump 125 GeV above the background.

are as follows: protons from ionized hydrogen gas are accelerated to 50 MeV by Linac 2, Linac 2 feeds the protons into the Proton Synchrotron Booster where the first proton bunches are formed and boosts the protons to 1.4 GeV, in turn, the Proton Synchrotron (PS) boosts them to 25 GeV, at this point the Super Proton Synchrotron (SPS) boosts them to 450 GeV before injection into the LHC as shown in Fig. 3.3.

The circular accelerators work in three steps. In the first step, proton bunches are injected into the ring. In the second step, proton bunch shapes and separations are refined, and the accelerator raises the proton energy. The last step is to pass the proton bunches to the next accelerator or, in the case of the LHC, cause them to collide at one of the designated collision points. When the LHC is performing

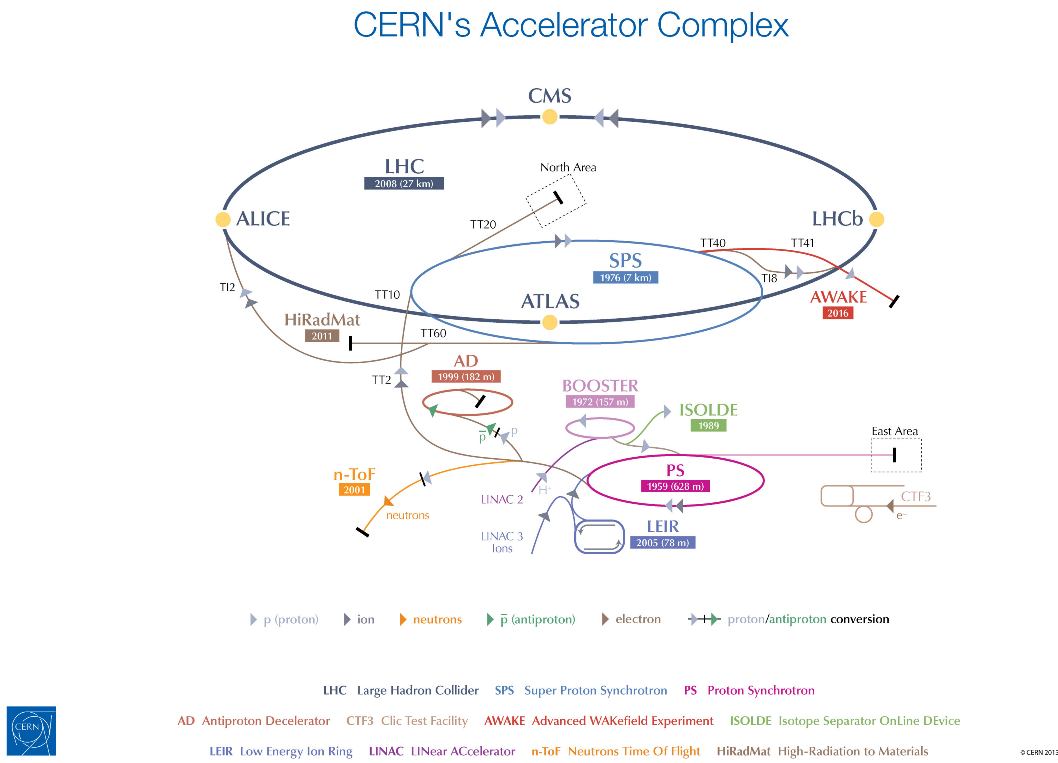


Figure 3.3: The CERN accelerator complex. Protons are accelerated by Linac 2, Proton Synchrotron Booster, Proton Synchrotron, and Super Proton Synchrotron before injection into the LHC.

nominally, each proton bunch contains about 1×10^{11} protons. Each bunch is spaced 25 ns apart, giving a collision frequency of 40 MHz. The beams within the LHC can be used for up to 10 hours before the number of protons in each bunch drops below a useful amount.

3.1.2 Collisions at the LHC

Collisions occur at four points around the LHC. The primary experiments at each collision point are CMS (Compact Muon Solenoid), ATLAS (A Toroidal LHC ApparatuS) [10], LHCb (LHC-beauty) [11], and ALICE (A Large Ion Collider Experiment) [12]. CMS and ATLAS are two large general purpose detectors. LHCb is specialized to detect and measure B mesons. ALICE is specialized for the characterization of heavy ion collisions, and studies the properties of of quark-gluon plasmas and other strong interaction processes. The collision points are widely spaced around the LHC. CMS is located in Cessy, France; ATLAS is located in Meyrin, Switzerland; LHCb is located in Ferney-Voltaire, France.

3.1.3 Particle Production at the LHC

Proton-proton collisions occur when bunches of protons on the order of 1×10^{11} are crossed at the afore mentioned collision points. Most of the protons in the bunch are not affected by the bunch crossing. About 20–60 inelastic collisions occur per bunch crossing in the nominal operation of the LHC during Run 2, as shown in Fig. 3.4 [13]. In an inelastic collision, constituents of the proton (partons), including both quarks and gluons, will interact. The interacting partons carry varying portions of the proton's momentum.

3.2 The CMS Detector

The CMS detector is designed to discover and measure the properties of the Higgs boson, more precisely measure standard model processes, and detect beyond standard model physics phenomena. As seen in Fig. 3.5, CMS is composed of many detectors working together. The signatures of these processes include electrons, muons, photons, and hadronic jets with a vary wide range of energies. The multiple subdetectors of CMS allow identification and measurement of these objects. Since operations began in 2010, the CMS detector has collected about $150\,\mathrm{fb}^{-1}$ of data at various center of mass energies as shown in Fig. 3.6. The LHC conducts several different collision programs including proton-proton, lead-lead, and proton-lead collisions. By far, most of the data collected has been proton-proton collisions with $6.1\,\mathrm{fb}^{-1}$ collected at a center of mass energy of $7\,\mathrm{TeV}$, $23.3\,\mathrm{fb}^{-1}$ at $8\,\mathrm{TeV}$, and $113\,\mathrm{fb}^{-1}$ at $13\,\mathrm{TeV}$.

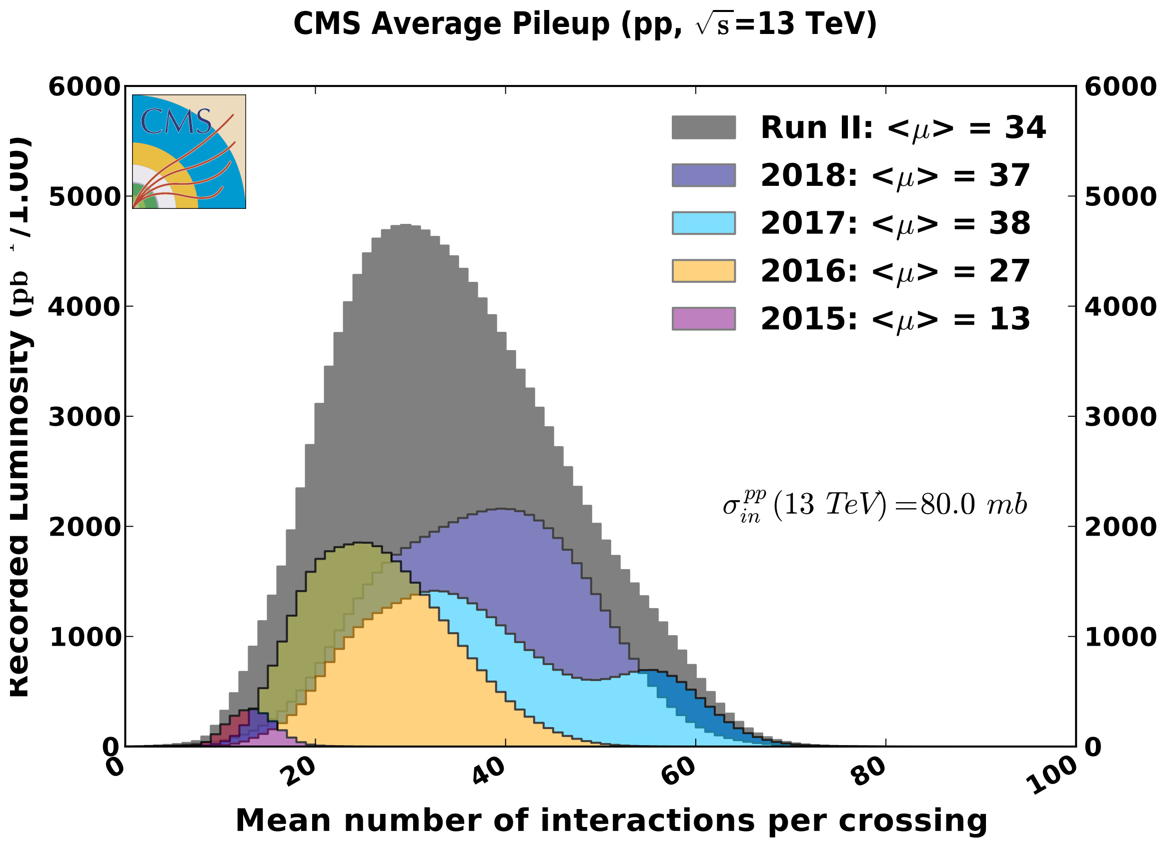


Figure 3.4: Distribution of the mean number of interactions per crossing for proton-proton collisions as measured by CMS during the LHC run 2.

The complex systems of CMS together collect collision information, select and preserve events of interest, and reconstruct the particles produced in the events. Overall the detector is roughly cylindrical with a length of 21.6 m and a diameter of 14.6 m. The cylindrical structure of the experiment naturally leads to the systems being roughly divided into a central portion called the “barrel” and the cylinder caps called “endcaps”. The total weight of the detector is about 12.500 t. A complete description of the original experiment design is available [15].

A central feature of CMS is the superconducting solenoid. It has an inner diameter of 6 m and is 12.5 m long. Within its interior, the solenoid produces a magnetic field with a nominal strength of 3.8 T. Three of the subdetectors are located within

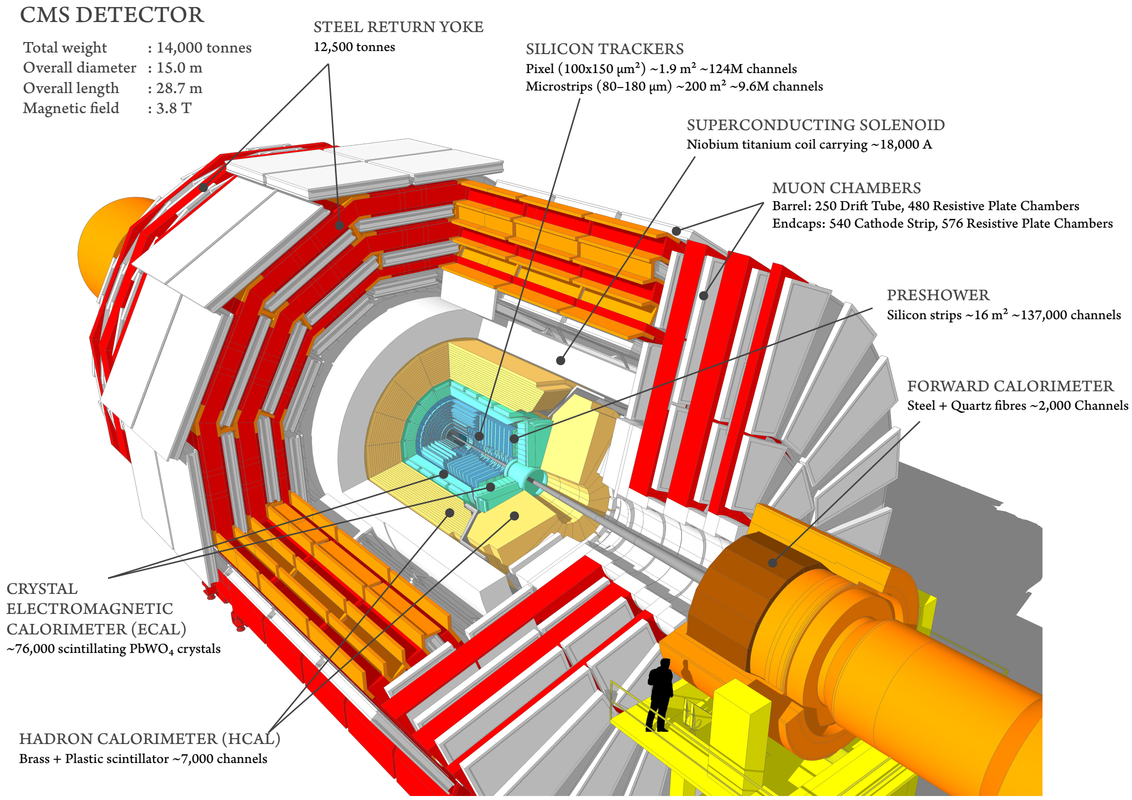


Figure 3.5: The Compact Muon Solenoid (CMS). This cutaway view shows the subsystems of the detector [14].

the volume of the solenoid itself: the pixel and strip silicon tracker, the barrel and endcaps of the Electromagnetic Calorimeter (ECAL), and the barrel and endcaps of the Hadron Calorimeter (HCAL).

In addition to the silicon tracker, ECAL, and HCAL within the volume of the solenoid, there are also muon detectors placed in the iron return yoke and an outer hadron calorimeter, as shown in Fig. 3.7. There are also forward calorimeters placed in the forward region just after the endcaps.

3.2.1 Particle Detection

There are three types of interactions that the subsystems use to detect and measure particles produced in high energy collisions.

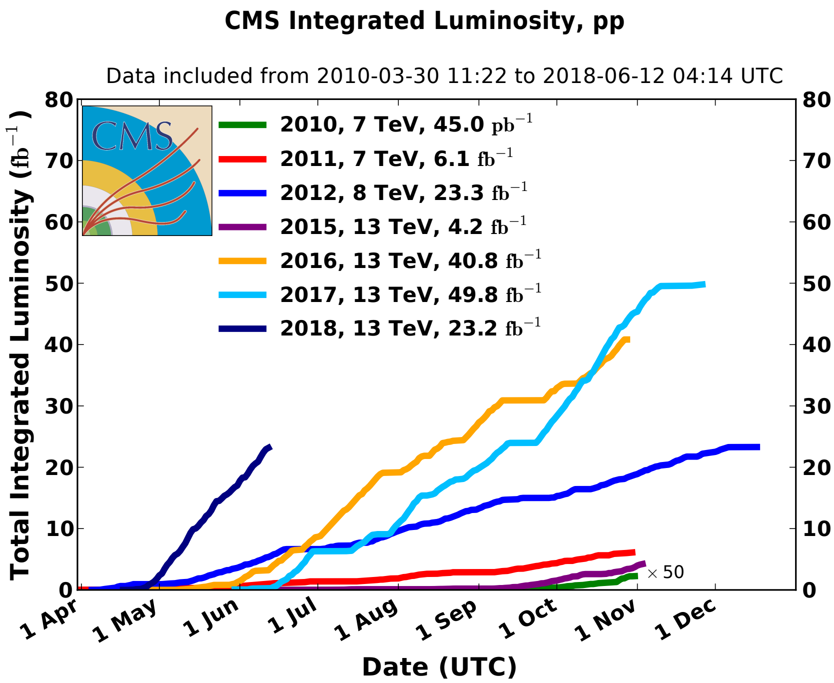


Figure 3.6: Integrated luminosity accumulated in each year of the LHC operation.

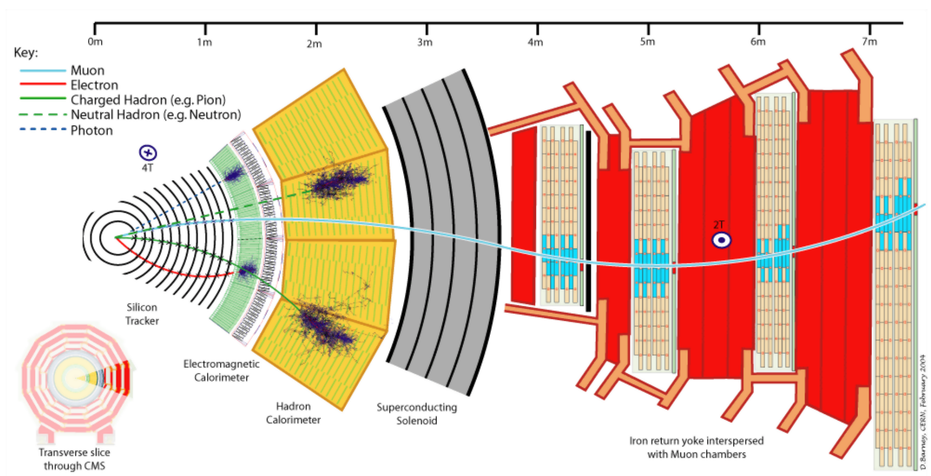


Figure 3.7: Transverse slice of CMS. The subdetectors and the interactions of collision products with the detectors are illustrated [16].

- Electrical detection of ionization signals: Subdetectors that use this type of interaction record when a passing charged particle ionizes some part of the material and the resulting ionization directly leads to an electrical signal. These subdetectors include the pixel and strip silicon detectors of the inner tracker, and the Drift Tubes (DT), Cathode Strip Chambers (CSC), and the Resistive Plate Chambers (RPC) of the muon detection system.

- Collection of light from scintillating material: Subdetectors that use this type of interaction use a scintillator which produces light in proportion to the energy given up by passing particles. These subdetectors include the electromagnetic calorimeter and the barrel, endcap, and outer hadron calorimeters. An absorbing material is used to slow and fragment particles to enhance interaction with the scintillator, and to pack more radiation lengths within a constrained volume.
- Collection of Cherenkov radiation: A subdetector that uses this type of interaction collects and measures the light given off as particles pass through a material at a speed greater than the local bulk speed of light. The forward hadron calorimeter is an example of this type of detector.

3.2.2 Inner Silicon Tracker

The silicon tracker is made of both pixel and strip elements as depicted in Fig. 3.8. It is the subdetector that lies closest to the interaction point. As charged particles pass through the tracker, it records their passage. The resulting cloud of hits is then analyzed to produce particle tracks and identify vertices. Reconstruction of the particle tracks and vertices is presented in Section 4.2.1. The inner tracker largely detects electrons, muons, and charged hadrons. Details of the inner tracker performance can be found in Ref. [17].

Located within the solenoid, the tracker is used to identify and measure the momentum of charged particles. It is also used to identify secondary vertices produced by the decay of long lived particles. Readout of the silicon tracker takes 6 clock cycles of 25 ns [19], and is not available to the low-level trigger (Level-1 trigger). Recording the data is initiated by the Level-1 trigger and is available to and used by the High Level Trigger (HLT).

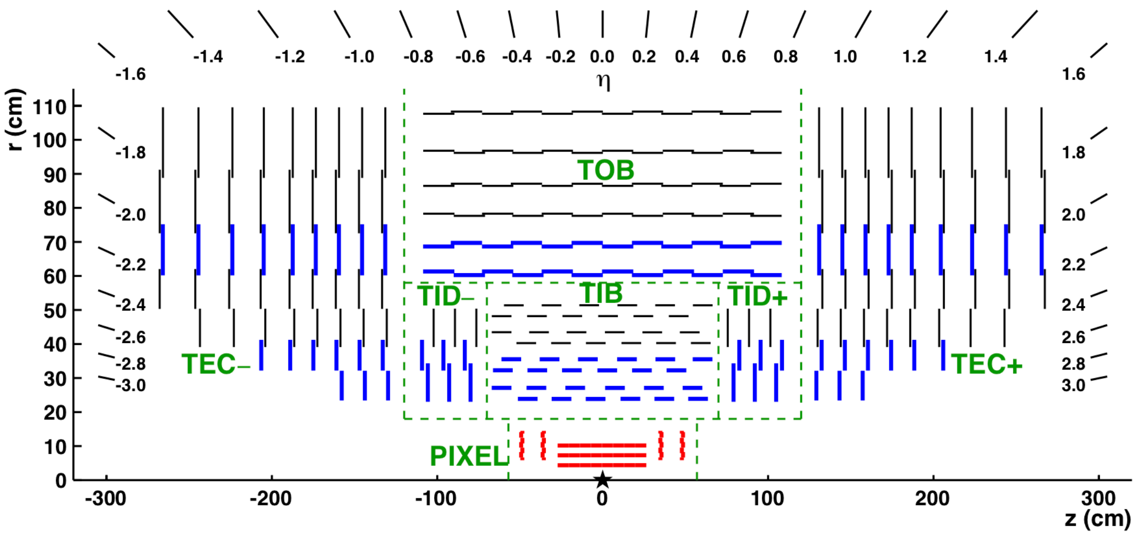


Figure 3.8: Schematic view of the inner silicon tracker. The vertical axis is the radial position from the beam line, and the horizontal axis is along the beam line and measures the distance from the interaction point. The cylindrical structure of the detector is found by rotating this view around the horizontal axis. The silicon pixel elements are shown in red nearest the interaction point. The silicon strip elements are shown in blue and black. [18]

The silicon pixel elements are three barrel layers (BPix) and two endcap disks (FPix) as seen in Figs. 3.8 and 3.9. These elements are found closest to the interaction point, providing coverage in the range $-2.5 < \eta < 2.5$. The three barrel layers are each 53 cm long and located 4.4 cm, 7.3 cm, and 10.2 cm from the beam line. The endcap disks are placed at $z = \pm 34.5$ cm and $z = \pm 46.5$ cm. They have an inner radius of 6 cm, and an outer radius of 15 cm.

The silicon strip elements are located further from the interaction point than the silicon pixel elements. As seen in Figs. 3.8 and 3.10, the strip elements are divided into four subsystems: the tracker inner barrel (TIB), tracker inner disks (TID+ and TID-), tracker outer barrel (TOB), and tracker endcaps (TEC+ and TEC-). The TIB and TIDs are in the volume immediately surrounding the pixel detector and extend out to a radius of 55 cm. The TIB is arranged into 4 barrel layers. Each TID consists of three discs. The TOB is made up of six barrel layers and immediately surrounds

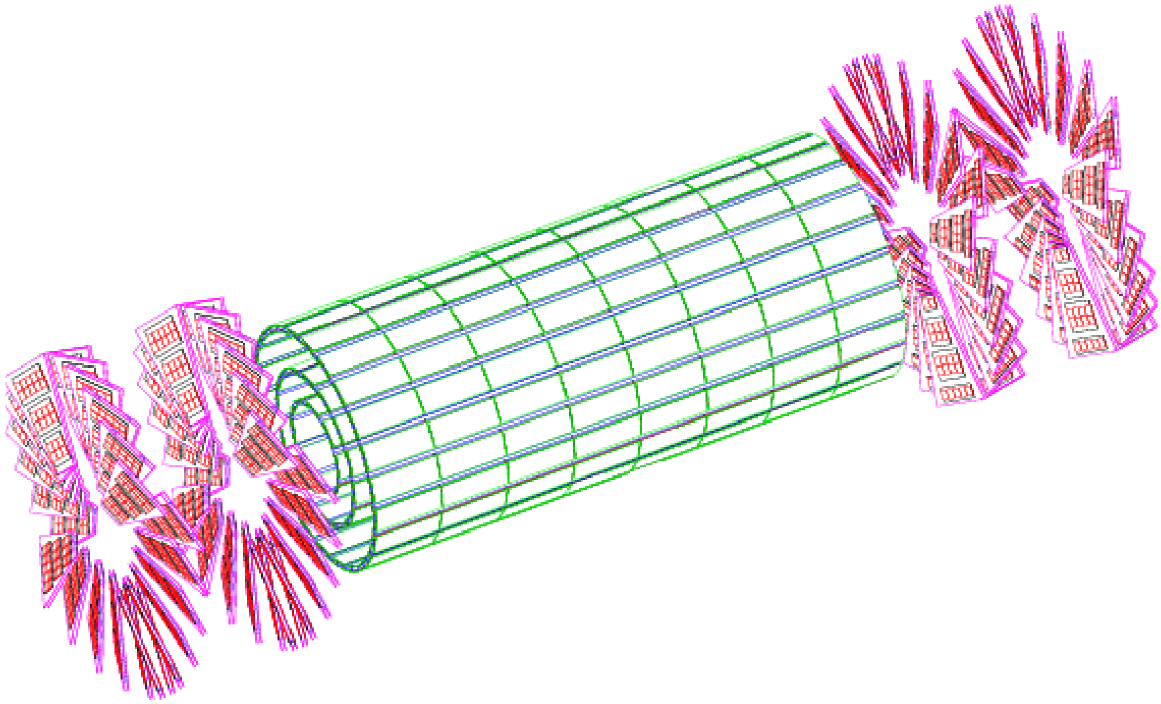


Figure 3.9: A view of the silicon pixel elements. The silicon detectors are arranged into three barrel layers and two discs at each end.

the TIB and TIDs; extending to a radius of 116 cm. TEC+ and TEC- cover the range 124 cm $< |z| <$ 282 cm. Each TEC consists of nine discs with outer radii up to 113.5 cm from the beam line.

The CMS silicon tracker is the largest semiconductor silicon detector ever constructed [20]. The pixel cells are each 100 µm by 150 µm. The BPix (FPix) contains 48 (18) million pixels providing a silicon semiconductor detection area of 0.78 m^2 (0.28 m^2). The strip trackers have a total of 9.3 million silicon strip elements providing a silicon semiconductor detection area of 198 m^2.

3.2.3 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) is made of lead tungstate ($PbWO_4$) crystals which serve as absorber, scintillator, and conductor of scintillation light directly to bonded photodetectors [21, 22]. There is a small lead and silicon preshower

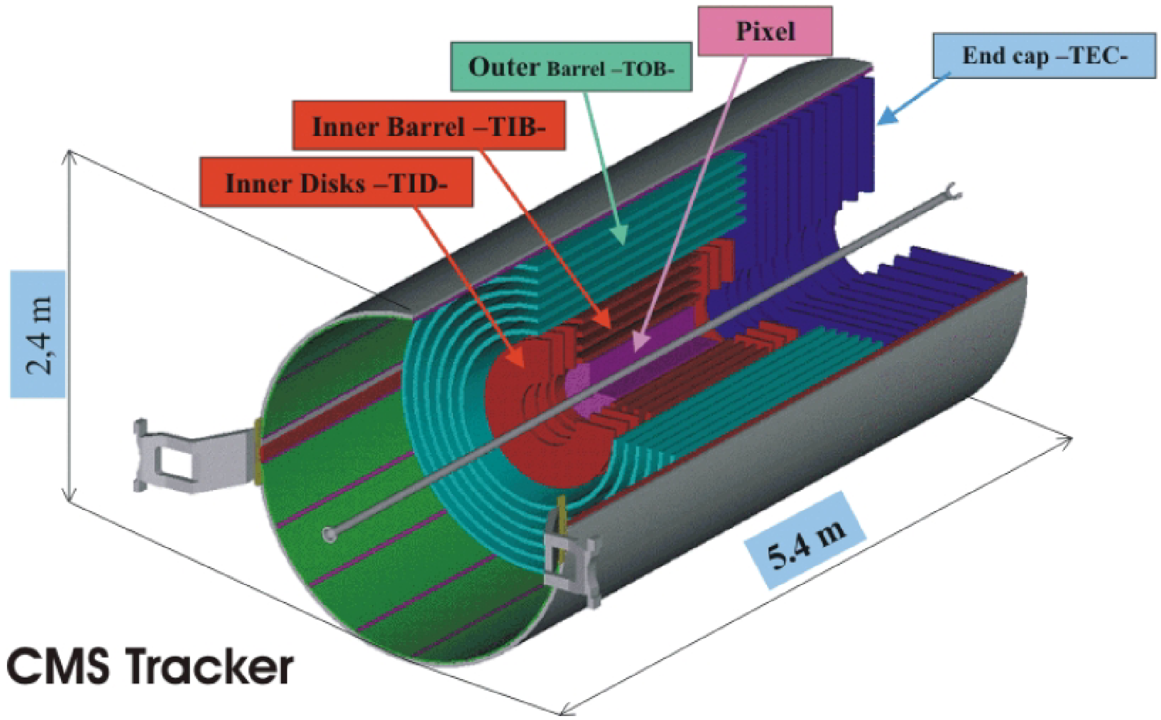


Figure 3.10: A view of the silicon strip elements. The silicon strips are organized into both barrel and endcap configurations.

sampling detector (ES) in the endcap region. The organization of the calorimeter is shown in Fig. 3.11. The electromagnetic calorimeter can collect energy from the passage of any charged particle, but it is particularly useful by absorbing and measuring electrons and photons. The reconstruction of photons is described further in Section 4.3.3 and Ref. [23]. Reconstruction of electrons in Section 4.3.1 and Ref. [24]. The electromagnetic calorimeter is located within the solenoid between the silicon tracker and hadron calorimeter.

The electromagnetic calorimeter contains $\sim$ 75000 scintillating crystals divided into a barrel (EB) and two endcap (EE) regions. Together these regions cover the range $|\eta| < 3.0$. The location of ECAL close to the interaction point and necessarily constrained in volume by its location within the solenoid between the tracker and hadron calorimeter guided the choice of lead tungstate crystals. The crystals have a high radiation tolerance and are able to withstand the high radiation environment

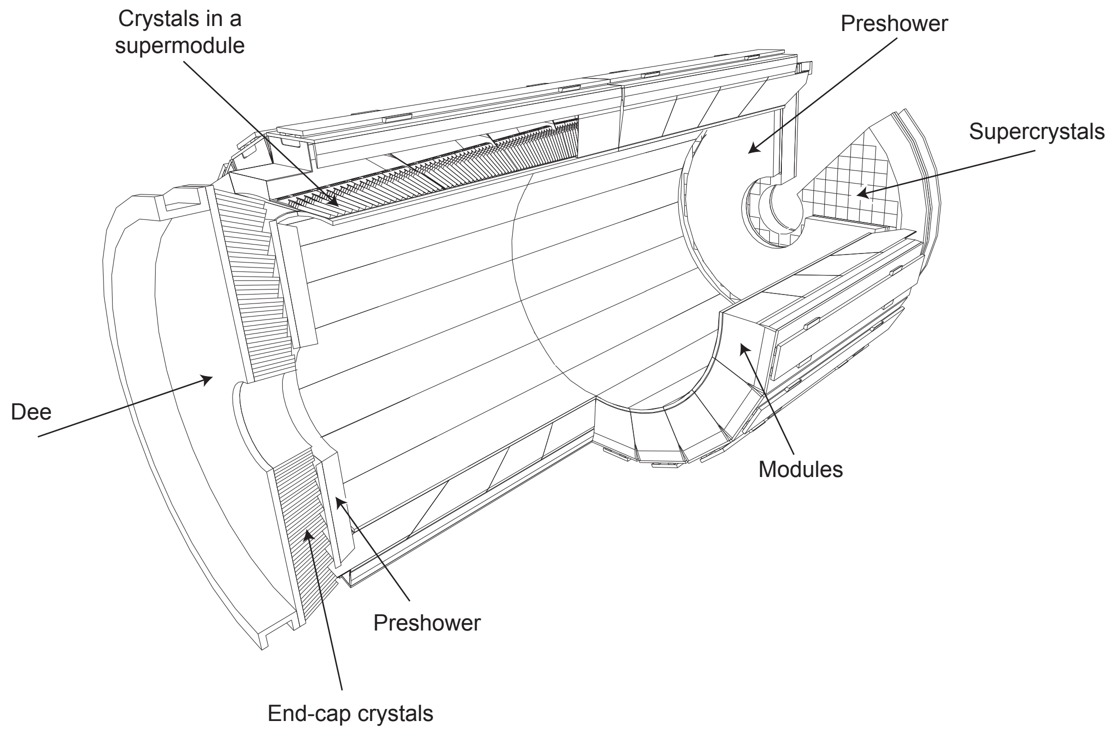


Figure 3.11: A cut away view of the Electromagnetic Calorimeter (ECAL). The lead tungstate ($PbWO_4$) crystals are arranged into barrel supermodules and endcap supercrystals.

especially at high η. Lead tungstate has a short radiation length ($X_0 = 0.85\,\mathrm{cm}$), which allows the complete absorption of electron and photon energies in a relatively small volume. The short relaxation time ($\sim 25\,\mathrm{ns}$) and small dimensions of the crystals provides a fast response time. The electromagnetic calorimeter induces and measures energy of particles that interact electromagnetically with it. Electrons and photons are completely absorbed by the ECAL; producing an electromagnetic shower. Electrons are absorbed because of the bremsstrahlung induced as the electrons pass near the lead and tungsten nuclei until they finally give up the last of their energy through ionization. Photons lose energy through electron pair production until they no longer have the energy to do so and give up the last of their energy through ionization. The scintillation light comes from the de-excitation of the ionized atoms in the crystal. The amount of scintillation light is proportional to the energy deposited in the crystals by the incident and absorbed particles. 1 MeV of deposited energy will result in the production of about 4.5 photoelectrons in both EE and EB.

The EB has an inner radius of 1.29 m and covers the range $|\eta| < 1.48$. It is made up of 61200 crystals constructed into 36 identical supermodules, each covering half of the barrel length. The crystals in the EB are 23 cm (27.1 radiation lengths) long and have a flat face towards the beam line of about 2.2 cm×2.2 cm. The crystals have a truncated pyramid design and aligned so that the seams between the crystals are not aligned with the trajectory of particles originating from the interaction point, preventing a particle from escaping undetected. EB provides 360-fold granularity in ϕ, and 85-fold granularity in η within the range $|\eta| < 1.48$.

The EEs extend the range of ECAL to $|\eta| < 3.0$. The crystals are arranged in x-y grid to form a circular shape. The individual crystals are 22 cm (25.8 radiation lengths) and have a front face about 2.86 cm×2.86 cm. Each endcap has 7324 crystals grouped into 5×5 structures called supercrystals.

The preshower detector (ES) is located in front of each EE and extends in the region $1.65 < |\eta| < 2.6$. The ES helps distinguish between high energy photons and neutral pions. The ES contains two active planes of silicon strip detectors oriented at right angles to one another provided both good positional accuracy and granularity.

