Grooving in the Shadows: A Search for Stealth and R-Parity Violating Supersymmetry in CMS Data using the ABCDisCoTEC Method

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Bryan James Crossman

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Nadja Strobbe

January, 2025

Acknowledgements

I am beyond grateful for all of the support that I have had while conducting the research for this thesis.

First and foremost, I want to thank my advisor, Nadja Strobbe, for her guidance and wealth of knowledge of the CMS experiment. When I first introduced myself to her, I told her "I know nothing about particle physics, coding in C++/Python, and electronics but I would love to work with you." She took me in without question and from that day on, has dedicated her time and effort to helping me succeed. I'm grateful for her useful insights on CMS and our analysis methods. Additionally, she has encouraged me to take big leaps, to do the hard things, and to never be afraid of failure. Her support was instrumental throughout the research, writing, and defense of this thesis, and I'm so thankful for all she has done for me.

Additionally, I would like to extend thanks to my "pseudo-adviser" Jeremy Mans. While many of the projects we worked together on are not featured in the content of this thesis, Jeremy has been a wonderful resource in the development of the HGCAL electronics testing. He has given me an appreciation of the intricacies of building a new detector and understanding how they work at the lowest level. I'm grateful for his guidance, insights, and patience throughout the years.

I would then like to thank the Stealth Stop analysis group for their collaboration in developing this analysis. The PIs—Jim Hirschauer, Aron Soha, and Nadja Strobbe—were instrumental in the conception, development, and publication of this analysis. I'm thankful for their great eyes for detail and insightful questions which helped guide me in my analysis journey. The other group members—Chris Madrid, Joshua Hiltbrand, Semra Turkcapar, and Kelvin Mei—spent countless hours with me working on code, discussing our findings, and preparing for talks. I can't thank each of them enough for

their efforts in this analysis and for the friendships we developed along the way.

The final CMS group I would like to thank are all of the UMN CMS graduate students, both past and present. I have learned so much through conversations with them about their own analyses and have many fond memories of our time in the office together. I would specifically like to thank Charlie Kapsiak for his years of friendship. Charlie has always made himself available both for technical discussions as well as keeping me sane through the thesis writing process.

My friends and family outside of the academic realm have also played a large role in this work, whether they know it or not. To my parents and brother—Julie, Scott, and Andrew Crossman—thank you for always believing in me and pushing me to be the best version of myself. Your constant support has allowed me to achieve goals much larger than I thought I could.

Finally, I want to thank my wife Kris for her support throughout my years of graduate school. While it was a difficult journey with many ups and downs, I couldn't imagine anyone else I would rather have by my side. Thank you for letting me pursue my dreams, for picking me up in my hardest moments, and for giving me the confidence to keep going. Though you may never read this thesis in full, know that I would not have been able to write it without you.

Dedication

To my wife, Kris

Abstract

A search for Stealth and R-Parity violating supersymmetry in final states with many jets and little to no missing transverse momentum is presented. This search uses a novel, neural-network-based technique for data-driven background estimation known as the ABCDisCoTEC approach. Events are classified using two independent neural network discriminators that are used to carry out a side-band extrapolation. No significant excess of events is observed for either the Stealth or R-Parity violating signal models. Upper limits are placed on the mass of the top squark at $M_{\tilde{t}} = 700$ GeV and $M_{\tilde{t}} = 930$ GeV for the R-Parity violating and Stealth supersymmetry models, respectively.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The theory of subatomic particles and their interactions is described by the standard model of particle physics. This theory is established as one of the most precise predictions of nature to date. However, the standard model is known to be incomplete as it is only a description of three of the four fundamental forces of nature, with gravity being the one exception. Additionally, observations of dark matter and dark energy have no explanation within the standard model. Therefore, if the standard model is to be the ultimate explanation of interaction on the quantum scale, it must be extended.

One such extension of the standard model is supersymmetry (or SUSY). This theory posits an extension to the standard model which doubles the number of fundamental particles. Touted as a solution to the "hierarchy problem" and "lack of naturalness" of the standard model, supersymmetry is a prime candidate for high energy particle physics searches. To this point, however, there is no experimental evidence that supersymmetric particles exist.

Searches for supersymmetry are most frequently carried out under the assumption that supersymmetric signatures would appear as large imbalances of momentum in particle collision events. However, as this glaring signature has not been found, it is possible that SUSY is hiding in the periphery of SUSY phase space. That is, supersymmetric decays may create detector signatures that are quite similar to known signatures of standard model decays. If this is the case, alternative search techniques must be developed to investigate unexplored territory.

This thesis describes one such search for "stealthy" supersymmetry. Two supersymmetric models are considered, both resulting in collision events that generate a large number of standard model particles. A previous search for events matching these supersymmetric decays found an interesting discrepancy [1]. As such, alternative background estimation techniques have been developed to probe the search region of interest in an alternative fashion.

This thesis will proceed as follows. A description of the Large Hadron Collider and Compact Muon Solenoid is given in chapter 2. Chapter 3 presents a discussion of modern particle physics theory including a detailed description of supersymmetry. A brief discussion of the motivation for the analysis as well as an introductory description of analysis techniques is given in chapter 4. Datasets, event simulation, and signal region definitions are described in chapter 5. The ABCDisCoTEC (ABCDisCo Training Enhanced with Closure) method—a general purpose, neural-network based analysis approch developed for this analysis—is described in detail in chapter 6. Chapter 7 describes the analysis strategy including discussion the application and validation of the ABCDisCoTEC method in the context of this SUSY search. The results of the analysis are shown in chapter 8. Finally, a discussion of the findings and future outlook is provided in chapter 9.

Chapter 2

Experimental Equipment

At the heart of experimental particle physics are the extraordinary machines which allow researchers to explore the world on the smallest scale. Such experiments are designed to observe everything from the decay of protons to ultra-high energy neutrinos originating outside of the solar system.

As will be outlined in section 3.4, supersymmetric particles are predicted to have masses around the TeV energy scale. So, a high energy particle accelerator is necessary to produce sufficient energy for the creation of such particles. Additionally, a robust detection system is needed to identify the signatures left by supersymmetric particles.

The following section will discuss the Large Hadron Collider and the CMS detector, the catalyst for searching for physics beyond the standard model and the means for observing new physical phenomena.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) stands at the forefront of the energy frontier of particle physics. The LHC is a 26.7 km circular proton-proton collider that resides under the border of France and Switzerland near Geneva. Two counter-rotating beams of protons collide at four main interaction points along the circumference of the accelerator. Each of these interaction points houses a separate detector which has the capability of observing the aftermath of these interactions.

The LHC is responsible for the highest energy particle collisions in a man-made

experiment. Groups of protons, called "bunches", with individual proton energies of 6.5 TeV collide with a center-of-mass energy of 13 TeV every 25 ns. The resulting collisions create a shower of fundamental particles that can be detected and studied.

The two general purpose particle detectors at the LHC are the ATLAS and CMS experiments. These two experiments were designed as complementary devices for measuring different decay channels of the Higgs boson (H), leading to differences in design choices. The ATLAS collaboration opted for a toroidal magnet and high-granularity liquid argon electromagnetic calorimeter for precise measurement of the $H \to \gamma\gamma$ channel. The CMS detector is outfitted with solenoid magnet which allowed for precise measurement of muon decay angle and momentum in the $H \to ZZ \to 4\mu$ channel, while also using an electromagnetic calorimeter for measurement of Higgs decays to photons. However, both detectors are able to observe the decay of the Higgs boson in various channels, providing a useful crosscheck. The combination of these two specialties led to the discovery of the Higgs boson in 2012 [2, 3]. Both detectors also are designed to look for a variety of new particle physics processes, like those originating from SUSY interactions.

The other major experiments measuring LHC collisions are the LHCb and ALICE detectors. LHCb is an experiment dedicated to observing and analyzing rare decays of B mesons while ALICE is focused on collisions of lead ions.

The journey of the protons is shown in figure 2.1. A series of accelerators gradually increases the energy of proton bunches as they travel through the accelerator complex. Negatively charged hydrogen ions are first accelerated to 160 MeV in the LINAC 4 linear accelerator. The two electrons are stripped from these ions, yielding bare protons, as they enter the Proton Synchrotron Booster (PSB). Protons are then accelerated to 2 GeV before injection into the Proton Synchrotron (PS). Then, protons are further accelerated to 26 GeV within the PS and are injected into the Super Proton Synchrotron (SPS). A final acceleration to 450 GeV occurs before the protons are injected into the LHC main ring. They will then reach their final energy of 6.5 TeV through acceleration in the LHC before collision [5].

One of the main limitations in proton energy is the strength of the dipole magnets bending protons along the path of the LHC. The circular orbit of protons — a necessary component to ensure continuous acceleration — can only be achieved given that the

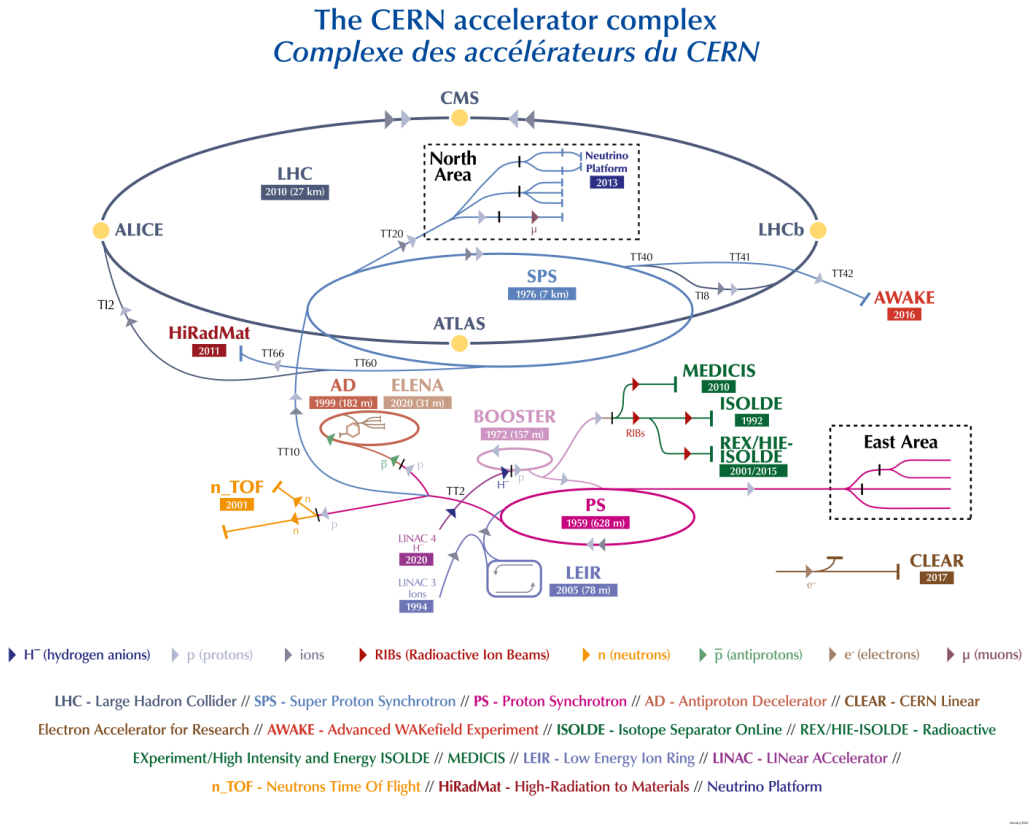

Figure 2.1: The accelerator chain at the LHC responsible for creating bunches of protons that are used in collisions. Protons are boosted through the LINAC, Booster (PSB), PS, and SPS before their final injection into the LHC. Adapted from [4].

bending force of the magnets is equal to the centrifugal force acting on the protons. That is:

$$F_{Lorentz} = evB \tag{2.1}$$

$$F_{Centrifugal} = \frac{\gamma m_0 v^2}{r} \tag{2.2}$$

$$\frac{p}{e} = Br, \tag{2.3}$$

where B is the magnetic field strength, r is the radius of the circular orbit, m_0 is the rest mass of the proton, v is the velocity of the proton, e is the electrical charge of a proton, p is the momentum of the proton, and γ is the Lorentz factor. The total bending angle of all magnets in the LHC must necessarily sum to a full rotation, meaning that:

$$2\pi = \frac{\int B d\ell}{Br}. \tag{2.4}$$

In the case of the LHC, 1232 dipole magnets with a length of 15 m are arranged around the circumference of the main ring. Therefore, the magnetic field strength required for keeping the 6.5 TeV protons in stable orbit is $B = 8.33T$ [6].

A large number of collisions need to be recorded in order to observe rare processes predicted by many theories. The luminosity ($\mathcal{L}$) of the collider dictates the total number of interactions that will be observed in our detector. This value is the proportionality constant defining the number of events of a given probability we expect to see per unit time:

$$\frac{dN_{event}}{dt} = \mathcal{L}\sigma_{event}, \tag{2.5}$$

where $\frac{dN_{event}}{dt}$ is the rate of event occurrence for a given process, $\mathcal{L}$ is the LHC luminosity, and σ_{event} is the cross section of the process of interest.

The LHC luminosity is determined by a number of factors, including the geometry of the beam and the beam emittance. It can be calculated using the following expression

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma}{4\pi\epsilon_n \beta} F, \tag{2.6}$$

where N_b is the number of protons per bunch, n_b is the total number of bunches per beam, f_{rev} is the revolution frequency of the beam, γ is the relativistic gamma factor,

ϵ_n is the normalized transverse beam emittance, β is the beta function at the collision point which describes the width of the beam, and F is a geometric reduction factor at the interaction point due to the crossing angle of the bunches. With 2808 total bunches of protons per beam containing $\sim 10^{11}$ protons each, the peak luminosity of the LHC at the CMS detector has a value of $\sim 10^{34} cm^{-2} s^{-1}$.

The physics reach of experiments at the LHC is limited by the total integrated luminosity, that is the luminosity integrated over the time of operation of the collider:

$$\mathcal{L}_{int} = \int_{\Delta t} \mathcal{L} dt, \tag{2.7}$$

where $\mathcal{L}_{int}$ is the total integrated luminosity, $\mathcal{L}$ is the instantaneous luminosity, and Δt is the total up-time of the LHC and CMS detector [7]. During Run 2 (the LHC operational period between 2016-2018), the CMS detector collected a total of $138 fb^{-1}$ of data.

The high-luminosity LHC (HL-LHC) is planned to begin data collection in 2029 as shown in the estimated timeline in figure 2.2. During long shutdown 3, the LHC will be upgraded to support an instantaneous luminosity 5 to 7.5 times greater than its current value. Experiments will be responsible for upgrading their detector systems in order to accommodate this increase in total collisions per bunch crossing (pileup). Currently, there are around 60 collisions per bunch crossing on average at the CMS and ATLAS interaction points. The HL-LHC is estimated to increase the number of pileup per bunch crossing to 200 [8].

2.2 The CMS Detector

The Compact Muon Solenoid (CMS) detector is a general purpose particle detector designed to measure and digitally reconstruct proton-proton collisions from the LHC.

The coordinate system for the CMS detector is shown in figure 2.3 starting with the x-axis, which points to the center of the LHC main ring. With the y-axis pointing vertically, the z-axis is directed along the beamline in the direction of the Jura mountains which reside to the west of the experiment. The positions of particles interacting with the detector are nominally defined in terms of a cylindrical coordinate system to match the shape of the experiment. The azimuthal angle ϕ is measured from the x-axis in the

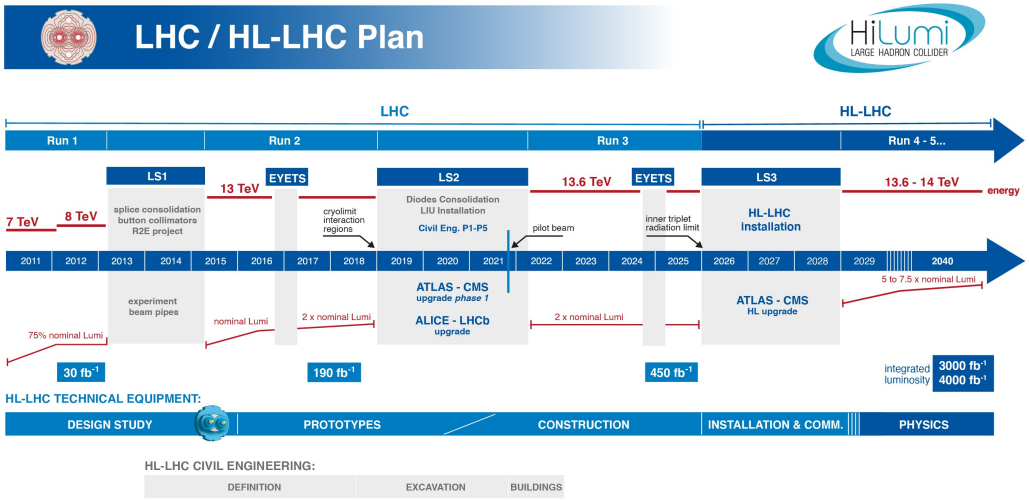

Figure 2.2: After the end of Run 3 in 2025, upgrades will take place to the LHC and the CMS detector to ready for operation at high luminosity. The first run of the HL-LHC is projected to begin in 2029 with an instantaneous luminosity of 5-7.5$\times 10^{-34} cm^{-2} s^{-1}$. Adapted from [9].

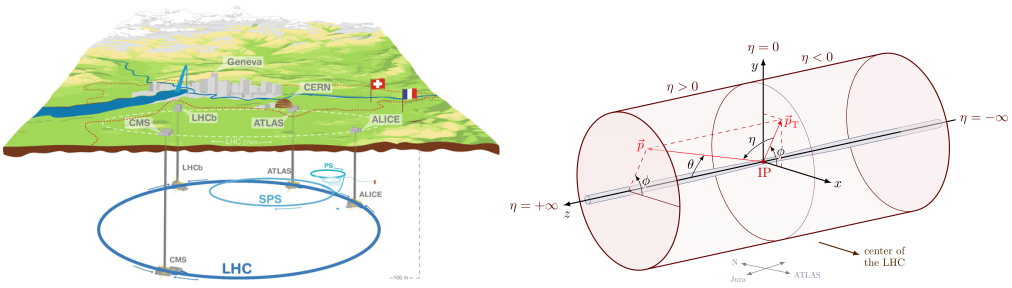

Figure 2.3: The CMS detector resides at the P5 interaction point of the LHC in Cessy, France (left). The coordinate system defined for the CMS detector is shown in the right diagram. Adapted from [11, 12].

x-y plane. The polar angle θ is defined as the angle from the z-axis. A transformation of the polar angle, called pseudorapidity (η), is defined as:

$$\eta = -\ln \tan \frac{\theta}{2}. \tag{2.8}$$

This angular definition is preferred, as opposed to the polar angle, because differences in pseudorapidity are Lorentz invariant under boosts in the z-direction. Additionally, the flux of particles in any pseudorapidity slice remains constant [10].

The main quantities of interest for objects produced in collisions are the transverse momentum (p_{T}) and transverse energy (E_{T}). Due to the composite nature of protons, the constituent quarks of the protons do not have the same momentum in the z-direction as the proton. Therefore, transverse momentum quantities are more useful in hadronic collisions as conservation of momentum in the transverse plane is calculable, whereas using conservation of total momentum would require knowledge of the momenta of all constituent quarks in both protons. Conservation of momentum implies that the negative vectorial sum of transverse momentum—known as missing transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$)—is a useful quantity for investigating particles which do not interact with the detector. Momentum reconstruction is made possible by the presence of the superconducting solenoidal magnet, producing a magnetic field of 3.8 T within the detector's inner volume. Charged particle momentum can be measured using the bending angle of the particle traversing the layers of the detector.

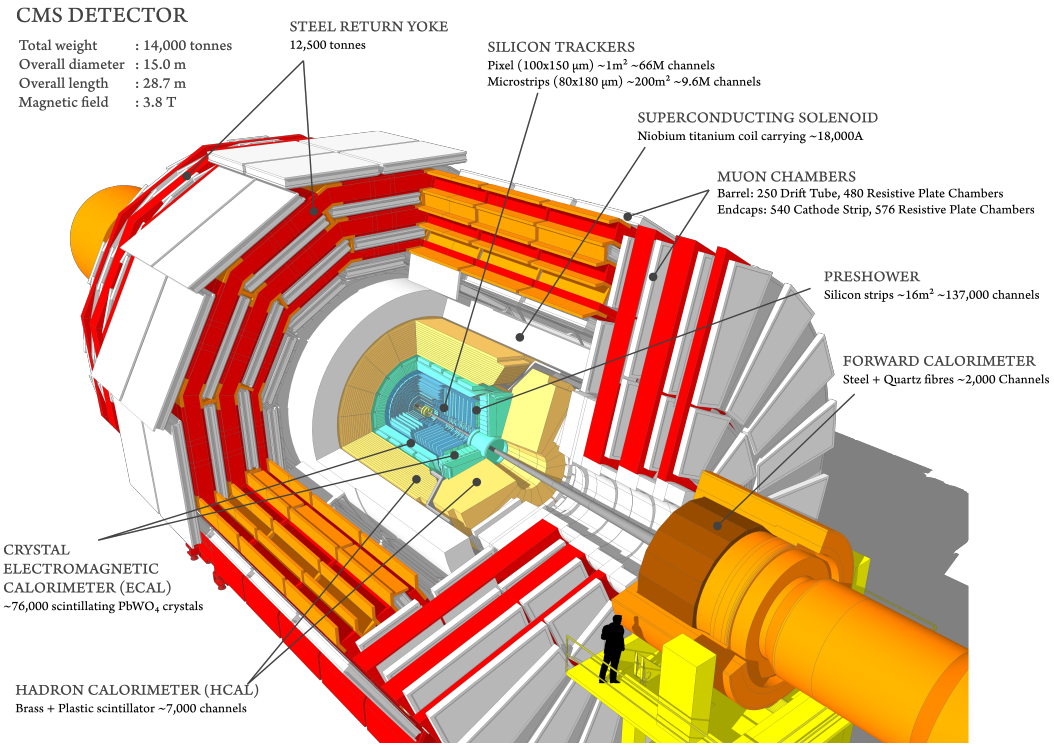

Figure 2.4: The CMS detector is a composite device which is made up of many subdetector systems. Each of the subdetectors is responsible for the identification and measurement of specific particles. Adapted from [3].

CMS is comprised of four main subdetector systems which are responsible for measuring different types of fundamental particles. The four subsystems in the CMS detector are the silicon tracker, the electromagnetic calorimeter, the hadronic calorimeter, and the muon system. For each event, the data from these subsystems are collected and aggregated to create a full picture of the particles produced in an event. The goal of the detector is to detect and reconstruct muons, electrons, photons, and hadrons (charged and neutral) so that events can be characterized based on the number and physical qualities of these objects. The following sections will outline how each of the subdetectors (shown in figure 2.4) measures these objects.

2.2.1 Tracker

One of the most critical aspects in event reconstruction is identifying the precise path of collision products in order to determine the primary vertex (PV) of the collision. The detector responsible for this tracking should also be precise enough to identify whether an object is a product of the collision of interest or from the incidental contact of other protons in the bunches (known as pileup). Therefore, the inner most layer of the CMS detector is a high-granularity silicon tracker which measures the path of charged particles originating from collisions.

In light of the search for the Higgs boson, the tracker was designed with the main focus of identifying muons and electrons from the decays of W and Z bosons. The decay signatures of $Z \to \ell\ell$ and $W \to \ell\nu_\ell$ are characterized by well isolated muons and electrons [13]. Thus, the tracker should be capable of determining if muons and electrons are produced without other objects in close proximity. Additionally, the tracker serves the purpose of identifying showers of electrons from photons, which behave similarly in electromagnetic calorimeters (see section 2.2.2 for more information). So, the tracker should see different signatures from the interaction of these two particles. Finally, the tracker is the closest detector to the interaction point, meaning that it receives a large dose of radiation. Therefore, the tracker must be made of radiation hard material that can withstand these interactions over a long period of time.

These criteria lead to the design choice of the tracker as a high-granularity, silicon detector. This detector is comprised of two layers containing pixels and silicon strip sensors which are responsible for measuring the position of charged particles a distance of less than 20 cm, 20 to 120 cm away from the beamline, respectively. The layers are combined to form a cylinder whose center resides at the main interaction point (IP) of the detector. A diagram of the Tracker subsystem is shown in figure 2.5.

The pixel tracker in the detector barrel is composed of four layers of silicon sensors at a radius of 29, 68, 109, and 160 mm from the beamline in the barrel region. Three layers of pixels are also included in the forward (endcap) regions of the detector. In total, there are 124 million readout channels with dimensions of 100 μm by 150 μm. The total coverage of the pixels is about 1 m^2 with a positional resolution of 11.9 μm in $r - \phi$ and 21 μm in z [14].

The pixel tracker is surrounded by silicon microstrip detectors from 20 to 110 cm

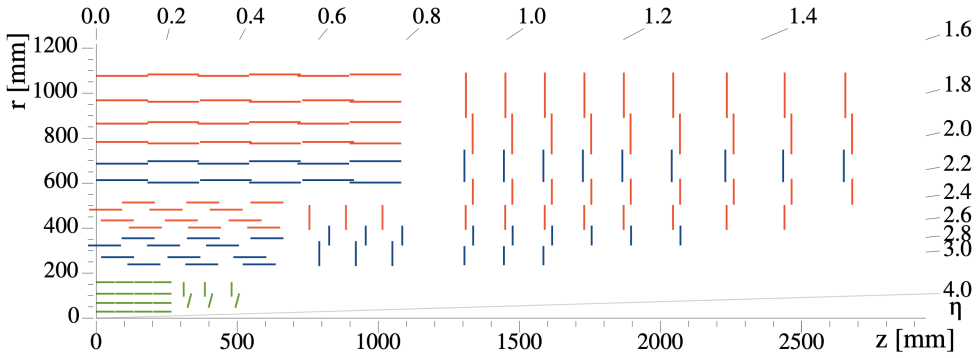

Figure 2.5: The CMS tracker is a cylindrical detector which makes up the inner-most layer of the CMS detector. The Tracker is responsible for collecting positional information of charged particles traversing through the detector. This positional information is then used to reconstruct the tracks of charged particles, which can be used to reconstruct their momentum. A schematic representation of the tracker subsystem is shown above with pixel layers shown in green and silicon strips shown in blue and red. Adapted from [14].

from the beamline. This portion of the detector is segmented into four regions: Tracker Inner Barrel (TIB), Tracker Outer Barrel (TOB), Tracker End Cap (TEC), and Tracker Inner Disk (TID). The TIB and TOB are made up of four and six layers of silicon strips with coverage extending from $|z| > 65, 110cm$, respectively. The first two layers of each are "stereo" layers, which allow for angular measurements in the $r - \phi$ and $r - z$ coordinates. The TID fills in the empty space left by the TIB and TOB with three small discs, the first two being stereo layers. Outer portions of the endcap are covered by the TEC, which adds an additional 9 layers. The first two and fifth layer of the TEC are also stereo layers. This system provides positional resolution from 35 μm (inner region of the TIB) to 500 μm (outer region of the TID) In total, the strip tracker contains 9.6 million silicon strips and covers an area of 200 m^2.

For a muon with 100 GeV of transverse momentum (p_{T}), the tracker is able to measure p_{T} to within 1-7% accuracy while having 10-20 μm and 100 - 1000 μm resolution on the transverse and longitudinal impact parameters, respectively. This yields a global track reconstruction efficiency of around 98% for $\eta < 2$ [10].

2.2.2 Electromagnetic Calorimeter

Another crucial aspect of particle identification and reconstruction is accurate measurement of a particle's energy. One way to glean such information with a particle detector is to use a calorimeter. These detectors measure either light or charge deposited by a particle as it is slowed down via interactions with the detector material. The total energy of a particle can then be reconstructed by summing the energy contributions from the total decay shower [15].

There are two main categories of calorimetery: electromagnetic and hadronic. The former is discussed in this section while the latter is discussed in 2.2.3.

Electromagnetic calorimeters are designed to measure the energy of incident photons and electrons (or positrons). At LHC energy levels ($\sim$ GeV - TeV), the most prominent interactions based on cross section are bremsstrahlung (for incident electrons) and pair production of electrons and positrons (for incident photons). Both interactions produce more electrons (positrons) or photons with reduced energy. Then, these secondary photons and electrons (positrons) interact again via these two processes, creating a cascade of particles known as an electromagnetic shower. Once the energy of the showering particles is sufficiently low, their energy will be absorbed through ionization and excitation as opposed to showering further. This energy can then be measured via charge and light capture.

The shape of electromagnetic showers in a material is dictated by the radiation length (X_0). This quantity describes the average length a particle must travel through a material to have its energy reduced by a factor of $\frac{1}{e}$. That is:

$$\langle E(x) \rangle = E_0 e^{-x/X_0}, \tag{2.9}$$

where $\langle E(x) \rangle$ is the average energy of a particle after traveling a distance x through a material and E_0 is the incident energy of the particle on the material. The radiation length can be determined using:

$$X_0 = \frac{716\text{g cm}^{-2}\, A}{Z(Z+1)\ln 287/\sqrt{Z}}, \tag{2.10}$$

where Z and A are the atomic number and weight of the material respectively.

These principles are utilized in the CMS Electromagnetic Calorimeter (ECAL) to measure the energies of photons and electrons (positrons). One of the primary reasons for including the ECAL in CMS is observation of $H \to \gamma\gamma$, $H \to ZZ^*$, and $H \to W^+W^-$. Additionally, high-energy electrons and photons could be indicators of new physics in the TeV energy regime [16].

The ECAL is the second-closest subdetector to the beamline in CMS. A diagram of the ECAL is shown in figure 2.6. It is comprised of 61200 lead tungstate ($PbWO_4$) crystals in the barrel and 7324 in the endcap, which are responsible for both stopping and measuring the energy of electrons and photons. This material was chosen due to its low $X_0 = 0.89cm$, meaning that electrons and photons are more likely to stop within the ECAL. This detector resides a distance of 129 cm from the beamline in the barrel and 314 cm from the center of the detector in z.

Light is collected by avalanche photodiodes (barrel) and vacuum phototriodes (endcap) which are well suited for the low light yields (30 γ/MeV) of $PbWO_4$. This light is converted to an electrical signal that can then be used for reconstructing the incident particle's energy. The ECAL Endcap (EE) also contains a pre-shower layer of silicon strip sensors to help identify the difference between neutral pions and photons (which have similar signatures).

ECAL is hermetically sealed from the perspective of photons and electrons, meaning that there is at least 10 X_0 of material for any path taken by a particle.

The energy resolution of the ECAL is sub-1% for electrons with an energy of > 20 GeV [10].

2.2.3 Hadronic Calorimeter

Energy measurement of hadrons happens in a similar (albeit more complicated) fashion to electromagnetic calorimeters. In fact, for hadronic energies considered at the LHC, most of the energy measured from hadronic showers comes from photons pair producing, leading to the exact formulation of calorimetry above. As hadrons interact with detector media, they interact both via the strong and (if charged) electromagenetic forces. This results in an intial hadronic shower that is primarily composed of protons, neutrons, and pions. However, as the cascade proceeds and secondary particle energy

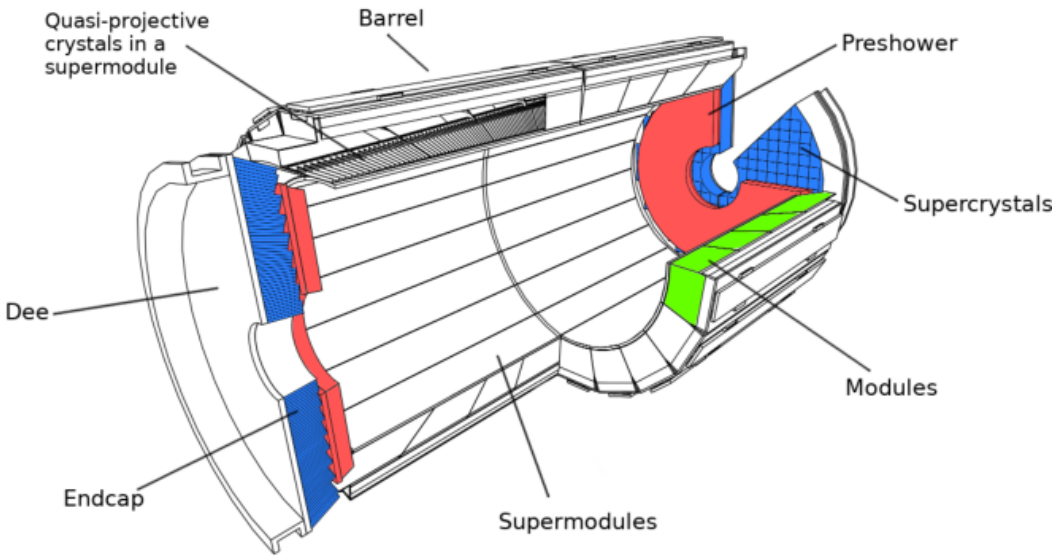

Figure 2.6: The Electromagnetic Calorimeter (ECAL) is the second inner-most subdetector in CMS. This detector is responsible for measuring the energy of electrons and photons produced in proton collisions. Adapted from [17].

drops, the cascade becomes dominated by neutrons, photons (originating from π_0 decays), and electrons/positrons. As noted in section 2.2.2, light from the electromagnetic portion of this decay chain can be measured and combined to form a portion of the incident hadron's energy. However, there is a significant fraction of energy which will be deposited into the detector via nuclear reactions. This energy cannot be measured effectively as the photons emitted from this interaction have a significant time delay ($\sim$1 μs). Therefore, modified techniques are needed to infer the energy and momentum of a particle interacting with a hadronic calorimeter.

There are two main purposes for measuring the energy of hadronic decays in a general purpose particle detector. The first is that all transverse momentum of the particles produced in a collision should sum to zero. Thus, any missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) could be a signature of new stable particles which do not decay to visible matter and do not interact with the detector. This is valuable for many dark matter and supersymmetric particle searches which posit the existence of such particles. Additionally, computing the energy of hadronic showers allows for better reconstruction

of electrons, photons, and muons. So, the inclusion of a hadronic calorimeter in CMS allows for searches for physics beyond the standard model as well as improves particle reconstruction [18].

The CMS hadronic calorimeter (HCAL) measures the energy and direction of hadronic showers of activity (called "jets"). Plastic scintillating tiles are used as the means for measuring scintillation light from particle showers. These are layered between brass absorption plates, which increase the interaction length of material in the HCAL to ensure hadrons shower within the HCAL volume. The plastic tiles are embedded with wavelength-shifting (WLS) fibers which are responsible for light readout. Light traveling on the WLS fibers is collected using hybrid photodiodes (HPDs) in order to convert to an electrical signal. During the Phase 1 upgrade during a year end shutdown in 2017-2018, the HPDs in the HCAL endcap were replaced with silicon photomultipliers (SiPMs), which allowed for better photon detection efficiency during the 2018 data taking period as well as finer granularity which improves radiation damage corrections.

The CMS HCAL is composed of four sub-components: the barrel (HB), endcap (HE), outer (HO), and forward (HF) detectors. The positioning of these four detectors is shown in figure 2.7. Combining these four components creates a hermetically sealed detector, allowing no jets to escape the detector without hitting active material and having at least a partial energy deposit. This allows for a proper reconstruction of the total E_{T} within an event as well as the total amount of $E_{\mathrm{T}}^{\mathrm{miss}}$.

HB is built using 2304 towers that each have a segmentation of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. The cylinder created by these towers is 32 towers long in the z-direction and 72 towers in the ϕ-direction. Each of these towers contains 15 layers of 5 mm brass plates separating the 3.7 mm scintillating tiles with an initial 9 mm layer of scintillator placed directly after the ECAL. Energies are read out longitudinally, generating one energy measurement per tower.

HO is placed just outside the magnet, adding to the hermeticity of the barrel section of HCAL. The purpose of this portion of the detector is to increase the thickness of the detector, catching any hadronic showers which penetrate through HB. Covering the region $-1.26 < \eta < 1.26$, HO is a collection of 10 mm scintillators with 30° segmentation in ϕ. These scintillators are organized into five rings, each wrapping around the detector at different η values. HO adds an additional 10 interaction lengths to the HCAL barrel.

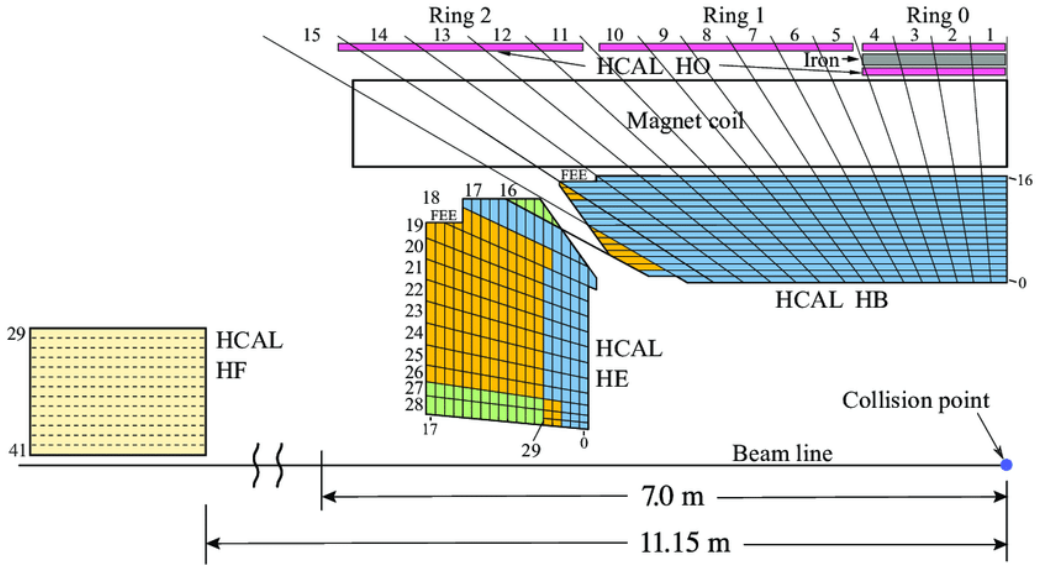

Figure 2.7: The CMS HCAL system is responsible for measuring the energy and direction of hadronic showers in all events. HCAL fills the entirety of the space between the outer edge of the ECAL and the inner radius of the super-conducting solenoid magnet. This diagram represents the segmentation of HCAL prior to the Phase-1 upgrade in 2019. The color of each detector segment corresponds to the associated depth—blue, yellow, green, and magenta for depths 1, 2, 3, and 4, respectively. Adapted from [19].

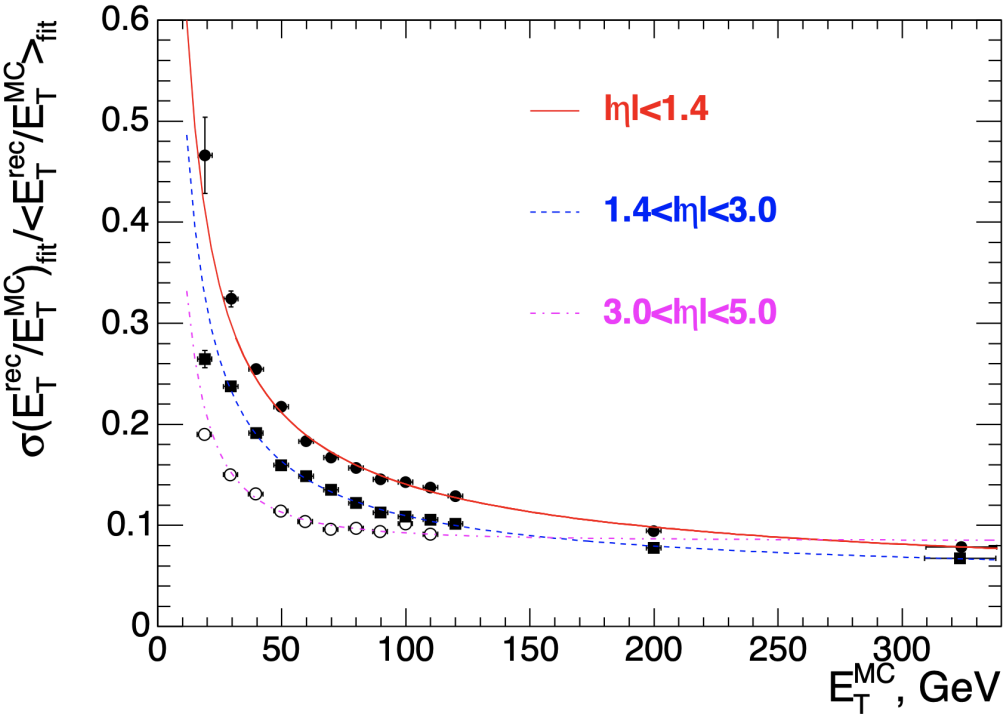

Figure 2.8: The transverse energy resolution of HCAL changes with respect to both position in $|\eta|$ as well as the incoming truth E_{T} of the jet being measured. Adapted from [18].

The HE section is made up of 14 η towers spanning $1.3 < |\eta| < 3.0$ on either end of the detector. There are a total of 2304 towers, each having varying segmentation in both ϕ and η based on their proximity to the beamline.

HF lies in the forward region of the detector with coverage of $3.0 < |\eta| < 5.0$. 900 towers are created using steel absorber plates with quartz fibers running through them along the beamline axis. The quartz fibers measure Cerenkov light created by hadronic showers.

The performance of HCAL is summarized in figure 2.8. The transverse energy of a 100 GeV jet can be measured with a resolution factor of ~ 0.15 in the barrel and ~ 0.1 in the endcap [18].

2.2.4 Muon System

The final particle that cannot be reconstructed via the measurements of the calorimeters is the muon. Muons preferentially decay to electrons and two neutrinos via the weak force. This interaction can be expressed as $\mu \rightarrow W^* \nu_\mu \rightarrow e\nu_e\bar{\nu}_\mu$. Since the mass difference between the muon and the W boson is relatively small (~ 25.4 MeV), muons possess a sizeable mean lifetime of 2.2 μs. Assuming these particles are traveling near the speed of light, this means that the average muon will traverse around 660 m before decaying. Therefore, most muons will travel through the entirety of the detector.

There are three primary interactions which cause energy loss for muons traveling through media. The first is excitation of atomic electrons, which (as noted in section 2.2.3) will cause photon emission that is too slow to measure. The second interaction is bremsstrahlung, which occurs via the exchange of a photon with a nucleus and then radiation of another photon. The final interaction is ionization of atomic electrons. For muons produced by LHC collisions, the dominant energy loss mechanism is ionization, which can be leveraged to measure and reconstruct muons [20].

The main motivation for including a muon detector within CMS was to identify the decay of the Higgs boson to four muons ($H \rightarrow ZZ^* \rightarrow 4\mu$). As muons will lose less energy while traversing the detector, this decay channel provides an excellent probe of the Higgs boson mass. Additionally, high energy muons could be a signature of new gauge bosons, such as heavy Z bosons (Z') [10].

The CMS muon system is the outer-most layer of the detector. A diagram of the different components of the muon system is shown in figure 2.9. It is responsible for tracking muons' flight as they pass through the steel flux return yoke of the CMS magnet. The muon system is layered into the steel of the yoke, and measures the track of the muon as it bends through the outer magnetic field.

