A Search for Scalar Top Quarks in Final States with High Jet Multiplicity and a Lepton with the CMS detector at the Large Hadron Collider

Kelvin Mei

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Physics
Advisor: Christopher Tully

September, 2020
© Copyright 2020 by Kelvin Mei.
All rights reserved.
Abstract

A search is conducted for scalar top quarks decaying in the Full Run II dataset for the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC) of 137.2 fb$^{-1}$ collected at a center of mass energy of $\sqrt{s} = 13$ TeV. The particular models being investigated involve pair production of scalar top quarks decaying either through an $R$-parity violating coupling or through a stealth sector into high jet multiplicity final states with low missing transverse energy. Novel machine-learning techniques have been applied, including gradient reversal, to extend sensitivity to a previously inaccessible and unexplored region of phase space. The highest observed local significance for the signal models investigated is 2.78 for the $R$-parity violating SUSY model where the scalar top quark mass is 400 GeV, with a local p value at $2.7 \times 10^{-3}$ and a signal strength of $0.21 \pm 0.07$ times the theoretical predicated cross section. Limits are set at the 95% confidence level for the $R$-parity violating models with the scalar top quark mass less than 675 GeV and for stealth SUSY models with the scalar top quark mass less than 975 GeV.
Acknowledgements

Anything is possible when you have the right people there to support you. - Misty Copeland

Completing a doctorate program has always been a difficult task, but given the times that the world and I personally have been going through, it has been more trying than I could have ever imagined when I began embarking on this journey six years ago. Through it all, however, many people have been my pillars of support and it is these people that I would like to thank in the space here.

First, I would like to express my deepest appreciation to my advisor Chris Tully. The roles he played in my doctorate career are too many to count, including advisor, teacher, mentor, co-worker, supporter, life coach, and more. His trust and confidence in my work exceeded even my own, and he provided me with the freedom, opportunity, and independence to reach for projects that not only challenged me, but also made me grow as a researcher and person. Our conversations always started with asking about each other’s well being and always ran until one of us had another appointment to attend already five minutes late. Between working with him on developing a computing course to early simulations of a new sub detector at CMS to hearing about his ingenuous geometries to solve problems for his PTOLEMY experiment, interactions with Chris were always interesting and learning experiences that expanded my physicist toolkit.

Next, I would have to thank both my analysis (StealthStop) and hardware (HCAL) teams that I have worked on in my collaboration, as the hours spent together made us grow into a family. For analysis, I want to specifically thank Jim Hirschauer, Owen Long, Aron Soha, and Nadja Strobbe for always being available to help me understand all the nuances
of a very complicated analysis, and it is not an understatement to say that there were
many. For hardware, I have to thank Pawel de Barbaro, Paolo Rumerio, Jim Hirschauer,
Dick Kellogg, Edward Laird, Joseph Pastika, Steven Goadhouse, and the late Danila Tsi
for their leadership, their high expectations, their trust, and their guidance as we worked
together to get a very ambitious upgrade completed successfully. In terms of mentorship
and guidance, I am grateful for the support and conversations with Princeton professors
Dan Marlow, Jim Olsen, Isobel Ojalvo, and Mariangela Lisanti, as well as John Paul Chou
and Sunil Somalwar, who were my undergraduate mentors at Rutgers University.

Up next, I have to thank my friends at CERN and Fermilab that made all-nighters
bearable, deadlines fun with coffee and laughter, and life in Geneva and Batavia an adven-
ture. There were my co-collaborator friends who worked in the trenches with me, including
Chris Madrid, Semra Turkcapur, Josh Hiltbrand, Caleb Smith, Ayse Bat, Vivan Nguyen,
and Laura Dodd. There were also the friends that made my doctorate years more enriching
and fun, but too many to name them all: the group affectionately known as the Ameri-
can Teenagers, the TRT BBQ crowd, the Thursday Wine and Cheese group, Lunch Bell
at Fermilab, Gates in Geneva, and the ENTP group. Then, there were the Princeton
students, including Lou Chen, Gillian Kopp, Stephanie Kwan, Justin Ripley, Sonia Zhang,
Brian Kraus, Matteo Ippoliti, Stevie Bergman, Jingyu Luo, and Stephane Cooperstein,
who made coming back to Princeton feel like coming back home.

Finally, I would like to thank the friends and family that stuck with me from the very
beginning of my doctorate career without being too bewildered that it took as long as it
did. The Hillbillies group (Paul Bergen, Devinn and Elias Boujaoude, Kaity Veenstra, Eric
Washkewicz, et al.), the Rutgers crew (Katherine Chang, Shruti Sriram, Nathan Mahani-
rananda, Sameen Jalal, Aaron Lin, et al.), HCRHS friends (Paige Meade, Phil Hallam,
et al.), and Chinese school friends (Karen, Fiona, and Wendy) and more, are all people I
wanted to thank immensely for their love and support throughout. And last, but not least,
my family: my dad Paul, my brothers Newton and Kent, and my aunt who is like my second
mother Bo You Fok, who were patient and acted as pillars of support and love throughout.
To my mom, Winnie
1957-2019
# Contents

Abstract iii

Acknowledgements iv

Contents vii

List of Figures xi

List of Tables xxvii

1 Introduction 1

  1.1 Fundamental Particles of the Standard Model . . . . . . . . . . . . . . . . . . 2
  1.2 Quantum Electrodynamics (QED) . . . . . . . . . . . . . . . . . . . . . . . . 7
  1.3 Quantum Chromodynamics . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
  1.4 Electroweak Theory . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
  1.5 Brout-Englert-Higgs Mechanism . . . . . . . . . . . . . . . . . . . . . . . . 22
  1.6 Current Status of Standard Model . . . . . . . . . . . . . . . . . . . . . . . 28

2 Beyond the Standard Model 33

  2.1 Issues with the Standard Model . . . . . . . . . . . . . . . . . . . . . . . . 33
    2.1.1 Gravity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
    2.1.2 Dark Energy and Dark Matter . . . . . . . . . . . . . . . . . . . . . . 34
    2.1.3 Matter-Antimatter Asymmetry . . . . . . . . . . . . . . . . . . . . . . 35
## 3 The Experiment 55

3.1 Large Hadron Collider (LHC) ............................. 55
  3.1.1 Proton-Proton Collider .............................. 56
  3.1.2 LHC Accelerator Complex ........................... 57

3.2 Compact Muon Solenoid (CMS) .......................... 61
  3.2.1 Tracker System ...................................... 64
  3.2.2 Electromagnetic Calorimeter ......................... 67
  3.2.3 Hadronic Calorimeter ............................... 69
  3.2.4 Magnet ................................................ 73
  3.2.5 Muon Systems ......................................... 74

3.3 Trigger ..................................................... 77
  3.3.1 L1 Trigger ............................................. 77
  3.3.2 HLT .................................................. 79
  3.3.3 Analysis Triggers ................................. 79

3.4 Event Simulation ......................................... 81

3.5 Simulation Samples ....................................... 83

## 4 Event Selection 86

4.1 Object Definition and Reconstruction .................. 86
  4.1.1 Electrons ........................................... 89
  4.1.2 Muons ............................................... 94
  4.1.3 Jets ............................................... 97
# Background

5.1 Correction of the simulation: scale factors

5.1.1 Lepton Scale Factors

5.1.2 b tag scale factor

5.1.3 Prefiring scale factor

5.1.4 QCD Renormalization and Factorization Scale Factor

5.1.5 Parton distribution function scale factor

5.1.6 Pileup scale factor

5.1.7 $H_T$ scale factor

5.2 $t\bar{t}$ estimation: fit function

5.2.1 Fit function motivation

5.2.2 Fit function implementation

5.3 QCD multijet estimation: control region

5.4 Other Backgrounds and Signal Simulation

# Systematic Uncertainties

6.1 $t\bar{t}$ + jets Systematics

6.1.1 Nominal systematic

6.1.2 Event weight based systematics

6.1.3 Bin migration systematics
List of Figures

1.1 Table of elementary particles in the Standard Model. The (spin 1/2) quark and lepton fermions are shown on the left and the (spin 1) gauge bosons and the (spin-0) Higgs boson are shown on the right [7] ....... 3

1.2 The measurement of the electric charge (left) and the color charge (right) as a function of the distance of the observer [10] .......... 16

1.3 The different types of one-loop corrections for a generic vector gauge field, where (a) is the fermion loop corrections, which contribute a positive term and (b) is the gauge field loop corrections, present only in the non-Abelian case [10] .......... 17

1.4 The results of different measurements of the running coupling constant for QCD over two orders of magnitude of $Q^2$ as of the end of 2019 [6] .......... 18

1.5 All the different electroweak gauge boson interactions terms, including (a) the three point interaction between $W^\pm$ with a photon or $Z$ boson, (b) the four point interaction of the $W$ bosons, and (c) the four point interaction between two $W$ bosons and two neutral bosons (two photons, two $Z$ bosons, or one $Z$ and one photon) [11] .......... 23

1.6 The shape of the Higgs potential in two dimensions when $\mu^2$ and $\lambda$ have opposite signs. The origin is no longer the location of the absolute minimum configuration, with the lowest energy state now located at $\sqrt{-\mu^2/2\lambda}$ [14] .......... 25

1.7 The measured cross section of different processes in the Standard Model conducted by the CMS experiment during Run I and II as of May 2020 [16] .......... 30
1.8 Fits to determine three different parameters of the Standard Model (top quark mass, W boson mass, and Higgs boson mass) as a result of precision measurements of other parameters in the Standard Model [17] (left). A visualization of the fit to determine the different parameters in the CKM matrix, with the constraints from different independent experiments shown in different shaded regions [20].

1.9 The signal strength modifiers for the the production and decay modes for the 125 GeV Higgs boson for the CMS experiment, presented as a ratio to the theoretical value predicted by electroweak theory [21].

2.1 The one-loop corrections the Higgs boson propagator arising, from left to right, the Higgs boson self coupling, the Yukawa coupling to fermions, and the coupling to the vector bosons [35].

2.2 Some of the tree level diagrams for interactions involving the Yukawa term for the top quark [42].

2.3 Proton decay through two R-parity violating couplings, where the first interaction is characterized by the \( \lambda'' \) coupling and the second interaction is characterized by a \( \lambda' \) coupling [49].

2.4 An example of a top squark decay into a top quark and a neutralino, and the neutralino decaying through an \( R \)-parity violating coupling to three jets.

2.5 The sequence of decays in stealth SUSY theories with the minimal number of particles [50].

2.6 The decay of the a bino to a singlino and a photon and the decay of a singlet to two gluons, both mediated by one-loop interactions of the \( Y \) and \( \bar{Y} \) vector-like messenger fields [50].
3.1 The blue LHC dipole magnet as it is lowered into the cavern to be in- stalled \cite{52} and the schematic detailing the different components. The beam separation distance (distance between one beam pipe and the other) is nom- inally 194mm at operating temperatures \cite{53}. 

3.2 The entire CERN accelerator complex, including the Large Hadron Col- lider \cite{54}. 

3.3 The LHC beam dump system with the septum magnet that separates the beam, the fast kicker magnet that deflects the beam from the ring, the dilu- tion magnet, which lowers the intensity of the beam, and the graphite block that the beam is directed to \cite{55}. 

3.4 A schematic of the CMS detector with the major subdetectors labeled \cite{58}. 

3.5 The CMS tracker system. In the center is the pixel detector, divided into the barrel and endcap regions, and right on the outside is the silicon strip detector system, composed of six different components \cite{56}. 

3.6 Comparison of the pixel detector before and after the Phase I upgrade \cite{59}. 

3.7 A schematic of the electromagnetic calorimeter showing the different compo- nents, including the barrel, the endcap, and the preshower \cite{60}. 

3.8 The electromagnetic calorimeter during installation in the barrel region (left). The translucent green color is a result of the lead tungstate crystals, shown bundled up side by side in a module (right) \cite{56}. 

3.9 Rate of degradation of the response in Layer 1 as a function of $\eta$ and total integrated luminosity during Run I and the beginning of Run II (left) \cite{61}. An array of silicon photomultipliers that will replace the hybrid photodiodes (middle) \cite{62}. A graduate student posing for a picture before continuing to install the front-end electronics in a readout box on the HCAL endcap (right). 

3.10 The number of layers in each channel readout before the upgrade (left) and after the Phase I upgrade between 2017 and 2019 \cite{62}. 

xiii
3.11 Picture of the solenoid during installation (left) and a schematic showing the five modules that make up the cold mass inside the cryostat and some of the supporting system features.

3.12 The muon system with all the subsystems explicitly defined. In the barrel, there are drift tubes (MB) and resistive plate chambers (RB). In the end-cap, there are cathode strip chambers (ME) and the resistive plate chambers (RE).

3.13 The cathode strip chamber modules placed where they will be installed in ME1 (left) and installation of an RPC module into the endcap (right).

3.14 The flow of information in the Level 1 Trigger.

3.15 A pictorial representation of a collision in 2016 at the CMS detector with many different tracks that may have arisen from one of the $O(10)$ vertices.

3.16 Feynman diagrams for the benchmark $R$-parity violating SUSY model using the $\lambda''_{112}$ coupling (left) and the benchmark stealth SUSY $SY \bar{Y}$ (right).

4.1 The reconstruction of an electron in CMS, including clusters in the ECAL and the hits in the tracker.

4.2 Definition of a tracker and global muon based on how their tracks are reconstructed, shown on a cross section slide of the CMS detector.

4.3 Different jet algorithms applied to a generated HERWIG event with additional soft particles added in. These algorithms are the $k_T$ algorithm (top left), the Cambridge-Aachen algorithm (top right), the SISCone algorithm (bottom left), and the anti-$k_T$ algorithm (bottom right).

4.4 The different stages of the jet energy corrections, applied successively from the reconstructed jets to the calibrated jets from data (top) and for simulation (bottom).
4.5 The pileup offset in both data and simulation, normalized by the average number of pileup interactions for AK4 CHS jets (left) and the resulting scale factor for the jets in different periods of data taking in 2016 (right) [118].

4.6 The jet response for jets of different transverse momenta. The jet response is relatively flat in the barrel region, but has a stronger $p_T$ dependence in the endcaps and forward regions [118].

4.7 The $\eta$ dependent L2L3 Residual correction plotted as a function of $\eta$, showing that only near the boundary between the endcap and the forward detector is the $\eta$-dependence of the jet response mismodeled [118].

4.8 The $p_T$ dependent L2L3 Residual correction plotted as a function of $p_T$ for jets with $|\eta| < 1.3$ in 2016 for the data taking periods G and H (end of the year) [118].

4.9 The distribution of the discriminator used for determining b jets using the CSV algorithm (left) and using the DeepCSV algorithm (right) [120].

4.10 The distribution of the invariant mass of the lepton and the b tagged jet, $M_{b,l}$, used in the loose semileptonic top tag for all four analysis time periods: 2016, 2017, 2018A, and 2018B. If there is more than one b tagged jet, the value closest to 105 GeV is used. Note that the signal cross sections are increased by a factor shown in the legend so that shapes can be prepared.

4.11 The signal region shown divided up by $N_j$ and $S_{NN}$ such that high $S_{NN}$ and high $N_j$ is signal enriched and low $S_{NN}$ and low $N_j$ is background enriched.