ECAL extends to $|\eta| < 3.0$, but differentiating between photons and electrons is only possible in the region covered by the tracker ($|\eta| < 2.5$). Additional information about ECAL performance is available in Ref. [21].

3.2.4 Hadron Calorimeter

The hadron calorimeter (HCAL) [25, 26] of the CMS detector [9] at the Large Hadron Collider (LHC) is composed of four major sub-detectors: the barrel hadron calorimeter (HB) [27], the endcap hadron calorimeter (HE) [28], the forward calorimeter (HF) [29] and the outer hadron calorimeter (HO) [30]. Figure 3.12 shows how these detectors are arranged within CMS as a whole.

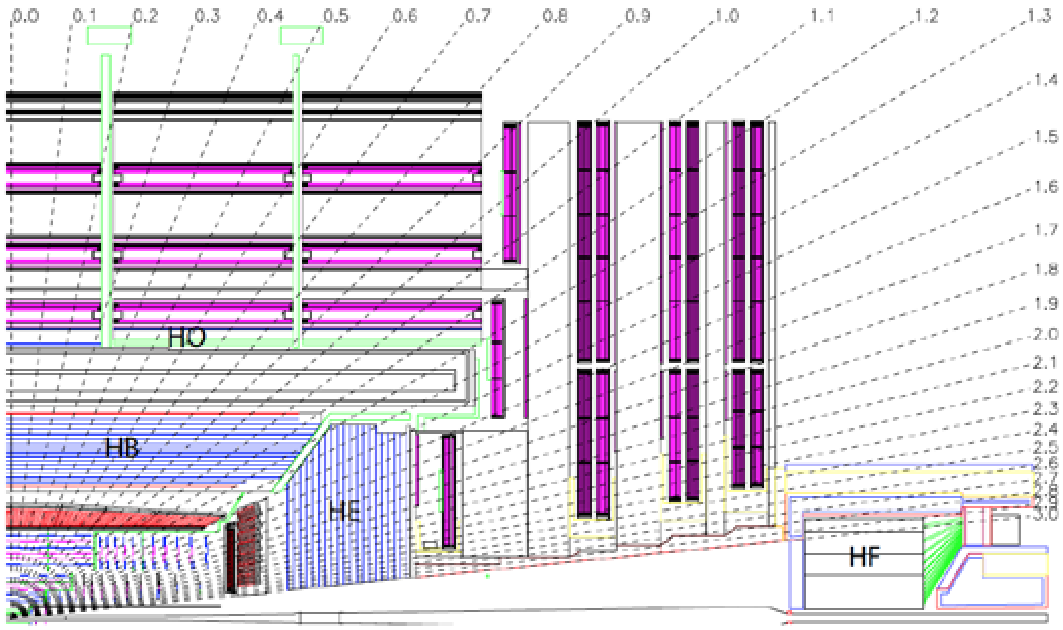


Figure 3.12: A schematic of the hadron calorimeter systems [9]. This view shows the hadron calorimeter detectors in relation to CMS as a whole. HCAL Barrel (HB) and HCAL Endcap (HE) are located within the volume of the solenoid. HCAL Outer (HO) is located in the barrel region of iron return yoke. HCAL Forward (HF) is located in the forward region beyond the muon endcaps.

There are two technologies used in the hadron calorimeter. HE, HB, and HO are brass and scintillator sampling calorimeters. The brass absorber induces particle showers, and the plastic scintillator produces light when the showering particles interact with it. The scintillator tiles are surrounded by a loop of wave-length shifting (WLS) fiber which collects the scintillation light. It is because of the frequency shifting that the light can be captured and carried away by the fiber, otherwise, the the angle of the light after is has entered through the sides of the fiber would be such that exit again through the fiber's sides. After the WLS fiber collects the scintillation light, clear fiber optic cable carries it to a photodetector. When CMS was first commissioned, hybrid photodiodes (HPDs) were used to measure the scintillator light. Before data were collected in 2018, silicon photomultipliers (SiPMs) replaced the

HPDs in the endcap hadron calorimeter. The HPDs in the barrel hadron calorimeter were replaced following the end of data taking in 2018.

The forward calorimeter uses a different technology. It is made from a large block of steel into which holes are drilled and threaded with quartz fiber. The steel helps induce showering. As particles pass through the quartz fibers, Cherenkov light is produced. The quartz fibers conduct the Cherenkov light to photomultiplier tubes.

HCAL cells are described by three parameters, the azimuthal angle, ϕ, the pseudorapidity, η, and a depth. See Fig. 3.13 for a description of the segmentation in the endcap and barrel. The figures includes the original segmentation, and the segmentations after the endcap and barrel updates. In the region, $|\eta| < 1.7$, the granularity of the HCAL cells is $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. In the region $1.7 < |\eta| < 3.0$, the granularity of the HCAL cells varies from $\Delta\eta \times \Delta\phi = 0.09 \times 0.174$ to $\Delta\eta \times \Delta\phi = 0.35 \times 0.174$.

Since the start of 13 TeV operation, upgrades have been planned and implemented to several hadron calorimeter subdetectors [31, 32]. These updates include improved photodetectors, electronics, and finer depth granularity.

3.2.4.1 HCAL Forward. HCAL forward is a sampling calorimeter with quartz fibers as the sampling medium encased within a steel absorber. Figure 3.14 is a schematic view of the HCAL forward assembly. The fibers are spaced 5 mm apart and the fibers associated with a particular $\eta \times \phi$ region are optically combined. As high energy particles pass through the quartz fibers, Cherenkov light is produced and then conducted by the fibers to PMTs. The Cherenkov light produced is proportional to particle flux passing through the quartz fibers. The steel block induces particle showers such that the energy of an incident particle is proportional to the number of the produced particles in the shower. There are two classes of quartz fiber in HF. They are called “long” and “short”. By comparing the the energy recorded in the short fibers to the energy recorded in the long fibers, a reasonable estimate of the contribution from electrons and photons vs. hadrons can be made. EE and HE only

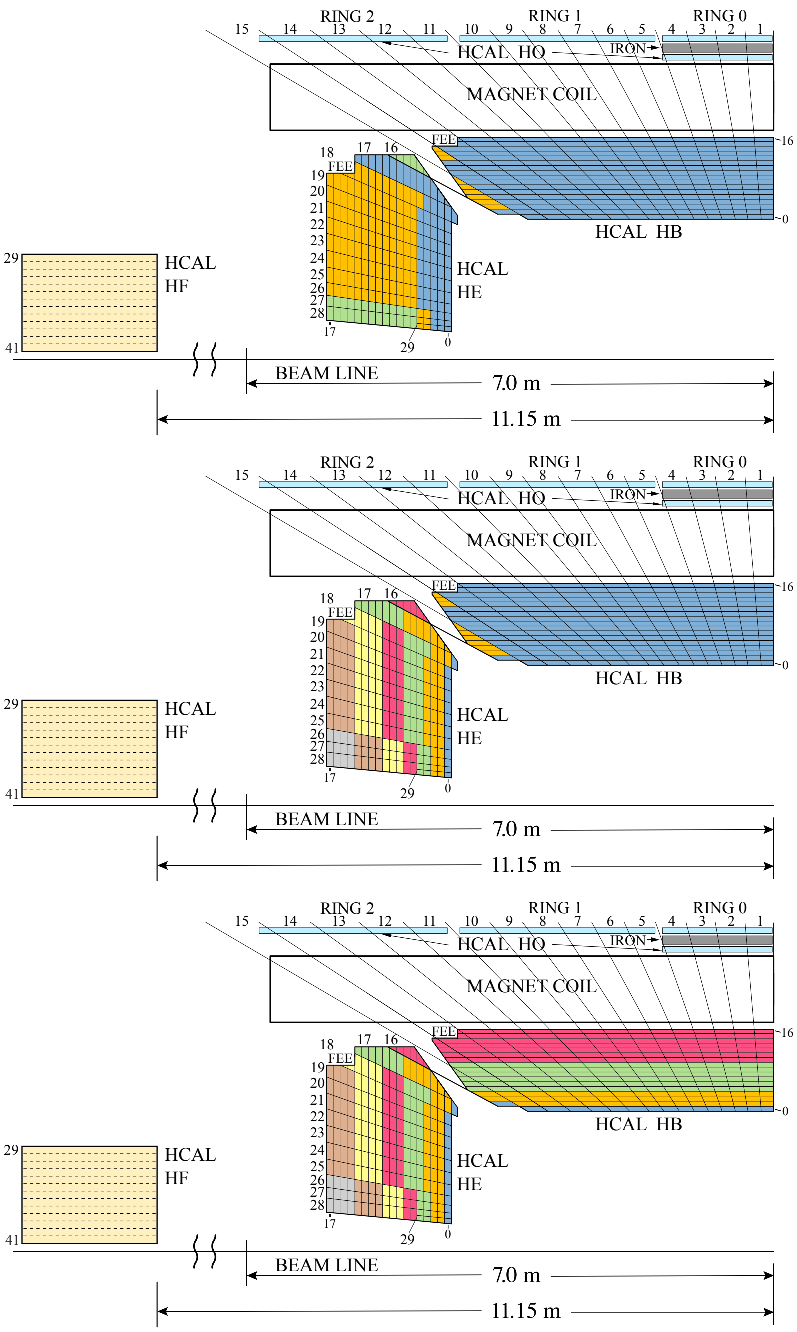


Figure 3.13: Hadron calorimeter segmentation. Individual scintillator tiles are combined into HCAL cells. The figures show segmentation of tiles into cells for the detector as originally configured, after the endcap upgrade, and including the proposed upgrade to the barrel.

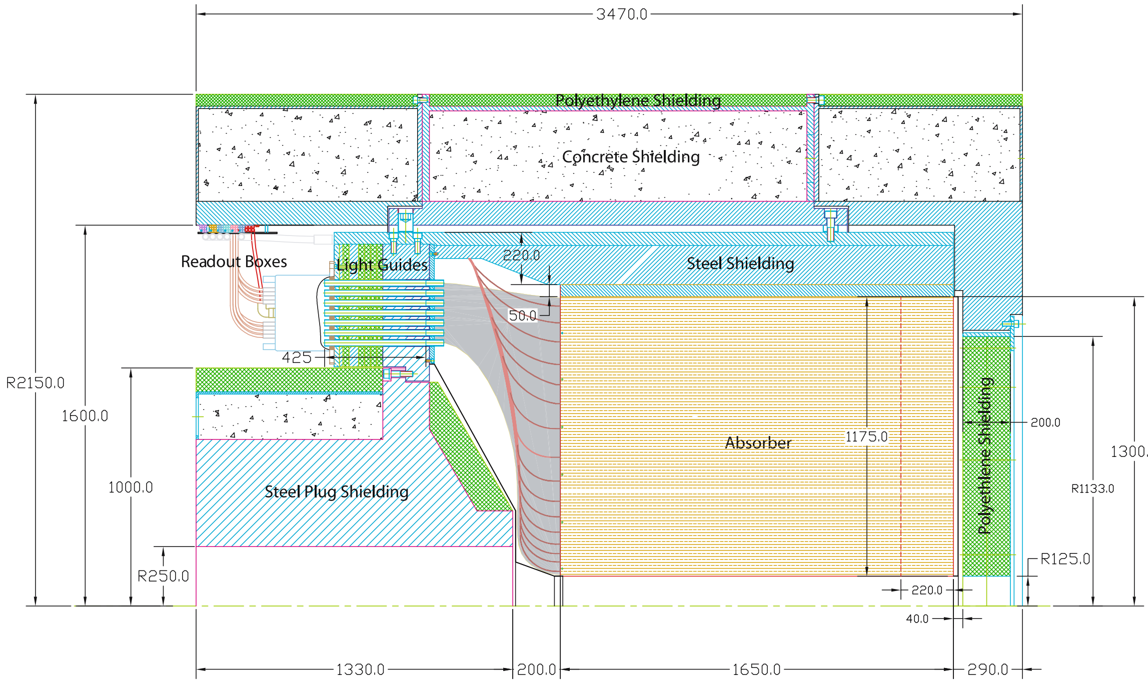


Figure 3.14: A cross sectional view of the HF. The steel absorber has an inner radius of 12.5 cm and an outer radius of 130 cm. The steel absorber extends along the beam line for 165 cm. Quartz fibers are threaded through the steel absorber, and passing particles will create Cherenkov light with the fibers. This light to is transmitted and recorded by PMTs. Note that in this diagram all of the given dimensions are in mm [29].

extends to $|\eta| < 3.0$, so HF, which covers $3.0 < |\eta| < 5.0$, is essential for measuring the total energy in the system, identifying transverse momentum imbalances, and identifying forward jets. The steel and quartz construction of the detector makes it resistant to the high radiation environment of the forward region.

HCAL forward was upgraded during the winter break of 2016–2017. This involved replacing the photomultipliers, the electronics, and improving the software.

It was discovered that one source of noise in the original HCAL forward was anomalous signals due to the direct impingement of particles on the photomultiplier tube. During Run 1, these anomalous signals could be effectively filtered both by looking at the energy of the surrounding cells and by looking at the timing of the signal with respect to the event. A large signal surrounded by low signal cells is likely an anomaly. Similarly, a signal created by a particle passing through the PMT will arrive sooner than the light from the quartz fibers. However, going forward these strategies would no longer be sufficient. With a higher number of collisions per event, isolation of an anomalous signal becomes harder to distinguish. The time between bunch crossing was decreased from 50 ns to 25 ns, and signals from adjacent events made rejection based on signal timing less useful. A replacement photomultiplier tube was selected that had two attributes that help reduce this noise. First, the replacement tube has a thinner window which reduces the number of particles that shower within the tube. Second, the replacement tube has its light collecting area divided into four anodes. An anomalous signal is likely to show up in only a single anode, and can be identified and corrected. In the current detector, the anodes are paired together, producing two channels. The reconstruction software was suitably modified to account for two channels of information covering the same HF region.

The electronics were upgraded to handle the greater number of channels, and the previous version of the charge integration and encoding chip (QIE) was replaced with a new version. All of the original HCAL subsystems originally used QIE8 chips. The QIE10 chip can encode charge information with greater granularity and includes intra-bunch crossing timing information. This information is helpful to accurately measure the increased light from higher luminosities and disentangle signal from overlapping events which had their temporal separation reduced from 50 ns to 25 ns. Not only does the new timing information help with signal reconstruction, it provides a further handle to identify detector noise.

3.2.4.2 HCAL Endcap. HCAL endcap is depicted in Fig. 3.15. HCAL endcap was upgraded during the year end technical shutdown between 2017 and 2018. This was preceded the previous year by an upgrade to only one of the HE modules (HEP17). The endcap was originally planned to be upgraded in LS2 starting in 2019; however, the upgrade was advanced because radiation damage to the scintillator, especially in

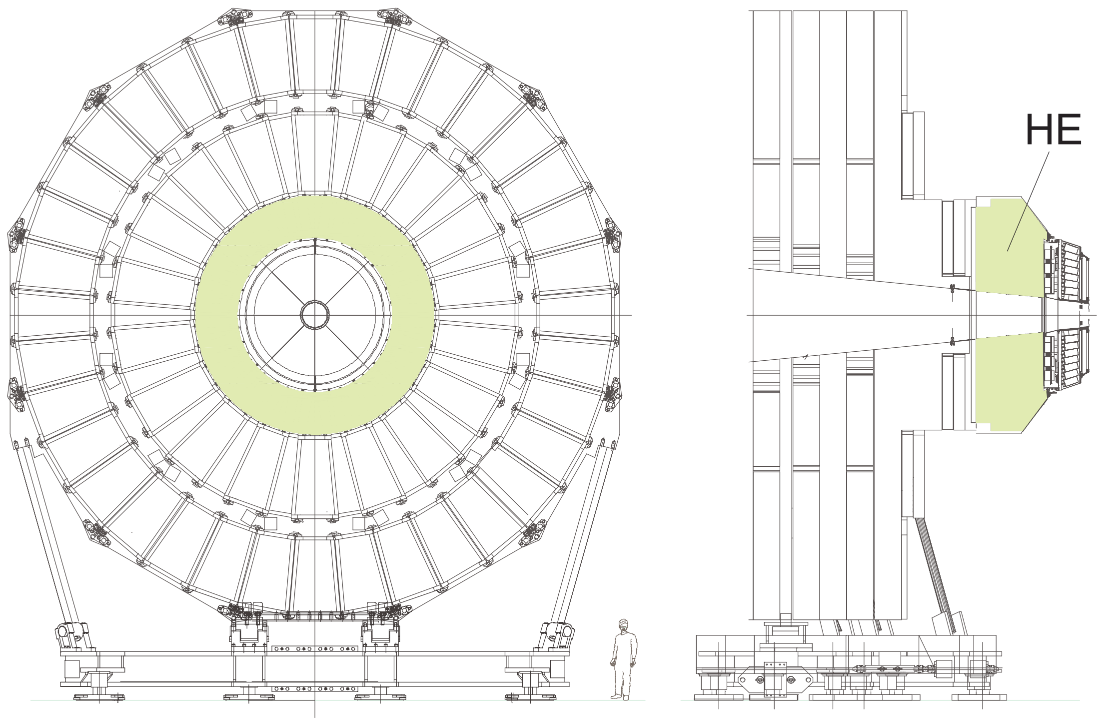


Figure 3.15: A view of the HCAL Endcap (HE) mounted on the endcap iron yoke. For the purposes of construction and maintenance, CMS is divided into several rings which can be pulled apart. The outer most rings support the endcap iron yoke, endcap muon chambers, the HCAL Endcap, and the ECAL Endcap.

the high $|\eta|$ region was dramatically reducing the energy resolution in the endcap, leading to fears that the data collected would dip below acceptable performance before the upgrade planned during LS2 could occur. Radiation damages the scintillator causing it to "darken" i.e. attenuate the amount of light coming from the scintillator. Since replacing the damaged scintillator is not feasible, the damage is mitigated by increasing the segmentation of the calorimeter depths as seen in Fig. 3.13. Previously the layers of scintillator had been optical connected into seven depths. The upgrade changes this so that the scintillator layers are optically connected into up to seven depths. The inner most layers are the most heavily damaged so, by increasing the depth granularity, the decrease in energy resolution can be confined to a smaller volume of the detector.

3.2.4.3 HCAL Barrel. HCAL barrel is cylindrically symmetrical and is located between EB and the solenoid. Figure 3.16 shows the the construction of an HCAL barrel wedge. It has an inner radius of 1.77 m and an outer radius of 2.95 m. In order to fit within the solenoid, a large proportion of HB is devoted to the brass absorber so as to maximize the number of interaction lengths within HB. There is a layer of scintillator placed before the first absorber of HB. This layer-0 is design to help identify earlier forming particle showers. Likewise there is a last layer of scintillator near the outside of HB, this layer-16 is designed to help identify late developing showers and characterize the spill of particle showers beyond HB. There are 15 layers interleaved within the brass absorber. In its original configuration, except for a couple of towers at higher η, all HB layers are optically combined into a single logical depth. During the upgrades happening during LS2, it will be reconfigured to have the tiles optically connected into four depths, the HPDs will be replaced with SiPMs, and among other hardware changes, the QIE8 chips will be replaced by QIE11 chips that provide greater granularity and additional timing information.

3.2.4.4 HCAL Outer. HCAL Outer (HO) is a complement to HB. As depicted in Fig. 3.17, HO extends the number of interaction lengths within the calorimeter volume. It ensures that high energy showers are sampled and measured beyond the HB. HO extends in the region $|\eta| < 1.4$. In the most central region there are two HO layers located at a radii 3.82 m and 4.07 m. In the farther barrel region there is a single HO layer at a radius of 4.07 m. HO originally used HPDs, but was the first system of the hadron calorimeter to switch to SiPMs.

3.2.5 Muon Detectors

Muons are one of the cleanest signals seen in CMS. They can be cleanly distinguished from other particles. With the magnetic field produced by the solenoid, their momenta can be precisely measured.

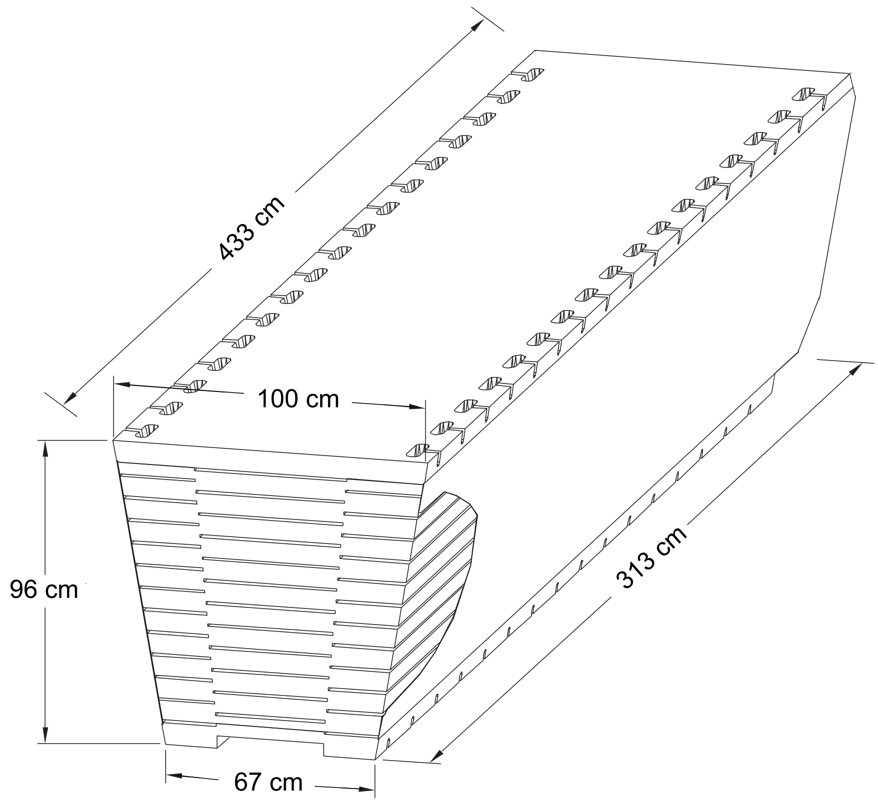


Figure 3.16: A schematic of an HB wedge showing the alternating layers of thick brass absorber and the thinner layers of active scintillator. HB is composed of 36 of these wedges in each of two half barrels. [27]

Muon detectors are embedded within the solenoid's iron return yoke as seen in Fig. 3.18 and are organized into muon barrel (MB) and muon endcap (ME) regions. Muons easily penetrate the elements of the detector, but their passage through a gas can be detected by the ionization they effect. The muon detectors are used in conjunction with the inner silicon tracker to identify and measure muons produced in particle collisions. Because cosmic ray muons can make it down to the detector cavern and through the detector, it is important to reconstruct a muon's track, and remove from analyses muons that do not originate from the interaction region. Such cosmic ray muons can be used to calibrate the alignment of the detector. Reconstruction of muons is further discussed in Section 4.3.1.

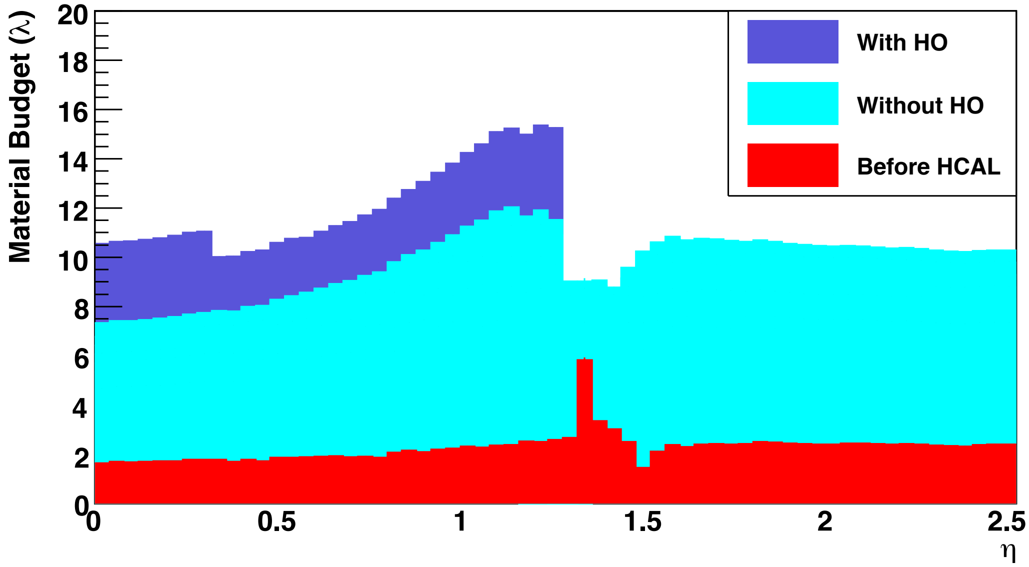


Figure 3.17: The number of nuclear interaction lengths provided by HB and HO as a function of η. As η increases, the amount of calorimeter material traversed by a particle increases. HO complements HB and helps to better contain particle showers by increasing the number of particles lengths contained within the calorimeter volume. The effect of the two HO layers in the central region is particularly noticeable.

Three types of gas ionization chambers are used for muon detection. Drift Tube chambers (DTs) measure the drift time of electrons produced by an ionization event to an anode wire to determine the position of the ionization event within the chamber. Cathode Strip Chambers (CSCs) consist of arrays of anode and cathode wires arranged at right angles to each other. Ionization of the gas within the chamber will cause electrons to move to the closest anode wire, and the ionized gas to the closest cathode wire. This provides a very accurate measurement of the location of the ionization event. Resistive Plate Chambers (RPCs) are double-gap chambers that consist of two parallel plates of plastic. The plates are oppositely charged and produce the electric field produces an electron avalanche when an ionization event occurs. The electrons are detected by metallic strips attached to the outside surfaces of the plastic plates. The RPCs provide very good time resolution and are paired with the DTs and CSCs.

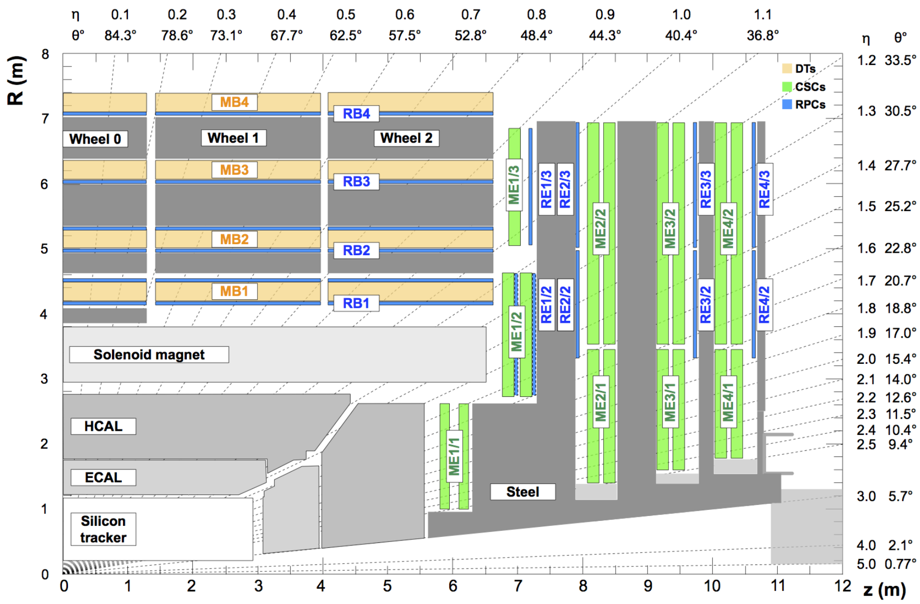


Figure 3.18: A schematic of the muon detectors. The detectors are embedded in the steel return yoke. Drift Tubes (DTs) and Resistive Plate Chambers (RPCs) are located in the barrel region. Cathode Strip Chambers (CSCs) and RPCs are located in the endcap regions. [33]

Within the MB, DTs paired with RPCs are used. DTs are appropriate for the environment of the MB region: a uniform magnetic field, low induced radioactivity, and lower muon flux. The MB region extends over the η range, $-1.2 < \eta < 1.2$.

Within the ME, CSCs paired with RPCs are used. CSCs are appropriate for the environment of the ME region: a non-uniform magnetic field, high induced radioactivity, and a higher muon flux. The CSCs extend the muon detection region to $-2.4 < \eta < 2.4$. The RPCs are only positioned in the region $-1.6 < \eta < 1.6$.

Using data collected during 2015 and 2016, it was found that the muon system provided a reconstructed hit spatial resolution of $\approx 50\,\mu\text{m}$–$300\,\mu\text{m}$, muon timing resolution $\approx 1.4\,\text{ns}$, and muon reconstruction and identification efficiency $> 96\%$. Further information about the performance of the muon system can be found in Ref. [33].

3.2.6 *Trigger and Data Acquisition System*

Collisions occur at a rate of 40 MHz. If data were saved for every collision, the resulting data rate would be about $12\,\mathrm{TiB\,s^{-1}}$. Not only is recorded data at such a high rate technologically impossible, it is also unneeded. Events of interest are identified and recorded thorough a two level triggering system. [34]. The first layer, the Level-1 (L1) trigger, is implemented in hardware and must decide within 4 µs if the event should be passed on to the the second layer, the HLT. The L1 trigger reduces LHC event rate of 40 MHz to an event rate of 100 kHz passed to the HLT. The second layer, the HLT, is implemented as software on a large data farm. The HLT must make a decision within 175 ms per event, and the selected events are recorded to the data storage at an average rate of 400 Hz.

The L1 trigger does not have full event information available to it. The calorimeters pass slightly simplified information to the L1 trigger which is used to construct estimates of electrons/photons, jets, total transverse momentum, and missing transverse momentum. The muon systems pass information to the L1 trigger which is used to construct muon candidates. The qualities of the event thus estimated can be combined in up to 128 ways in order to identify events of particular interest. Among such conditions are events with large missing transverse momentum, events with isolated muons, or events with a number of jets that pass a p_t a certain threshold. Figure 3.19 is a block diagram that shows how information from the calorimeters and muon systems are collected, processed, and eventually acted upon.

The HLT receives the full information available to the event, and performs event reconstruction comparable to the offline event reconstruction, albeit with tight timing restrictions. With the tracker information available and the full information of the calorimeters, physics objects such as electrons, photons, muons, and jets are reconstructed and made available for trigger evaluation. All events are uniquely identified

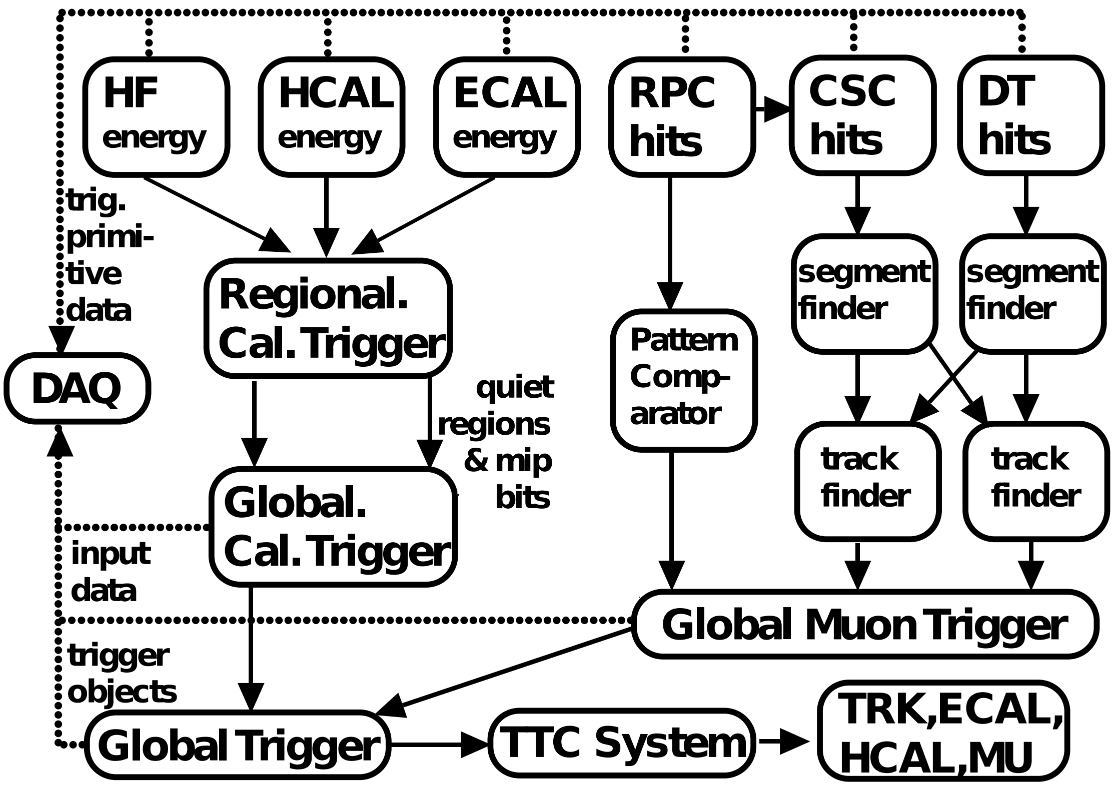


Figure 3.19: Overview of the L1 Trigger. Only the forward, barrel, and endcap hadron calorimeters; the barrel and endcap electromagnetic calorimeters; and the RPCs, DTs, and CSCs are available to the L1 trigger. L1 must decide within 4 µs if an event is to be readout and sent to the HLT. [34]

by a run number, lumi section, and event number. Selected and saved events will be referenced by this identification throughout the processing and analysis processes.

CHAPTER FOUR

Event Simulation and Reconstruction

Chapter Three was devoted to the physical construction of the experiment, its detectors, and the information saved from an event. In this chapter I will discuss the processing that takes place to convert the signal from the detector into a reconstructed event. We need a large number of simulated events representing both standard model processes and processes predicted in possible extensions to the standard model. I will discuss this event simulation first, up to the simulated detector response at which point simulated events and real events are reconstructed using the same software.