There are three different types of detectors that comprise the muon system: Drift Tubes (DT), Cathode Strip Chambers (CSC), and Resistive Plate Chambers (RPC). Each of these utilize the principle of measuring the amount of gas ionized when a muon interacts with the detector. The barrel muon system is made up of five wheels which circle the detector at different η values. These wheels include both DTs and RPCs and have coverage of $|\eta| < 1.2$. The endcap muon system, which is layered with CSCs and RPCs, can measure muons within $|\eta| < 2.4$. Each of these detectors is layered and

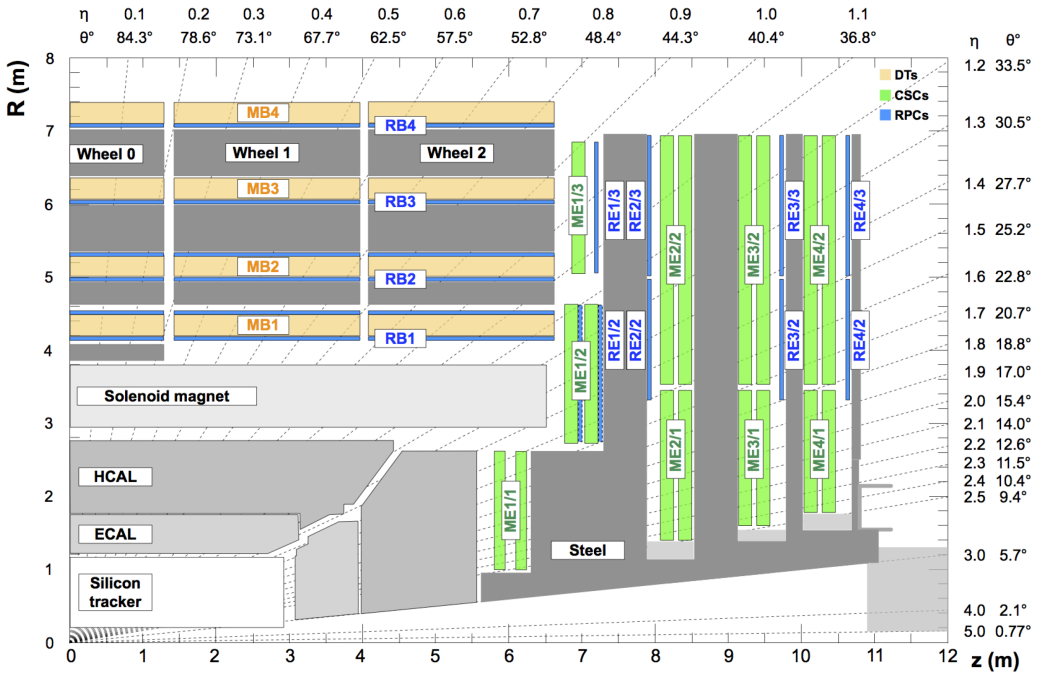

Figure 2.9: The Muon System includes four separate types of gas ionization measurement devices: drift tubes, cathode strip chambers, and resistive plate chambers. These components reside outside the super-conducting magnetic coil and provide tracking information for muons exiting the detector volume. Adapted from [21].

allows for positional measurements of traversing muons.

The transverse momentum of a muon can be identified by considering the bending of its path due to the CMS magnet. By knowing the precise magnetic field strength, one can compute the momentum by identifying the path of the muon via positional measurements. Both the positional information from the tracker as well as the muon system are used for this computation. This results in a total momentum resolution of $\Delta p/p \approx 1\%$ for 100 GeV muons in both the barrel and the endcap [22].

2.2.5 Triggering

Not all proton collisions result in interesting events for particle physics analysis. Only certain signatures will contain physical properties of interest for searches for new physics. Additionally, the computational storage needed for saving every event is unfeasible. The collision rate of the LHC (assuming one collision per bunch crossing and all bunches filled) is 40 MHz. If an event was written for each collision, the total data volume from a single second of data taking would require $O(10^7)$ TB of storage [23]. Therefore, the CMS experiment has developed a triggering system to reduce the total number of events which will be fully reconstructed and stored for further analysis.

The CMS trigger system is responsible for selectively accepting events which are useful for physics analysis and reducing the total data volume to a reasonable size for conducting such analyses. The Trigger is made up of two stages: the Level-1 Trigger (L1) and High Level Trigger (HLT). A diagram of the CMS trigger architecture is shown in figure 2.10.

After digitization of the detector signals, the L1 system identifies the presence of muons, electrons, photons, τ leptons, jets, and $E_{\mathrm{T}}^{\mathrm{miss}}$ in an event. Muon candidates are identified via information from the muon system while the other object candidates are identified using information from the two calorimeters. No information from the tracker is provided to the L1 trigger. The latency of the L1 trigger is 3.8 μs, meaning that information about an event is stored, processed, and either accepted or rejected within 3.8 μs. A user-defined set of 400 selection criteria are applied to an event with a logical "OR" to determine if an event should be considered by the HLT for full reconstruction. Event acceptance is determined by a combination of the muon global trigger and the calorimeter trigger object candidates. This process reduces the data rate to around 100

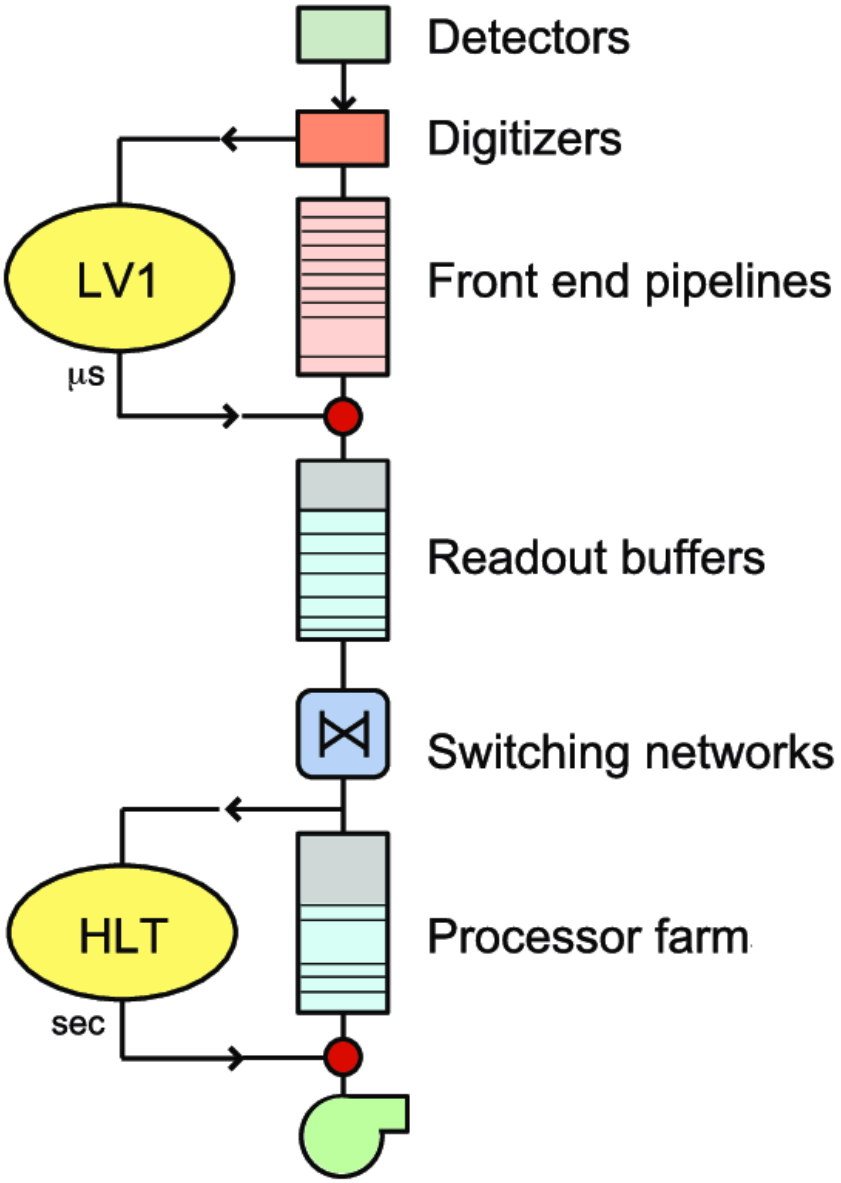

Figure 2.10: The CMS trigger system is comprised of two layers: the L1 and HLT systems. L1 is a hardware based trigger which reduces the overall data rate to 100 kHz. HLT is comprised of a computer farm which runs a fast reconstruction algorithm. After the HLT, the data rate is reduced to around 1 kHz. Adapted from [24].

kHz for events that will be processed by the HLT.

HLT computation takes place on commercial CPUs inside an off-detector computing farm. Events are fully reconstructed using an optimized reconstruction algorithm that is around 100 times faster than the offline reconstruction (more information in section 2.2.6). This algorithm contains a series of reconstruction and filtering modules that are run in succession to identify if an event matches one of hundreds of paths defining interesting topologies. The HLT has a reduction factor of around 100, reducing the overall data rate to 1 kHz [24].

2.2.6 Reconstruction Algorithm

CMS analyses are based on the idea of an "event" which is a full description of the particles which originated from a single proton-proton collision. The digitized detector signals described in the previous sections can be pieced together to identify physics "objects" which allow for easy computation of higher-level variables. This results in reconstruction of charged particle paths (tracks), origins (vertices), and reconstructed particle interactions with calorimeters (hits). The CMS reconstruction algorithm, Particle Flow (PF), identifies which particle type created such signatures and describes the kinematic behavior of each of these objects [25].

The response of each of the subdetectors based on the type of particle traversing the detector is shown in figure 2.11. The path of charged particles will be measured by the tracker, which also allows for charge reconstruction due to the presence of the magnetic field. Electrons and photons shower in the ECAL and are absorbed, allowing for precise energy reconstruction of these objects via energy clusters. Charged and neutral hadrons begin showering in the ECAL and are fully absorbed in the HCAL. The combination of these two energy measurements (as well as tracking information for charged hadrons) is used to recreate the energy of these particles in software. Both muons and neutrinos produce no noticeable signal within either calorimeter. Muons are measured by the tracker and the muon system, yielding additional path information as they traverse the detector's external magnetic field. Neutrinos, being electrically neutral, will traverse the entirety of the detector without leaving a signature.

There are four main classes of physics objects in CMS analyses:

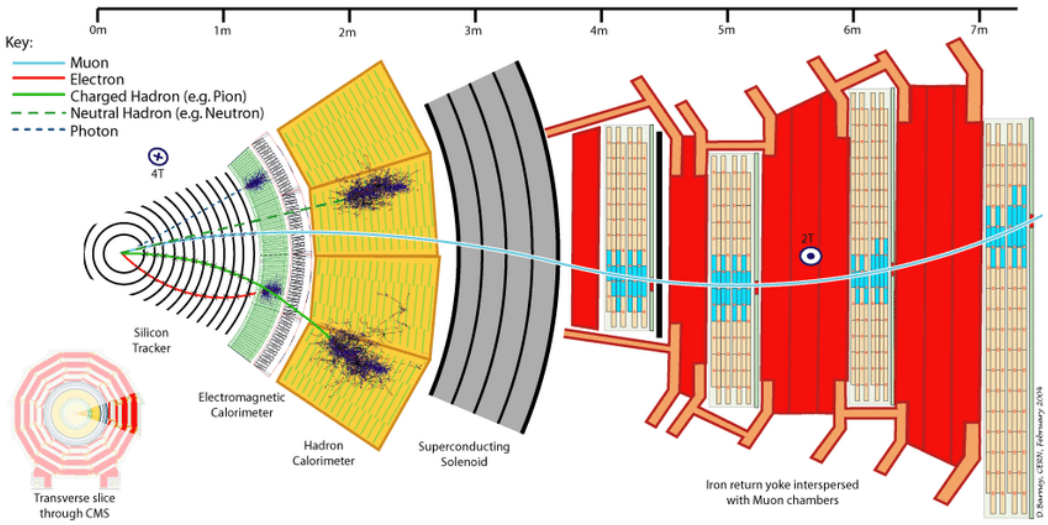

Figure 2.11: Particles interact with multiple subdetectors within CMS. These signatures in combination can help determine which type of particles were created in a collision as well as their momentum four-vector. This information allows for the reconstruction of digital recreations of particles as software objects which can be used for physics analysis. Adapted from [25].

- *Jets* are showers of hadronic activity which are indicators of quarks, gluons, or heavy τ leptons created in the initial interaction. These showers contain hadrons and photons which are measured both in ECAL and HCAL. A combination of tracker and calorimeter information is used to recreate these objects. Additionally, missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$) is reconstructed as a separate object via conservation of momentum.

- *Isolated photons* are identified by energetic hits in ECAL without an associated path from the tracker.

- *Electrons* leave hits within ECAL (similar to photons), but also leave a path within the tracker due to their charged nature.

- *Muons* interact in both the tracker and the muon system, allowing for precise momentum measurements due to their bending angle in the CMS magnet.

Additional information about the properties of hadronic showers allows for the tagging of jet objects. For example, jets can appear from the decay of a B meson, which have a long lifetime. This leads to a secondary vertex that can be identified in reconstruction and used to identify the presence of a b quark in an event. Additionally, other taggers can be used to determine whether a jet originated from a quark or a gluon. Identification of the particle from which a jet showered adds information about the event topology which can be useful in analysis-level selections.

The PF algorithm makes use of all information from each of the subdetectors in order to reconstruct objects. This *linking algorithm* associates information from the tracker, calorimeters, and muons system in order to create a full picture of particles' paths through the detector. Additionally, this combination of information allows for energy-momentum reconstruction for each constituent, creating a relativistic momentum four-vector for each object. Analysis therefore entails computing high-level variables using these objects for identification of physically interesting events.

The other important feature of the PF algorithm is identifying which particles originated from the primary vertex for mitigation of pileup interactions. Within each bunch crossing, there are on average ≈ 40 proton-proton collisions. However, most of these collisions are from incidental contact of protons, creating a noisy background of low

energy objects. This background is defined as *pileup interactions* which must be filtered out of the final event description. By using tracking information to infer if an object originated from the collision of interest, the PF algorithm is also responsible for mitigating the effects of these ancillary interactions.

Chapter 3

Theory

The best description of how fundamental particles interact at the subatomic level is known as the standard model (SM) of particle physics. This theory describes the most fundamental building blocks of nature — the fermions — which make up the entirety of the visible universe, and the force carriers — the bosons — which mediate interactions between them. From this theory, one can make predictions about the behavior of systems of particles, and even posit the existence of new particles or new physics. The most recent success of this framework is the prediction of the existence of the Higgs boson, which was observed at the LHC in 2012 by the ATLAS and CMS experiments [2, 3]. The following sections will discuss the standard model in more detail as well as shortcomings of the standard model, which prompt ideas for possible extensions to the theory.

3.1 The Standard Model

The fundamental particles that make up the basis for the SM are shown in figure 3.1. There are twelve spin-$\frac{1}{2}$ fermions separated into two groups known as the quarks and the leptons. The primary distinction between these groups is that the leptons do not interact via the strong force while the quarks do. The six quarks are organized into three generations, each containing one up-type and one down-type quark: (up, down), (charm, strange), and (top, bottom). Constituents in successive generations have the same physical properties except for their mass. Six leptons comprise the remainder of the

fermions in three similar generations containing a charged lepton and a neutrino. The electron (e), muon (μ), and tau (τ) are the three charged leptons which are identical besides mass. Additionally, each charged lepton has a neutrino counterpart in the electron, muon, and tau neutrino (ν_e, ν_μ, and ν_τ, respectively). Each fermion has an antiparticle counterpart which has opposite electromagnetic charge (and color charge for antiquarks). Thus, there are 24 fundamental fermions which make up the entirety of the visible universe.

Fermions interact via the exchange of four gauge bosons that represent three of the four fundamental forces of nature. The standard model is a quantum field theory that is based on the gauge symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$. This symmetry group includes the eight generators for the strong interaction ($SU(3)_C$) and the four generators for the electroweak interaction ($SU(2)_L \times U(1)_Y$). These generators physically manifest as the eight gluons (g) of the strong force, the $W^\pm$ and Z bosons of the weak force, and the photon (γ) of the electromagnetic force.

The symmetries of the standard model dictate which quantities are conserved by each type of interaction. These symmetries are discussed further in section 3.2. For the strong force, this quantity is known as "color charge" and it can take on three values: red, green, or blue. Thus, only quarks and gluons interact via the strong force as they are the only particles in the SM with intrinsic color charge. Strong interactions are further detailed in section 3.2.1. The $SU(2)_L \times U(1)_Y$ symmetry group introduces the conserved quantities of weak isospin and hypercharge. These quantities dictate the allowed interactions of both the massive gauge bosons ($W^\pm$ and Z) and the photon. Through the spontaneous symmetry breaking of the Higgs mechanism, this symmetry physically manifests as $U(1)_Q$, or conservation of electric charge [27]. Interactions via the weak and electromagnetic force are detailed in section 3.2.2.

The final aspect of the SM is the Higgs mechanism which is responsible for giving mass to the fermions and massive gauge bosons. The mechanism for this is described in section 3.2.3.

Though the SM has produced extremely accurate predictions about the nature of fundamental particle interactions, it still is unable to account for a few phenomena. Theoretical extensions to the SM could ameliorate these issues, making it an even better description of the universe. A few of these shortcomings are described in section 3.3.

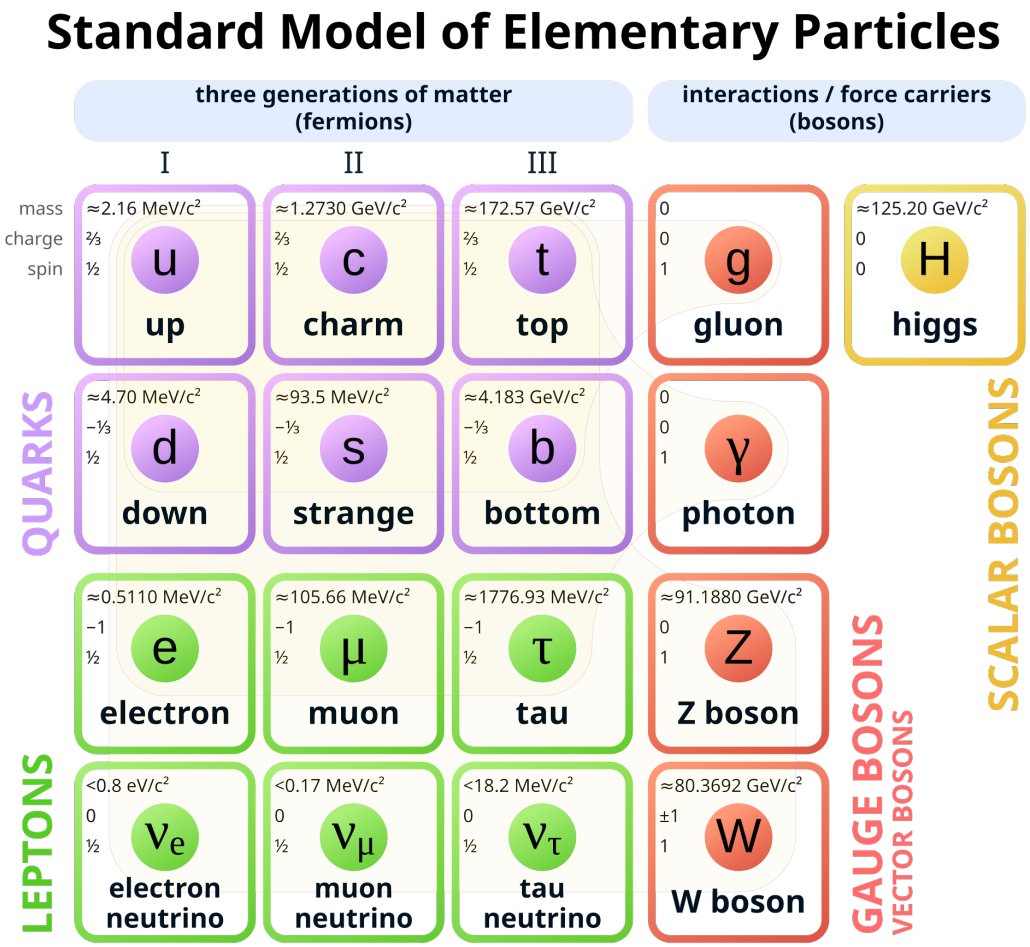

Figure 3.1: The standard model of particle physics explains how fundamental particles interact via three of the four known forces. The SM is comprised of 12 fermions—6 quarks (purple) and 6 leptons (green)—and their antiparticles which interact through the exchange of 5 bosons. The quarks and leptons are comprised of three generations of successively higher mass yet identical quantum numbers. The gauge bosons (red) mediate interactions via the strong, weak, and electromagnetic forces between fundamental particles. The only scalar boson (yellow) of the SM is the Higgs boson, which couples to all particles that have mass. Adapted from [26].

3.2 Formalism of the Standard Model

The standard model is based on a branch of physics known as quantum field theory (QFT) which melds together the ideas of quantum mechanics and relativity. QFT posits that all particles are manifestations of fields that exist at every point in space-time. Interactions of particles are represented by interactions of these fields.

The fields of all free particles and their interactions are described by the *standard model Lagrangian density* (or simply *Lagrangian*, for short). Guided by the principle of least action, observable quantities such as interaction probabilities (known as interaction cross sections) and other observables can be calculated using integration of the Lagrangian. The SM Lagrangian includes terms for three of the four fundamental forces of nature, with gravity being the one exception. It also provides kinetic terms which describe the mass of each of the fundamental particles.

Generation of the SM Lagrangian begins with consideration of the observed symmetries for each of the strong, weak, and electromagnetic forces. Each of these symmetries is associated with a representation group describing how transformations (interactions) occur. As a result of Noether's Theorem, each symmetry of the SM provides a new conserved quantity [27].

The first of these symmetries is represented by the Poincaré group. This symmetry states that the laws of physics should be the same under all spatial rotations, translations, and changes of reference frame (or velocity) [28]. Thus, all particle interactions must conserve energy-momentum. This is known as a global symmetry as it does not depend on the space-time position of any particles [28].

Other symmetries of the standard model are internal meaning that they are associated with the inherent properties of a particle rather than its position in space. Each of the three forces described by the SM can be represented by a symmetry group. For the strong force, this group is $SU(3)_C$ where the subscript C denotes the symmetry associated with color charge, the conserved quantity of the strong interaction. This symmetry can be described as invariance under the rotation of color charge for all strongly interacting particles [28]. The electroweak force which is the unification of the weak and electromagnetic forces is described by the group $SU(2)_L \times U(1)_Y$. Here, the subscripts L and Y denote the conserved quantities of weak isospin and hypercharge [28].

$SU(2)_L$ symmetry represents the rotation of left-handed isospin doublets such as interchanging all up-type quarks with their down-type isospin counterpart, or equivalently interchanging charged leptons with their neutrino counterpart. The $U(1)_Y$ symmetry group is associated with invariance when interchanging positive and negative electric charge [28]. The standard model can thus be represented by the gauge symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$.

The use of conserved quantities allows for a heuristic description of particle physics. That is, particles and their interactions can be fully described by associating quantum numbers to each of the fundamental particles and ensuring that all interactions follow the conservation laws ascribed by the symmetries. In this paradigm, the SM can be fully described in terms of the particles which make up the matter fields and the exchange of force carriers. Then, all possible interactions can be described by associating the exchange of force carriers with the generators of the symmetry group. Only interactions which obey the conservation laws outlined in the SM are physically possible.

3.2.1 The Strong Sector

The theory of interactions via the strong force is described using quantum chromodynamics (QCD). QCD is a non-Abelian gauge field theory which describes the interactions of the quarks via the exchange of massless gluons, the mediators of the strong force.

The group symmetry of $SU(3)_C$ describing the strong interaction gives rise to a conserved quantity known as color charge. Only fundamental particles which carry color charge are able to interact via the strong force. A new quantum number describing color charge is carried by all quarks which can take on three values: red, green, or blue (or equivalently for antiquarks, antired, antigreen, or antiblue). Gluons carry two color charge quantum numbers: one color and one anticolor. There are a total of 8 distinct gluon mediators which represent the 8 generators of the $SU(3)_C$ symmetry group [27].

There are two main features of the strong sector which dictate how interactions occur. The first of these is that the strong force is said to be asymptotically free [27]. That is, interactions between particles become weaker as the energy scale increases or equivalently, the length scale of the interaction decreases. This means that at larger length scales ($\geq$ 1 fm), it becomes more energetically favorable for a quark-antiquark pair to appear instead of increasing the distance between the two existing quarks [29].

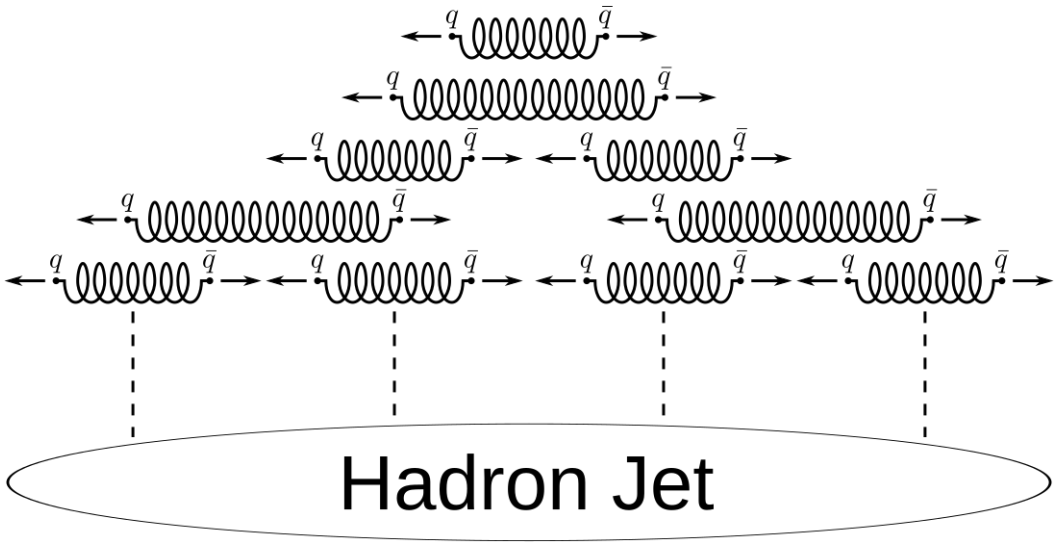

Figure 3.2: Initially free quarks and gluons produced in particle interactions quickly undergo the process of hadronization. Due to color confinement, it is energetically favorable for quarks separated by a distance to produce a quark-antiquark pair which can form colorless mesons or baryons. This process creates a cascade of hadronic particles which are identified in particle detectors as *jets*. Adapted from [30].

Strongly interacting particles are therefore forced into bound states which are colorless, known as color confinement [27]. Only particles which have a color charge of rgb (or $\bar{r}\bar{g}\bar{b}$) or color and anticolor are able to exist in a stable state. These include both baryons which are composite states of three quarks that contain a (anti-)red, (anti-)green, and (anti-)blue quark and mesons which are doublets states including one color and one anticolor of the same charge.

The principles of color confinement imply that free quarks cannot exist in nature. So, initially free quarks which are produced in interactions quickly undergo a process known as hadronization. An illustration of hadronization is shown in figure 3.2. In this process, quarks separate until it is energetically favorable for a new pair of quarks to be produced. These pairs combine with others to produce colorless hadrons. This process continues until all quarks have sufficiently low energy to settle into a colorless hadron. Sprays of hadronic matter of this nature are identified as *jets* in particle detectors and are the signature of quarks originating from the initial interaction.

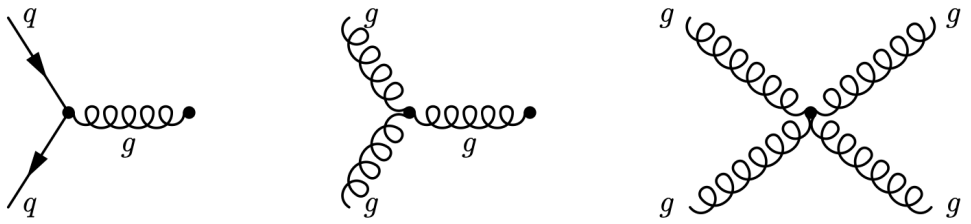

Figure 3.3: All possible strong interaction vertices are shown above with solid lines indicating quarks and helices indicating gluons. Note that strong interactions can occur between all particles that carry a color charge; therefore, tri-linear and quadri-linear couplings of gluons are also allowed in the SM. Adapted from [31].

The set of all possible strong interactions is shown in figure 3.3. Strong interactions between two quarks occur through the process of color exchange where a gluon "carries away" the color of one quark and "exchanges" it with another quark. This amounts to the color charge quantum numbers of the two quarks changing by one unit in the interaction. Additionally, as gluons carry color charge, self interactions of gluons are also predicted by QCD.

3.2.2 The Electroweak Sector

The final two forces described by the standard model are the electromagnetic and weak force. It was shown by Glashow, Salam, and Weinberg that the two forces are actually the manifestation of a single, electroweak (EW) force at energies above 246 GeV [32, 33, 34]. This work lead to the prediction and discovery of the neutral Z boson by the UA1 and UA2 experiements at CERN [35, 36].

Electroweak theory is based on the gauge symmetry group $SU(2)_L \times U(1)_Y$. The resulting conserved quantities of these symmetries are weak isospin (denoted by $I_W^{(3)}$) and weak hypercharge (denoted by Y). These are related to electromagnetic charge via the following relation:

$$Q = I_3 + \frac{Y}{2} \tag{3.1}$$

where I_3 is the third component of weak isospin and Y is the hypercharge [27]. Left-handed chiral up-flavored quarks and neutrinos carry a weak isospin of $I_3 = +\frac{1}{2}$ while

down-flavored quarks and charged leptons carry $I_3 = -\frac{1}{2}$ (and vice versa for the anti-fermions). These particles form weak isospin doublets which are the mathematical objects acted on by the generators of the $SU(2)_L$ group. Right-handed chiral fermions form singlets in weak isospin space and therefore cannot interact with the charged mediators of the weak force. Weak hypercharge arises from interactions with a neutral bosonic field and behaves similarly to electromagnetic charge.

There are three gauge bosons due to the generators of the $SU(2)_L$ gauge symmetry group, which are commonly denoted as $W_\mu^{(i)}$ with $i \in \{1, 2, 3\}$. It is enticing to want to associate these bosons with the commonly known $W^\pm$ and Z bosons of the weak interaction; however, the Z boson couples to right-handed particles while operations of the $SU(2)_L$ only permit interactions of left-handed particles [27]. Therefore, the neutral $W^{(3)}$ is not the physical mediator of the neutral current weak interactions.

The physical bosons of electromagnetic (EM) and weak interactions arise from spontaneous symmetry breaking of $SU(2)_L \times U(1)_Y \to U(1)_Q$ through interaction of the bosonic fields of EW theory with the Higgs field. These bosonic fields are the three $W_\mu^{(i)}$ fields of $SU(2)_L$ and the neutral B_μ associated with the $U(1)_Y$ symmetry. In this formulation, the SM bosons can be expressed as linear combinations of the four EW fields. The physical $W^\pm$ bosons of the SM are:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^{(1)} \mp i W_\mu^{(2)}) \tag{3.2}$$

Physical Z and γ bosons can only be realized through the Higgs Mechanism (discussed further in section 3.2.3). For now, note that the two neutral fields can be expressed as linear combinations of the two neutral fields of the EW sector:

$$\begin{aligned} A_\mu &= +B_\mu \cos\theta_W + W_\mu^{(3)} \sin\theta_W \\ Z_\mu &= -B_\mu \sin\theta_W + W_\mu^{(3)} \cos\theta_W \end{aligned} \tag{3.3}$$

Where A_μ and Z_μ are the fields representing the γ and Z bosons of the SM and θ_W is known as the weak mixing angle. Thus, all physical bosons of the electromagnetic and weak sector can be recovered from the unified EW theory.

One important aspect of weak interactions is the idea of flavor-changing charged current interactions. Due to observation of flavor-changing decays in kaons, Cabibbo

posited that the eigenstates for charged weak interactions via quarks are not the same as the quark mass eigenstates [37]. This lead to the conception of the Cabibbo-Kobayasi-Maskawa (CKM) matrix which is a transformation matrix between the quark mass eigenstates and the weak charged current interaction eigenstates. That is:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{3.4}$$

where d', s', b' are the weak eigenstates and V_{xy} are vertex factors describing the probability of observing a flavor-changing interaction between quark of flavor x and y. Thus, weak interactions with quarks allow for flavor mixing between quark generations.

All of the allowed interaction vertices for the electroweak sector are shown in figure 3.4. Any charged particle (including the $W^{\pm}$ bosons) are allowed to couple to the photon field. Note that all EM interactions explicitly preserve electric charge. Charged weak sector interactions are allowed between isospin doublet partners while any fermion couples to the Z boson. Finally, interactions between the bosonic fields of the electroweak sector are also allowed.

3.2.3 The Higgs Mechanism

The symmetries of the standard model described in the previous section have lead to extremely accurate predictions about fundamental particle interactions. However without any augmentation, the SM has no explanation for the masses of the gauge bosons of the weak force or for the masses of the fermions. Including a kinetic term in the SM Lagrangian would violate the $SU(2)_L$ symmetries associated with the weak force, rendering the predictions of the model unreliable. Therefore, there is a need for another mechanism that can recover the massive nature of these particles.

The idea to introduce a new field in order to restore the masses of the weak gauge bosons is attributed to Higgs, Englert, and Brout [38, 39]. Their theory introduced a new, complex scalar field ϕ to the SM which would be responsible for spontaneous symmetry breaking of the EW sector at low energies. This field is subject to a potential

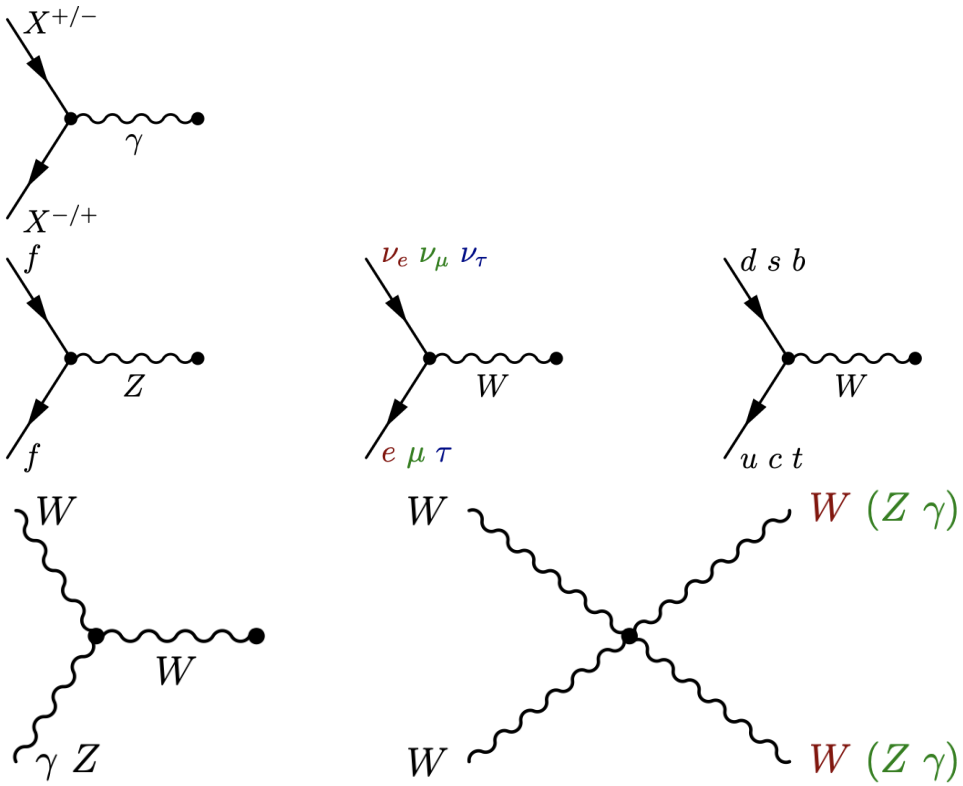

Figure 3.4: All allowed SM interactions between fermions and EW bosons are shown above. Note that all fermions can interact with the $W^{\pm}$ and Z bosons while only charged particles interact with the photon. There are additionally couplings between the bosons of the weak and EM sectors. Note that f and X refer to any fermion and any electrically charged particle, respectively. Adapted from [31].

of the form:

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 \tag{3.5}$$

which for $\mu^2 < 0$ exhibits an infinite set of minima with a non-zero vacuum expectation value. A graphical representation of this potential is shown in figure 3.5

Including the interactions of the new scalar field into the SM Lagrangian results in a few phenomena. The possible SM interaction vertices between the Higgs boson and other fields are shown in figure 3.6. First, a new kinetic term including only self-interactions of the Higgs field describes a new, spin-0 boson [27]. This is known as the

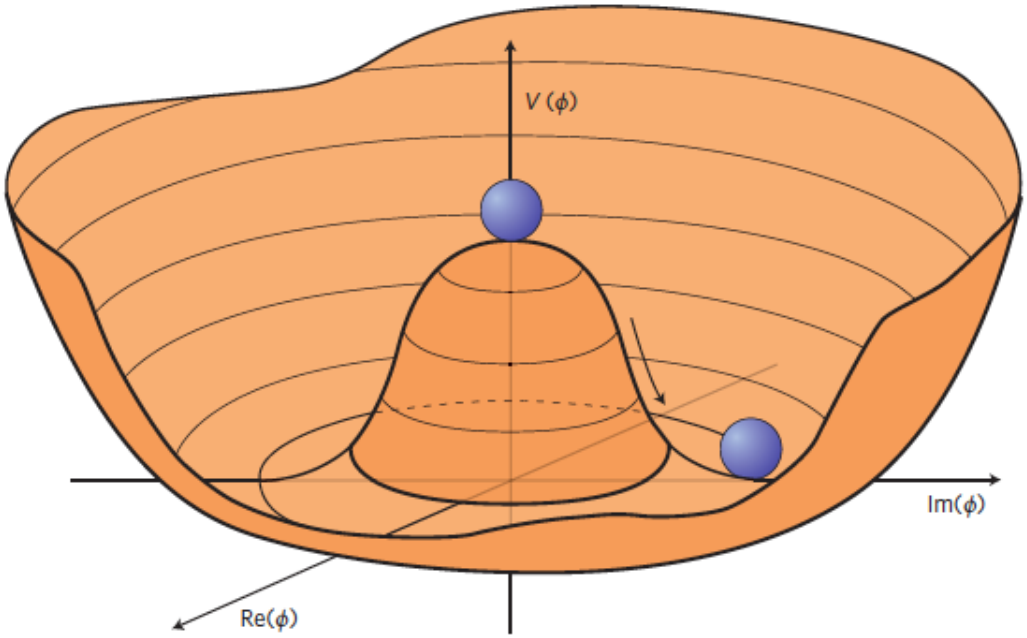

Figure 3.5: The Higgs potential has an infinite set of minima where $\phi^\dagger\phi = \frac{v^2}{2} = \frac{-\mu^2}{2\lambda}$. This minimum v corresponds to the non-zero vacuum expectation value (VEV) of the potential. The fact that the Higgs potential has a non-zero VEV leads to the spontaneous breaking of the local EW gauge symmetry. Adapted from [40].

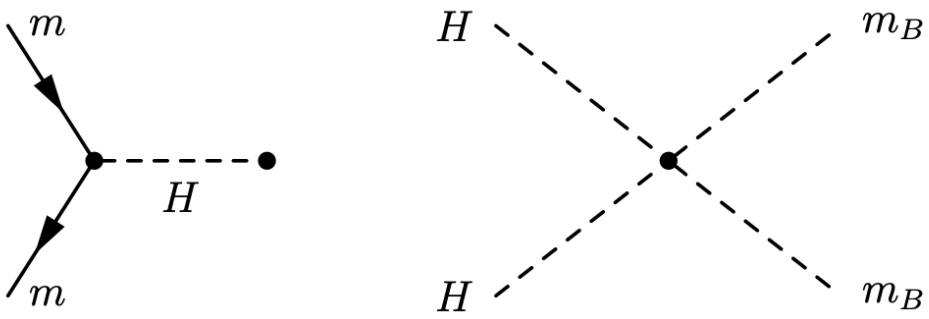

Figure 3.6: Interaction vertices are shown for the couplings of the Higgs boson to both fermionic (left) and bosonic (right) matter. Note that the strength of these interactions is proportional to the mass of each of the particles. Adapted from [31].

Higgs boson, which is the mediator of interactions in the Higgs sector. Additionally, interactions of the Higgs field with the EW bosons results in the inclusion of mass terms for each of the weak vector bosons. The mass values of these bosons can be related to the EM or weak coupling constant and the constants of the Higgs potential:

$$m_W = \frac{1}{2} g_W v,\ m_Z = \frac{1}{2} v \sqrt{g_W^2 + g'^2},\ m_\gamma = 0 \tag{3.6}$$

where $v = 246$ GeV is the Higgs potential VEV, g_W is the weak coupling strength, and g' is the EM coupling strength. Thus, the spontaneous symmetry breaking of the Higgs mechanism recovers the masses of the EW sector bosons. Finally, terms including interactions between the Higgs field and the fermions of the SM are also included in the Lagrangian. This gives rise to the masses of the fermions as each fermion couples to the Higgs field with a strength proportional to its mass.

3.3 Shortcomings of the SM

Though the SM has made extremely accurate predictions about the interactions of fundamental particles, it still lacks a description of a few observed phenomena. Most notably, there is no explanation for gravity in the SM. Additionally, observations of galactic rotation curves and gravitational lensing effects hint at a minimally interacting form of matter known as "dark matter" which has no SM equivalent [41]. This section

Figure 3.7: All fermions and bosons contribute corrections to the Higgs boson mass through quantum loops. Fermions (shown on the left) and bosons (right) contributions to the mass differ by a factor of -1. Adapted from [42].

will outline one particular shortcoming of the standard model which can be resolved through supersymmetry as a hypothesized extension to the SM.

3.3.1 The Hierarchy Problem

The SM has withheld the test of searches for new physics up to an energy regime of around $\sim$ 1 TeV. However, it is apparent that at a small enough length scale (or equivalently, a high enough energy scale), gravitational effects can no longer be ignored in particle interactions. These effects become important at an energy scale known as the Planck mass, $\Lambda_P = (8\pi G_{Newton})^{-1/2} = 2.4 \times 10^{18}$ GeV [42]. Thus, there are 16 unexplored orders of magnitude which could potentially be hiding new physics.

Consider the effects that a new particle would have on the particles of the SM. Given that the particle is massive, it must necessarily couple with the Higgs boson which results in corrections to the Higgs boson mass. These corrections enter through the loop diagrams in figure 3.7.

The mass correction from a fermion or boson with a mass Λ is given by:

$$\begin{aligned} \Delta m_{H,f}^2 &= -\frac{|\lambda_f|^2}{8\pi^2}\Lambda_{UV}^2 + ... \\ \Delta m_{H,S}^2 &= \frac{\lambda_S}{16\pi^2}\Lambda_{UV}^2 + ... \end{aligned} \tag{3.7}$$

where Λ_{UV} is the ultraviolet cutoff energy scale at which the SM is no longer an accurate prediction of nature. Note that the coupling of the fermion or bosons, λ_f or λ_S, increases

as the new particles mass increases. Any particle heavier than the top quark (i.e. greater than ~ 200 GeV) would have a value of $\lambda \gtrapprox \mathcal{O}(1)$

Considering that the corrections to the Higgs mass are of order $\Lambda_{UV}^2 \sim \mathcal{O}(10^{30})$, it is difficult to imagine that all additional terms would result in a Higgs mass of 125 GeV. However, if a symmetry existed which relates fermions and bosons, the difference in sign of the mass corrections could be leveraged to generate an exact cancellation of all new contributions. This theorized symmetry is known as *superysymmetry* and will be discussed further in section 3.4.