4.12 On the left is the ratio of events after and before the HEM failure for the $\eta$-$\phi$ of good muons and on the right is the equivalent plot for good electrons. Binning is roughly chosen to match the $\eta$-$\phi$ coverage of RBXs. Looking in the region of HEM 15/16, specifically for good electrons, there is an excess when comparing data taken after the HEM failure to data taken before.
4.13 The percentage of total background after the baseline selection is applied for 2016 (left), 2017 (middle), and 2018 (right). The category other consists of diboson, triboson, single top + jets, W + jets, and Drell-Yan + jets events.

4.14 The normalized shape for $t\bar{t} + \text{jets}$ and some signal models for 2016 for input variable distributions. The top row shows leading jet $p_T$, leading jet $\eta$, leading jet $\phi$, leading jet mass. In the middle row is the FWM for moments 2, 3, 4 and 5, from left to right. The bottom row shows the shape of the three JMTE.

4.15 On the left is the event $\beta_z$ for events in the SingleMuon data set and on the right is the equivalent plot for the SingleElectron data set. The black distribution corresponds to all events recorded in 2018A and the red distribution corresponds to all events recorded in 2018B.

4.16 On the left is the 2nd Fox-Wolfram moment for events in the SingleMuon data set and on the right is the equivalent plot for the SingleElectron data set. The black distribution corresponds to all events recorded in 2018A and the red distribution corresponds to all events recorded in 2018B.

4.17 On the left is the 2nd eigenvalue of the jet momentum tensor for events in the SingleMuon data set and on the right is the equivalent plot for the SingleElectron data set. The black distribution corresponds to all events recorded in 2018A and the red distribution corresponds to all events recorded in 2018B.

4.18 Example of a fully connected neural network with arbitrary number of input variables, two hidden layers with arbitrary number of nodes, and an output layer that has two nodes corresponding to the prediction of the classifier (signal or background in this case).

4.19 A diagram that shows a network with a gradient reversal layer.
4.20 For each $N_j$ region: (top left) shape of $S_{NN}$ for $t\bar{t}$ + jets, (top right) shape of MVA output of all signal models together, and (bottom) ROC curve of these distributions for the no gradient reversal case.

126

4.21 For each $N_j$ region: (top left) shape of $S_{NN}$ for $t\bar{t}$ + jets, (top right) shape of MVA output of all signal models together, and (bottom) ROC curve of these distributions with gradient reversal applied.

126

4.22 The architecture of the neural network. Starting from the input layer, the network moves to a lambda layer that scales all inputs to have a mean of 0 and variance of 1, followed by a hidden layer that is then batch normalized, a common technique used in neural networks. Next, there is another hidden layer that has dropout performed and is connected to an output layer that is supervised for the signal vs background classification. Finally, to remove $N_j$ dependence, this output layer is connected to a gradient reversal layer followed by a hidden layer, and finally another output layer that is supervised for classifying $N_j$.

128

4.23 The overall $S_{NN}$ score distribution for all the estimated backgrounds and the data for 2016 (left) and 2017 (right). A few representative signal samples’ $S_{NN}$ distributions are also overlaid with the cross section increased so that the shapes are visible. The uncertainty band incorporates the statistical uncertainty and the two largest systematic uncertainties added in quadrature.

129

4.24 Bin division with respect to $N_j$ and MVA value, in the ideal scenario (left), and the actual scenario (right). The discrepancy is a result of the residual dependence on $N_j$ in the training.

130

4.25 (Left) Number of signal events for different RPV signal models in each NN bin. (Right) The sum in quadrature over each $N_j$ bin of the significance metric as defined in Equation 4.11 per NN bin.

133
4.26 Dividing up the $t\bar{t} +$ jets simulation into the four NN bins using bin boundaries defined only using the $N_j = 7$ bin (left) and defined by each individual $N_j$ bin (right).

5.1 The muon track reconstruction efficiency as a function of $p_T$ for 2016 (left), 2017 (middle), and 2018 (right) derived using the tag-and-probe method in Drell-Yan simulation samples and data [131].

5.2 The muon reconstruction and “Medium” identification efficiency for muons with $|\eta| < 0.9$ on the left and for muons with $2.1 < |\eta| < 2.4$ on the right [132].

5.3 The muon isolation scale factor binned by $p_T$ and $|\eta|$ for a mini-isolation cut of 0.2 [132].

5.4 Trigger efficiency for the isolated muon triggers binned in $p_T$ (left) and $|\eta|$ (right) in both data and the $t\bar{t} +$ jets simulation for 2016.

5.5 Trigger efficiency scale factor for events that pass the isolated muon triggers binned by $p_T$ and $|\eta|$.

5.6 The electron $\epsilon_{\text{track,rec+ID}}$ binned by $p_T$ for different $|\eta|$ ranges on the left and $\epsilon_{\text{isolation}}$ binned by $p_T$ and different $|\eta|$ ranges on the right [132].

5.7 Trigger efficiency for the isolated electron triggers binned in $p_T$ (left) and $\eta$ (right) in both data and the $t\bar{t} +$ jets simulation for 2016.

5.8 Trigger efficiency scale factor for events that pass the isolated electron triggers binned by $p_T$ and $|\eta|$.

5.9 ECAL pre-firing scale factor distribution for the $t\bar{t} +$jets sample (left) and the RPV SUSY model with $m_{\tilde{t}}$ of 550 GeV (right). The y-axis refers to the binning by a multivariate neural net output that will be described in Sec. 4.3.

5.10 The dependence on the cross section of $e^+e^- \rightarrow \text{hadrons}$ relative to the hard process of $e^+e^- \rightarrow qq$ on the QCD renormalization scale. Here $Q$ is taken to be the reference scale of the mass of the Z boson [137].

5.11 The PDF set (NNPDF3.1) used for the generation of the simulation for a scale of $\mu^2 = 10 \text{GeV}^2$ on the left and $\mu^2 = 10^4 \text{GeV}^2$ [140].
5.12 Number of events, measured in luminosity, as a function of the mean number of interactions in 2016 for simulation (left) and for the recommendation derived in data (right) [143].

5.13 The distribution of pileup energy density without the pileup weight (left) and with the pileup weight (middle). The corrected pileup energy density function from the luminosity POG (right) [142].

5.14 The $H_T$ distribution in the signal region selection of all simulation backgrounds and data in 2016 (left), 2017 (middle), and 2018 (right). The data to simulation ratio is shown in the bottom, with the normalization in the simulation done to the predicted theoretical cross section.

5.15 The $H_T$ distribution in the signal region selection of all simulation backgrounds and data for the $N_j = 5$ (left), $N_j = 6$ (middle), and $N_j = 7$ (right) events in 2016. The data to simulation ratio is shown in the bottom, with the normalization in the simulation done to the predicted theoretical cross section.

5.16 The data over simulation ratio in the $H_T$ distribution for $N_j = 8$ in 2016 (left). A direct fit to this distribution is compared to the extrapolated $H_T$ scale factor (middle) and then, the ratio of one over the other is used to determine an $H_T$ scale factor error used to calculate a systematic (right).

5.17 The linear fit to extrapolate the exponential fit parameters, $a(N_j)$ (left) and $b(N_j)$ (middle), of the data to simulation ratio in the $N_j = 5$, 6, and 7 bins to higher $N_j$ bins. The resulting $N_j$ dependent $H_T$ correction is shown on the right.

5.18 The $H_T$ distribution in the signal region selection of all simulation backgrounds and data in 2016 (left), 2017 (middle), and 2018 (right). The data to simulation ratio is shown in the bottom, with the normalization in the simulation done to the predicted theoretical cross section.
5.19 The two different jet scaling patterns in a study measuring the cross section of jets in addition to a Z boson, with the staircase scaling (left) and Poisson scaling (right) [148].

5.20 The overall $N_j$ distribution where the Poisson scaling dominates in the lower additional jet multiplicity regime and the staircase scaling for the high additional jet multiplicity regime. [144].

5.21 The $N_j$ distribution and the $S_{NN}$ distribution of the QCD multijet dominated control region selection.

5.22 The $N_j$ distribution of the QCD multijet simulation in the signal region (SR) with a good muon (blue), in the SR with a good electron (green), and in the control region (CR) with a non-isolated muon (red). All shapes are normalized to unit area.

5.23 The normalized $N_j$ distributions for the different components of the Other background compared to $t\bar{t} +$ jets and $t\bar{t} + X$ in 2016 (left) and 2017 (right).

6.1 The $t\bar{t} +$ jets shape in simulation for data collected in 2016, 2017, 2018A, and 2018B divided into the four NN bins. The finite sample contributed to deviations between the shapes at the highest $N_j$ bins.

6.2 The “nominal” shape systematic for the year 2016 derived as a ratio of the fitted $N_j$ distribution in each NN bin to the fitted total $N_j$ distribution.

6.3 All event weight based shape systematics (up variation) for the year 2016 derived as a ratio of the $N_j$ distribution in each NN bin to the total $N_j$ distribution, and then further divided by $R_{t,\text{nom}}$.

6.4 All bin migration based shape systematics for the year 2016 derived as a ratio of the $N_j$ distribution in each NN bin to the total $N_j$ distribution, and then further divided by $R_{t,\text{nom}}$.
6.5 All $t\bar{t} + $ jets modeling parton shower based shape systematics for the year 2016 derived as a ratio of the $N_j$ distribution in each NN bin to the total $N_j$ distribution, and then further divided by $R_{i,\text{nom}}$. $CR$ is the color reconnection systematic, $hdamp$ is the ME-PS matching systematic, and $UE$ is the underlying event systematic.

6.6 The data over simulation ratio in the $H_T$ distribution for $N_j = 8$ in 2016 (left). A direct fit to this distribution is compared to the extrapolated $H_T$ scale factor (middle) and then, the ratio of one over the other is used to determine an $H_T$ scale factor error used to calculate a systematic (right).

6.7 The alternate versions of the $H_T$ scale factor in 2016 used to derive the two systematics: one to address the low statistics in the high $H_T$ tail (left) and one to address potential mis-modeling of the $N_j$ dependence (right).

6.8 The leading jet mass distribution without (left) and with (right) the $H_T$ scale factor in 2017. The jet mass is one of the input variables to the neural net.

6.9 The “up” fluctuation for the four $H_T$ systematics for the year 2016. The first systematic ($8j$ Extrapolation) utilizes the error from comparing the fit to the extrapolation in the $N_j = 8$ $H_T$ distribution, the high $H_T$ tail systematic addresses mis-modeling at high $H_T$, the constant $N_j$ systematic addresses the $N_j$ dependence of the scale factor, and the no $H_T$ SF addresses shape changes as a result of the scale factor.

6.10 The agreement in the data versus simulation in data collected in 2016 for the $S_{NN}$ distribution per $N_j$, with (from left to right) $N_j = 7$, 8, and 9 on the top row and $N_j = 10$ and $N_j \geq 11$ on the bottom row. The $\geq 11$ bin was chosen as a result of simulation sample size.

6.11 The shape differences in the $S_{NN}$ distribution per $N_j$ relative to the total $S_{NN}$ distribution for data from the control region collected in 2016, with (from left to right) $N_j = 7$, 8, and 9 on the top row and $N_j = 10$ and $N_j \geq 11$ on the bottom row.
6.12 The shape differences in the $S_{NN}$ distribution per $N_j$ relative to the total $S_{NN}$ distribution for the QCD multijet simulation in the control region in 2016, with (from left to right) $N_j = 7, 8, 9$ on the top row and $N_j = 10$ and $N_j \geq 11$ on the bottom row. ............................................................... 184

6.13 The shape differences in the $S_{NN}$ distribution per $N_j$ relative to the total $S_{NN}$ distribution for the $t\bar{t} +$ jets simulation in the signal region in 2016, with (from left to right) $N_j = 7, 8, 9$ on the top row and $N_j = 10$ and $N_j \geq 11$ on the bottom row. ............................................................... 184

6.14 The $S_{NN,All}/S_{NN,N_j=i}$ comparisons between data in the control region, QCD multijet simulation in the control region, and $t\bar{t} +$ jets simulation in the signal region for the year 2016. ............................................................... 185

6.15 The nominal shapes for the $N_j$-$S_{NN}$ correlation systematic derived using the data in the control region and the $t\bar{t} +$ jets simulation in the signal region. 187

6.16 The fits to the 2016 $N_j$-$S_{NN}$ correlation data-driven systematic with the one standard deviation band showing the flexibility of the systematic shape that the fit can pull. The shapes are for NN bin D1 (top left), D2 (top right), D3 (bottom left), and D4 (bottom right). ............................................................... 188

6.17 The data compared to the simulation in the signal region for 2016 with the leading jet (ranked by $p_T$) in the top left, the subleading jet in the top middle, and so on until the sixth jet in the bottom right. ............................................................... 189

6.18 The data compared to the simulation in the control region for 2016 with the leading jet (ranked by $p_T$) in the top left, the subleading jet in the top middle, and so on until the sixth jet in the bottom right. ............................................................... 190

6.19 The data compared to the simulation in the control region for 2016 with the leading jet (ranked by $p_T$) in the top left, the subleading jet in the top middle, and so on until the sixth jet in the bottom right. ............................................................... 191
6.20 The data compared to the simulation in the control region for 2016 with
the leading jet (ranked by $p_T$) in the top left, the subleading jet in the top
middle, and so on until the sixth jet in the bottom right. ........................ 192

6.21 The shape systematic for the year 2016 derived from scaling the jet $p_T$ and
jet mass for each $t\bar{t}$ + jets simulation event based on a scale factor derived
in the control region. ....................................................................... 193

6.22 All event weight based shape systematics (up variation) for the year 2016 for
the $t\bar{t}$ + X backgrounds. ............................................................ 195

6.23 All event weight based shape systematics (up variation) for the year 2016 for
the Other backgrounds. ................................................................... 196

6.24 The jet energy resolution/correction and $H_T$ shape systematics for the year
2016 for the $t\bar{t}$ + X backgrounds. .................................................. 196

6.25 The jet energy resolution/correction and $H_T$ shape systematics for the year
2016 for the Other backgrounds. ....................................................... 197

6.26 All event weight based shape systematics (up variation) for the year 2016 for
the RPV SUSY signal model with $m_{\tilde{t}} = 400$ GeV. ..................... 198

6.27 The jet energy resolution/correction and $H_T$ shape systematics for the year
2016 for the RPV SUSY signal model with $m_{\tilde{t}} = 400$ GeV. .............. 198

6.28 All event weight based shape systematics (up variation) for the year 2016 for
the SYY SUSY signal model with $m_{\tilde{t}} = 600$ GeV. ......................... 198

6.29 The jet energy resolution/correction and $H_T$ shape systematics for the year
2016 for the SYY SUSY signal model with $m_{\tilde{t}} = 600$ GeV. ............... 199

7.1 Sample distribution for $q_0$ generated with toy studies for the Higgs dis-
covery analysis. This is then used as the probability distribution function

\[
(f(q_0|0, \theta_{\text{obs}}^\text{d})dq_0) \quad 156. \quad ......... ......... ....... \quad 206
\]

7.2 Sample $\tilde{q}_r$ distributions for the Higgs discovery analysis for the background
only hypothesis when $r = 0$ in blue and the signal + background hypothesis
when $r = 1$ in red. Here $\mu$ is the same as $r$ in the body of the text [156]. 209
7.3 A typical data drawn from a Gaussian (left) compared to the Asimov data set for the same Gaussian distribution (right).

7.4 Background only fit to the pseudo data generated from the t\(\bar{t}\) + jets simulation for 2016.

7.5 Signal + background fit to the pseudo data generated from the t\(\bar{t}\) + jets simulation for 2016. The signal model event yield is scaled to the measured signal strength in this plot.