4.1 Event Simulation

Simulated events are used to validate reconstruction, motivate study design by simulating events from signal models, and aid in standard model background estimation. The process of producing a simulated event takes several steps: event generation, detector simulation, digitization, and feed forward into the standard reconstruction chain.

4.1.1 Event Generation

An event is not simulated by a single program. The needs of the simulation can be divided into different physical interactions and different time and length scales. Monte Carlo simulations are run to simulate the hard scattering processes expected in LHC collisions. The final state particles of these first simulations are then run through additional software packages that simulate particle shower/hadronization and decays of long lived products.

There are a variety of software packages available to simulate these processes. The development of the packages and the tuning of their configuration to match the observed results in data is a continuous and ongoing effort. For the analysis presented in this thesis, the inelastic scattering processes are typically simulated by MadGraph5_aMC@NLO [35, 36] or POWHEG [37–40] and the simulation of particle showers/hadronization is performed by the PYTHIA v8 [41] program.

4.1.2 Detector Simulation

The simulated inelastic collisions and the subsequent particle shower, hadronization, and decay of long lived products described in the previous section will produce a list of stable particles. Here stable means that the particle lifetimes are long compared to the time and distance scales of the detector and detector response. The next step is to simulate the detector response as these particles interact with the detector. This requires both a simulation of the particle interaction with the physical structure of the detector as well as the physical processes that elicit a response from the detection elements. Current particle physics experiments commonly use the Geant4 package [42], and this is the case for my analysis as well. Geant4 has functionality to handle physical geometry of detectors, including material composition, tracking and interactions of particles passing through detectors, and simulation of the detector response.

Geant4 simulation of the detector includes particle interactions with the physical elements of the experiment including detectors, absorbers, structural elements, etc., and the ability to simulated how particle paths will be affected by the experimental layout. When a particle passes through the active area of a detector element, Geant4 is able to produces a simulated detector response which can then be further processed into a simulation of the electrical response that we would observe coming from the experiment itself. As shown in [43], Geant4 simulates the response of the CMS detector quite well.

4.1.3 *Digitization*

Custom software takes the results of the GEANT4 simulation of the detector response and translates this physical response into a format that is identical to the data streams recorded from the physical detector.

4.1.4 *Reconstruction of Simulated Events*

The events are reconstructed using the same software that does reconstruction of real events collected from the detector as described in Section 4.2.

4.1.5 *Fastsim*

GEANT4-based simulations are very complex. The ability to produce Monte Carlo samples at a faster rate greatly aids analysis work. CMS fastsim [44] is an algorithm that is able to achieve simulation results similar to those from GEANT4 Monte Carlo. The algorithm simulates events about 100 times faster than the full simulation and reconstruction chain. In the analysis presented in Chapter Five, fastsim samples were essential to create the multiple signal samples needed to cover the wide parameter space included in the study.

Fastsim takes as its input the list of particles created by the event generator including the mother-daughter relationships of particles in a decay chain. The trajectories of stable/quasi-stable particles is simulated as they traverses the magnetic field of the detector, and the parts of the various subdetectors with which they might interact. The geometry of the detector materials is simplified for this simulation. Figure 4.1 shows the silicon tracker geometry used by fastsim. As the particles move along their simulated trajectory, the quasi-stable particles will have simple simulations of their decays performed based on their lifetimes and branching ratios. The decay products are added to the list and allowed to move along a trajectory and decay as well. The simulation of PU is done by reading in pre-generated PU profiles. The

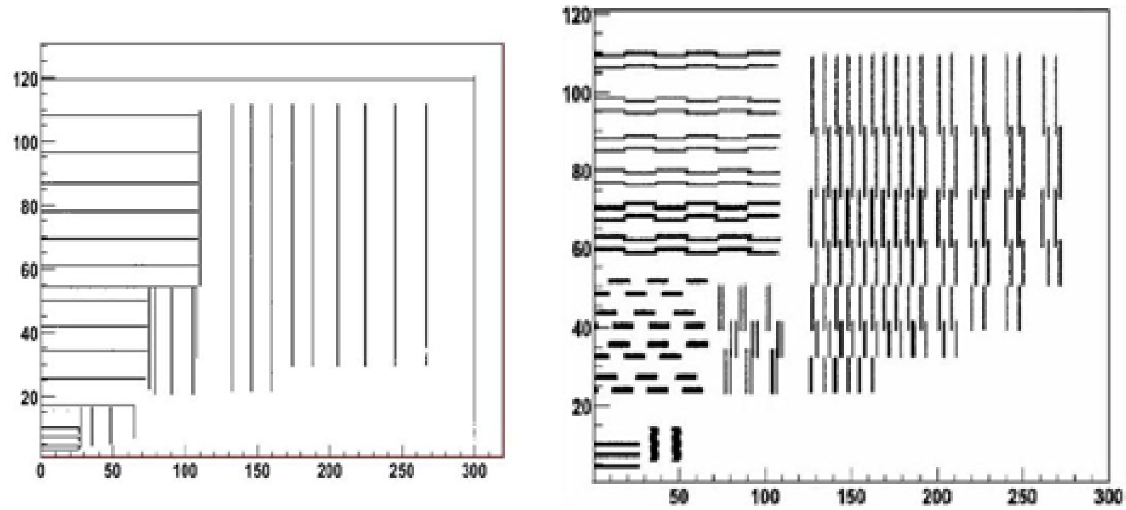


Figure 4.1: Simplified tracker geometry for fastsim MC. Fastsim uses a simplified tracker geometry for simulating particle interactions with the tracker. Reconstruction of the event uses the common geometry used in event reconstruction.

electronic response of the particle interactions is then created using some simplified models.

4.2 Event Reconstruction

Event reconstruction begins with forming tracks and vertices with hits in the inner tracker and muon chambers, and forming reconstructed clusters from energy deposits in the calorimeters. The tracks reveal the trajectory of charged particles and how the trajectory is bent by the magnetic field of the solenoid. This is enough information to determine the momentum and charge of the particle associated with a track. The calorimeters record the position and energy of the electrons, photons, and hadron absorbed by the calorimeter. Correlating these fairly basic objects from all of the detector's layers allows us to identify final state particles through an algorithm known as particle flow [45]. Figure 4.2 shows a simulated dijet event overlaid with reconstructed tracks and calorimeter clusters, and with jets formed by clustering PF particles. This section will describe how tracks and vertices are formed from hits in the inner tracker and muon system and how energy deposits in the calorimeters

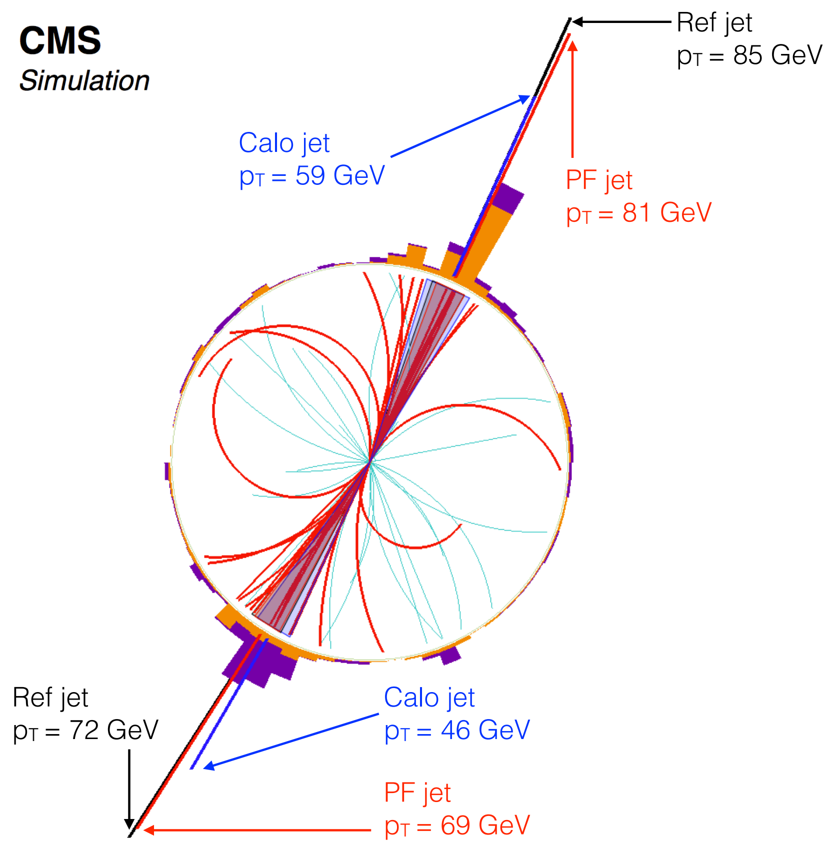


Figure 4.2: Simulated dijet event with reco objects and PF jets. The simulated tracks and calorimeter responses are shown. Calo jets are formed by clustering calorimeter hits. PF jets are formed by clustering PF particles reconstructed using not only the calorimeter hits but also reconstructed charged particle tracks and muons.

are determined. Section 4.3 will describe how the particle flow algorithm synthesizes this information together in order to form physics objects such as electrons, muons, photons, and jets.

4.2.1 Tracks

As charged particles traverse the elements of the silicon tracker, a series of hits is recorded along the trajectory of the particle. Tracks are made by identifying which hits were induced by the single same particle. Forming tracks is a very challenging problem given the large number of charged particles produced during each bunch crossing, and the huge number of hits produced by those particles. Track reconstruction is handled by an iterative process that handles the combinatorial problem through successive rounds of track identification and fitting. The earliest iterations identify the most unambiguous tracks and remove the hits associated with them. Successive iterations are able to identify less obvious tracks from amongst the reduced set of hit. This algorithm allows for the reconstruction of very complicated particle topologies with identification efficiencies of up to 95% [17].

At the start of every iteration, the algorithm generates a set of seeds. The seeds are pairs or triplets of hits which are good candidates to have been generated by the same particle. Only hits from the pixel layers go into the generation of the seeds. It would be computationally impossible to evaluate every triplet or pair combination. Restraints are imposed to reduce the number of possible seeds to be evaluated. Essentially, the algorithm looks at every hit in the outer(inner) layer of the pixel detector, and then based on the range of possible trajectories that a charged particle could follow and intersect with the hit, it looks for matching hits in the inner(outer) layer of the pixel detector. The precise method of producing these track seeds varies from iteration to iteration.

The track seeds are fed into the Combinatorial Track Finder (CTF) algorithm [17]. The CTF algorithm produces track candidates through a series of steps where additional hits are added to the track candidate and the parameters of the candidate are refined. The algorithm estimates the momentum of the track suggested by each seed then progressively looks for additional hits that would be associated with that trajectory working outward from the pixel layers to the silicon strip layers. It refines the estimated trajectory parameters as additional hits are added to the candidate track. It also merges tracks together as it becomes apparent that they parts of the same trajectory of a single charged particle. When a complete track is formed from interaction point to the outer layers of the tracker, a particle trajectory is refit from the proposed set of hits and is accepted if the track meets certain quality standards.

The first iterations look for high momentum tracks that originate from the beam spot. Later iterations use relaxed constraints, and may accept tracks that do not originate from the beam spot. Tracks originating outside of the beam spot may be caused by the decay of collision products away from the point of the initial interaction.

Electrons have very high chance ($\sim$ 85%) of radiating a photon as it passes through the material of the tracker (this radiation is known as bremsstrahlung). When an electron emits bremsstrahlung radiation, it's momentum changes and there will be a kink in it's trajectory. An algorithm known as the Gaussian Sum Filter (GSF) [46] is used to identify tracks that may correspond to an electron trajectory with a kink in the middle. The GSF algorithm checks track candidates produced by the CTF algorithm that failed the quality criteria to see if a track can be formed consistent with an electron plus bremsstrahlung hypothesis.

In each bunch crossing there are many proton-proton collisions. Each individual collision is seen as a vertex or source of tracks. These primary vertices are identified by looking for the common source of good quality tracks that are also located in a position consistent with the beam spot.

Muon tracks are reconstructed using the same algorithms, but the track seeds are produced from hits in the muon chambers located in the outer parts of the detector, and the track is then developed inward towards the inner tracker.

4.2.2 Calorimeter Clusters

The calorimeters measure both the energy and position of the absorbed particles. The particles are absorbed by causing them to decay into a particle shower. This shower can penetrate into many depths of the calorimeter and widen to cover adjacent calorimeter cells. In order to accurately identify and measure the energy of these particles, the response of the calorimeter cells needs to be clustered together.

Clustering is performed separately for HCAL and ECAL. The clustering algorithm identifies seeds which are calorimeter cells that exceed a specific energy threshold. Superclusters are then formed by joining adjacent calorimeter cells that have smaller deposits of energy. The superclusters may be associated with particle showers of absorbed particles. A maximum likelihood fit of Gaussian distributions is made to the cells of the supercluster, and the number, distribution, and energy of the absorbed particles is estimated.

4.3 Particle Flow Algorithm

Physics analyses can be done with the reconstructed tracks and calorimeter clusters already described. The tracks can give sufficient evidence of the identity and momentum of charged particles, and the calorimeter clusters can be used to identify and measure hadronic jets. The particle flow algorithm synthesizes information from all of the detector subsystems to produce a much more refined and accurate reconstruction. The algorithm takes as inputs the inner tracker and muon tracks and the calorimeter clusters. It links them together to form PF candidates. The PF candidates are iteratively refined and identified.

Tracks are linked to calorimeter clusters by extrapolating the trajectory of the charged particle into the calorimeters and looking for a matching clusters. If a GSF track is linked to an ECAL cluster, then the trajectory of the bremsstrahlung photon is estimated and a link is made to an ECAL cluster that may exist along that trajectory. ECAL and HCAL clusters are linked if they overlap. Tracks from the inner tracker can also be linked to hits that follow the candidate particle's trajectory in the muon system that haven't already been assign to a muon track. After these links are made and the initial PF candidates formed, an iterative process identifies and refines the candidates. At each step of particle flow identification, candidates that have been assigned an identity are removed from consideration for the following steps. The identity of the PF candidate is used to apply object specific energy corrections based on the known detector response to different species of particles.

4.3.1 Muons

The first particles to be classified by the particle flow algorithm are muons. Two types of PF candidates may prove to be a muon. The muon tracks formed from hits in the muon system, and inner tracker tracks that have been linked to hits in the muon system. The vast majority of particles that make it from the interaction point to produce hits in the muon system are muons, but there is a fraction of hadrons that can make it that far and produce hits in the muon system. In order to be identified as a muon, an isolation criteria is applied. The energy of the adjacent tracks and calorimeter clusters within a small cone around the muon candidate is summed together, if the candidate muon's energy is significantly larger than the summed energy of the cone, the candidate is identified as a muon. An example of the reconstruction of a pair of muons can be found in Fig. 4.3.

4.3.2 Electrons

After muons have been identified, the particle flow algorithm looks for electrons. Electrons are found using the information provided by both the tracker and the calorimeters. Electron identification is a multi-step process that attempts to find and refine tracks with clusters in ECAL consistent with an electron. Two types of PF candidates are evaluated as possible electrons. The first is a candidate that has an ECAL cluster without a significant matching cluster in HCAL and linked to an inner tracker track. The second class of candidates are the tracks produced by the GSF algorithm that are linked to an ECAL cluster. The process of classifying this candidates as electrons involves refitting tracks, reclustering ECAL clusters, and identifying bremsstrahlung ECAL clusters under the hypothesis that the candidate

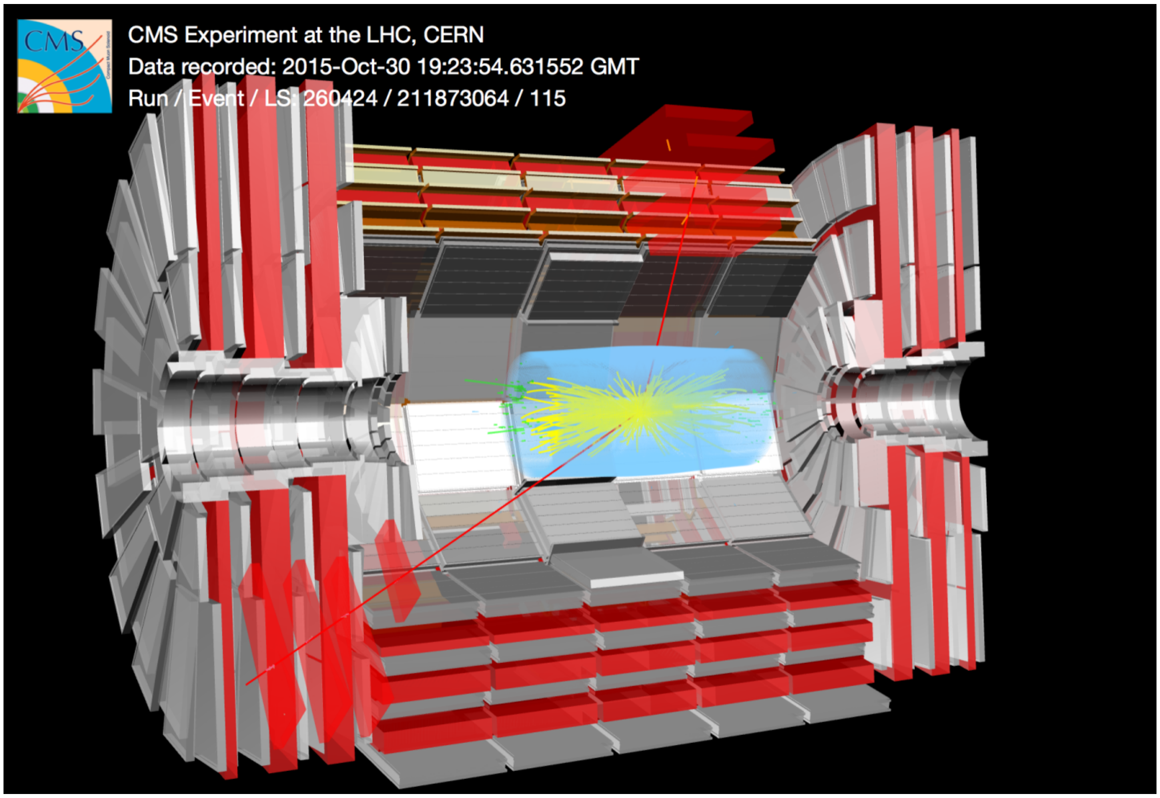


Figure 4.3: A collision event with two reconstructed muons. Reconstruction of the muons combines information from the inner tracker and the muon chambers. In this example, one muon has been reconstructed in the barrel region and the other muon has been reconstructed in the endcap region.

is an electron. The resulting PF candidate is classified as an electron if it meets the appropriate quality criteria.

4.3.3 Photons

The response of the detector to photons share qualities with the response to electron and the response to neutral hadrons. Like electrons, photons will deposit most of their energy in the ECAL. Like neutral hadrons, the calorimeter clusters made by the photons will not be linked to a track. In the particle flow algorithm, photons are identified after the electrons. Photon candidates are simply formed if an

ECAL cluster is not associated with a track and the ratio of the ECAL energy to the HCAL energy is consistent with a photon.

4.3.4 Neutral and Charged Hadrons

Once the isolated muons, electrons and photons have been identified, the remaining particles to be identified are the charged and neutral hadrons that arise from hadronization of the collision products. During the process of identifying these particles, it becomes necessary at times to consider that there may be non-isolated muons, electrons, and photons as yet unidentified in the remaining PF candidates.

The simplest identification of a neutral hadron is a link between ECAL and HCAL clusters that does not have any tracks linked to it, and is not identified as a photon. The simplest identification of a charged hadron is when the momenta of tracks linked to ECAL and HCAL clusters are comparable to the energy deposited in the calorimeter.

If the calorimeter energy exceeds the total momentum of the linked tracks, then the algorithm tests the hypothesis that there is a photon or neutral hadron in close proximity to the charged hadron. As before, the photon and neutral hadron can be distinguished by the relative amounts of excess energy in the ECAL and HCAL.

If the calorimeter energy is significantly less than the total momentum of the linked tracks, the algorithm tests the hypothesis that a muon produced one of the tracks. If a muon is constructed, its momentum is removed from the sum of the tracks. After such muons a reconstructed, if the calorimeter energy is still significantly less that the total momentum of the linked tracks, it is assumed that there are some misreconstructed tracks linked to the calorimeter cluster. The tracks are removed until the sum of the track momenta is consistent with the calorimeter energy, or there are no more tracks linked to the cluster.

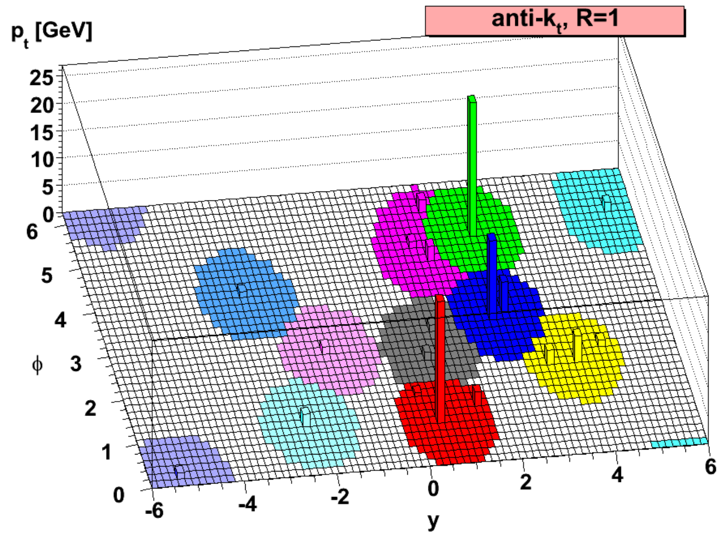


Figure 4.4: Jet clustering with the anti-k_{T} algorithm. This plot is from the paper describing the anti-k_{T} algorithm [47].

4.4 Other Physics Objects

After the particle flow algorithm has produced its collections of reconstructed and identified particles, those collections are used to construct additional objects used in physics analyses.

4.4.1 Jets

Identifying jets is a fundamental part of understanding particle physics collisions. Final state quarks and gluons will hadronize and produce a jet of particles.

To identify a jet, an algorithm is used to cluster particles or particle candidates. In CMS, during Run 2, the most commonly used jet collections use the anti-k_{T} algorithm [47]. Two typical distance parameters of the anti-k_{T} algorithm used, characterizing the size of jet, are 0.4 and 0.8. Those jets clustered using the anti-k_{T} algorithm with the distance parameters 0.4 and 0.8 are referred as AK4 and AK8 jets, respectively. Fig. 4.4 shows an example of the clustering performed by the anti-k_{T} algorithm.

4.4.2 Heavy-flavor Jets

It is possible to identify jets that are likely the product of a heavy particle such as a bottom or top quark. Identifying top quarks can be especially challenging, and the methods used to identify them in this analysis are discussed in Section 5.4 and a more powerful approach to identifying top quarks that decay into three distinguishable jets is discussed in Chapter Six.

In general, it is possible to look at the properties of jet constituents in order to characterize its origin. One such method used to identify bottom quark jets is the combined secondary vertex algorithm (CSVv2) [48, 49], used in the analysis described in Chapter Five, which utilizes variables such as the impact parameters of charged-particle tracks, the properties of reconstructed decay vertices, and combinations thereof.

4.4.3 Missing Transverse Momentum

Missing transverse momentum is an important handle in looking for new physics beyond the standard model. The magnitude of the missing transverse momentum is denoted by $p_{\mathrm{T}}^{\mathrm{miss}}$. Another commonly used term in the high energy physics community and in some literature is MET (Missing Transverse Energy). Since at the time of the proton collision there is no net momentum in a transverse direction, a transverse momentum imbalance in the detector indicates that there may have been particles produced that did not interact with the detector. Neutrinos are a natural source of momentum imbalance, and are the only standard model particle that will not interact with the detector. Any type of weakly interacting dark matter would not interact with the detector, including some types of SUSY particles.

There are L1 triggers that look for transverse momentum imbalances based on the results of the calorimeters. More refined versions of $p_{\mathrm{T}}^{\mathrm{miss}}$ are calculated after event reconstruction and object identification with the particle flow algorithm. In general

$p_{\mathrm{T}}^{\mathrm{miss}}$ will be calculated by finding the negative of the vector sum of the PF particles in the transverse direction.

In searches that look at all-hadronic final states, a variation on $p_{\mathrm{T}}^{\mathrm{miss}}$ called $H_{\mathrm{T}}^{\mathrm{miss}}$ or MHT can be used. MHT is the inverse of the vector sum of the transverse momentum of all reconstructed jets that pass a certain energy threshold.

In addition to true imbalances of transverse momentum due to non-interacting particles, there is a certain amount of imbalance that can be caused by poorly reconstructed particles. These include cases of momentum mismeasurement and particles that are not identified or correctly identified by the particle flow algorithm.

CHAPTER FIVE

Search for Supersymmetry in All-Hadronic Final States Using Top Quark Identification

Searches for supersymmetric models are complicated by the large backgrounds from standard model processes. Stringent limits have already been established on the production of Standard Model superpartners, and the remaining possible cross sections are often too low to be observed while superimposed on the stand model background. In general then, a SUSY search is going to need two well developed aspects: selection criteria that greatly enriches the proportion of signal events, and precise predictions of the standard model backgrounds.

5.1 Search Overview

An all-hadronic search is a good match for searching for R-parity conserving SUSY. In events that produce SUSY particles, at least two LSPs will be produced and escape reconstruction as $p_{\mathrm{T}}^{\mathrm{miss}}$. In a hadron collider such as the LHC, charged leptons are often produced through Drell-Yan processes, and also through the decay of W bosons to a charged lepton and a neutrino. By excluding events with isolated leptons, we are able to reduce the number of events that contain a W boson that decays leptonically. This reduces the the significance of a natural imbalance in transverse energy due to the production of neutrinos.

If masses of color-charged BSM particles are sufficiently light, it is reasonable that they will produced by the strongly interacting partons of a proton-proton collision. Color-charged BSM particles would decay to strongly interacting standard model particles, giving good motivation to look for events with a large number of jets and all-hadronic events.

5.2 Signal Models

There are numerous families of theories that extend the standard model through the introduction of a supersymmetric relationship of standard model particles with as yet undiscovered partners [50]. As mentioned in Chapter Two, there are physical and theoretical features of SUSY theories that lead us to think that such BSM phenomena may be produced and observed in the current experiment. Although a broken symmetry, there is no a priori reason to believe that the symmetry is broken to the extent that the masses of the SM particles and the SUSY partners are wildly different or inaccessible. Indeed, SUSY particles at the electroweak scale allow a natural theory free of excessive fine tuning of the parameters. The astronomical observation of dark matter and the likely possibility that dark matter is a particle fit well with a stable LSP in an R-parity conversing theory. Such a stable LSP would leave an identifiable signature in the event as $p_{\mathrm{T}}^{\mathrm{miss}}$. The simplest extensions include the fewest number of new particles. Such a model will introduce a single new particle for each of the fermions and gauge bosons, and a couple of additional Higgs fields. Even such a simple model is unwieldy as it introduces an order of magnitude more unfixed parameters into the model.

Signal scenarios for SUSY are considered in the context of simplified models [51–55]. For direct top squark pair production, the simplified model denoted “T2tt” is examined. In this model, each top squark $\widetilde{\mathrm{t}}$ decays to a top quark and the LSP: $\widetilde{\mathrm{t}} \to \mathrm{t}\widetilde{\chi}_1^0$. For top squark production through gluino decay, the models described in the following two paragraphs are considered.

In the model denoted “T1tttt,” pair-produced gluinos each decay to an off-shell top squark and an on-shell top quark. The off-shell top squark decays to a top quark and the LSP. The gluino decay is thus $\widetilde{\mathrm{g}} \to \mathrm{t}\bar{\mathrm{t}}\widetilde{\chi}_1^0$. The T1tttt model provides sensitivity to situations in which the top squark is too heavy to be produced directly while the gluino is not. In the “T1ttbb” model, pair-produced gluinos each decay via an off-shell top or bottom squark as $\widetilde{\mathrm{g}} \rightarrow \mathrm{t}\bar{\mathrm{t}}\widetilde{\chi}_1^0$ (25%), $\widetilde{\mathrm{g}} \rightarrow \bar{\mathrm{t}}\mathrm{b}\widetilde{\chi}_1^+$ or its charge conjugate (50%), or $\widetilde{\mathrm{g}} \rightarrow \mathrm{b}\bar{\mathrm{b}}\widetilde{\chi}_1^0$ (25%), where $\widetilde{\chi}_1^+$ is the lightest chargino. The mass difference between the $\widetilde{\chi}_1^+$ and the LSP is taken to be $\Delta m(\widetilde{\chi}_1^+, \widetilde{\chi}_1^0) = 5\,\mathrm{GeV}$. Thus the $\widetilde{\chi}_1^+$ is taken to be nearly mass degenerate with the $\widetilde{\chi}_1^0$, representing the expected situation if the two particles appear within the same SU(2) multiplet [53]. The $\widetilde{\chi}_1^+$ subsequently decays to the LSP and an off-shell W boson. The T1ttbb model provides sensitivity to mixed states of top and bottom squarks.

In the model denoted "T5tttt," the mass difference between the top squark and the LSP is $\Delta m(\widetilde{\mathrm{t}}, \widetilde{\chi}_1^0) = 175\,\mathrm{GeV}$. Pair-produced gluinos each decay to a top quark and an on-shell top squark. The top squark decays to a top quark and the LSP. This model provides sensitivity to a region that is difficult to probe with the T2tt model because of the similarity between the properties of T2tt signal and $\mathrm{t}\bar{\mathrm{t}}$ background events when $\Delta m(\widetilde{\mathrm{t}}, \widetilde{\chi}_1^0)$ approximately equals the top quark mass (m_{t}). The "T5ttcc" model is similar to the T5tttt model except it assumes $\Delta m(\widetilde{\mathrm{t}}, \widetilde{\chi}_1^0) = 20\,\mathrm{GeV}$ and the top squark decays to a charm quark and the LSP. Note that decay to a charm quark and an LSP represents the dominant decay mode of a top squark when its decay to a top quark and an LSP is kinematically disallowed and the chargino is sufficiently heavy that stop to b quark and chargino is suppressed. The choice of $\Delta m(\widetilde{\mathrm{t}}, \widetilde{\chi}_1^0)$ has little effect on the final results for the T5ttcc model so long as $\Delta m(\widetilde{\mathrm{t}}, \widetilde{\chi}_1^0)$ remains below m_{t}. The T5ttcc model provides sensitivity to scenarios in which the top squark is kinematically unable to decay to an on-shell top quark.

The signal scenarios are illustrated in Fig. 5.1. They exhibit common features, such as the presence of multiple top quarks and two LSPs.

5.3 Preparing an All-Hadronic Sample

To prepare an all-hadronic event sample, the following three types of event vetoes are used: isolated lepton vetoes, an isolated charged-particle veto, and an isolated charged hadron veto.

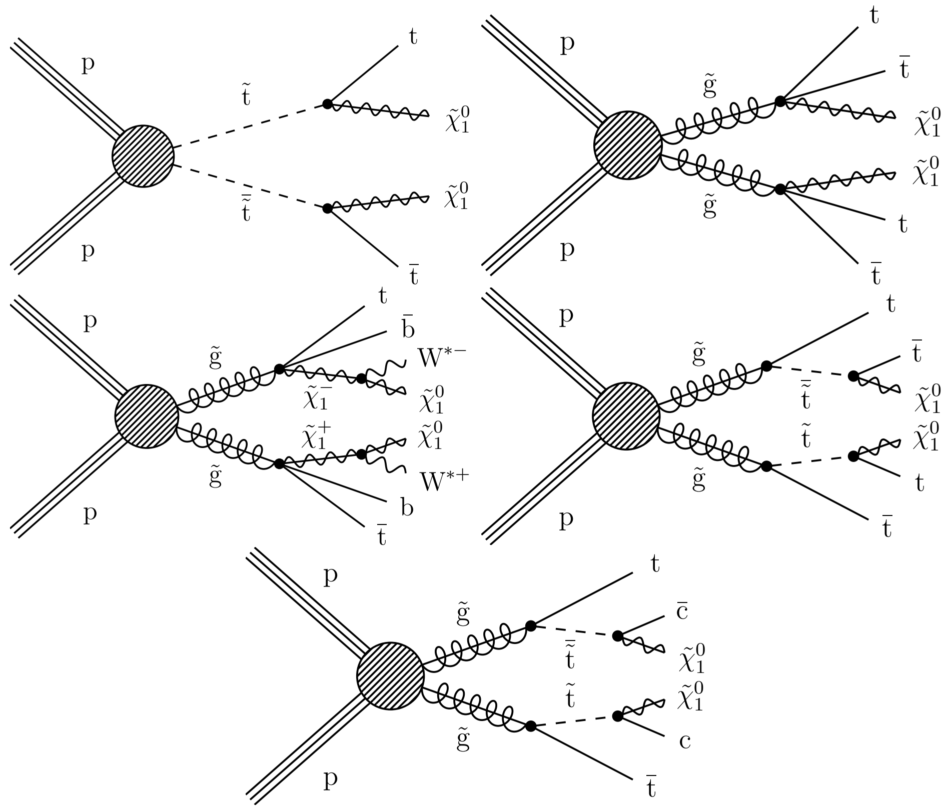


Figure 5.1: Diagrams representing the simplified models of direct and gluino-mediated top squark production considered in this analysis: the T2tt model (top left), the T1tttt model (top right), the T1ttbb model (middle left), the T5tttt (middle right), and the T5ttcc model (bottom).