3.4 Supersymmetry

Supersymmetry (SUSY) is an extension to the SM which theorizes a previously unidentified symmetry between fermions and bosons. SUSY transformations take the form:

$$Q\,|\text{Boson}\rangle = |\text{Fermion}\rangle\,,\ Q\,|\text{Fermion}\rangle = |\text{Boson}\rangle\,; \tag{3.8}$$

That is, the operator of the supersymmetric transformation (Q) changes bosonic states into fermionic states and vice versa. This operator acts on supermultiplets which contain both a fermion and boson that are superpartners.

SUSY suggests that for every standard model particle, there is a supersymmetric superpartner particle that differs by 1/2 unit of spin. These new particles and their SM counterparts are displayed in figure 3.8 All bosonic particles in the supersymmetric sector are denoted with an s prepending the name of it's SM counterpart, and the fermionic particles' supersymmetric particles are identified with the suffix *ino*. So, the supersymmetric partners of the fermions and bosons are the *sfermions* (e.g. squark, slepton, etc.) and *bosinos* (e.g. photino, gluino, etc.). The sfermions are all scalar particles, meaning that they are spin-0 particles. Gauginos of the standard model are all spin-$\frac{1}{2}$ particles except for the superpartner of the (hypothesized) graviton, which carries a spin value of $\frac{3}{2}$. Other than difference in spin, supersymmetric counterparts carry the same quantum numbers as their superpartner.

In the minimally supersymmetric standard model (MSSM), five separate Higgs boson mass eigenstates are required to preserve the electroweak gauge symmetry of the SM. This results in a two charged ($H^{\pm}$), two neutral Higgs bosons (H^0 and h^0), and the

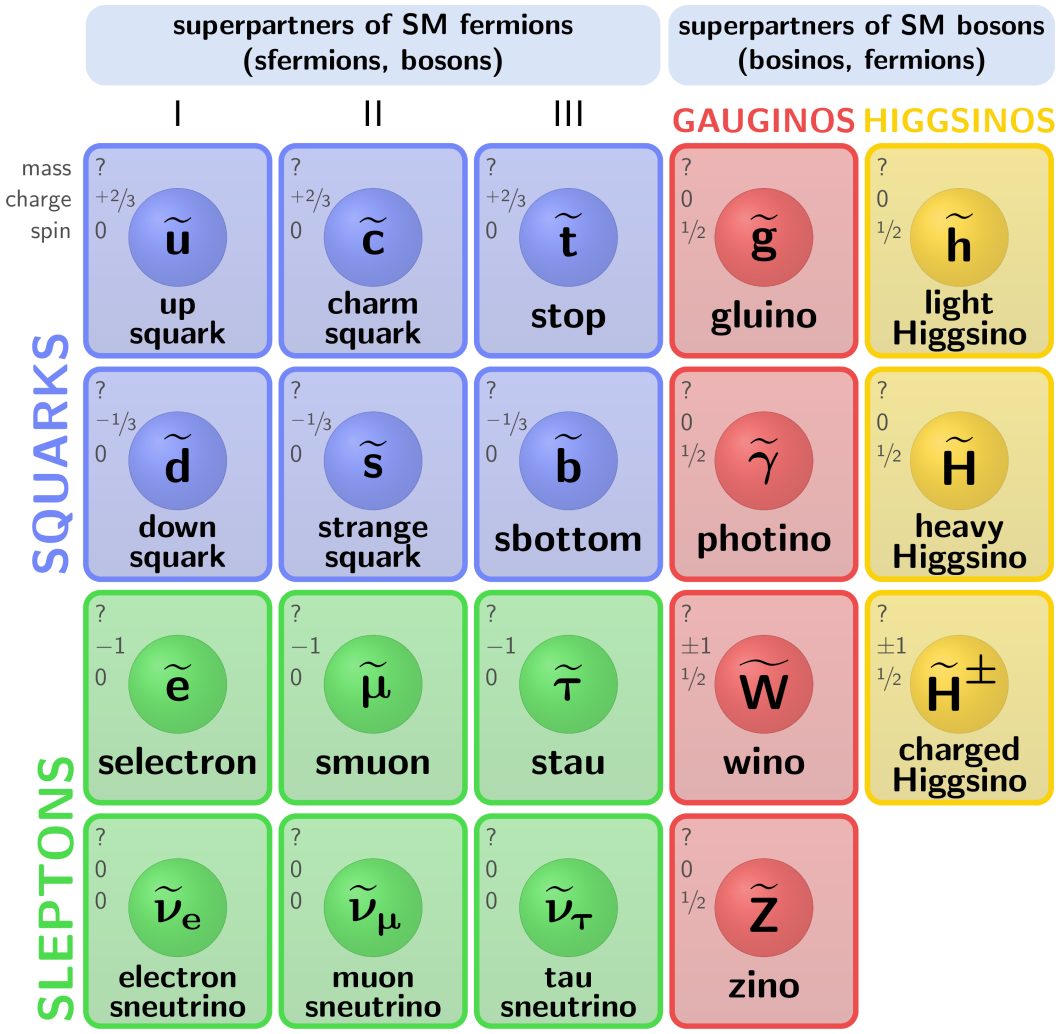

Figure 3.8: SUSY posits that there is a nearly identical superpartner for each standard model particle. Adapted from [43]

.

CP-odd A^0. Additionally, note that the higgsinos and bosinos outlined in figure 3.8 are the eigenstates of the gauge representation. The physical MSSM bosinos are linear combinations of these sparticles and are known as the neutralinos (χ_i^0 with $i \in \{1, 2, 3, 4\}$) and charginos ($\chi_j^{\pm}$ with $j \in \{1, 2\}$) [42]. The lightest of the neutralinos (χ_1^0) is assumed to be stable only if R-parity (described below) is conserved.

The effects of this extension to the SM become apparent when considering its effect on the Higgs boson mass corrections of equation (3.7). If SUSY is an exact symmetry, then SM particles and their superpartners would be mass degenerate. Thus, the difference in sign between the bosonic and fermion loop mass corrections would yield an exact cancellation. This is an enticing theory as it would resolve the hierarchy problem without fine tuning of the SM parameters.

However, if superpartners which are mass degenerate with their standard model counterpart existed, they would have been detected in experiment. As no discoveries have been made, it must be the case that if SUSY exists, it is not an exact symmetry. That is, SUSY must be a broken symmetry such that the masses of the superpartners are different than their SM counterparts. This can be accomplished by asserting that the symmetry breaking is soft. That is, the symmetry breaking mass terms of the SUSY Lagrangian must depend on some mass scale m_{soft} which produces corrections to the Higgs mass that do not diverge as $m_{\text{soft}} \to 0$ [42]. Therefore, it is possible to have sparticles which are heavier than their SM counterparts without ruining the benefits of SUSY.

As a consequence of supersymmetry, a number of baryon and lepton number violating interactions are now possible. This implies that the proton is no longer stable as SUSY interactions allow for the process $p \to e^+ + \pi^0$ as shown in figure 3.9.

However, no experimental observations of proton decay have been made. Therefore, if SUSY exists, there must be some mechanism that prevents this from occurring.

Introducing a new conserved quantity known as R-parity re-establishes the stability of the proton. R-parity is defined as:

$$P_R = (-1)^{3(B-L)+2s} \tag{3.9}$$

where B and L are the number of baryon and lepton number and s is the spin of a

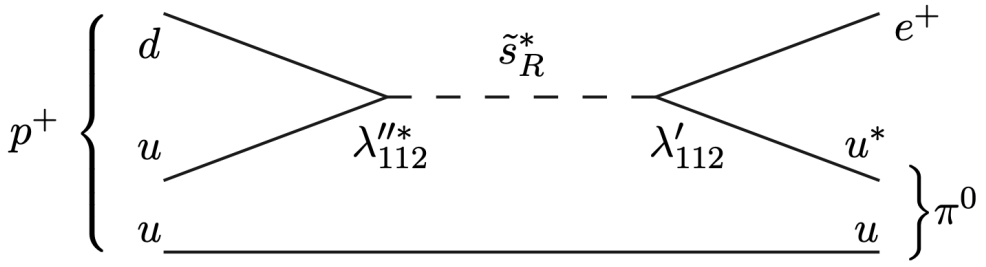

Figure 3.9: Without a constraint on R-parity, the proton is allowed to decay via an off-shell strange squark. However, no such decays of the proton have been observed to this point. Imposing R-parity conservation on all MSSM interactions removes this decay channel. Adapted from [42].

particle in the interaction [42]. With this definition, all sparticles incur an R-parity value of $P_R = -1$ while all SM particles have the value $P_R = +1$ Thus, imposing that R-parity is conserved in all interactions is equivalent to conserving the number of SM and SUSY sector particles in the final state. This implies that the lightest sparticle is stable (known as the lightest supersymmetric particle or LSP).

3.4.1 R-Parity Violating Supersymmetry

Many experiments have attempted to identify possible supersymmetric interactions by targeting the LSP. The interaction (or lack thereof) between the stable LSP and active material in detectors would lead to an imbalance in the amount of transverse moment (p_{T}) from a particle collision. Searches for missing transverse momentum (or $E_{\mathrm{T}}^{\mathrm{miss}}$) have failed to identify evidence of supersymmetric decays resulting in an LSP. As a result, alternative theories about how SUSY could present itself have been theorized.

One theory hypothesizes that the LSP has not been identified because it is allowed to decay via R-parity violating (RPV) decay modes with small couplings. This would indicate that the LSP could decay into a SM only final state, eluding all searches for SUSY via identification of $E_{\mathrm{T}}^{\mathrm{miss}}$. Certain RPV decays can be allowed while also preserving the stability of the proton given that no interactions occur which simultaneously violate baryon and lepton number [44].

The allowed interactions of RPV SUSY are shown in figure 3.10. Three three couplings are introduced which allow decays from supersymmetric states into strictly SM particles: λ and λ' indicate the coupling strength of interactions which violates lepton number and λ'' indicates an interaction which violated baryon number. Note that these interactions allow superpartners of the SUSY sector to decay to strictly SM particles. These RPV decay modes can result in fully SM final states that could be hidden among the SM background at particle collider experiments.

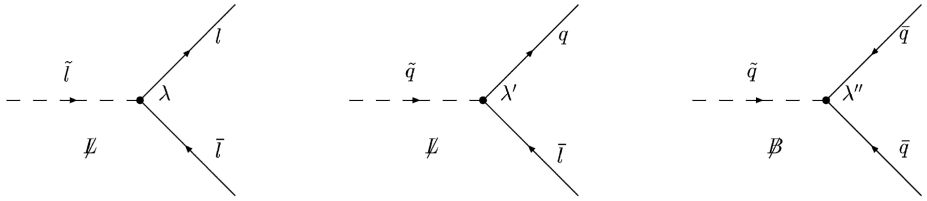

Figure 3.10: The three tri-linear interactions between quarks, leptons, squarks, and sleptons are shown above. All interactions result in final states which contain only SM quarks or leptons. Adapted from [44].

The RPV decay signature of interest for this analysis assumes that the one of the λ'' is non-zero, allowing for decays of superpartners to strictly standard model quarks [44]. Therefore, top squark pair production would result in final states indicated in figure 3.11. These final states contain SM particles only, which will manifest as hadronic jets in the detector. Additionally, the final state contains two top quarks.

3.4.2 Stealth Supersymmetry

Alternative theories assume that SUSY could be eluding traditional search techniques in other fashions. One such theory describes "stealthy" SUSY where decays of supersymmetric particles result in final states with low $E_{\mathrm{T}}^{\mathrm{miss}}$. This would explain how searches for SUSY in high $E_{\mathrm{T}}^{\mathrm{miss}}$ final states have been unable to see any hints of new physics.

Such final states manifest by assuming that there is a hidden sector of particles which act as a portal between the SUSY sector and the SM sector [45]. The simplest such model includes a new scalar particle known as the singlet (S) and its superpartner the singlino ($\tilde{S}$). These two particles are nearly mass degenerate in the case that SUSY

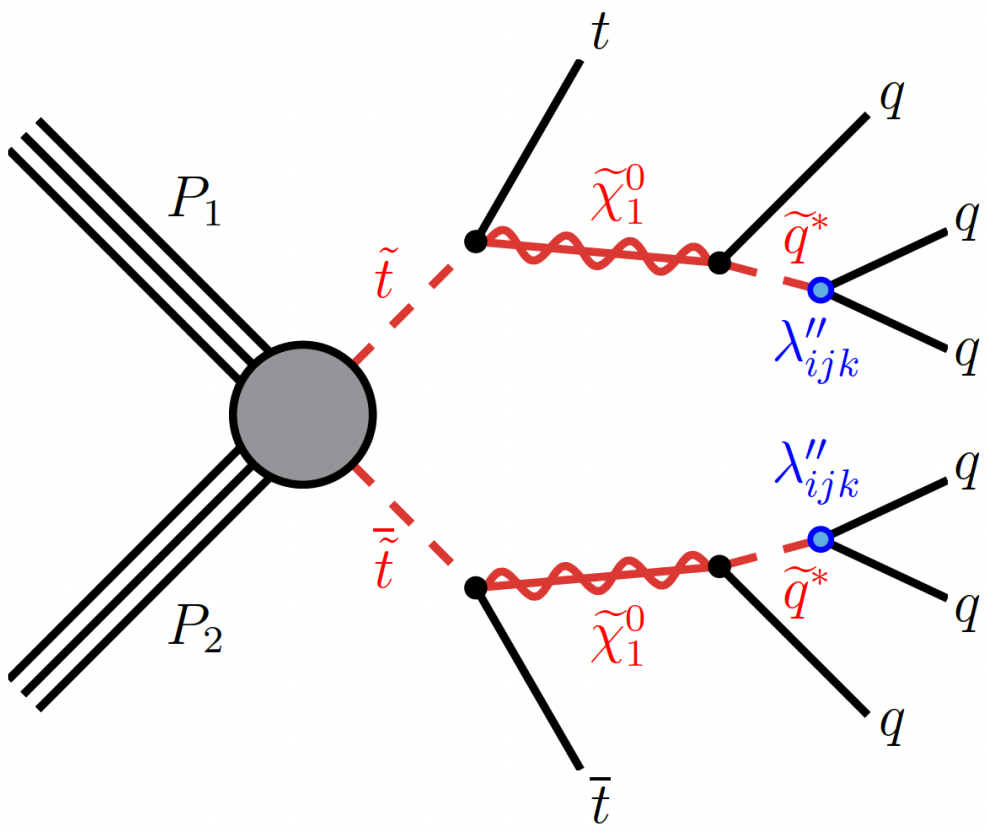

Figure 3.11: The decay of two top squarks via the RPV couplings are shown in the Feynman diagram above. Note that the RPV couplings between the off-shell squark and the two quarks strictly violates R-parity conservation. The final state from this decay contains two top quarks and up to six quarks from the hard process.

is broken at a low energy scale [46]. Thus, the decay of the singlino to a singlet and gravitino—the supersymmetric partner of the hypothesized graviton—would result in a light gravitino with minimal momentum evading detection. As the gravitino is stable, R-parity is still conserved in this model. The singlet will decay primarily to a pair of gluons with a branching fraction of $\sim 100\%$ [45]. The result of production and decay of SUSY particles will thus produce final states with little to no $E_{\mathrm{T}}^{\mathrm{miss}}$ and many SM sector particles.

The model considered in this analysis is known as the Stealth $SY\bar{Y}$ model. The

lightest ordinary superpartner (LSOP) is assumed to be the top squark, which is allowed to decay via the quadra-linear interaction shown in figure 3.12. Interactions between the LSOP and the singlet occur only through loops of new vectorlike particles Y and $\bar{Y}$. The blue blob in figure 3.12 represents an effective vertex for the decay $\tilde{t} \to t\tilde{g}^* \to t\tilde{S}g$.

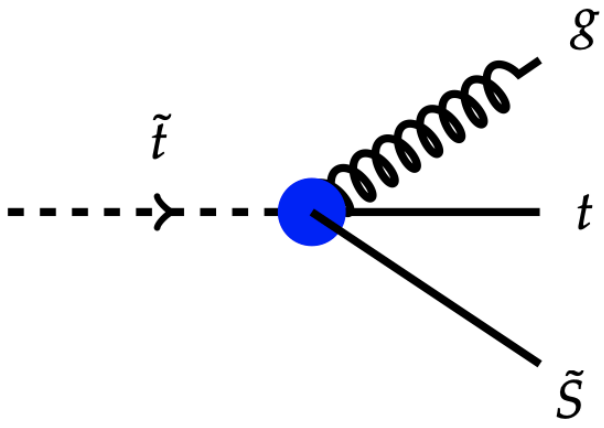

Figure 3.12: Stealth $SY\bar{Y}$ SUSY allows for the top squark to decay via a gluon, top quark, and singlino. Adapted from [45].

The decay of theorized top squark pairs produced via proton-proton collisions is shown in figure 3.13. Similar to the RPV model discussed in section 3.4.1, the Stealth $SY\bar{Y}$ decay of the top squarks results in SM only final states save for the gravitino which leaves no identifiable signature. Thus, a search for this signature can be carried out by looking for an excess of events in a search region with two tops, many light flavored jets, and no additional $E_{\mathrm{T}}^{\mathrm{miss}}$.

3.5 Standard Model Backgrounds

In order to identify signal events in data, it is necessary to understand what known SM processes can mimic a signal-like detector response. For the two signal models considered, the primary background of interest is top/anti-top production in combination with extra jets from initial- or final-state radiation, referred to as $t\bar{t} + jets$. A tree-level diagram for $t\bar{t}$ production is shown in figure 3.14.

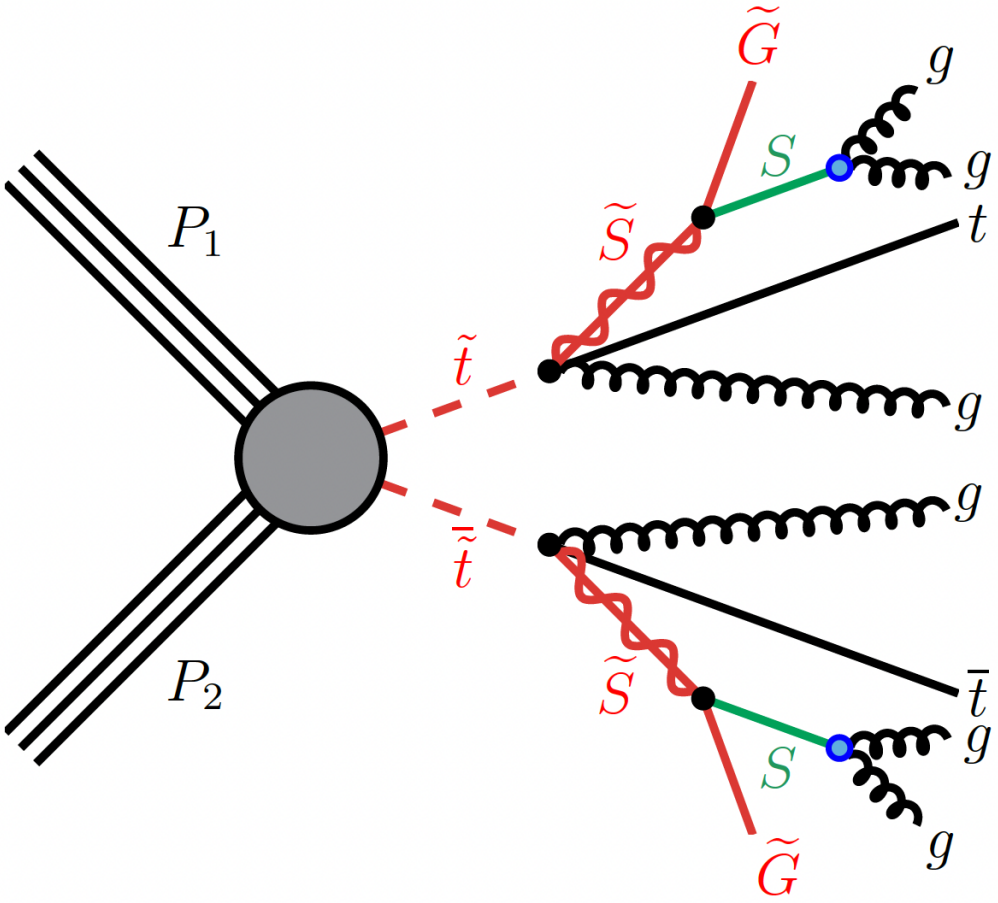

Figure 3.13: Top squarks decaying via the Stealth $SY\bar{Y}$ signal model are shown above. The gravitino ($\tilde{G}$) is assumed to be light and invisible as a result of the near-mass-degeneracy of S and $\tilde{S}$. The final state from this interaction has two top quarks and six hard process jets from gluons.

When a pair of top quarks is produced, the decay proceeds through a bottom quark and a W boson. The W boson has the possibility of decaying to two quarks of the first two generations (u, d, s, c) or to a charged lepton and the corresponding flavor neutrino. It is expected that top quarks will decay hadronically (i.e. to only quarks) about 66.5% of the time, with the remaining decays resulting in leptons [20].

Similarly, the top quarks in the RPV and Stealth $SY\bar{Y}$ models are expected to have

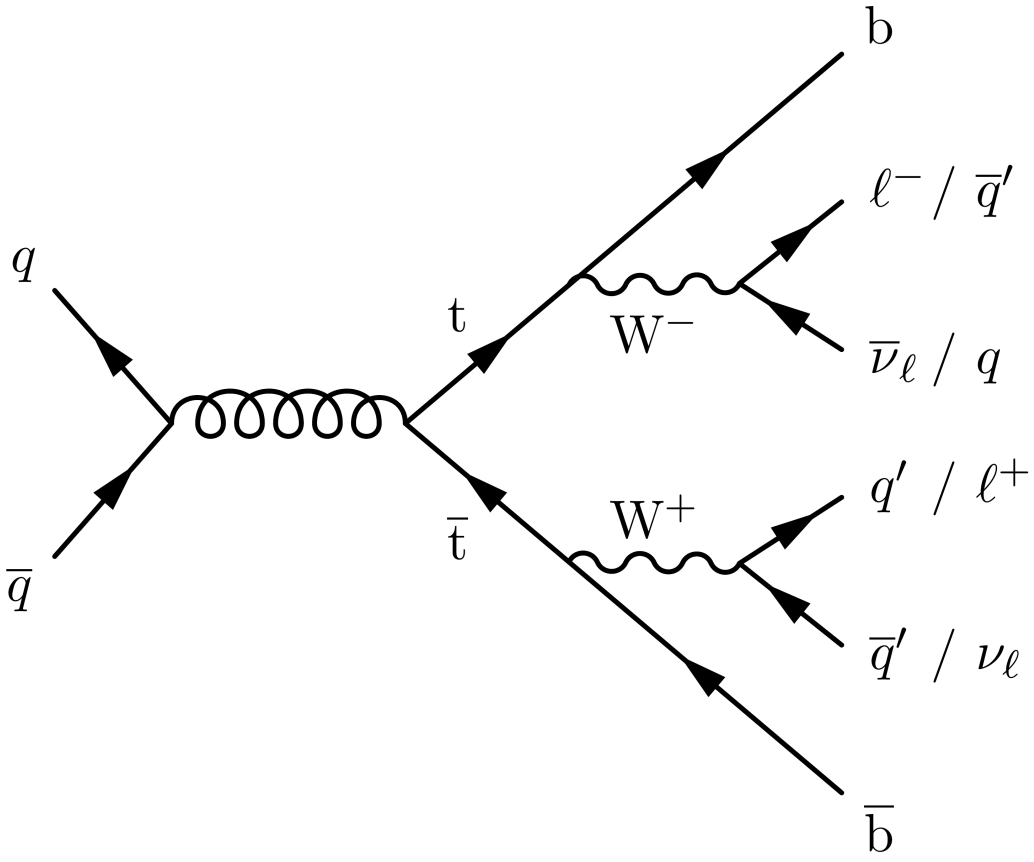

Figure 3.14: Pair production of top quarks results in a final state which looks like signal events. Top quarks decay to a W boson and bottom quark, with the W boson subsequently decaying to either a charged lepton and neutrino or to two quarks of the first two generations. Additional jets can be produced through initial- and final-state radiation of gluons. Adapted from [12].

the same decay modes as those in $t\bar{t}$. Comparing the final state particles in figures 3.11 and 3.13 to figure 3.14, it is expected that signal events would contain up to six more jets than $t\bar{t}$ events. However, additional quarks and gluons can be radiated by either the incoming protons or the outgoings top quarks and hadronic decay products through the mechanisms of initial- and final-state radiation (ISR and FSR, respectively) [47]. These processes are shown in figure 3.15. The result is $t\bar{t} + jets$ events with more jets than expected from $t\bar{t}$ events at the tree level, and therefore can manifest as signal-like

events.

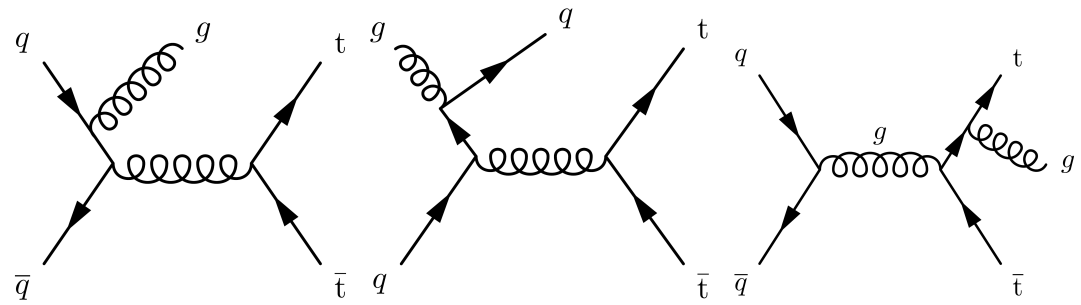

Figure 3.15: Quarks and gluons originating from ISR (left and center) and FSR (right) result in $t\bar{t} + jets$ events which have a similar number of jets to the two signal model processes. This SM process is the most similar standard model background to the signal models presented and, thus, will present challenges during background estimation. Adapted from [12].

Thus, $t\bar{t} + jets$ is the largest background in this analysis due to its similarity to the two signal topologies presented above. Understanding this background and developing selections which mitigate its presence in the search region are two of the most crucial aspects of this search.

Chapter 4

Analysis Overview

The goal of searches for new particles is to determine if there is a significant amount of data which cannot be ascribed to a known particle physics process. In the case of this analysis, this would amount to checking if the number of high jet multiplicity events expected from known SM processes matches with data. The general strategy for carrying out a particle physics search is described below:

1. Design a set of signal region selection criteria to maximize sensitivity to signal
2. Estimate the number of background events expected in the region of interest from known SM processes
3. Predict the number of signal events expected in the search region using Monte Carlo simulation of the signal process
4. Estimate statistical and systematic uncertainties on these estimates due to statistical fluctuations as well as known detector issues, physics modeling shortcomings, or other systematic effects
5. Compare the prediction with uncertainty to data and determine whether there is an excess of events consistent with the presence of signal events

Designing such an analysis for the RPV and Stealth $SY\bar{Y}$ models begins with identification of unique aspects of these signal processes which differentiate their detector signatures from background events. The most differentiating aspect is the number of

jets (N_{Jets}) produced by $\tilde{t}$ decays, which is on average larger than for SM backgrounds. The total number of jets produced in signal topologies is either 6, 8, or 10, depending on the decay channel of the W bosons produced by the top quarks. This is in contrast to $t\bar{t}+jets$ events, which only result in 2, 4, or 6 jets from the hard process. Thus, N_{Jets} is one of the most important variables and will be used as the basis of the background prediction in order to gain signal sensitivity.

Additional methods can be used to identify signal events in data. Namely, jets originating from the decay of a $\tilde{t}$ are generally expected to have larger p_{T} than those from tops due to the high mass of the $\tilde{t}$. So, the sum total of all transverse momenta for jets in each event (known as H_{T}) can be used as an additional handle for signal identification. This feature additionally results in different angular separation of objects in events for signal. Lastly, signal events are expected to have two bottom quarks coming from the decay of the top quarks. Tagging jets produced from B mesons and applying selection cuts on their kinematic behavior can allow further reduction of background in our signal region. Placing selections on these criteria generates a search region with reduced background.

Three separate signal region definitions are defined based on the number of leptons present in events. These are the fully-hadronic (0ℓ), semi-leptonic (1ℓ), and fully-leptonic (2ℓ) channels which mirror the three possible decay modes of the two top quarks in signal events. Different techniques are used to define the baseline signal region selections in each of the channels, which are discussed in section 5.3.

Even after these selections are applied, there are still far too many background events in the search region for the analysis ($O(100)$ times more background than signal). So, a neural network based approach is used to classify events as either signal-like or background-like. This improves the sensitivity of the analysis by taking into account correlations between object and event level variables which cannot be utilized in a simple selection-based analysis.

There are additional difficulties associated with the signal topologies of interest that must be overcome. Namely, accurately simulating the number of jets becomes increasingly difficult for high-N_{Jets} events. The majority of additional jets are produced via ISR and FSR, which become more difficult to model as the number of additional jets increases. Therefore, a data-driven background estimation method is needed to ensure

that the number of expected background events in the high-N_{Jets} regime is properly represented.

Additionally, the similarity between these two signal and $t\bar{t} + jets$ processes reduces the possibility of using a signal-free control region for background estimation. To combat this, a common high energy physics analysis strategy called the "ABCD" method is used. This strategy entails imposing selections on two uncorrelated analysis variables, creating four separate regions. A transfer factor is then computed for each of the three signal channels in order to generate a prediction for the number of events in the final signal region. In this search, a novel approach to the ABCD method is used where the two independent variables are the output of two neural networks. This strategy is discussed in further detail in chapter 7.

The final results of the analysis are determined by a binned maximum-likelihood fit to the N_{Jets} distribution. This fit determines if there is evidence that events consistent with the decay of $\tilde{t}$ via the RPV or Stealth $SY\bar{Y}$ exist in data. Additionally, upper limits on the cross-section of proton-proton collisions to these signal topologies are calculated. The fitting procedure is detailed in section 7.4.

Chapter 5

Event Selection

5.1 Data and Simulated Samples

Accurate reconstruction and identification of analysis level objects is imperative for performing a valid high-energy physics analysis. First, the Particle Flow algorithm (described in section 2.2.6) is used to convert the digital signals created by particle interactions with the detector into analysis level physics objects. Then, objects are filtered using analysis-specific definitions. These definitions are called "object definitions" and are used as the baseline criteria for establishing quality requirements for different particle species. Finally, selection criteria are placed on low- and high-level analysis variables to establish a low-background signal region to carry out the analysis. This results in a final signal region containing only signal and background events which match certain criteria characteristic of the signal topologies of interest.

The following sections discuss the motivations for the signal region selections of this analysis. They delineate the process for data collection, generation of simulated data sets, selecting objects, and defining signal region selections. Note that this is only a baseline selection which reduces the number of background events in the signal region to a reasonable level. Further optimizations are applied to reduce background for the final analysis, which are discussed in chapter 7.

5.1.1 Data

This analysis searches for R-parity violating and Stealth SUSY in CMS data collected in the 2016, 2017, and 2018 run periods. This time period is all part of the second major data taking campaign of the LHC and, referred to as "Run 2". In comparison to the first data collection period ("Run 1"), the center-of-mass energy for collisions was increased by nearly a factor of 2 to $\sqrt{s} = 13$ TeV. The total amount of data available for analysis is measured by the integrated luminosity ($\mathcal{L}_{int}$). This corresponds to a total integrated luminosity of 138 fb^{-1} for the Run 2 data taking period. A plot of the total integrated luminosity over Run 2 is shown in figure 5.1.

Note that detector conditions change from year to year which leads to different calibrations for offline event reconstruction. Thus, each year of data is analyzed separately with the correct per-year calibrations applied before combination in the final analysis. Years are further subdivided based on post-facto corrections of malfunctioning detector equipment. Most notably, the 2016 data taking period is split into two separate eras due to a malfunctioning preamplifier chip in the silicon tracker [49]. This is accommodated by changing the offline reconstruction operating conditions for the period of data taking before and after the problem occurred. Thus, 2016 data is split into 2016preVFP (before adjustment of the feedback preamplifier bias voltage) and 2016postVFP to accommodate the difference in run conditions.

Three datasets are used in this analysis, corresponding to the possible decays of top quarks in signal events. The `JetHT` dataset contains events with a large amount of hadronic activity (H_{T}), corresponding to events with many jets which matches the signature of top squarks decaying fully hadronically. `SingleMuon` and `SingleElectron` datasets contain events with at least one muon/electron passing minimum p_{T} and quality requirements. These datasets are used to search for top squarks decaying to a final state with one or two leptons. Each of these datasets are used for each of the data taking years in the final analysis. The integrated luminosity for physics analysis purposes by era is 19.5, 16.8, 41.5, and 59.8 fb^{-1} for 2016preVFP, 2016postVFP, 2017, and 2018, respectively.

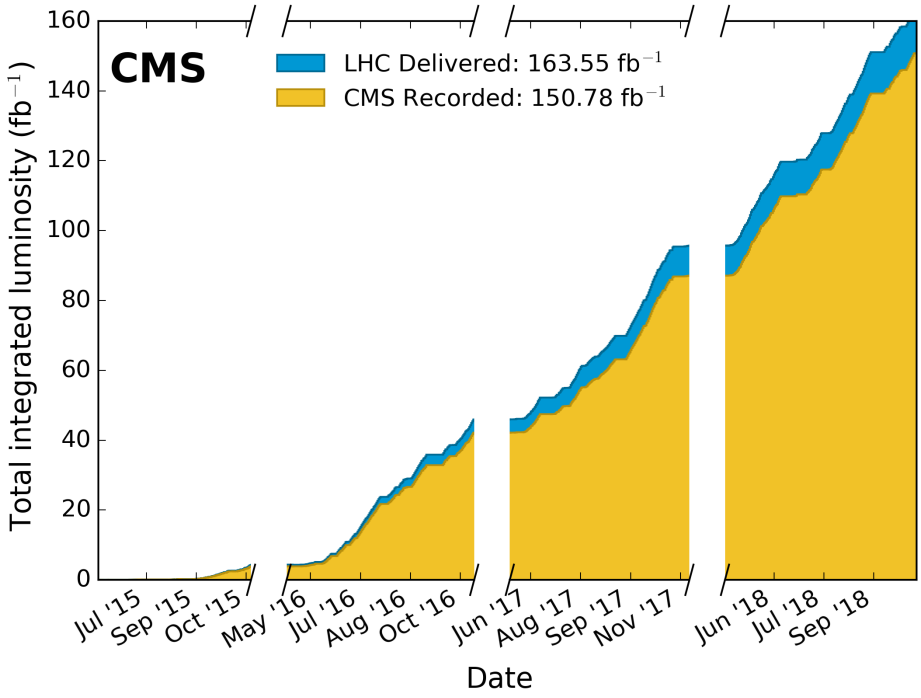

Figure 5.1: The integrated luminosity is shown for the Run 2 data taking period. Two distributions are shown corresponding to the luminosity delivered by the LHC and that recorded by CMS. The difference in these distributions is attributed to detector downtime. Note that data taking in 2015 was intended for establishing the data collection procedure and is therefore not used in this analysis. Gaps in the distributions correspond to technical stops between data taking years. Adapted from [48].

5.1.2 Simulation

Simulated samples are produced for estimation of both the signal and background event contributions to the total signal region event yield. The process for simulating collisions and their interaction with the detector is as follows:

1. Model how protons will interact to initiate a given physics process
2. Simulate the passage of the resulting products from the initial interaction as they travel through detector media
3. Recreate the expected response of the detector to incident particles
4. Reconstruct the physics objects from the simulated detector response information

This results in a simulated dataset that can be used as a direct comparison to data. Ideally, simulated events would match data perfectly, but this is often not the case due to detector malfunction, mismodeling of the underlying physics, or other factors. These are effects are mitigated using scale factors which adjust the relative contribution of a simulated event (or weight) based on dedicated measurements. Uncertainty estimates on these scale factors are included as systematic uncertainties on the background and signal event yield estimates.

Events are first generated using either POWHEG [50, 51, 52, 53, 54, 55] or MadGraph [56, 57, 58, 59] to varying degrees of precision. These programs are responsible for producing the correct interaction probability for a specified set of incoming and outgoing particles. Parton distribution functions are included in this calculation to simulate the internal momentum of the proton constituents which are responsible for the interaction. These are supplied by the NNPDF group [60]. Then, the program PYTHIA is used to simulate the hadronization and showering of the final state quarks and gluons which will be identified as jets [61]. PYTHIA is additionally responsible for simulating the decays of the top squarks in signal events. GEANT4 is then used to simulate the interaction of particles with the detector and determine how the interactions would be digitized into electrical signals [62]. Finally, the digitized signals are run through the ParticleFlow algorithm to generate an object-level description of the events.

Often, it is valuable to produce more simulated events than would be expected in data such that rare processes are represented more frequently. In doing so, the relative

weight of a simulated event must be less than an equivalent data event (defined to have event weight of 1). Thus, simulated events are scaled based on a ratio of the number of events generated and the true interaction probability times the luminosity. That is:

$$w_P = \frac{\sigma_P \mathcal{L}_{Data}}{N_P} \tag{5.1}$$

where P denotes a given process, σ is the interaction cross-section, $\mathcal{L}_{Data}$ is the integrated luminosity of the analysis dataset, and N is the number of events generated. Note that this scaling can also increase the relative importance of an event if the number of events simulated is less than the number of expected events in data.

Cross section calculations for the RPV and Stealth $SY\bar{Y}$ signal models are computed to next-to-next-to-leading-order (NNLO) plus next-to-next-to-leading-logarithm (NNLL) precision [63, 64]. For the RPV model, the mass of the $\tilde{\chi}_1^0$ is set to 100 GeV during generation. $SY\bar{Y}$ events contain both the singlet and singlino, which are assumed to have a mass splitting of 10 GeV and $M_{\tilde{S}} = 100$ GeV. For both models, signal samples are generated with $M_{\tilde{t}}$ between 300-1400 GeV in steps of 50 GeV. Standard model processes have cross sections which are calculated to NLO or NNLO precision [65, 66, 67, 68, 69, 70, 71].

5.2 Object Definitions

Object definitions are a set of criteria that an object must pass in order to be considered within the analysis. Selections are imposed in order to prevent "fake" objects from incorrectly contributing to background or signal event yields. These criteria are based on quality of reconstruction, object p_{T}, and location in the detector. Only PF candidate objects which pass the object definition requirements will be considered in the analysis.

Reconstruction quality cuts are based on pre-established thresholds defined by CMS physics object groups. In most cases, three sets of quality selections are defined internally for objects in CMS: loose, medium, and tight. The quality level chosen for most CMS analyses are based on recommendations from physics analysis groups guided by expected event kinematics for the signal models of interest.

There are three additional criteria which should be highlighted. First, objects have a minimum p_{T} requirement. This is imposed to remove soft objects which are the result

of pileup interactions. Additionally, a selection is placed on $|\eta|$ such that all objects are within the region of coverage of the tracker. This is important as it ensures that objects used in the analysis are reconstructed based on criteria from more than one subdetector. Finally, objects are required to be isolated within the detector. That is, they should be well separated from other objects to ensure that they are not the product of some secondary decay.

Below are the object definitions for electrons, muons, and jets. Additionally, definitions are provided for top quarks which decay before interacting with the detector. However, a machine learning algorithm is used infer the number of top quarks in an event. This is useful for improving the signal selection efficiency for the fully-hadronic decay channel.

5.2.1 Electrons

Electrons are subject to the "tight" identification selection [72] to increase the purity of the sample. Also, electrons are further selected based on their p_{T} and $|\eta|$. Electron isolation is calculated using the mini-isolation (miniIso) algorithm [73] which is a useful quantity for determining isolation in high occupancy events. Last, impact parameter cuts are applied to enforce that the electron arises from the primary interaction vertex using two dimensions—the transverse (d_0) and longitudinal (dz) distances from the primary vertex as given for both barrel and endcap regions. The relevant selections for "good" electrons are shown in table table 5.1.

Table 5.1: Good electron selection criteria for the signal region are given below.

Good Electron Selections
"Tight" cut-based electron ID
$p_{\mathrm{T}} > 30(37)$ GeV for 2016 (2017/2018)
$\|\eta\| < 2.4$
miniIso < 0.1
$\|d_0\| < 0.05$ (0.10) cm for barrel (endcap)
$\|d_z\| < 0.10$ (0.20) cm for barrel (endcap)

5.2.2 Muons

Muons are identified using the "medium" identification selection [74]. Also, muons are further selected based on their transverse momentum and pseudorapidity given as $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$. Mini-isolation for muons [75] is defined like that for electrons. Finally, there are impact parameter cuts, as given in table 5.2, which are based on transverse (d_B) and longitudinal (d_z) distances from the primary vertex.

To create control regions for the QCD multijet background estimation, non-isolated muons are specifically defined for the analysis. Non-isolated muon selection criteria with the other selections above are listed in table 5.2.

Table 5.2: Good muon selection criteria for the signal regions as well as Non-isolated muon selection criteria for the QCD control region are shown below.

Good Muon Selections	**Non-isolated Muon Selections**
"Medium" cut-based muon ID	"Medium" cut-based muon ID
$p_{\mathrm{T}} > 30$ GeV	$p_{\mathrm{T}} > 55$ GeV
$\|\eta\| < 2.4$	$\|\eta\| < 2.4$
miniIso < 0.2	miniIso > 0.2
$\|d_B\| < 0.2$ cm	$\|d_B\| < 0.2$ cm
$\|d_z\| < 0.5$ cm	$\|d_z\| < 0.5$ cm

5.2.3 Jets

Jets are clustered with the anti-k_T algorithm [76], using a distance parameter of 0.4 (AK4) and charged hadron subtraction. Exact p_{T} and η requirements on jets for the individual channel event selections can be found in section 5.3. Additionally, each jet passes "JetID" [77] requirements to ensure the quality of jet reconstruction. Finally, any jet that is within an angular separation (ΔR) of 0.4 of an identified and isolated lepton and whose transverse momentum is within 100% of the p_{T} of the lepton, is removed from consideration.

For determining if a jet originates from a b quark, the discriminant value computed by the deep flavor (DeepFlavour) algorithm [78] is used. The "medium" working point with a 1% mistag rate is chosen for classifying a jet as a b jet.

5.2.4 Tops

Considering the final state of the signal process, the two top quark decays represent non-trivial resonances that can be identified and used to reject many-jet background events such as those originating from QCD multijet processes. To reduce the QCD multijet background, a hadronic top tagger developed for [79] is implemented as a part of the signal region selection to tag hadronically decaying top quarks.

Top quarks preferentially decay though the process $t \to b + W$ with a branching fraction near 1 [20]. Top quarks can then perform a fully-hadronic decay if the W boson decays into two quarks. Thus, identification of a top quark amounts to determining if there are two quarks which originated from a W boson in the vicinity of a b jet.

The tagger used in this analysis considers two categories of top quarks based on their p_{T}, as shown in figure 5.2. Low-p_{T} tops can be identified as three AK4 jets, with one jet originating from a b quark. In the high top p_{T} (or boosted) regime, the three constituent jets become collimated and can be clustered together in an AK8 jet which uses a cone of distance parameter $\Delta R < 0.8$. These two categories are known as resolved and merged tops, respectively.

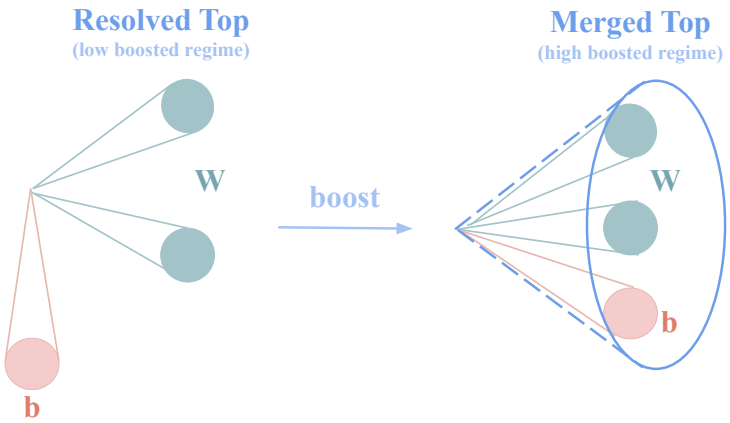

Figure 5.2: Resolved and merged top quarks are identified using different tagging algorithms. Resolved top quarks consists of three individual AK4 jets while merged top quarks consists of three jets which are clustered into a single AK8 jet.