7.6 Background only fit to the pseudo data generated from the t\(\bar{t}\) + jets simulation with the stealth SUSY model with \(m_{\tilde{t}} = 600\) GeV injected at full theoretical cross section for 2016.

7.7 Table of pulls for the background only fit to the pseudo data generated from the t\(\bar{t}\) + jets simulation with the stealth SUSY model with \(m_{\tilde{t}} = 600\) GeV injected at full theoretical cross section for 2016, where the value for any non-fit parameters are in standard deviations. The uncertainty presented here is the larger value if the errors are asymmetrical.

7.8 Signal + background fit to the pseudo data generated from the t\(\bar{t}\) + jets simulation with the stealth SUSY model with \(m_{\tilde{t}} = 600\) GeV injected at full theoretical cross section for 2016.

7.9 Table of pulls for the signal + background fit to the pseudo data generated from the t\(\bar{t}\) + jets simulation with the stealth SUSY model with \(m_{\tilde{t}} = 600\) GeV injected at full theoretical cross section for 2016, where the value for any non-fit parameters are in standard deviations. The uncertainty presented here is the larger value if the errors are asymmetrical.

7.10 Local p value measured for pseudo data with the RPV SUSY model with \(m_{\tilde{t}} = 550\) GeV injected with fits conducted using the stealth SUSY signal shapes. The term “observed” is used to express that the pseudo data is generated and input to the fit in the same way as real data would be.
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.11</td>
<td>Local p value measured for pseudo data with the $1\sigma$ JEC, JER, and FSR scale factors applied to the $t\bar{t} + \text{jets}$ simulation with RPV SUSY signal models (left) and for pseudo data with an artificial signal injected using the stealth SUSY signal shapes (right). The term “observed” is used to express that the pseudo data is generated and input to the fit in the same way as real data would be.</td>
</tr>
<tr>
<td>7.12</td>
<td>Fits to NN bin 1 for the datasets in 2016, 2017, 2018A, and 2018B from left to right.</td>
</tr>
<tr>
<td>7.13</td>
<td>The local p value for different $m_{\tilde{t}}$ values for the RPV SUSY model (left) and the stealth SUSY SYY model (right) for pseudo data generated with the simulation in 2016, 2017, 2018A, and 2018B with the signal injected at full theoretical cross section.</td>
</tr>
<tr>
<td>7.14</td>
<td>Expected upper limit on the cross section as a function of $m_{\tilde{t}}$ for the RPV SUSY model and the stealth SUSY SYY model on the left and right, respectively. The green and yellow bands show the 1 and 2$\sigma$ uncertainty bands. The red line shows the theoretical top squark cross section as a function of its mass.</td>
</tr>
<tr>
<td>8.1</td>
<td>Fitted background prediction and observed data counts for 2016, 2017, 2018A and 2018B in each of the four NN bins. The signal distributions normalized to the theoretical cross section for the RPV SUSY model with $m_{\tilde{t}} = 400,\text{GeV}$ and the stealth SUSY model with $m_{\tilde{t}} = 600,\text{GeV}$ are overlaid.</td>
</tr>
<tr>
<td>8.2</td>
<td>List of nuisance parameters that contribute to the likelihood fit in the background only fit. Only nuisance parameters that have a pull with an absolute value greater than 0.2 are listed here.</td>
</tr>
<tr>
<td>8.3</td>
<td>The “nominal” shape systematic for the year 2017 derived as a ratio of the fitted $N_j$ distribution in each NN bin to the fitted total $N_j$ distribution in simulation.</td>
</tr>
</tbody>
</table>
8.4 Fitted background prediction and observed data counts for 2016, 2017, 2018A
and 2018B in each of the four NN bins. In this particular fit, the $N_j$ shapes
for the stealth SUSY model with $m_{\tilde{t}} = 600$ GeV were allowed to be pulled,
with its theoretically predicted $N_j$ distribution at full cross section shown in
pink.

8.5 Fitted background prediction and observed data counts for 2016, 2017, 2018A
and 2018B in each of the four NN bins. In this particular fit, the $N_j$ shapes
RPV SUSY model with $m_{\tilde{t}} = 400$ GeV were allowed to be pulled, with its
theoretically predicted $N_j$ distribution at full cross section shown in pink.

8.6 List of nuisance parameters that contribute to the likelihood fit in the signal +
background fit using the stealth SUSY SYY signal model with $m_{\tilde{t}} = 600$ GeV.
Only nuisance parameters that have a pull with an absolute value greater
than 0.2 are listed here.

8.7 List of nuisance parameters that contribute to the likelihood fit in the back-
ground only fit using the RPV SUSY signal model with $m_{\tilde{t}} = 400$ GeV. Only
nuisance parameters that have a pull with an absolute value greater than 0.2
are listed here.

8.8 The local $p$ value for different $m_{\tilde{t}}$ values for the RPV SUSY model (left) and
the stealth SUSY SYY model (right) for the observed data in 2016, 2017,
2018A, and 2018B.

8.9 The negative log likelihood scan as a function of the signal strength $r$ shown
for the 2016 data (top left), the 2017 data (top right), 2018S data (bottom
left), 2018B data (bottom right). All other nuisance parameters are

8.10 Observed upper limit on the cross section as a function of $m_{\tilde{t}}$ for the RPV
SUSY model and the stealth SUSY SYY model on the top and bottom,
respectively. The green and yellow bands show the 1 and 2$\sigma$ uncertainty
bands. The red line shows the theoretical top squark cross section as a
function of its mass.
List of Tables

1.1 The fermion content of the Standard Model further subdivided by their quantum numbers: electric charge $Q$, third component of weak isospin $T_3$, weak hypercharge $Y$, and color charge ........................................ 4

1.2 The boson content of the Standard Model further subdivided by their quantum numbers: electric charge $Q$, third component of weak isospin $T_3$, weak hypercharge $Y$, and color charge. \(^1\) There are eight linear color-anticolor combinations corresponding to the eight different gluons. ........................................ 5

2.1 The chiral supermultiplets in the Minimal Supersymmetric Standard Model, with their spin 0 components, spin 1 components and their Standard Model transformations ........................................ 45

2.2 The vector superfields in the Minimal Supersymmetric Standard Model, with their spin 1/2 components, spin 1 components and their Standard Model transformations ........................................ 45

3.1 Muon triggers used by year ........................................ 80

3.2 Electron triggers used by year ........................................ 80

4.1 Electron tight identification selection criteria ........................................ 91

4.2 Effective area values used for isolation calculations \(^{[104]}\) ........................................ 92

4.3 Electron offline selection for the signal region ........................................ 94

4.4 Muon offline selection for the signal and control region ........................................ 97
4.5 Jet identification criteria that need to be passed by all jets to reject fake jets arising from HCAL and ECAL noise. ........................................ 105
4.6 The exact $\eta$, $\phi$, and $p_T$ requirements for jets and leptons that are used to define the HEM event veto. ........................................ 115
4.7 Input variables to the neural net, totaling 39 variables per event ............ 117
4.8 MVA bin edges per $N_j$ bin for all four data sets .......................... 132
5.1 Total number of weighted QCD multijet simulation events in the SR and the CR, along with the transfer factor and its uncertainty for each year .... 164
6.1 The RMS values of the jet $p_T$ distributions for data in the control region and QCD multijet simulation by jet rank (ranked by $p_T$) in 2016, followed by the ratio of the two. ...................................................... 191
6.2 The overall relative size of all systematics over all years for the $t\bar{t} +$ jets background, the non $t\bar{t} +$ jets background, and a representative signal model (RPV $m_{\tilde{t}} = 550$ GeV) ................................................. 202
7.1 Fit parameters for the $t\bar{t}$ shape for 2016 simulation. ....................... 215
7.2 Fit parameters for the $t\bar{t}$ shape for 2016 simulation with the stealth SUSY SYY model with $m_{\tilde{t}} = 600$ GeV injected at full cross section. ............. 219
7.3 Fit parameters for the $t\bar{t} +$ jets shape for the D1 bin in 2016, 2017, 2018A, and 2018B. ............................................................... 222
7.4 Best fit signal strength and expected significances for the RPV SUSY model as a function of $m_{\tilde{t}}$, for the combination of 2016, 2017, 2018A, and 2018B. ...................................................... 223
7.5 Best fit signal strength and expected significances for the stealth SUSY SYY model as a function of $m_{\tilde{t}}$, for the combination of 2016, 2017, 2018A, and 2018B. ...................................................... 224
8.1 Fit parameters for the $t\bar{t} +$ jets shape in the fit to the data, for the combined background only fit for all four time eras. ........................... 230
8.2 Fit parameters for the $t\bar{t} + \text{jets}$ shape for the combined signal + background fits to the data sets in 2016, 2017, 2018A, and 2018B, using the signal model of stealth SUSY SYY with $m_{\tilde{t}} = 600$ GeV and RPV SUSY with $m_{\tilde{t}} = 400$ GeV. 234

8.3 Best fit signal strength and observed significances for the RPV SUSY model as a function of top squark mass, for the combination of 2016, 2017, 2018A, and 2018B. 235

8.4 Best fit signal strength and observed significances for the stealth SUSY SYY model as a function of top squark mass, for the combination of 2016, 2017, 2018A, and 2018B. 236
Chapter 1

Introduction

From a macroscopic viewpoint, the study of modern particle physics arose from the fundamental questions humanity has been asking themselves since the beginning of time. *What makes up the Earth, the Sun, and the stars? What are humans made of? Can everything, even light, be broken down to the same few elements?*

This curiosity of what things were comprised of naturally led to the concept of a basic component of the universe, one that can be obtained by dividing a piece of anything as many times as possible until it could no longer be divided. While this idea of a small basic part (a particle) of matter was conceived simultaneously in both ancient India and Greece, this object is known today as an “atom,” termed by Leucippus of Miletus and Democritus after the Greek word for “cannot be cut.” This paved the way for the desire to search for this smallest part of matter, and, eventually to attempts to further divide the atom into even smaller components more fundamental than the atom. These attempts proved fruitful, as the title of “most fundamental particle” migrated from the atom to the protons, neutrons, and electrons in the late 1800s to early 1900s, and then to the quarks, leptons, and gauge bosons in the mid 20th century. The theoretical framework of these quarks, leptons, and gauge bosons, known as the Standard Model, encompasses the current understanding of the most basic components of the universe, and is the basis for all modern particle physics. Since its inception, the Standard Model has made a multitude of predictions that have been
validated by experimental evidence, including the famous discovery of the Higgs boson on the 4th of July in 2012.

Thus, this section focuses on describing the Standard Model, its history, and its successes, as it provides a good starting point to the motivation behind the studies presented afterwards. The section is divided into six subsections, beginning with the particle content of the Standard Model in Section 1.1. Next, in Section 1.2, quantum electrodynamics (QED) is explained with a focus on fundamental concepts applicable to all forces in the Standard Model, such as symmetries and gauge invariance. Generalization of these concepts, and then its applicability to the strong force and quantum chromodynamics will be the focus of Section 1.3, followed by the components of electroweak theory and the massive vector gauge bosons presented in Section 1.4. Finally, the Brout-Englert-Higgs mechanism is discussed in Section 1.5, and a survey of current experimental limits on the Standard Model will follow in Section 1.6.

The information found in this chapter is covered in further detail in the following resources: [2–6].

1.1 Fundamental Particles of the Standard Model

While the concept of having one indivisible particle is alluring for its compactness, there is actually an array of particles that are inherently different that make up the known universe, and so this subsection aims to provide context as to how these particles relate to each other. The table of elementary particles in the Standard Model are shown in Figure 1.1 which shows all the fundamental particles currently discovered along with a few of its fundamental properties, such as mass, electric charge, and spin. The table is conveniently further subdivided into two larger categories, with the spin 1/2 fermions on the left and the spin 0 and spin 1 bosons on the right.
Figure 1.1: Table of elementary particles in the Standard Model. The (spin 1/2) quark and lepton fermions are shown on the left and the (spin 1) gauge bosons and the (spin-0) Higgs boson are shown on the right [7].

The fermions are the particles that make up all visible matter in the universe, and are grouped into three columns for the three analogous generations of increasing mass from left to right. The first generation comprises of the up quark, the down quark, the electron, and the electron neutrino. The vast majority of all known matter that humans interact with are comprised of quarks from the first generation, as the up and down quarks combine and act as the valence quarks for protons and neutrons, and all atoms are formed with a combination of protons, neutrons, and electrons.

The second generation comprises of the charm quark, the strange quark, the muon, and the muon neutrino. It can be argued that the discovery of these particles ushered in the modern, more messy, era of particle physics, since their existence destroyed the possibility that the neat early framework of just protons, neutrons, and electrons was the entire reality. For example, the discovery of muons from cosmic radiation and its decay to
electrons engendered studies into the beta emission spectra of muon decays, which ultimately led to the discovery of the neutrino. Similarly, the study of neutral kaons, which involve strange quarks, was one of the first indications of CP violation, where the neutral kaon was able to convert to its antiparticle and vice versa, but not at equivalent rates. Thus shattered the idea that all physics is completely symmetric even at the fundamental scale. A few of the questions arising from the study of the second generation, including the mass of neutrinos and CP violation, continue to be frontiers in research in the field today.

The third generation of fermions are the top quark, the bottom quark, the tau lepton, and the tau neutrino. This collection of fermions is notable in that they are particularly heavy, with the top quark having a mass around five orders of magnitude larger than the up quark and two orders of magnitude larger than the bottom quark, which is in the same generation. The discrepancy between the masses is currently still not well understood, but the heavier masses allow these particles to serve as useful probes of Higgs physics because the Higgs interacts with other particles proportional to their masses. Furthermore, as some of these fermions have longer life times, particularly the tau lepton and the bottom quark, their displaced decays have provided both challenges and novel ways of detection in modern particle physics experiments.

<table>
<thead>
<tr>
<th>Fermions</th>
<th>Generation</th>
<th>Spin</th>
<th>Q</th>
<th>$T_3$</th>
<th>Y</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quarks</td>
<td>1 2 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(u^d)_L$</td>
<td>$c_L$</td>
<td>$(t^b)_L$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{3}$</td>
<td>r,g,b</td>
</tr>
<tr>
<td>$u_R$</td>
<td>$c_R$</td>
<td>$t_R$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{3}$</td>
</tr>
<tr>
<td>$d_R$</td>
<td>$s_R$</td>
<td>$b_R$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{3}$</td>
</tr>
<tr>
<td>Leptons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\nu^\mu e)_L$</td>
<td>$(\nu^\tau \mu)_L$</td>
<td>$(\nu^\tau \tau)_L$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>$\nu_e$</td>
<td>$\nu_\mu$</td>
<td>$\nu_\tau$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
</tr>
</tbody>
</table>

Table 1.1: The fermion content of the Standard Model further subdivided by their quantum numbers: electric charge $Q$, third component of weak isospin $T_3$, weak hypercharge $Y$, and color charge.
In addition to belonging to a generation, there are further interesting trends and relations between the fermions, which is demonstrated in the more comprehensive Table 1.1. The table makes it clear that all fermions in the same row have exactly the same fundamental charges (i.e., electric, isospin) and differ only by the mass. Next, implicitly taken into account with each entry in the fermion table is its corresponding anti-particle, which have equal and opposite charges. For example, the anti-particle content of the first generation are the anti-up quark, anti-down quark, the positron, and the electron anti-neutrino. Tab. 1.1 also stresses that each of the six quarks actually represent three different versions: one for each color charge (r,g,b). Finally, the chirality of the particles was glossed over in Fig. 1.1, but is actually an important factor as to whether the particle can interact via this weak force, as shown more explicitly in Tab. 1.1 where the right-handed particles have a charge of zero for the third component of the weak isospin $T_3$. While the details of the importance of chirality will be further explored in Sec. 1.4 in the discussion of the weak force, some immediate things to note is that right-handed quarks and leptons do not interact via the weak force. Another note about chirality is that the neutrinos shown in Fig. 1.1 only represent the left-handed version, as no right-handed neutrinos have been discovered.