The isolated lepton vetos reject events with isolated electrons or muons. The "isolation sum" of electron and muon candidates is defined as the scalar p_{T} sum of PF candidates in a cone of radius $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ around the candidate's trajectory, where ϕ is the azimuthal angle and the sum excludes the electron or muon candidate. The cone size is 0.2 for $p_{\mathrm{T}} \leq 50\,\mathrm{GeV}$, 0.05 for $p_{\mathrm{T}} \geq 200\,\mathrm{GeV}$, and decreases in inverse proportion to the lepton p_{T} for $50 < p_{\mathrm{T}} < 200\,\mathrm{GeV}$. This decreasing cone size for increasing lepton p_{T} accounts for the greater collimation of an object's decay products as its Lorentz boost increases. The isolation sum is corrected for contributions from pileup using an estimate of the pileup energy in the cone [56]. Electron and muon candidates are considered to be isolated if their relative isolation, i.e., the ratio of the isolation sum to the candidate p_{T}, is less than 0.1 and 0.2, respectively.

An event that survives the isolated lepton vetoes are subjected to an isolated charged-particle track veto. This veto suppresses events with a hadronically decaying τ lepton and and isolated electrons or muons not identified as such. Tracks considered for this veto must have $p_{\mathrm{T}} > 5\,\mathrm{GeV}$, $|\eta| < 2.5$, and relative track isolation less than 0.2. The relative track isolation is defined analogously to the relative isolation of electrons and muons but is computed using charged PF candidates only, that appear within a fixed cone of $\Delta R = 0.3$ around the track. To preserve signal efficiency, the isolated-track veto is applied only if the transverse mass m_{T} [57] of the isolated track-$\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ system is consistent with W boson decay [1], namely $m_{\mathrm{T}} < 100\,\mathrm{GeV}$. The isolated-track veto reduces background from events with a leptonically decaying W boson by about 40%.

Following application of the above two vetoes, a significant fraction of the remaining SM background arises from events with a hadronically decaying τ lepton (τ_{h}). A charged-hadron veto is applied to reduce this background. The charged-hadron veto eliminates events that contain an isolated PF charged hadron with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$, $|\eta| < 2.5$, and $m_{\mathrm{T}} < 100\,\mathrm{GeV}$. To be considered isolated, the relative isolation of the charged hadron, defined as in the previous paragraph, must be less than 0.1.

5.4 Top Quark Identification

Identifying top quarks is a central feature of the analysis. Chapter Six will discuss in depth a proposed algorithm for the identification of top quarks that decay to three distinctly reconstructed jets. This algorithm will provide advantages over the method used in this analysis and other generally used methods. In this section, the top quark tagging algorithm used in the analysis is described.

---

[1] m_{T} is invariant mass of the lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$, in an event with $E_{\mathrm{T}}^{\mathrm{miss}}$ largely due to a single neutrino from the leptonic decay of a W boson, m_{T} will cut off at the W boson mass.

It is designed to provide high reconstruction efficiency over the full range of top quark p_{T} in the considered signal models. A common strategy [58, 59] for tagging hadronically decaying top quarks is to cluster jets with the AK8 algorithm and then to test whether the jet is consistent with having three subjets, as expected for the $\mathrm{t} \to \mathrm{bq\overline{q}'}$ decay of a highly Lorentz-boosted top quark. Although these algorithms are efficient at large top quark p_{T}, for $p_{\mathrm{T}} < 400\,\mathrm{GeV}$ top quarks are more efficiently reconstructed by combining three individual AK4 jets, an approach known as "resolved" top quark tagging. To obtain high reconstruction efficiency over a wide range of top quark p_{T}, we employ both types of algorithms and, in addition, consider top quark decays in which the decay products of the W boson are contained within an AK8 jet. To fully reconstruct the top quark in the latter case, an AK8 jet corresponding to the W boson decay is combined with an AK4 jet.

To identify high-p_{T} top quarks, AK8 jets with $p_{\mathrm{T}} > 400\,\mathrm{GeV}$ are selected. The mass of the jet is corrected with the soft-drop method [60,61] using angular exponent $\beta = 0$, soft cutoff threshold $z_{\mathrm{cut}} < 0.1$, and characteristic radius $R_0 = 0.8$, where the values of β, z_{cut}, and R_0 are those recommended in Ref. [62] for AK8 jets. The soft-drop algorithm reclusters the AK8 jet into subjets using the Cambridge–Aachen algorithm [63, 64]. This reclustering removes soft radiation, which can bias the jet mass determination. To be considered as a top quark candidate, the soft-drop mass must lie between 105 and 210 GeV. The N-subjettiness variables τ_N [65] are used to determine the consistency of the jet with having three subjets. More details on this algorithm can be found in Ref. [58]. To be consistent with having three subjets, the requirement $\tau_3/\tau_2 < 0.65$ is imposed. This requirement is made on the basis of optimization studies [62].

To avoid overlap between the top-tagged AK8 jets (denoted "monojets") and the AK4 jets that are used to reconstruct resolved ("trijets") or partially merged ("dijets") top quarks, AK4 jets matched to the top-tagged AK8 jet are removed from the list of AK4 jets used in the reconstruction of the dijet and trijet categories. An AK4 jet is considered matched if it lies within $\Delta R < 0.4$ of one of the soft-drop subjets of the tagged AK8 jet.

For the dijet category of top quark decays, we employ a similar technique to identify the jet from the hadronic W boson decay. An AK8 jet with $p_{\mathrm{T}} > 200\,\mathrm{GeV}$ must have a soft-drop corrected mass between 65 and 100 GeV. To be consistent with having two subjets, the requirement $\tau_2/\tau_1 < 0.6$ is imposed. This requirement corresponds to the "high-purity pruning" criterion of Ref. [62]. The AK8 jet is combined with a loose AK4 jet to form a top quark candidate. The candidate must have a mass between 100 and 250 GeV, both jets must appear within a cone of radius $\Delta R = 1$ around the direction of their summed p_{T} vector, and the ratio of the soft-drop corrected AK8 jet mass to the top quark candidate mass must lie between $0.85\,(m_{\mathrm{W}}/m_{\mathrm{t}})$ and $1.25\,(m_{\mathrm{W}}/m_{\mathrm{t}})$, with m_{W} the W boson mass. If more than one top quark candidate is found using the same AK8 jet, the combination with mass closest to m_{t} is chosen. The AK4 jet used to form the top quark candidate, and all AK4 jets matched to within $\Delta R < 0.4$ of the soft-drop subjets from the AK8 jet, are removed from the list used to reconstruct the trijet category.

The trijet sample of top quark candidates is formed by combining three loose AK4 jets. The three jets must appear within a cone of radius $\Delta R = 1.5$ around the direction of their summed p_{T} vector, no more than one of the three jets can be b tagged, and the trijet mass must lie between 100 and 250 GeV. The cone size is chosen to be $\Delta R = 1.5$ because the background becomes very large for larger ΔR values. The final trijet top quark sample is defined by applying the results of a random forest boosted decision tree [66] to the selected combinations. The random forest is trained with simulation using trijet combinations that satisfy the above criteria. Simulated samples of $\mathrm{t\bar{t}}$ and $\mathrm{Z}(\nu\overline{\nu})$+jets events are used for this purpose. In the $\mathrm{t\bar{t}}$ simulation, one top quark decays hadronically and the other semileptonically. Signal top quarks are defined as trijet combinations in the $t\bar{t}$ simulation for which each of the three jets is matched to a distinct generator-level hadronically decaying top quark decay product within $\Delta R < 0.4$, and whose overall momentum is matched to the generator-level top quark momentum within $\Delta R < 0.6$. Background combinations are defined as trijet combinations in the $t\bar{t}$ sample with no jet matched to a generator-level hadronically decaying top quark decay product, and as trijet combinations in the $Z(\nu\overline{\nu})$+jets sample. If more than one background combination is found in an event, all combinations are used.

The variables considered in the random forest algorithm are the mass of the trijet system, the mass of each dijet combination, the angular separation and momenta of the jets in the trijet rest frame, the b tagging discriminator value of each jet, and the quark-versus-gluon-jet discriminator [67] value of each jet. To reduce correlations with the top quark p_{T} and thus to prevent overtraining in this variable, the p_{T} spectra of signal and background triplet combinations are flattened through reweighting. The random forest performance is improved by replacing the kinematic variables in the laboratory frame with their equivalents in the trijet rest frame, and by sorting jets according to their momenta in the trijet rest frame so that the highest (lowest) momentum jet is most (least) likely to originate from a b quark.

Trijet top quark candidates are selected by requiring the random forest discriminator value to exceed 0.85. This value is chosen based on optimization studies involving the full limit-setting procedure.If two or more selected trijets share one or more AK4 jets, only the combination with the largest discriminator value is retained.

All top quark candidates must have $|\eta| < 2.0$. The final set consists of the non-overlapping candidates from the three reconstruction categories. The total efficiency of the algorithm, including a breakdown into the three categories, is shown in Fig. 5.2. The efficiency is determined using T2tt signal events with a top squark mass of 850 GeV and an LSP mass of 100 GeV, based on the number of generator-level

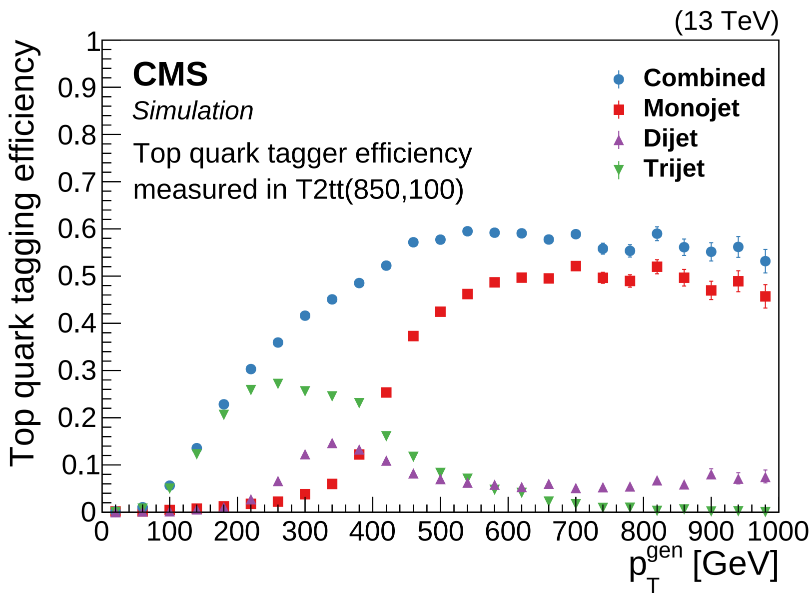


Figure 5.2: Efficiency of the top quark tagger as a function of generator-level top quark p_{T} for the monojet (red boxes), dijet (magenta triangles), and trijet (green upside-down triangles) categories and for their combination (blue circles), as determined using T2tt signal events with a top squark mass of 850 GeV and an LSP mass of 100 GeV. The vertical bars indicate the statistical uncertainties.

hadronically decaying top quarks that are matched to a reconstructed top quark candidate divided by the total number of generator-level top quarks that decay hadronically. Similar results are found using SM $\mathrm{t\bar{t}}$ events. The matching between the generator-level and reconstructed top quarks requires the overall reconstructed top quark to be matched to the generator-level top quark within $\Delta R < 0.4$. The misidentification rate varies between 15 and 22% as a function of $p_{\mathrm{T}}^{\mathrm{miss}}$, with an average of about 20%, as determined using simulated $\mathrm{Z}(\nu\overline{\nu})$+jets events after applying selection criteria similar to those used for the data (Section 5.5): $N_{\mathrm{j}} \geq 4$, $N_{\mathrm{b}} \geq 1$, $p_{\mathrm{T}}^{\mathrm{miss}} > 250\,\mathrm{GeV}$, and no isolated electron or muon with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$.

Relative to Ref. [68], the top quark tagging algorithm has been improved by using AK8 jets for the monojet and dijet categories, rather than strictly AK4 jets, and through implementation of the random forest tree for the trijet category. These improvements provide a factor of two reduction in the top quark misidentification rate while maintaining a similar efficiency.

5.5 Event Selection and Search Regions

Our study is an inclusive search for events containing $p_{\mathrm{T}}^{\mathrm{miss}}$ and reconstructed top quarks. Events in this search are collected with a trigger that requires $p_{\mathrm{T}}^{\mathrm{miss}} > 100\,\mathrm{GeV}$ and $H_{\mathrm{T}}^{\mathrm{miss}} > 100\,\mathrm{GeV}$, where $H_{\mathrm{T}}^{\mathrm{miss}}$ is the magnitude of the vector p_{T} sum of the jets. After the selection criteria, explained further in this section, this trigger is fully effective at capturing events that meet these requirements. Filters are applied to every event designed to remove events with excessive noise. These filters look for effects from known sources of noise within the detector itself and noise related to the condition of the particle beam.

The events are subjected to the lepton, isolated-track, and charged-hadron vetoes of Section 5.3. To improve the rejection of background, the two tight AK4 jets with highest p_{T} must have $p_{\mathrm{T}} > 50\,\mathrm{GeV}$. Events are required to have $N_{\mathrm{j}} \geq 4$, $N_{\mathrm{b}} \geq 1$, $N_{\mathrm{t}} \geq 1$, $p_{\mathrm{T}}^{\mathrm{miss}} > 250\,\mathrm{GeV}$, and $H_{\mathrm{T}} > 300\,\mathrm{GeV}$.

The QCD multijet background mostly arises when the p_{T} of one of the highest p_{T} jets is under measured, causing $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ to be aligned with that jet. This under measurement can occur because of jet misreconstruction or, in the case of semileptonic b or c quark decays, an undetected neutrino. To reduce this background, requirements are placed on the azimuthal angle between $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and the three loose AK4 jets with highest p_{T}, denoted j_1, j_2, and j_3 in order of decreasing p_{T}. Specifically, we require $\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}}, j_1) > 0.5$, $\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}}, j_2) > 0.5$, and $\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}}, j_3) > 0.3$.

The m_{T2} variable [68–70] is used to reduce background from $\mathrm{t\bar{t}}$ events. This variable is designed to provide an estimate of the transverse mass of pair-produced heavy objects that decay to both visible and undetected particles. It has a kinematic upper limit at the mass of the heavy object undergoing decay. Thus the upper limit for SM $\mathrm{t\bar{t}}$ events is m_{t}, while the upper limit for TeV-scale squarks and gluinos is much larger. If there are two tagged top quarks in an event, m_{T2} is calculated using the pair of tagged top quarks and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$. If there are more than two tagged top quarks, we compute m_{T2} for all combinations and choose the combination with the smallest m_{T2}. If there is only one tagged top quark, we construct a proxy for the other top quark using the highest p_{T} b tagged jet as a seed. If a b tagged jet is not available, because there is only one b tagged jet in the event and it is part of the reconstructed top quark, the highest p_{T} jet is used as the seed. The seed jet is combined with a loose AK4 jet to define the top quark proxy if the resulting pair of jets has a mass between 50 and 220 GeV and if the two jets appear within $\Delta R = 1.5$ of each other; otherwise the seed jet by itself is used as the top quark proxy. The proxy is combined with the tagged top quark and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ to determine m_{T2}. Irrespective of the number of tagged top quarks, we require $m_{\mathrm{T2}} > 200\,\mathrm{GeV}$.

The search is performed in 84 non-overlapping search regions. Regions with $N_{\mathrm{b}} \leq 2$ and $N_{\mathrm{t}} \leq 2$ use N_{b}, N_{t}, $p_{\mathrm{T}}^{\mathrm{miss}}$, and m_{T2} as the binned search variables. Regions with $N_{\mathrm{b}} \geq 3$ or $N_{\mathrm{t}} \geq 3$ use N_{b}, N_{t}, $p_{\mathrm{T}}^{\mathrm{miss}}$, and H_{T}. The reason H_{T} is used for these latter regions, and not m_{T2}, is that in events with many jets, the jets from the decay of a particular heavy object may not always be correctly associated with that object, causing the distribution of m_{T2} to be broad and relatively flat. We find that H_{T} provides better discrimination between signal and background for $N_{\mathrm{b}} \geq 3$ or $N_{\mathrm{t}} \geq 3$. The 84 regions in m_{T2} versus $p_{\mathrm{T}}^{\mathrm{miss}}$ or in H_{T} versus $p_{\mathrm{T}}^{\mathrm{miss}}$ are illustrated in Fig. 5.3. The boundaries between the regions were determined through sensitivity studies.

Recently, analyses of data produced by LHC experiments have been encouraged to share data in a form compatible with reinterpretation [71]. Reinterpretation of physics analyses cannot result in a discovery. Excess events above standard model backgrounds is model independent. However when an analysis reports limits placed

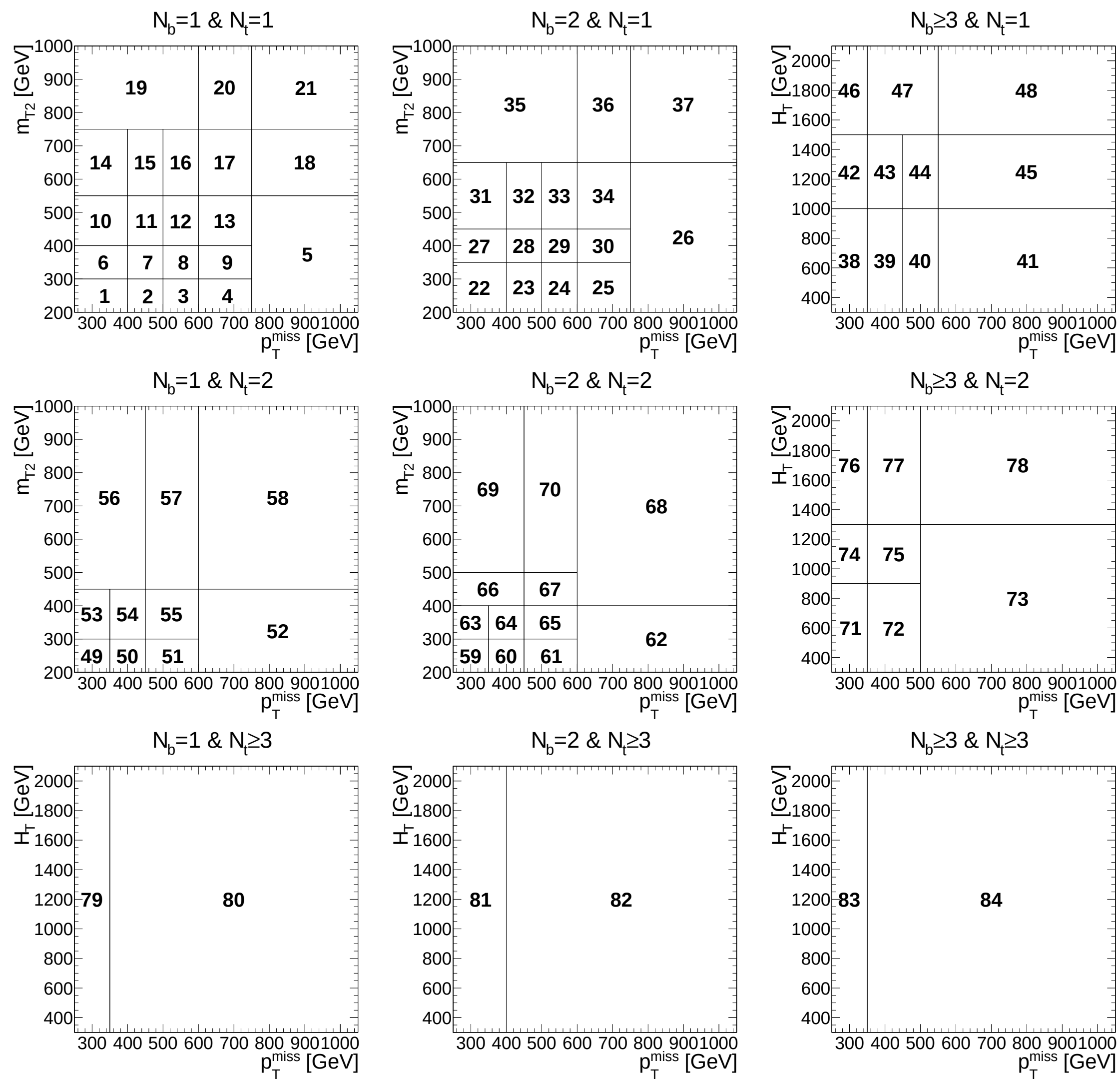


Figure 5.3: Search region definitions in the kinematic variables. The highest $p_{\mathrm{T}}^{\mathrm{miss}}$, m_{T2}, and H_{T} regions are open-ended, e.g., $p_{\mathrm{T}}^{\mathrm{miss}} > 750\,\mathrm{GeV}$ and $m_{\mathrm{T2}} > 750\,\mathrm{GeV}$ for search region 21.

Table 5.1: Definition of the aggregate search regions.

Region	N_t	N_b	m_{T2} [GeV]	p_T^{miss} [GeV]	Motivation
1	≥1	≥1	≥200	≥250	Events satisfying selection criteria
2	≥2	≥2	≥200	≥250	Events with $N_t \geq 2$ and $N_b \geq 2$
3	≥3	≥1	≥200	≥250	Events with $N_t \geq 3$ and $N_b \geq 1$
4	≥3	≥3	≥200	≥250	T5tttt; small $\Delta m(\widetilde{g}, \widetilde{\chi}_1^0)$ and $m_{\widetilde{\chi}_1^0} < m_t$
5	≥2	≥1	≥200	≥400	T2tt; small $\Delta m(\widetilde{t}, \widetilde{\chi}_1^0)$
6	≥1	≥2	≥600	≥400	T2tt; large $\Delta m(\widetilde{t}, \widetilde{\chi}_1^0)$
Region	N_t	N_b	H_T [GeV]	p_T^{miss} [GeV]	Motivation
7	≥1	≥2	≥1400	≥500	T1ttbb, T5ttcc; large $\Delta m(\widetilde{g}, \widetilde{\chi}_1^0)$
8	≥2	≥3	≥600	≥350	T1tttt; small $\Delta m(\widetilde{g}, \widetilde{\chi}_1^0)$
9	≥2	≥3	≥300	≥500	T1/T5tttt, T1ttbb; intermediate $\Delta m(\widetilde{g}, \widetilde{\chi}_1^0)$
10	≥2	≥3	≥1300	≥500	T1/T5tttt; large $\Delta m(\widetilde{g}, \widetilde{\chi}_1^0)$

on new physics, it does so in terms of specific models and parameter spaces. Reinterpretation can take the event selections and background estimations of an analysis and apply it to other physics models in order to place limits on those particular models.

Additional search bins were defined, and a simplified tagger produced to assist in reinterpretation. These bins are aggregates of the standard bins, and chosen to be sensitive to different models as shown in Table 5.1. The bins are not orthogonal, meaning that the bins may overlap, and are not appropriate for combined interpretation. These bins, their definitions, predicted and observed yields are presented in Table B.4. The process of defining the aggregate search bins, and their sensitivity to the various simplified SUSY models previously presented is further explained in Appendix C.

5.6 Background Estimations

The significant background events that remain after the baseline selections are applied include events that include W bosons that decay into leptons in which the charged lepton is not reconstructed (lost lepton background) or in which the W decays to a τ lepton which decays to hadronic jets, events that include Z bosons that decay into neutrinos, QCD multijet events. The estimates of these backgrounds all use various data driven estimation methods. Control samples are generated from events not included in the search bins. The control sample distribution are transformed through validated methods to provide estimates of the background events in the search bins. These backgrounds and a brief description of the methods used to estimate them are described in the following subsections.

When using a data driven method of estimating backgrounds, it is important to understand and mitigate the effects of signal contamination, i.e. potential contributions of signal events to the control samples. If signal events end up in the control sample, then the background estimation may swallow up signal events in the search regions. The mitigation of the signal contamination in the control samples and consequently background estimations are also discussed below.

5.6.1 Events that Include W Bosons Decaying Leptonically

The signal signatures of interest are all-hadronic final states. The isolated lepton veto significantly cuts away the non all-hadronic background. However, this cut by itself does not eliminate all events with a lepton in the final state. Not every event will be correctly reconstructed. As such an event with an isolated electron or muon may not have them identified and reconstructed. This is the “lost lepton” problem. As depicted in Fig. 5.4, τ leptons are massive enough that significant number of its decays result in hadronic jets, and not subject to the lepton veto. This is the “hadronic-τ” problem.

Control regions are generated using data events from the same triggers that produce the search regions. All of the same selection cuts are applied except for the lepton veto and the isolated track veto. Instead events must include exactly one reconstructed, isolated electron or muon. Signal contamination is reduced by imposing a requirement that m_{T} be less that 100 GeV (consistent with W boson decay).

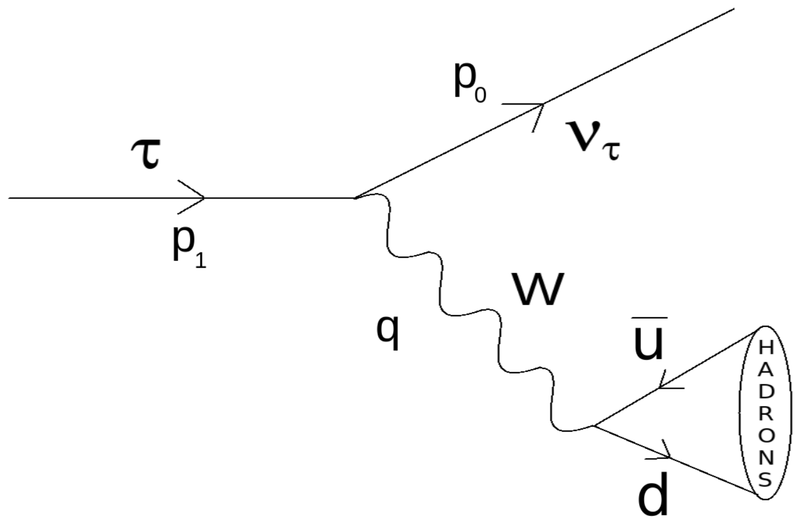


Figure 5.4: A tree level diagram of a τ lepton decaying to a hadronic jet [72]. Unlike the lighter electron and muon, the τ lepton is massive enough that a significant number of its decay modes include quarks in the final state.

The resulting control region events are binned the same for events in the search region. The estimate of this background for each search bin is performed by taking the sum of the control region events for each bin and multiplying it by a bin specific translation factor. This translation factor is derived from simulation. The translation factor is calculated for each bin by taking the ratio of the number of lost lepton and hadronic-τ events in the search region and the number single electron and single muon events in the control region. The simulated events are made to be as close to data as possible by correcting for mismodeling initial state (ISR) jet spectrum [73] and differences between data and simulation of b tagging efficiency [74], muon and electron identification [24], and isolation cut efficiency.

The translation factor method is validated in a side band region exclusive from the search bins. This side band region are bins defined as $N_{\mathrm{t}} = 0$, $N_{\mathrm{b}} \geq 2$, and $\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}}, j_{1,2,3,4}) > 0.5$. Figure 5.5 shows that the prediction of the translations factor method are consistent with data within the statistical uncertainties.

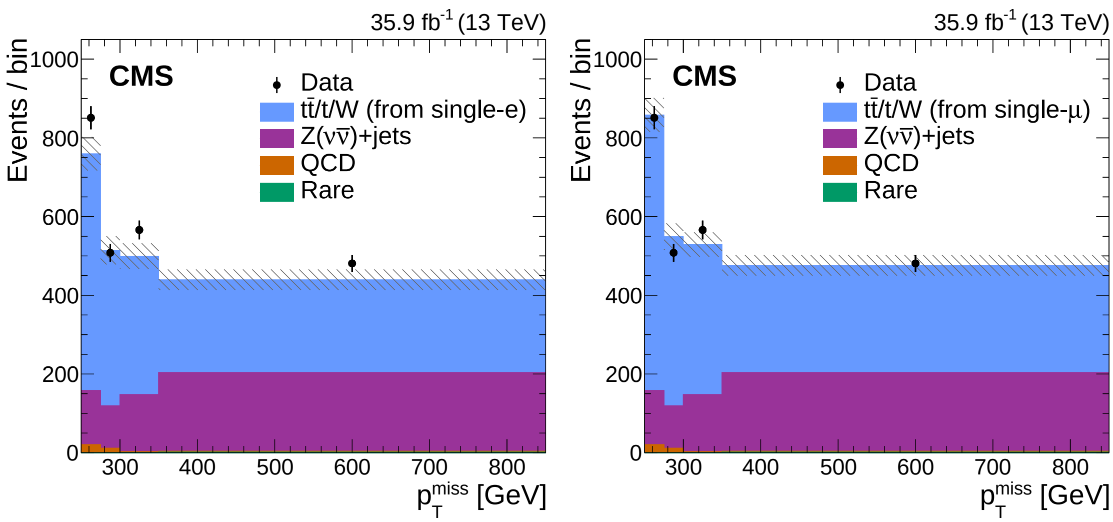


Figure 5.5: Distribution of $p_{\mathrm{T}}^{\mathrm{miss}}$ in the sideband data sample in comparison to predictions for SM processes. The prediction for $\mathrm{t\bar{t}}$, single top quark, and W+jets events is obtained using translation factors applied to a single-electron control sample (left) or to a single-muon control sample (right). The hatched bands indicate the statistical uncertainties in the total SM prediction. Note that the data and the predictions for all backgrounds except that for $\mathrm{t\bar{t}}$, single top quark, and W+jets events are identical between the left and right plots.

5.6.2 Events that include Z Bosons Decaying to Two Neutrinos

Z boson to two neutrino backgrounds are estimated from simulated $\mathrm{Z}(\nu\overline{\nu})$+jets events with corrections derived from a dimuon control region. The simulated background has two corrections applied. The first correction is shape correction to the N_{j} distribution. The second correction applies an overall normalization factor.

These correction factors are derived from the correction factors that would need to be applied to a similar Z boson to dimuon control region. Events from both observation and simulation are selected that have two oppositely charged muons. One muon must have $p_{\mathrm{T}} > 50\,\mathrm{GeV}$ in order to match the single muon trigger requirement that collects these events in data. The other muon must have a mass with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, and the dimuon invariant mass must be between 81 GeV and 101 GeV, i.e. within $\pm 10\,\mathrm{GeV}$ of the Z boson mass. Both simulated and observed events are then reconstructed as if the muons were neutrinos; producing samples of proxy $\mathrm{Z} \to \nu\overline{\nu}$

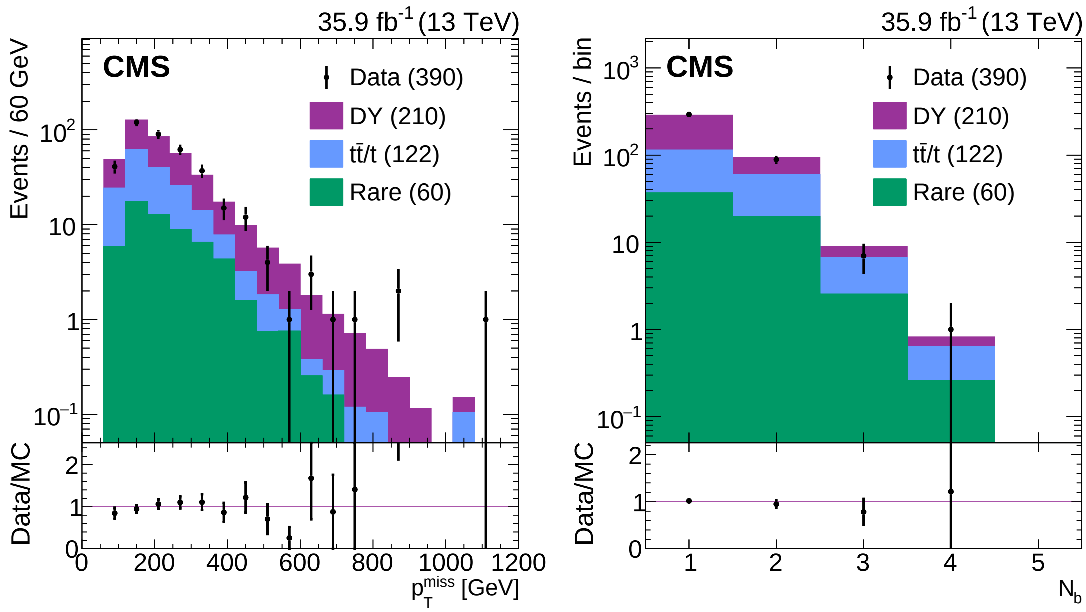


Figure 5.6: The $p_{\mathrm{T}}^{\mathrm{miss}}$ (left) and N_{b} (right) distributions of data and simulation in the loose dimuon control sample after applying a correction, as described in the text, to account for differences between the data and simulation for the N_{j} distribution. The lower panels show the ratio between data and simulation. Only statistical uncertainties are shown. The values in parentheses indicate the integrated yields for each component.

events. The procedure follows by deriving the corrections for the proxy simulated events to match the shape and normalization of the proxy observed events.

The N_{j} shape correction is derived in a "loose" region that has the same $\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}}, j_{1,2,3})$, H_{T}, and N_{t} requirements as the signal region, but has a less stringent $p_{\mathrm{T}}^{\mathrm{miss}} > 100\,\mathrm{GeV}$ requirement and with no requirement on N_{b}. The shape correction as a function of N_{j} is the ratio between the number of observed events in this control region (with the number of non Drell-Yan events subtracted using simulation) over the number of simulated Drell-Yan events. The agreement between the observed and simulated events in this dimuon control region as described is present in Fig. 5.6.

The second correction is to the overall normalization of the background estimate. This is done with the same dimuon samples as before, but with tighter criteria. This sample must meet all of the criteria applied to the signal regions except that there is no N_{b} requirement and only the isolated electron veto is applied. The normalization correction is the ratio of the observed number of events minus simulated non-Drell-Yan backgrounds over the number of simulated events.

5.6.3 QCD Multijet Events

The proton-proton collisions of hadron colliders frequently produce events with a large number of jets through QCD interactions. These events are called QCD multijet events. An event is a QCD multijet event if is produced only through QCD processes and does not belong to any of the other background event categorizations. A key characteristic of a QCD multijet event is that it usually does not contain any significant true $p_{\mathrm{T}}^{\mathrm{miss}}$. Hadronic jets may include some number of leptons in the shower, and so some neutrinos may be produced. These neutrinos will be soft, carrying little of the total momentum away. If such an event passes the $p_{\mathrm{T}}^{\mathrm{miss}}$ selection, this is because there has been some sort of misreconstruction of the event such as the mismeasurement of jet momenta.