So, two separate tagging algorithms are used to determine the presence of either

resolved or merged tops. The working points for these taggers are optimized specifically for signal sensitivity in the fully-hadronic analysis channel. Top candidates which pass the working point selection criteria are used in the signal region event selection for this channel.

5.3 Signal Region Definition

The signal region selection criteria are shown in table 5.3. The primary feature separating the three regions is the number of isolated leptons ($N^{\text{iso}}_{\text{leptons}}$) in the final state. Categorizing by number of isolated leptons improves signal sensitivity in the leptonic channels where fewer background events are expected than in the 0ℓ channel. An additional selection requires exactly zero non-isolated leptons which allows for orthogonality between the signal regions and QCD control region (described in section 7.2.1).

Table 5.3: Below are the signal region selection criteria for the 0ℓ, 1ℓ, and 2ℓ channels (left to right). The three channels are differentiated by both number of isolated leptons and number of jets in the final state.

Selection criteria		0ℓ	1ℓ	2ℓ
$N^{\text{iso}}_{\text{leptons}}$		0	1	2, oppositely charged
$N^{\text{non-iso}}_{\text{muon}}$		0	0	0
H_{T} (GeV)		> 500	> 500	> 500
N_{Jets}	($p_{\text{T}} > 30$ GeV)	≥ 8	≥ 7	≥ 6
	($p_{\text{T}} > 45$ GeV)	≥ 6	–	–
$N_{\text{b jets}}$	($p_{\text{T}} > 30$ GeV)	≥ 2	≥ 1	≥ 1
	($p_{\text{T}} > 45$ GeV)	≥ 1	–	–
N_t		≥ 2	–	–
$M_{b\ell}$ (GeV)		–	$> 50, < 250$	–
$M_{\ell\ell}$ (GeV)		–	–	< 81 or > 101
$\Delta R_{\text{b jets}}$		> 1	–	–

The most distinguishing feature for identifying signal events in data is the number of jets (N_{Jets}) in the events. Note that the object definition of jets changes in the 0ℓ channel in order to match trigger requirements. A number of the triggers also include b jets. So, a selection is placed on the number of b jets ($N_{\text{b jets}}$) in events as well.

As top squarks are assumed to have large mass, the pair will be produced nearly at rest. Thus, the jets originating from the top squark decay should exhibit large p_{T} as a consequence of conservation of momentum. Therefore, signal events are expected to have large $H_{\mathrm{T}} = \sum_{jets} p_{\mathrm{T}}$.

For the 0ℓ channel, there are two channel specific selections. First, all 0ℓ signal region events should contain at least two top quarks ($N_t \geq 2$) identified by the tagging algorithms detailed in section 5.2.4. Additionally, the two b jets should have a ΔR value of greater than 1 to avoid events with pairs of collimated b quarks arising from gluon splitting. Both of these selections are imposed to reduce the QCD background contribution to the signal region.

The 1ℓ channel includes a selection on the invariant mass of the b jet and isolated lepton. This mass should loosely fall near the mass of the top quark in signal events. This will remove events which have a b jet and lepton that do not originate from a top quark decay.

Finally, the 2ℓ channels selection includes a Z boson mass peak veto. That is, the invariant mass of the two isolated leptons should not be within 10 GeV of the Z boson mass. This reduces the number of background events containing pair produced leptons known as Drell-Yan interactions.

Events which pass the signal region selection are sorted into four process categories. $t\bar{t} + jets$ events correspond to top anti-top production with extra jets from ISR and FSR. QCD multijet events have strong interactions which are able to produce a high number of jets due to gluon radiation or gluon splitting. $t\bar{t} + X$ events are top anti-top production in conjunction with a boson ($X = Z, W, H$). The other category includes all other SM processes.

Pie charts representing the relative contribution of known SM processes for the three channels are shown in figure 5.3. After filtering all simulated events through these three sets of selection criteria, the dominant background for all channels is $t\bar{t} + jets$. Methods for further improving signal sensitivity in the three channels as well as the final background estimation techniques are presented in chapter 7.

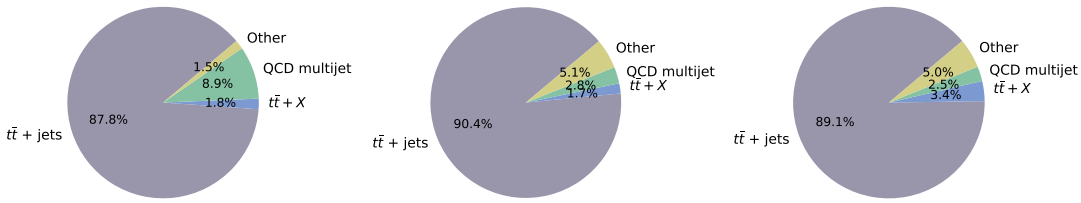

Figure 5.3: The composition of backgrounds in the three signal regions (left to right: 0ℓ, 1ℓ, and 2ℓ) is shown below. The largest background for all three channels is $t\bar{t} + jets$.

Chapter 6

The ABCDisCoTEC Neural Network

As mentioned in chapter 4, this analysis presents a unique challenge as it is difficult to accurately simulate events with large jet multiplicity. Additionally, the signal models of interest closely resemble the $t\bar{t} + jets$ process, making signal events difficult to distinguish from background. Therefore, this analysis employs a novel, neural-network-based background estimation technique to simultaneously improve signal sensitivity and establish a means for performing a background estimation using data events.

The "ABCDisCoTEC" method (ABCDisCo Training Enhanced with Closure) leverages a novel category of loss functions which enforces independence of model outputs in order to establish a background estimate. The following sections will outline the motivation, construction, training, and results of the ABCDisCoTEC model used in this analysis. This method is used to generate a background estimate for the $t\bar{t} + jets$ background which does not rely on potentially mismodeled simulated events.

6.1 ABCD Method

The ABCD method is a common background estimation method used within particle physics analyses to determine an event yield prediction in a blinded signal region by leveraging two, independent analysis variables. A diagram setting the stage for the

ABCD method is shown in figure 6.1. Two independent variables are selected as the basis of the signal and validation region definitions. A selection is placed on these variables which partitions the 2D analysis plane into four orthogonal regions. Ideally, the two variables selected would be distinguishing in that events which surpass the thresholds placed in both dimensions are more likely to be signal events than background.

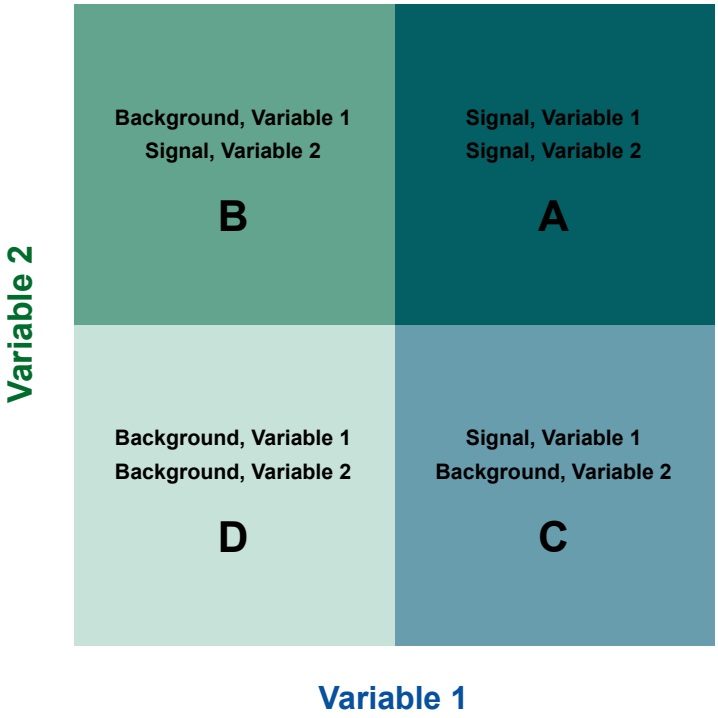

Figure 6.1: The ABCD method makes use of two independent variables to define the signal and validation regions. Selections are placed on these two variables, creating four separate regions. The goal of this method is to predict the number of background events in the signal-like A region in data by leveraging a transfer factor defined by the event counts in the other three regions.

Given that the two basis variables are independent and signal contamination is low, an unbiased transfer factor (known as the closure relation) can be used to derive the number of background events in the A region. The closure relation is defined as:

$$N_{A,Pred.} = \frac{N_{B,Obs.} N_{C,Obs.}}{N_{D,Obs.}} \tag{6.1}$$

where $N_{A,Pred.}$ is the number of predicted background events in the A region and $N_{X,Obs.}$ with $X \in \{B, C, D\}$ is the number of events observed in data in the B, C, and D regions. If signal events are present, the event yield observed in data in the A region would be larger than that predicted via the closure relation.

Validation of the ABCD method entails ensuring that the two basis variables are independent and create a plane which closes well. This ensures that the background estimate from the ABCD method is a good representation of the true number of events in the A region. This can be done by monitoring the non-closure of the analysis plane, defined as:

$$\not{C} = 1 - \frac{N_{B,Obs.}N_{C,Obs.}}{N_{A,Obs.}N_{D,Obs.}} \tag{6.2}$$

where $\not{C}$ is the non-closure of the plane given a particular set of selections on the two variables. Note that this metric is zero when $N_{A,Pred.} = N_{A,Obs.}$ This validation is first carried out using simulated events, where the event yield in the A region can be checked without bias. A correction factor for non-closure in simulation is defined as the inverse of the closure; that is:

$$\alpha = \frac{N_{A,Sim.}N_{D,Sim.}}{N_{B,Sim.}N_{C,Sim.}} \tag{6.3}$$

where the $Sim.$ subscript indicates event yields taken from simulated data.

Therefore, the final background estimate for the number of data events in the A region is given by:

$$N_{A,Corr.} = \alpha N_{A,Pred.} \tag{6.4}$$

where the subscript $Corr.$ corresponds to the corrected background prediction in data.

In practice, mismodeling of the basis variable distributions between data and simulated events results in some amount of non-closure. This is accounted for by deriving a systematic uncertainty which modifies the event yield prediction proportional to the effect of the mismodeling on the prediction.

An idealized scenario for performing the ABCD method is shown in figure 6.2. Signal should peak in the A region indicating that the two basis variables are distinguishing. Addtionally, the background distribution should peak in the D region and fall smoothly in the direction of increasing either basis variable.

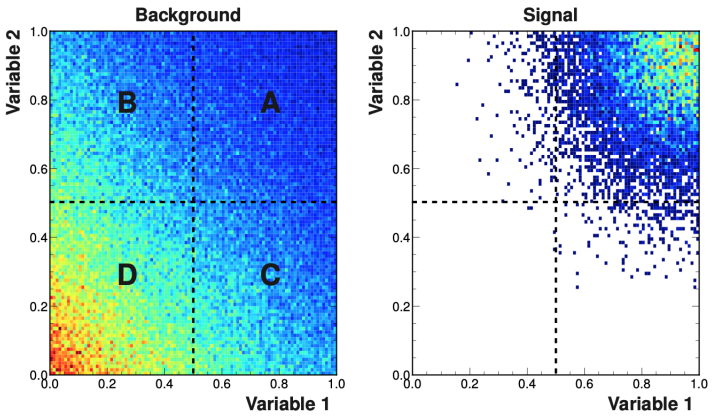

Figure 6.2: The two variables selected to create the ABCD regions should result in a smoothly decaying background distribution with the maximum falling in the D region. Additionally, the distribution for signal should reside almost completely in the A region with few signal events spilling into the B, C, and D regions.

6.2 ABCDisCoTEC Method

The majority of neural networks intended for classification train a single discriminant to label events as either signal or background. However, to make use of the ABCD method for background estimation, the network used in this analysis generates two independent discriminants. The ABCDisCoTEC network uses both a custom architecture and loss function to accomplish this task.

This method is inspired by and an extension of previous work on the ABCDisCo method [80]. This technique involved creating two discriminants with independence enforced using a distance correlation loss. This work extends this approach by including a closure-specific loss function, further improving the methods performance.

The ABCDisCoTEC neural network structure is given in section 6.2.1. Section 6.3 outlines the four loss components which control the learning. Finally, section 6.4.5 lists the various hyperparameters which are tuned to find an optimal training.

6.2.1 Neural Network Structure

The ABCDisCoTEC model is constructed and trained using TensorFlow 2.2.0 with a Keras 2.4.3 backend [81, 82]. The network model can be constructed exclusively using Keras built-in functions and classes. The full model structure used in our analysis is shown in figure 6.3.

The ABCDisCoTEC model is a fully-connected dense neural network (DNN) containing four separate "modules" which each serve to accomplish a distinct learning task. This model is nearly sequential with the one caveat being the separation of the two classification modules for creating the two independent classifiers.

The model is constructed using one input layer, four hidden layers, and two output layers. Each hidden layer contains 200 nodes using rectified linear unit (ReLU) activation functions. Batch normalization and dropout layers are included throughout the model in order to prevent gradient explosion as well as overtraining. Additionally, there is a single concatenation layer which merges inputs for the classification modules as well as three duplication layers which triplicate the classification outputs for ease of loss calculation.

The first module of the network is the "Mass Regression" (MR) module. In the MR module, a regression is performed on the inputs of the network in order to estimate the mass of the top squarks (or top quark, for background events) in each event. The predicted top squark mass is then concatenated with the original inputs and passed to the classification modules. The truth labels for this regression are the invariant mass of the generator level top squarks (in the case of signal) or top quarks (background) in the event. This module contains two hidden layers.

The intent of the MR module is twofold. First, it provides the two discriminants with an additional, high-level variable for classification. Additionally, this information is valuable in that it provides a solution for classifying signal events which vary in top squark truth mass. This is important as the top squark mass will dictate the kinematic behavior of its decay products, and therefore the overall kinematics of the event. Thus, the MR module improves performance on under represented signal mass hypotheses.

The next module in the model is the "Pre-Classification" module containing one hidden layer. The intent of this module is to expand the inputs passed to the classification layers. Inclusion of this layer improves both signal vs. background identification

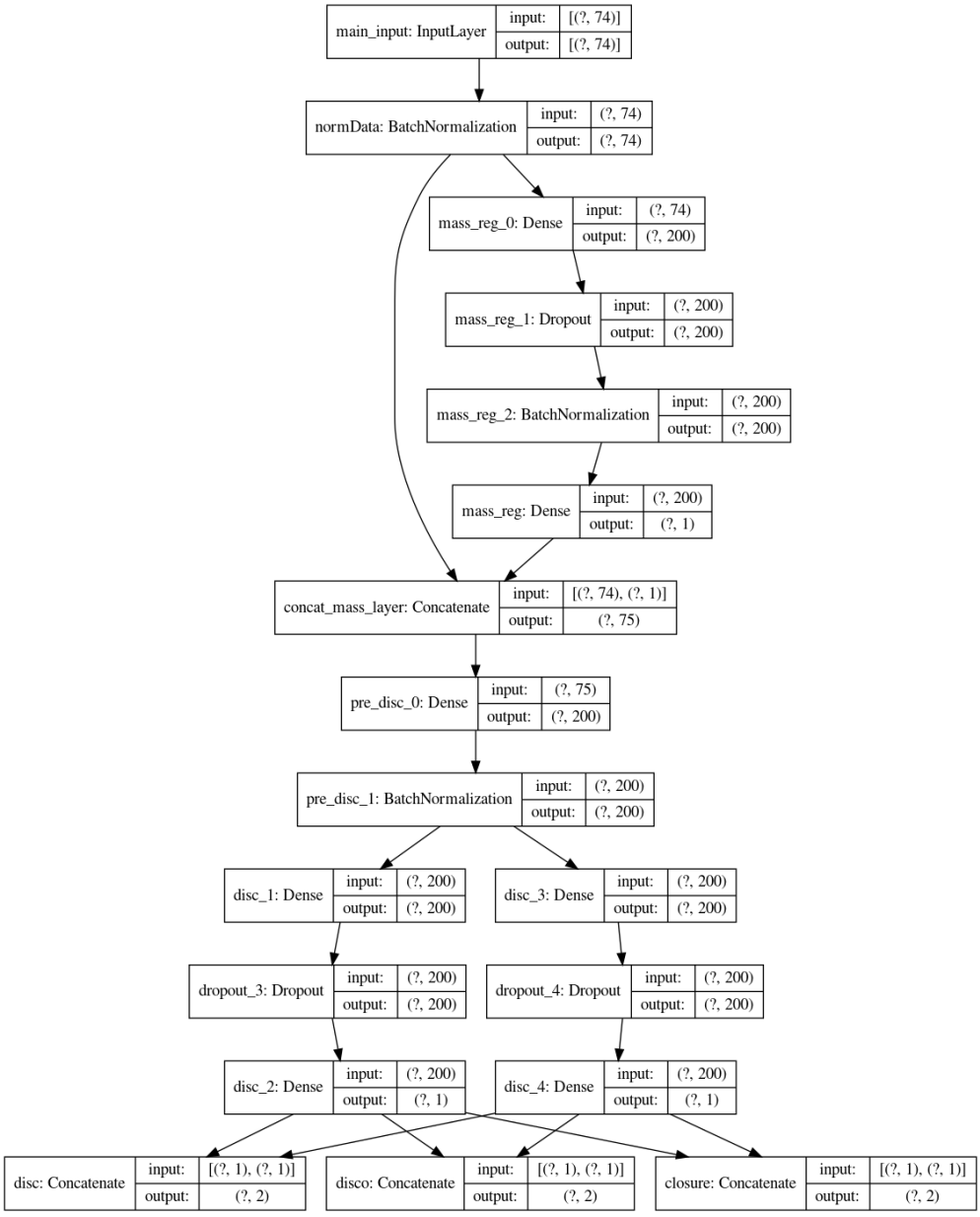

Figure 6.3: The central feature of the ABCDisCoTEC model is the two separate binary classification modules (starting at layers disc_1 and disc_3). Each of these modules is responsible for classifying events as either signal or background. The two discriminants from these models will serve as the basis variables for the ABCD method.

accuracy as well as decorrelation of the two output discriminants. The resulting output of this module is 200 nodes, where each node represents some combination of the physical inputs to the network.

The distinguishing feature of the network is the two, separate discriminant modules. Each path of the network will output a separate discriminant variable which will be trained to identify signal events from background events. Both paths contain a single fully-connected hidden layer. Both paths end with a fully-connected output layer with a single sigmoid activation node. The network is trained such that both discriminants should output one for signal events and zero for background events. The two discriminants are concatenated to give a size two output array containing the two discriminant values.

6.3 Loss Functions

The loss function dictating network learning contains four components: signal vs. background discrimination, closure enforcement, decorrelation, and mass regression. A linear combination of these four components defines the learning objective.

The total loss function for the model is:

$$L_{Total} = \lambda_{BCE} L_{BCE} + \lambda_{Closure} L_{Closure} + \lambda_{DisCo} L_{DisCo} + \lambda_{MR} L_{MR} \tag{6.5}$$

The λ_i scaling parameters dictate the relative strength of each of the loss components. Network optimization involves determining the ideal values of these parameters such that the resulting training shows low non-closure and good classification ability for both discriminants.

Each of the individual components of the loss function is responsible for implementing a separate constraint on the classification. Though each of these terms is given a separate functional form, it is important to note that they are not explicitly independent. As an example, the minimum for the signal vs. background discrimination loss term would be to explicitly label all background events with a zero (denoting a "background-like" event) and all signal events with a one for both discriminants. However, the closure enforcement and decorrelation portions of the loss function require at

least some background events to be given a 'signal-like" label by one of the discriminators in order to approach a closure loss minimum. There are a number of effects of this nature that must be considered when training this network. Many of these effects can be mitigated by carefully choosing the λ values for each of the loss terms.

In the following sections, each individual loss term will be described in detail. In addition, the relative strength of these parameters (i.e. λ values when compared to other losses) is described to give a sense of scale.

6.3.1 Signal vs. Background Discrimination Loss

The primary goal of the neural network is separating background and signal events into two populations. As shown in figure 6.2, the background distribution ideally would take the form of a two-dimensional exponential decay with its maximum value residing at $(d_1, d_2) = (0, 0)$. A distribution of this nature would have low non-closure as well as good background rejection in the A region. The signal events should reside near $(d_1, d_2) = (1, 1)$ with limited signal contamination in the B, C, and D regions.

As a means for performing this task, a binary cross entropy (BCE) loss term is included in the overall loss function for this training. The functional form for the BCE loss term is:

$$L_{BCE,j} = -\frac{1}{N}\sum_{i=1}^{N} y_{ij}\log\left(p(y_{ij})\right) + (1 - y_{ij})\log\left(1 - p(y_{ij})\right) \tag{6.6}$$

where N is the total number of events and $y_i, p(y_i)$ are the true label and predicted label for event i, respectively. The values for the predicted label are taken from the output of the neural network while the true labels are given to the network for each event in the training sample. The output of the neural network is determined using a sigmoid activation function allowing for event classification values between 0 and 1. The individual discriminator loss functions are bounded between zero and one, where a loss value of zero is synonymous with perfect identification of the correct label for all signal and background events in a given batch. The total loss responsible for signal and background discriminations will be the sum of the loss associated with the two discriminators:

$$\lambda_{BCE} L_{BCE} = \lambda_{BCE}(L_{BCE,d_1} + L_{BCE,d_2}) \tag{6.7}$$

The loss term shown in equation (6.7) is bounded between $(0, 2\lambda_{BCE})$.

The relative strength of this loss function compared to the others is dictated by λ_{BCE}. The value of λ_{BCE} is set such that the overall loss contribution (i.e. $\lambda_{BCE}L_{BCE}$) is around 1-2 orders of magnitude larger than the other loss components.

6.3.2 Distance Correlation Loss

In order to be used for the ABCD method, the two discriminant outputs of the network must be decorrelated. Thus, a dedicated loss term is included in the overall loss function which drives this decorrelation.

The main function driving the decorrelation of the two discriminants is the distance correlation (DisCo) function [83]. While other metrics exist which quantify the linear correlation between two variables (e.g. Pearson correlation), DisCo also accounts for non-linear correlations. In fact, two finite, real-valued variables have a DisCo value of 0 if and only if the two variables are independent [83]. This makes distance correlation an ideal metric for enforcing independence of the two discriminants created by this method.

DisCo is described in equation (6.8):

$$\mathrm{dCor}^2(X,Y) = \frac{\mathrm{dCov}^2(X,Y)}{\sqrt{\mathrm{dVar}^2(X)\mathrm{dVar}^2(Y)}} \tag{6.8}$$

The distance covariance (dCov) is a metric which is calculated on a pair of real valued or vector valued random variables (X, Y). Let (X_k, Y_k) for $k = 1, 2, ..., n$ be a statistical sample from the sets X and Y.

The first step of the calculation is to compute the $n \times n$ distance matrices $A_{j,k}$ and $B_{j,k}$, where n is the total number of samples:

$$\begin{aligned} a_{j,k} &= ||X_j - X_k||,\ j,k = 1,2,...,n \\ b_{j,k} &= ||Y_j - Y_k||,\ j,k = 1,2,...,n \\ A_{j,k} &= a_{j,k} - \bar{a}_{j.} - \bar{a}_{.k} + \bar{a}_{..} \\ B_{j,k} &= b_{j,k} - \bar{b}_{j.} - \bar{b}_{.k} + \bar{b}_{..} \end{aligned} \tag{6.9}$$

Where $\bar{a}_{j.}$ and $\bar{a}_{.k}$ are the jth-row and kth-column mean and $\bar{a}_{..}$ is the grand mean of

the distance matrix. The distance covariance ($dCov$) and distance variance ($dVar$) can then be defined using the matrices above:

$$\begin{aligned} \mathrm{dCov}_n^2(X,Y) &= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n} A_{j,k}B_{j,k} \\ \mathrm{dVar}_n^2(X) = \mathrm{dCov}_n^2(X,X) &= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n} A_{j,k}^2 \end{aligned} \tag{6.10}$$

An empirical understanding of the difference between DisCo and Pearson Correlation can be gained from figure 6.4. DisCo is able to quantify both linear and non-linear correlations while Pearson correlation can only identify linear correlations. This feature is necessary for enforcing the statistical independence of the two discriminants.

Independence of the two classification outputs of the ABCDisCoTEC network is controlled by the DisCo loss component:

$$\lambda_{DisCo}L_{DisCo} = \lambda_{DisCo}\mathrm{dCor}_n^2(d_1,d_2) \tag{6.11}$$

where the inputs to the DisCo component of the loss are the two discriminant values distributions from a batched sample of events. Note that there is no need for signal events to have independent distributions for the two discriminants; so, one could imagine only applying the DisCo loss function to background labeled events. However, applying the DisCo loss to background events alone tends to result in undesirable background distribution shapes. Applying the DisCo loss to both background and signal results in a decoupling of L_{BCE} and L_{DisCo} which results in more reliable trainings.

Distance correlation values are bounded between $(0,1)$ and thus, the final loss values are bounded such that $L_{DisCo} \in (0, \lambda_{DisCo})$. In order to preserve the signal vs. background discriminant performance, the scaling parameter λ_{DisCo} is set such that $\frac{\lambda_{DisCo}L_{DisCo}}{\lambda_{BCE}L_{BCE}} \sim \frac{1}{100}$.

6.3.3 Closure Loss

An additional loss function is included which will inform the network of the closure constraint required by the ABCD method. This loss component (in a similar fashion to L_{DisCo}) uses the two discriminants from the binary classifiers to estimate non-closure.

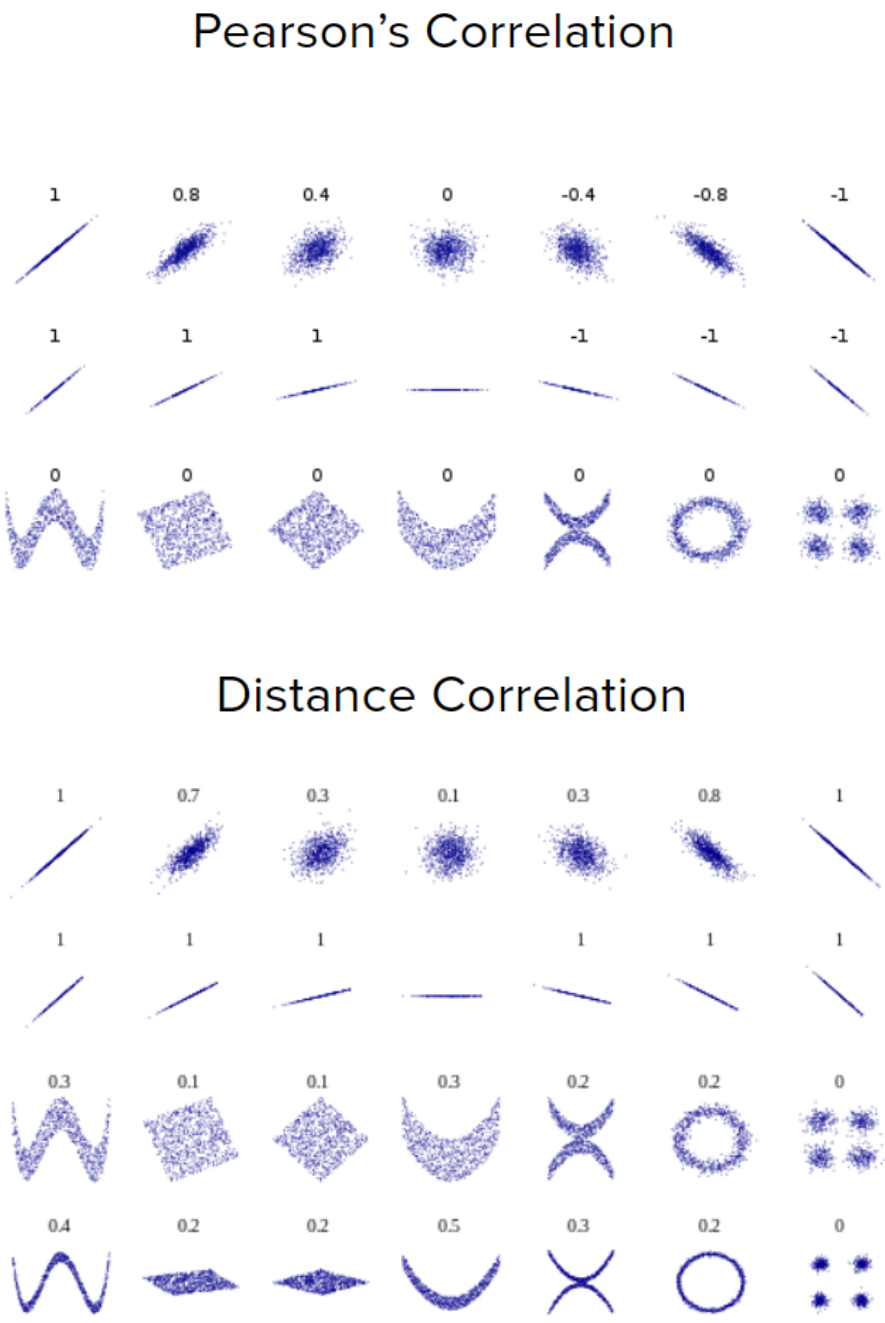

Figure 6.4: A comparison of Pearson correlation (top) and DisCo (bottom) values for a sampling of two dimensional distributions. DisCo returns a non-zero value for the bottom row of distributions where Pearson correlation is zero. This characteristic of DisCo makes it an ideal metric for enforcing independence between the two discriminants produced by the ABCDisCoTEC model. Adapted from [84].

For each batch, two random discriminant values from zero to one are chosen to define the ABCD regions. Event counts in each region are computed via integration and then used to determine the non-closure loss.

Random ABCD bin edges are chosen to ensure good closure across the entire double discriminant plane rather than enforcing closure only on a single set of ABCD region definitions. This way, there is no preferential treatment of any choice of bin edges during training which is important for ensuring a robust background estimation.

In order to preserve the differentiability of the closure loss component, a sigmoid function is used to calculate the number of events in each of the A, B, C, and D regions. This is crucial for the gradient descent algorithm which is responsible for network learning. The number of events in each of the regions as used by the loss is given by:

$$\begin{aligned} N_A &= \sum_{i=1}^{N} \sigma(100(d_{1,i} - B_1))\sigma(100(d_{2,i} - B_2)) \\ N_B &= \sum_{i=1}^{N} \sigma(100(d_{1,i} - B_1))\sigma(100(B_2 - d_{2,i})) \\ N_C &= \sum_{i=1}^{N} \sigma(100(B_1 - d_{1,i}))\sigma(100(d_{2,i} - B_2)) \\ N_D &= \sum_{i=1}^{N} \sigma(100(B_1 - d_{1,i}))\sigma(100(B_2 - d_{2,i})) \end{aligned} \tag{6.12}$$

Where N is the total number of events in the batch, $\sigma(x)$ is the sigmoid function centered at $1/2$, and B_1, B_2 are the random bin edges used to determine the region boundaries.

The prediction for the number of background events in the A region is then determined by:

$$N_{A,Pred} = \frac{N_B N_C}{N_D} \tag{6.13}$$

Given that the final two-dimensional plane predicted by the network should satisfy the closure relationship, the predicted number of events in the A region should match the actual number of events:

$$N_{A,Pred} = N_A \tag{6.14}$$

Through algebraic manipulation, the following equation is equivalent to the requirement imposed by the ABCD method.

$$N_A N_D - N_B N_C = 0 \tag{6.15}$$

Thus, if the closure relation is satisfied by the distribution of events in the two discriminant plane, then equation (6.15) will also be zero. The full closure loss term for the ABCDisCoTEC neural network can then be expressed as:

$$\lambda_{Closure} L_{Closure} = \lambda_{Closure} \left(\frac{N_A N_D - N_B N_C}{N_A N_D + N_B N_C} \right)^2 \tag{6.16}$$

The numerator of equation (6.16) is set to ensure that each batch adheres to the closure relation. The overall scale of the loss function and the possible minima are controlled by the denominator. The closure loss function is bounded between zero and one for ease of scaling. The denominator regularizes the behavior of the loss function with respect to each of the four regions. That is, if any one of the event counts in a particular region becomes too large or too small, the overall loss will increase.

The ideal scaling of this loss term is found to be $\frac{\lambda_{Closure} L_{Closure}}{\lambda_{BCE} L_{BCE}} \sim \frac{1}{10}$. Note that both the closure and distance correlation loss functions accuracy are highly correlated with batch size and learning rate. For a full discussion of these effects, see Section 6.4.

6.3.4 Mass Regression Loss

The final component of the loss function is an analysis specific mass regression term. In both signal decay topologies considered, a heavy top squark will decay to produce many light flavored jets and/or leptons. Using the four vector information of these (as well as additional information from other object- and event-level variables), the network infers the mass of the heaviest resonance in the events. This mass value will then be fed into the binary classification module of the network as an additional input variable.

The mass regression label of the top squark (or in the case of $t\bar{t} + jets$ background events, the top quark) is computed for each of the events in the input dataset using generator level object information. The generator level objects are matched (via ΔR and p_T matching) to their reconstructed counterparts. Then, the four vectors of objects

decaying from one of the top squarks (quarks) are summed with one another to create a pseudo-top squark (quark) four vector. The invariant mass of this reconstructed object is used as the true mass regression label for an event.

The mass regression loss function is the mean squared error loss of the true ($M_{\tilde{t},true}$) and predicted ($M_{\tilde{t},pred.}$) invariant mass averaged over all N events:

$$\lambda_{MR}L_{MR} = \lambda_{MR}\frac{1}{N}\sum_{i=1}^{N}(M_{\tilde{t},true,i} - M_{\tilde{t},pred,i})^2 \tag{6.17}$$

The output of the mass regression module of the network is concatenated with the other inputs before the two discriminant modules. The optimal value of this loss term is found to be $\frac{\lambda_{MR}L_{MR}}{\lambda_{BCE}L_{BCE}} \sim \frac{1}{100}$.

6.4 Training

The following sections outline the training procedure for the ABCDisCoTEC model. This includes both analysis specific considerations as well as general practices for training a multi-objective neural network.

6.4.1 Datasets

For the purpose of this analysis, the ABCDisCoTEC model will be used to generate a data-driven background estimation for the $t\bar{t} + jets$ event yield in the signal region. Though other minor backgrounds exist in this region (notably QCD multijet and rare multi-boson production processes), their contribution to the total event yield in the search region is small enough that they can be effectively estimated using Monte Carlo (MC) simulation. Thus, the background sample used for training and validation contains only $t\bar{t}+jets$ events. All training and validation samples are composed of MC generated events and do not contain actual detector data.

In addition to the nominal $t\bar{t} + jets$ simulated samples, systematic variations are made at generator level in an attempt to address any mismodeling of the physical process. These additional $t\bar{t} + jets$ variation samples will be included in the training sample in order to ensure that the network is robust to any fluctuation of behavior between Monte Carlo simulation and data. Physical effects which are addressed by

these systematic uncertainties include the modeling of initial- (ISR) and final-state (FSR) radiation, matrix element/parton shower matching, underlying event tune, color reconnection, as well as variations associated with jet energy scale (JES) and resolution (JER). These additional training samples result in a network which is more robust to differences between simulation and data.

Both the nominal and systematically varied $t\bar{t} + jets$ events are assigned a truth label of "background" (numerically, zero). All background categories are combined and randomly mixed to create the total background sample.

Signal samples are directly taken from Monte Carlo simulation. Separate samples are created for different top squark mass hypotheses from 300 GeV to 1400 GeV in steps of 50 GeV. All signal events are given a truth label of "signal" (numerically, one), irrespective of the top squark mass. All signal categories are combined to create the signal sample.

Once the full background and signal sample sets have been created, they are split using an 80/10/10 training/test/validation split. In addition, the background and signal samples are split corresponding to the three signal selection regions based on the number of leptons in the final state: zero, one, or two leptons. A separate network is trained for each of the three channels and for each of the two signal models, resulting in six separate networks for the analysis.

6.4.2 Input Variables

Distributions for each input variable are first checked to ensure good modeling agreement between Monte Carlo simulation and data. Any variables which show a significant difference or exhibit any sort of mismodeling trend are rejected from consideration. A table of the input variables for the fully-hadronic, semi-leptonic, and fully-leptonic channels are shown in Table 6.1.

The input variables include both low- and high-level event information. Low-level inputs include the momentum-energy four vector information for both jets and leptons, jet flavor tagging discriminators, and pseudo-object four vectors constructed by summing together the lowest p_{T} ranked jets in an event. All four vectors are translated to the center-of-mass frame before being used as inputs. Jets and leptons are ranked by p_{T}. It is observed that there is considerable disagreement between jet p_{T} in data

Table 6.1: Neural network input variables for the three channels. The ranges denote the p_t ranked objects that are fed into the network.

Input Variables		
Fully-Hadronic (0l)	Semi-Leptonic (1l)	Fully-Leptonic (2l)
Jet η and ϕ (1-7)	Jet η and ϕ (1-6)	Jet η and ϕ (1-5)
Jet p_t/H_T (1-7)	Jet p_t/H_T (1-6)	Jet p_t/H_T (1-5)
Jet DeepJet Discriminators (1-7)	Jet DeepJet Discriminators (1-6)	Jet DeepJet Discriminators (1-5)
Combined 8th+ Jet η and ϕ	Combined 7th+ Jet η and ϕ	Combined 6th+ Jet η and ϕ
Combined 8th+ Jet p_t/H_T	Combined 7th+ Jet p_t/H_T	Combined 6th+ Jet p_t/H_T
Lepton 4-Vector (1-2)	Lepton 4-Vector (1)	N/A
Fox-Wolfram Moments (2-5)	Fox-Wolfram Moments (2-5)	Fox-Wolfram Moments (2-5)
Jet Momentum-Energy Tensor eigenvalues (0-2)	Jet Momentum-Energy Tensor eigenvalues (0-2)	Jet Momentum-Energy Tensor eigenvalues (0-2)
Stop Hemisphere 4-Vector (1-2)	Stop Hemisphere 4-Vector (1-2)	Stop Hemisphere 4-Vector (1-2)

and simulated events. However, the fractional jet p_{T} with respect to the total hadronic transverse momentum (H_{T}) shows better agreement. As this variable encodes much of the same information as jet p_{T}, it is used as an input feature for the network training in place of jet p_{T} and H_{T} separately. Three examples of distributions of jet $\frac{p_{\mathrm{T}}}{H_{\mathrm{T}}}$ are shown in figure 6.5.

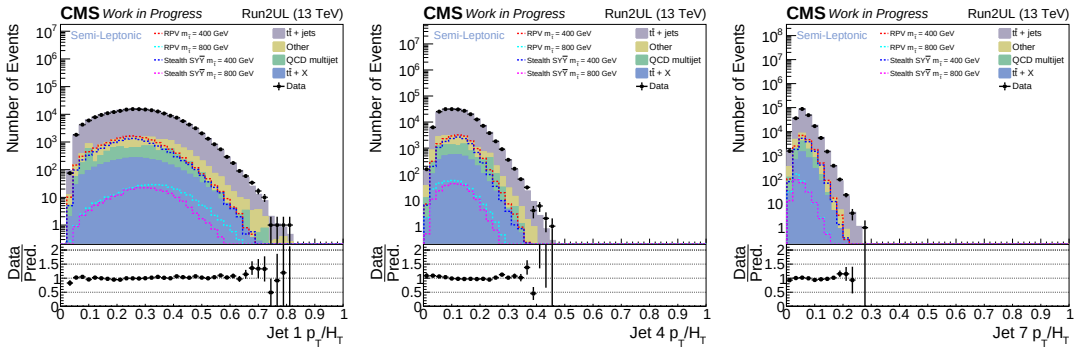

Figure 6.5: Mismodeling of jet p_{T} is reconciled by dividing by H_{T} which results in much better agreement between data and simulation. Jet $\frac{p_{\mathrm{T}}}{H_{\mathrm{T}}}$ distributions are shown for the first, fourth, and seventh highest p_{T} ranked jets for events under the 1ℓ channel selection for all background categories. The same distributions are shown for the two signal models under two different $M_{\tilde{t}}$ hypotheses. The ratio plots in the bottom panels show good agreement between data and simulation.

The high-level variables include Fox-Wolfram moments (FWM), Jet Momentum-Energy Tensor eigenvalues (JMT), and event hemisphere objects which represent the two largest objects in each event (either top squark or top quark). The two former

variables describe the overall kinematic behavior of the event as well as energy flow through the decay chain. The latter variable includes high level information about the pair produced top squarks (quarks). Examples of FWM, JME, and reconstructed stop quark variables are shown in figure 6.6 for both simulated background and data.

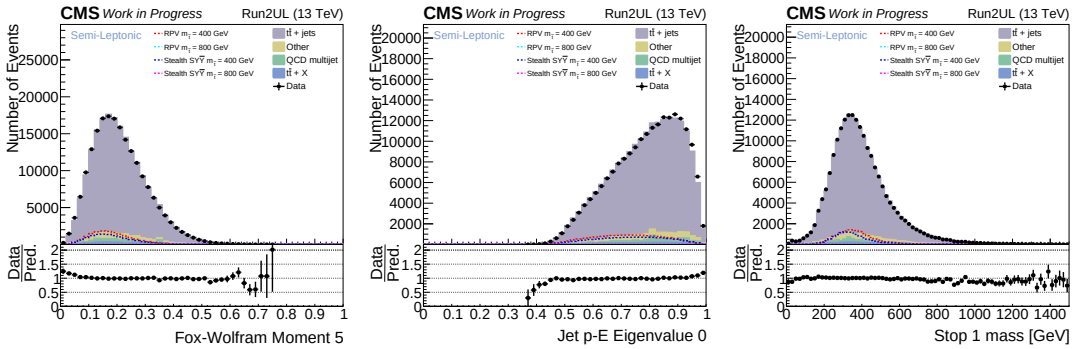

Figure 6.6: Data and simulation distributions are shown for the FWM, JMT, and top squark variables are shown above. Each of these variables provide discrimination power as the representative signal distributions take a slightly different shape than background. Additionally, these variables show good agreement between data and simulation in the bulk of the distribution, making them good candidates for NN inputs.

6.4.3 Batch Construction

The signal and background training samples suffer from imbalance in two categories. First, the number of signal events used in the training ($\sim 1M$ events) is much smaller than the number of background events ($\sim 100M$). Usually, imbalanced datasets result in poor classification performance for the under represented class. Additionally, events with high jet multiplicity are far more rare than their low jet multiplicity counterparts. To account for this two-fold imbalance, batches are constructed using oversampling and undersampling such that all categories are equally represented.

The size of batches is chosen such that the values of the $L_{Closure}$ and L_{DisCo} are representative of the overall training sample. A batch size of 4096 is found to give the best balance between decorrelation and discrimination performance.

Batches are split such that each has an equal population of signal and background events. This entails randomly choosing 2048 events from each of the two categories per

batch. Thus, background events are undersampled by a factor of 0.01. Doing so results in better performance for high $M_{\tilde{t}}$ signal models that have the fewest training events.

Both the background and signal categories are further subdivided into five equal samples split by N_{Jets}. This results in a network with performance that is independent of the number of jets in an event. Therefore, batches are split into ten subsamples of around 400 events based on signal/background and N_{Jets} categories. The overall performance of the network with respect to N_{Jets} is shown in section 6.5.