<table>
<thead>
<tr>
<th>Bosons</th>
<th>Mass</th>
<th>Spin</th>
<th>$Q$</th>
<th>$T_3$</th>
<th>$Y$</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photon</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>$W^\pm$</td>
<td>80.4</td>
<td>1</td>
<td>$\pm 1$</td>
<td>$\pm 1$</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>$Z^0$</td>
<td>91.2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Gluon</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>(r,g,b) + $\bar{(r,g,b)}$</td>
</tr>
<tr>
<td>Higgs</td>
<td>125.2</td>
<td>0</td>
<td>0</td>
<td>$-\frac{1}{2}$</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1.2: The boson content of the Standard Model further subdivided by their quantum numbers: electric charge $Q$, third component of weak isospin $T_3$, weak hypercharge $Y$, and color charge. $^1$ There are eight linear color-anticolor combinations corresponding to the eight different gluons.

Unlike the fermions, the bosons have integer spin and are divided into two subcategories: the vector (spin 1) bosons and the scalar (spin 0) boson, as summarized in Tab. 1.2. Bosons
are the force carrying particles in the Standard Model and are in charge of the interactions between the fermions.

The vector bosons are responsible for three of the four fundamental forces of nature: the electromagnetic force, the weak force, and the strong force. The electromagnetic force is carried by the massless photon, is represented mathematically as a U(1) symmetry, and is responsible for the familiar phenomena of light and electricity. To interact via the electromagnetic force, fermions need to have a non zero electromagnetic charge, which includes all up type quarks (up, charm, and top) with electric charge +2/3, all down type quarks (down, strange, and bottom) with electric charge −1/3, and all leptons (electron, muon, and tau lepton) with charge −1, while excluding all neutrinos. The weak force is transmitted by the massive W± and Z bosons in a SU(2) symmetry, and collectively these bosons are responsible for phenomena such as flavor changing decays and CP violation. All fermion flavors can interact via the weak force, but the chirality matters since weak isospin must be conserved. This means that interactions with the W± bosons only occur amongst left-handed particles and right-handed anti-particles, whereas the Z boson, which has zero weak isospin, can interact with right-handed particles as well. Finally, the strong force is carried by the eight massless gluons in the SU(3) symmetry, and are responsible for quark confinement and asymptotic freedom, making atomic nuclei feasible energetic states despite electromagnetic repulsion between protons. Only fermions possessing a color charge can interact via the strong force, and so only the quarks (up, down, charm, strange, top, and bottom) and the gluons interact via the strong force.

The only remaining particle is the Higgs boson, which assumes its role as the only scalar boson in the Standard Model with two main functions. The first is to be the particle field that is in charge of providing the masses to the W± and Z bosons through spontaneous symmetry breaking, a process that will be explained in further detail in Sec. 1.5. The second function is to provide the masses for all the fermions through Yukawa interactions, where fermions with higher Yukawa couplings, like the top quark, are more massive.
1.2 Quantum Electrodynamics (QED)

In the context of quantum field theory and particle physics, quantum electrodynamics is one of the most successful theories, with Tomonaga, Schwinger, and Feynman winning the Nobel Prize in Physics in 1965 for their contributions to the theory. Quantum electrodynamics is able to describe different electromagnetic phenomena and continues to predict experimental results to extraordinary precision, such as the existence of the anomalous magnetic dipole moment, which as of 2012 has a measurement that matches the theoretically calculated value up to less than a part per trillion \(10^{-12}\) \[8\]. While earlier versions of this theory, and local gauge quantum field theories in general, had issues with higher order divergences, that for example, led to calculations showing the mass of the electron would be infinite, techniques like renormalization helped establish these theories as both viable and integral in comprehending high energy phenomena. Given that QED can be described as the simplest of the gauge theories that describe the forces incorporated in the Standard Model, QED serves as an easier platform to emphasize important properties for all local gauge theories. It is with this generalization in mind that the mathematics of QED is elucidated with more detail here.

On a pure mathematical level, quantum electrodynamics can be categorized as an Abelian gauge theory with a U(1) symmetry with one gauge field and one spin 1/2 fermion, normally taken to be the electron. Already, this is a simplification since the reality is that there will be extra terms for each additional fermion in the Standard Model; however, those terms will be analogous to the ones presented here using the electron, with additional multiplicative factors depending on the charge. As such, the explanation here will consider the electron and its corresponding anti-particle, the positron, as the sole fermions in the theory. The Lagrangian for QED with one fermion anti-fermion pair is then shown in Eq. [1.1]

\[
\mathcal{L} = -\frac{1}{4} F_{\mu\nu}^2 + i \bar{\psi} \slashed{D} \psi - m \bar{\psi} \psi
\]  

(1.1)

The first term is the kinetic term associated with the gauge boson, in this case, the field representing the photon; the second term represents the interaction between the Dirac
fermion and the field through the covariant derivative $D$; and the third term is the mass term for the fermion. Each component of the Lagrangian will now be described in more detail.

In the first term, also known as the Lagrangian for the free electromagnetic field, $F_{\mu\nu}$ is an antisymmetric rank-2 field tensor that gives the field strength of the classical electromagnetic field. This quantity is related to the covariant four-potential $A_\mu$ in classical electromagnetism through:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

(1.2)

As is true also in classical electromagnetism, the electromagnetic fields represented in $F_{\mu\nu}$ are not uniquely defined by the potential $A_\mu$. Any transformation of the form:

$$A'_\mu(x) = A_\mu(x) + \partial_\mu \alpha(x)$$

(1.3)

where $\alpha(x)$ is some arbitrary function, leaves $F_{\mu\nu}$ unchanged because of the anti-symmetrization of the indices in the definition of $F_{\mu\nu}$. This freedom in defining this potential without changing the Lagrangian and the underlying physics is known as gauge symmetry. Since $\alpha(x)$ is a function of the location in space time, this particular gauge symmetry is a local gauge symmetry.

Gauge symmetry is a very important concept in modern particle physics, as all the forces in the Standard Model are most simply described by introducing local gauge symmetries and gauge fields. From a mathematical point of view, local gauge symmetries represent a redundancy in the description of the physics of the situation. In choosing a specific gauge, certain unphysical pathologies of the gauge field can be sequestered from the theory, and other properties of the system can be emphasized.

A concrete example of this is the quantum electrodynamics potential $A_\mu$, which has four degrees of freedom. The photon that it represents, however, only has two transverse polarization states, and so there are two “redundant” degrees of freedom. One way to get the classical quantization of the photon in the free electromagnetic field from this Lagrangian is by taking the equation of motion and “choosing” a gauge. The equation of motion, obtained
using the Euler-Lagrange formalism, is:

$$\partial_{\mu} \left( \frac{\partial L}{\partial (\partial_{\mu} A_{\nu})} \right) = -\partial_{\mu} F^{\mu\nu} = -\partial_{\mu} (\partial^{\mu} A^{\nu} - \partial^{\nu} A^{\mu}) = \partial^{2} A^{\nu} - \partial_{\mu} \partial^{\nu} A^{\mu} = 0$$  \hspace{1cm} (1.4)

If one imposes the Lorenz gauge of $\partial_{\mu} A^{\mu} = 0$, the second term drops out, and we get an equation very similar to the Klein-Gordon equation for scalars and the Dirac equation for spinors:

$$\partial^{2} A^{\mu} = 0$$  \hspace{1cm} (1.5)

The solution to the vector version of the Klein-Gordon equation is a series of plane wave equations that can be represented by:

$$A^{\mu} = \epsilon^{\mu} e^{iq \cdot x} = j^{\nu}$$  \hspace{1cm} (1.6)

where $\epsilon$ is the polarization vector with four degrees of freedom and $q$ is the four-momenta vector. By applying the Lorenz gauge condition on these solutions, one degree of freedom of $\epsilon$ is removed by $q_{\mu} \epsilon^{\mu} = 0$. Finally, by observing that the photon is massless and that states where the polarization is longitudinal have negative norm, one can further apply the Coulomb gauge condition ($\epsilon^{0} = 0$ or $\vec{\epsilon} \cdot \vec{q} = 0$) to remove the remaining degree of freedom to get the familiar photon wave functions.

Given this exact procedure of imposing two gauge conditions, one can make the argument that the initial description of quantum electrodynamics with its extra degrees of freedom is unphysical, and that the theory is only physical after imposing these restrictions. While this is partially true, one can imagine conducting the calculations without imposing a gauge, and coming up with a different way in handling the nonphysical states. More complicated prescriptions in dealing with gauge-fixing, such as in BRST quantization [9], introduce extra ghost fields and anti fields that when applied to QED result in similar results as presented above. While these techniques are not integral in understanding QED, they are absolutely necessary when applied to non-Abelian gauge theories. What is important here is that fundamentally, the extra degrees of freedom in gauge field theories can be cleverly utilized in order to only pick out the physical aspects of the theory.
The second term in the Lagrangian is the Dirac term for the fermion where \( \mathcal{D} \) is the gamma matrix multiplied by the gauge covariant derivative, written more explicitly as 
\[
\gamma^\mu (\partial_\mu + iqeA_\mu),
\]
where \( q \) is the electric charge of the fermion. For the election, \( q = -1 \), but all results can be generalized for up type quarks and down type quarks by setting \( q \) to \( 2/3 \) and \(-1/3 \), respectively. The covariant derivative is important here, since without the extra field term in the derivative, this term is not invariant under gauge transformations. More explicitly, by applying the following transformations:
\[
A_\mu \to A_\mu + \partial_\mu \alpha(x) \quad \text{and} \quad \psi \to e^{-iqe\alpha(x)}\psi \quad (1.7)
\]
the covariant derivative transforms as:
\[
D_\mu \psi = \partial_\mu \psi + iqeA_\mu \psi \\
\to \partial_\mu (e^{-iqe\alpha(x)}\psi + iqe(A_\mu + \partial_\mu \alpha(x))(e^{-iqe\alpha(x)}\psi)) \\
= e^{-ie\alpha(x)}D_\mu \psi \quad (1.9)
\]
where the extra exponential phase term from the gauge covariant derivative is cancelled by the corresponding exponential on the \( \bar{\psi} \) term. Of the three terms in the Lagrangian, this term is the only one where both the fermion field and boson field are together, and thus this term also describes the interaction between the two. This can be seen by looking at the equations of motion, which taking the derivatives for the Euler-Lagrange equation with respects to the field \( A_\mu \) results in:
\[
\partial_\mu F^{\mu\nu} = q e\bar{\psi}\gamma^\nu \psi \quad (1.11)
\]
where the right hand side of the equation is the conserved charged density current produced by the fermion.

Having a mathematical description for the interactions between the boson field and the fermion field, one can now address the global symmetry in the Lagrangian. If \( \alpha(x) \) is set to be a constant, the symmetry described by the constant phase \( \alpha \) is no longer local nor a gauge symmetry. Instead, the Lagrangian and action remain invariant to this constant phase change, and this is described as a global symmetry similar to other space time symmetries.
Thus, this global symmetry can be described by the first Noether’s theorem, meaning there is a conserved quantity associated with the symmetry. Taking the time component of the charged current and integrating it over all space time, one gets the conserved electric charge $Q$, such that:

\[
Q = q e \int d^3 x \bar{\psi}(x) \gamma^0 \psi(x)
\]  

(1.12)

This is the classical conserved electric charge as a result of Noether’s theorem, and it arises naturally from the interaction term in QED. In other gauge theories, there will also conserved charges and currents, but because they are non-Abelian, they will be a bit more complex and involve generators of the Lie group associated with those theories.

Finally, the third term is the explicit mass term for the fermion; without this term in quantum electrodynamics, the fermions would be massless. In QED, where the only symmetry is the U(1) symmetry discussed here, the appearance of a mass term is perfectly allowed with no consequences. However, with additional gauge symmetries in which the left-handed and right-handed fermions are charged and transform differently under the generators associated with those symmetries, this is no longer the case since explicit mass terms treat the left-handed and right-handed components equally. This will be explored in more detail in the following sections.

### 1.3 Quantum Chromodynamics

Quantum chromodynamics (QCD) is the gauge theory used to describe the strong force, which consists of the gluon gauge fields and all of the quark fields, since they are the only fundamental fermions with color charge. The mathematical representation of QCD can be described as a non-Abelian compact Lie special unitary group of degree three, SU(3). The non-Abelian nature of QCD results in some substantial differences in the physical interpretation and calculations when compared to the Abelian U(1) QED case, but there are many similarities as well. For one, the overall structure of the Lagrangian remains unchanged, consisting of kinetic field terms, interaction terms, and mass terms.
To understand the mathematical structure of QCD, it is important to take a step back and give a brief description of the Lie algebra associated with Yang-Mills theories, of which QCD is a specific example of. A Lie group G is defined by a set of generators $T^a$ that satisfy the following relations:

$$[T^a, T^b] = i f^{abc} T^c \quad \text{and} \quad Tr(T^a T^b) = \frac{1}{2} \delta^{ab} \quad (1.13)$$

where the indices $a, b, c$ run over the number of generators in the particular Lie group and $f^{abc}$ are the fully anti-symmetric structure constants. For the SU(N) group, there are by definition $N^2 - 1$ generators; in QCD, there are eight generators. Each of these generators corresponds to an infinitesimal transformation given by:

$$\hat{U}(\delta \vec{\alpha}) = 1 - i \sum_j^N \delta \alpha_j T^j + O(\delta \alpha^2) \quad (1.14)$$

and any transformation under this group can be expressed as an infinite sum of these smaller transformations:

$$\hat{U}(\vec{\alpha}) = \lim_{n \to \infty} \left( 1 - i \sum_j^N \delta \alpha_j T^j \right)^n = \exp \left( i \sum_j^N \alpha_j T^j \right) \quad (1.15)$$

A more familiar and concrete example of a symmetry described by Lie algebra would be for the spin representation of a fermion, which is SU(2) and the generators $T^a$ are often represented by the Pauli matrices. In this example, the transformation $\hat{U}$ is simply rotations in the spin space, with the SU(N) generalization having analogous “rotations” in n dimensional space. Having defined the transformation, the goal is now to generate a Lagrangian invariant to transformations described by this operator.