This kind of mismeasurement is not usually significant, and only a small fraction of QCD multijet events pass the $p_{\mathrm{T}}^{\mathrm{miss}}$ requirement of this analysis. However, due to the large production rate of QCD multijet events, even a small fraction of such events results in a non-negligible background. As discussed in Section 5.5, this background is further suppressed by the requirement that $\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}}, j_1) > 0.5$, $\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}}, j_2) > 0.5$, and $\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}}, j_3) > 0.3$, where j_1, j_2, and j_3 are the three leading jets in the event.

The QCD multijet background which remains after the events selections is estimated using a combination of data and simulation. A data control region is defined using the same selection criteria for the search regions defined in Section 5.5; however, the $p_{\mathrm{T}}^{\mathrm{miss}}$ requirement is reduced to 200 GeV and the $\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}}, j_{1,2,3})$ requirement is inverted. The resulting control sample is predominately composed of QCD multijet events with $p_{\mathrm{T}}^{\mathrm{miss}}$ caused by jet mismeasurement. The sample is depleted of our expected signal events, and includes a small number of other standard model

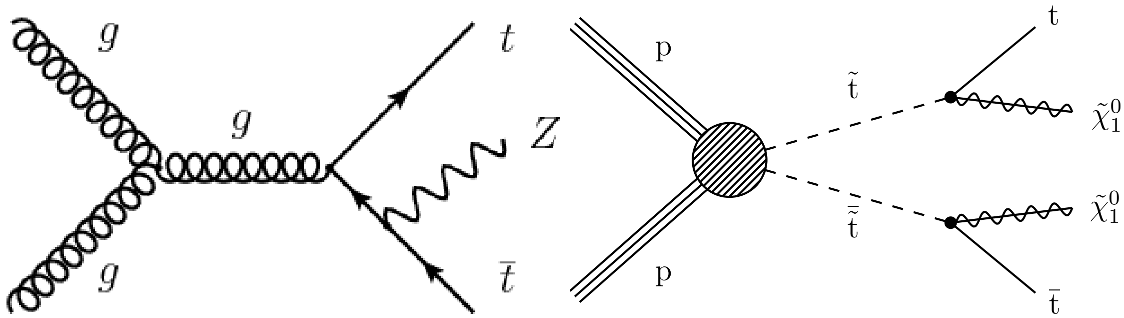


Figure 5.7: Comparison of a $t\bar{t}Z$ event and a SUSY event with the same signature. If the Z boson decays to two neutrinos then in both cases two top quarks are produced and non-detected particles carry away momentum.

background events. The contamination of this region by the other standard model backgrounds is removed by subtracted the estimated number of such events. These estimates are made using the methods of Section 5.6.1 as applied to this CR.

The QCD multijet background in each of the search bins is made by multiplying the yield in the control region by translation factors derived from simulation. The estimate is validated by comparing this data driven estimated background to predictions made by direct simulation of the QCD background. The difference between the two methods is added as a systematic uncertainty to final event yields in each search bin (varies from 30% to 500% depending on the search bin). Other sources of uncertainty from this background estimation include the statistical uncertainty of the simulation derived translation factors (varies from 30% to 300% depending on the search bin), and uncertainty due to the subtraction of the other standard model backgrounds from the data control region (varies from 2% to 50%).

5.6.4 A Note on Smaller Background Contributions

There are some smaller standard model backgrounds. The number of such events is small and have little effect on the final result. The estimation of such backgrounds is taken directly from simulation. These backgrounds include $t\bar{t}Z$, $t\bar{t}W$, WW, WZ, ZZ, WWW, WWZ, WZZ, and ZZZ. Of these, $t\bar{t}Z$ is largest of these small backgrounds.

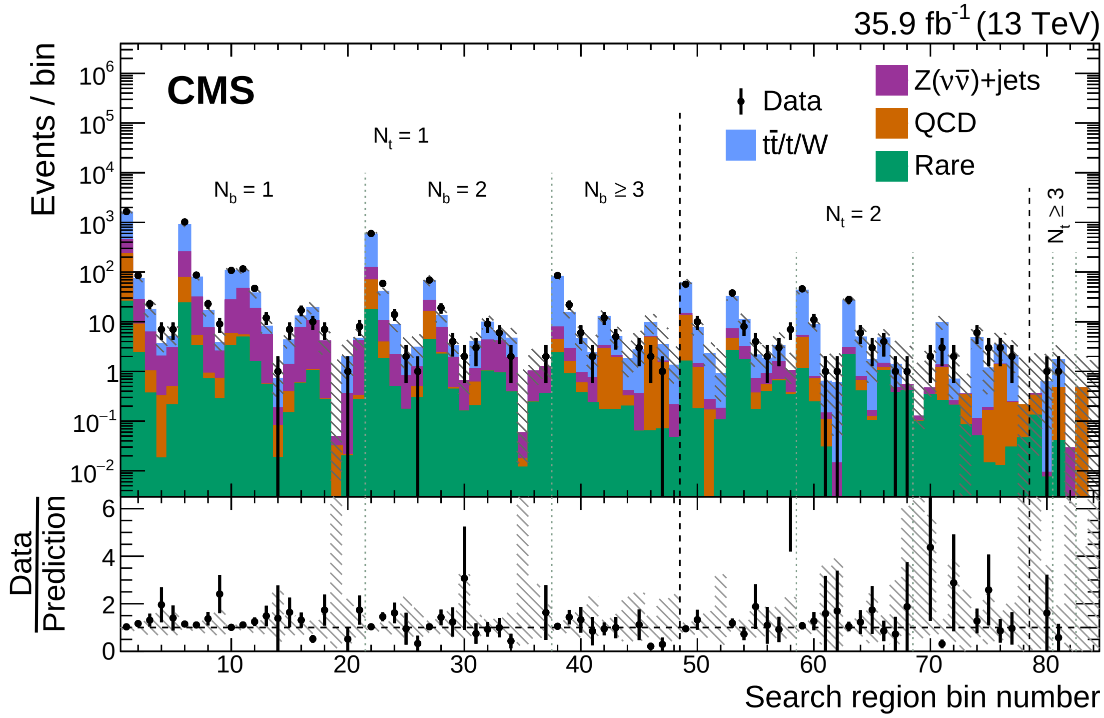


Figure 5.8: Observed event yields (black points) and prefit SM background predictions (filled solid areas) for the 84 search regions, where "prefit" means there is no constraint from the likelihood fit. The lower panel shows the ratio of the data to the total background prediction. The hatched bands correspond to the total uncertainty in the background prediction.

The $t\bar{t}Z$ background is particularly interesting because it looks very much like the signals of interest as seen in Figure 5.7.

5.7 Search Results and Interpretations

The number of observed events and the predicted number of SM background events in each of the 84 search regions are summarized in Fig. 5.8. Numerical values are given in Tables B.1–B.3 of Appendix B. The corresponding results for the aggregate search regions are presented in Fig. 5.9, with numerical values in Table B.4.. No statistically significant excess of events in data above the background predictions is observed. The largest source of background is typically the production of $t\bar{t}$ or W+jets events, followed by Z($\nu\overline{\nu}$)+jets production. The Z($\nu\overline{\nu}$)+jets background can

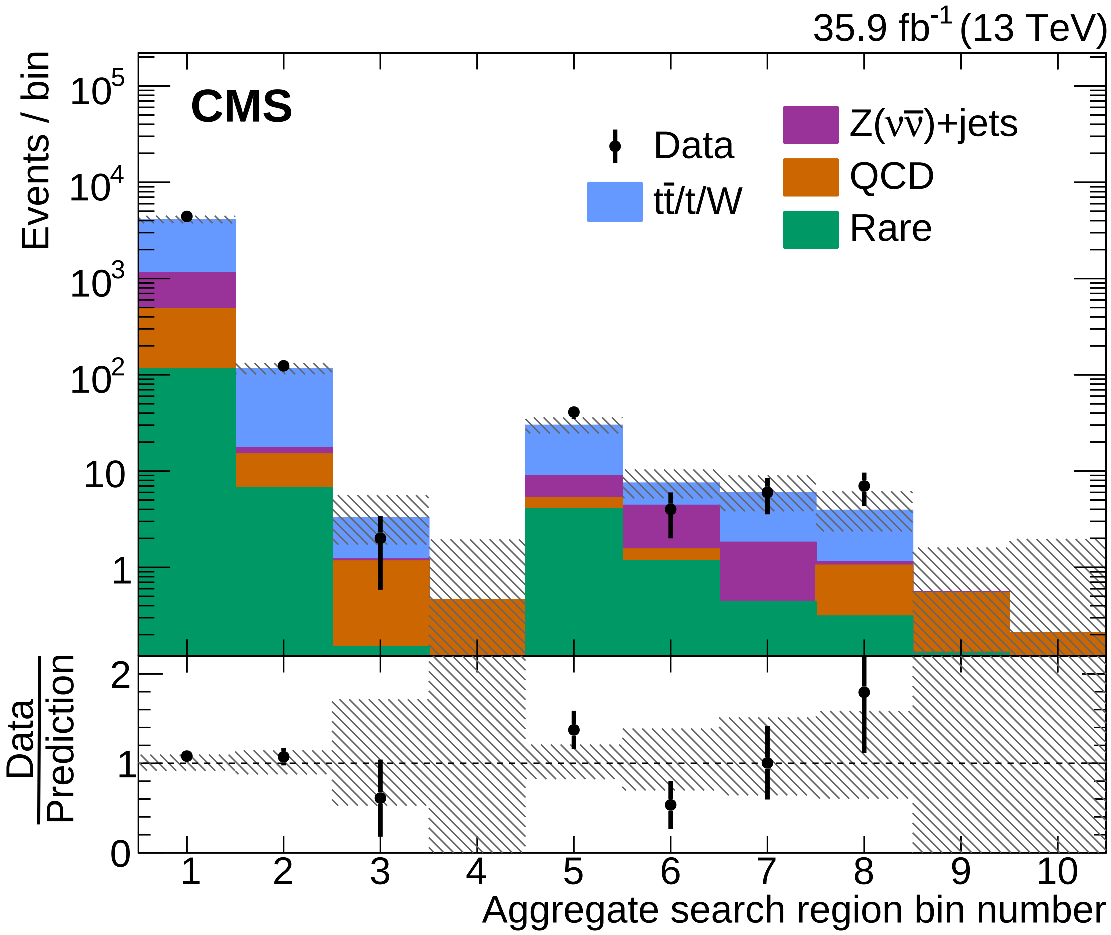


Figure 5.9: Observed event yields (black points) and prefit SM background predictions (filled solid areas) for the 10 aggregate search regions, where "prefit" means there is no constraint from the likelihood fit. The lower panel shows the ratio of the data to the total background prediction. The hatched bands correspond to the total uncertainty in the background prediction.

be dominant in search regions with a high $p_{\mathrm{T}}^{\mathrm{miss}}$ threshold where other backgrounds become highly suppressed because they lack a true source of $p_{\mathrm{T}}^{\mathrm{miss}}$. The contributions of the QCD multijet and rare backgrounds are small in all regions.

5.7.1 Setting Exclusion Limits

Exclusion limits are derived for the signal models of Section 5.2 using a binned likelihood fit to the data. The likelihood function is given by the product of Poisson probability density functions, one for each search region and for each of the corresponding regions of the single-electron, single-muon, and QCD data control samples, that account for the background predictions and signal yields. The uncertainties are treated as nuisance parameters with log-normal probability density functions. Correlations between search regions are taken into account. Upper limits at 95% confidence level (CL) on the SUSY production cross sections are calculated using a modified frequentist approach with the $\mathrm{CL_s}$ criterion [75, 76] and asymptotic results for the test statistic [77, 78]. Signal models for which the 95% CL upper limit on the production cross section falls below the theoretical cross section, based on next-to-leading order (NLO) plus next-to-leading logarithm (NLL) calculations [79], are considered to be excluded by the analysis.

The uncertainties in the signal modeling are determined individually for each search region and account for the following sources: the statistical uncertainty in the simulated event samples, the integrated luminosity (2.5% [80]), the lepton and isolated-track veto efficiencies (up to 6.8%), the b tagging efficiency (up to 21%), the trigger efficiency (up to 2.6%), the renormalization and factorization scales (up to 3.5%), the ISR modeling (up to 46%), the jet energy scale corrections (up to 34%), the top quark reconstruction efficiency (up to 14%), and the modeling of the fast simulation compared with the full simulation for top quark reconstruction and mistagging (up to 24%). All uncertainties except those from the statistical precision of the simulation are treated as fully correlated between search regions. Signal contamination is handled by using simulation to estimate the number of signal events that pass into the CR regions of the data driven background estimates. These events are subtracted from the signal yields, resulting in a "reduced" signal efficiency. Signal contamination is significant only for the single-lepton control samples used in the prediction of the $\mathrm{t\bar{t}}$ or W+jets backgrounds and is negligible for the dimuon and inverted-$\Delta\phi$ control samples used in the prediction of the $\mathrm{Z}(\nu\overline{\nu})$+jets and the QCD multijet backgrounds.

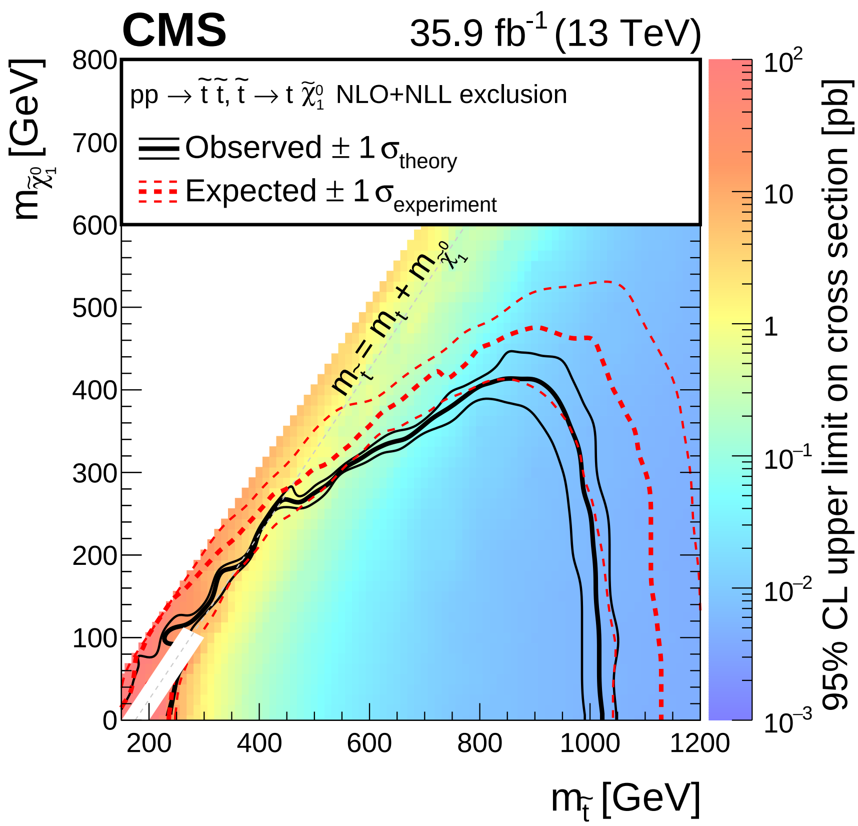


Figure 5.10: The 95% CL upper limit on the production cross section of the T2tt simplified model as a function of the top squark and LSP masses. The solid black curves represent the observed exclusion contour with respect to NLO+NLL signal cross sections and the change in this contour due to variation of these cross sections within their theoretical uncertainties. [79]. The dashed red curves indicate the mean expected exclusion contour and the region containing 68% of the distribution of expected exclusion limits under the background-only hypothesis. No interpretation is provided for signal models for which $|m_{\tilde{t}} - m_{\tilde{\chi}_1^0} - m_{t}| \leq 25\,\text{GeV}$ and $m_{\tilde{t}} \leq 275\,\text{GeV}$ because signal events are essentially indistinguishable from SM $t\bar{t}$ events in this region, rendering the signal event acceptance difficult to model.

Figure 5.10 shows the 95% confidence level exclusion limits obtained for the T2tt model of direct top squark pair production: top squark masses up to 1020 GeV and LSP masses up to 430 GeV are excluded.

The exclusion contour is the intersection of two surfaces. One surface is the NLO+NLL signal cross section as a function of the two SUSY parameters. In the case of this T2tt model, it is a function of the top squark and LSP masses. The other surface is the observed 95% confidence level upper limit as a function of the same two parameters. An area of the parameter space is excluded if the 95% CL when the upper limit is lower than the predicted NLO+NLL cross section.

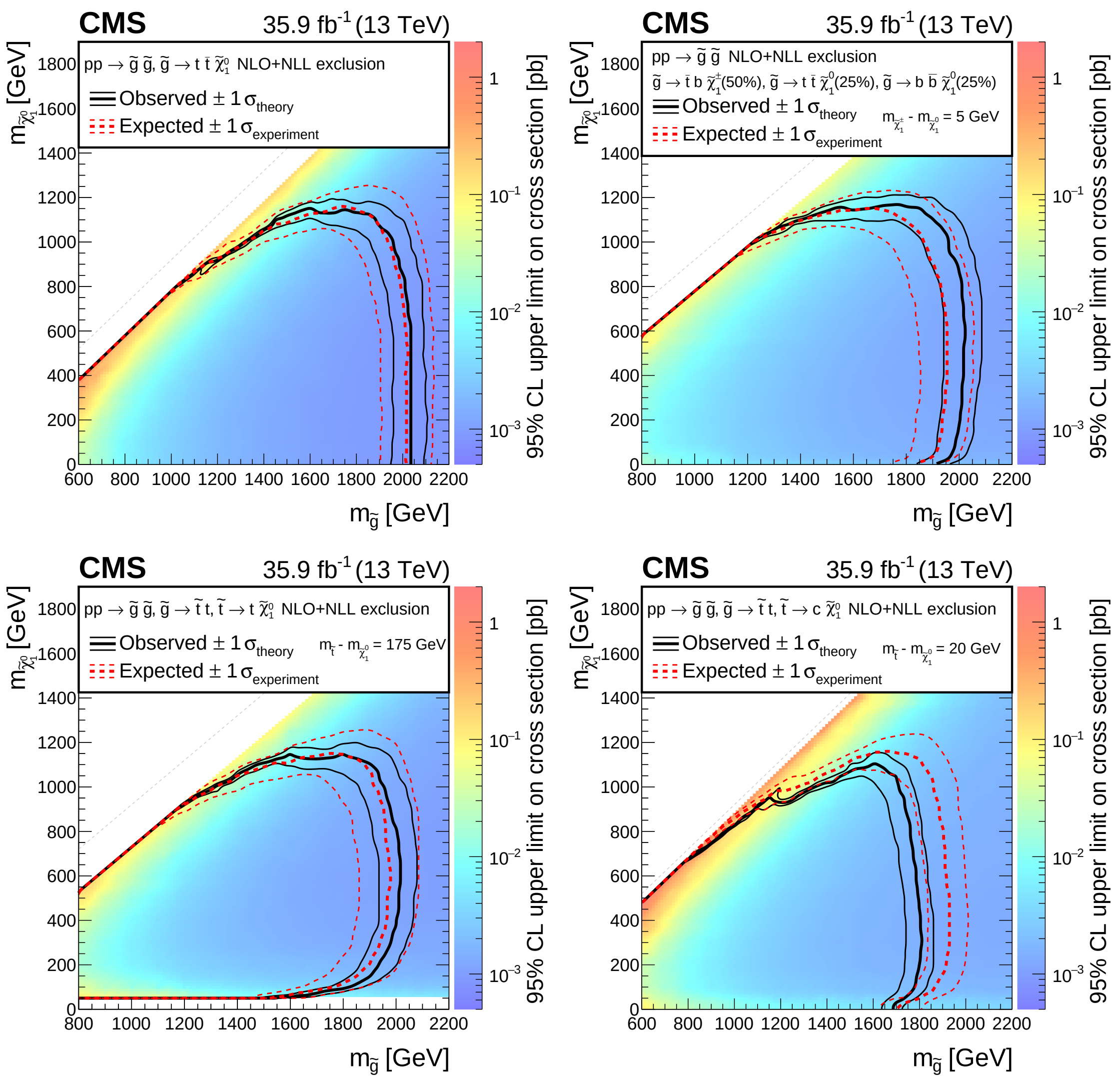


Figure 5.11: The 95% CL upper limit on the production cross section of the T1tttt (upper left), T1ttbb (upper right), T5tttt (bottom left), and T5ttcc (bottom right) simplified models as a function of the gluino and LSP masses. The meaning of the curves is explained in the Fig. 5.10 caption. Limits are not given for the T5tttt model for $m_{\tilde{\chi}_1^0} < 50\,\mathrm{GeV}$ for the reason explained in Appendix A.

The results for the four models of gluino pair production, T1tttt, T1ttbb, T5tttt, and T5ttcc, are shown in Fig. 5.11. Gluino masses up to 2040 GeV and LSP masses up to 1150 GeV are excluded for the T1tttt model, with corresponding limits of 2020 and 1150 GeV for the T1ttbb model, 2020 and 1150 GeV for the T5tttt model, and 1810 and 1100 GeV for the T5ttcc model. The limits on the gluino mass are somewhat lower for the T1ttbb model than for the T1tttt model because of the smaller average number of top quarks in T1ttbb signal events make it harder to distinguish them from background events. The lower limit of up to 2040 GeV obtained for the gluino mass in the T1tttt model improves the corresponding limits of Refs. [81,82] by around 100 GeV, while the limit on the gluino mass of up to 1810 GeV obtained for the T5ttcc model improves that presented in Ref. [83] by 560 GeV. The strategy of looking at all-hadronic events and using an effective top tagger to identify events with higher top multiplicity is an effective means to explore these particular MSSM models and is sensitive in general to new physics phenomena that produces top quarks. It is complementary to other searches that look at other signatures.

In the case of the T5tttt model, the exclusion curves significantly weaken as $m_{\widetilde{\chi}_1^0}$ approaches zero. This is a consequence of the kinematics of the $\widetilde{\mathrm{t}} \to \mathrm{t}\widetilde{\chi}_1^0$ decay. Only a small portion of the top squark momentum is transferred to the LSP if the LSP is lighter than the top quark. In such events there is a very small $p_{\mathrm{T}}^{\mathrm{miss}}$ resulting in few events passing the $p_{\mathrm{T}}^{\mathrm{miss}}$ selection. The number of events passing the selection criteria becomes comparable to signal contamination correction. Further information about the statistical consequence of these effects are presented in Appendix A. The statistical treatment becomes unreliable for for the T5tttt model when $m_{\widetilde{\chi}_1^0} < 50\,\mathrm{GeV}$, and thus the results for this region are not included

CHAPTER SIX

Deep Learning Algorithm for Resolved Top Quark Decays

6.1 Introduction

As discussed in Chapter Five, it is natural to look for evidence of top squark production through the excess of events containing top quarks. Depending on the specifics of the model, such top quarks may be highly boosted, or not boosted much at all. Identifying top quarks in an event depends on how highly boosted it is. At higher momenta, the decay products of the top quark will be collimated into a narrower region and are likely to be clustered together. Algorithms which look at the substructure of wide jets are effective. At lower momenta, the decay products do not resolved into a single jet, but resolve more cleanly into two or three jets as depicted in Fig. 6.1.

When the top quark decays into products which are resolved into three separate jets, it presents a particularly challenging case. A combinatorial explosion of top candidates occurs when events with a large number of jets must be tested for the presence of one or more top quarks. As the number of possible combination of jets within an event increases, the likelihood of misidentifying random jets as the decay of a top quark increases dramatically. Figure 6.2 shows as an example four possible variables that could be used to distinguish a top candidate that corresponds to a true top quark, and a candidate that does not. It is important to recognize that the distributions for these variables are distinct between top candidates that correspond to real top quarks, and top candidates that do not correspond to top quarks. The great challenge lies in that there are no simple cuts on these variables that will produce both good top quark identification yields and little misidentification. In aggregate we can see how the distributions differ; however, for any particular candidate, the

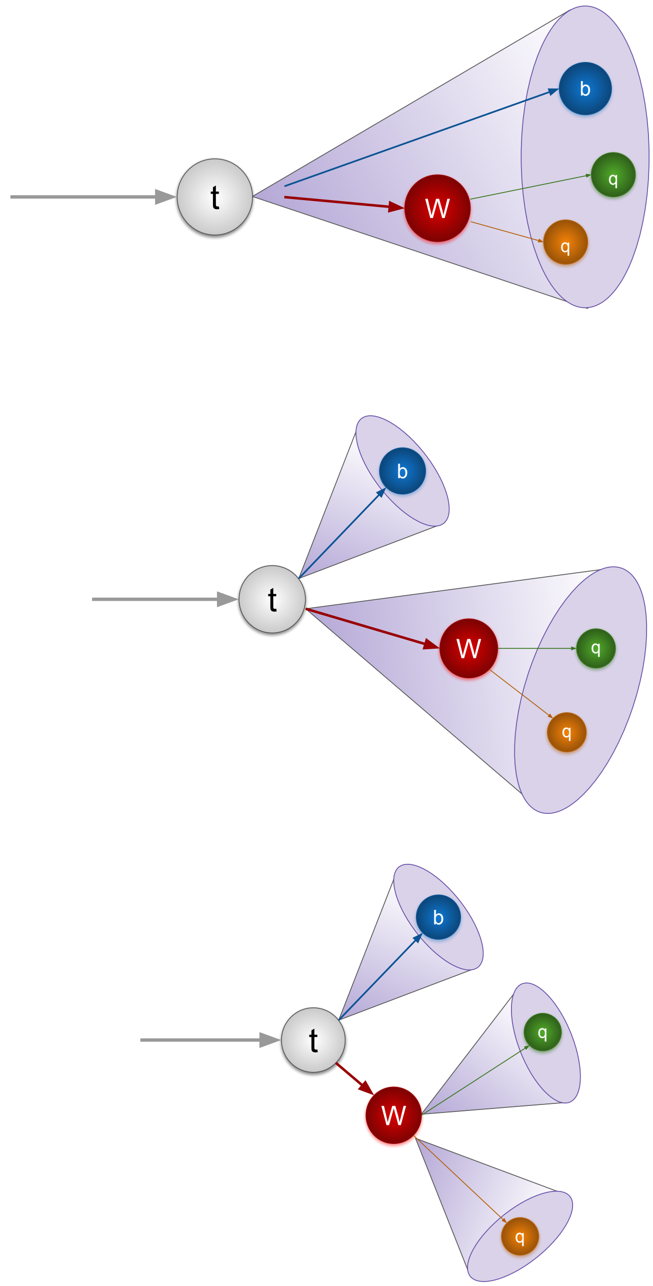


Figure 6.1: Signatures of an all-hadronic top quark decay. The branching ratio for a top quark to decay to a W boson and a bottom quark is about 90%. It is an all-hadronic decay if the W boson decays to two quarks. Depending on how highly boosted the top quark is, the decay products may be collimated in a single wide cone (top), the the W boson products may be collimated in a single wide cone and the b quark jet resolved in a separate cone (middle), or the jets of all three quarks may be resolved separately (bottom).

relationship between the variable values and proper classification is complex and non-obvious.

Resolved taggers have been used before. In the analysis presented in Chapter Five, a BDT based resolved top quark tagger was combined with a merged tagger that used jet substructure and a simple dijet tagger that looked for objects consistent with the top quark mass. The innovation of the currently described resolved top quark tagging algorithm is to consider many more top candidate variables than before. The previously used MVA techniques cannot effectively handle such an increase in input variables. This tagging algorithm uses a dense neural network (DNN) implemented in TensorFlow [84]. Figure 6.3 shows a schematic of such a neural network.

6.2 Variables

The inputs to the the resolved tagger algorithm are divided into two categories: The jet variables used to describe each jet individually, and candidate variables that describe the entire top quark candidate or combinations of multiple jets. In total, the tagger algorithm accepts 82 input variables; 25 variables for each jet, and 7 variables for the candidate as a whole, as summarized in Table 6.1.

6.2.1 Candidate Variables

The top candidate variables are fairly straightforward and represent a fairly basic description of the candidate itself. The invariant mass of the top candidate is a powerful, but incomplete, identifier of the top quark. This algorithm makes no assumptions about the identity of the three jets and treats each symmetrically. Thus the dijet mass and dijet θ separation of each pair of jets is considered. Notice that a dijet invariant mass near the W boson mass is another powerful, but incomplete, identifier of the top quark.

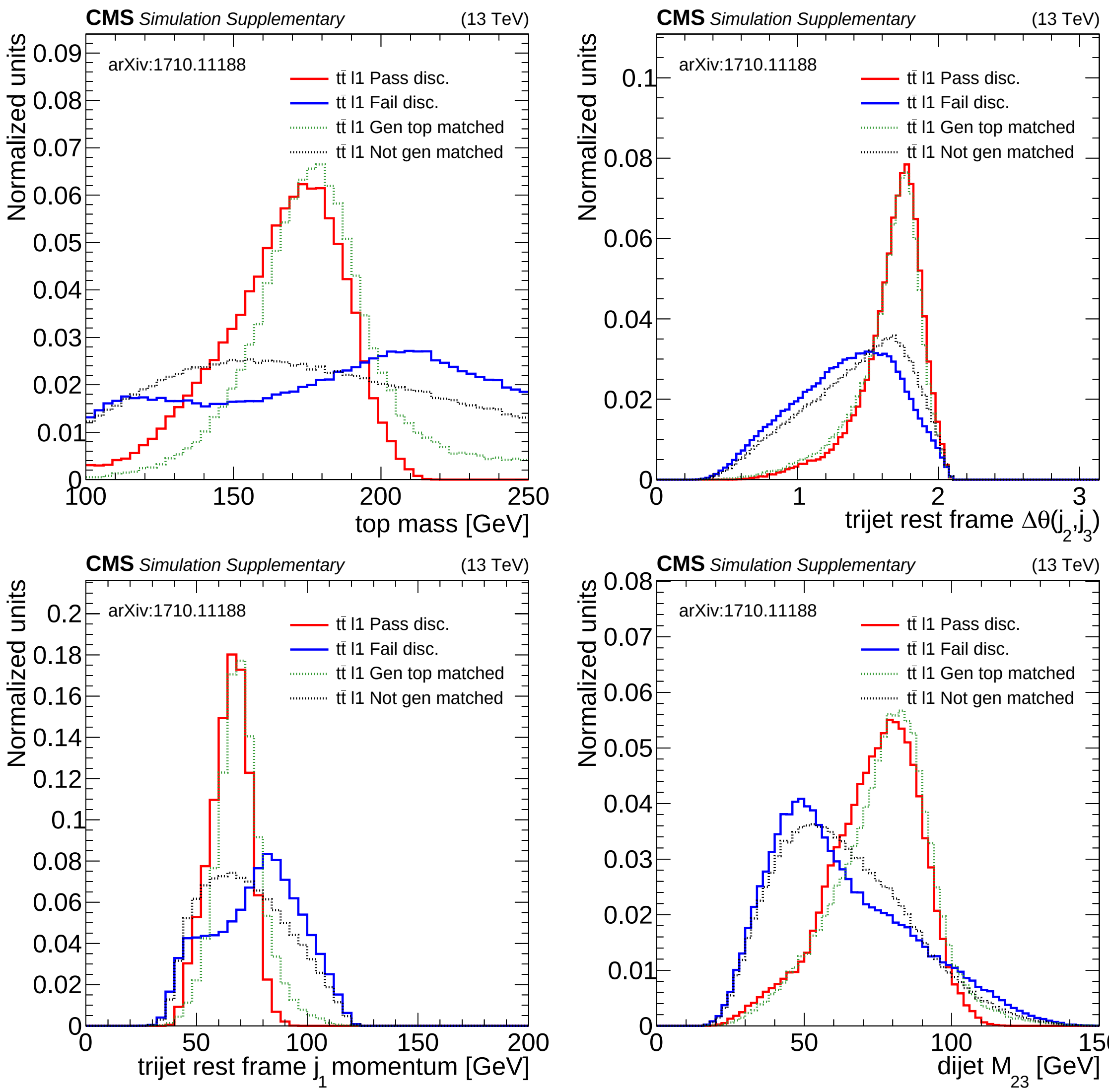


Figure 6.2: Top quark candidate variables. The variables that describe the top quark candidate have distinct distributions. However, there are no simple cuts on the variables that will both minimize misidentification and yield good top quark identification.

Table 6.1: Variables used by the resolved top quark identification algorithm. There are two classes of variables: variables used to describe each of the three jets (jet variables), and variables that describe the top quark candidate or combinations of multiple jets (candidate variables). Variables used by the tagger described in Chapter Five are marked BDT. Variables used by the tagger described in this chapter are marked TF.