Additionally, the initial bias of the sigmoid output layers for the two discriminant modules is set to:

$$b_{Output} = \log \frac{N_{Sig}}{N_{BG}} \tag{6.18}$$

This biasing factor scales the initial output of the sigmoid to $\sim N_{Sig}/N_{BG}$. This procedure provides the network with a good starting position for learning how to enforce the closure relationship.

6.4.4 Optimizer

The Adam optimizer [85] is chosen for minimization of the loss function during training. All options are set to the Keras defaults except for the starting learning rate that has been set to 0.0001, which has been determined that a lower learning rate greatly improves the non-closure performance of the network.

6.4.5 Training Hyperparameters

A table of model and training related hyperparameters is shown in Tab. 6.2. These hyperparameters differ from the λ parameters as they modify either the model structure or the training loop for the model. Optimal values for these hyperparameters are also shown in the table.

6.5 ABCDisCoTEC Results

The following sections show the results from training the ABCDisCoTEC neural network to classify signal and background events. Results from only a single model and channel

Table 6.2: Hyperparameters associated with network training. A grid search of each parameter is carried out to determine the best training.

Hyper parameter	Description	Optimal Value
Input Nodes	Number of nodes included in layer before splitting to discriminant layers	~ 200
Disc. Nodes	Number of nodes in each discriminant hidden layer	~ 200
Mass Regression Nodes	Number of nodes in each mass regression hidden layer	~ 200
Input Layers	Number of hidden layers before splitting to discriminant layers	1
Disc. Layers	Number of hidden layers for each discriminant portion	1
Mass Regression Layers	Number of hidden layers for mass regression	1
Dropout	Fraction of weights to drop at each dropout layer	0.3
Batch Size	Number of events per batch	4096
Epochs	Number of training epochs	~ 100
Learning Rate	Initial value of Adam learning rate for minimization	0.0001

(Steath $SY\bar{Y}$ 1ℓ) are shown. More results for both models and all lepton channels are shown in appendix A.

6.5.1 Signal vs. Background Discrimination

The primary goal of the ABCDisCoTEC neural network is classifying signal-like events with discriminant values near one. An example of how the model classifies background and signal events is shown in figure 6.7. Figure 6.8 shows a comparison of the $M_{\tilde{t}} = 550$ GeV signal model and $t\bar{t} + jets$ background separated by N_{Jets}. Note that the events shown in these distributions are part of the validation set which are not included in the training.

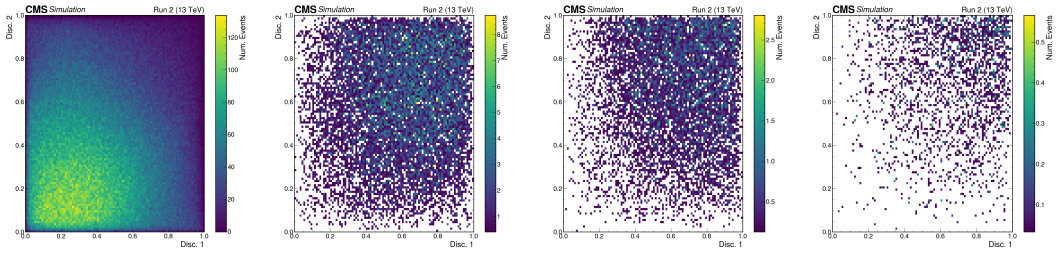

Figure 6.7: A comparison of the two-dimensional discriminant distributions for background (left) and three $SY\bar{Y}$ signal mass hypotheses: 350, 550, and 850 GeV (center left to right, respectively). Note that it is much easier to differentiate high $M_{\tilde{t}}$ signal events from background due to differences in kinematic behavior of objects.

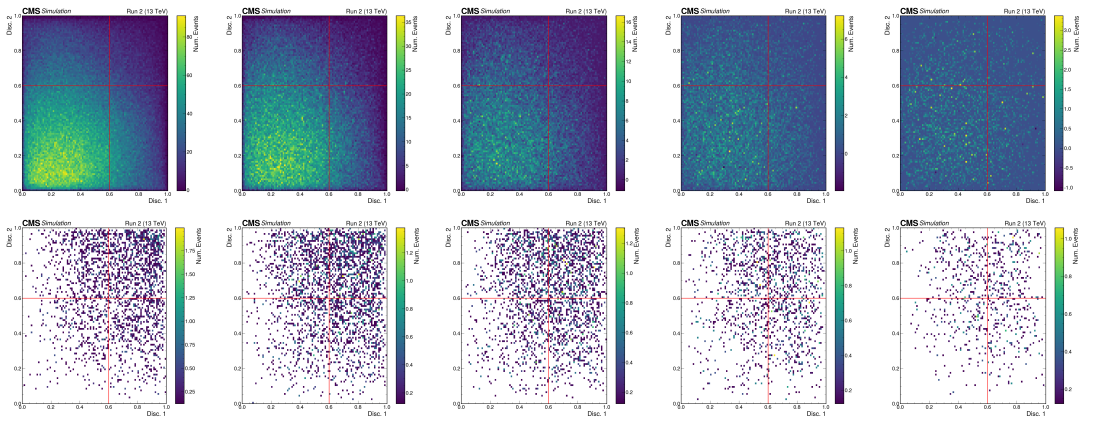

Figure 6.8: The two-dimensional discriminant distributions are shown for $t\bar{t} + jets$ background (top) and the $SY\bar{Y}$ signal model (bottom) separated by N_{Jets}. The plots increase in jet multiplicity from 7 to 11 jets (left to right). The red dashed lines in the figures show possible cuts which could be placed on the two discriminants to create the four ABCD regions. However, these are not the final bin edges used in the analysis.

The model shows good separation between signal and background with the background distribution in the lower left corner of the discriminant distributions while the majority of signal events reside in the upper right. Additionally, figure 6.8 shows similar behavior across all N_{Jets} categories, demonstrating that the custom batching procedure explained in section 6.4.3 has mitigated the potential bias due to a limited number of high N_{Jets} events.

The classification power of the network determines how well the network can distinguish signal from background events. One method for quantifying the classification power of the network is to compute the receiver operating characteristic (ROC) curve. This shows the performance of a classification model by comparing the true positive rate (TPR) and false positive rate (FPR) for different classification thresholds. These quantities are defined as:

$$TPR = \frac{TP}{TP + FN} \tag{6.19}$$

$$FPR = \frac{FP}{FP + TN} \tag{6.20}$$

where TP, TN, FP, and FN are defined as the number of events classified as true

signal events, true background events, false signal events, and false background events, respectively. A metric for interpreting the classification ability of the network is the integral over FPR between zero and one, also known as the area under the ROC curve (AUC). A perfect classifier has an AUC value of unity as it has a TRP value of one for any value of FPR. In contrast, a classifier which randomly assigns labels has an AUC score of 0.5.

As the ABCDisCoTEC NN generates two discriminants for each events, the ROC curve must be computed independently for each classifier. The ROC curves in figure 6.9 show the classification power of both discriminators. The performance of the two discriminants for individual $M_{\tilde{t}}$ hypotheses is shown in figure 6.10.

For both figures, a comparison of the ROC curve for the training set and validation set is made. The similarity of the two curves displays how well the two discriminants generalize to events which were not used during training. This comparison informs whether the network under performs on the validation set, which is a strong signature of network overtraining. In this case, no overtraining is seen for either discriminant or for any of the signal mass hypotheses as the training and validation curves coincide well with one another.

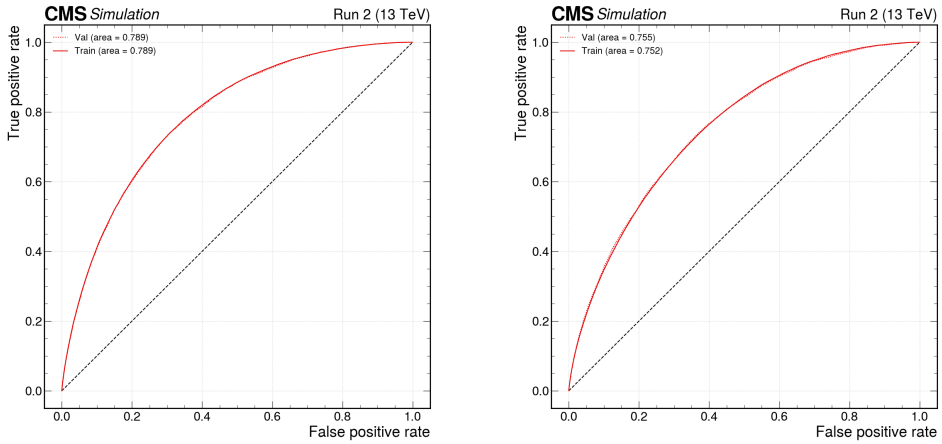

Figure 6.9: ROC curves shown for both discriminant 1 and discriminant 2 exemplify the classification power of the $SY\bar{Y}$ 1ℓ neural network. Note that these distributions only consider the discrimination power of each classifier independently.

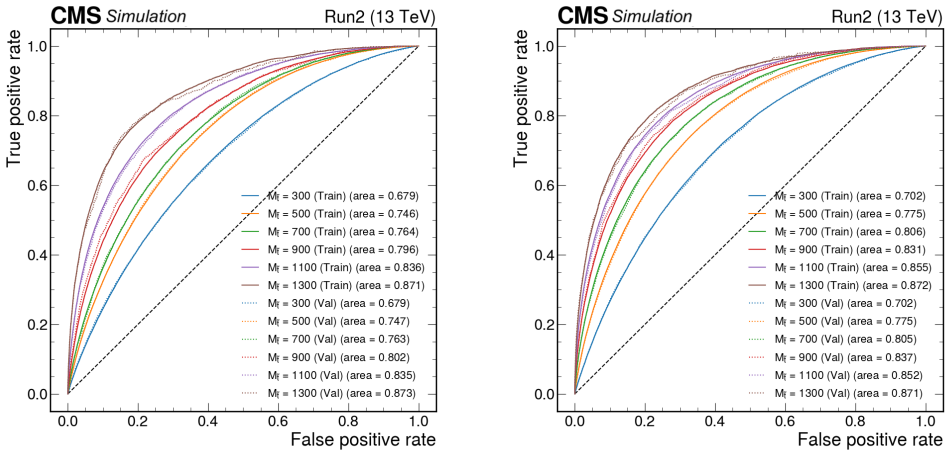

Figure 6.10: The classification performance of the network is highly dependent on the mass of the top squark. As $M_{\tilde{t}}$ increases, the input distributions for signal and background become more separated which leads to easier classification of signal-like events.

6.5.2 Closure Performance

The validity of the ABCD background estimation relies on the fact that the two basis variables are independent, which implies that they adhere to closure relationship shown in equation (6.1). Non-closure—as defined in equation (6.2)—is the metric used for determining how well the prediction for the number of background events in the A region matches with the actual number of events in simulation. If the total number of events predicted by the closure relation in the A region ($N_{A,Pred}$) matches the observed number of events, this metric is zero. Therefore, the final background estimation would ideally have a low value of $\not{C}$ to ensure that the number of $t\bar{t} + jets$ events is accurately predicted.

The distributions in figure 6.11 display the non-closure for each N_{Jets} bin separately. Each bin in these distributions shows the value of $\not{C}$ when choosing a given set of discriminator values as the boundaries for defining the ABCD regions. Discriminator values are scanned in steps of 0.05 in both dimensions.

Figure 6.12 shows an example of the closure performance for a single choice of bin edges where the actual and predicted N_{Jets} distributions in the A region are plotted.

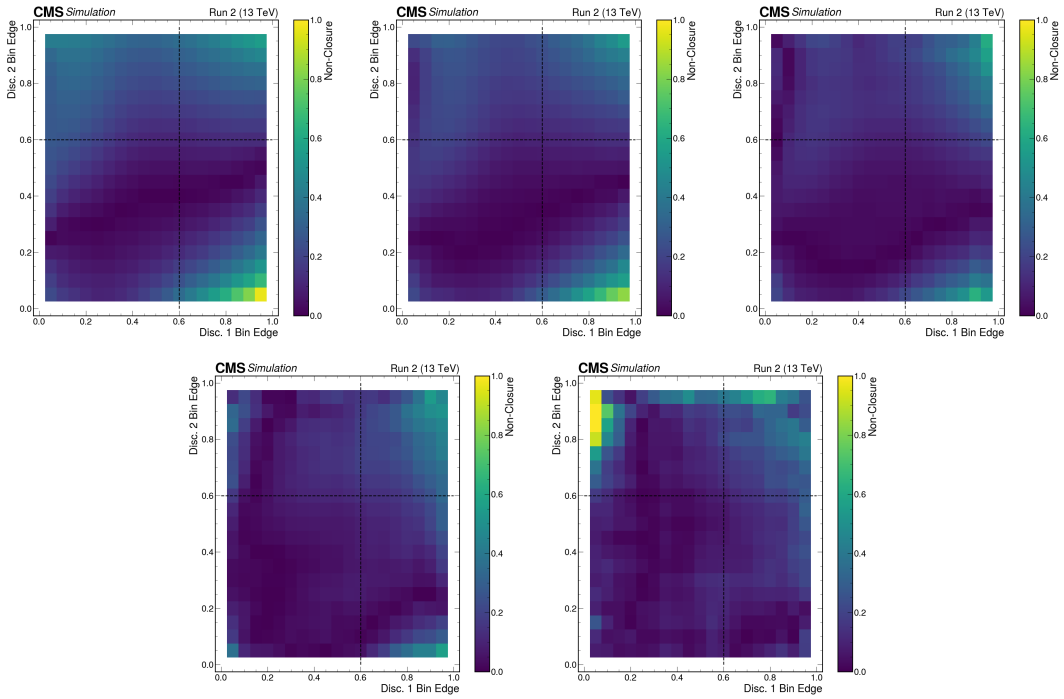

Figure 6.11: The non-closure ($\not\subset$) is shown for a grid of ABCD boundary values ranging from zero to one in steps of 0.05 for both discriminants. Plots are shown from left to right in order of increasing N_{Jets} from 7 to 11+. For all N_{Jets} and for the majority of the boundary value selections, the prediction from the closure relation exhibits $\not\subset$ values below 20%.

These distributions show how well the prediction matches for bin edges randomly selected at $(d_1, d_2) = (0.6, 0.6)$. The ratio plot shown in the bottom panel represents the non-closure for this choice of bin edges for each of the N_{Jets} bins.

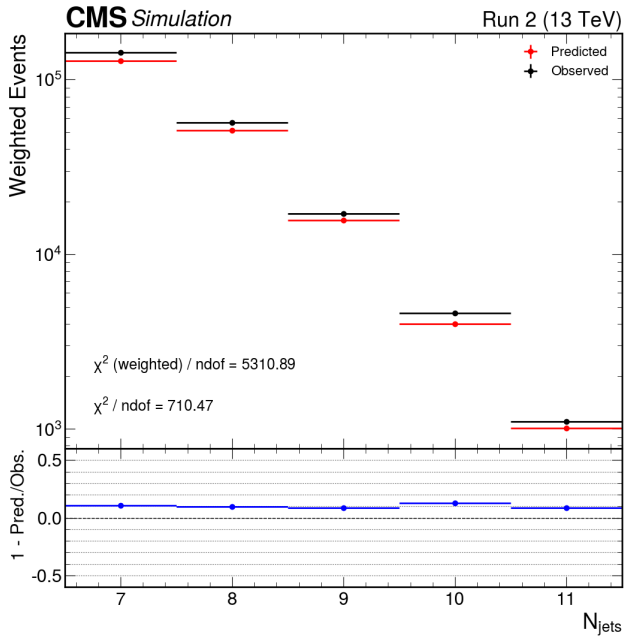

Figure 6.12: A comparison of the actual N_{Jets} distribution to the one predicted by the ABCD closure relation for bin edges of $(d_1, d_2) = (0.6, 0.6)$ The non-closure (shown in the bottom panel) is $\leq 12\%$ for all N_{Jets} bins.

Chapter 7

Analysis

As described in chapter 4, the main challenge in this analysis is predicting the $t\bar{t} + jets$ background which is poorly modeled at high N_{Jets}. In order to circumvent these issues, a data-driven background estimation technique is used to determine the $t\bar{t} + jets$ event yield in the signal region. The neural network described in Chapter 6 is used to estimate this background. Section 7.1 discusses the methods for computing the final background estimation and the estimation of systematic uncertainties using the two outputs of this network.

Additionally, the QCD multijet background modeling is hindered by non-perturbative calculations, meaning that it is difficult to precisely predict the production cross-section. Therefore, this background is predicted using a control region orthogonal to the signal region. A transfer factor is defined which is applied in order to determine the number of QCD events in the signal region in data. The event yield for the other two background categories (TTX and Other) as well as the predictions for signal are taken directly from simulation. The procedure for estimating these background is presented in section 7.2.

These predictions are then used as inputs for a statistical model to determine if signal is present as well as set upper limits on the top squark pair production cross section for the models. Systematics uncertainties based on physics mismodeling and detector effects are estimated and applied as fit inputs. These systematic uncertainties are further discussed in section 7.3.

A description of the formalism for extracting results from fits to data is shown in section 7.4.

7.1 $t\bar{t} + jets$ Background Prediction

The background estimation for $t\bar{t} + jets$ via the ABCD method includes a number of steps to correct for discrepancies between the ABCD prediction from simulation and in data. This section will outline procedure for defining the ABCD regions as well as producing the corrections and systematic uncertainties used for this estimation. In the following sections, each of these steps is explained in more detail.

First, selections are placed on the two discriminants to define the signal-like region (A) and validation regions (B, C, D). For these choices of bin edges, a closure correction is computed using equation (6.3) to ensure that the predicted and observed number of events in A are identical in simulation.

After this correction is applied to the A region $t\bar{t} + jets$ prediction, systematic uncertainties are calculated based on closure behavior in validation regions in data. Three validation regions are defined orthogonally to the A region which are used to derive a systematic uncertainty to correct for differences in closure performance between data and simulation. These regions will also be used for estimating systematic uncertainties based on physics and detector modeling.

With all corrections and systematic uncertainties defined, an optimization is run to determine the (d_1, d_2) selections which lead to optimal signal sensitivity. The final choice of bin edges is decided based on signal sensitivity and signal production cross section upper limit as calculated from the fit on three separate signal mass hypotheses using simulation only.

The final product is a set of (d_1, d_2) selections defining the ABCD regions as well as a closure correction and systematic uncertainties correcting for mismodeling of the ABCD closure. This procedure is carried out for each of the three analysis channels. Additionally, difference choices of bin edges will be made to optimize for low- or high-mass top squarks.

7.1.1 Neural Network Performance

For each of the three channels (called 0ℓ, 1ℓ, and 2ℓ) and for both signal models (RPV and Stealth $SY\bar{Y}$), a separate ABCDisCoTEC neural network is trained. The following section will show the performance for only one of the six neural networks (RPV 1ℓ) unless

otherwise specified. The results for each of the six networks are shown in appendix A with references to the equivalent plots in each of the figure captions in this section.

Neural Network Training Samples

Background events used for training are those that pass the 0ℓ, 1ℓ, and 2ℓ baseline selections as defined in chapter 5. Signal events are simulated RPV or Stealth $SY\bar{Y}$ events with varying $M_{\tilde{t}}$ between 300 GeV and 1250 GeV that pass the baseline selections. In addition to nominal $t\bar{t} + jets$ events, several variations are included in the training set to improve network robustness to mismodeling in simulation. Events with variations to the ME-PS matching scale are used as well as events generated with tune variations, and events where color reconnection is allowed in the parton shower. For all of these events and for all signal events, jet energy resolution and jet energy scale variations of the events are also added to the training set.

The batches of events are prepared so that the network sees equal populations of signal and background events as well as N_{Jets} categories while training. A batch size of 2048 is used where half of the events are background and half are signal. For each half, five N_{Jets} categories are equally represented, i.e. ~ 200 events each for all N_{Jets} categories (0ℓ: 8-12+ jets, 1ℓ: 7-11+ jets, 2ℓ: 6-10+ jets).

Neural Network Performance

Using the aforementioned configuration and setup, the six neural networks are trained as specified in section 6.4 and the resulting performance plots are obtained. Shown in figure 7.1 is the output of the mass regression for $t\bar{t} + jets$ as well as four signal mass points for each of the six networks. In all networks, the regression is able to discern the difference between $t\bar{t} + jets$ and each signal and predicts a mass close to the truth value. In general, the predicted mass is lower than the truth mass, regardless of signal mass point. However, this artifact is not troubling as the predicted mass is used as an additionally input to the network. Thus, it only matters that the network is able to effectively distinguish between different $M_{\tilde{t}}$ events and $t\bar{t} + jets$ events.

ROC curves are shown for each N_{Jets} categories in figure 7.2. The networks performs equally well for either of the two classifiers and across all N_{Jets} bins. Figure 7.3 show that performance varies slightly with $M_{\tilde{t}}$ with the higher masses out-performing the

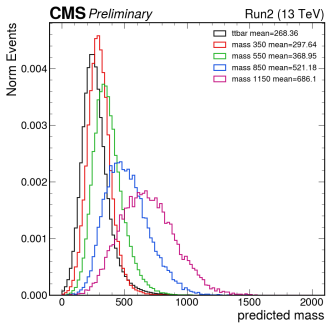

Figure 7.1: Mass regression output for background and signal for both the RPV 1ℓ trained neural networks. Results for all channels and models are shown in figure A.1.

lower ones. This is expected as low mass signal events are kinematically more similar to $t\bar{t} + jets$ events than high mass signal events.

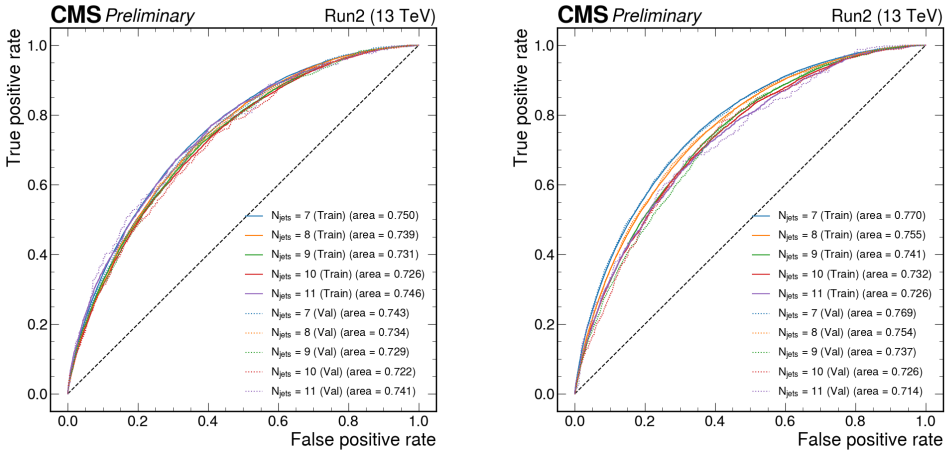

Figure 7.2: For the RPV signal model, ROCs separated by N_{Jets}. Disc. 1 (left) and Disc. 2 (right) are shown for the RPV 1ℓ channel. Results for all channels and models are shown in figure A.3.

A qualitative understanding of the 2D discriminant plane is revealed in figure 7.4 for both $t\bar{t} + jets$ and RPV events with $M_{\tilde{t}} = 550$ GeV. The background has a desirable shape, peaking near (0.0, 0.0) with both discriminants decaying smoothly towards (1.0,

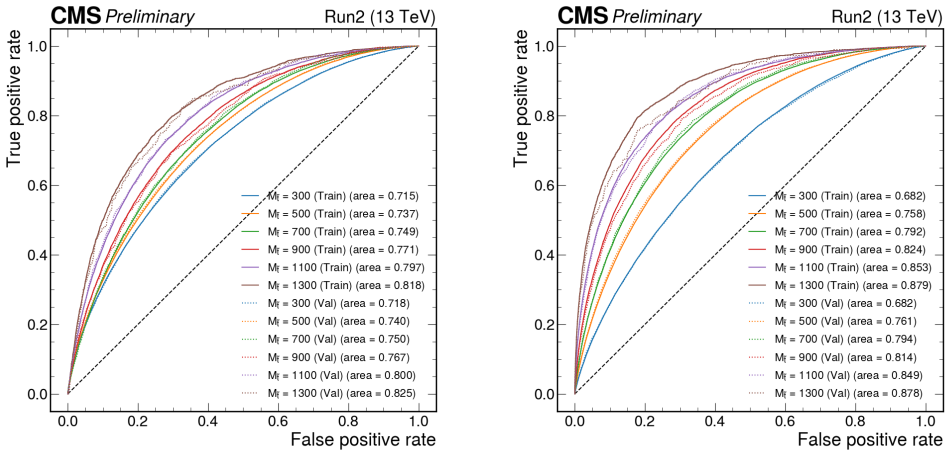

Figure 7.3: For the RPV signal model, ROCs separated by signal mass hypothesis. Disc. 1 (left) and Disc. 2 (right) are shown for the RPV 1ℓ channel. Results for all channels and models are shown in figure A.5.

1.0). This behavior is conducive to performing the ABCD method. In addition, the signal peaks near the (1.0, 1.0) corner, showing separation from background. Both of these behaviors for background and signal also hold across all N_{Jets} bins.

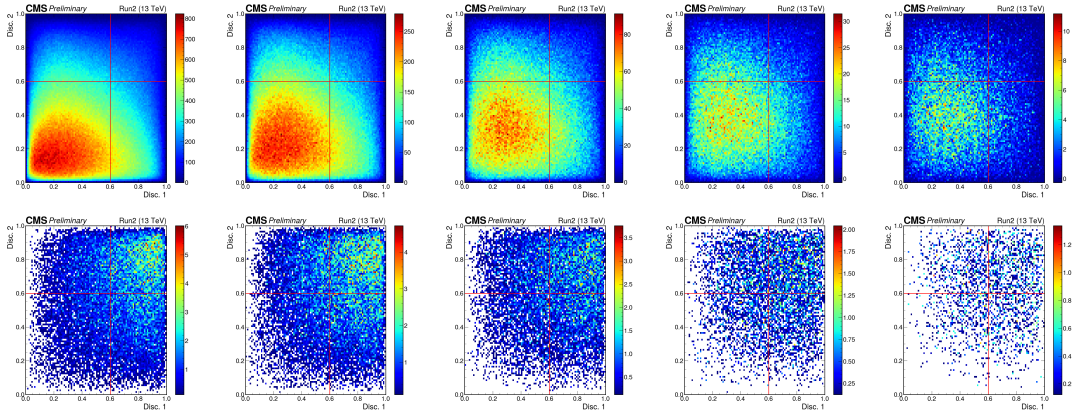

Figure 7.4: For the 1ℓ channel, the 2D discriminant plane for $t\bar{t} + jets$ (top), RPV $M_{\tilde{t}} = 550$ (bottom) separated in N_{Jets} from $N_{\text{Jets}} = 7$ (left) to $N_{\text{Jets}} = 11+$ (right). Results for all channels and models are shown in figure A.10.

7.1.2 Optimizing the ABCD Bins

Once the models are trained, the boundaries defining the ABCD regions must be chosen to finalize the background estimation. The optimization procedure determines the (d_1, d_2) selection criteria which allow for the highest signal sensitivity. Therefore, it is important to not only consider the statistical significance obtained for a given bin edge choice, but also the effects of both systematic and statistical uncertainties.

Rather than choosing an ad hoc optimization metric which may not take into account all the nuances of the analysis, the full fit is run on simulated data (pseudo-data) for a large number of ABCD bin edge choices. The optimal ABCD boundaries can then be chosen based on the metrics of signal sensitivity and upper limit on the top squark production cross section. Signal sensitivity is determined by injecting signal events into pseudo-data and computing the statistical significance of the observed excess of events. Upper limits on the cross section are determined using pseudo-data containing only background events (i.e. no signal injection). The main advantage of this approach is that all systematic and statistical uncertainties will be considered appropriately. Additionally, the presence and effect of signal in the B, C, or D regions is automatically accounted for using this procedure. This is crucial to model accurately as signal contamination in these regions will lead to a larger $t\bar{t} + jets$ background estimate in the A region.

In the context of the ABCDisCoTEC neural network, the ABCD bin edges can take on any value between 0 and 1. Some bin edge choices can be immediately excluded – namely, those too close to the extrema of the discriminant ranges — as they will result in low event yields for both background and signal in some of the ABCD regions. Additionally, bin edge choices which reside closer to the bulk of background events than signal events ($d_1, d_2 \lesssim 0.4$) are likely to result in large background event yields in the A region, leading to poor sensitivity. Considering these points, candidate ABCD bin edges are scanned for in the range of $d_1, d_2 \in (0.4, 0.9)$.

The procedure for choosing the optimal set of bin edges is detailed below:

1. Step through the range of $d_1, d_2 \in (0.4, 0.9)$ in steps of 0.02

2. Integrate the event yields of each background and signal category in simulation and populate the per-process background estimation for each set of bin edges

3. Calculate and apply the appropriate systematic and statistical uncertainties for the full fit

4. Run both an asymptotic limit and significance fits using Higgs Combine

More details on systematic uncertainties and the fitting procedure are outlined in section 7.3 and section 7.4, respectively.

This procedure is repeated for each of the search channels and for each of the signal models independently. The procedure is also separately performed for the $M_{\tilde{t}} =$ 400, 600, and 800 GeV hypotheses. These mass hypotheses span the region of SUSY phase space where the analysis is expected to have the greatest sensitivity. A requirement of at least 3 events per analysis bin is applied when considering each bin edge choice in order to exclude results with statistically limited prediction ability. The results of this procedure are shown in figures 7.5 and 7.6, corresponding to the expected cross section upper limits and significances for the RPV model.

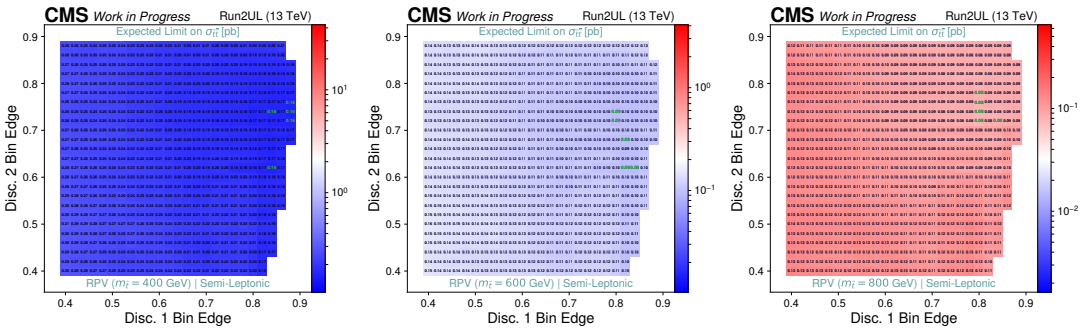

Figure 7.5: For each ABCD bin edge choice, the expected limit using pseudo-data. Limit values are shown for RPV 400, 600, and 800 (top to bottom) and for the RPV 1ℓ channel. Blue values are expected limits where the signal model/mass hypothesis could be excluded, while red values are limits where the signal model/mass hypothesis could not be excluded. The top five choices are highlighted in green. Results for all channels and models are shown in figure A.12.

When making the final choice of ABCD bin edges for each model and channel, both the signal sensitivity (significances) and ability to exclude the highest signal mass hypotheses (expected limits) are considered. Therefore, two different choices of ABCD bin edges are used: one for achieving optimal significance values at low masses (based

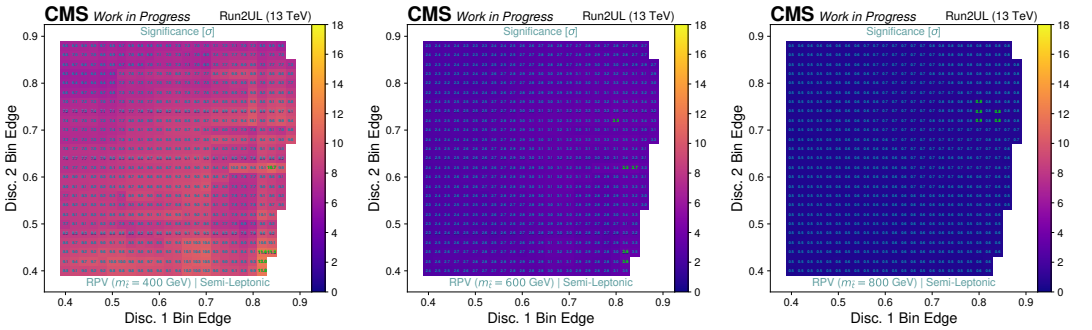

Figure 7.6: For each ABCD bin edge choice, the expected significance using pseudo-data. Significance values are shown for RPV 400, 600, and 800 (top to bottom) and for the RPV 1ℓ channel. The top five choices are highlighted in green. Results for all channels and models are shown in figure A.13.

on $M_{\tilde{t}} = 400$ GeV) and one for obtaining optimal expected limits at high masses (based on $M_{\tilde{t}} = 800$ GeV). These two choices for the ABCD bin edges are subsequently referred to as the low-mass and high-mass optimizations, respectively.

The "crossover" point between the low-mass and high-mass optimization is determined by performing the fit for all values of $M_{\tilde{t}}$ with each optimization's bin edges, and observing at which $M_{\tilde{t}}$ the measured significances are approximately the same for either optimization. For the NN trained on RPV, the low-mass optimization bin edge choice is used up to and including the $M_{\tilde{t}} = 600$ GeV mass point, while for the NN trained on Stealth $SY\bar{Y}$, the low-mass optimization bin edge choice is used up to and including the $M_{\tilde{t}} = 650$ GeV mass point. The plots used to determine the two optimal choices of bin edges are shown in figures 7.7 and 7.8 which compare expected limits and significances for the RPV model. The values of the ABCD bin edges for the two optimizations for each of the channels and signal models are shown in table 7.1.

7.1.3 Validating the ABCD Prediction

To this point, all results from the ABCDisCoTEC networks have been generated using simulated events. However, the final background estimation will use actual collision data. Therefore, it is imperative to quantify the extent to which the simulated samples are an accurate representation of data.

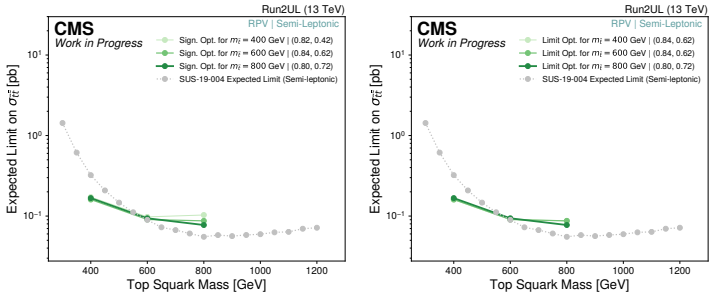

Figure 7.7: **RPV:** Expected limits when choosing the ABCD bin edges based on either best signal significance (upper) or best expected limit (lower). Each shade of color represents the ABCD bin edges chosen based on a particular $M_{\tilde{t}}$. Results for all channels and models are shown in figure A.16.

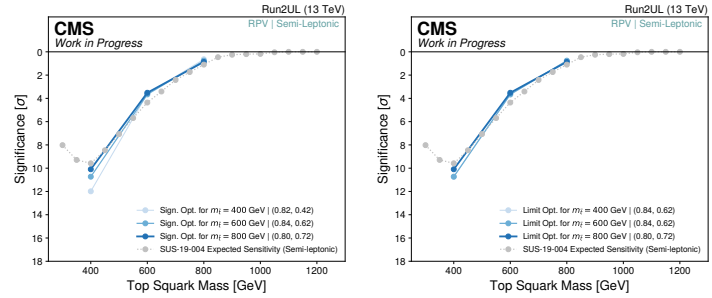

Figure 7.8: **RPV:** Signal significance when choosing the ABCD bin edges based on either best signal significance (upper) or best expected limit (lower). Each shade of color represents the ABCD bin edges chosen based on a particular $M_{\tilde{t}}$. Results for all channels and models are shown in figure A.17.

Table 7.1: Final ABCD bin edge choices obtained from the optimization procedure. The low-mass optimization bin edges are used for masses at or below 600 GeV for the RPV signal model and at or below 650 GeV for the Stealth $SY\bar{Y}$ signal model. The high-mass optimization bin edges are used for masses higher than either of those values.

	Low-Mass Opt.	**High-Mass Opt.**
RPV 0ℓ	(0.52, 0.54)	(0.74, 0.80)
RPV 1ℓ	(0.84, 0.42)	(0.80, 0.72)
RPV 2ℓ	(0.52, 0.58)	(0.50, 0.50)
$SY\bar{Y}$ 0ℓ	(0.76, 0.70)	(0.54, 0.56)
$SY\bar{Y}$ 1ℓ	(0.44, 0.42)	(0.68, 0.82)
$SY\bar{Y}$ 2ℓ	(0.40, 0.42)	(0.48, 0.48)

A series of steps are carried out to determine how well simulation represents data. These both correct for known non-closure in simulation and estimate systematic uncertainties to account for sources of non-closure in data. The procedure for carrying out this validation is three-fold:

1. Compute and apply a simulation-based closure correction to the A region background prediction

2. Determine a data-based non-closure systematic uncertainty to account for differences in network modeling between simulation and data

3. Check if any additional systematic uncertainties are needed stemming from the $t\bar{t} + jets$ variational samples

The final product of this validation is the full background estimate for $t\bar{t} + jets$ in the A region with appropriate systematic uncertainties. The following sections will detail how this procedure is implemented for each of the six channel and model combinations.

Defining the Validation Regions

The first step in validating the background estimation for $t\bar{t} + jets$ is to determine how well the closure of the ABCD relation translates from simulation into data. To do so, non-closure is computed in data using three validation regions. Each of these regions is defined to determine how well the ABCD closure relation is satisfied in different areas of the ABCD plane. Notably, these regions are defined such that they have limited predicted signal event yield. This requirement prevents any corrections or systematic uncertainties from having a large contribution due to the presence of signal.

Figure 7.9 shows the three validation regions. Each validation region can be further subdivided as shown in the right column of the figure, and the event distributions in each of the sub-regions can be used to cross check the non-closure.

Evaluation of the non-closure in each one of these regions is carried out while moving both the central and outer boundaries towards the final ABCD region definition. This process happens in steps of 0.05 in either of the two discriminant directions. For the validation region III, the bin edges are stepped symmetrically in both directions.

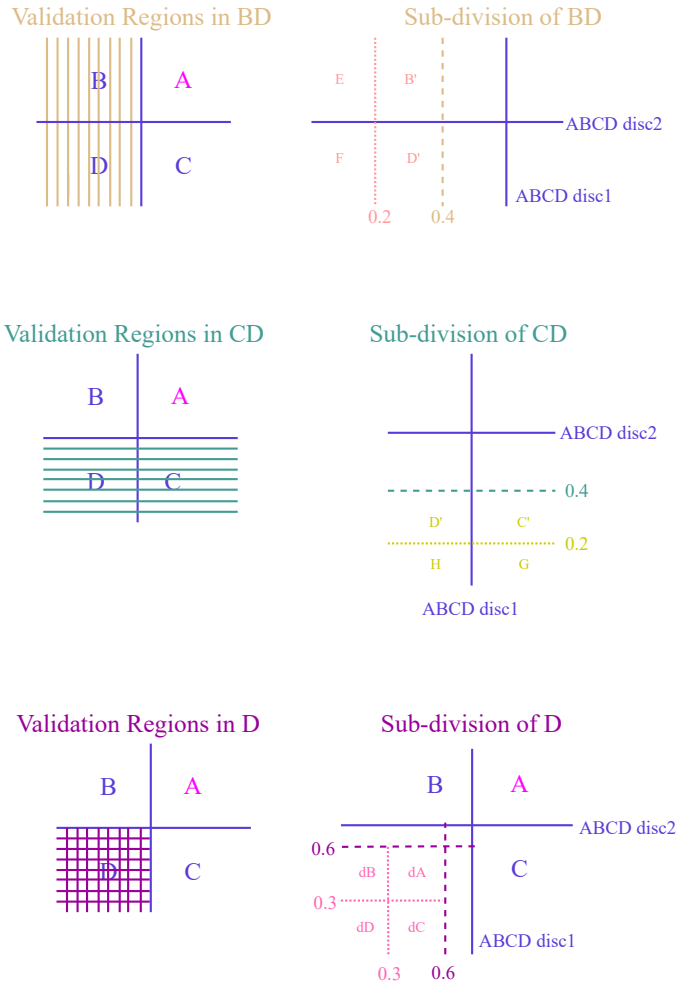

Figure 7.9: Three validation regions are used to determine the performance of the ABCDisCoTEC models in data. Validation region I (top) and validation region II (middle) span the entire domain of a single discriminant. Validation region III (D) (bottom) divides the plane symmetrically along the two discriminants.

$t\bar{t} + jets$ Closure Correction

The two discriminators generated by the ABCDisCoTEC network have a non-negligible amount of non-closure in simulation that could carry over to the background estimate in data. Thus, a closure correction is derived to adjust the background prediction in data based on the amount of non-closure observed in simulation. Assuming that the data and simulation have similar values of non-closure, this procedure should improve the accuracy of the final $t\bar{t} + jets$ background estimate.

A simulation-based closure correction (α) — as quantified in equation (7.1) — is calculated using $t\bar{t} + jets$ simulated events. This correction can be computed and applied for any set of ABCD boundaries including the validation regions defined in the previous section. This correction is applied for the final background prediction using the optimized ABCD region definitions in data.

$$\alpha = \frac{1}{\text{ABCD Closure}} = \left.\frac{N_{\text{obs.},A}}{N_{\text{pred.},A}}\right|_{\text{MC}} \tag{7.1}$$

$t\bar{t} + jets$ Systematics

The background prediction for $t\bar{t}+jets$ is finalized by estimating systematic uncertainties for possible discrepancies between simulation and data. There are two categories of these systematic uncertainties: those arising from differences in neural network inference between data and simulation (called data-based systematic uncertainties) and those which can be associated with mismodeling of simulated events (called simulation-based systematic uncertainties). The former is handled by determining the scale of non-closure in the validation regions for data. The latter is determined by assessing how $t\bar{t} + jets$ samples with different generation parameters affect the ABCD closure. The procedure for estimating these systematic uncertainties is outlined in the following sections.

Data-based Systematics

A data-based systematic uncertainty is derived to account for any non-closure in data after applying the simulation-based closure correction. This value is derived in the lowest N_{Jets} bin for each channel due to the low signal contamination in this region. This systematic uncertainty is then applied to all N_{Jets} bins to correct for residual

non-closure in data.

To estimate the data-based systematic uncertainty, the closure relation is used to predict the number of events in the upper-rightmost region of each validation region. This prediction is then compared to the number of events observed in this region to determine non-closure. Closure in data after applying the simulation based correction (C_{Data}) is calculated as described in equation (7.2). Residual non-closure values are only analyzed for validation boundary values where the estimated signal contamination in the validation region's effective A region is less than 5%. This prevents the presence of possible signal in data from biasing the value of the systematic uncertainty.

$$C_{Data} = \alpha \frac{N_{\text{pred.},A}}{N_{\text{obs.},A}}\bigg|_{\text{Data}} \tag{7.2}$$

If the closure correction works well, then the corrected data closure will be near a value of 1.0 for any of the validation regions. However, a systematic uncertainty can be estimated in order to correct any residual non-closure that is seen when performing this test. The data-based non-closure systematic is calculated as the reciprocal of the maximum corrected data closure value for all validation regions.

The validation region scanning procedure is conducted for data, and the corrected data closure values are shown in figure 7.10 for the low mass optimization and figure 7.11 for the high mass optimization. The reciprocal of these values are applied to the $t\bar{t}+jets$ prediction as a systematic uncertainty in the final fit. The final values of the systematic uncertainties for the two optimizations are shown in table 7.2.