The approach taken here is to use the nominal gauge theory Lagrangian from QED given by Eq. 1.1 and generalize it to the non-Abelian case by redefining how each term in the Lagrangian transforms. This nominal Lagrangian is reproduced here:

$$L = -\frac{1}{4} F_{\mu\nu}^2 + i \bar{\psi} \slashed{D} \psi + m \bar{\psi} \psi \quad (1.16)$$

There are three terms where the transformation must be defined: $\psi, A_\mu, F^{\mu\nu}$. The transformation of the fermion field $\psi$ is the easiest, as it simply requires applying the transformation
operator to the fermion field:

$$\psi \rightarrow U(\bar{\alpha})\psi$$  \hspace{1cm} (1.17)

Next, with the foresight from QED, a covariant derivative can be defined that will help aid in cancelling out additional terms coming from the transformations involving the gauge field and the fermion field. The covariant derivative is a little more complicated than the QED case since there are eight generators, meaning that the covariant derivative must take the form:

$$D_\mu = \partial_\mu - igA_\mu^aT^a$$  \hspace{1cm} (1.18)

where $g$ is a coupling constant associated with the strength of the strong force and the summation over the index $a$ from one to the number of generators is implied. With this definition of the covariant derivative, the transformations of the gauge fields, one for each generator, are given by:

$$A_\mu^aT^a \rightarrow U(\bar{\alpha})A_\mu^aT^aU(\bar{\alpha})^{-1} - \frac{i}{g}(\partial_\mu U(\bar{\alpha}))U(\bar{\alpha})^{-1}$$  \hspace{1cm} (1.19)

where this definition is such that the covariant derivative term transforms properly under SU(3):

$$D_\mu = \partial_\mu - igA_\mu^aT^a \rightarrow D'_\mu = \partial_\mu - ig\left[U A_\mu^a T^a U^{-1} - \frac{i}{g}(\partial_\mu U)U^{-1}\right]$$

$$= \partial_\mu + \left[\partial_\mu(UU^{-1})\right] - igU A_\mu^a T^a U^{-1} - (\partial_\mu U)U^{-1}$$

$$= \partial_\mu + U(\partial_\mu U^{-1}) + (\partial_\mu U)U^{-1} - i g U A_\mu^a T^a U^{-1} - (\partial_\mu U)U^{-1}$$

$$= U\left[U^{-1}\partial_\mu + (\partial_\mu U^{-1}) - i g A_\mu^a T^a U^{-1}\right]$$

$$= U\left[U^{-1}\partial_\mu - i g A_\mu^a T^a\right] U^{-1} = U D_\mu U^{-1}$$

Here, the explicit dependence of the transformation on $\bar{\alpha}$ is suppressed and the transformation is unitary because it is the transformation of the SU(3) group. Finally, the last step is
to define the kinetic terms for the gauge field potentials as:

\[ F_{\mu\nu}^a T^a = \frac{i}{g} [D_{\mu}, D_{\nu}] = \frac{i}{g} [\partial_{\mu} + iA_{\mu}^b T^b, \partial_{\nu} + iA_{\nu}^c T^c] \]

\[ = \left[ \partial_{\mu} A_{\nu}^a - \partial_{\nu} A_{\mu}^a - i[A_{\mu}^b, A_{\nu}^c] \right] T^a \]

\[ = \left[ \partial_{\mu} A_{\nu}^a - \partial_{\nu} A_{\mu}^a - g f_{abc} A_{\mu}^b A_{\nu}^c \right] T^a \]

The first two terms are familiar from the QED Lagrangian, whereas the last term is new and a result of the non-Abelian nature of QCD. This term can also be shown to transform properly under SU(3) as \( UF_{\mu\nu}^a T^a U^{-1} \):

\[ F_{\mu\nu}^a T^a = \frac{i}{g} [D_{\mu}, D_{\nu}] \rightarrow \frac{i}{g} \left[ UD_{\mu} U^{-1} U D_{\nu} U^{-1} - U D_{\nu} U^{-1} U D_{\mu} U^{-1} \right] \]

\[ = \frac{i}{g} \left[ UD_{\mu} D_{\nu} U^{-1} - U D_{\nu} D_{\mu} U^{-1} \right] \]

\[ = \frac{i}{g} U[D_{\mu}, D_{\nu}] U^{-1} = UF_{\mu\nu}^a T^a U^{-1} \]

Now all the ingredient are available to write down the QCD Lagrangian, with the notation change of \( F_{\mu\nu} = F_{\mu\nu}^a T^a \):

\[ \mathcal{L} = -\frac{1}{4} F_{\mu\nu}^2 + i \bar{\psi} \slashed{D} \psi - m \bar{\psi} \psi \]

As expressed in the introduction in the section, the form of the Lagrangian is almost exactly the same as that for QED. Some of the conclusions in the QED case are generalizable here as well. All the gluon fields remain massless like the photon, as explicit mass terms for the gluon gauge fields are also not gauge invariant under SU(3). The fermion-gluon interaction term arising from the covariant derivative has a similar interpretation, allowing for gluon interactions with two quarks. In addition, the conserved currents associated with the global symmetries ensure that the color charge current density has to be conserved in any strong force interactions. Finally, the explicit mass term for the fermions is allowed in QCD, as QCD treats the left-handed quarks and the right-handed quarks equally.

The major difference between QCD and QED is that gluons can interact with themselves and photons cannot. Particularly, when breaking down the kinetic gauge field terms into
its components:

$$-\frac{1}{4} F_{\mu\nu}^a F_{\mu\nu}^a = -\frac{1}{2} \partial^\mu A_\nu^a (\partial_\mu A_{\mu}^a - \partial_\nu A_{\mu}^a)$$

$$- g f_{abc} \partial^\mu A_\nu^b A_{\mu}^c - \frac{g^2}{4} f_{abc} f_{ade} A_\mu^b A_\nu^c A_{\mu}^d A_{\nu}^e$$

there exists two sets of terms that explicitly have the gluon gauge fields interacting with each other in three point and four point interactions that are nonexistent in the QED case. Moreover, the gluon fields carry color charge (one color and one anti-color charge), but the photon does not carry electric charge.

There are some physical consequences of the self-interaction of gluons that may be very surprising. The first is the concept of asymptotic freedom, where the strong force is negligible between two colored particles at very small distances, but gains in strength as the two colored particles are pulled apart. This goes against the intuition of what a force does, as the electromagnetic force is known to get weaker at larger distances.

A good analogy for understanding asymptotic freedom on an intuitive basis is often framed in the context of measuring a point charge from a distance as an observer. For QED, the point charge polarizes the vacuum, essentially pulling virtual electron-positron pairs out of the vacuum for short amounts of time to not violate the Heisenberg uncertainty principle. These electron-positron pairs in the vacuum polarization tend to align so that if the point charge is positive, the electron is closer to the point charge and so “shields” part of the charge from an observer. As the observer moves further away, there is more vacuum between the observer and the point charge, and so there is more “screening,” resulting in feeling a weaker charge.

For the QCD, the effect is often described as the opposite of the case for QED, and the effect is known as anti-screening. There are still quark-antiquark pairs that are produced in the vacuum that screen the color charge, but the gluon interactions overwhelm the screening effect of the virtual quark-antiquark pairs, resulting in an enhanced strong force as distance increases. As an observer tries to probe a color charged particle by pulling it away and isolating it, the gluons continue to enhance the strong force holding the particle back, until
Charge screening in QED

• In QED, a charged particle like the electron is surrounded by a cloud of virtual photons and \( e^+ e^- \) pairs continuously popping in and out of existence. Because of the attraction of opposite charges, the virtual positrons tend to be closer to the electron and screen the electron charge, as is indicated in the figure. This is analogous to the polarization of a dielectric medium in the presence of a charge and is called vacuum polarization.

• This gives rise to the notion of an effective charge \( e(r) \) that becomes smaller with larger distance.

• One says that the beta function \( \beta(r) \equiv d(e(r)) / dr \ln r \) is positive in QED.

Charge screening in QCD

• Likewise, the QCD vacuum consists of virtual \( q \bar{q} \) pairs, and if this would be all, the charge screening mechanism would be the same as in QED, with a positive beta function.

• However, due to the gluon self-coupling, the vacuum will also be filled with virtual gluon pairs as is indicated in the figure. Because the gluon cloud carries color charge, it turns out that the effective charge becomes larger with larger distance; the beta function is negative. This effect is called antiscreening.

• It turns out that the negative contribution wins over the positive contribution, so that the QCD beta function is negative, and the effective strong coupling becomes small at short distances.

Figure 1.2: The measurement of the electric charge (left) and the color charge (right) as a function of the distance of the observer.

Mathematically, this effect can be understood by considering the one-loop corrections to the gauge field propagator. For the gluon field propagator in QCD, there are one-loop corrections from the gluon self-interaction vertex and from the interaction with the quarks, as shown in Fig. 1.3. Since the momenta of the particle in the loop can be infinitely large, the addition of the contributions of these diagrams into the propagator results in divergences and infinities in the calculation. These infinities are only further exacerbated when considering higher order corrections.

These infinities can be dealt with through different methods used for regularizing the infinities present in these integrals and renormalizing the theory, but at the cost of introducing a reference energy scale \( \mu^2 \). These techniques essentially sequesters the infinities from the calculations by measuring the coupling constant at this reference energy scale, and assuming that the strength of an interaction at any other energy scale characterized by \( Q^2 \), can be extrapolated from the value at \( \mu^2 \) perturbatively, known as a running coupling.
Figure 1.3: The different types of one-loop corrections for a generic vector gauge field, where (a) is the fermion loop corrections, which contribute a positive term and (b) is the gauge field loop corrections, present only in the non-Abelian case [10].

The infinities that were present in the integrals can then be understood as already factored into the value at the reference energy scale. For QCD, the running coupling constant is then described (at the one-loop level) by:

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + \beta_0 \alpha_s(\mu^2) \ln(Q^2/\mu^2)}$$

where the function $\beta_0$ characterizing how the extrapolation from the reference energy to another energy scale is done. For the one loop corrections in QCD, the beta function is:

$$\beta_0 = \frac{-2n_f + 11N_c}{12\pi}$$

with $N_c$ is the number of colors in QCD and $n_f$ is the number of quark flavors in the Standard Model.

It turns out that the first term in the beta equation arises from the one-loop corrections from the fermion-gluon interaction terms (diagram (a) in Fig. 1.3), and the second term arises from the one-loop corrections resulting from the gluon self-interaction terms (diagram (b) in Fig. 1.3). If the first term is larger than the second term in the beta equation, then $\beta_0$ is negative, and the coupling strength increases as $Q^2$ increases, and so the interaction is stronger at smaller distance scales. If the second term is larger than the first term in the beta equation, then $\beta_0$ is positive, and the interaction strength decreases as $Q^2$ increases, corresponding to stronger interactions at larger distance scales.

This latter case is exactly the case for QCD, as $N_c = 3$ for the three color charges in SU(3) and $n_f = 6$ for the six flavors of quarks, resulting in $\beta_0 = 21/12\pi$. This corresponds
Figure 1.4: The results of different measurements of the running coupling constant for QCD over two orders of magnitude of $Q^2$ as of the end of 2019 [6].

to having a stronger strong force at larger distance scales and provides a mathematical representation of the effects of anti-screening from the gluon self-interactions. From this equation, it is clear that in order to probe the effects of the strong force, a very high energy probe has to be used. This also means that in normal circumstances, color charged particles coalesce together in such a way that they remain colorless from an outside observer. Experimental evidence has shown that the above calculation works, as the plot of the running coupling constant as a function of $Q^2$ with the reference energy scale $\mu = m_Z \approx 90 \text{ GeV}$ shown in Fig. 1.4 shows good experimental agreement with theory.

For an example of the type of physics that would have resulted if the $\beta_0$ term was negative, recall that in QED, there are corrections to the photon propagator from loops involving electron-positron pairs (diagram(a) in Fig. 1.3), but no photon self-interaction. Thus, while the constants are different for QED, the sign of the second term remains unchanged, with $\beta_0 = -1/3\pi$ for QED. Therefore, there is no anti-screening in QED, and the electromagnetic interaction is observed to be weaker at longer distances.
1.4 Electroweak Theory

Having derived the Lie algebra necessary to understand QCD and having developed the mathematical framework for QED, all the mathematics for electroweak theory has already been described. From a mathematical perspective, electroweak theory is just the Lie algebra for SU(2) coupled with a U(1) symmetry, though with the additional requirement that the SU(2) symmetry only acts on the left-handed particles (SU(2)_L). The latter is derived from an experimentally observed fact that the W bosons do not interact with the right-handed fermions.

The standard notation used to describe the fermion content in electroweak theory is as follows:

\[ Q_i = \begin{pmatrix} u_i \\ d_i \end{pmatrix}_L \quad u_i = u_{R,i} \quad d_i = d_{R,i} \quad L_i = \begin{pmatrix} \nu_i \\ e_i \end{pmatrix}_L \quad e_i = e_{R,i} \] (1.26)

where the index \( i \) runs over the three generations, \( Q_i \) is the left-handed quark doublet, \( u_i \) is the right-handed up type quark singlet, \( d_i \) is the right-handed down type quark singlet, \( L_i \) is the left-handed lepton doublet, and \( e_i \) is the right-handed lepton singlet. The term singlet and doublet refers to how these particles are charged under the SU(2)_L symmetry, where the fermions in the doublet have weak isospin charge \( T = \frac{1}{2} \). Furthermore, the fermion on the top position of the doublet has isospin up (\( T_3 = \frac{1}{2} \)), and the fermion on the bottom position of the doublet has isospin down (\( T_3 = -\frac{1}{2} \)). The right-handed fermions do not have an isospin charge (\( T_3 = 0 \)) since they are not coupled to the weak isospin symmetry.

The Lagrangian is derived in a similar way to the one in QCD, with transformations under SU(2)_L for the left-handed quark and lepton doublets characterized by the three generators. More explicitly, using the representation of the generators as the Pauli matrices \( (\sigma^a) \), this transformation under SU(2)_L can be represented as:

\[ Q \rightarrow Q' = \exp \left( -\frac{i\alpha(x) \cdot \sigma^a}{2} \right) Q \] (1.27)

\[ L \rightarrow L' = \exp \left( -\frac{i\alpha(x) \cdot \sigma^a}{2} \right) L \] (1.28)

For the U(1) symmetry, the transformation has the same structure as the QED case and
applies to all the fermions. For this SU(2)_L \otimes U(1)_Y symmetry, the covariant derivative is given by:

\begin{align*}
D_\mu &= \partial_\mu - ig_1 \frac{Y}{2} B_\mu, \quad \text{right-handed particles} \\
D_\mu &= \partial_\mu - ig_2 T_a W^a_\mu - ig_1 \frac{Y}{2} B_\mu, \quad \text{left-handed particles}
\end{align*}

where \( W^a_\mu \) are the gauge fields with \( a = 1, 2, 3 \) for the three generators of the SU(2) symmetry, \( T_a \) is just the Pauli matrices divided by two \( \frac{\sigma^a}{2} \), \( B_\mu \) is the gauge field associated with the U(1) symmetry, \( Y \) is the weak hypercharge associated with the U(1) symmetry, and \( g_1 \) and \( g_2 \) as the gauge coupling constants for the electroweak interaction. The electroweak Lagrangian is then given by:

\[ L_{EW} = -\frac{1}{4} W^{\mu \nu}_i W^i_{\mu \nu} - \frac{1}{4} B_\mu B^{\mu \nu} + \bar{Q}_i i \slashed{D} Q_i + \bar{L}_i i \slashed{D} L_i + \bar{u}_{R,i} i \slashed{D} u_{R,i} + \bar{d}_{R,i} i \slashed{D} d_{R,i} + \bar{e}_{R,i} i \slashed{D} e_{R,i} \]

Here the first difficulty with chirality arises in the mass terms with the fermions, which is not gauge invariant given that the left-handed and right-handed particles have different covariant derivatives. To see this explicitly, the typical mass term for a Dirac fermion is given by:

\[ -m \bar{\psi} \psi = -m \bar{\psi} (P_L + P_R)(P_L + P_R) \psi \]

\[ = -m (\bar{\psi}_R \psi_L + \bar{\psi}_L \psi_R) \]

where the fields for the left-handed and right-handed particle are intricately tied together. Thus, all the fermions have to gain mass through a different mechanism in order for this SU(2)_L \otimes U(1)_Y symmetry to hold.