Variable	Description	BDT	TF
	Candidate-level variables		
m_{cand}	Invariant mass of the top candidate	✓	✓
m_{j12}	Invariant mass of jets 1 and 2 combination.	✓	✓
m_{j13}	Invariant mass of jets 1 and 3 combination.	✓	✓
m_{j23}	Invariant mass of jets 2 and 3 combination.	✓	✓
$\Delta\theta_{12}$	Transverse angle of jets 1 and 2		✓
$\Delta\theta_{13}$	Transverse angle of jets 1 and 3		✓
$\Delta\theta_{23}$	Transverse angle of jets 2 and 3		✓
	Jet-level variables		
m_{jet}	Invariant mass of the jet	✓	✓
p_{jet}	Momentum of the jet		✓
CSV	B-tag discriminator	✓	
CvsL	Charm vs. light discriminator	✓	
DeepCSVb	Component of the DeepCSV discriminator		✓
DeepCSVbb	Component of the DeepCSV discriminator		✓
DeepCSVc	Component of the DeepCSV discriminator		✓
DeepCSVcc	Component of the DeepCSV discriminator		✓
DeepCSVl	Component of the DeepCSV discriminator		✓
qgAxis1	Component of the quark-gluon discriminator	✓	✓
qgAxis2	Component of the quark-gluon discriminator		✓
qgMult	Component of the quark-gluon discriminator	✓	✓
qgPtD	Component of the quark-gluon discriminator	✓	✓
ChargedHadronMultiplicity	Number of PF charged hadrons		✓
NeutralHadronMultiplicty	Number of PF neutral hadrons		✓
PhotonMultiplicity	Number of PF photons		✓
ElectronMultiplicty	Number of PF electrons		✓
MuonMultiplicty	Number of PF muons		✓
HFEMEnergyFraction	Fraction of jet energy from HF EM		✓
HFHadronEnergyFraction	Fraction of jet energy from HF Hadron		✓
ElectronEnergyFraction	Fraction of jet energy from electrons		✓
PhotonEnergyFraction	Fraction of jet energy from photons		✓
ChargedEMEnergyFraction	Fraction of jet energy from charged EM particles		✓
ChargedHadronEnergyFraction	Fraction of jet energy from charged hadrons		✓
MuonEnergyFraction	Fraction of jet energy from muons		✓
NeutralEMenergyFraction	Fraction of jet energy from neutral EM particles		✓
NeutralHadronEnergyFraction	Fraction of jet energy from neutral hadrons		✓

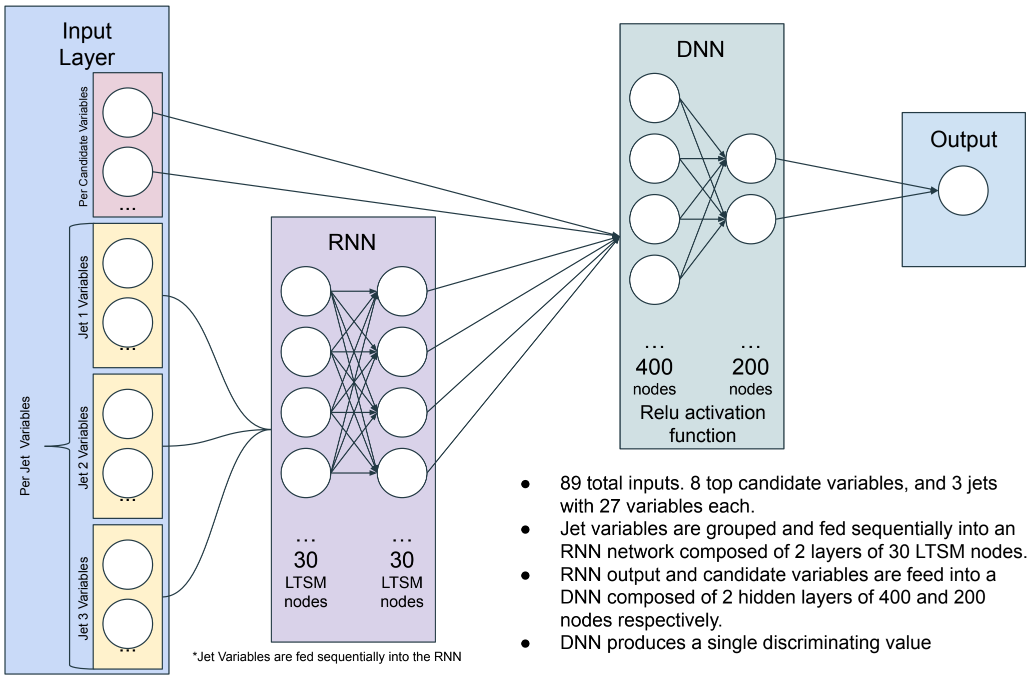


Figure 6.3: An example of the topology of a neural network used to classify hadronically decaying top quarks. It features an recurrent neural network (RNN) to combine jet level variables, the results of the RNN and object level variables are fed into a DNN.

6.2.2 Jet Variables

The algorithm gains much of its additional discriminating power by evaluating and combining information about the individual jets. This includes results from the b-tagging algorithm, the quark-gluon discriminator, and information about the particle content of the jet.

6.3 Quantitative Evaluation of the Algorithm

The output of the neural network, called the discriminator, is a continuous variable between zero and one. In a sense, it is a description of how "top-like" the candidate is. An ideal tagger would always return zero every non-top candidate and one for every candidate that truly corresponds to a top quark. The algorithm makes a final decision about the identity of the top quark candidate by comparing the discriminator produced to some set value.

Figure 6.4 shows the spectrum of discriminator values in a simulated sample that by construction contains exactly one top quark that decays all-hadronically. Specifically, these are simulated $t\bar{t}$ events in which one top quark decays to a bottom quark and two light quarks and the other top quark decays to a bottom quark, a charged lepton, and a neutrino. These simulated tops are not guaranteed to decay to three resolved jets, but many do. Figure 6.5 shows the spectrum of discriminator values in a simulated sample that by construction contains no hadronically decaying top quarks. Specifically, these are simulated $t\bar{t}$ events in which both top quarks decay to a bottom quark, a charged lepton, and a neutrino. Note that there is a sharp peak near zero, and there is broad representation over the rest of the discriminator values. The spectra peaks again near one. In this particular sample, there are two top quarks present. The b quarks produced by the top quark decays make it fairly easy to produce a candidate that looks very similar to a hadronically decaying top quark. However, notice that the spectrum in Fig. 6.5 peaks short of the peak in Fig. 6.4. A discriminator value is chosen that simultaneously maximizes top quark identification while minimizing misidentification.

A Receiver Operating Characteristic (ROC) curve is a plot that graphically represents the classifying power of a system [85]. As the discriminator is varied, the true prediction rate (TPR) and fake prediction (FPR) are calculated and the curve is generated. A perfect classifier will have an 100% true prediction rate and an 0% fake prediction rate for any discriminator values between its extrema. The ideal is never achieved, but the relative strength of two classifiers can be evaluated by looking at how closely the ROC curve approaches this ideal. Figure 6.6 shows an example of ROC curves. This is an example taken from studies into the the most effective

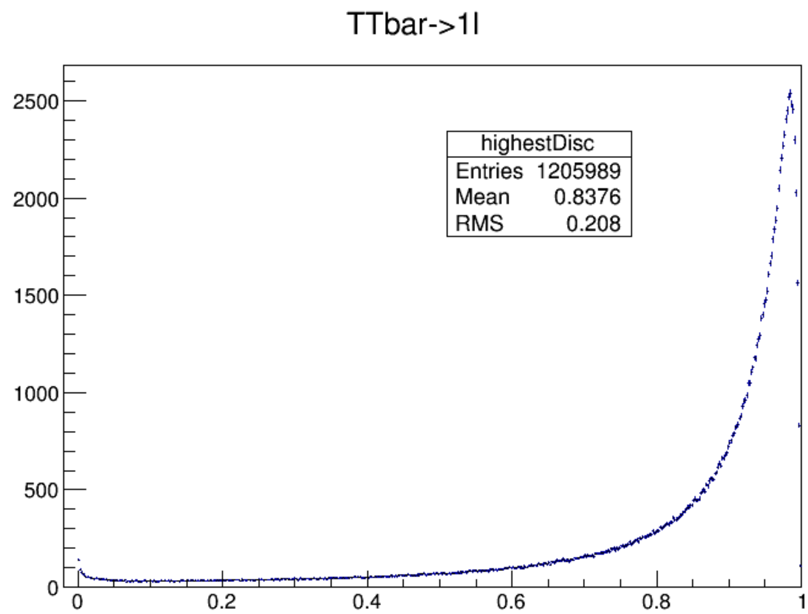


Figure 6.4: Spectrum of discriminator values in a simulated semi-leptonic $t\bar{t}$ sample. Every event contains exactly one top quark that decays all-hadronically. Although not every event will contain a top quark that decays to three resolved jets, many do. Note that there is a tight peak near one.

network structure. In this particular example the number of hidden layers and nodes per layer is varied.

This was part of an extensive campaign to identify the most effective network structure, training procedures, and choice of input variables. The relative strength of each classifier was taken to be the integral under the ROC curve. The configuration of the neural network is controlled through configuration files which are easily varied programmatically. Hundreds of different configurations were tested and automatically assessed.

6.4 Training

The network is trained primarily on MC events with some data events. The training of the network requires truth information about the top candidates presented to it. We can justify the use some training from data events by carefully choosing a

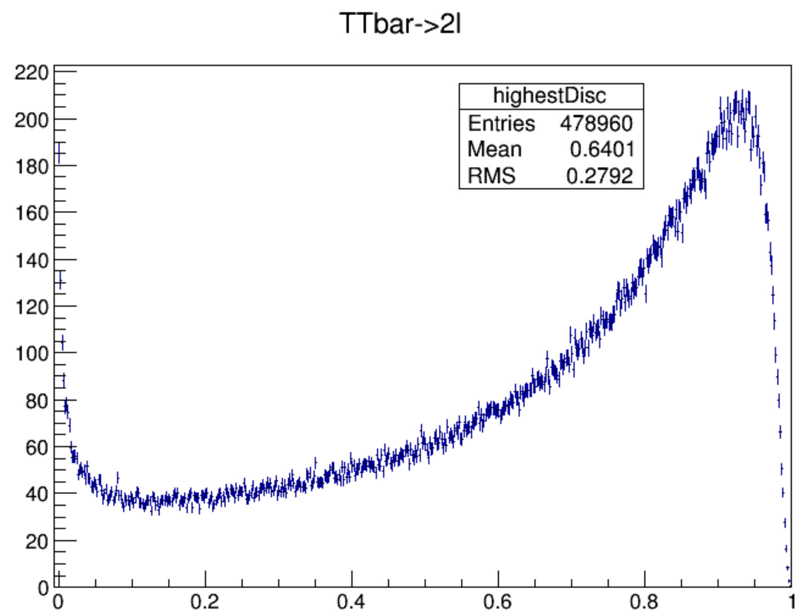


Figure 6.5: Spectrum of discriminator values in a simulated dilepton $t\bar{t}$ sample. There are no true resolved hadronically decaying top quarks in this sample. Note that this spectrum has a sharp peak near zero, and is widely spread over the rest of the values.

control region that is depleted of top quarks such that the top candidates formed from such events are highly likely not to be top quarks. Training a neural network is robust against some degree of impurity of the object labels, and some light amount impurity may even aid in the training process because it helps protect against overtraining as described in [86].

During validation of the tagger, we discovered that the performance of the tagger in events dominated by the direct production of jets from gluon hadronization differed in MC and data. Upon investigation we determined that this was likely due to a type of overtraining in the model. Specifically, the tagger was picking up artifacts of the MC that were not present in data. Two strategies were employed to minimize this difference. First, the tagger was retrained with additional candidates taken directly from data in a control region carefully defined to remove events with top quarks. This control region is defined in Table 6.2. The simulated events are compared to events

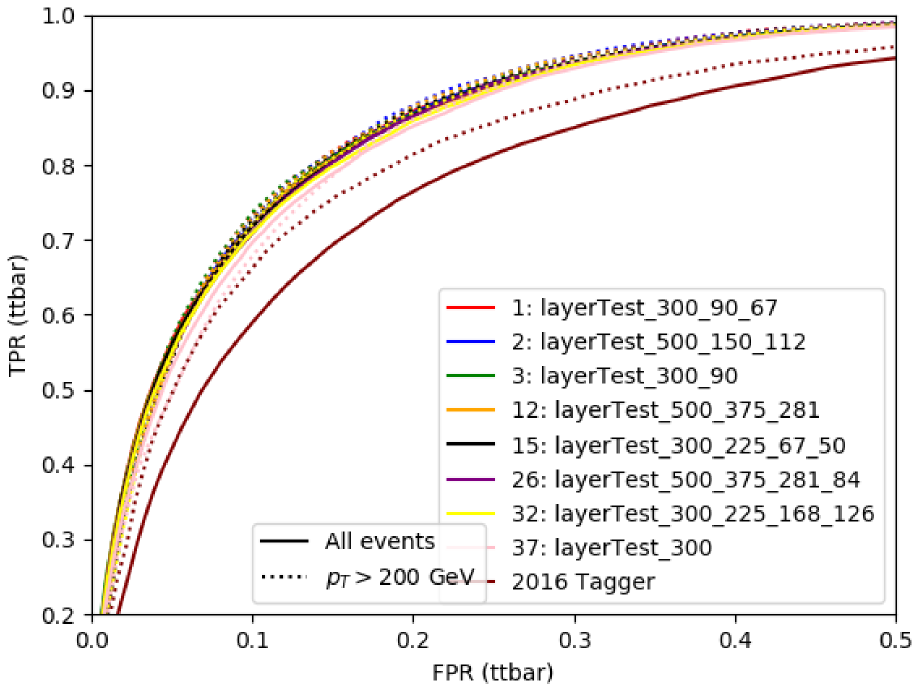


Figure 6.6: ROC curves for several different topologies. A broad swath of input variables and network topologies were investigated. This figure shows the ROC curves for various networks that have been varied by the number of DNN layers and number of nodes in each layer.

Table 6.2: Definitions of the QCD control region (QCD CR) designed to be used for the top quark tagger misidentification rates. The control region is defined to be depleted in $t\bar{t}$ events, so most of the events should come from QCD multijet processes.

QCD CR
At least 4 jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$
Lepton Veto
$H_{\mathrm{T}} > 1000\,\mathrm{GeV}$

collected by the detector and included in the JetHT dataset. The H_{T} requirement is present to match the trigger conditions that produce this dataset.

The second strategy was to employ a method of domain adaptation first described in [86].

6.4.1 Domain Adaptation

QCD multijet events are particularly challenging because the simulated events are not as well modeled as other types of interactions. Domain adaptation is a modification of the fitness function such that not only is correctly identifying the object rewarded, but behavior that recognizes the differences in object domain is penalized. Specifically, the model was still trained using the truth information present in the MC samples, but it was also presented with samples taken from data, and penalized for being able to discern the MC samples from the data samples.

6.5 Validation

The tagger is validated in regions orthogonal to the search regions expected in a all-hadronic search. The tagger efficiency is evaluated in a region enriched in $\mathrm{t\bar{t}}$ events in which one of the top quarks decays to a b quark, a lepton, and a neutrino (semi-leptonic $\mathrm{t\bar{t}}$ CR which is defined in Table 6.3). Figure 6.7 shows the agreement between simulation and data in this control region. Based on the simulation results, we expect that over 80% of the events in this CR will be semileptonic $\mathrm{t\bar{t}}$ events. The tagger fake rate is evaluated in a region depleted in events containing top quarks, and selected to be dominated by jets produced by QCD multijet processes. Figure 6.8 shows the agreement between simulation and data in this control region. Based on simulation, we expect that less than 1% of the events in this control region contain a top quark.

Validation of the tagger is performed by comparing the tagging rate in simulation and data. Tagging efficiency is modeled by the tagging rate in the $\mathrm{t\bar{t}}$ CR. The tagging rate is defined as the number of events that contain a tagged top candidate over the total number of events. Under the assumption that each event contains a resolved top quark, this tagging rate can be interpreted as a top tagging efficiency. Figure 6.9 shows this tagging efficiency as a function of the top candidate p_{T}. In this case, the working point has been chosen to give an average 5% misidentification rate.

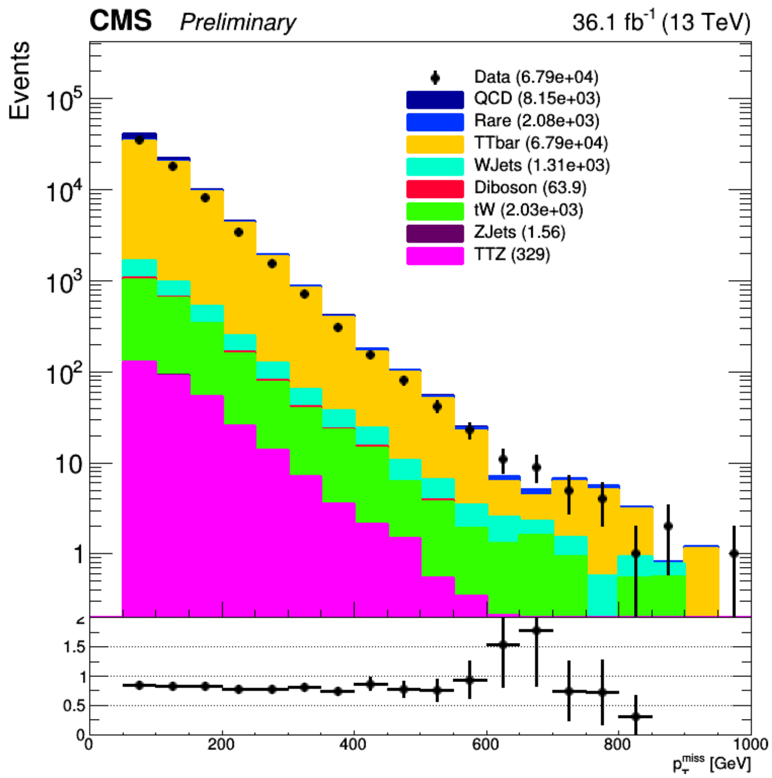


Figure 6.7: The $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution of events in the $\mathrm{t\bar{t}}$ CR for both observed and simulated events. This CR is designed to be enriched in $\mathrm{t\bar{t}}$ events that decay semileptonically. It compares observed events gathered using the single muon triggers to simulated events. This region is used to validate tagging efficiency. The region is made up of over 80% semileptonic $\mathrm{t\bar{t}}$ events.

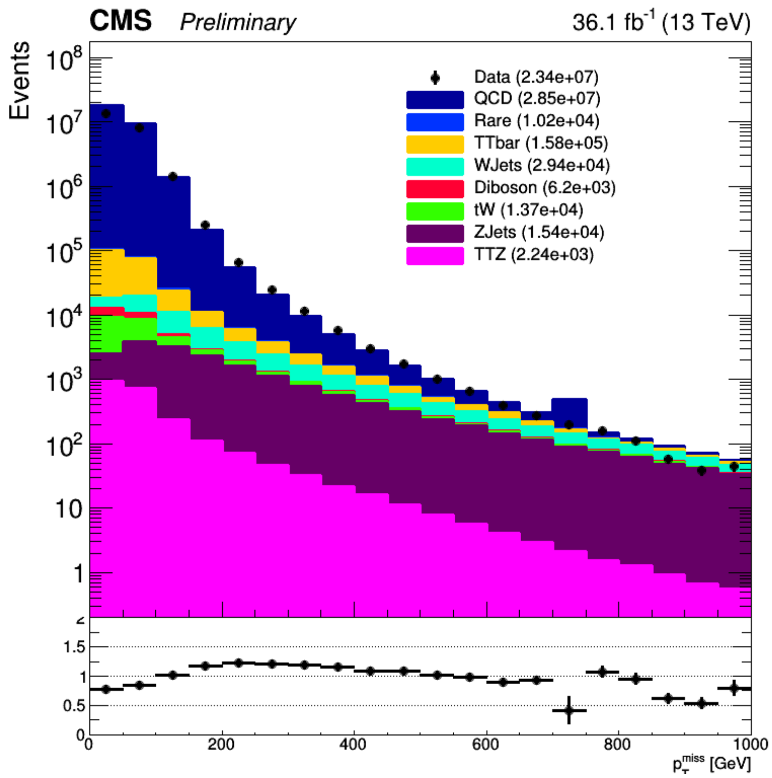


Figure 6.8: The $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution of events in the QCD CR for both observed and simulated events. This CR is designed to be depleted in $\mathrm{t\bar{t}}$ events. It compares observed events gathered using the JetHT triggers to simulated events. This region is used to validate the top misidentification mistagging. Less than 1% of the events in this control region contain a top.

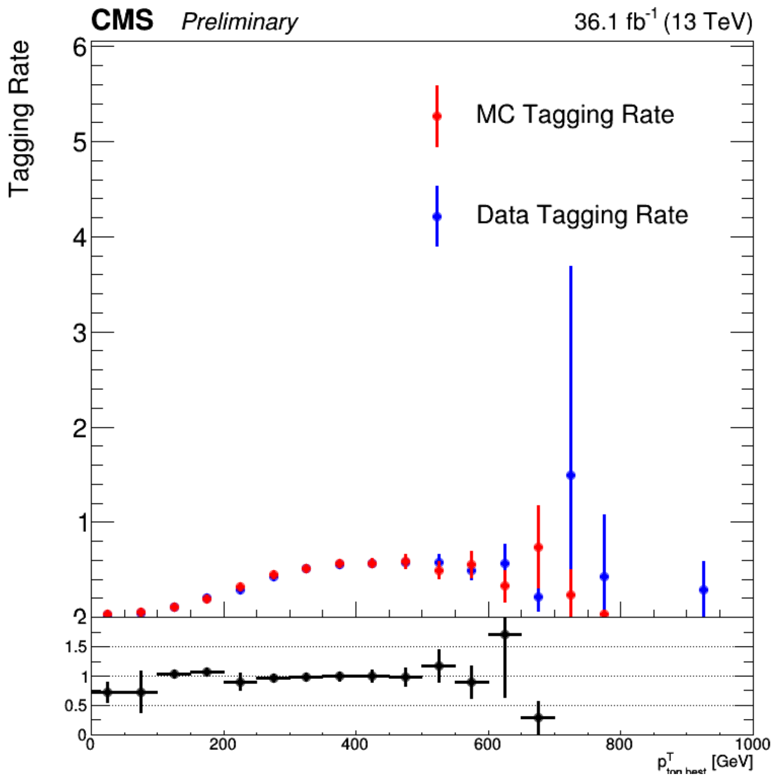


Figure 6.9: Top quark identification rate for both observed and simulated events. Tagging efficiency for each top p_{T} bin is taken to be the number of events with an identified top divided by the total number of events. The ratio of the tagging rates is also provided. The working point was chosen to give an overall 5% mistag rate.

Table 6.3: Definitions of the $t\bar{t}$ control region ($t\bar{t}$ CR or sometimes also referred to as semileptonic $t\bar{t}$ CR) designed to be used for the top quark tagger identification efficiency studies. The control region is defined to be enriched in $t\bar{t}$ events that decays semi-leptonically. The $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement is present to match the trigger conditions that produce data for this CR.

Semileptonic $t\bar{t}$ CR
At least 4 jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $\lvert\eta\rvert < 2.4$
A single lepton with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$
At least 1 b-tagged jet
$\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}}, j1) > 0.5$,
$\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}}, j2) > 0.5$,
$\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}}, j3) > 0.3$
At least one jet with $\Delta R(j, \mathrm{lepton}) < 1.5$
At least one jet such that $30\,\mathrm{GeV} < m(\mathrm{jet{+}lepton}) < 180\,\mathrm{GeV}$
$\Delta\phi(\mathrm{lepton}, E_{\mathrm{T}}^{\mathrm{miss}}) < 0.8$
$m_{\mathrm{T}} < 100\,\mathrm{GeV}$
$H_{\mathrm{T}} > 250\,\mathrm{GeV}$
$E_{\mathrm{T}}^{\mathrm{miss}} > 250\,\mathrm{GeV}$

The misidentification rate is modeled by the tagging rate in the QCD CR. Again, the tagging rate is defined as the number of events that contain a tagged top candidate over the total number of events. Under the assumption that no event contains a true resolved top quark, this tagging rate can be interpreted as a misidentification rate. Figure 6.10 shows this misidentification rate as a function of the top candidate p_{T}. In this case, the working point has been chosen to give an average 5% misidentification rate.

6.6 Results

Figure 6.11 shows that the ROC curve for the resolved top tagger presented here is an improvement over previous results. This offers encouragement that this approach will improve the sensitivity of future analyses that use identified top quarks. The improved ROC curves mean that for a particular acceptable level of misidentification, the identification rate of top quarks is increased. Or alternatively, if the

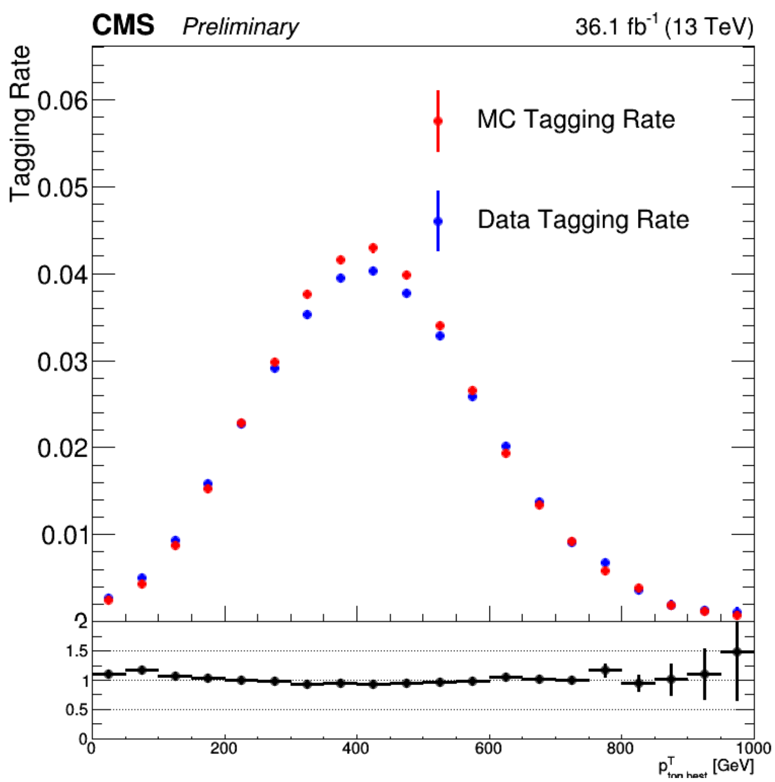


Figure 6.10: Top quark misidentification rate for both observed and simulated events. The top quark misidentification rate is taken to be the number of events with an identified top divided by the total number of events. This shows the mistag rate in both data and MC, and plots the ratio of the tagging rates. The working point was chosen to give an overall 5% mistag rate.

the identification rate is held constant, the number of misidentified top quarks is decreased. Validation of the top identification algorithm shows reasonable agreement between data and MC.

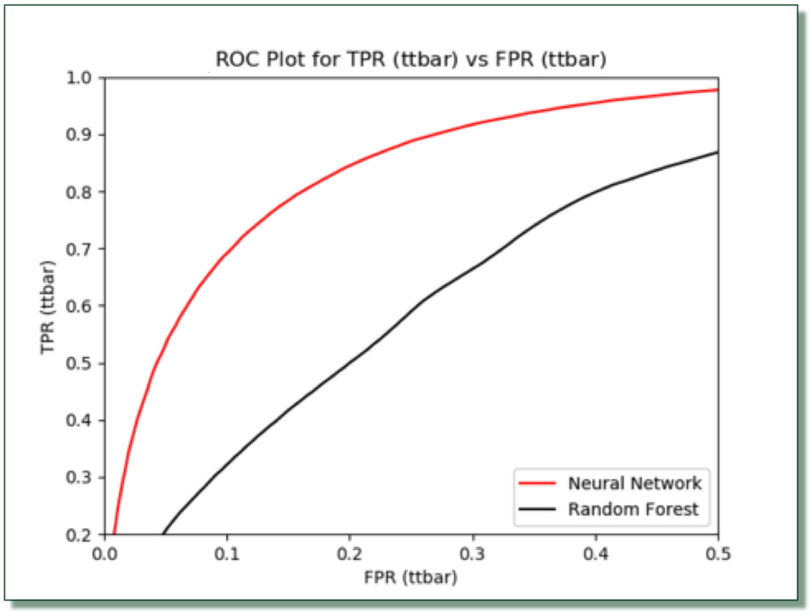


Figure 6.11: Performance of the neural network based top quark tagger compared to the previous BDT based tagger.

CHAPTER SEVEN

Conclusions

Top quark production and identification at the LHC remains an active topic of investigation. I have presented the results of search that looked at events with all-hadronic final states using data collected in 2016 at a center of mass energy of 13 TeV. This data sample collected by the CMS detector represents collisions corresponding to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$.

The search was designed to be sensitive to beyond standard model physics that produce top and bottom quarks and unreconstructed transverse momentum. Specifically, the analysis looks for direct and gluino-mediated top squark production in R-parity conserving SUSY. Event selection criteria were designed to reduce the standard model backgrounds relative to this class of signal events. The remaining standard model backgrounds were estimated through various data-driven methods.

No significant deviations from the standard model predictions were observed. In the context of the MSSM models presented, 95% confidence level upper-bounds on the observed production cross-sections and NLO+NLL calculations of the lower-bounds of the production cross sections allow us to exclude top squark masses up to 1020 GeV and LSP masses up to 430 GeV in the direct top quark pair production model. For models of gluino-mediated top squark production, gluinos with masses up to 1810 to 2040 GeV and LSPs with masses up to 1100 to 1150 GeV are excluded.

Accurate identification of top quark production provides significant power in this type of search. Future searches will benefit from more powerful and accurate top quark identification methods. I have also presented a new resolved top tagging algorithm based on a neural network that shows improved performance over the previously used top quark identification algorithms.

APPENDICES

APPENDIX A

T5tttt Interpretation

As seen in Fig. A.1 and also as discussed in Section 5.7, the exclusion limits for the T5tttt_DM175 model weaken substantially at low $m(\widetilde{\chi}_1^0)$ values. This T5tttt_DM175 model is the same as the T5tttt model discussed in Chapter Five. DM175 emphasizes that in this simplified model $\Delta m(\widetilde{\mathrm{t}}, \widetilde{\chi}_1^0)$ is fixed to 175 GeV. As $m(\widetilde{\chi}_1^0)$ approaches 0 GeV, the top squark mass follows. These lighter top squarks will be more highly boosted, and when it decays ($\widetilde{\mathrm{t}} \to \mathrm{t}\widetilde{\chi}_1^0$), the top quark will carry larger fraction of the top squark momentum than the lighter $\widetilde{\chi}_1^0$. In such scenarios, the $p_{\mathrm{T}}^{\mathrm{miss}}$ spectrum is much softer. There are two consequences of this softer $p_{\mathrm{T}}^{\mathrm{miss}}$ spectrum that weaken the limits in this region:

- The signal acceptance reduces, especially for all-hadronic signal events. Because of the softer $p_{\mathrm{T}}^{\mathrm{miss}}$ spectra, fewer signal events satisfy the $p_{\mathrm{T}}^{\mathrm{miss}}$ selection.
- Signal events with leptons are affected less. In signal events that include decays to leptons, part of the $p_{\mathrm{T}}^{\mathrm{miss}}$ comes from the neutrinos produced.

This latter effect influences background prediction in two ways.

- Increased signal contamination (i.e. signal events that fall in the 1-lepton control samples used for the top/W background estimation)
- Increased relative fraction of “lost-lepton” signal events among total signal events that fall into search regions

In order to gain some insight to these effects and understand the strength of our analysis in the context of these model regions, limits were computed for several signal mass points with different configurations as shown in Table A.1. The limits are reported in term of the parameter μ, the signal strength. In the context of limit

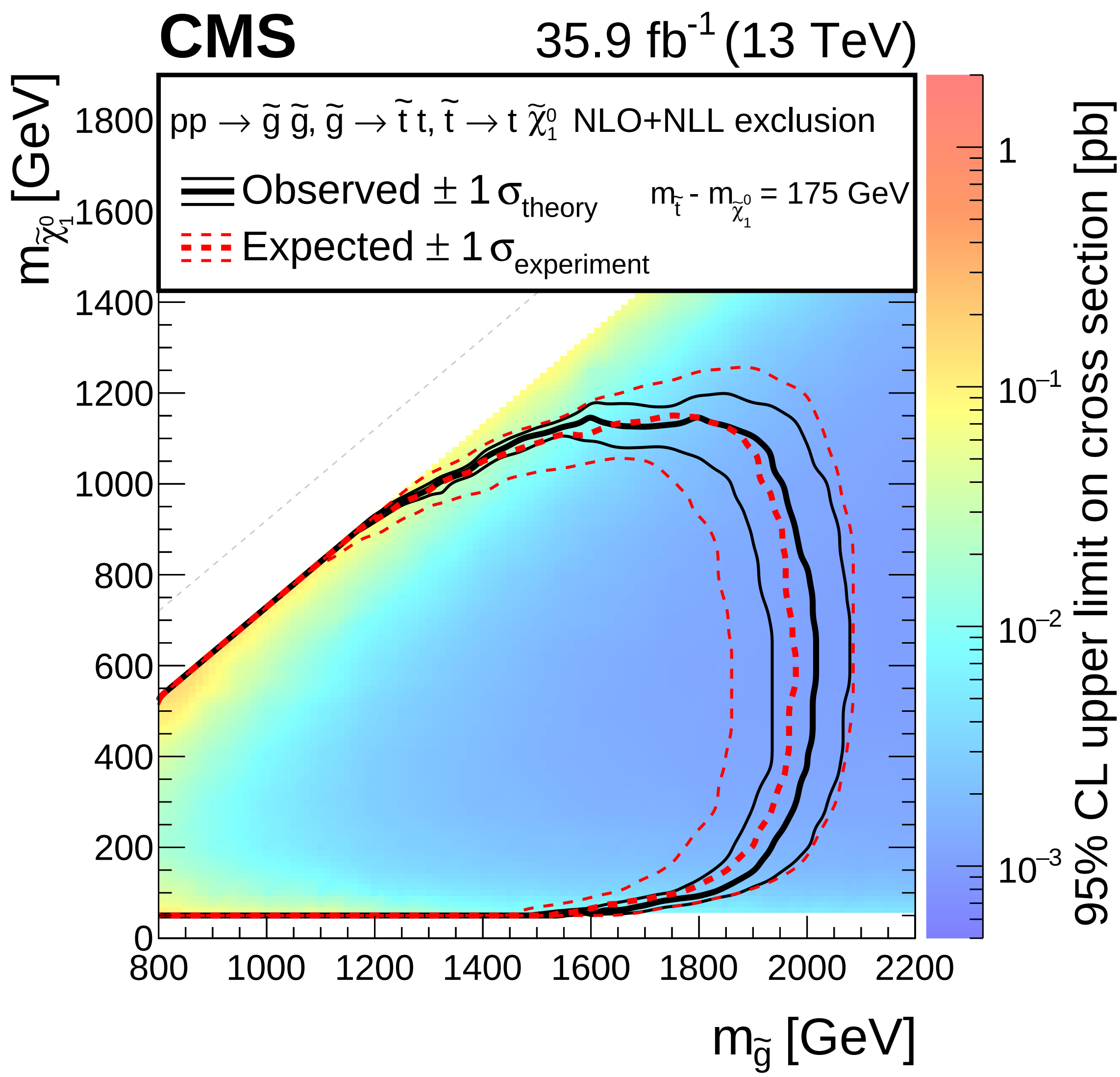


Figure A.1: The 95% CL upper limit on the production cross section of the T5tttt simplified model as a function of the gluino and LSP masses. The solid black curves represent the observed exclusion contour with respect to NLO+NLL signal cross sections and the change in this contour due to variation of these cross sections within their theoretical uncertainties. [79]. No exclusion is set for $m(\tilde{\chi}^0_1) \leq 50\,\mathrm{GeV}$.

setting for a particular model, models are excluded when $\mu < 1$. Some observations of this table include:

- Signal contamination subtraction has large effects. Small variations of the model in this part of the model parameter space means that a significant, but relatively poorly estimated, number of signal events start to leak into the control region. This washes out the statistical significance of the data driven estimates of the standard model background based on these control regions.
- Although the analysis is designed primarily to search for all-hadronic signal events, a non-negligible number of signal events come from "lost-lepton" events where a top quark has decayed leptonically but the charged lepton was not reconstructed. Selecting only hadronic signal events at the gen-level level weakens the limits because it fails to include these events that will contribute to the total number of events observed in data.
- Two $p_{\mathrm{T}}^{\mathrm{miss}}$ values, gen-$p_{\mathrm{T}}^{\mathrm{miss}}$ vs reco-$p_{\mathrm{T}}^{\mathrm{miss}}$, make only minor differences. The default results fall in between them. The specific conclusion here is that the imperfections of the $p_{\mathrm{T}}^{\mathrm{miss}}$ measurement based on the particle flow reconstruction does not have a major effect on the interpretation of these signal models.