Simulation-based Systematics

As the closure correction is derived in simulation, there is inherent uncertainty in the calculated value related to mismodeling of simulated events and how the network inferences on events with different feature values. This uncertainty is directly linked to the uncertainty in the physics modeling (variations in POWHEG and PYTHIA parameters) and detector modeling (jet energy correction and jet energy resolution variations) of the simulation. Investigating how these variations affect the closure correction, α, at high N_{Jets} provides additional confidence that deviations in data vs. simulation modeling do not affect the overall shape of the background prediction for $t\bar{t}+jets$.

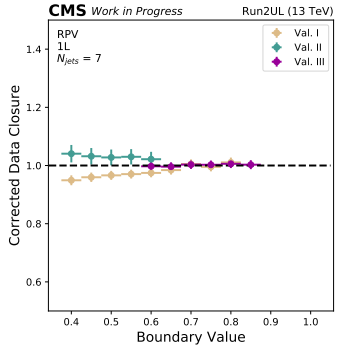

Figure 7.10: Corrected data closure values are plotted for the low mass optimizations as a function of the boundary value defining the three validation regions. The corrected data closure value is computed as in equation (7.2). The maximum value of corrected data closure for any of the three validation regions is used in computing the data-based systematic uncertainty per channel and model. Results for all channels and models are shown in figure A.20.

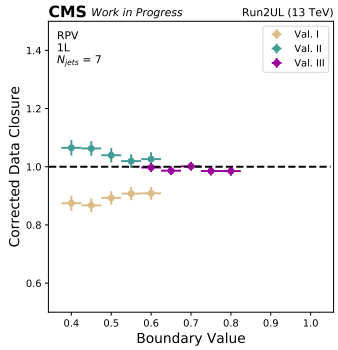

Figure 7.11: Corrected data closure values are plotted for the high mass optimizations as a function of the boundary value defining the three validation regions. The corrected data closure value is computed as in equation (7.2). The maximum value of corrected data closure for any of the three validation regions is used in computing the data-based systematic uncertainty per channel and model. Results for all channels and models are shown in figure A.21.

Table 7.2: Data-based non-closure systematic uncertainty values are shown for all channels, models and optimization. Note that these values are derived by taking the maximum absolute deviation from one of all corrected data closures in figure 7.10 and figure 7.11. These systematic uncertainties are then computed by taking the reciprocal of the maximum corrected data closure. All values are applied multiplicatively to the $t\bar{t} + jets$ prediction in the final fit.

Model	**Channel**	**Low Mass Opt.**	**High Mass Opt.**
	0ℓ	0.920	0.921
RPV	1ℓ	1.062	1.201
	2ℓ	1.072	0.873
	0ℓ	1.231	1.265
Stealth $SY\bar{Y}$	1ℓ	0.982	1.088
	2ℓ	1.076	1.153

A separate closure correction is derived for each variation of the $t\bar{t}+jets$ background in the signal region. The systematic uncertainty for each variation is computed by taking the ratio between the closure correction using the nominal $t\bar{t} + jets$ sample and the closure correction computed with a variation of the $t\bar{t} + jets$ events, e.g.:

$$\delta_{t\bar{t}+jets\ Var.} = \frac{\alpha_{Var.}}{\alpha_{Nom.}} \tag{7.3}$$

where $\alpha_{Var.}, \alpha_{Nom.}$ are the simulation based closure corrections computed using the variational and nominal $t\bar{t} + jets$ samples, respectively.

The closure correction also accounts for the statistical uncertainty based on the finite number of events in each of the A, B, C, and D regions as can be inferred from equation (7.1). As a comparison, the size of simulation-based systematics for each $t\bar{t} + jets$ variation are compared to the statistical uncertainty on the nominal closure correction. Figure 7.12 shows the values of the closure correction ratios for each of the $t\bar{t}+jets$ variation samples for the RPV model using the high mass optimization. For the other optimization and model, all of the variations fall within the statistical uncertainty band of coverage. These ratio values are explicitly the value of the potential systematic uncertainty that could be applied to the $t\bar{t} + jets$ prediction. The gray band is the statistical uncertainty on the nominal closure correction divided by the nominal closure

correction itself (to allow direct comparison with the $t\bar{t} + jets$ variation systematics).

The modeling systematics are separated into three categories: top quark modeling, event-level, and detector systematic uncertainties. These are considered separately since some of these categories contain correlated statistical uncertainties with the nominal $t\bar{t} + jets$ sample. Thus, only the statistical uncertainty for top quark modeling systematics uncertainties are considered when determining coverage by the nominal statistical uncertainty.

Nearly all of the closure correction ratios for each of the variational $t\bar{t} + jets$ samples fall within the grey band for the statistical uncertainty, implying that including an additional systematic uncertainty for these variations is unnecessary. The one caveat to this is the FSR variation which strays outside of the grey envelope. Thus, this will be incorporated as an additional systematic uncertainty in the final fit. This systematic uncertainty is applied multiplicatively to the $t\bar{t} + jets$ prediction in the A region on a per-N_{Jets} basis.

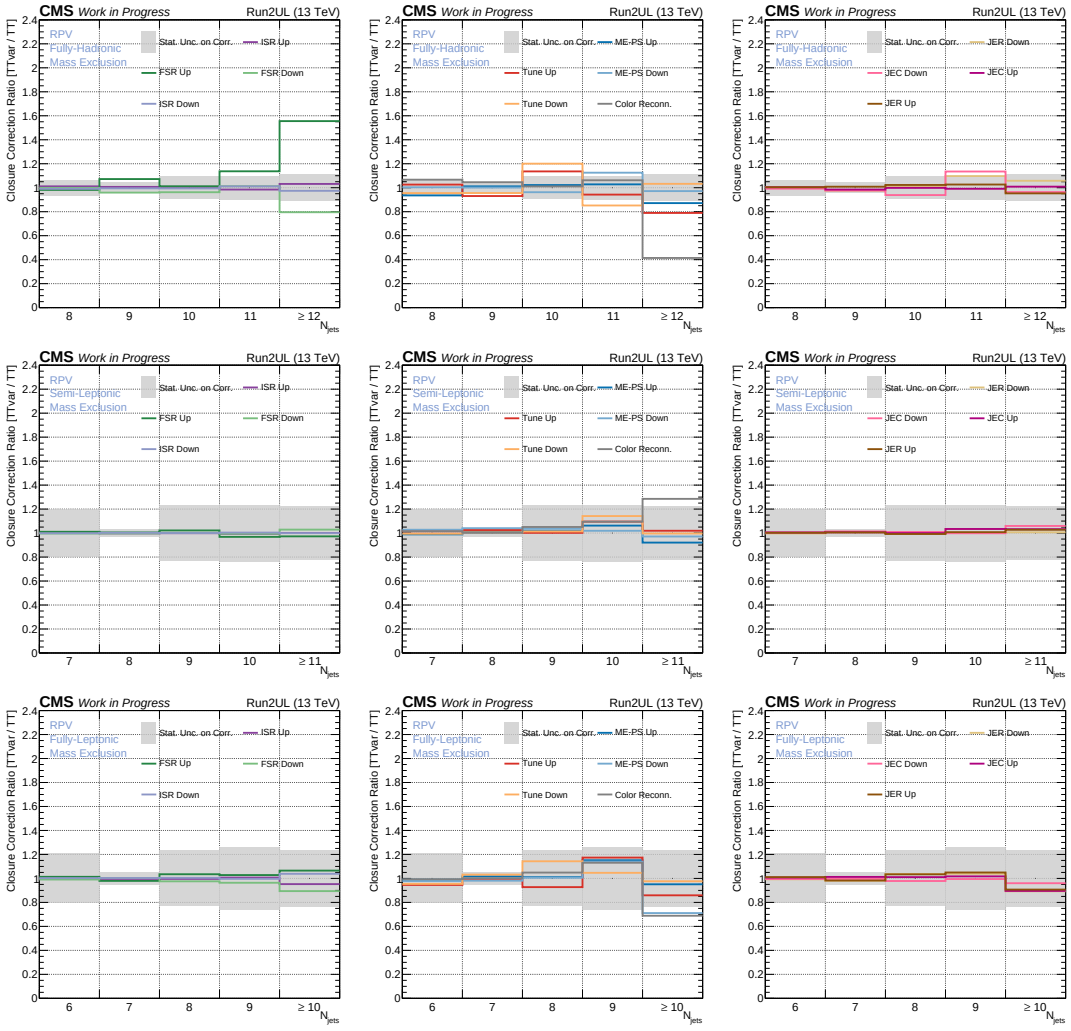

Figure 7.12: For the 0ℓ (top), 1ℓ (middle) and 2ℓ (bottom) channels with the RPV high mass optimization, the closure correction ratios for event based (left), detector based (middle), and top modeling based (right) variations are shown. The gray band given on this plot is the pure statistical uncertainty on the $t\bar{t}+jets$ prediction (from simulation). Note that for the RPV/Stealth $SY\bar{Y}$ low mass optimization and for the stealth $SY\bar{Y}$ high mass optimization, none of the closure correction ratios for any variation fall outside of the grey band, implying that no additional systematic uncertainty need be applied to correct for them. The FSR variational sample displays a systematic uncertainty value larger than the statistical uncertainty for the 0ℓ channel. Therefore, an additional systematic uncertainty is applied to the $t\bar{t}+jets$ prediction for FSR.

7.2 Minor Background Prediction

The ABCD regions established in section 7.1 are used to define the analysis bins for each of the other minor backgrounds. For the TTX and Other background categories, estimates can be taken directly from simulation as the production cross-section for these processes yield a miniscule number of events in comparison to $t\bar{t}+jets$. However, additional work is needed to determine an accurate prediction for the QCD multijet background. Notably, simulation of QCD multijet events is difficult due to the non-perturbative nature of quantum chromodynamics. This leads to large event weights and, in turn, large statistical uncertainty. To circumvent this issue, the background prediction for the QCD multijet process is computed using a data control region. This procedure is detailed in section 7.2.1.

7.2.1 QCD Multijet Background Prediction

Estimation of the QCD multijet background is carried out using a control region which is defined orthogonally to the signal region for each channel. This method produces an indirect estimate of the number of QCD multijet events in the signal region through the use of transfer factors. Events in the control region are required to have at least one non-isolated muon in order to pass the selection. Additionally, the number of good leptons in each of the control regions is required to be zero. All other selection criteria for the control region are the same as the signal region to ensure that events in the control region are kinematically similar to signal-like events.

Figure 7.13 displays the N_{Jets} distribution for the control region selections in the 0ℓ, 1ℓ, and 2ℓ cases. The inversion of the good lepton and non-isolated muon cuts results in a control region which is enriched in QCD multijet events. Additionally, the agreement between data and simulation distributions is poor, as indicated by the ratios in the bottom panels. To mitigate this issue, the QCD multijet background is predicted separately per N_{Jets} bin.

The background prediction proceeds by calculating a transfer factor, TF_i between the signal and control regions:

$$TF_i = \frac{N_i^{SR}}{N_i^{CR}} \tag{7.4}$$

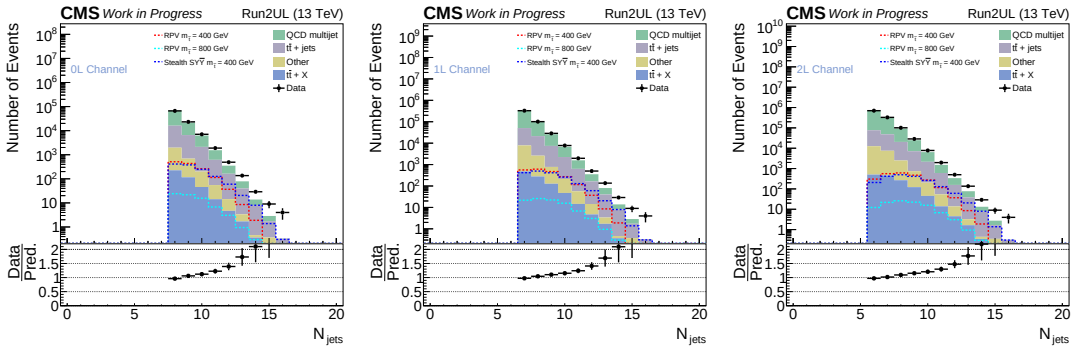

Figure 7.13: The N_{Jets} distributions are shown for the three QCD control region selections (left to right: 0ℓ, 1ℓ, and 2ℓ). The discrepancy in agreement between data and simulated events at high N_{Jets} necessitates individual transfer factor computation per N_{Jets} bin.

where N_i^{SR} is the QCD multijet event yield in the signal region in simulation and N_i^{CR} is the event yield in the control region in simulation. The subscript i denotes the N_{Jets} bin in which the transfer factor is derived. The number of expected QCD multijet events in the signal region can be derived by taking:

$$N_{\text{Data},i}^{SR} = TF_i N_{\text{Data},i}^{CR} \tag{7.5}$$

where $N_{\text{Data},i}^{SR}, N_{\text{Data},i}^{CR}$ are the number of QCD multijet events in the signal region and control region in data, respectively. In practice, some N_{Jets} bins only have a small number of QCD multijet events in them, leading to statistical fluctuations in the transfer factor values from bin to bin. Thus, a simple polynomial fit is first performed on the signal and control region N_{Jets} distributions before computing the transfer factor. The values of these transfer factors are shown in figure 7.14.

Note that the fundamental assumption in this procedure is that:

$$\frac{N_i^{SR}}{N_i^{CR}} = \frac{N_{\text{Data},i}^{SR}}{N_{\text{Data},i}^{CR}} \tag{7.6}$$

This assumption can only be broken if modeling disagreement between the data and simulation differs in the signal and control regions. Therefore, additional systematic

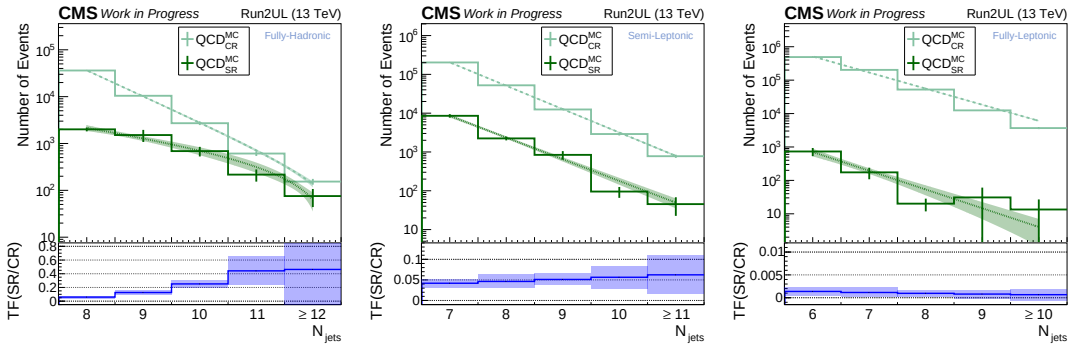

Figure 7.14: The QCD control region transfer factors are shown for each of the three channels (left to right: 0ℓ, 1ℓ, and 2ℓ).

uncertainties are estimated to combat any differences in modeling between the two. These are discussed further in section 7.3.

7.3 Systematic Uncertainty Estimation

One of the most crucial aspects of high-energy physics analyses is understanding and estimating systematic uncertainties in a reasonable fashion. This serves as a test of the fundamental assumptions of the analysis. Systematic uncertainties are a way to determine how the results of an analysis would change given that the understanding of some aspect of the analysis is not correct. As an example, one could estimate how an analysis would change under the assumption that the N_{Jets} distribution for the analysis region is slightly different in data than in simulation due to mismodeling of FSR. By estimating and applying systematic uncertainties, analyzers can ensure that their result is robust to misunderstandings of the underlying physics or detector modeling.

In this analysis, the most basic assumption is that the two discriminant variables are independent and can be used to derive the background estimation in the A region for $t\bar{t} + jets$. Thus, all systematic uncertainties are estimated independently for each of the ABCD regions. Additionally, this analysis depends heavily on the modeling of the N_{Jets} spectrum for all background and signal processes. Therefore, uncertainties will also be estimated separately for each N_{Jets} bin. In all, systematic uncertainties are computed for each of the 20 ABCD regions by N_{Jets} bins independently.

7.3.1 Systematic Uncertainty Functional Form

All systematic uncertainties in this analysis take the form of a histogram containing 20 bins, one for each ABCD region and N_{Jets} bin. These histograms represent the $\pm 1\sigma$ variations in the source of the systematic uncertainty. For detector and physics modeling effects, these are computed as the ratio of the event yield between the nominal and varied event yield:

$$R_{Sys.}(i,j) = \frac{N_{Sys.}(i,j)}{N_{Nom.}(i,j)}, \tag{7.7}$$

where $R_{Sys.}(i,j)$ is the systematic uncertainty ratio and $N_{Sys.}(i,j)$ and $N_{Nom.}(i,j)$ are the event yields in the systematically varied and nominal case for the ith ABCD region and the jth N_{Jets} bin.

Each of these ratios adjust the background estimate for a given process as such:

$$N(i,j) = N_0(i,j) R_{Sys.}(i,j)^{\theta} \tag{7.8}$$

where N is the final event yield and N_0 is the nominal event yield estimate. Here, θ represents a nuisance parameter within the fit which is an additional parameter indicating the "strength" of a given systematic uncertainty. A θ value of ± 1 indicates a systematic uncertainty which is pulled by a factor of $\pm 1\sigma$ within the fit. Each systematic uncertainty inherits an independent nuisance parameter.

7.3.2 Sources of Uncertainty

For the $t\bar{t} + jets$ background which is estimated using data, the main systematic uncertainties are those which correct for differences in ABCD closure between data and simulation. These are the data-based non-closure and FSR closure correction systematic which are discussed in section 7.1.3. Note that both of these uncertainties are only applied in the A region as they are computed based on ratios of the closure corrections. No other systematic uncertainties are applied for this background due to the fact that background estimates in the A region are all derived using data events.

For minor backgrounds (QCD mulitjet, TTX, and Other) and signal, systematic uncertainties are computed for various physics modeling discrepancies using simulation. Most notably, systematic uncertainties are calculated for event generation parameters which are responsible for establishing the number of jets in an event. These are initial state radiation (ISR), final state radiation (FSR), (μ_R, μ_F) scales, parton distribution function modeling, and pileup reweighting. Systematic uncertainties which address these modeling parameters are computed by varying each of the parameters up and down by a factor of $\sqrt{2}$ and $1/\sqrt{2}$. Then, a ratio of the nominal background event yield and the up/down variation background estimations are taken for each of the background collections. This results in an estimate for the $\pm 1\sigma$ systematic uncertainty per ABCD region and N_{Jets} bin. Each of these systematic uncertainties are applied multiplicatively to the background estimate for a given process and a log-normal nuisance parameter is included in the final fit.

Other systematic uncertainties are applied for differences in detector modeling between data and simulation. Note that these effects can be caused by difference in detector resolution as well as issues with detector operation that are not properly modeled in simulated events. Detector resolution effects include lepton identification and triggering, jet energy scale (JES), jet energy resolution (JER), bottom and top quark tagging,

and jet triggering. Additionally, a systematic uncertainty is estimated for a timing issue within the ECAL sub-detector (known as prefiring). This timing issue caused energy from previous bunch crossings to be associated with the one after which can potentially cause inaccurate reconstruction of events. Therefore, a systematic uncertainty is applied which accounts for the probability of this phenomenon occurring.

The scale of each of the systematic uncertainties described above is shown in table 7.3 for the RPV search with the low-mass ABCD bin edge optimization. The systematic uncertainties for the other three search/optimization combinations have similar scales.

Table 7.3: Magnitude of systematic uncertainties for the 0ℓ, 1ℓ, and 2ℓ channels based on the RPV-trained NN and ABCD bin boundaries optimized for low $M_{\tilde{t}}$. Reported values are in units of % and ranges correspond to the 16th and 84th percentile for the value of a systematic uncertainty across all applicable analysis regions (ABCD regions and N_{Jets} categories). The maximum value for a given systematic uncertainty across these regions is shown in parentheses. The "Other" category includes QCD multijet, TTX, and rare multiboson processes. The systematic uncertainties for the RPV signal shown correspond to the sample with $M_{\tilde{t}} = 550$ GeV.

Source of	0ℓ			1ℓ			2ℓ		
uncertainty	$t\bar{t}+jets$	Other	RPV	$t\bar{t}+jets$	Other	RPV	$t\bar{t}+jets$	Other	RPV
PDF	–	< 1 (1)	< 1	–	< 1 (1)	< 1	–	< 1 (2)	< 1
(μ_{R}, μ_{F}) scales	–	1–8 (16)	0–1 (2)	–	0–5 (13)	0–1 (2)	–	0–7 (19)	0–1 (2)
FSR	1–3 (3)	1–14 (56)	1–12 (18)	0–1 (3)	1–12 (26)	1–7 (15)	0–6 (9)	2–21 (100)	1–9 (16)
ISR	–	0–10 (17)	1–4 (5)	–	0–8 (15)	1–4 (5)	–	3–11 (17)	0–4 (5)
Pileup	–	0–2 (12)	0–1 (3)	–	0–1 (22)	< 1 (3)	–	0–9 (26)	0–1 (9)
Nonclosure	3	–	–	5	–	–	7	–	–
κ Stat. Unc.	1–4 (7)	–	–	1–7 (8)	–	–	5–15 (19)	–	–
QCD TF	–	1–4 (16)	–	–	< 1 (1)	–	–	< 1	–
JES	–	4–18 (100)	0–10 (27)	–	1–18 (100)	1–14 (19)	–	4–100 (100)	1–15 (25)
JER	–	0–8 (23)	0–2 (4)	–	0–5 (35)	0–1 (4)	–	0–20 (45)	0–3 (9)
b tagging	–	0–1 (7)	1–2 (3)	–	0–1 (3)	< 1	–	0–1 (7)	< 1 (2)
t tagging	–	26–33 (42)	26–31 (34)	–	–	–	–	–	–
Jet trigger	–	0–1 (1)	< 1 (1)	–	–	–	–	–	–
Lepton ID	–	–	–	–	3–3 (4)	3–3 (3)	–	5–6 (6)	5–5 (6)
Prefiring	–	0–2 (7)	0–2 (3)	–	0–3 (6)	0–3 (4)	–	0–3 (11)	0–2 (4)
Integrated Luminosity	–	1.6	1.6	–	1.6	1.6	–	1.6	1.6
Theoretical Cross Section	–	20	–	–	20	–	–	20	–

7.4 Fitting Procedure

The final results of this analysis are created using the statistical framework COMBINE [86] which is based on the ROOT and ROOFIT packages [87, 88]. COMBINE generates and fits a likelihood function based on the number of events observed in the signal region while taking into account the background estimations and systematic uncertainties described in the previous sections. In doing so, COMBINE determines the likelihood that signal events exists within the observed data.

7.4.1 Constructing the Likelihood

Statistical analysis in high energy physics experiments is carried out by determining the probability that a given observation is made in data under certain initial conditions of the physics model being investigated. This can be represented using a likelihood function, $\mathcal{L}(\vec{\Phi})$ which depends on the parameters of the model. The likelihood is equivalent to the probability of an observation data given a set of model parameters:

$$\mathcal{L}(\vec{\Phi}) = p(data; \vec{\Phi}) \tag{7.9}$$

The likelihood can be factorized into two portions:

$$\mathcal{L} = \mathcal{L}_{Primary} \cdot \mathcal{L}_{Auxiliary} \tag{7.10}$$

where $\mathcal{L}_{Primary}$ is the probability of observing an event count in data given a set of model parameters, and $\mathcal{L}_{Auxiliary}$ corresponds to the external constraints on the parameters of the model. The primary likelihood encodes information regarding probability that the primary observable (in this analysis, that is the signal strength parameter r) takes on a given value. The auxiliary likelihood pertains to the probability that any other parameters (called *nuisance parameters*) have a certain value.

This analysis uses a binned approach for computing the likelihood. Thus, $\mathcal{L}_{Primary}$ is formulated using the product of Poisson distributions for each of the bins given the number of expected events in each bin:

$$\mathcal{L}_{Primary}(\vec{\Phi}) = \prod_{bins} Poiss(n_{obs,bin}; n_{exp,bin}(\vec{\Phi})) \tag{7.11}$$

That is, given the expected number of events, equation (7.11) defines the probability of observing a given distribution in data.

The auxiliary likelihood describes the probability that nuisance parameters take on a given value. Given a set of E previous observations y, the auxiliary likelihood can be written as:

$$\mathcal{L}_{Auxiliary} = \prod_{e=1}^{E} p_e(y_e; \nu_e) \tag{7.12}$$

where p_e signifies the probability of ν_e taking on a certain value given the previous observation y_e. In practice, $\mathcal{L}_{Auxiliary}$ encodes information about constraints on the systematic uncertainties for the analysis.

The crux of the analysis is determining how $n_{exp,bin}$ depends on the model parameters and deciding the proper constraints to place on them. In most analyses – and in this one – the binned likelihood is defined using templates, which are pre-defined histograms taken from either simulation or data describing the expected number of events in each bin. The number of expected events for a single bin can then be written as a function of the parameter of interest and the effects of the nuisance parameters:

$$n_{exp}(r, \vec{\nu}) = \max\left(0, \sum_{p} M_p(r) N_p(\vec{\nu}) \omega_p(\vec{\nu}) + E(\vec{\nu})\right) \tag{7.13}$$

where p indexes a given physics process and M_p, N_p, ω_p, and E represent the normalization effect of the parameter of interest, the nuisance parameters, expected number of events from the template, and the statistical uncertainty from the samples. M_p is a simple multiplicative factor which takes the value of r if a processes is signal, otherwise 1. N_p represents the product of all systematic uncertainties for a given process with $\vec{\nu}$ dictating their relative strength.

Therefore, the final likelihood function takes the form:

$$\mathcal{L} = \prod_{c=1}^{N_c} \prod_{b=1}^{N_b^c} Poiss(n_{cb}; n_{cb}^{exp}(r, \vec{\nu})) \cdot \prod_{e=1}^{N_E} p_e(y_e; \nu_e) \tag{7.14}$$

where c indexes over channels, b indexes over bins, and e indexes over nuisance parameters. The final fit entails determining the maximum likelihood values of the parameter

of interest, r, as well as the nuisance parameters for the observation made in data.

7.4.2 Fit Setup

Construction of the likelihood for the fit is carried out using COMBINE. The background estimates described in the previous sections are included as histograms binned based on both N_{Jets} and ABCD regions, resulting in 20 analysis bins per channel. These templates are included in the fit as ω_p in equation (7.13).

The ABCD relation is applied in the fit to determine the $t\bar{t} + jets$ normalization in the A region per N_{Jets} bin. This is computed using rate parameters for the B, C, and D region $t\bar{t} + jets$ event yield normalization. These parameters have a uniform constraint meaning that they are allowed to float between zero and ten times their nominal value. A rate parameter for the normalization in the A region is constrained using the closure relationship as computed with the rate parameters for B, C, and D.

The QCD multijet background estimate is computed via the transfer factor method described in section 7.2.1. This procedure results in the 20 N_{Jets} by ABCD binned predictions for the number of QCD events in each analysis bin. All other background and signal processes are taken directly from simulation

Systematic uncertainties are included for each of the sources specified in section 7.3. For each source, a 20 bin histogram is created by taking the ratio of the up/down variational and nominal samples. The nuisance parameters for these systematic uncertainties are subject to a Gaussian constraint. Finally, statistical uncertainties on simulated samples are included on a per-bin and per-process basis.

A description of all parameters included in the fits are shown in table 7.4.

Table 7.4: A collection of different parameters in the final fit for $t\bar{t} + jets$, QCD, TTX, Other backgrounds and Stealth $SY\bar{Y}$ model with $M_{\tilde{t}} = 550$ GeV. Including systematic uncertainties, components of the background prediction, and correction factors. Abbreviated names for each source of uncertainty are explained in the text.

Systematic Uncertainty	Parameter Name	**Correlated per**	**Calculation**
Integrated Luminosity	lumi	All Bins, All Processes	1.6%
Theoretical Cross Section	Other, TTX	Bin, Process	20%
Pileup	pu	Bin, Process	Up/Down variations
Prefiring	prf	Bin, Process	Up/Down variations
Lepton ID/trigger	lep	Bin, Process	Up/Down variations
Jet trigger	jet	Bin, Process	Up/Down variations
b tagging	btg	Bin, Process	Up/Down variations
Top tagging	ttg	Bin, Process	Up/Down variations
(μ_R, μ_F) scales	scl	Bin, Process	Up/Down variations
PDF	pdf	Bin, Process	Up/Down variations
FSR	fsr	Bin, Process	Up/Down variations
ISR	isr	Bin, Process	Up/Down variations
JES	JEC	Bin, Process	Up/Down variations
JER	JER	Bin, Process	Up/Down variations
MC Stat.	mcStat	Uncorrelated	Per bin raw event number and avg. event weight
QCD CR Stat.	mcStat	Uncorrelated	Per bin event yield and $\mathrm{TF}_i = \frac{N_i^{SR}}{N_i^{CR}}$
QCD CR TF Stat.	QCD_TF	Bin, Process	Propagation of stat. unc. for TF from MC
$t\bar{t} + jets$ Non-Closure (Post-Correction)	CorrectedDataClosureA	$t\bar{t} + jets$ A region	(Closure Correction $* N_{\mathrm{pred.},A}$)/$N_{\mathrm{obs.},A}\vert_{\mathrm{Data}}$
Fit Components	Parameter Name	**Applied to**	**Value**
$t\bar{t} + jets$ Background Prediction in B,C,D	beta, gamma, delta	$t\bar{t} + jets$ each B, C, D region	0 to 10 times MC prediction
$t\bar{t} + jets$ Background Prediction in A	-	$t\bar{t} + jets$ A region	$N_{\mathrm{pred.},A} = (N_{\mathrm{pred.},B}\ N_{\mathrm{pred.},C})$ / $N_{\mathrm{pred.},D}$
$t\bar{t} + jets$ Closure Correction Factor	-	$t\bar{t} + jets$ A region	$N_{\mathrm{obs.},A}$ / $N_{\mathrm{pred.},A}$

Chapter 8

Results and Interpretation

Using the statistical framework outlined in section 7.4, final results can be extracted from the analysis. These results are computed using the likelihood function generated from the background predictions for each of the processes in the signal region. The two main results of interest are the p-values describing the probability that signal exists within the observed data as well as the upper limits on the top squark production cross-section.

To claim discovery of a new particle or decay process in high-energy physics, the observation made must have a discrepancy of at least five standard deviations (colloquially referred to as the "5σ" threshold). Numerically, this standard represents a $5 \cdot 10^{-7}$ probability that the observed excess of events is due to statistical fluctuations in the data. P-values can be expressed in terms of the number of standard deviations that the observation is from the SM-only prediction. Thus, p-value distributions are shown describing the size of any observed excesses.

Signal production cross-section limits are computed by determining confidence bands on the signal strength parameter and converting this value to a cross-section. Certain model hypotheses can be excluded by comparing these bands to the theoretical cross-section prediction. As the model is parameterized by $M_{\tilde{t}}$, this procedure places an exclusion limit on the mass of the top squark.

The following sections outline the findings from the analysis. Section 8.1 outlines the statistical methods used for computing both p-values and cross-section limits. Then in section 8.2, the final results of the analysis are presented with accompanied by a

physical interpretation of the findings.

8.1 Statistical Methods

COMBINE can be used in different modes of operation in order to extract results from the analysis. These modes of operation can be used for not only finding the maximum likelihood values of each of the parameters in the fit but also to determine confidence intervals on these parameters. The confidence intervals are crucial for making final statements about the likelihood that signal exists in data as well as setting limits on the top squark production cross section. These methods are further discussed below.

8.1.1 Maximum Likelihood Fitting

Maximum likelihood fits are a means for generating a point estimate for the model parameters that have the highest probability of generating the observed data. This includes fitting both the parameter of interest (the signal strength parameter r) and the nuisance parameters, $\vec{\nu}$. Therefore, the maximum likelihood values of the parameters are defined as:

$$(\hat{r}, \hat{\vec{\nu}}) = \underset{r,\vec{\nu}}{\operatorname{argmax}}\, \mathcal{L}(r, \vec{\nu}) \tag{8.1}$$

These values are generated through a minimization algorithm known as Minuit [89].

Maximum likelihood fits are used to determine the post fit N_{Jets} by ABCD region distributions for each of the background and signal components. These values are compared to data to show the maximum likelihood signal strength parameter found by the fit.

8.1.2 Profile Likelihood and Confidence Intervals

Another important aspect of the fitting procedure is determining uncertainties on fit parameters. These uncertainties are necessary for making claims about discovery as well as for setting exclusion limits. For example, the uncertainty on the signal strength parameter r dictates the statistical significance of the final results.

To compute uncertainties on fit parameters, COMBINE uses a method known as *profiling* to determine the likelihood for any given value of the parameter of interest.

That is:

$$\mathcal{L}(r) = \max_{\vec{\nu}} \mathcal{L}(r, \vec{\nu}). \tag{8.2}$$

The profile likelihood is useful in determining confidence intervals as it can be determined how often a range of r values would contain the true maximum likelihood value $\hat{r}$. The profile likelihood ratio compares the value of the profile likelihood for any value of the parameter of interest to the maximum likelihood value as:

$$\Lambda \equiv \frac{\mathcal{L}(r, \vec{\nu}(r))}{\mathcal{L}(\hat{r}, \hat{\vec{\nu}})} \tag{8.3}$$

where the numerator represents the profile likelihood and the denominator is the maximum likelihood value [90].

Confidence intervals are defined as the range of values for r where the profile likelihood ratio meets a given cutoff criteria γ_{CL}. For technical purposes, the negative logarithm of Λ is used as the metric for making comparisons to the cutoff. Confidence intervals on the signal strength parameter can then be defined as:

$$\{r\}_{CL} = \{r : -\log(\Lambda) < \gamma_{CL}(r)\}. \tag{8.4}$$

In practice, this interval is determined by first finding the best fit value $\hat{r}$ by minimizing the negative log-likelihood. Then, different values of r_i are profiled to determine the nearest points where $-\log(\Lambda) = \gamma_{CL}$ in either direction of scanning [91]. For all fits in this analysis, γ_{CL} corresponds to the 95% confidence level that the $\hat{r}$ resides in the interval.

8.1.3 Claiming Discovery

Claims about the existence of a new particle or process can be assessed by comparing the likelihood of the background only hypothesis to the signal+background hypothesis. The test statistic used for this calculation is:

$$q_0 = \begin{cases} 0 & \hat{r} < 0 \\ -2\log\left(\frac{\mathcal{L}(r=0)}{\mathcal{L}(\hat{r})}\right) & \hat{r} \geq 0 \end{cases} \tag{8.5}$$

where the test statistic limits claims which correspond to negative signal strengths. The value of this test statistic for the observed data can then be compared to the distribution for the SM-only test statistic to derive a p-value [92]. Note that the criteria for claiming discovery is a p-value of $5 \cdot 10^{-7}$, meaning that there is an 0.00005% chance that a statistical fluctuation in the data could manifest a result as extreme as the seen in the experiment..

8.1.4 Limit Setting

One can place limits on the signal strength parameter by using the profile likelihood ratio to construct a test statistic:

$$t_r = -2 \log \left(\frac{\mathcal{L}(r, \vec{\nu})}{\mathcal{L}(\hat{r}, \hat{\vec{\nu}})} \right) \tag{8.6}$$

where limits can be set by determining where t_r takes on a value corresponding to the confidence interval cutoff, γ_{CL}.

However, two safeguards need to be applied to this test statistic to protect from undesirable results. First, the test statistic should only establish upper limits on the signal strength parameter as the test of interest is determining whether new physics exists. Also, the test statistic should not allow for a limit with a negative value. Therefore, the final test statistic is given as:

$$q_0 = \begin{cases} -2 \log \left(\frac{\mathcal{L}(r)}{\mathcal{L}(r=0)} \right) & \hat{r} < 0 \\ -2 \log \left(\frac{\mathcal{L}(r=0)}{\mathcal{L}(\hat{r})} \right) & 0 \geq \hat{r} \geq r \\ 0 & r < \hat{r} \end{cases} . \tag{8.7}$$

Note that if this test statistic is used to derive limits on the p-value, there is a 5% chance that the experiment places a limit on the signal strength parameter even if it has no sensitivity. Therefore, high energy physicists make use of the CL_s criterion [93, 94], which requires that:

$$\frac{p_s}{1 - p_b} < 1 - CL \tag{8.8}$$

where p_s is defined as the p-value for the observation under the signal plus background

hypothesis and p_b is the p-value in the background only hypothesis and is taken from the opposite side of the test statistic distribution. Note that the left side of equation (8.8) tends to one for small values of r, mitigating limits set by small values of the parameter of interest.

8.2 Analysis Results

The final results of the analysis are shown in the following sections. First, maximum likelihood fits are performed on the observation in data using the templates described in section 7.4. The results of these fits show the prediction for all components of background with their respective uncertainties. Next, the p-value for each of the signal mass hypotheses are calculated and displayed. This is to make a claim about the presence of signal in data. Finally, upper limits are placed on the signal production cross section as a function of $M_{\tilde{t}}$.

Note that each set of results is computed for each of the channels individually as well as for the combination of all three channels.

8.2.1 Fit Distributions

Predicted event distributions are shown for background-only fits which have the signal strength parameter fixed to zero. The maximum likelihood values of each background process are displayed to show how well the background-only hypothesis matches the observed data. These distributions are shown for the 0ℓ, 1ℓ, and 2ℓ channels in figure 8.1 and figure 8.2 for the RPV and Stealth $SY\bar{Y}$ signal models, respectively. Data and the background prediction agree for almost all bins, signifying that the SM-only prediction is an adequate representation of data in most cases.

8.2.2 P-Value Distributions

Using the prescription outlined in section 8.1.2, p-values are computed for each signal mass hypothesis for both models. These results are shown in figure 8.3 and figure 8.4.

Note that the results for the 0ℓ and 1ℓ channel favor the SM-only hypothesis as they have p-values near 0.5 for all $M_{\tilde{t}}$ values. However, a $\sim 2\sigma$ local excess is observed for the majority of $M_{\tilde{t}}$ hypotheses in the 2ℓ channel for both the RPV and Stealth $SY\bar{Y}$

model. The maximum local significance value for the Stealth $SY\bar{Y}$ 2ℓ channel occurs for $M_{\tilde{t}} = 550$ GeV and has a value of 3.1σ.

The 2ℓ channel results arise due to a few factors. First, figure 8.1 and figure 8.2 indicate an excess in the $N_{\text{Jets}} = 9$ bin for the A and D region, accompanied by a deficit of events in the B and C regions. However, if signal events existed in data, this trend would persist for all N_{Jets} bins. Additionally, the 2ℓ channel suffers from a small expected event yield which amplifies the impact of any statistical fluctuations in data. Thus, though the results indicate an interesting excess of events, it is likely a statistical artifact rather than a real signature of signal events in data.

8.2.3 Setting Cross Section Limits

The results in section 8.2.2 show that the data are consistent with the SM-only hypothesis due to the lack of sufficient evidence to claim discovery. Therefore, fits are conducted in order to determine upper limits on the top squark pair production cross section for the RPV and Stealth $SY\bar{Y}$ models. These limits are used to set exclusions on the top squark mass to indicate regions of phase space where this analysis would have observed these signatures if they existed in data.

Cross section limits are calculated via the methodology described in section 8.1.4. The results from these fits are the mean expected upper limit on the signal strength parameter as well as the 68% and 95% confidence intervals for this value. The confidence intervals represent the one and two standard deviation bands on the expected cross section. Additionally, the observed upper limit on the cross section is calculated via a maximum likelihood fit to data. The cross section limit plots parameterized by $M_{\tilde{t}}$ are shown in figure 8.5 and figure 8.6 for the RPV and Stealth $SY\bar{Y}$ models, respectively.

If the SM-only hypothesis is correct, the observed and mean expected limits should coincide with one another. However, if there is an excess observed, the observed limit would lie above the expect limit with the distance of separation indicating the statistical significance of the observed excess.

For the two models, these results are used to determine a lower limit on the mass of the top squark decaying via the RPV or Stealth $SY\bar{Y}$ signatures. Mass limits are determined by comparing the observed cross section limit to the theoretical cross section of either model. The $M_{\tilde{t}}$ value where the two curves intersect indicates the highest mass

which this analysis has sufficient sensitivity to claim exclusion. Top squarks with a mass of $M_{\tilde{t}} = 700$ GeV and below decaying to a top and neutralino via the RPV coupling are excluded at a 95% confidence level. Additionally, an upper exclusion limit of $M_{\tilde{t}} = 930$ GeV is placed on top squarks decaying via the Stealth $SY\bar{Y}$ prescription.

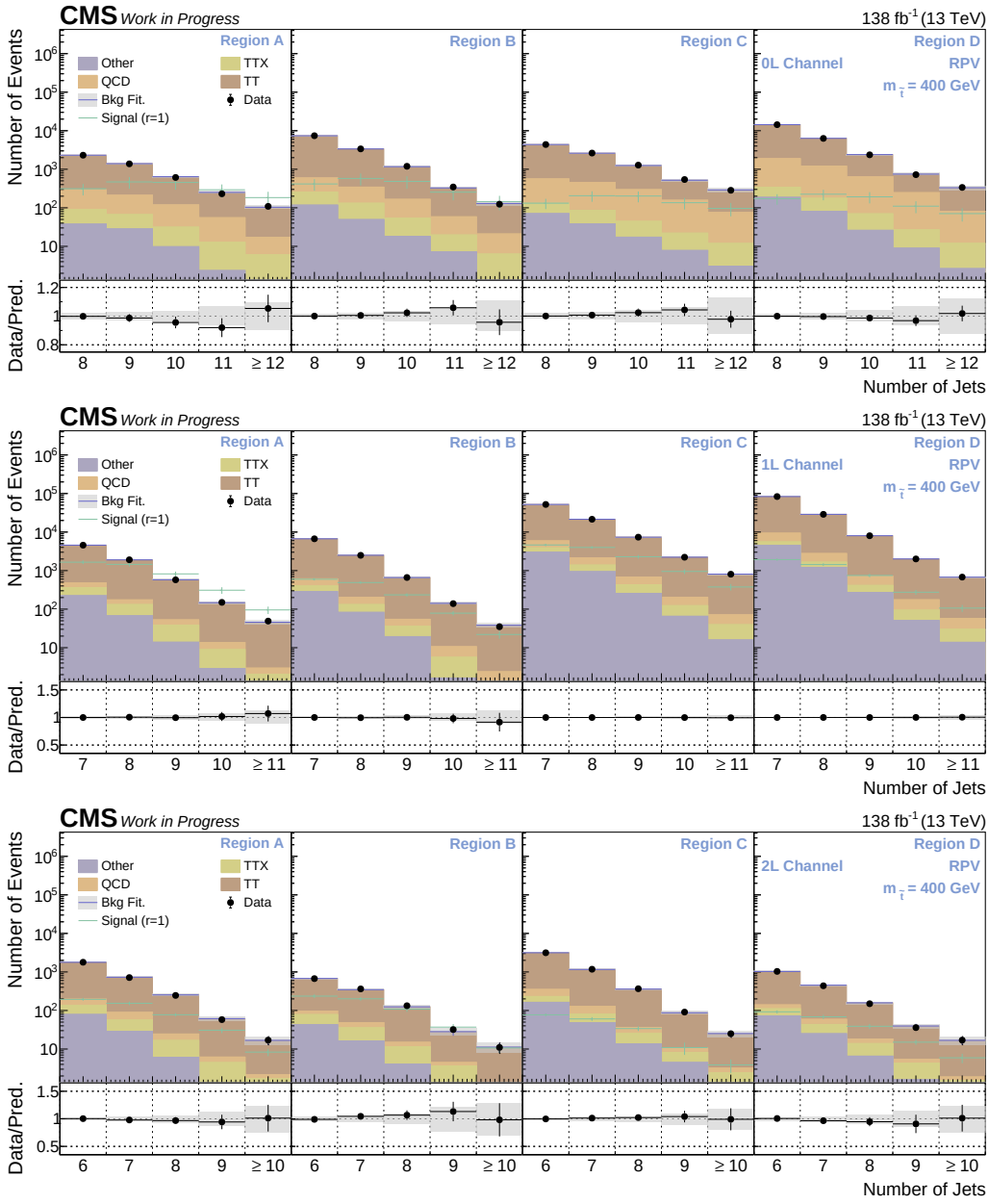

Figure 8.1: Background-only fit distributions are shown for the RPV model with $M_{\tilde{t}} = 400$ GeV for the 0ℓ, 1ℓ, and 2ℓ (top to bottom) channels. The background fit (shown in blue) includes both statistical and systematic uncertainty as calculated by the fit. The N_{Jets} by ABCD region distribution is also shown for signal (green) with a fixed signal strength of $r = 1$ for reference. Note that there is good agreement between the background prediction and data in almost all bins.