Next, it is important to recover the physics associated with the weak force and the electromagnetic force from this Lagrangian. Naively, one might expect the W and Z bosons to be associated with the three generators of the SU(2)_L symmetry, and the photon to be the gauge boson from the U(1)_Y symmetry, but this is not the case. For one, the
covariant derivative for the lepton doublet includes the neutrino, which interacts with the
gauge field associated with the \( U(1)_Y \) symmetry. Empirically, neutrinos do not have electric
charge and therefore will not interact with the photon, and so the \( B \) field cannot represent
electromagnetism. To solve this issue, the gauge bosons represented by the fields in this
Lagrangian must mix in a way to produce the \( W^\pm, Z, \) and photon of the weak force and
electromagnetism.

One way to figure out how these gauge fields mix is to look at the conserved currents
associated with each of the gauge boson fields of the \( SU(2)_L \) symmetry. Since \( W^\pm \) are the
only bosons that are charged under electromagnetism, any charged current that changes
the electric charge of the particles must be associated with these bosons. Using the Pauli
matrix representation for the generators,

\[
\begin{align*}
\sigma_1 &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\
\sigma_2 &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \\
\sigma_3 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}
\end{align*}
\] (1.31)

expanding out the interaction terms, and noting that the conserved current is

\[
J_a^\mu = L_i \gamma^\mu \sigma_a^i L_i + Q_i \gamma^\mu \sigma_a^i Q_i 
\] (1.32)

one can then calculate the three conserved currents associated with the three gauge fields.

Explicitly, the conserved current associated with \( L_i \) is calculated as:

\[
J_1^\mu = (\bar{\nu}, \bar{e}) \gamma^\mu \frac{\sigma_1}{2} \begin{pmatrix} \nu \\ e \end{pmatrix}
\] (1.33)

\[
J_2^\mu = \frac{1}{2} (\bar{\nu} \gamma^\mu e + \bar{e} \gamma^\mu \nu), \\
J_3^\mu = \frac{1}{2} (\bar{e} \gamma^\mu e - \bar{\nu} \gamma^\mu \nu)
\] (1.34)

where all the left-handed subscripts are implied for all leptons and neutrinos and right-
handed subscripts are implied for all anti-leptons and anti-neutrinos. Observing the terms in
the conserved currents, only the first two gauge fields couple the lepton with the neutrino (or
in the case of the quarks, the up type quark with the down type quark). Thus, interactions
of the fermions involving these two gauge fields change the electric charge by \( \pm 1 \), meaning
these two gauge fields must mix to produce the W bosons. Since both of these gauge fields have one $\bar{e}$ term and one $e$ term of equal weight, the natural mixing for the observed W bosons are:

$$W^{\pm}_\mu = \frac{1}{\sqrt{2}}(W^1_\mu + iW^2_\mu)$$ \hspace{1cm} (1.35)

This leaves the third gauge field $W^3$ and the B field to mix to become the gauge fields associated with the Z boson and the photon. The mixing of these two fields are not as straightforward and is often cast in terms of the Weinberg angle $\theta_W$:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos(\theta_W) & \sin(\theta_W) \\ -\sin(\theta_W) & \cos(\theta_W) \end{pmatrix} \begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix}$$ \hspace{1cm} (1.36)

The value of $\theta_W$ is then tuned such that the conserved current for the photon field $A_\mu$ does not interact with the neutrino. Explicitly solving the resulting set of equations gives the relationship between the gauge coupling for the electromagnetic force relative to the two gauge couplings in the electroweak theory as:

$$e = g_2 \sin(\theta_W) = g_1 \cos(\theta_W)$$ \hspace{1cm} (1.37)

with the relative strength of the two gauge couplings in the electroweak theory being:

$$\sin(\theta_W) = \frac{g_1}{\sqrt{g_1^2 + g_2^2}}$$ \hspace{1cm} (1.38)

Finally, there are many gauge field interaction terms, especially given the additional mixing of the $W^a$ and $B$ fields to produce the $W$, $Z$, and photon fields. The Feynman diagrams with all the gauge boson self-interaction terms are shown in Fig. 1.5 with all the expected interactions given the need for conservation of charge. There are no trilinear or quartic couplings amongst just Z bosons and photons.

### 1.5 Brout-Englert-Higgs Mechanism

Up until this point, the concept of the gauge boson masses have been purposefully left out in the discussion of the development of the electroweak theory of the Standard Model. This
is because the introduction of masses in a chiral gauge theory can be tricky since explicit mass terms are not inherently gauge invariant; however, by introducing extra degrees of freedom into the gauge theory, one can allow for masses to arise naturally. The explicit delineation of this mechanism in the Standard Model is credited to Robert Brout, François Englert, and Peter Higgs in their seminal papers in 1964 [12,13], with the latter two winning the Nobel Prize for this work in 2013.

From a mathematical perspective, a good place to start is the electroweak part of the Standard Model Lagrangian, with only the electron doublet used represented here to simplify expressions:

\[
L_{\text{EW}} = -\frac{1}{4} W_{\mu\nu}^a W_{\alpha\beta}^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B_{\alpha\beta}^{\mu\nu} + \bar{L}_i (i D_{\mu} \gamma^{\mu}) L_i + \bar{e}_{R,i} (i D_{\mu} \gamma^{\mu}) e_{R,i} \tag{1.39}
\]

where \( i = 1, 2, 3 \) for the three generations, \( a = 1, 2, 3 \) for the three W electroweak fields, and \( \mu \) and \( \nu \) are the standard Lorentz indices. As a recap, the electroweak field strengths are given by:

\[
W_{\mu\nu}^a = \partial_{\mu} W_{\nu}^a - \partial_{\nu} W_{\mu}^a + g_2 \epsilon^{abc} W_{\mu}^b W_{\nu}^c \tag{1.40}
\]

\[
B_{\mu\nu} = \partial_{\mu} B_{\nu} - \partial_{\nu} B_{\mu} \tag{1.41}
\]

and the covariant derivatives are given by:

\[
D_{\mu} L_L = (\partial_{\mu} - ig_2 T_a W_{\mu}^a - ig_1 \frac{Y}{2} B_{\mu}) L_L \tag{1.42}
\]

\[
D_{\mu} e_{R} = (\partial_{\mu} - ig_1 \frac{Y}{2} B_{\mu}) e_{R} \tag{1.43}
\]
with $T_a$ as the generators of the SU(2)$_L$ gauge group and $g_1$ and $g_2$ are once again, the coupling constants for the electroweak interaction. The Higgs mechanism then posits the existence of a complex scalar field doublet with weak hypercharge ($Y = \frac{1}{2}$):

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ H + i\phi_0 \end{pmatrix}$$

(1.44)

which adds a contribution to the Lagrangian:

$$\mathcal{L}_S = (D^\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi), \quad \text{with} \quad V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda(\Phi^\dagger \Phi)^2$$

(1.45)

The first term in the scalar complex field Lagrangian is the standard kinetic term associated with the scalar field, and the second term is the scalar potential associated with this field. The majority of the phenomenology of the Higgs mechanism arises from the relative signs of $\mu^2$ and $\lambda$. Per particle physics convention, $\lambda$ is often taken to be positive, and so the two interesting cases are where $\mu^2 > 0$ or $\mu^2 < 0$.

If $\mu^2 > 0$, then the potential takes the shapes of a quartic function where the lowest energy state is given by:

$$\langle 0 | \Phi | 0 \rangle = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

(1.46)

which comes as no surprise as the potential is zero or positive everywhere and so the minimum vacuum state is where the potential is equal to zero at the origin. In this case, $\mu^2$ acts as the coefficient to the mass term for this scalar, and one can interpret this doublet as a new complex scalar field introduced into the Lagrangian, corresponding to four real scalar particles with the same mass, each with one degree of freedom. However, if $\mu^2 < 0$, the minimum energy state is no longer located at the origin, and the potential takes on a “Mexican hat” shape, as shown in Fig. 1.6. The name is derived from the shape of the potential projected in 2D space when $\lambda$ and $\mu^2$ have opposite signs, and the most poignant feature is that the origin is no longer the lowest energy state.

In this potential, the scalar field acquires a vacuum expectation value (vev) by moving to this new minimum, but by doing so, breaks the symmetry of the system. The fact that the lowest energy state moves from the origin to this new value is known as spontaneous
Figure 1.6: The shape of the Higgs potential in two dimensions when $\mu^2$ and $\lambda$ have opposite signs. The origin is no longer the location of the absolute minimum configuration, with the lowest energy state now located at $\sqrt{-\mu^2/2\lambda}$ \[14\].

symmetry breaking–below some energy scale, the system naturally moves towards the lowest energy state away from the origin.

The consequences of spontaneous symmetry breaking is the physics in the electroweak sector that we observe, including the masses of the gauge bosons. To see this more explicitly, the gauge is chosen such that only the neutral component of the scalar field acquires the vacuum expectation value, or

$$\langle 0 | \Phi | 0 \rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \quad \text{with} \quad v = \sqrt{-\frac{\mu^2}{\lambda}}$$ \[1.47\]

This choice of gauge is a little arbitrary, but in choosing this gauge, only the neutral component of the scalar doublet acquires a vev. This choice preserves the electromagnetic charge, and so the U(1) symmetry associated with electromagnetism remains a symmetry of the system. Expanding around the vev, the scalar field can be expressed as

$$\Phi(x) = \exp\left( i \frac{\sigma_i}{2} \theta^i(x) \right) \times \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}$$ \[1.48\]

where the four fields of the scalar doublet is now re-parametrized into the three $\theta^i(x)$ fields and the $H(x)$ field. These three $\theta^i(x)$ fields are the massless Goldstone boson fields
associated with spontaneous symmetry breaking \cite{15}. Given that the Lagrangian is locally SU(2)$_L$ invariant, one can choose the unitary gauge such that any dependence on $\theta^i(x)$ is rotated away, i.e. $\theta^i(x) = 0$. Imposing the unitary gauge, the kinetic term of the scalar field can be expanded using this definition of the scalar field with $\theta^i(x) = 0$ and the covariant derivative. In doing this, it is easier to redefine the three $W$ and $B$ fields in to the more familiar $W^\pm$, $Z$, and $A$ fields using the follow definitions:

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu)$$ (1.49)

$$Z_\mu = \frac{1}{\sqrt{g^2_1 + g^2_2}}(g_2 W^3_\mu - g_1 B_\mu)$$ (1.50)

$$A_\mu = \frac{1}{\sqrt{g^2_1 + g^2_2}}(g_2 W^3_\mu + g_1 B_\mu)$$ (1.51)

Then, the kinetic term becomes

$$|D_\mu \Phi|^2 = \frac{1}{2}(\partial_\mu H)^2 + \frac{1}{2}g^2_2 (v + H)^2 W^{\pm\mu}W^{-\mu} + \frac{1}{8}(v + H)^2(g^2_1 + g^2_2)Z_\mu Z^\mu$$ (1.52)

Looking at the breakdown of the kinetic term, immediately there are quadratic terms in both the $W_\mu$ and $Z_\mu$ fields, but none for the $A_\mu$ fields. Interpreting the coefficients in the quadratic terms as the mass terms, the masses of the $W_\mu$ and $Z_\mu$ bosons naturally come out as a relationship between the gauge couplings and the vev.

$$M_W = \frac{1}{2}vg_2$$ (1.53)

$$M_Z = \frac{1}{2}v\sqrt{g^2_1 + g^2_2}$$ (1.54)

$$M_A = 0$$ (1.55)

Thus, the three Goldstone boson modes are “eaten” up by the $W$ and $Z$ bosons, manifesting themselves in the mass of the $W$ and $Z$ bosons. More explicitly, the $W^{\mu\nu}$ and $B^{\mu\nu}$ fields originally had two degrees of freedom each, being spin 1 massless particles, and now the $W$ and $Z$ bosons each have three degrees of freedom. The remaining degree of freedom from the original complex scalar doublet $H(x)$ is now just a simple real scalar, and is dubbed the Higgs boson in the context of the Standard Model. The photon associated with the $A_\mu$ field remains massless and keeps its two degrees of freedom.
Another measurable quantity from the Higgs mechanism is the mass of the Higgs boson. The part of the Lagrangian that contains the $H(x)$ terms are:

$$\mathcal{L}_H = \frac{1}{2}(\partial_\mu H)(\partial^\mu H) - \lambda v^2 H^2 - \lambda v H^3 - \frac{\lambda}{4} H^4$$  \hspace{1cm} (1.56)

Taking the quadratic term again as the mass term, the mass of the Higgs boson is then

$$m_H = \sqrt{2\lambda v}$$  \hspace{1cm} (1.57)

To complete the picture, the Higgs field also assists in providing mass terms for the fermions that arise naturally in new Yukawa terms added to the Lagrangian. Terms of the form:

$$\mathcal{L}_Y = c_1(\bar{u}, \bar{d})_L \left( \phi^0 \right) d_R + c_2(\bar{u}, \bar{d})_L \left( -\phi^- \right) u_R + c_3(\bar{\nu}_e, \bar{e})_L \left( -\phi^0 \right) e_R + h.c.$$  \hspace{1cm} (1.58)

where the second term uses the complex conjugate scalar field to $\Phi$, are invariant under the gauge symmetry and couples the new complex scalar doublet to the fermions. Propagating the results of spontaneous symmetry breaking through the Higgs mechanism for these terms, the Yukawa terms simplify quite nicely to

$$\mathcal{L}_Y = \frac{1}{\sqrt{2}} (v + H(x))(c_1 \bar{d} d + c_2 \bar{u} u + c_3 \bar{e} e)$$  \hspace{1cm} (1.59)

And so the fermions obtain a mass in the form of

$$m_d = \frac{c_1 v}{\sqrt{2}} \quad m_u = \frac{c_2 v}{\sqrt{2}} \quad m_e = \frac{c_3 v}{\sqrt{2}}$$  \hspace{1cm} (1.60)

where $u$ stands for the up type quark, $d$ stands for the down type quark, and $e$ stands for the lepton. Neutrino masses are not accounted for in the Standard Model since there are no right-handed neutrinos in the Standard Model. The three sets of constants ($c_1$, $c_2$, $c_3$) are not determined by the Standard Model and are measurable quantities that are experimentally determined. It is interesting to point out that there is nothing restricting the up type quarks and the down type quarks to be coupled to the same scalar particle. Some extensions to the Standard Model posit that there are actually two types of Higgs
bosons at work: one for the down type quarks and the leptons and one for the up type quarks.

Another result of these Yukawa terms is that there are now three mass matrices in flavor space: one for the down type quarks, one for the up type quarks, and one for the leptons. If one were to diagonalize these mass matrices through a rotation in flavor space, one finds that in general, it takes a different rotation to diagonalize each mass matrix, resulting in three different rotation matrices. These rotations do not affect interactions with the Z boson, since interactions with the Z boson are between two quarks of the same flavor, but they do introduce extra multiplicative factors in interactions terms with the W boson. Another way to look at this is that the weak eigenstates and the mass eigenstates for a particular fermion are not the same, and so the weak eigenstate down type quarks are actually quantum admixtures of all three mass eigenstate down type quarks. This is encapsulated in the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which is just a matrix relating the down type quark weak eigenstates with those of the mass eigenstates.

\[
\begin{pmatrix}
  d' \\
s' \\
b'
\end{pmatrix} =
\begin{bmatrix}
  V_{ud} & V_{us} & V_{ub} \\
  V_{cd} & V_{cs} & V_{cb} \\
  V_{td} & V_{ts} & V_{tb}
\end{bmatrix}
\begin{pmatrix}
  d \\
s \\
b
\end{pmatrix}
\] (1.61)

Given the requirement of unitarity, the CKM matrix has three degrees of freedom in deciding the terms in the matrix, as well as one extra degree of freedom for a CP-violating phase angle.