These observations are discussed more in detail below.

The size of the signal contamination with respect to the signal yields for the analysis baseline selection in the case of the T5tttt_DM175 signal can be found in Fig. A.2. The size of the signal contamination increases rapidly as the $m(\widetilde{\chi}_1^0)$ approaches zero. We also studied the size of the signal contamination with respect to the signal yields as a function of search bins, which are shown in Fig. A.3. In the case of $m(\widetilde{\chi}_1^0) = 1\,\mathrm{GeV}$, there are high signal yield bins where signal contamination exceeds the signal yield, while the signal yield is larger than the signal contamination in most of bins for $m(\widetilde{\chi}_1^0) = 50$ and $100\,\mathrm{GeV}$. The signal contamination being larger than

Table A.1: T5tttt limits on the signal strength μ based on $35.9\,\mathrm{fb}^{-1}$ of data with different settings. Default uses the average signal yields computed based on the reconstructed $p_{\mathrm{T}}^{\mathrm{miss}}$ (reco-$p_{\mathrm{T}}^{\mathrm{miss}}$) and generator-level $p_{\mathrm{T}}^{\mathrm{miss}}$ (gen-$p_{\mathrm{T}}^{\mathrm{miss}}$). The numbers in the gen-$p_{\mathrm{T}}^{\mathrm{miss}}$ and reco-$p_{\mathrm{T}}^{\mathrm{miss}}$ columns are the signal yields computed using the respective choices of $p_{\mathrm{T}}^{\mathrm{miss}}$. The numbers in the "No signal contamination" column are computed without the subtraction of signal contamination, i.e. the signal events with a lepton that contaminate the estimation of the top/W background. The numbers in the "Gen hadronic" columns are obtained by using only the all-hadronic signal events by vetoing the electron and muon from the W or τ decays at the generator-level.

T5tttt model $(m(\widetilde{\mathrm{g}}), m(\widetilde{\mathrm{t}}))$ GeV	Default	gen-$p_{\mathrm{T}}^{\mathrm{miss}}$	reco-$p_{\mathrm{T}}^{\mathrm{miss}}$	No signal contamination	Gen hadronic	Gen hadronic gen-$p_{\mathrm{T}}^{\mathrm{miss}}$
(1200,1)	0.66	0.76	0.57	0.21	4.89	7.46
(1200,50)	0.25	0.30	0.21	0.14	0.46	0.75
(1200,100)	0.08	0.09	0.07	0.06	0.10	0.13
(1400,1)	1.32	1.33	1.26	0.52	12.00	16.44
(1400,50)	0.44	0.48	0.40	0.31	0.86	1.32
(1400,100)	0.19	0.20	0.18	0.15	0.24	0.30
(1700,1)	3.67	4.08	3.39	2.01	40.12	50.25
(1700,50)	1.63	1.82	1.52	1.15	2.66	3.80
(1700,100)	0.74	0.77	0.70	0.62	0.95	1.18

the signal yield hits the limit of the way how the signal contamination is treated in our statistical analyses for interpretations based on the "reduced efficiency" method. Refer back to Section 5.7.1 for a description of how exclusion limits are set. The "reduced efficiency" method is the practice of subtracting the simulated contamination from the model's signal yields. If the simulated contamination is comparable to the expected signal yield, then the limit setting procedure no longer works.

This led us to show the T5tttt_DM175 interpretations only for $m(\widetilde{\chi}_1^0) \geq 50\,\mathrm{GeV}$. This complication may be eliminated in the future analysis if we consider the signal contributions in both signal and control regions independently in the statistical interpretation.

The composition of the signal was also studied for this model. The fraction of signal events after baseline which are all-hadronic top decays at the generator-level

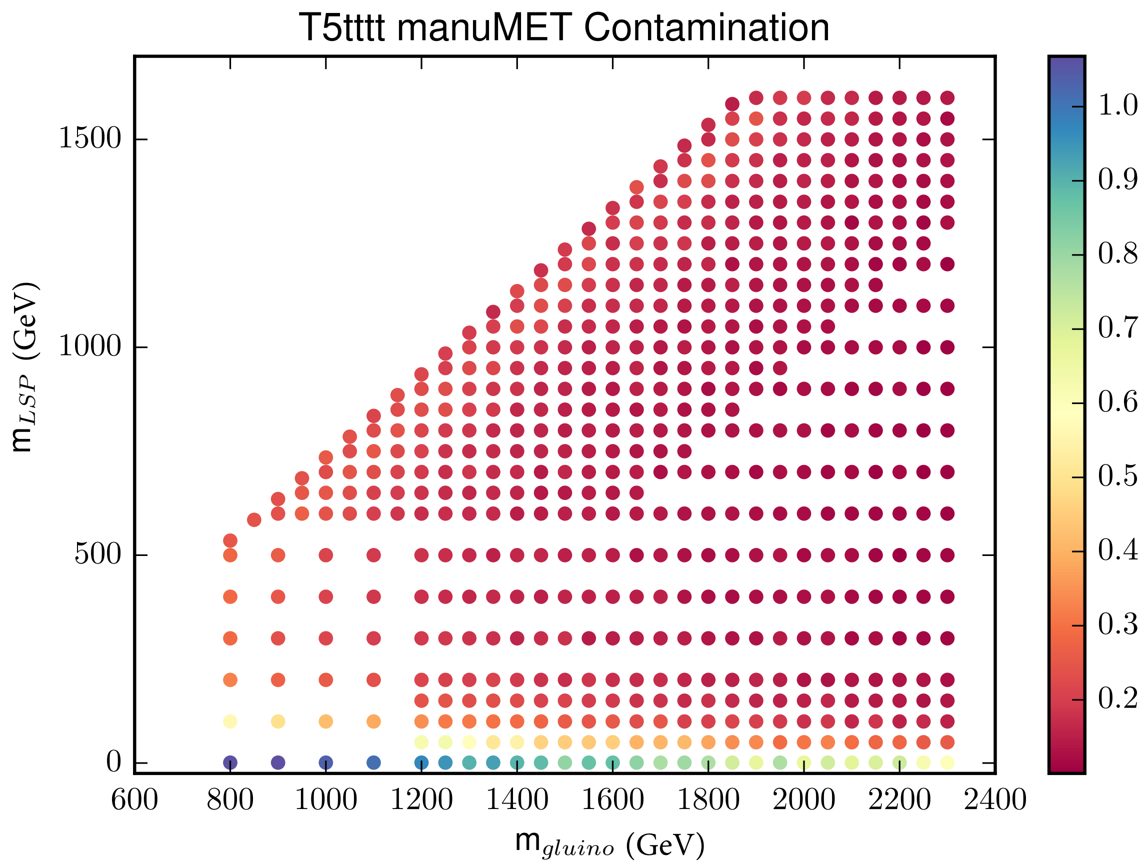


Figure A.2: Proportion of signal contamination in the T5tttt_DM175 model space. The ratio is the signal contamination yield with respect to the signal yields for the analysis baseline selection. The ratio is relatively insignificant except for m_{LSP} close to 0 GeV

are shown in Fig. A.4. The fraction of all-hadronic events is about 60% overall, i.e. 40% of the signal events have leptons that are lost either due to out-of-acceptance or mis-reconstruction/mis-identification/non-isolation. This doesn't really cause an overlap with leptonic SUSY searches, as leptonic searches are unlikely to use these "lost-lepton" signal events. It is interesting to note that this fraction of lost-lepton signal events increases very rapidly toward low $m(\widetilde{\chi}^0_1)$.

We also studied the impact of potential $p_{\mathrm{T}}^{\mathrm{miss}}$ mis-measurement in signal events simulated with fastsim by comparing gen-$p_{\mathrm{T}}^{\mathrm{miss}}$ and reco-$p_{\mathrm{T}}^{\mathrm{miss}}$. In general, gen-$p_{\mathrm{T}}^{\mathrm{miss}}$ and reco-$p_{\mathrm{T}}^{\mathrm{miss}}$ are close (see Fig. A.4), and the fraction of T5tttt signal events with

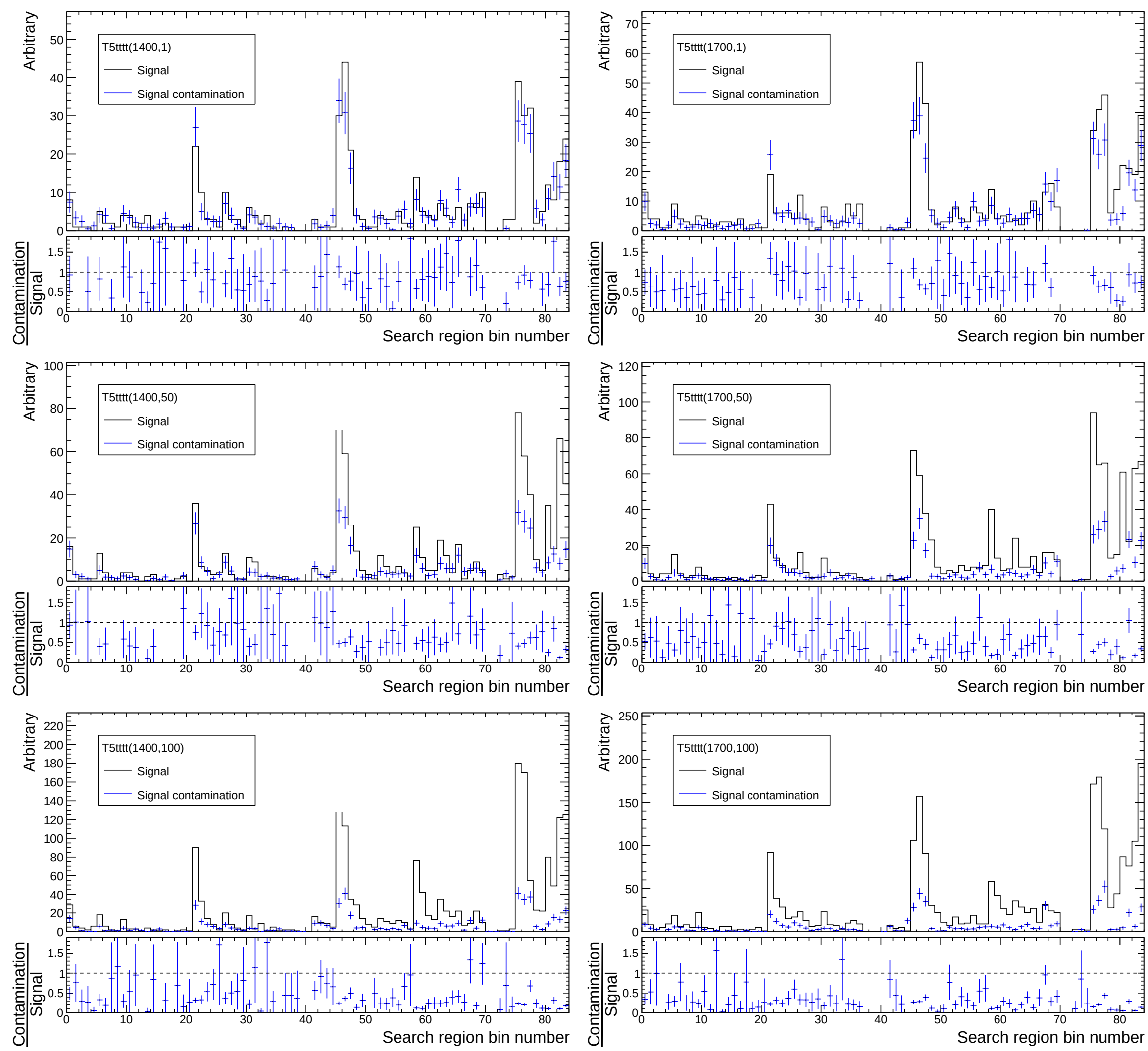


Figure A.3: The size of the signal contamination with respect to the signal yields for the analysis baseline selection in the case of the T5tttt_DM175 signal for six different mass points: [left] (1400,1), (1400,50), (1400,100), [right] (1700,1), (1700,50), (1700,100) (GeV).

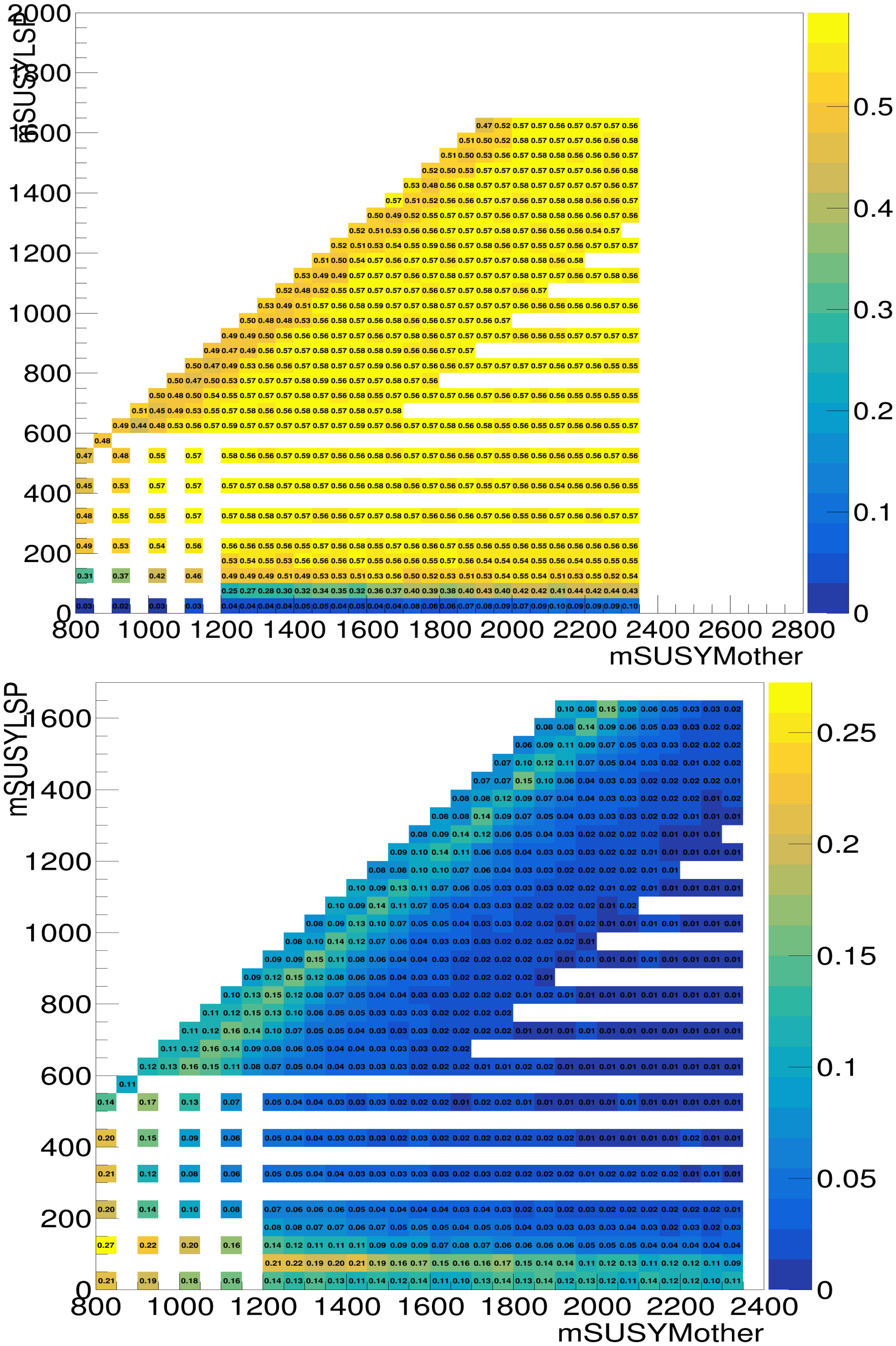
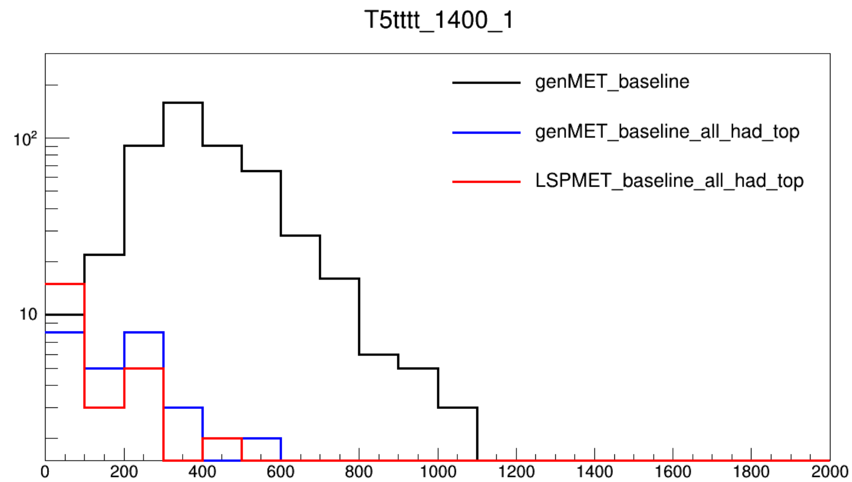


Figure A.4: Fraction of T5tttt signal events from either all-hadronic top decays of with mismatched gen-level and reco-level $p_{\mathrm{T}}^{\mathrm{miss}}$ as functions of the gluino and LSP masses in units of GeV. (Top) The fraction of T5tttt signal events after baseline which are all-hadronic top decays at the generator-level. Notice that for m_{LSP} close to 0 GeV, the fraction of gen-level all-hadronic events is reduced significantly. (Bottom) The fraction of T5tttt signal events with reco-$p_{\mathrm{T}}^{\mathrm{miss}} > 250$ GeV but gen-$p_{\mathrm{T}}^{\mathrm{miss}} < 250$ GeV. There is some $p_{\mathrm{T}}^{\mathrm{miss}}$ mismeasurement which becomes more pronounced at low m_{LSP}; however, studies showed that this didn't have a significant effect on limit setting in this region compared to other effects.

reco-$p_{\mathrm{T}}^{\mathrm{miss}} > 250\,\mathrm{GeV}$ but gen-$p_{\mathrm{T}}^{\mathrm{miss}} < 250\,\mathrm{GeV}$ is only around 10%; however, this fraction increases somewhat to $\sim 20\%$ toward low $m(\widetilde{\chi}_1^0)$. The gen-$p_{\mathrm{T}}^{\mathrm{miss}}$ distribution was studied after baseline selections are applied in signal events, which can be seen in Fig. A.5. There is some spill-over of events with gen-$p_{\mathrm{T}}^{\mathrm{miss}} < 250\,\mathrm{GeV}$ which is expected due to intrinsic $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution. The contribution of all-hadronic signal events at the generator-level is rather small. And in all-hadronic events, sizable $p_{\mathrm{T}}^{\mathrm{miss}}$ contributions come from non-$\widetilde{\chi}_1^0$ sources (e.g. neutrinos from b decays).



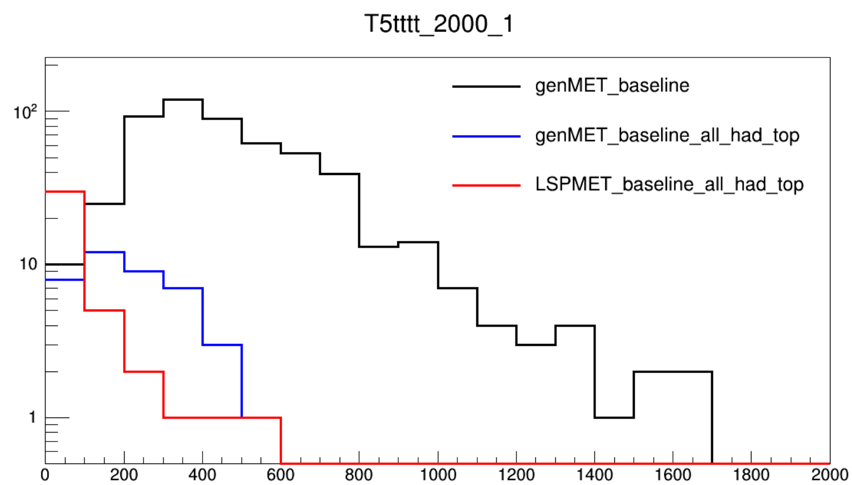


Figure A.5: The gen-level $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution in the T5tttt signal event for the signal mass points $m(\widetilde{\chi}_1^0) = 1\,\mathrm{GeV}$ and $m(\widetilde{\mathrm{g}}) = 1400\,\mathrm{GeV}$ and $m(\widetilde{\mathrm{g}}) = 2000\,\mathrm{GeV}$. The vertical axis is the number of events, and the horizontal axis is the event $p_{\mathrm{T}}^{\mathrm{miss}}$ in units of GeV. (Black) all signal events after baseline; (Blue) all-hadronic events at the generator-level; and (Red) the vector-sum of $\widetilde{\chi}_1^0$ p_{T}'s. The (Red) plot is the signal that the analysis is designed to capture. Notice that in this model region, the non all-hadronic signal completely swamps the all-hadronic signal.

APPENDIX B

Background Predictions

Tables B.1–B.3 present the prefit predictions for the number of standard model background events in each of the 84 search regions, along with the number of observed events. "Prefit" means that there is no constraint from the likelihood fit discussed in Section 5.7. The corresponding information for the ten aggregate search regions is presented in Table B.4.

Table B.1: The observed number of events and the total background prediction for search regions with $N_\mathrm{t} = 1$ and $N_\mathrm{b} = 1$. The first uncertainty in the background prediction is statistical and the second is systematic.

Search region	N_t	N_b	m_{T2} [GeV]	$p_\mathrm{T}^\mathrm{miss}$ [GeV]	Data	Predicted background
1	1	1	200–300	250–400	1649	$1600 \pm 30^{+130}_{-140}$
2	1	1	200–300	400–500	85	$73^{+7}_{-6}\,^{+12}_{-9}$
3	1	1	200–300	500–600	23	$18^{+4}_{-3}\,^{+6}_{-4}$
4	1	1	200–300	600–750	7	$3.6^{+1.9}_{-0.8}\,^{+1.9}_{-0.8}$
5	1	1	200–550	≥750	7	$5.0^{+2.4}_{-1.1}\,^{+1.9}_{-1.2}$
6	1	1	300–400	250–400	1020	$890 \pm 20^{+80}_{-80}$
7	1	1	300–400	400–500	87	$79^{+7}_{-6} \pm 9$
8	1	1	300–400	500–600	23	$17^{+4}_{-2} \pm 3$
9	1	1	300–400	600–750	9	$3.7^{+2.2}_{-0.8}\,^{+1.6}_{-0.9}$
10	1	1	400–550	250–400	108	$107^{+8}_{-7} \pm 10$
11	1	1	400–550	400–500	116	$105^{+7}_{-6} \pm 10$
12	1	1	400–550	500–600	47	$38^{+5}_{-4} \pm 7$
13	1	1	400–550	600–750	12	$8.1^{+2.4}_{-1.2} \pm 1.9$
14	1	1	550–750	250–400	1	$0.7^{+1.0}_{-0.3}\,^{+0.4}_{-0.2}$
15	1	1	550–750	400–500	7	$4.3^{+2.0}_{-1.1} \pm 0.8$
16	1	1	550–750	500–600	17	$13^{+3}_{-2} \pm 3$
17	1	1	550–750	600–750	10	$19^{+3}_{-2} \pm 4$
18	1	1	550–750	≥750	7	$4.0^{+1.5}_{-0.3} \pm 1.8$
19	1	1	≥750	250–600	0	$0.1^{+1.7}_{-0.1} \pm 0.1$
20	1	1	≥750	600–750	1	$1.9^{+2.2}_{-1.0}\,^{+0.9}_{-0.8}$
21	1	1	≥750	≥750	8	$4.6^{+1.6}_{-0.5} \pm 1.9$

Table B.2: The observed number of events and the total background prediction for search regions with $N_{\mathrm{t}} = 1$ and $N_{\mathrm{b}} \geq 2$. The first uncertainty in the background prediction is statistical and the second is systematic.

Search region	N_{t}	N_{b}	m_{T2} [GeV]	$p_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	Data	Predicted background
22	1	2	200–350	250–400	596	$580 \pm 20 \pm 60$
23	1	2	200–350	400–500	59	$41\,^{+6}_{-5}\,^{+6}_{-5}$
24	1	2	200–350	500–600	14	$8.7\,^{+3.4}_{-2.1} \pm 1.3$
25	1	2	200–350	600–750	2	$2.1\,^{+2.7}_{-0.8} \pm 0.5$
26	1	2	200–650	$\geq$750	1	$3.0\,^{+2.4}_{-1.0}\,^{+0.9}_{-0.6}$
27	1	2	350–450	250–400	69	$67\,^{+6}_{-5}\,^{+18}_{-14}$
28	1	2	350–450	400–500	19	$13\,^{+4}_{-2} \pm 3$
29	1	2	350–450	500–600	4	$3.2\,^{+2.1}_{-0.9} \pm 1.0$
30	1	2	350–450	600–750	2	$0.6\,^{+1.4}_{-0.1} \pm 0.3$
31	1	2	450–650	250–400	3	$4.0\,^{+2.0}_{-1.1}\,^{+0.7}_{-0.9}$
32	1	2	450–650	400–500	9	$9.7\,^{+2.7}_{-1.8}\,^{+2.1}_{-2.0}$
33	1	2	450–650	500–600	6	$6.0\,^{+1.6}_{-0.9} \pm 1.9$
34	1	2	450–650	600–750	2	$4.6\,^{+2.6}_{-1.3} \pm 1.2$
35	1	2	$\geq$650	250–600	0	$0.06\,^{+1.03}_{-0.03} \pm 0.03$
36	1	2	$\geq$650	600–750	0	$1.0\,^{+1.8}_{-0.1} \pm 0.5$
37	1	2	$\geq$650	$\geq$750	2	$1.2\,^{+1.1}_{-0.1} \pm 0.5$
Search region	N_{t}	N_{b}	H_{T} [GeV]	$p_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	Data	Predicted background
38	1	$\geq$3	300–1000	250–350	85	$81\,^{+9}_{-8} \pm 7$
39	1	$\geq$3	300–1000	350–450	22	$15\,^{+5}_{-3} \pm 2$
40	1	$\geq$3	300–1000	450–550	6	$4.5\,^{+3.4}_{-1.7} \pm 0.8$
41	1	$\geq$3	300–1000	$\geq$550	2	$2.4\,^{+2.9}_{-1.0}\,^{+1.0}_{-0.7}$
42	1	$\geq$3	1000–1500	250–350	12	$13\,^{+4}_{-3} \pm 2$
43	1	$\geq$3	1000–1500	350–450	5	$5.0\,^{+2.7}_{-1.7} \pm 1.1$
44	1	$\geq$3	1000–1500	450–550	0	$1.8\,^{+2.3}_{-1.0} \pm 0.4$
45	1	$\geq$3	1000–1500	$\geq$550	3	$2.7\,^{+3.9}_{-1.4}\,^{+0.6}_{-0.5}$
46	1	$\geq$3	$\geq$1500	250–350	2	$9.6\,^{+3.4}_{-2.2} \pm 3.3$
47	1	$\geq$3	$\geq$1500	350–550	1	$3.4\,^{+2.3}_{-1.2}\,^{+3.4}_{-1.5}$
48	1	$\geq$3	$\geq$1500	$\geq$550	0	$1.3\,^{+1.8}_{-0.7} \pm 0.3$

Table B.3: The observed number of events and the total background prediction for search regions with $N_{\mathrm{t}} \geq 2$. The first uncertainty in the background prediction is statistical and the second is systematic.

Search region	N_{t}	N_{b}	m_{T2} [GeV]	$p_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	Data	Predicted background
49	2	1	200–300	250–350	57	$60\ ^{+6}_{-5} \pm 11$
50	2	1	200–300	350–450	10	$7.5\ ^{+2.5}_{-1.7}\ ^{+1.8}_{-1.4}$
51	2	1	200–300	450–600	0	$2.2\ ^{+1.4}_{-0.8}\ ^{+0.8}_{-0.5}$
52	2	1	200–450	≥600	0	$0.9\ ^{+2.0}_{-0.6}\ ^{+0.4}_{-0.3}$
53	2	1	300–450	250–350	38	$32\ ^{+5}_{-4} \pm 3$
54	2	1	300–450	350–450	8	$11\ ^{+3}_{-2} \pm 2$
55	2	1	300–450	450–600	4	$2.1\ ^{+1.7}_{-0.7}\ ^{+0.8}_{-0.5}$
56	2	1	≥450	250–450	2	$1.8\ ^{+1.5}_{-0.6} \pm 0.4$
57	2	1	≥450	450–600	3	$3.3\ ^{+2.7}_{-1.1} \pm 0.9$
58	2	1	≥450	≥600	7	$1.0\ ^{+1.2}_{-0.1} \pm 0.5$
59	2	2	200–300	250–350	46	$43 \pm 5^{+5}_{-6}$
60	2	2	200–300	350–450	11	$8.7\ ^{+2.7}_{-1.9}\ ^{+1.4}_{-1.3}$
61	2	2	200–300	450–600	1	$0.6\ ^{+1.6}_{-0.4}\ ^{+0.3}_{-0.2}$
62	2	2	200–400	≥600	1	$0.6\ ^{+1.7}_{-0.5} \pm 0.2$
63	2	2	300–400	250–350	28	$27\ ^{+5}_{-4} \pm 3$
64	2	2	300–400	350–450	6	$4.9\ ^{+2.9}_{-1.6} \pm 0.9$
65	2	2	300–400	450–600	3	$1.7\ ^{+2.4}_{-1.0}\ ^{+0.6}_{-0.5}$
66	2	2	400–500	250–450	4	$4.7\ ^{+2.3}_{-1.2}\ ^{+0.7}_{-0.8}$
67	2	2	400–500	450–600	1	$1.4\ ^{+2.7}_{-0.7}\ ^{+0.4}_{-0.6}$
68	2	2	≥400	≥600	1	$0.5\ ^{+2.7}_{-0.1} \pm 0.2$
69	2	2	≥500	250–450	0	$0.1\ ^{+1.4}_{-0.1} \pm 0.1$
70	2	2	≥500	450–600	2	$0.5\ ^{+2.2}_{-0.1} \pm 0.1$
Search region	N_{t}	N_{b}	H_{T} [GeV]	$p_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	Data	Predicted background
71	2	≥3	300–900	250–350	3	$9.6\ ^{+3.0}_{-2.1} \pm 1.7$
72	2	≥3	300–900	350–500	2	$0.7\ ^{+2.0}_{-0.4} \pm 0.2$
73	2	≥3	300–1300	≥500	0	$0.3\ ^{+0.5}_{-0.3}\ ^{+0.3}_{-0.2}$
74	2	≥3	900–1300	250–350	6	$4.7\ ^{+2.9}_{-1.7}\ ^{+0.7}_{-0.9}$
75	2	≥3	900–1300	350–500	3	$1.2\ ^{+1.6}_{-0.7} \pm 0.4$
76	2	≥3	≥1300	250–350	3	$3.5\ ^{+2.1}_{-1.2} \pm 1.4$
77	2	≥3	≥1300	350–500	2	$2.1\ ^{+2.1}_{-1.0}\ ^{+0.4}_{-0.5}$
78	2	≥3	≥1300	≥500	0	$0.2\ ^{+1.7}_{-0.3} \pm 0.2$
79	≥3	1	≥300	250–350	0	$0.3\ ^{+2.0}_{-0.3} \pm 0.2$
80	≥3	1	≥300	≥350	1	$0.6\ ^{+1.6}_{-0.5} \pm 0.2$
81	≥3	2	≥300	250–400	1	$1.7\ ^{+1.5}_{-0.7}\ ^{+0.6}_{-0.5}$
82	≥3	2	≥300	≥400	0	$0.1\ ^{+2.2}_{-0.1} \pm 0.1$
83	≥3	≥3	≥300	250–350	0	$0.5\ ^{+1.5}_{-0.4} \pm 0.5$
84	≥3	≥3	≥300	≥350	0	$0.0\ ^{+1.6}_{-0.0}\ ^{+0.1}_{-0.0}$

Table B.4: The observed number of events and the total background prediction for the aggregate search regions. The first uncertainty in the background prediction is statistical and the second is systematic.