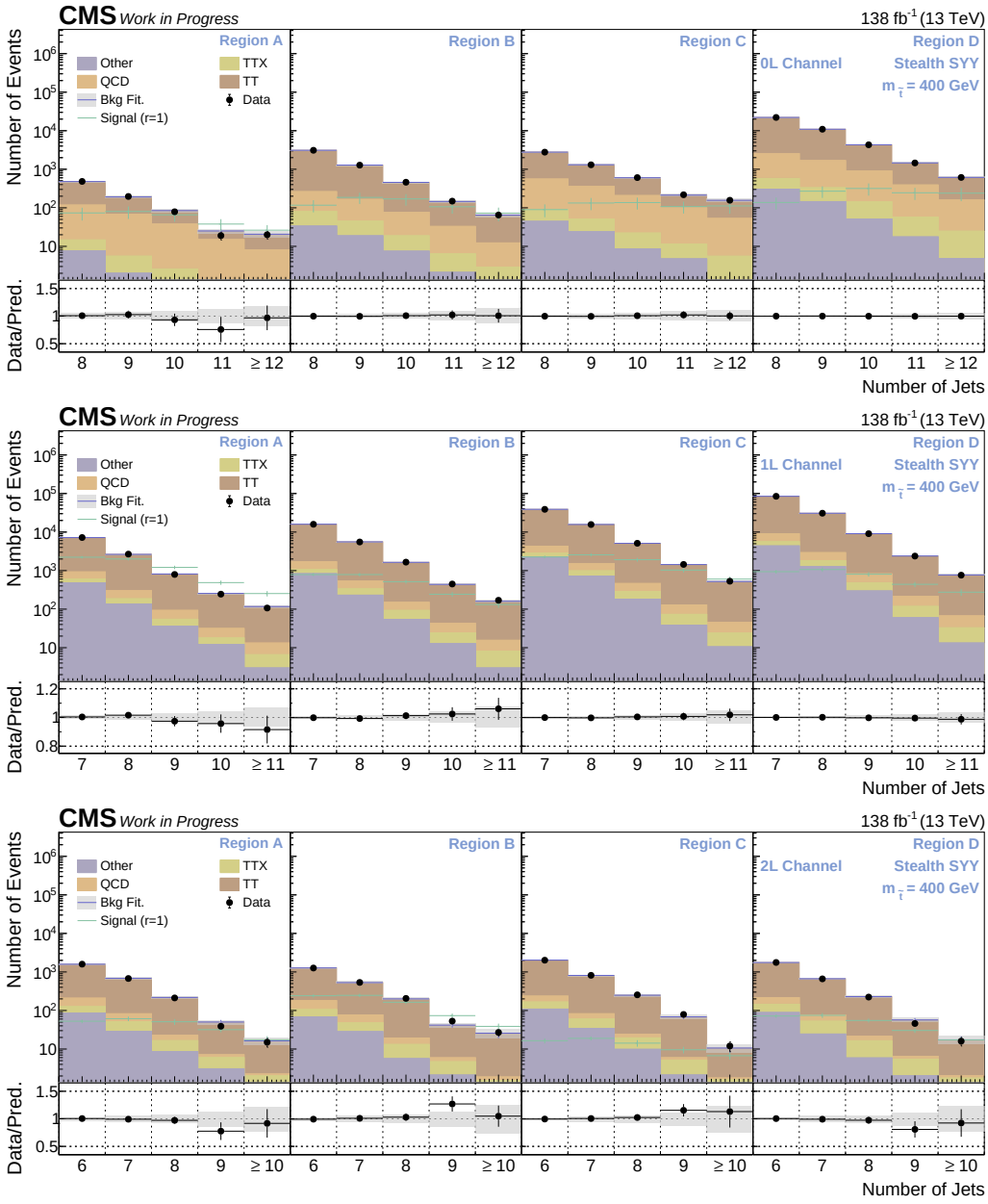

Figure 8.2: Background-only fit distributions are shown for the Stealth $SY\bar{Y}$ model with $M_{\tilde{t}} = 400$ GeV for the 0ℓ, 1ℓ, and 2ℓ (top to bottom) channels. The background fit (shown in blue) includes both statistical and systematic uncertainty as calculated by the fit. The N_{Jets} by ABCD region distribution is also shown for signal (green) with a fixed signal strength of $r = 1$ for reference. Note that there is good agreement between the background prediction and data in almost all bins.

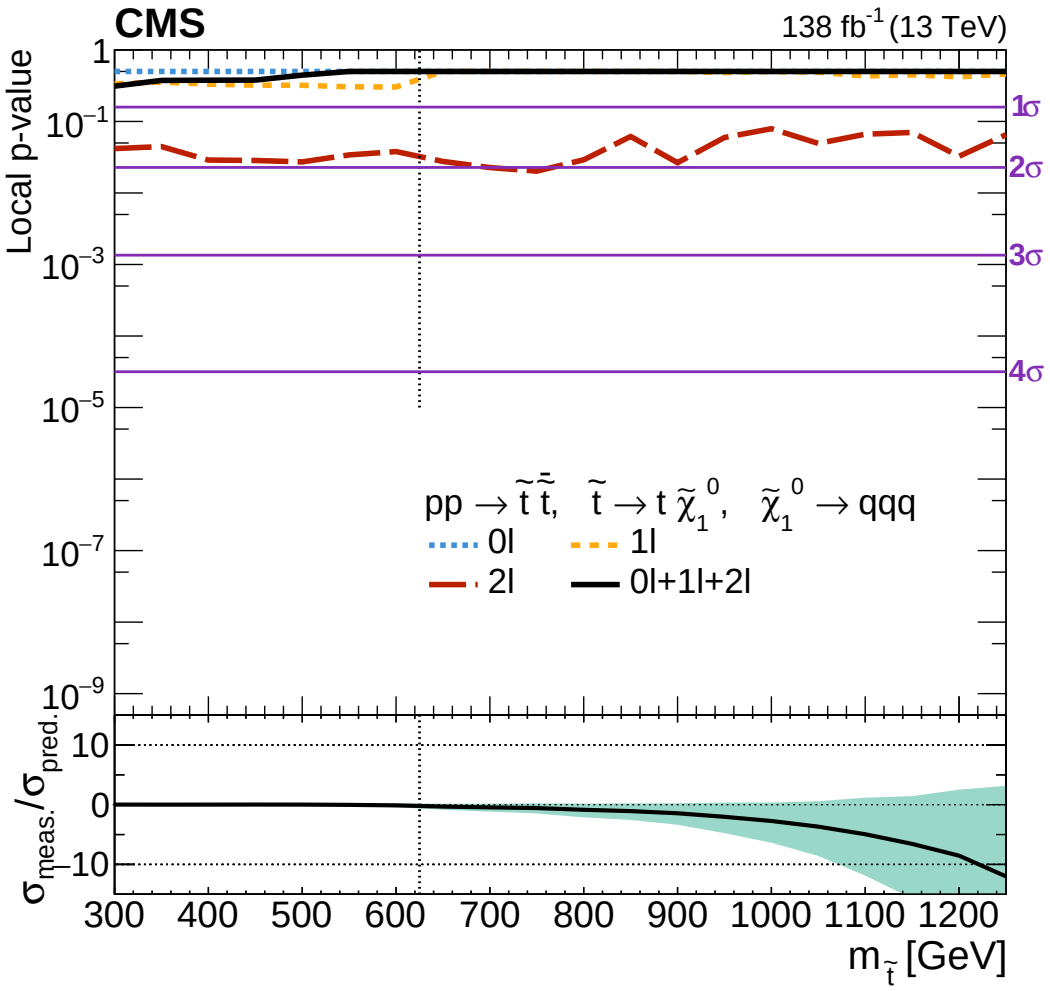

Figure 8.3: P-value distributions for the RPV model which outline the statistical significance of any excesses seen in data. Distributions are shown for fits to each of the three analysis channels individually as a function of $M_{\tilde{t}}$. The p-values for the three channel combination fit are shown as the black curve. The bottom panel shows the ratio of the signal production cross section predicted and measured which is equivalent to the best fit value for r.

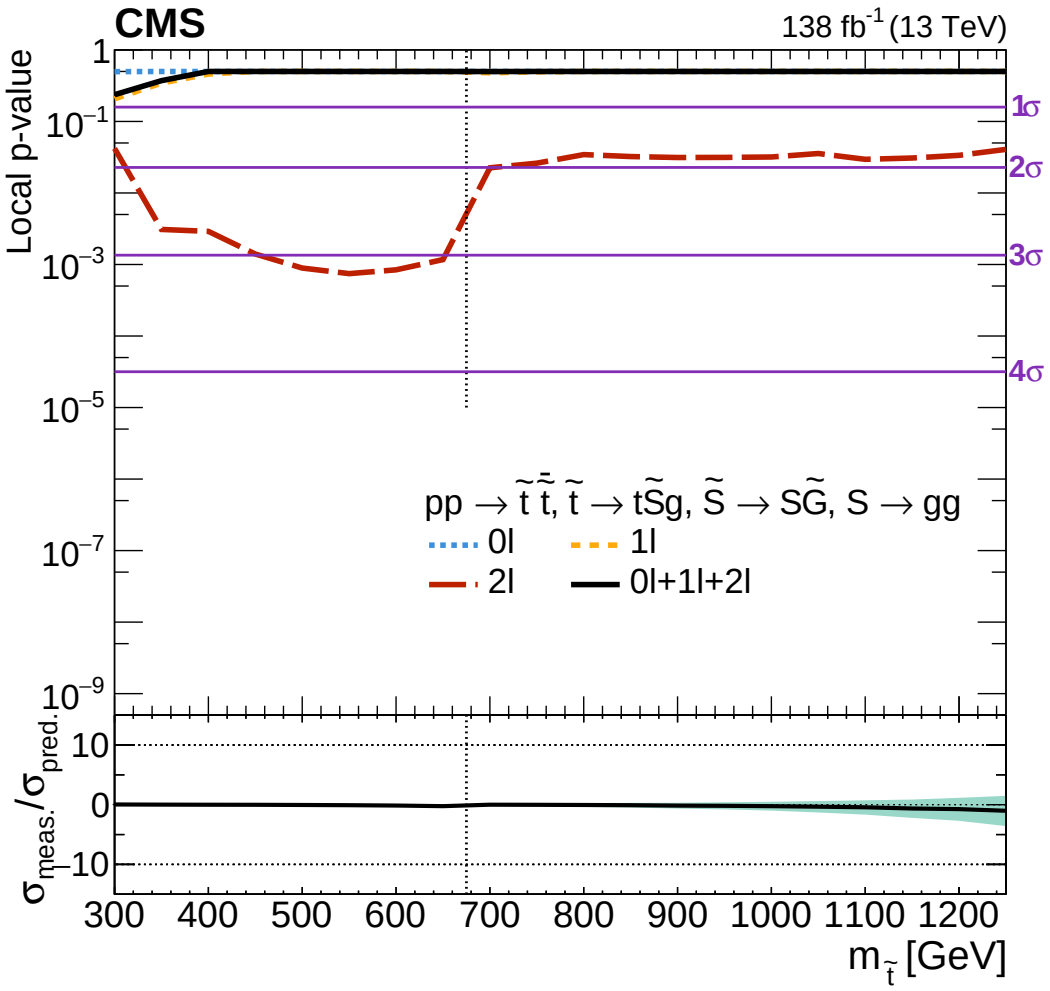

Figure 8.4: P-value distributions for the Stealth $SY\bar{Y}$ model which outline the statistical significance of any excesses seen in data. Distributions are shown for fits to each of the three analysis channels individually as a function of $M_{\tilde{t}}$. The p-values for the three channel combination fit are shown as the black curve. The bottom panel shows the ratio of the signal production cross section predicted and measured which is equivalent to the best fit value for r.

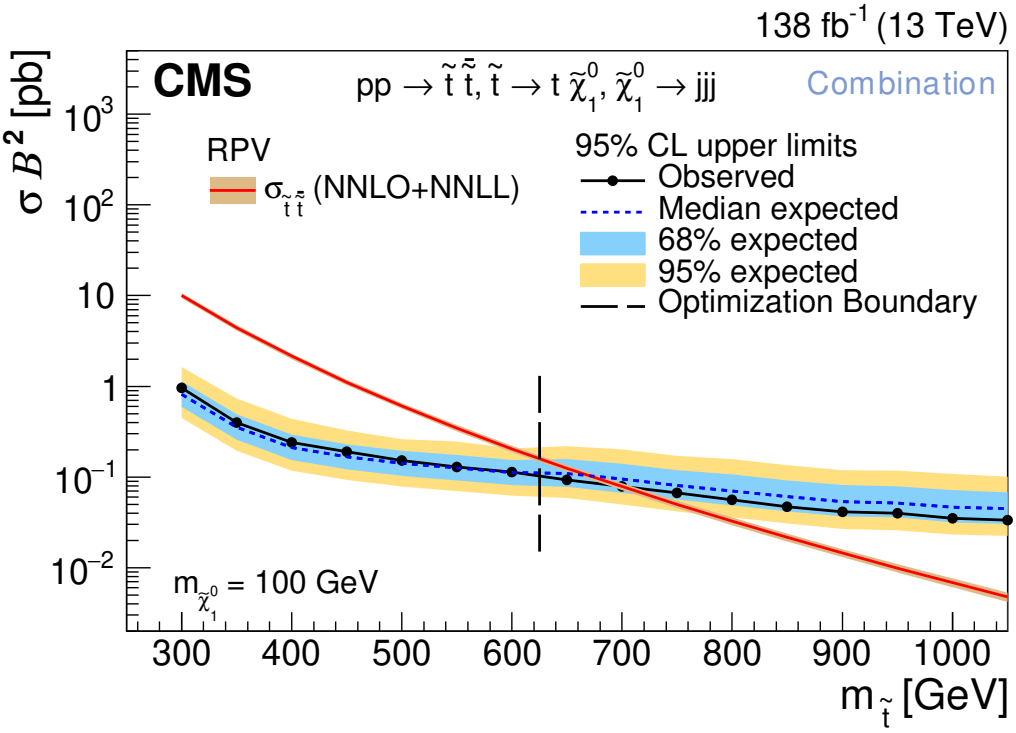

Figure 8.5: Upper limits on the production cross section for top squarks decaying via the RPV model are shown as a function of $M_{\tilde{t}}$. Blue and yellow bands represent the 68% and 95% confidence intervals on the upper limit, respectively. Note that the limits shown are computed using the three channel combination fit. The theoretical cross section (assuming 100% branching fraction for RPV decays) is shown in red with associated uncertainty tan.

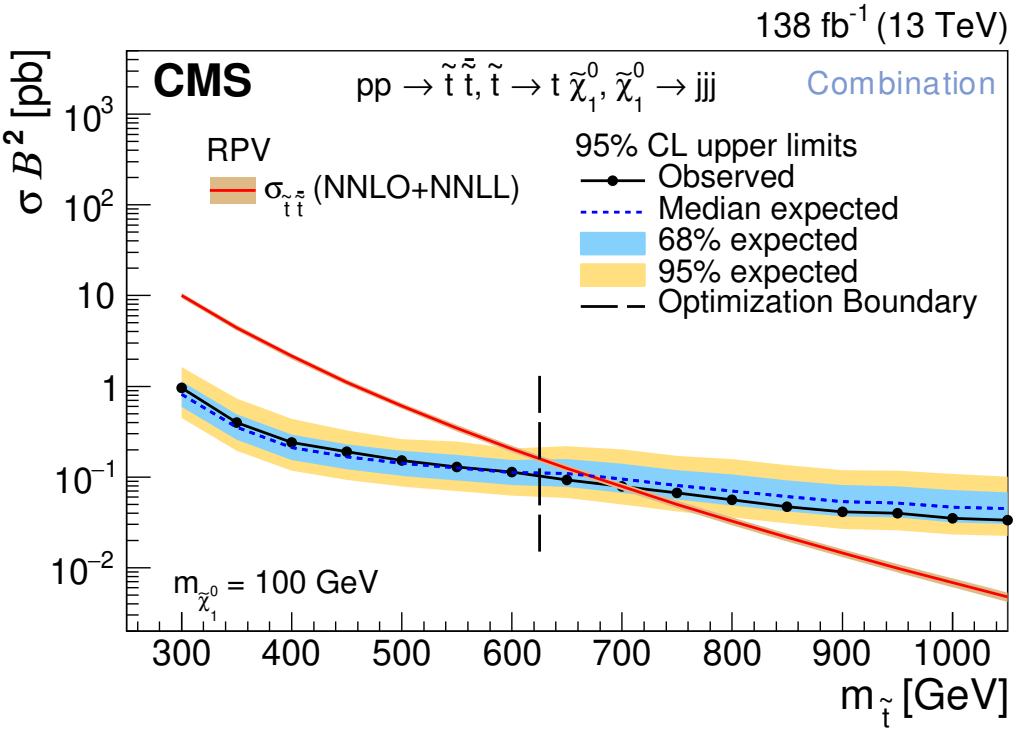

Figure 8.6: Upper limits on the production cross section for top squarks decaying via the Stealth $SY\bar{Y}$ model are shown as a function of $M_{\tilde{t}}$. Blue and yellow bands represent the 68% and 95% confidence intervals on the upper limit, respectively. Note that the limits shown are computed using the three channel combination fit. The theoretical cross section (assuming 100% branching fraction for Stealth $SY\bar{Y}$ decays) is shown in red with associated uncertainty tan.

Chapter 9

Conclusion and Discussion

An analysis searching for top squarks decaying via the Stealth $SY\bar{Y}$ and RPV models has been presented. The two signal models of interest have final states with a large number of light flavored jets which are produced in tandem with a pair of top quarks. A neural-network-based approach to the ABCD method is used to differentiate possible signal events from the SM process $t\bar{t}+jets$, which can mimic the signal final state via initial and final state radiation. Six networks are trained to identify possible signal events in three search channels defined by the expected decay modes of top quarks—either zero lepton, one lepton, or two leptons in the final state. The models are validated using both simulated particle collision events as well as data in validation regions orthogonal to the final signal region. The main sources of systematic uncertainty are associated with the assumption that the neural networks are able to produce two independent discriminants which can identify signal events from SM backgrounds. These uncertainties are estimated using the data validation regions based on the magnitude of nonclosure in the ABCD plane. The final results of the analysis are computed using a log-likelihood fit to assess whether there is an excess of events above the SM expected background which is consistent with RPV or Stealth $SY\bar{Y}$ decays of pair produced top squarks.

There is not sufficient evidence to claim discovery of either of these supersymmetric decay models in LHC collision data. There is evidence of an excess in the 2ℓ channels for both models with a magnitude ranging from 2-3 standard deviations above the expected background; however, this is most likely due to statistical fluctuations present in this search channel. Future searches could be designed to probe this phase space in a more

dedicated fashion to confirm this hypothesis. Limits are placed on $M_{\tilde{t}}$ which indicate which areas of model phase space have been excluded. Specifically, these limits are $M_{\tilde{t}} \geq 700$ GeV for the RPV model and $M_{\tilde{t}} \geq 930$ GeV for the Stealth $SY\bar{Y}$ model. These results will guide future analyzers to develop analyses which probe untouched areas of supersymmetric phase space.

Additionally, this analysis probed a previously observed excess at $M_{\tilde{t}} = 400$ GeV for the RPV model. While this excess cannot be confirmed by the results of this analysis, this analysis has improved the reach of this search by increasing signal sensitivity to the lowest top squark mass signals by a factor of ~ 2. This is in part due to the use of a data-driven background estimation using the ABCDisCoTEC method. By removing reliance on simulated modeling of extra jets in $t\bar{t} + jets$ events, this version of the analysis is sensitive to a different and less impactful set of systematic uncertainties. This analysis has additionally extended the previous search into the fully-hadronic and fully-leptonic decay channels, a task which requires careful analysis construction due to the high presence of QCD multijet events as well as low expected signal region event yield, respectively. Using the ABCDisCoTEC method, these regions have been explored despite the difficulties that they present.

In addition to the physics results from this analysis, a novel neural-network-based technique for performing the ABCD method has been developed. The ABCDisCoTEC model can be generalized for other applications in high energy physics for data-driven background estimation. This method shows promise in establishing a bias-free estimation of backgrounds while utilizing the prediction power of any number of differentiating analysis variables. This is specifically useful for analyses which require multivariate approaches for difficult to identify signals or which are hindered by poorly modeled simulated events.

References

[1] CMS Collaboration. Search for top squarks in final states with two top quarks and several light-flavor jets in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. D*, 104(3):032006, 2021, 2102.06976.

[2] 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*, 716:1–29, 2012, 1207.7214.

[3] CMS Collaboration. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012, 1207.7235.

[4] Ewa Lopienska. The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022. 2022. General Photo.

[5] Michael Benedikt, Paul Collier, V Mertens, John Poole, and Karlheinz Schindl. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.

[6] B.J. Holzer. Introduction to particle accelerators and their limitations. *CERN Yellow Reports*, pages Vol 1 (2016): Proceedings of the 2014 CAS–CERN Accelerator School: Plasma Wake Acceleration, 2016.

[7] Werner Herr and B Muratori. Concept of luminosity. 2006.

[8] CMS Collaboration. *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2020.

[9] Markus Zerlauth and Oliver Bruning. Status and prospects of the hl-lhc project. page 615, 01 2024.

[10] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical design report. CMS. CERN, Geneva, 2006.

[11] Giordon Stark. *The Large Hadron Collider and the ATLAS Detector*, pages 27–46. Springer International Publishing, Cham, 2020.

[12] Izaak Neutelings. Cms coordinate system, 2017.

[13] CMS Collaboration. The CMS tracker system project: Technical Design Report. 1997.

[14] CMS Collaboration. Development of the CMS detector for the CERN LHC Run 3. *JINST*, 19(05):P05064, 2024, 2309.05466.

[15] Christian Wolfgang Fabjan and F Gianotti. Calorimetry for Particle Physics. *Rev. Mod. Phys.*, 75:1243–1286, 2003.

[16] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997.

[17] Prasanna Kumar Siddireddy. The cms ecal trigger and daq system: electronics auto-recovery and monitoring. *arXiv: Instrumentation and Detectors*, 2018.

[18] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997.

[19] CMS Collaboration. Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13$ TeV. *JINST*, 15(05):P05002, 2020, 1910.00079.

[20] Particle Data Group, P A Zyla, et al. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8):083C01, 08 2020, https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf.

[21] CMS Collaboration. Overview of the CMS Detector Performance at LHC Run 2. *Universe*, 5(1):18, 2019.

[22] CMS Collaboration. The Phase-2 Upgrade of the CMS Muon Detectors. Technical report, CERN, Geneva, 2017. This is the final version, approved by the LHCC.

[23] V. Brigljevic, G. Bruno, E. Cano, S. Cittolin, A. Csilling, D. Gigi, F. Glege, R. Gomez-Reino, M. Gulmini, J. Gutleber, C. Jacobs, M. Kozlovszky, H. Larsen, I. Magrans de Abril, F. Meijers, E. Meschi, S. Murray, A. Oh, L. Orsini, L. Pollet, A. Racz, D. Samyn, P. Scharff-Hansen, C. Schwick, P. Sphicas, V. O'Dell, I. Suzuki, L. Berti, G. Maron, N. Toniolo, L. Zangrando, A. Ninane, S. Erhan, S. Bhattacharya, and J. Branson. The CMS Event Builder. COMPACT MUON SOLENOID. *eConf C*, 0303241:WEPT003, 2003. Conference CHEP03 Subj-class: Instrumentation and Detectors.

[24] V. Khachatryan et al. The cms trigger system. *Journal of Instrumentation*, 12(01):P01020, jan 2017.

[25] CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017, 1706.04965.

[26] Wikimedia Commons. Standard model of elementary particles, 2024. [Online; accessed 23-July-2024].

[27] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.

[28] Scott Willenbrock. Symmetries of the standard model. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics in $D \geqq 4$*, pages 3–38, 10 2004, hep-ph/0410370.

[29] Jeff Greensite. *An introduction to the confinement problem*, volume 821. 2011.

[30] Wikimedia Commons. File:quark confinement.svg, 2009. [Online; accessed 26-July-2024].

[31] Jack Lindon. Particle Collider Probes of Dark Energy, Dark Matter and Generic Beyond Standard Model Signatures in Events With an Energetic Jet and Large Missing Transverse Momentum Using the ATLAS Detector at the LHC, 2020. Presented 30 Oct 2020.

[32] Sheldon L. Glashow. The renormalizability of vector meson interactions. *Nucl. Phys.*, 10:107–117, 1959.

[33] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519:367–377, 1968.

[34] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.

[35] G. Arnison et al. Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider. *Phys. Lett. B*, 126:398–410, 1983.

[36] P. Bagnaia et al. Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p}p$ Collider. *Phys. Lett. B*, 129:130–140, 1983.

[37] Nicola Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531–533, 1963.

[38] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, 1964.

[39] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, 1964.

[40] John Ellis. Topics in Higgs Physics. *CERN Yellow Rep. School Proc.*, 2:1, 2018, 1702.05436.

[41] Gianfranco Bertone and Dan Hooper. History of dark matter. *Rev. Mod. Phys.*, 90(4):045002, 2018, 1605.04909.

[42] Stephen P. Martin. A Supersymmetry primer. *Adv. Ser. Direct. High Energy Phys.*, 18:1–98, 1998, hep-ph/9709356.

[43] Izaak Neutelings. Standard model, Mar 2024.

[44] R. Barbier, C. Bérat, M. Besançon, M. Chemtob, A. Deandrea, E. Dudas, P. Fayet, S. Lavignac, G. Moreau, E. Perez, and Y. Sirois. R-parity-violating supersymmetry. *Physics Reports*, 420(1):1–195, 2005.

[45] JiJi Fan, Rebecca Krall, David Pinner, Matthew Reece, and Joshua T. Ruderman. Stealth supersymmetry simplified. *Journal of High Energy Physics*, 2016(7), July 2016.

[46] JiJi Fan, Matthew Reece, and Joshua T. Ruderman. Stealth Supersymmetry. *JHEP*, 11:012, 2011, 1105.5135.

[47] Ansgar Denner, Stefan Dittmaier, Stefan Kallweit, and Stefano Pozzorini. NLO QCD corrections to off-shell ttbar production at hadron colliders. *PoS*, LL2012:015, 2012, 1208.4053.

[48] CMS Collaboration. CMS Luminosity - Public Results, July 2024.

[49] Simulation of the Silicon Strip Tracker pre-amplifier in early 2016 data. 2020.

[50] Paolo Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004, hep-ph/0409146.

[51] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP*, 11:070, 2007, 0709.2092.

[52] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010, 1002.2581.

[53] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction. *JHEP*, 09:126, 2007, 0707.3088.

[54] Emanuele Re. Single-top Wt-channel production matched with parton showers using the POWHEG method. *Eur. Phys. J. C*, 71:1547, 2011, 1009.2450.

[55] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO single-top production matched with shower in POWHEG: s- and t-channel contributions. *JHEP*, 09:111, 2009, 0907.4076.

[56] 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*, 07:079, 2014, 1405.0301.

[57] Johan Alwall et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J. C*, 53:473–500, 2008, 0706.2569.

[58] Rikkert Frederix and Stefano Frixione. Merging meets matching in MC@NLO. *JHEP*, 12:061, 2012, 1209.6215.

[59] Pierre Artoisenet, Rikkert Frederix, Olivier Mattelaer, and Robbert Rietkerk. Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations. *JHEP*, 03:015, 2013, 1212.3460.

[60] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J. C*, 77:663, 2017, 1706.00428.

[61] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159, 2015, 1410.3012.

[62] S. Agostinelli et al. GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003.

[63] Christoph Borschensky, Michael Krämer, Anna Kulesza, Michelangelo Mangano, Sanjay Padhi, Tilman Plehn, and Xavier Portell. Squark and gluino production cross sections in pp collisions at $\sqrt{s} = 13$, 14, 33 and 100 TeV. *Eur. Phys. J. C*, 74:3174, 2014, 1407.5066.

[64] Wim Beenakker, Christoph Borschensky, Michael Krämer, Anna Kulesza, and Eric Laenen. NNLL-fast: predictions for coloured supersymmetric particle production at the LHC with threshold and Coulomb resummation. *JHEP*, 12:133, 2016, 1607.07741.

[65] Michal Czakon and Alexander Mitov. Top++: A program for the calculation of the top-pair cross-section at hadron colliders. *Comput. Phys. Commun.*, 185:2930, 2014, 1112.5675.

[66] P. Kant, O. M. Kind, T. Kintscher, T. Lohse, T. Martini, S. Mölbitz, P. Rieck, and P. Uwer. HATHOR for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions. *Comput. Phys. Commun.*, 191:74, 2015, 1406.4403.

[67] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, and M. Wiedermann. HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR. *Comput. Phys. Commun.*, 182:1034, 2011, 1007.1327.

[68] T. Gehrmann, M. Grazzini, S. Kallweit, P. Maierhöfer, A. von Manteuffel, S. Pozzorini, D. Rathlev, and L. Tancredi. W^+W^- production at hadron colliders in next to next to leading order QCD. *Phys. Rev. Lett.*, 113:212001, 2014, 1408.5243.

[69] John M. Campbell and R. Keith Ellis. An update on vector boson pair production at hadron colliders. *Phys. Rev. D*, 60:113006, 1999, hep-ph/9905386.

[70] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Vector boson pair production at the LHC. *JHEP*, 07:018, 2011, 1105.0020.

[71] Ye Li and Frank Petriello. Combining QCD and electroweak corrections to dilepton production in FEWZ. *Phys. Rev. D*, 86:094034, 2012, 1208.5967.

[72] https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedElectronIdentificationRun2.

[73] https://twiki.cern.ch/twiki/bin/viewauth/CMS/SUSLeptonSF#Electrons.

[74] https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2.

[75] https://twiki.cern.ch/twiki/bin/viewauth/CMS/SUSLeptonSF#Muons.

[76] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, apr 2008.

[77] CMS Collaboration. Jet algorithms performance in 13 TeV data. Technical report, CERN, Geneva, 2017.

[78] E. Bols, J. Kieseler, M. Verzetti, M. Stoye, and A. Stakia. Jet flavour classification using DeepJet. *Journal of Instrumentation*, 15(12):P12012–P12012, dec 2020.

[79] CMS Collaboration. Search for top squark production in fully hadronic final states in proton-proton collisions at $\sqrt{s} = 13$ tev. *Phys. Rev. D*, 104:052001, Sep 2021.

[80] Gregor Kasieczka, Benjamin Nachman, Matthew D. Schwartz, and David Shih. Automating the ABCD method with machine learning. *Phys. Rev. D*, 103(3):035021, 2021, 2007.14400.

[81] Martín Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[82] François Chollet et al. Keras. `https://keras.io`, 2015.

[83] Gábor J. Székely, Maria L. Rizzo, and Nail K. Bakirov. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6), December 2007.

[84] Terence Shin. Introducing distance correlation, a superior correlation metric., Feb 2021.

[85] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017, 1412.6980.

[86] Aram Hayrapetyan et al. The CMS statistical analysis and combination tool: COMBINE. Submitted to *Comput. Softw. Big Sci.*, 2024, 2404.06614.

[87] Wouter Verkerke and David P. Kirkby. The RooFit toolkit for data modeling. *eConf*, C0303241:MOLT007, 2003, physics/0306116.

[88] I. Antcheva et al. ROOT — a C++ framework for petabyte data storage, statistical analysis and visualization. *Comput. Phys. Commun.*, 180:2499, 2009, 1508.07749.

[89] F. James and M. Roos. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput. Phys. Commun.*, 10:343–367, 1975.

[90] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011, 1007.1727.

[91] Vardan Khachatryan et al. Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV. *Eur. Phys. J. C*, 75:212, 2015, 1412.8662.

[92] Luc Demortier. P values and nuisance parameters. In *Statistical issues for LHC physics. Proceedings, Workshop, PHYSTAT-LHC, Geneva, Switzerland, June 27-29, 2007*, page 23, 2008.

[93] Alexander L. Read. Presentation of search results: The CL_s technique. *J. Phys. G*, 28:2693, 2002.

[94] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth. A*, 434:435, 1999, hep-ex/9902006.

Appendix A

Plots for All Channels and Models

The following section includes plots for all channel (0ℓ, 1ℓ, and 2ℓ) and model (RPV and Stealth $SY\bar{Y}$) combinations. This appendix is meant to mirror chapter 7 in its content.

A.1 Mass Regression Performance

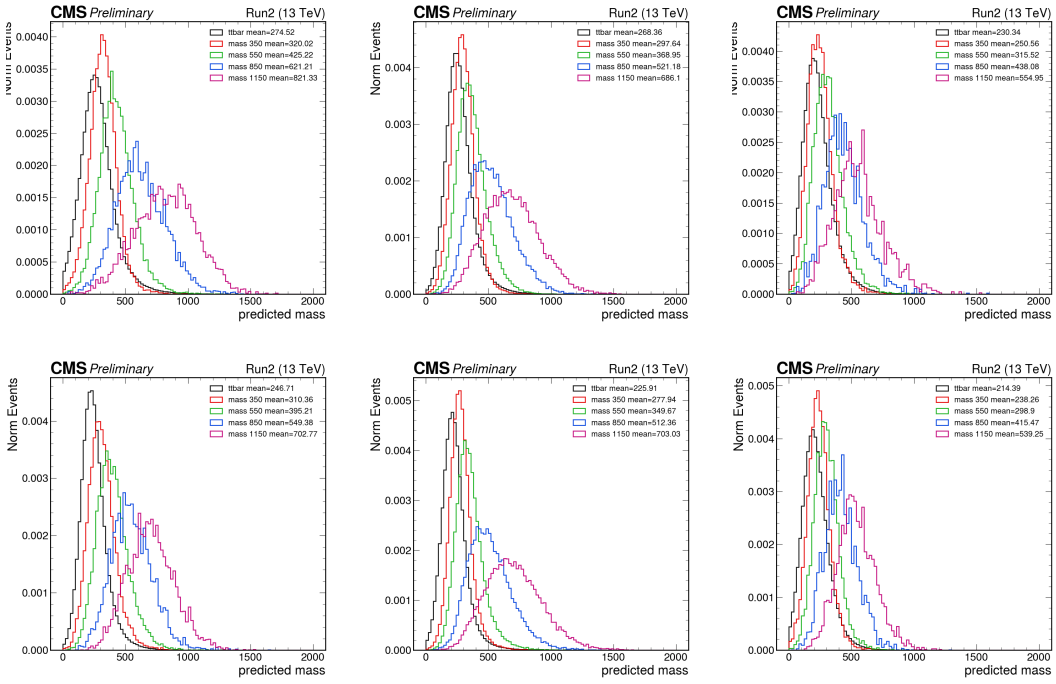

Figure A.1: Mass regression output for background and signal for both the RPV (top) and Stealth $SY\bar{Y}$ (bottom) trained neural networks. From left to right, distributions are shown for for the 0ℓ, 1ℓ, and 2ℓ channels.

A.2 Classification Performance

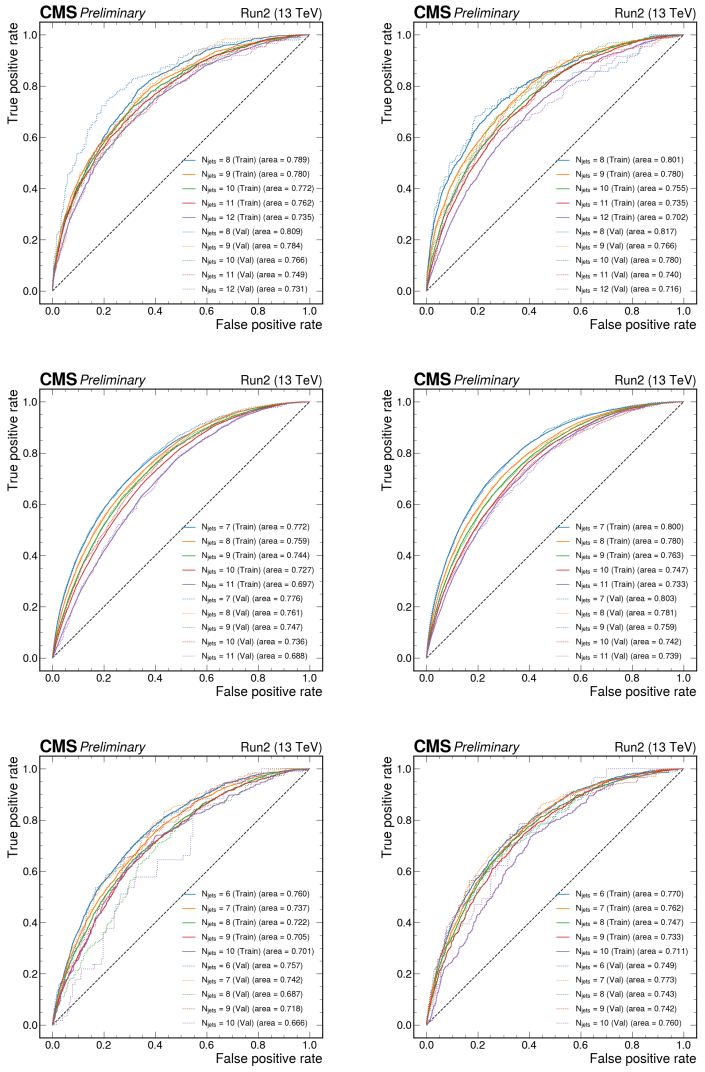

Figure A.2: For the $SY\bar{Y}$ signal model, ROCs separated by N_{Jets}. Disc. 1 (left) and Disc. 2 (right) are shown for the 0ℓ, 1ℓ, and 2ℓ channels (top to bottom).

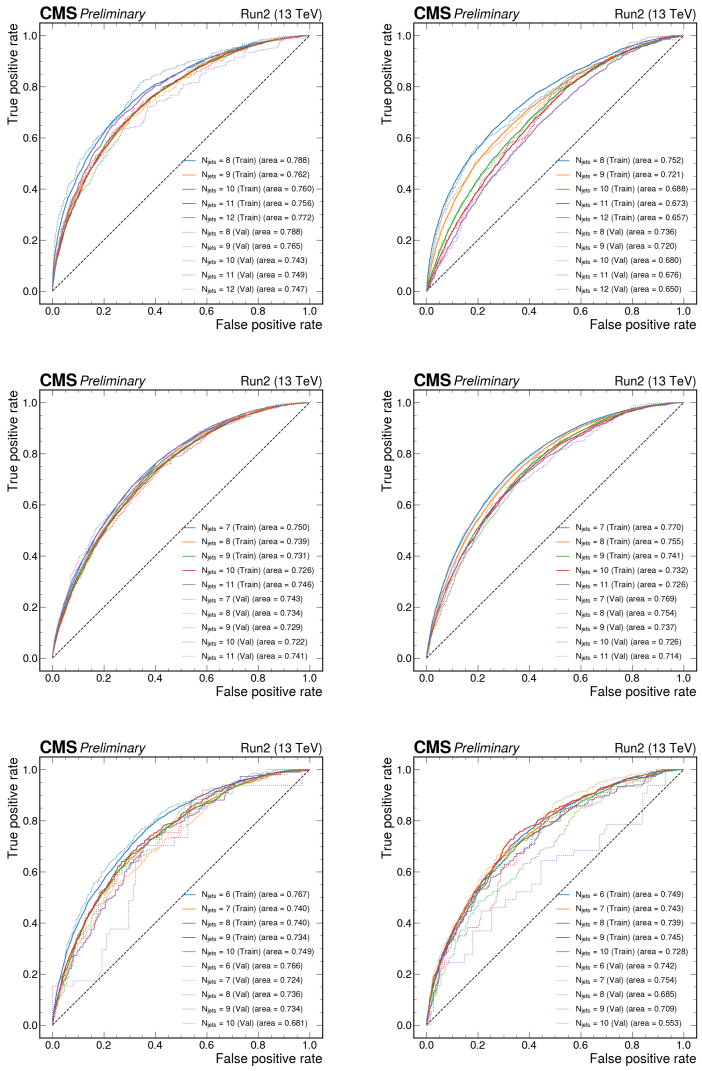

Figure A.3: For the RPV signal model, ROCs separated by N_{Jets}. Disc. 1 (left) and Disc. 2 (right) are shown for the 0ℓ, 1ℓ, and 2ℓ channels (top to bottom).

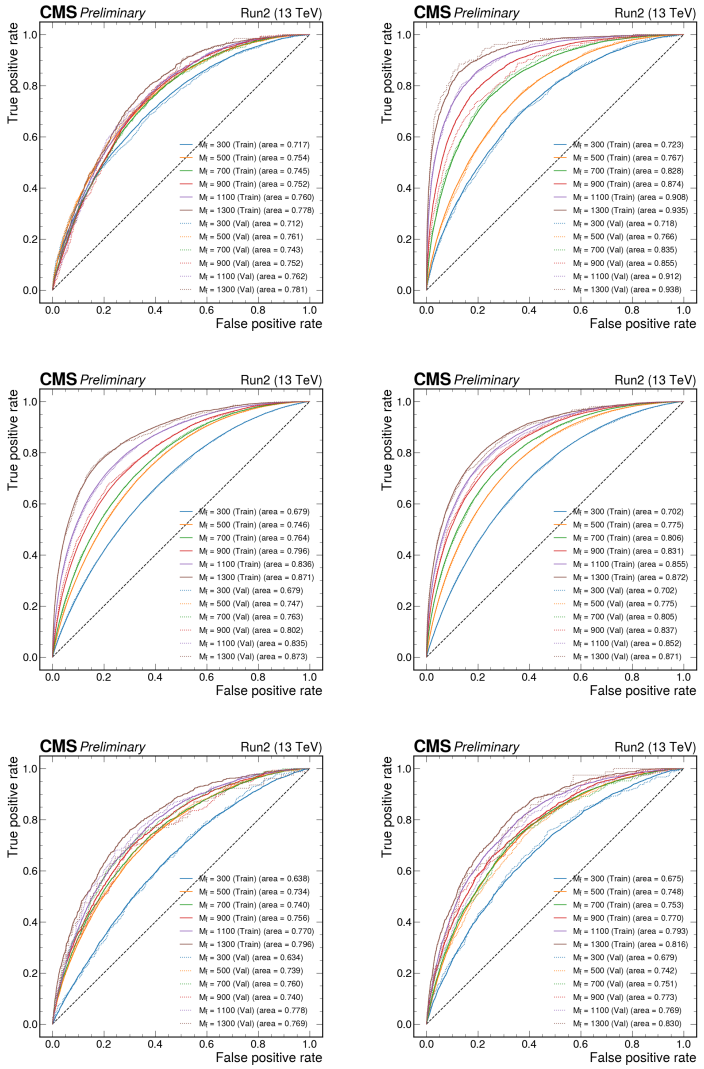

Figure A.4: For the $SY\bar{Y}$ signal model, ROCs separated by signal mass hypothesis. Disc. 1 (left) and Disc. 2 (right) are shown for the 0ℓ, 1ℓ, and 2ℓ channels (top to bottom).