1.6 Current Status of Standard Model

The success of the Standard Model in experimental predictions has been unparalleled in particle physics, making accurate predictions ranging from the rate of different processes relative to each other to the anomalous magnetic moment of leptons to high precision. While a full comprehensive collection of all the experimental results predicted by the Standard
Model would be too extensive, a few of them are placed here to give a general feel for the success of the model.

One of the most amazing things about the Standard Model is its ability to predict the rate of phenomena that are very commonplace and those that are very rare. This is shown in the production cross section plot in Fig. 1.7, where the cross section on the y axis is a measure of the rate at which certain processes occur in high energy collisions. Particularly for Fig. 1.7, the measurements were conducted by the Compact Muon Solenoid experiment (CMS) across two different runs of data collected at a total of three different center of mass energies (7 TeV, 8 TeV, and 13 TeV). The results shown here are updated for all public results up until the spring of 2020, and the plots shows an amalgamation of many different experiments measuring precisely the rate of different interactions. This plot clearly shows the breath of the results, as the gap between the cross section measurement for the inclusive cross section for W + jets, which has a cross section of $\mathcal{O}(10^{5})$pb, and the production of W + $\gamma\gamma$ with a cross section of $\mathcal{O}(10^{-3})$pb span eight orders of magnitude. All of these cross sections measurements that were sensitive enough to reach the theoretical values agree with the Standard Model prediction within two standard deviations, with most at less than one standard deviation away.

Another success of the Standard Model involve measurements for the different parameters that are not calculable from first principles in the Standard Model, which include the fermion masses, the mass of the Higgs, the gauge coupling constant strengths, and the free parameters for the CKM matrix. Even though all the traditional Standard Model parameters have been measured either directly or indirectly by experiments at this point in time, global fits are still conducted to make sure that the results are internally consistent. The GFitter group [17] does extensive fits on the different electroweak parameters, leaving out measurements for one of the parameters and determining what the experiment predicts the actual value of the parameters to be. One specific example is in determining the mass of the W boson and the top quark with respects to the mass of the Higgs boson, as shown in the left plot in Fig. 1.8. In this plot, for the fit for the W boson mass, it removes all exper-
Figure 1.7: The measured cross section of different processes in the Standard Model conducted by the CMS experiment during Run I and II as of May 2020 [16].

Experimental evidence probing the W boson mass, and then does a global fit with the remaining Standard Model parameters (without the Higgs boson mass) to predict the value of the W boson. This is also done for the top quark mass, and then the fit results produces a 2D phase space that predicts the top quark mass and W boson mass, shown in grey. Tests like these are important to check whether the result of the different experiments measuring different components of the Standard Model agree.

Another set of fits to experimentally measured quantities are for the different matrix elements in the CKM matrix. As noted in the previous section, there are four free parameters in the CKM matrix when the unitarity conditions are imposed on the matrix element values, and these values have to be measured in experiments. One way of parameterizing these degrees of freedom is through the Wolfenstein parameter set \( \{ \lambda, A, \bar{\rho}, \bar{\eta} \} \) [18, 19] that obeys the following relations:

\[
\chi^2 = \frac{|V_{us}|^2}{|V_{ud}|^2 + |V_{us}|^2}, \quad A^2 \lambda^4 = \frac{|V_{cb}|^2}{|V_{ud}|^2 + |V_{us}|^2}, \quad \bar{\rho} + i\bar{\eta} = -\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \quad (1.62)
\]
Another successful prediction of the Standard Model is none other than the existence of the Higgs boson, with its discovery in 2012 and many of its properties measured throughout the past decade. To date, the collider experiments at the Large Hadron Collider have unambiguously discovered the Higgs boson in many of its various production modes, like gluon-gluon fusion, vector boson fusion, and associated production with a vector boson. Furthermore, the Higgs boson has been discovered through various decay modes, including to two tau leptons and to two b quarks for the first time between 2017 and 2018. A summary of the coupling constants for the production and decay modes for the Higgs boson

Figure 1.8: Fits to determine three different parameters of the Standard Model (top quark mass, W boson mass, and Higgs boson mass) as a result of precision measurements of other parameters in the Standard Model (left). A visualization of the fit to determine the different parameters in the CKM matrix, with the constraints from different independent experiments shown in different shaded regions.

A global fit with these parameters was conducted in the $\hat{p} - \hat{\eta}$ plane using CKMFitter, shown in the right plot of Fig. 1.8. This fit utilizes the constraints on the different CKM matrix elements from different experiment measurements, including single top quark decays and rare meson decays, and places them all into one plot. All the measurements from the different experiments overlap nicely within one to two standard deviations. Fits like the above two continue to show that experiments show little to no significant deviations from the Standard Model, even when accounting for many different experiments in vastly different contexts.
Figure 1.9: The signal strength modifiers for the production and decay modes for the 125 GeV Higgs boson for the CMS experiment, presented as a ratio to the theoretical value predicted by electroweak theory \cite{21}, as of 2020 for the CMS experiment are shown in Fig. 1.9 and all the measured couplings are consistent with the Higgs boson as presented in this chapter with a mass of 125 GeV.
Chapter 2

Beyond the Standard Model

Having developed the framework of the Standard Model and its far-reaching predictions, this chapter begins by elucidating some of the inconsistencies between the observed physical reality and the Standard Model in Sec. 2.1. The theory of supersymmetry is then introduced as an example of beyond the Standard Model (BSM) physics that may address some of these issues in Sec. 2.2, with a focus on the particle content of the Minimal Supersymmetric extension to the Standard Model (MSSM) in Sec. 2.3. Two specific subsets of supersymmetry are then introduced, as these theories will provide the basis for the signal models sought after in this analysis, $R$-parity violating SUSY in Sec. 2.4 and stealth SUSY in Sec. 2.5.

2.1 Issues with the Standard Model

With all the success of the Standard Model, there exists a few questions and inconsistencies within the Standard Model that lead physicists to believe that it is not the final complete framework. Some of these shortcomings are related to curiosities as to why certain parameters are the way they are, i.e. why is the mass of the Higgs at the $O(100)$ GeV scale, whereas others are related to the complete absence of the prediction of certain effects, such as dark matter and dark energy. There also seems to be a few coincidences and facts in the formulation of the Standard Model that are not highly motivated, such as why are there
three generations of fermions, why are there only three colors in quantum chromodynamics, why is only the electroweak interaction chiral, and why is the electric charge quantized in units of e/3? The combination of these questions and the known discrepancies in the following section contribute to the zeal in which the physics community continues to push the boundaries of particle physics, despite the success of the Standard Model.

2.1.1 Gravity

One of the most glaring issues with the Standard Model is that it only provides a framework for three of the four observed fundamental forces: the electromagnetic, the weak, and the strong force, leaving out gravity completely. In fact, it may even be surprising that, given the existence of gravity and its importance for effects on the macroscopic scale, so many of the Standard Model predictions are so accurate and do not require corrections because of gravity, since energy concentrated at very small distances can warp space time. This is not the case, however, as it turns out that the gravitational force is orders of magnitude weaker than the other forces explained by the Standard Model. Thus, it is natural to ask both (1) why is gravity so much weaker than the other fundamental forces (the hierarchy problem) and (2) is it possible to both have the framework of general relativity and a framework of quantum gravity that is consistent. Attempts to unify gravity with the Standard Model generally need the energy scales of spontaneous symmetry breaking and the cosmological constant to be about the same [23], but these two constants are $10^{56}$ orders of magnitude apart. There are some other attempts to bring in gravity to the Standard Model with superstring theories and effective theories, such as the minimal supergravity models (mSUGRA) [24], with the latter currently sought after in experimental studies at the LHC.

2.1.2 Dark Energy and Dark Matter

Somewhat related to the gravity problem is just the absence of a significant portion of the universe in the description of the Standard Model. Experiments observing the universe in
the cosmos have demonstrated that the composition of the vast majority of the universe is made up of dark energy and dark matter. From observations on the cosmic microwave background to those involving supernovae, it seems that the energy associated with the acceleration of the universe is driven by dark energy \cite{25,26}, and increasing evidence, such as the observed rotation curves of galaxies \cite{27}, suggests the existence of dark matter. Results from the Planck collaboration results in 2018 \cite{28} estimates the relative proportions of dark energy, dark matter, and regular matter to be around 69% dark energy, 26-27% dark matter, and about 4-5% regular matter. Since the Standard Model can only make predictions about regular matter, this leaves over 90% of the universe unexplained.

2.1.3 Matter-Antimatter Asymmetry

Focusing just on the 5% of matter that is observable, another natural question arises as to why much of what is observed is made up of matter and not anti-matter. The majority of theories involving the Big Bang is such that during the inception of the universe, equal amounts of matter and anti-matter should have been created. Given that annihilation requires equal amounts of matter and anti-matter, there has to exist some mechanism to explain why matter dominates in the universe over anti-matter instead of just a universe of just uniform energy.

According to Sakharov in his 1967 paper \cite{29}, there are three conditions that are necessary in order for there to be matter-antimatter asymmetry in the universe:

- The universe had to go through a period where it was not in thermal equilibrium, so that one process can dominate and the reverse process was suppressed

- There must exist an interaction that violates conservation of baryon or lepton number

- There must be an interaction that particles undergo that violates CP symmetry

Of the three conditions, only the last condition is found in the Standard Model, with CP violation discovered in rare meson decays. However, the CP violation in the Standard Model is generally agreed to be too small to explain the matter dominance, and so it is accepted
that there needs to be further sources of CP violation. With respects to the first condition, there do exist some theories that have interaction terms that do not conserve baryon and lepton number, such as in supersymmetry and certain Grand Unification Theories \[23\], and many of these theories are prime candidates for BSM physics.

2.1.4 Neutrino Masses/Oscillations

One experimental observation that contradicted the Standard Model during its inception was neutrino oscillations. It was observed in several experiments \[30–32\] that a beam of neutrinos of the same flavor produced at one source will have an admixture of all three flavors of neutrinos when measured at a distance away from the production source. The best explanation for why neutrino flavors seem to change is similar to the explanation of how the weak force can change the flavor of quarks: the mass eigenstates and the weak eigenstates of the neutrinos must not be the same. This implies, however, that at least two of the neutrinos must have mass, since the oscillation amplitude calculated in this case is proportional to the \(\sin^2(a(L, E)\Delta m^2)\), where \(a\) is a function of the length between measurements and \(E\) is the energy.

Now, the fundamental question is through what mechanism do these neutrinos obtain mass and why are these masses so much smaller than the other fermions? One set of theories tries to explain the lightness of the neutrinos through the seesaw mechanism. In the seesaw mechanism, a sterile right-handed neutrino is introduced for each flavor left-handed neutrino. When the mass matrix for these neutrino pairs is diagonalized, one of the eigenvalues gets larger as the other eigenvalue gets smaller, resulting in a very light left-handed neutrino and a very heavy right-handed neutrino, which can have masses approaching the Planck scale. Since the sterile right-handed neutrinos are heavy, they may be inaccessible through direct production in collider experiments, but many different experiments, such as COBRA and CUORE, are searching for indirect evidence for these theories through searches for neutrinoless double beta decay \[33\].
2.1.5 Hierarchy Problem

One interesting feature of the Standard Model is the so-called “hierarchy problem,” which fundamentally asks the question: why is the electroweak energy scale $O(100)\text{ GeV}$, but the Planck scale is $O(10^{19})\text{ GeV}$? If the Standard Model was truly the complete theory for all energy scales until the need for quantum gravity (which is assumed to be at or near the Planck scale), then why is the electroweak scale not also the Planck scale? Fundamentally, the disparity between the two scales is not in itself a problem with the model, but it does address more of a general disbelief that the universe is the way it is because of some very “fine-tuned” parameters. There are some proponents of the anthropic principle that claims that there is some inherent randomness in determining the free parameters of the model of the universe, and it may just be random chance that the current universe is the way it is, but this is overwhelmingly cynical, as there are different classes of theories that can help to explain the differences in a more “natural” way. One example of a “natural” scale difference of many orders of magnitude is with the mass of the proton and the Planck scale, where the difference in scale can be understood through dimensional transmutation [34].

There are some reasons to expect that this hierarchy problem can prove to be an issue in the calculations of the masses of scalar particles, particularly the Higgs boson in the Standard Model. While the bare mass of the Higgs boson is not a calculable quantity in the Standard Model, as it is just one of the experimentally determined constants, one can use the Standard Model framework to calculate corrections to the Higgs mass term arising from one-loop diagrams. Of these corrections, one can imagine dividing them up into three different categories, one-loop corrections arising from the coupling of the Higgs boson to itself, the coupling of the Higgs boson to the fermions, and the coupling of the Higgs boson to the vector bosons (W and Z bosons), as shown in Fig. 2.1.

If the one-loop calculations are calculated out for the three terms, the parameters are about:

$$\Delta m_h^2 = \left(6\lambda - 6y_f^2 + \frac{1}{4}(9g_1^2 + 3g_2^2)\right)\frac{\Lambda^2}{32\pi^2}$$  \hspace{1cm} \text{(2.1)}$$

where $\lambda$ is the quartic coupling of the scalar potential for the Higgs field, $y_f$ are the Yukawa...
couplings, $g_1$ and $g_2$ are the weak gauge boson couplings, and $\Lambda$ is the UV scale cutoff introduced to regulate the loop integral. $\Lambda$ can also be interpreted as the scale in which there is new BSM physics. Plugging in the Planck scale for $\Lambda$, the corrections terms to the mass of the Higgs boson is on the order of $10^{38}$ GeV$^2$ with the bare mass being on the order $10^4$ GeV$^2$, which is a vast difference.

On the other hand, one can imagine that this might be a problem for the other particles in the Standard Model: surely, there are one-loop corrections for the fermion propagator involving the Higgs and vector bosons as well, and so the fermion masses could be in danger of these large corrections. However, given the chiral symmetry of the fermions, these corrections only grow as the logarithm of $\Lambda$ and not $\Lambda^2$, which protects the fermion masses from the worst of these corrections.

Finally, it is important to point out that in Eq. 2.1, the coefficients for the scalar and vector boson correction terms are positive, whereas the coefficients for the fermion correction terms are negative. Thus, one way to solve the gauge hierarchy problem, as this is sometimes dubbed, is to introduce particles that have similar couplings, but are off by spin 1/2, such that these quadratic terms can cancel. In this way, the Higgs mass is not so much just a coincidence of nature, but a result of this specific cancellation that can be physically motivated. While this would involve introducing a whole new set of particles, at least one corresponding to each Standard Model particle, this concept did eventually became one of the main motivations behind supersymmetric theories.
2.2 Supersymmetry Basics

Supersymmetry \cite{24,36,37} refers to a class of theories that extends the Standard Model with many different attractive characteristics. For one, the particle content that supersymmetric models suggests can help to account for the hierarchy and fine-tuning problems mentioned in the previous section. Furthermore, a subclass of supersymmetric theories, particularly in $R$-parity conserving theories, can also potentially providing a candidate for dark matter in the form of the neutralinos. Supersymmetry also provides a mechanism in which the electroweak force and the strong forces couplings can all merge at the Planck scale.