Search region	N_{t}	N_{b}	m_{T2} [GeV]	$p_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	Data	Predicted background
1	$\geq$1	$\geq$1	$\geq$200	$\geq$250	4424	$4100 \pm 50^{+390}_{-340}$
2	$\geq$2	$\geq$2	$\geq$200	$\geq$250	124	$116 \pm 8^{+15}_{-12}$
3	$\geq$3	$\geq$1	$\geq$200	$\geq$250	2	$3.3^{+2.0}_{-1.1}\,^{+1.2}_{-1.1}$
4	$\geq$3	$\geq$3	$\geq$200	$\geq$250	0	$0.5^{+1.4}_{-0.4} \pm 0.5$
5	$\geq$2	$\geq$1	$\geq$200	$\geq$400	41	$30^{+4}_{-3}\,^{+5}_{-4}$
6	$\geq$1	$\geq$2	$\geq$600	$\geq$400	4	$7.5^{+2.1}_{-1.2}\,^{+2.0}_{-1.9}$
Search region	N_{t}	N_{b}	H_{T} [GeV]	$p_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	Data	Predicted background
7	$\geq$1	$\geq$2	$\geq$1400	$\geq$500	6	$6.0^{+2.7}_{-1.5} \pm 1.5$
8	$\geq$2	$\geq$3	$\geq$600	$\geq$350	7	$3.9^{+2.1}_{-1.2} \pm 0.9$
9	$\geq$2	$\geq$3	$\geq$300	$\geq$500	0	$0.6^{+1.0}_{-0.4} \pm 0.4$
10	$\geq$2	$\geq$3	$\geq$1300	$\geq$500	0	$0.2^{+1.8}_{-0.3} \pm 0.2$

APPENDIX C

Aggregate Search Bins for Reinterpretation

In the analysis presented in Chapter Five, the results are interpreted in the context of several simplified SUSY models; however, it is possible to reinterpret the results in the context of other BSM theories [87]. One of the current CMS collaboration goals is to facilitate reinterpretation of the analysis results by non-CMS physicists. The goal is not to provide perfect replication, such a goal is not possible without heavy involvement of the CMS collaborators. The goal is to provide the information and tools necessary for re-interpreters to take the published results and re-interpret them without heavy involvement from CMS collaborators.

The information provided is designed to be useful for phenomenologists interested in re-interpreting the results of the analysis in the context of other physics models. Such re-interpretation requires that someone be able to produce the hadronized simulation results of their BSM target signal, be able to cluster jets, run a fast simulation package, and know how to process the contents of root files. With this knowledge and skill they should be able to estimate the event yield for their targeted signal for the bins defined in the analysis and compare the signal yields to the data yields and background predictions from the CMS analysis. The re-interpreter can evaluate whether the data presented in the analysis exclude their signal model or not.

Re-interpretation of the results of many exclusive search bins could be problematic, as it is not trivial for people to implement so many search bins. Therefore, we have defined much smaller numbers of non-exclusive aggregated search bins as shown in Table 5.1. Re-interpretation would be able to use the "best" aggregate search region for their particular BSM signal. This results in some loss of sensitivity compared to using all the exclusive search bins in the analysis, but is easier for an outside physicist to work with. In order to provide a reduced set of search bins with good sensitivity to various alternative signal models, an iterative process of defining and identifying sensitive search bins took place. As part of this process, the sensitivity of the search bins was evaluated at the various mass points in the parameter space of our simplified models. Fig. C.1 shows a step of this process. For each simulated mass point, the 95% CL upper limit on the production cross section was calculated based on the results of each individual search bin. Fig. C.1 shows the aggregate search bin that provided the most stringent limit at that particular mass point. The search bin numbers shown in this figure do not correspond to the final reported aggregate search bins as these plots are from the exploration step of defining the aggregate search bins. This step helped identify which bin definitions were the most broadly useful, and which bins were superfluous.

In order to process alternate signal events in a manner consistent with the treatment that produced the reported background and data yields, b-tagging efficiency is made available here [88], and the simplified top tagger used to produce the reported values for the aggregate search bins is made available here [89].

It would be useful to create exclusion curves based on the aggregate search bin analysis similar to the exclusion curves based on the full search bin analysis. Since the aggregate search bins are not orthogonal and contain overlapped events, the calculation of the 95% CL production cross section upper limit is not based on the combination of the search bins, but a separate limit is calculated based on each single aggregate bin. In producing the exclusion curves shown here, for each mass point in the model's parameter space, the single aggregate bin with the most stringent limit is used. Figs. C.2–C.6 show comparisons between the limits calculated using the full analysis and limits calculated by choosing the strongest aggregate bin for each mass point in the models. The figures are for respectively the T2tt, T1tttt, T1ttbb, T5tttt. T5ttcc models.

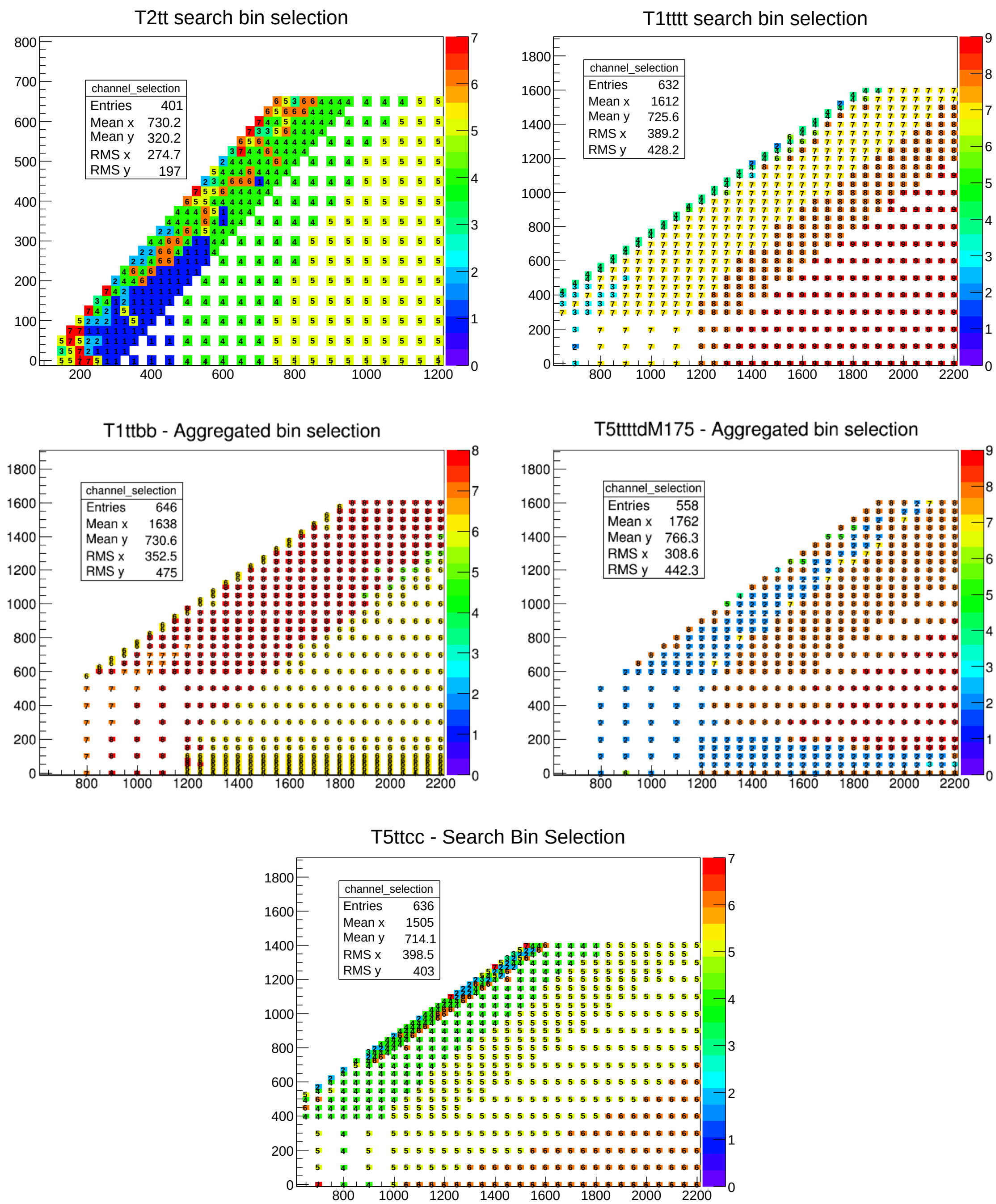


Figure C.1: The most sensitive aggregate search bins in the SUSY models presented in Chapter Five. At each mass point the most sensitive aggregate search bins is identified. The search bin numbers in these plots do not correspond to the search bin numbers presented in the text. These plots part of the testing to identify the most useful aggregate search bins, and once final selection was made, the chosen aggregate search bins were assign their current numbers. The T2tt model (top left), the T1tttt model (top right), the T1ttbb model (middle left), the T5tttt (middle right), and the T5ttcc model (bottom).

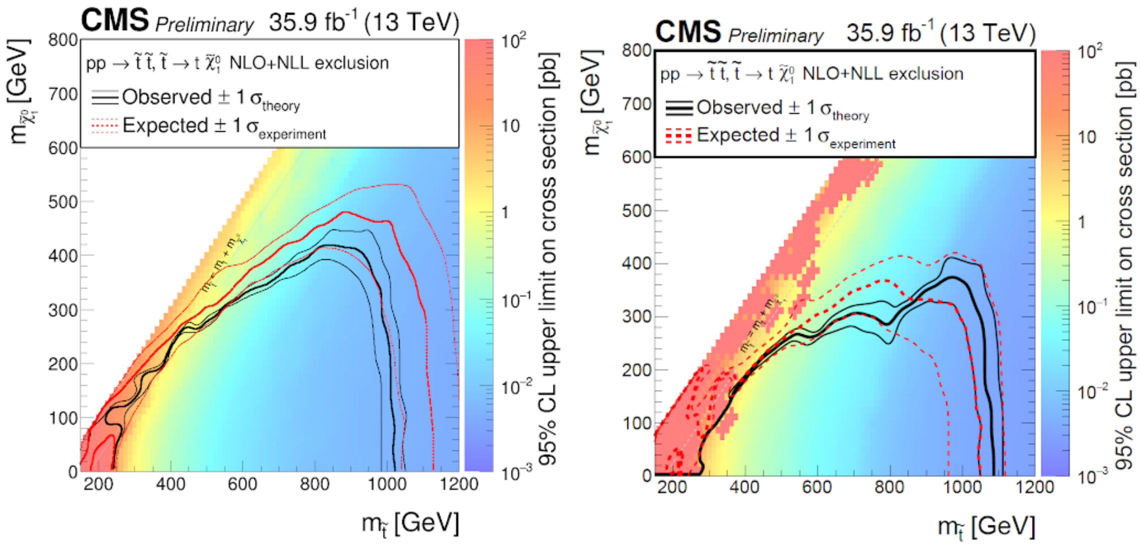


Figure C.2: Exclusion curve based on standard search bins (left) and aggregate search bins (right) for the T2tt model.

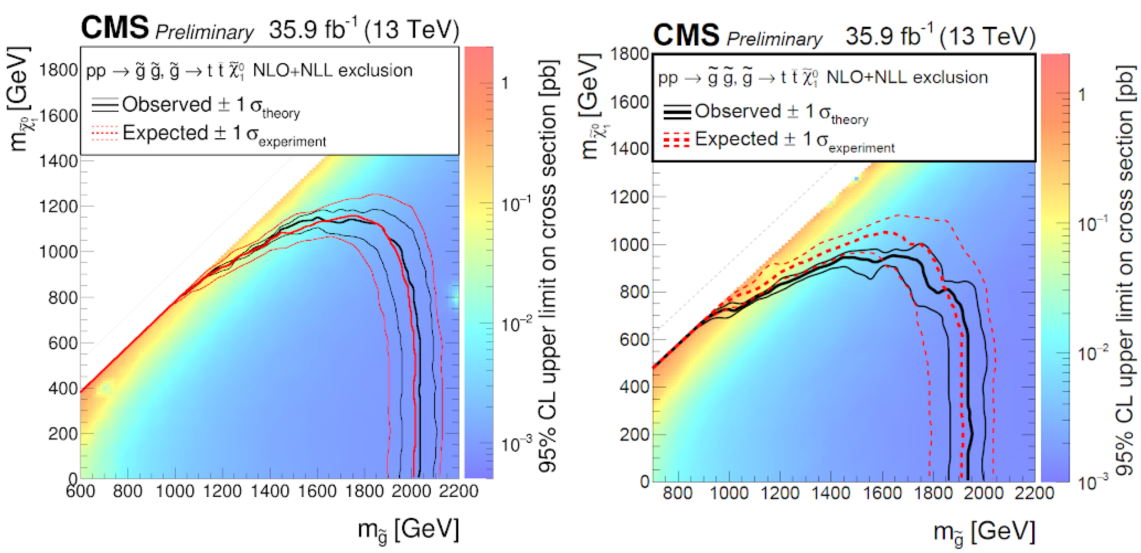


Figure C.3: Exclusion curve based on standard search bins (left) and aggregate search bins (right) for the T1tttt model.

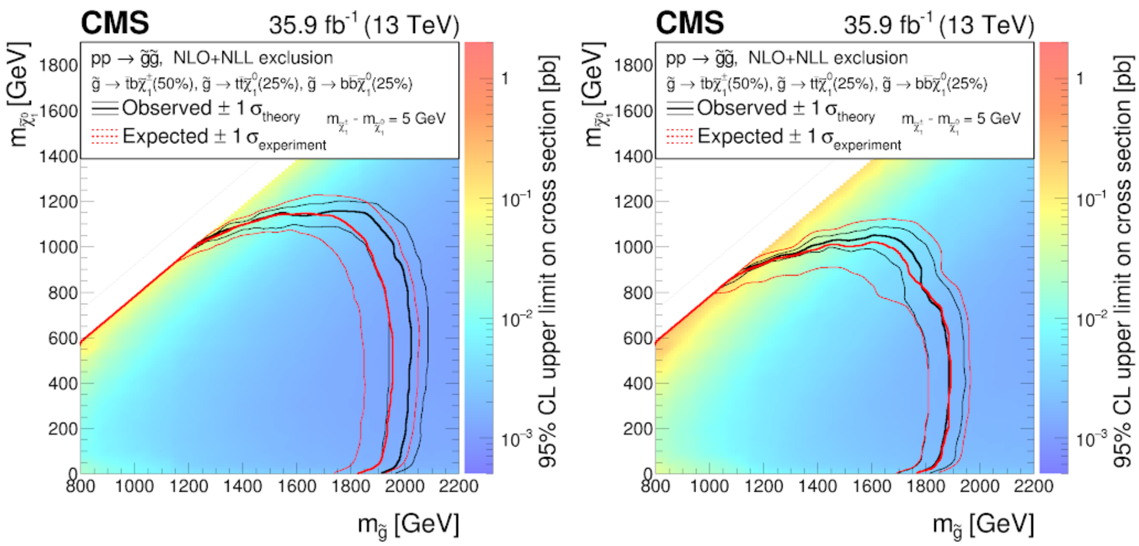


Figure C.4: Exclusion curve based on standard search bins (left) and aggregate search bins (right) for the T1ttbb model.

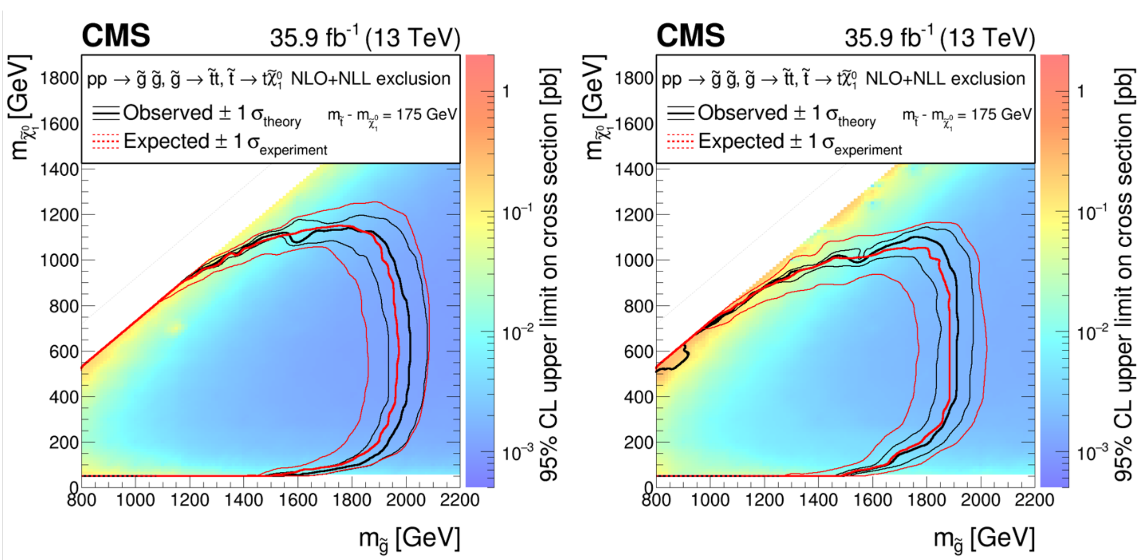


Figure C.5: Exclusion curve based on the standard search bins (left) and aggregate search bins (right) for the T5tttt model.

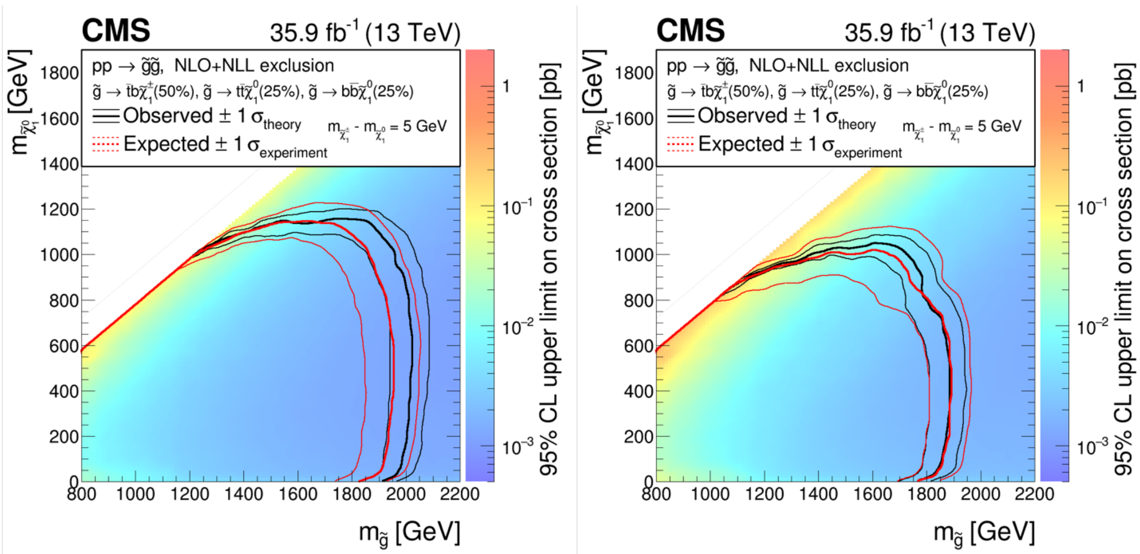


Figure C.6: Exclusion curve based on the standard search bins (left) and aggregate search bins (right) for the T5ttcc model.

BIBLIOGRAPHY

[1] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett. B*, vol. 716, p. 1, 2012.

[2] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys. Lett. B*, vol. 716, p. 30, 2012.

[3] “Worth the wait,” Jul 2012. https://www.economist.com/graphic-detail/2012/07/04/worth-the-wait

[4] “Standard Model — Wikipedia, the free encyclopedia.” https://en.wikipedia.org/wiki/Standard_Model

[5] M. Newby, “Galaxy rotation curve.” https://milkyway.cs.rpi.edu/download/images/gal_rotation_curve

[6] B. Dunbar, “A clash of clusters provides another clue to dark matter.” https://www.nasa.gov/mission_pages/chandra/multimedia/photos08-111.html

[7] D. I. Kazakov, “Beyond the standard model: In search of supersymmetry,” in *2000 European School of high-energy physics, Caramulo, Portugal, 20 Aug-2 Sep 2000: Proceedings*, 2000, pp. 125–199.

[8] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*, ser. CERN Yellow Reports: Monographs, 2004. http://cds.cern.ch/record/782076

[9] CMS Collaboration, “The CMS experiment at the CERN LHC,” *JINST*, vol. 3, p. S08004, 2008.

[10] ATLAS Collaboration, “The ATLAS experiment at the CERN Large Hadron Collider,” *JINST*, vol. 3, p. S08003, 2008.

[11] LHCb Collaboration, “The LHCb detector at the LHC,” *JINST*, vol. 3, p. S08005, 2008.

[12] ALICE Collaboration, “The ALICE experiment at the CERN LHC,” *JINST*, vol. 3, p. S08002, 2008.

[13] CMS Collaboration, “CMS luminosity - public results.” http://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults

[14] CMS Collaboration, “Cutaway diagrams of CMS detector,” http://cds.cern.ch/record/2665537

[15] G. Bayatian *et al.*, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*, ser. Technical Design Report CMS, 2006. http://cds.cern.ch/record/922757

[16] CMS Collaboration, "Interactive slice of the CMS detector," Aug 2016. http://cds.cern.ch/record/2205172

[17] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker," *JINST*, vol. 9, no. 10, p. P10009, 2014.

[18] CMS Collaboration, "Precision measurement of the structure of the CMS inner tracking system using nuclear interactions," *JINST*, vol. 13, no. 10, p. P10034, 2018. http://cds.cern.ch/record/2629890

[19] D. Kotliński *et al.*, "The control and readout systems of the cms pixel barrel detector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 565, no. 1, p. 73, 2006, proceedings of the International Workshop on Semiconductor Pixel Detectors for Particles and Imaging. http://www.sciencedirect.com/science/article/pii/S0168900206007200

[20] "World's largest silicon tracking detector installed at cern." http://www.sciencedaily.com/releases/2007/12/071218115622.htm

[21] CMS Collaboration, "Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV," *JINST*, vol. 8, no. 09, p. P09009, Sep 2013.

[22] CMS Collaboration, "Performance of the CMS electromagnetic calorimeter during the LHC Run II and its role in precision physics measurements," CMS-CR-2018-162, Aug 2018. http://cds.cern.ch/record/2637093

[23] CMS Collaboration, "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV," *JINST*, vol. 10, no. 08, p. P08010, Aug 2015.

[24] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV," *JINST*, vol. 10, no. 06, p. P06005, Jun 2015.

[25] CMS Collaboration, "CMS hadron calorimeter technical design report," CERN, CERN/LHCC 97-31, 1997. http://cds.cern.ch/record/357153/

[26] CMS Collaboration, "Calibration of the CMS hadron calorimeters with proton-proton collision data at $\sqrt{s} = 13$ TeV," 2019, to be submitted to JINST. http://cms.cern.ch/iCMS/analysisadmin/cadilines?line=PRF-18-001

[27] CMS HCAL Collaboration, "Design, performance, and calibration of CMS hadron-barrel calorimeter wedges," *Eur. Phys. J.*, vol. C55, p. 159, 2008.

[28] CMS HCAL Collaboration, “Design, performance, and calibration of CMS hadron endcap calorimeters,” 2008. http://cds.cern.ch/record/1103003

[29] CMS HCAL Collaboration, “Design, performance and calibration of the CMS forward calorimeter wedges,” *Eur. Phys. J.*, vol. C53, p. 139, 2008.

[30] CMS HCAL Collaboration, “Design, performance, and calibration of the CMS hadron-outer calorimeter,” *Eur. Phys. J.*, vol. C57, p. 653, 2008.

[31] CMS Collaboration, “CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter,” CERN-LHCC-2012-015. CMS-TDR-10, Sep 2012. http://cds.cern.ch/record/1481837

[32] CMS Collaboration, “Performance of the prototype readout system for the CMS endcap hadron calorimeter upgrade,” CMS-CR-2015-291, Nov 2015. http://cds.cern.ch/record/2065593

[33] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV,” *JINST*, vol. 13, no. 06, p. P06015, 2018, all the figures and tables can be found at http://cms-results.web.cern.ch/cms-results/public-results/publications/MUO-16-001. http://cds.cern.ch/record/2313130

[34] CMS Collaboration, “The CMS trigger system.” *JINST*, vol. 12, no. 01, p. P01020, 2017, all the figures and tables can be found at http://cms-results.web.cern.ch/cms-results/public-results/publications/TRG-12-001/index.html. http://cds.cern.ch/record/2212926

[35] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations,” *JHEP*, vol. 07, p. 079, 2014.

[36] J. Alwall *et al.*, “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions,” *Eur. Phys. J. C*, vol. 53, p. 473, 2008.

[37] P. Nason, “A new method for combining NLO QCD with shower Monte Carlo algorithms,” *JHEP*, vol. 11, p. 040, 2004.

[38] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method,” *JHEP*, vol. 11, p. 070, 2007.

[39] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX,” *JHEP*, vol. 06, p. 043, 2010.

[40] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method," *Eur. Phys. J. C*, vol. 71, p. 1547, 2011.

[41] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "A brief introduction to PYTHIA 8.1," *Comput. Phys. Commun.*, vol. 178, p. 852, 2008.

[42] GEANT4 Collaboration, "GEANT4—a simulation toolkit," *Nucl. Instrum. Meth. A*, vol. 506, p. 250, 2003.

[43] CMS Collaboration, "Validation of Geant4 simulation with test-beam and Run-2 data," Jul 2018. http://cds.cern.ch/record/2629853

[44] S. Abdullin, P. Azzi, F. Beaudette, P. Janot, and A. Perrotta, "The fast simulation of the CMS detector at LHC," *J. Phys. Conf. Ser.*, vol. 331, p. 032049, 2011.

[45] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector," *JINST*, vol. 12, no. 10, p. P10003, 2017, all the figures and tables can be found at http://cms-results.web.cern.ch/cms-results/public-results/publications/PRF-14-001. http://cds.cern.ch/record/2270046

[46] W. Adam, R. Frühwirth, A. Strandlie, and T. Todor, "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC," 2005. http://cds.cern.ch/record/815410

[47] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_{t} jet clustering algorithm," *Journal of High Energy Physics*, vol. 2008, no. 04, p. 063, 2008.

[48] CMS Collaboration, "Identification of b quark jets at the CMS experiment in the LHC Run 2," CMS Physics Analysis Summary CMS-PAS-BTV-15-001, 2016. http://cds.cern.ch/record/2138504

[49] CMS Collaboration, "Identification of b-quark jets with the CMS experiment," *JINST*, vol. 8, p. P04013, 2013.

[50] H. E. Haber and L. Stephenson Haskins, "Supersymmetric Theory and Models," in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics : Anticipating the Next Discoveries in Particle Physics (TASI 2016): Boulder, CO, USA, June 6-July 1, 2016*, WSP. WSP, 2018, pp. 355–499.

[51] J. Alwall, P. Schuster, and N. Toro, "Simplified models for a first characterization of new physics at the LHC," *Phys. Rev. D*, vol. 79, p. 075020, 2009.

[52] J. Alwall, M.-P. Le, M. Lisanti, and J. G. Wacker, "Model-independent jets plus missing energy searches," *Phys. Rev. D*, vol. 79, p. 015005, 2009.

[53] LHC New Physics Working Group Collaboration, "Simplified models for LHC new physics searches," *J. Phys. G*, vol. 39, p. 105005, 2012.

[54] D. Alves, E. Izaguirre, and J. G. Wacker, “Where the sidewalk ends: jets and missing energy search strategies for the 7 TeV LHC,” *JHEP*, vol. 10, p. 012, 2011.

[55] CMS Collaboration, “Interpretation of searches for supersymmetry with simplified models,” *Phys. Rev. D*, vol. 88, p. 052017, 2013.

[56] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas,” *Phys. Lett. B*, vol. 659, p. 119, 2007.

[57] UA1 Collaboration, “Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV,” *Phys. Lett. B*, vol. 122, p. 103, 1983.

[58] CMS Collaboration, “Top tagging with new approaches,” CMS Physics Analysis Summary CMS-PAS-JME-15-002, 2016. http://cds.cern.ch/record/2126325

[59] ATLAS Collaboration, “Identification of high transverse momentum top quarks in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector,” *JHEP*, vol. 06, p. 093, 2016.

[60] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, “Towards an understanding of jet substructure,” *JHEP*, vol. 09, p. 029, 2013.

[61] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, “Soft drop,” *JHEP*, vol. 05, p. 146, 2014.

[62] CMS Collaboration, “Jet algorithms performance in 13 TeV data,” CMS Physics Analysis Summary CMS-PAS-JME-16-003, 2017. http://cds.cern.ch/record/2256875

[63] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, “Better jet clustering algorithms,” *JHEP*, vol. 08, p. 001, 1997.

[64] M. Wobisch and T. Wengler, “Hadronization corrections to jet cross sections in deep-inelastic scattering,” 1998. http://arxiv.org/abs/hep-ph/9907280

[65] J. Thaler and K. Van Tilburg, “Identifying boosted objects with N-subjettiness,” *JHEP*, vol. 03, p. 015, 2011.

[66] T. K. Ho, “Random decision forests,” *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1, p. 278, 1995.

[67] CMS Collaboration, “Performance of quark/gluon discrimination in 8 TeV pp data,” CMS Physics Analysis Summary CMS-PAS-JME-13-002, 2013. http://cds.cern.ch/record/1599732

[68] CMS Collaboration, “Search for supersymmetry in the all-hadronic final state using top quark tagging in pp collisions at $\sqrt{s} = 13$ TeV,” *Phys. Rev. D*, vol. 96, p. 012004, 2017.

[69] C. G. Lester and D. J. Summers, “Measuring masses of semi-invisibly decaying particles pair produced at hadron colliders,” *Phys. Lett. B*, vol. 463, p. 99, 1999.

[70] A. Barr, C. Lester, and P. Stephens, “A variable for measuring masses at hadron colliders when missing energy is expected; m_{T2}: the truth behind the glamour,” *J. Phys. G*, vol. 29, p. 2343, 2003.

[71] H. E. Haber and L. Stephenson Haskins, “Supersymmetric theory and models,” in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics : Anticipating the Next Discoveries in Particle Physics (TASI 2016): Boulder, CO, USA, June 6-July 1, 2016.* WSP, 2018, p. 355.

[72] A. Rodríguez-Sánchez and A. Pich, “Confronting hadronic tau decays with non-leptonic kaon decays,” *Nucl. Part. Phys. Proc.*, vol. 300, p. 131, 2018.

[73] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV,” *JINST*, vol. 12, p. P02014, 2017.

[74] A. Sirunyan *et al.*, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV,” *JINST*, vol. 13, no. 05, p. P05011, 2018.

[75] T. Junk, “Confidence level computation for combining searches with small statistics,” *Nucl. Instrum. Meth. A*, vol. 434, p. 435, 1999.

[76] A. L. Read, “Presentation of search results: the CL_s technique,” *J. Phys. G*, vol. 28, p. 2693, 2002.

[77] ATLAS and CMS Collaborations, “Procedure for the LHC Higgs boson search combination in summer 2011,” ATL-PHYS-PUB-2011-011, CMS NOTE-2011/005, 2011. http://cds.cern.ch/record/1379837

[78] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics,” *Eur. Phys. J. C*, vol. 71, p. 1554, 2011.

[79] C. Borschensky, M. Krämer, A. Kulesza, M. Mangano, S. Padhi, T. Plehn, and X. Portell, “Squark and gluino production cross sections in pp collisions at $\sqrt{s} = 13$, 14, 33 and 100 TeV,” *Eur. Phys. J. C*, vol. 74, p. 3174, 2014.

[80] CMS Collaboration, “CMS luminosity measurements for the 2016 data taking period,” CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017. http://cds.cern.ch/record/2257069

[81] CMS Collaboration, “Search for supersymmetry in multijet events with missing transverse momentum in proton-proton collisions at 13 TeV,” *Phys. Rev. D*, vol. 96, p. 032003, 2017.

[82] CMS Collaboration, "Search for new phenomena with the M_{T2} variable in the all-hadronic final state produced in proton-proton collisions at $\sqrt{s} = 13$ TeV," *Eur. Phys. J. C*, vol. 77, p. 710, 2017.

[83] CMS Collaboration, "Search for physics beyond the standard model in events with two leptons of same sign, missing transverse momentum, and jets in proton-proton collisions at $\sqrt{s} = 13$ TeV," *Eur. Phys. J. C*, vol. 77, p. 578, 2017.

[84] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. http://www.tensorflow.org/

[85] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, p. 861, 2006, ROC Analysis in Pattern Recognition. http://www.sciencedirect.com/science/article/pii/S016786550500303X

[86] Y. Ganin and V. Lempitsky, "Unsupervised Domain Adaptation by Backpropagation," 2014. http://arxiv.org/abs/1409.7495

[87] "(Re)interpreting the Results of New Physics Searches at the LHC - INSPIRE-HEP," 2018. http://inspirehep.net/record/1667587

[88] CMS Collaboration, "CMS SUSY Results: Objects Efficiency." https://twiki.cern.ch/twiki/bin/view/CMSPublic/SUSMoriond2017ObjectsEfficiency

[89] CMS Collaboration, "Stand alone top tagger." https://github.com/susy2015/TopTagger/releases/tag/StandaloneRelease_v1.0.1