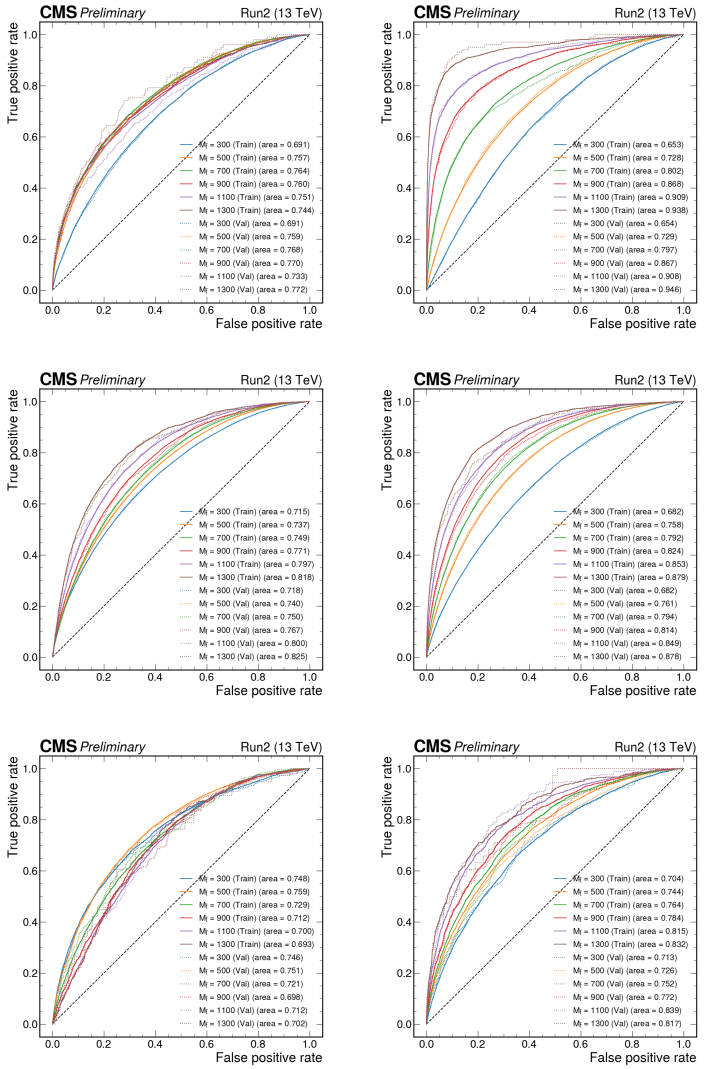

Figure A.5: For the RPV signal model, ROCs separated by signal mass hypothesis. Disc. 1 (left) and Disc. 2 (right) are shown for the 0ℓ, 1ℓ, and 2ℓ channels (top to bottom).

A.3 Background and Signal Distributions

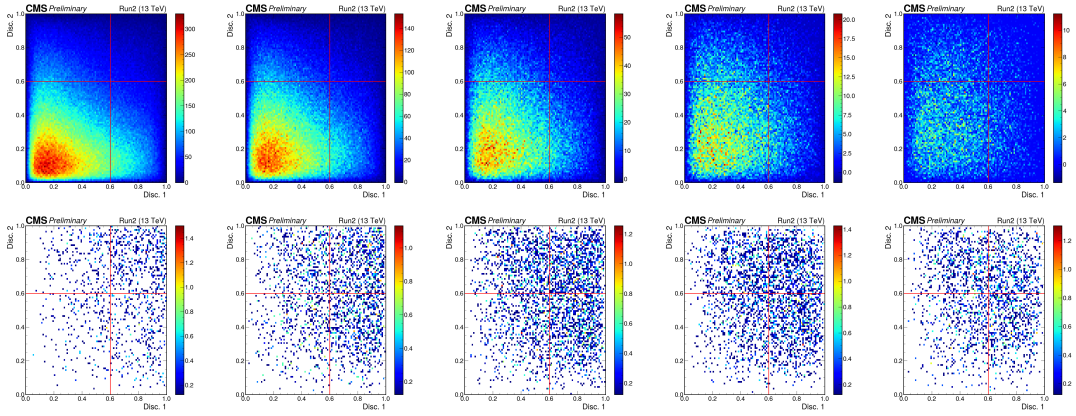

Figure A.6: For the 0ℓ channel, the 2D discriminant plane for $t\bar{t} + jets$ (top), Stealth $SY\bar{Y}$ $M_{\tilde{t}} = 550$ (bottom) separated in N_{Jets} from $N_{\text{Jets}} = 8$ (left) to $N_{\text{Jets}} = 12+$ (right).

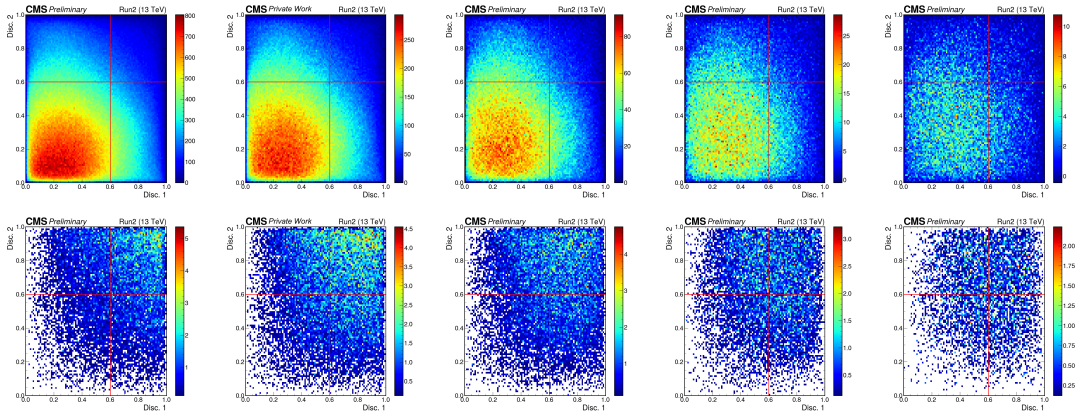

Figure A.7: For the 1ℓ channel, the 2D discriminant plane for $t\bar{t} + jets$ (top), Stealth $SY\bar{Y}$ $M_{\tilde{t}} = 550$ (bottom) separated in N_{Jets} from $N_{\text{Jets}} = 7$ (left) to $N_{\text{Jets}} = 11+$ (right).

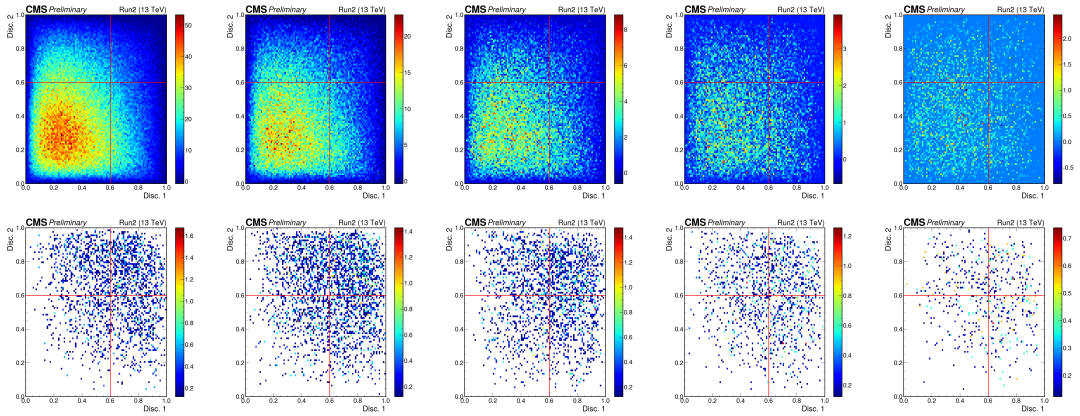

Figure A.8: For the 2ℓ channel, the 2D discriminant plane for $t\bar{t} + jets$ (top), Stealth $SY\bar{Y}$ $M_{\tilde{t}} = 550$ (bottom) separated in N_{Jets} from $N_{\text{Jets}} = 6$ (left) to $N_{\text{Jets}} = 10+$ (right).

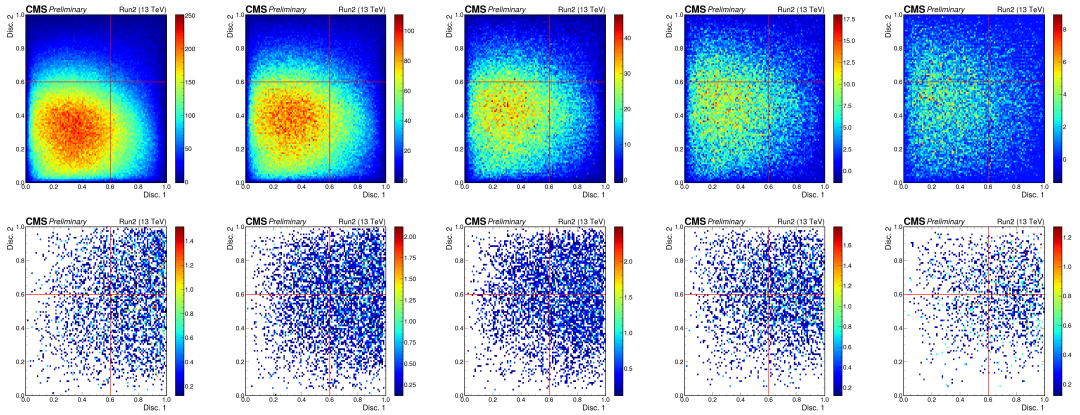

Figure A.9: For the 0ℓ channel, the 2D discriminant plane for $t\bar{t} + jets$ (top), RPV $M_{\tilde{t}} = 550$ (bottom) separated in N_{Jets} from $N_{\text{Jets}} = 8$ (left) to $N_{\text{Jets}} = 12+$ (right).

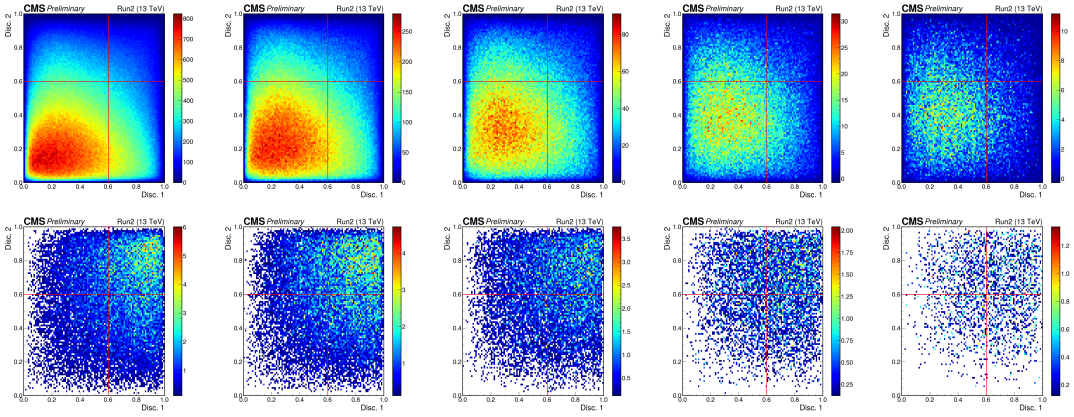

Figure A.10: For the 1ℓ channel, the 2D discriminant plane for $t\bar{t} + jets$ (top), RPV $M_{\tilde{t}} = 550$ (bottom) separated in N_{Jets} from $N_{\text{Jets}} = 7$ (left) to $N_{\text{Jets}} = 11+$ (right).

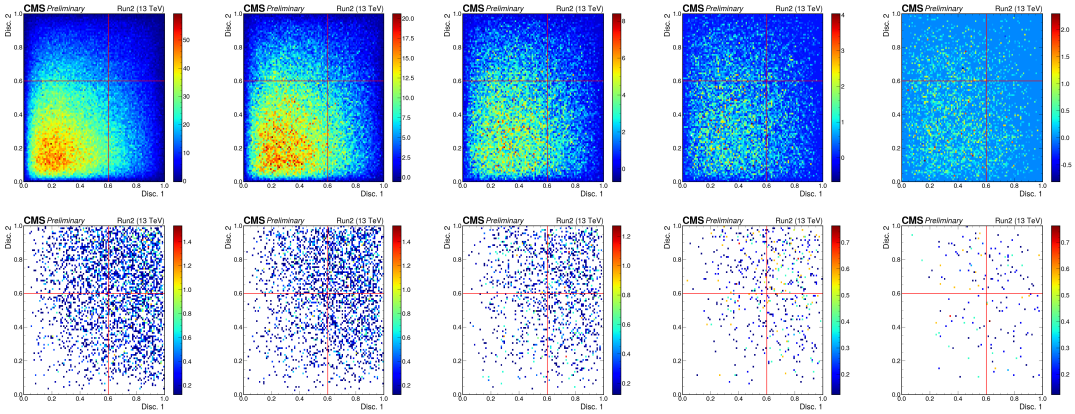

Figure A.11: For the 2ℓ channel, the 2D discriminant plane for $t\bar{t} + jets$ (top), RPV $M_{\tilde{t}} = 550$ (bottom) separated in N_{Jets} from $N_{\text{Jets}} = 6$ (left) to $N_{\text{Jets}} = 10+$ (right).

A.4 Optimization

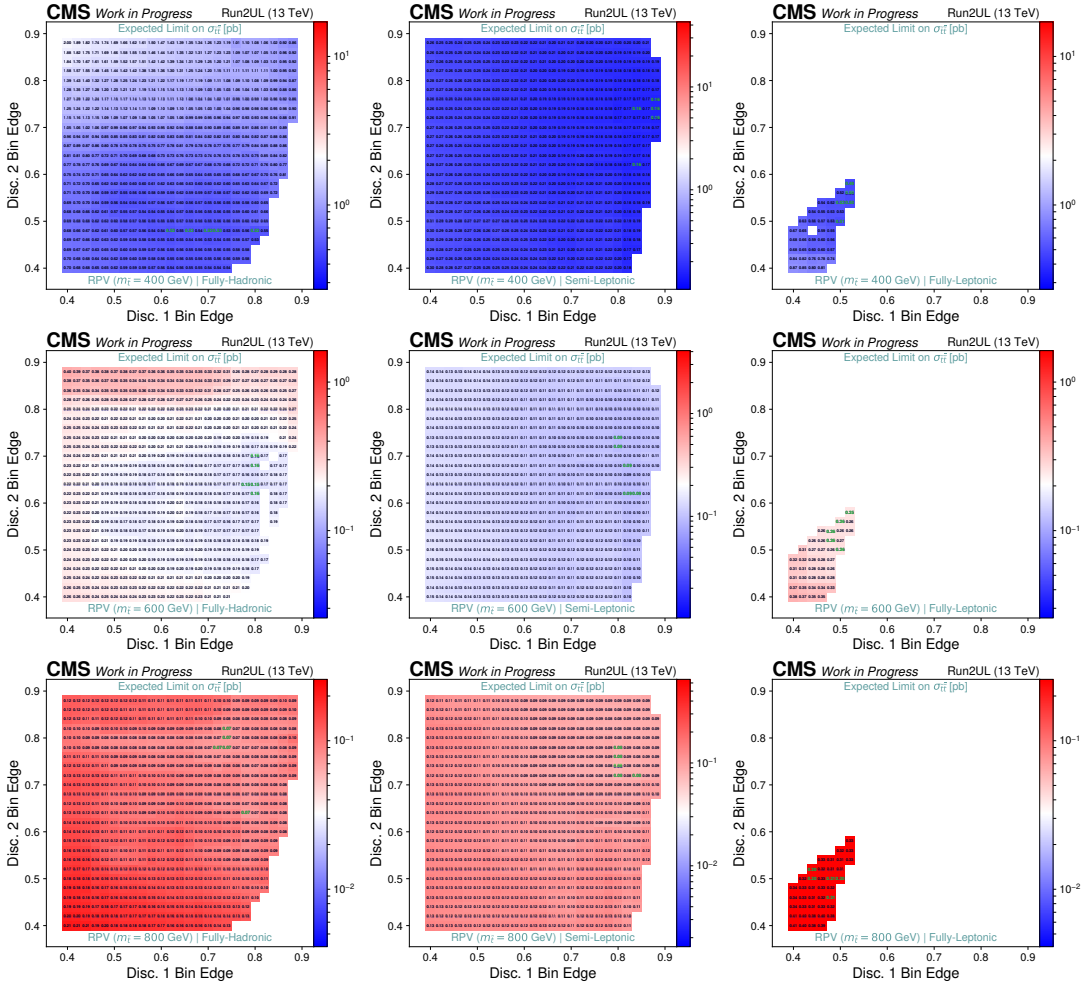

Figure A.12: **RPV:** For each ABCD bin edge choice, the expected limit using pseudo-data. Limit values are shown for RPV 400, 600, and 800 (top to bottom) and for the three channels: 0ℓ, 1ℓ, and 2ℓ (left to right). Blue values are expected limits where the signal model/mass hypothesis could be excluded, while red values are limits where the signal model/mass hypothesis could not be excluded. The top five choices are highlighted in green.

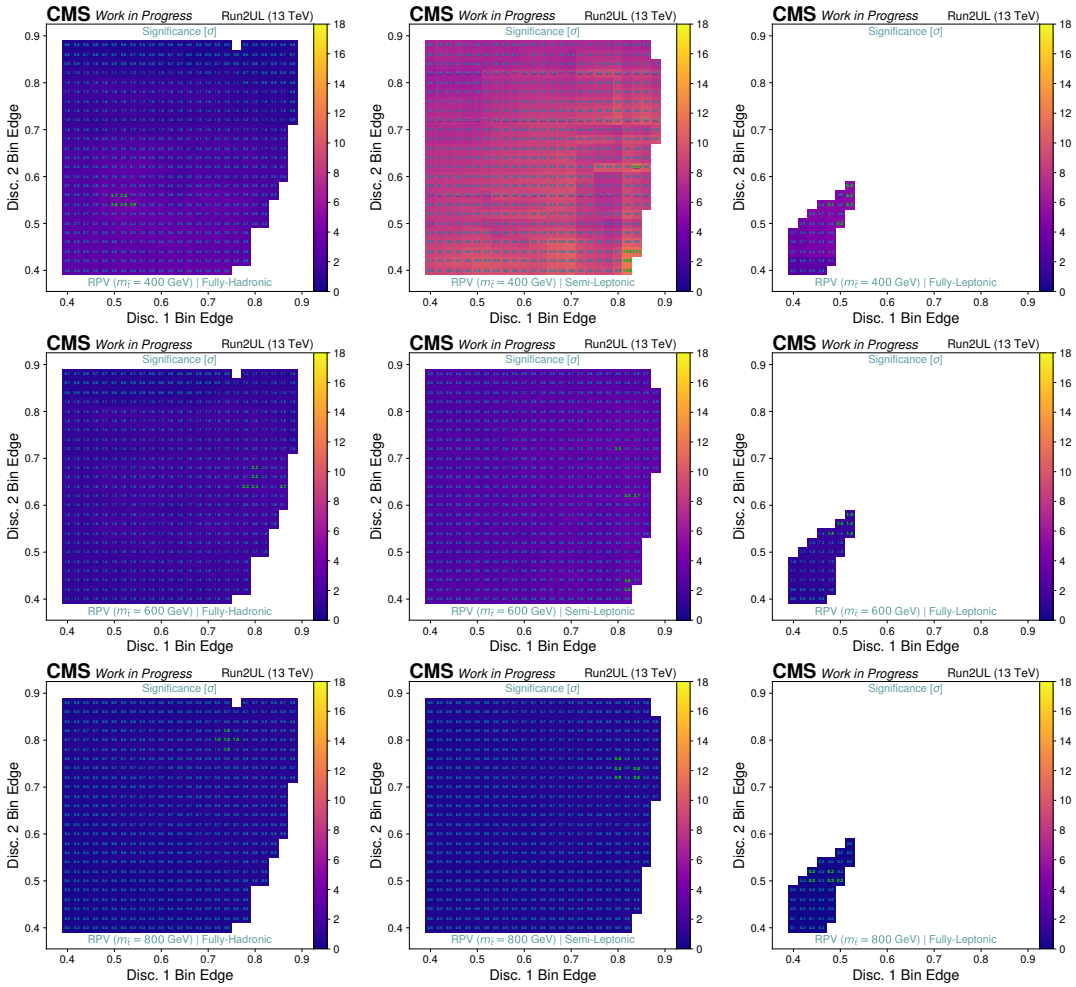

Figure A.13: **RPV:** For each ABCD bin edge choice, the expected significance using pseudo-data. Significance values are shown for RPV 400, 600, and 800 (top to bottom) and for the three channels: 0ℓ, 1ℓ, and 2ℓ (left to right). The top five choices are highlighted in green.

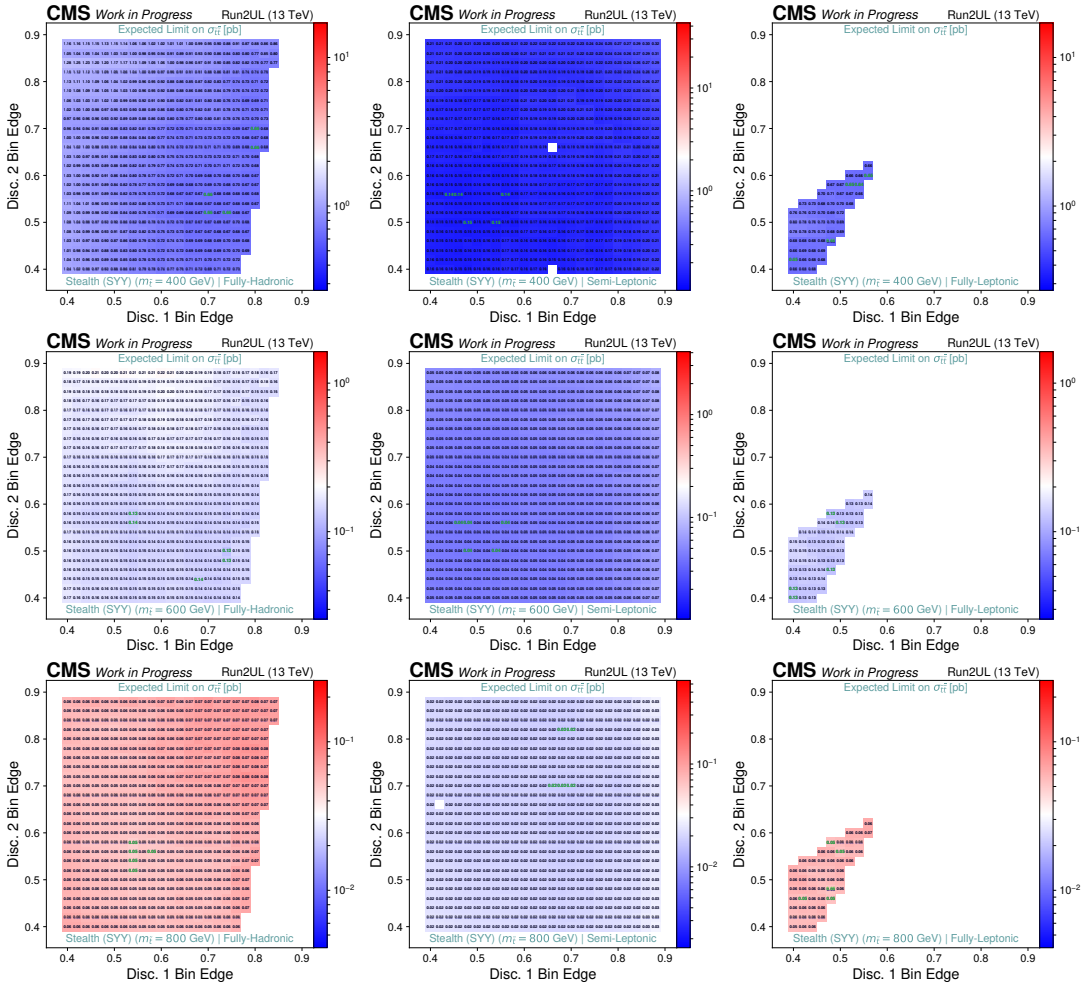

Figure A.14: **Stealth $SY\bar{Y}$:** For each ABCD bin edge choice, the expected limit using pseudo-data. Limit values are shown for Stealth $SY\bar{Y}$ 400, 600, and 800 (top to bottom) and for the three channels: 0ℓ, 1ℓ, and 2ℓ (left to right). Blue values are expected limits where the signal model/mass hypothesis could be excluded, while red values are limits where the signal model/mass hypothesis could not be excluded. The top five choices are highlighted in green. The top five choices are highlighted in green.

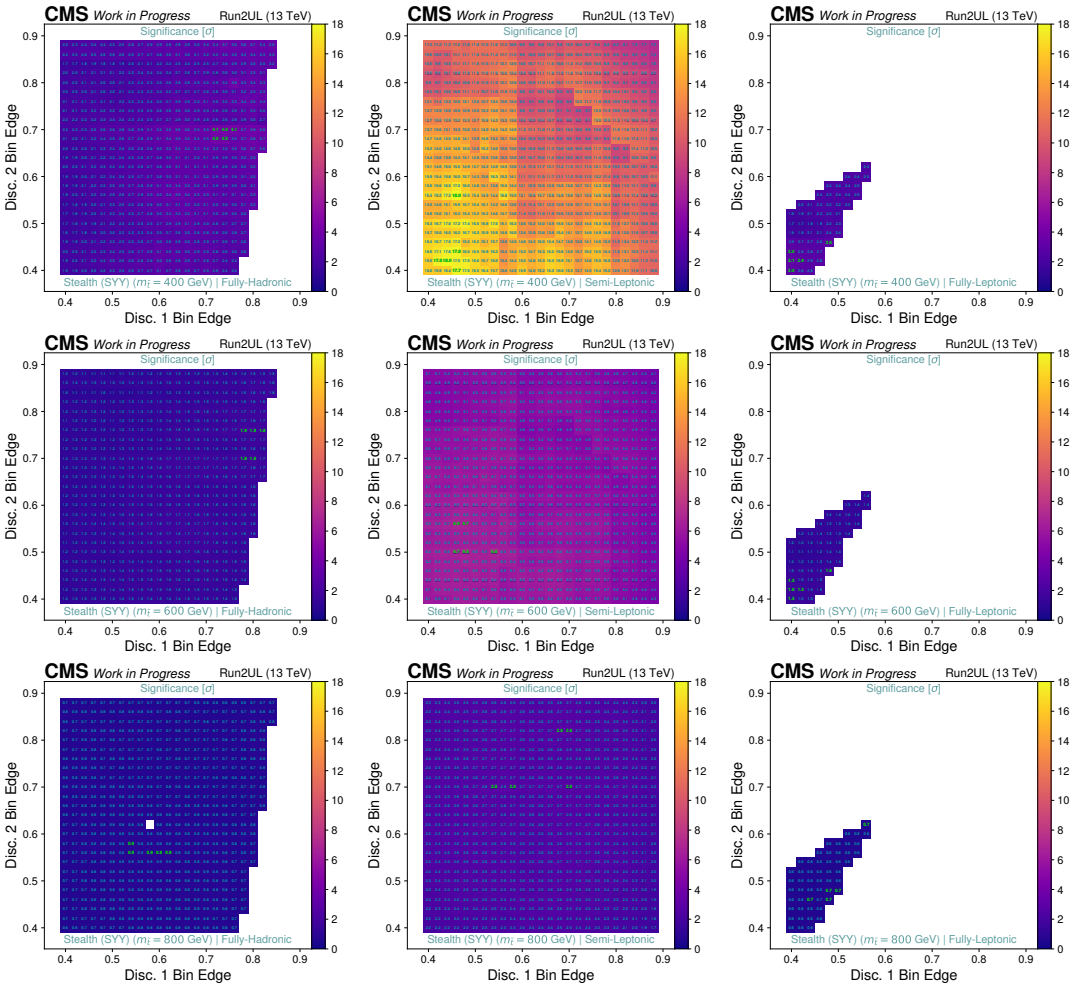

Figure A.15: **Stealth $SY\bar{Y}$:** For each ABCD bin edge choice, the expected significance using pseudo-data. Significance values are shown for Stealth $SY\bar{Y}$ 400, 600, and 800 (top to bottom) and for the three channels: 0ℓ, 1ℓ, and 2ℓ (left to right). The top five choices are highlighted in green.

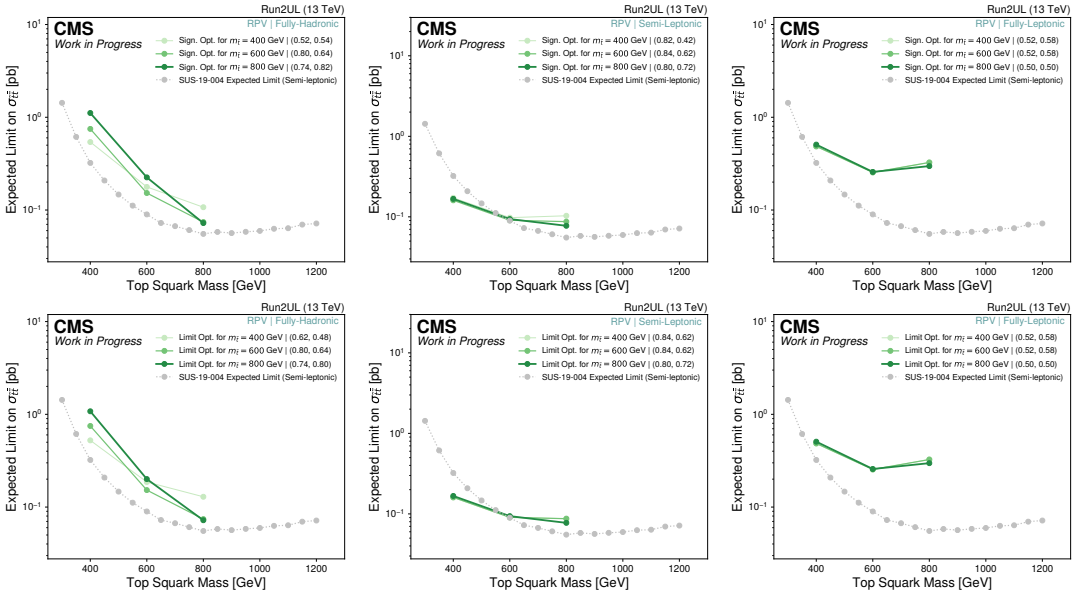

Figure A.16: **RPV:** Expected limits when choosing the ABCD bin edges based on either best signal significance (upper) or best expected limit (lower). Each shade of color represents the ABCD bin edges chosen based on a particular $M_{\tilde{t}}$. From right to left, the three separate analysis channels are shown (0ℓ, 1ℓ, 2ℓ, respectively).

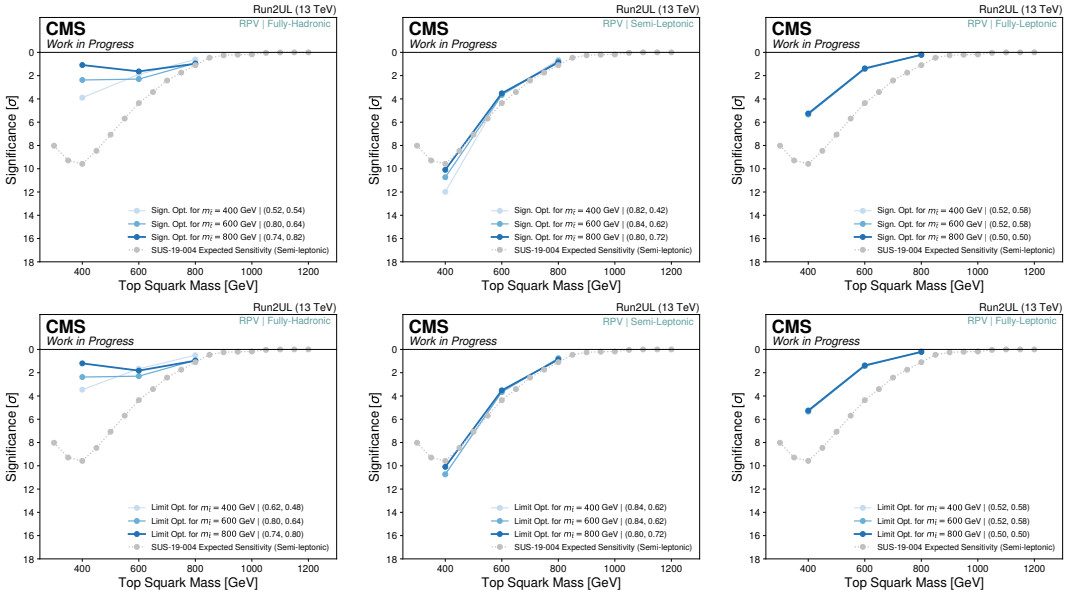

Figure A.17: **RPV:** Signal significance when choosing the ABCD bin edges based on either best signal significance (upper) or best expected limit (lower). Each shade of color represents the ABCD bin edges chosen based on a particular $M_{\tilde{t}}$. From right to left, the three separate analysis channels are shown (0ℓ, 1ℓ, 2ℓ, respectively).

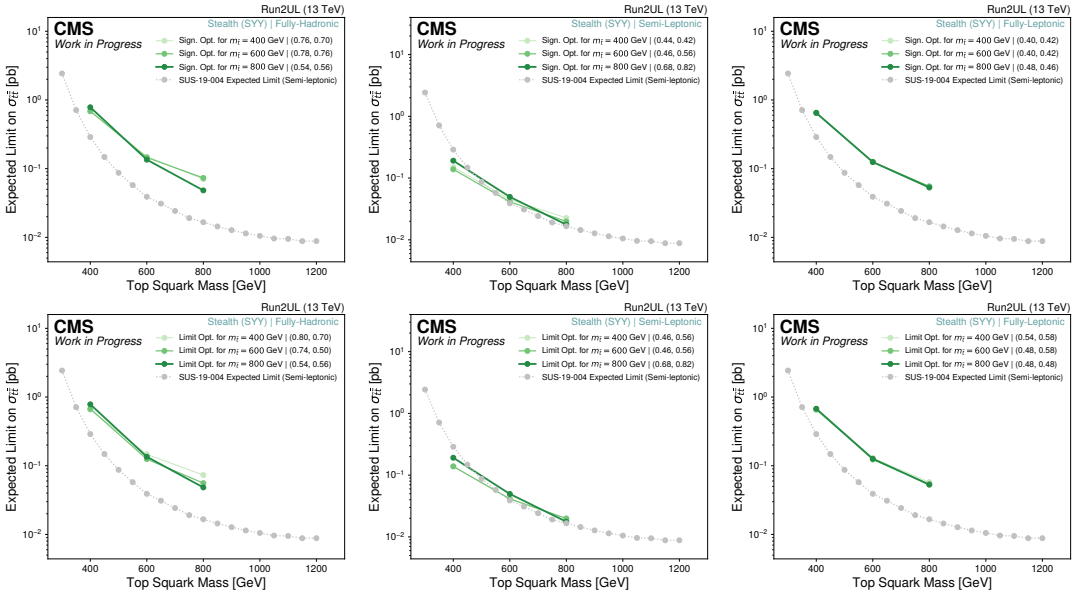

Figure A.18: **Stealth $SY\bar{Y}$:** Expected limits when choosing the ABCD bin edges based on either best signal significance (upper) or best expected limit (lower). Each shade of color represents the ABCD bin edges chosen based on a particular $M_{\tilde{t}}$. From right to left, the three separate analysis channels are shown (0ℓ, 1ℓ, 2ℓ, respectively).

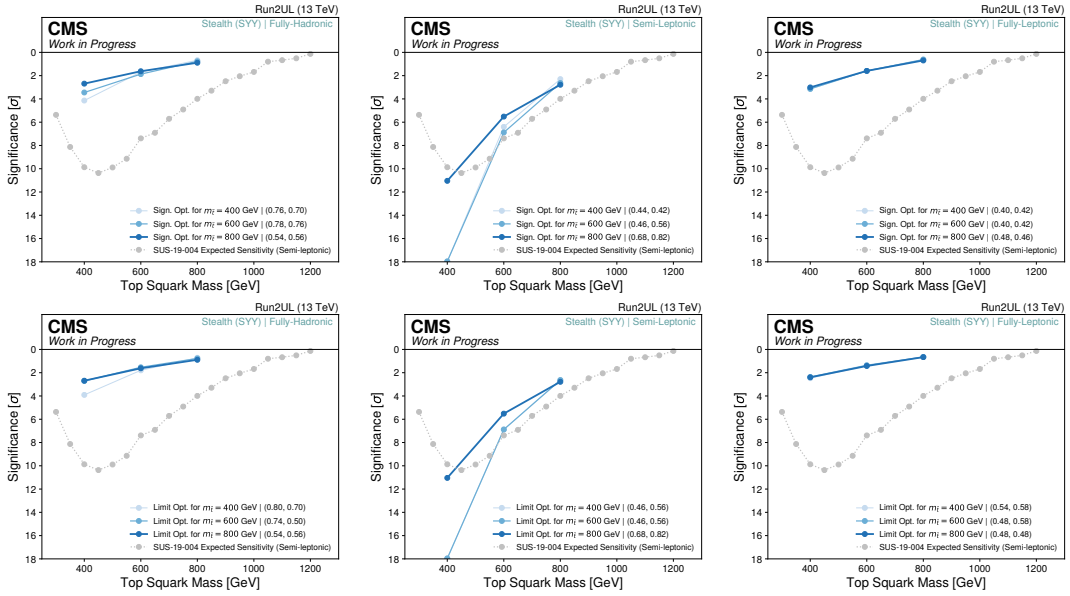

Figure A.19: **Stealth $SY\bar{Y}$:** Signal significance when choosing the ABCD bin edges based on either best signal significance (upper) or best expected limit (lower). Each shade of color represents the ABCD bin edges chosen based on a particular $M_{\tilde{t}}$. From right to left, the three separate analysis channels are shown (0ℓ,1ℓ, 2ℓ, respectively).

A.5 Data-Based Systematic Estimation

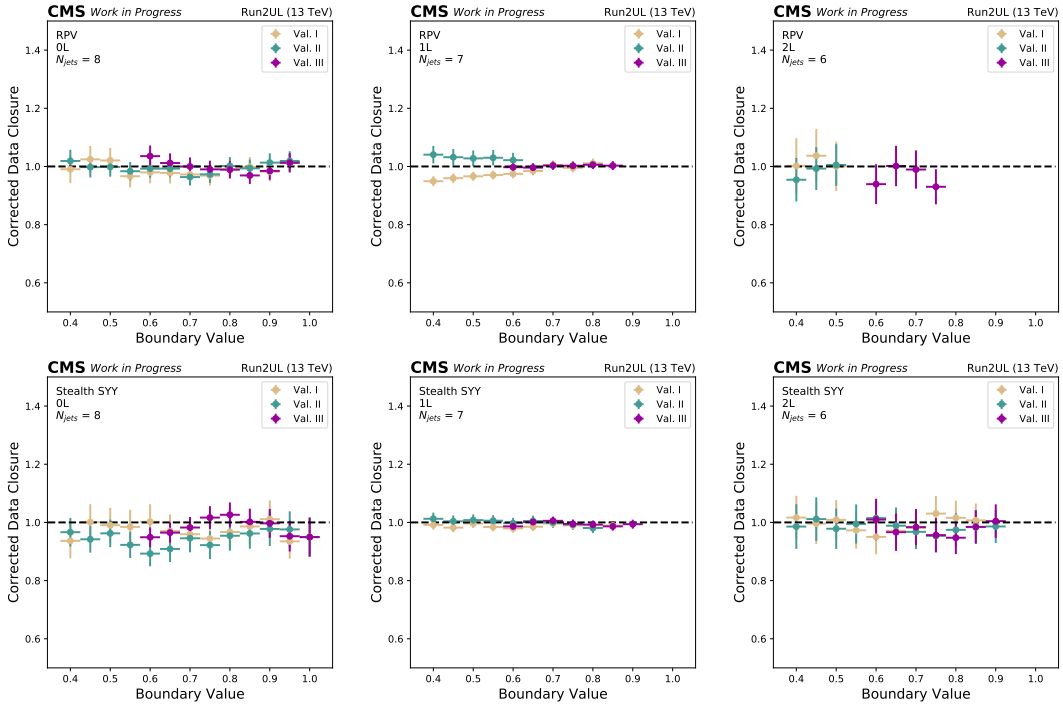

Figure A.20: Corrected data closure values are plotted for the low mass optimizations as a function of the boundary value defining the three validation regions for the three channels (left to right: 0ℓ, 1ℓ, and 2ℓ) and both signal models (top to bottom: RPV and Stealth $SY\bar{Y}$). The corrected data closure value is computed as in equation (7.2). The maximum value of corrected data closure for any of the three validation regions is used in computing the data-based systematic uncertainty per channel and model.

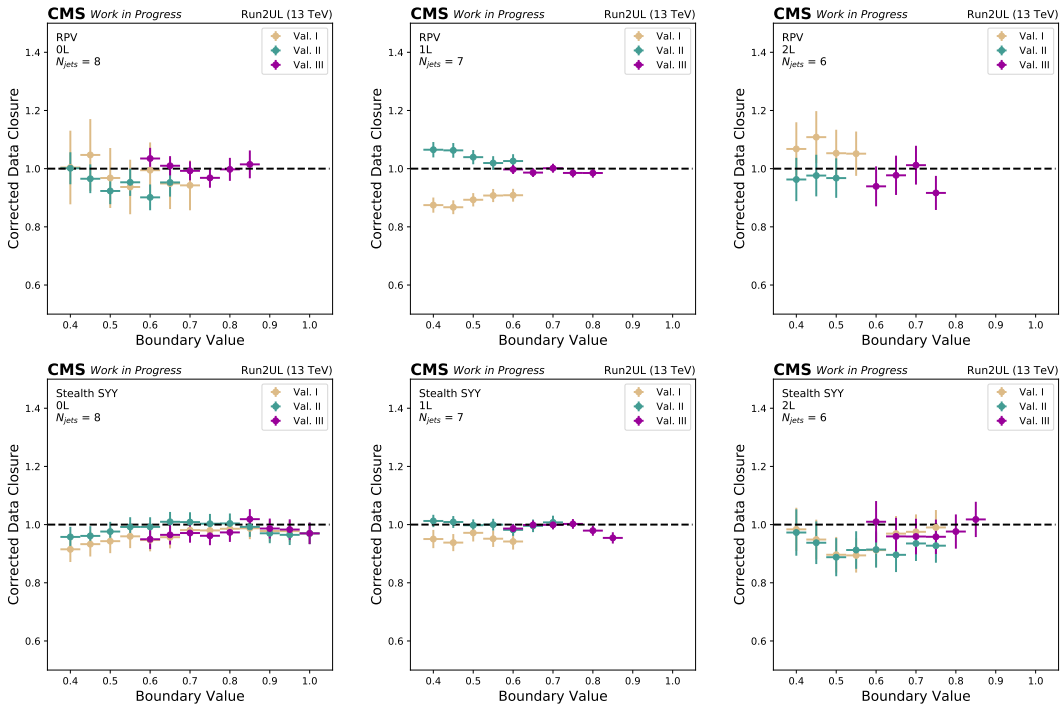

Figure A.21: Corrected data closure values are plotted for the high mass optimizations as a function of the boundary value defining the three validation regions for the three channels (left to right: 0ℓ, 1ℓ, and 2ℓ) and both signal models (top to bottom: RPV and Stealth $SY\bar{Y}$). The corrected data closure value is computed as in equation (7.2). The maximum value of corrected data closure for any of the three validation regions is used in computing the data-based systematic uncertainty per channel and model.