At the very core, supersymmetry posits a space-time symmetry that relates fermions with bosons, such that for every fermion, there is a boson with all other quantum numbers constant, and vice versa. Mathematically, one can represent supersymmetry as an operator, denoted here with $Q$, that acts as an anti-commuting spinor with the following properties:

\begin{align}
Q |\text{Boson} \rangle &= |\text{Fermion} \rangle \\
Q |\text{Fermion} \rangle &= |\text{Boson} \rangle
\end{align}

These relationships suggest that this operator must carry spin $\frac{1}{2}$ as a result of conservation of spin, and thus the supersymmetric operator is fermionic in nature.

Having given a general idea of what supersymmetry posits, the next step is to turn the crank like was done for each symmetry in the Standard Model: derive a Lagrangian that involves a set of fields, define the transformations of these fields with respects to these operators, and ensure that the Lagrangian is unchanged under transformations of these operators. However, the procedure developed for the Standard Model involving introducing a gauge field and fermion fields, is a little clumsy because after all, the Standard Model Lagrangian treats fermions and gauge bosons rather differently, whereas supersymmetry points to a fundamental relationship between these two components. A more convenient way to define the action and Lagrangian of supersymmetry would thus involve defining a new algebra introducing a new object, known as the superfield, which pairs up each Standard Model particle with its supersymmetric partner into one object. In the following,
a full derivation of supersymmetry is beyond the scope of the paper, but major components
necessary for the understanding of the phenomenological aspects of supersymmetry will be
discussed. The following discussion is a summary of the more in-depth derivations that can
be found in the following resources [38–41].

First, it is often convenient to do calculations in superspace, which is a manifold that
acts as an extension of space time by adding four fermionic components to the familiar four
bosonic spacetime coordinates of $t, x, y, z$. The new superspace is labelled by the coordinates
$(x^\mu, \theta^\alpha, \theta^{\dagger}_\dot{\alpha})$, where the latter two components are constant complex anti-commuting two-
component spinors that have dimensions of $[\text{mass}]^{-1/2}$. The latter two coordinates are also
known as Grassmann variables, and since they are anti-commutating, it means that:

$$\{\theta_{i\alpha}, \theta_{j\beta}\} = 0 \quad (2.4)$$

Differentiation and integration with respects to these variables are defined by:

$$\frac{d\theta_{i\alpha}}{d\theta_{j\beta}} = \delta_{i^{\dot{\alpha}}}^{\beta} \delta_{j}^{\alpha} \quad (2.5)$$

$$\int d\theta_{i} = 0 \quad (2.6)$$

$$\int d\theta_{i\alpha} d\theta_{j\beta} = \delta_{ij} \delta_{\alpha\beta} \quad (2.7)$$

Because the Grassmann variables are anti-commutating, all terms involving Grassmann
variables can at most be quadratic, and integration and differentiation are equivalent to
each other. From this definition of superspace, a generic superfield then has the following
terms:

$$S(x, \theta, \theta^{\dagger}) = a + \theta \xi + \theta^{\dagger} \xi^{\dagger} + \theta \theta b + \theta^{\dagger} \theta^{\dagger} c + \theta^{\dagger} \sigma^{\mu} \theta v_{\mu} + \theta^{\dagger} \sigma^{\mu} \theta \eta + \theta \theta \theta^{\dagger} \zeta + \theta \theta \theta^{\dagger} \theta^{\dagger} d \quad (2.8)$$

Having defined the generic superfield and the new superspace, the next thing to define is the
algebra associated with this superspace, which is the super-Poincaré algebra with $N = 1$:

$$\{ Q_{\alpha}, Q^{\dagger}_{\dot{\alpha}} \} = -2\sigma^{\mu}_{\alpha\dot{\alpha}} P_{\mu} \quad (2.9)$$

$$\{ Q_{\alpha}, Q_{\beta} \} = \{ Q^{\dagger}_{\dot{\alpha}}, Q^{\dagger}_{\dot{\beta}} \} = 0 \quad (2.10)$$

$$[P^\mu, Q_{\alpha}] = [P^\mu, Q^{\dagger}_{\dot{\alpha}}] = 0 \quad (2.11)$$
where $P^\mu$ is the four-momenta operator that generates spacetime translations, and $Q$, and its corresponding Hermitian conjugate $Q^\dagger$, are the differential fermionic symmetry operators. The sigma matrices are the familiar Pauli matrices since the Grassmann variables are two component spinors.

One immediate consequence of supersymmetry is that since $Q$ commutes with the space time generator $P^2$, the particle and its superpartner must have the same mass under supersymmetry. This means that supersymmetry must be a broken symmetry if supersymmetry is to describe the physics that is observed since there is currently no experimental evidence of such particles at these accessible energy scales.

The next step in defining supersymmetry is to define the chiral covariant derivative, which is analogous to the space time derivatives, but for the Grassmann variables. Again the anti-commuting nature of these variables prove to be a bit of an issue, since

$$\delta_\epsilon \left( \frac{\partial S}{\partial \theta^\alpha} \right) \neq \frac{\partial}{\partial \theta^\alpha} (\delta_\epsilon S)$$

meaning that the derivative as defined in this way is not supersymmetric covariant or that the derivative acting on a superfield ends up with an object that is not a superfield. To address this, the covariant derivatives are defined as:

$$D_\alpha = \frac{\partial}{\partial \theta^\alpha} - i (\sigma^\mu \theta^\dagger)_\alpha \partial_\mu, \quad D^\alpha = -\frac{\partial}{\partial \theta^\alpha} + i (\theta^\dagger \sigma^\mu)^\alpha \partial_\mu, \quad (2.13)$$

$$\bar{D}_{\dot{\alpha}} = \frac{\partial}{\partial \theta^\dagger_{\dot{\alpha}}} - i (\bar{\sigma}^\mu \theta)_{\dot{\alpha}} \partial_\mu, \quad \bar{D}^{\dot{\alpha}} = -\frac{\partial}{\partial \theta^\dagger_{\dot{\alpha}}} + i (\theta \sigma^\mu)_{\dot{\alpha}} \partial_\mu \quad (2.14)$$

The covariant derivative also allows for the definition of one of two superfields present in supersymmetric extensions of the Standard Model: the chiral superfield, which is the type of superfields that encompasses the Higgs bosons and all of the fermions. These chiral superfields pair spin 0 fields (represented by a complex scalar $\phi$) with spin 1/2 particles (represented by a two-component fermion field $\psi$). Chiral superfields are defined in this superspace formalism as any superfield $\Phi(x, \theta, \theta^\dagger)$ that obeys:

$$\bar{D}_{\dot{\alpha}} \Phi = 0, \quad \text{chiral superfield (left-handed)} \quad (2.15)$$

$$D_\alpha \Phi = 0, \quad \text{anti-chiral superfield (right-handed)} \quad (2.16)$$
A more explicit form of the chiral superfield,

\[ \Phi = \phi(x) + i \bar{\theta}^\dagger \bar{\sigma}^\mu \partial_\mu \phi(x) + \frac{1}{4} \theta \theta \theta \theta \bar{\theta}^\dagger \bar{\theta}^\dagger \partial_\mu \partial^\mu \phi(x) + \sqrt{2} \theta \psi(x) - \frac{1}{\sqrt{2}} \theta \theta \theta \theta \bar{\sigma}^\mu \partial_\mu \phi(x) + \theta \theta F(x) \]  

(2.17)

where \( F \) is an auxiliary field that is introduced in order for the symmetry algebra to close off-shell (i.e. keep the number of bosonic and fermionic degrees of freedom equal to each other).

The second type of superfield present in supersymmetric extensions to the Standard Model are the superfields that include the gauge bosons, known as vector or real superfields. Vector superfields pair up massless spin 1 bosons with two degrees of freedom with a spin 1/2 Weyl fermion, also with two degrees of freedom. Bosons that are part of vector superfields have to be massless in order for the theory to be renormalizable, which means that the massive gauge bosons have to obtain mass through spontaneous symmetry breaking. These superfields are defined using the condition that:

\[ V = V^* \]  

(2.18)
To complete this brief introduction to supersymmetry, one must now introduce the dynamical structure within the supersymmetric Lagrangian by looking for superfield terms that are invariant to supersymmetric transformations. The problem of finding these terms may seem difficult at first, but one observation makes it clear which terms to hone in on. For any given term of the superfield, if after integrating over the fermionic coordinates, the superfield term transforms as a total spacetime derivative, then it is true that that specific superfield term is invariant under supersymmetry, as an integration over all space time of a total spacetime derivative can only differ by a constant. In renormalizable supersymmetric theories, there are two terms that are invariant under supersymmetry: the D-term and F-term. The D-term arises from real superfields, where the $\theta \theta \bar{\theta} \bar{\theta}$ component of the superfield transforms as a total spacetime derivative. Mathematically, the Lagrangian density is obtained by integrating over the Grassmann variables, such that:

$$[V]_D = \int d^2 \theta d^2 \bar{\theta} V(x, \theta, \bar{\theta}) = \frac{1}{2} D + \frac{1}{4} \partial_{\mu} \partial^\mu a$$  \hspace{1cm} (2.24)$$
where the second term will integrate to zero when integrated over spacetime. For chiral superfields, the term that is invariant under supersymmetry is the F-term, which is written as:

$$[\Phi]_F = \int d^2 \theta d^2 \bar{\theta} \delta^2 (\bar{\theta}) \Phi = F$$  \hspace{1cm} (2.25)$$
and to ensure that the action is real, normally this term is taken with its complex conjugate.

The goal now is to find the correct D and F terms for combinations of chiral and real superfields, that give the correct relationships that the particles in the Standard Model have. Recall that in the general gauge theory, there were fermion terms, interaction terms in the covariant derivative, and gauge boson self interaction terms. Equivalent version of these terms can be found in supergauge theory, where the chiral superfields play the role of the fermion terms and real superfields play the role of the gauge bosons fields.

For the equivalent fermion terms, notice that the product of chiral superfields are also chiral superfields, and so any holomorphic function $W(\Psi_i)$ has F terms that can be used in the Lagrangian. These terms appear in the MSSM Lagrangian as terms relating three
different chiral superfields with Yukawa couplings. As for interaction terms between the real and chiral superfields, one can imagine a chiral superfield that transforms under a matrix generator $T_\alpha^i$, such that

$$\Phi_i \to (e^{2i g_\alpha T^a})_i^j \Phi_j, \quad \Phi^* i \to \Phi^* j (e^{2i g_\alpha T^a})_j^i$$

(2.26)

where $\Omega^a$ is the supergauge transformation parameter that is a chiral superfield with gauge couplings $g_\alpha$. For each one of these generators, there is a corresponding vector superfield $V^a$, just like in non-Abelian gauge theory where there is a vector gauge boson associated to each generator. Given that, there are $D$ terms of the form that can be used in the Lagrangian, such that

$$L = [\Psi^* i (e^{2g_\alpha T^a V^a})_i^j \Psi_j] D$$

(2.27)

Finally, there are real superfield kinetic terms and self-interaction terms of the form:

$$\mathcal{W}_a = \frac{1}{4} \bar{D} D (e^{-2g_\alpha T^a V^a}) D (e^{2g_\alpha T^a V^a})$$

(2.28)

$$L = \frac{1}{4} [\mathcal{W}_{\alpha \beta} \mathcal{W}_{\alpha \beta}]_F$$

(2.29)

### 2.3 Minimal Supersymmetric Standard Model

The Minimal Supersymmetric Standard Model, known as the MSSM, is the natural extension to the Standard Model after imposing supersymmetry with a more than doubling of the particle content. Each fermion in the Standard Model, as well as the two Higgs bosons, are now included in chiral superfields, as shown in Tab. 2.1, whereas each original gauge boson in the Standard Model are now included into vector superfields, summarized in Tab. 2.2. Tab. 2.1 defines all of the chiral superfields in terms of left-handed Weyl spinors, and so the right-handed quarks are listed here through their conjugates.

Supersymmetric partners of Standard Model particles are often represented by a tilde over the corresponding symbol representing the Standard Model particle. While the supersymmetric partners of the fermions are spin 0, and thus are not left handed or right handed, it is common to keep the designation of their Standard Model partner in the symbol.
Supersymmetric partners of the fermions are called squarks and sleptons, where the “s” in front stands for scalar, whereas the fermion partners of the bosons append “-ino” to the name, collectively known as gauginos. In the MSSM, there are actually two Higgs superfields, one for the up type fermions with hypercharge $Y = +1/2$ and one for the down type fermions with hypercharge $Y = -1/2$. This is necessary for two reasons: (1) in order for the anomaly contributions from the higgsinos to cancel, the trace must be zero, and so there must be higgsinos of opposite hypercharges and (2) the conjugate field of the scalar Higgs used in the Standard Model when promoted into a superfield only has the right chirality to give masses to either the up type or down type quarks.

Table 2.1: The chiral supermultiplets in the Minimal Supersymmetric Standard Model, with their spin 0 components, spin 1 components and their Standard Model transformations

<table>
<thead>
<tr>
<th>Chiral</th>
<th>Spin 0</th>
<th>Spin 1/2</th>
<th>$SU(3)_c, SU(2)_L, U(1)_Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>squarks,quarks</td>
<td>$Q$</td>
<td>$(\bar{u}_L, \bar{d}_L)$</td>
<td>$(u_L, d_L)$</td>
</tr>
<tr>
<td>($\times 3$ families)</td>
<td>$\bar{U}$</td>
<td>$\bar{u}_R^+$</td>
<td>$u_R^+$</td>
</tr>
<tr>
<td>sleptons,leptons</td>
<td>$L$</td>
<td>$(\bar{\nu}, \bar{e}_L)$</td>
<td>$(\nu, e_L)$</td>
</tr>
<tr>
<td>($\times 3$ families)</td>
<td>$\bar{E}$</td>
<td>$\bar{e}_R^+$</td>
<td>$e_R^+$</td>
</tr>
<tr>
<td>Higgs, higgsinos</td>
<td>$H_u$</td>
<td>$(H_u^+ H_u^0)$</td>
<td>$(\tilde{H}_u^+ \tilde{H}_u^0)$</td>
</tr>
<tr>
<td></td>
<td>$H_d$</td>
<td>$(H_d^0 H_d^-)$</td>
<td>$(\tilde{H}_d^0 \tilde{H}_d^-)$</td>
</tr>
</tbody>
</table>

Table 2.2: The vector superfields in the Minimal Supersymmetric Standard Model, with their spin 1/2 components, spin 1 components and their Standard Model transformations

<table>
<thead>
<tr>
<th>Gauge</th>
<th>Spin 1/2</th>
<th>Spin 1</th>
<th>$SU(3)_c, SU(2)_L, U(1)_Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>gluino, gluon</td>
<td>$\tilde{g}$</td>
<td>$g$</td>
<td>$(8, 1, 0)$</td>
</tr>
<tr>
<td>winos, W bosons</td>
<td>$\tilde{W}^\pm$</td>
<td>$W^\pm$</td>
<td>$(1, 3, 0)$</td>
</tr>
<tr>
<td>bino, B boson</td>
<td>$\tilde{B}^0$</td>
<td>$B^0$</td>
<td>$(1, 1, 0)$</td>
</tr>
</tbody>
</table>

The next step to define the MSSM is to give the superpotential:

$$W = y_u \bar{U}_R Q_L H_u - y_d \bar{D}_R Q_H d - y_L \bar{E}_L L_H d + \mu H_u H_d$$ (2.30)
where $y_u, y_d, y_e$ are the Yukawa coupling parameters with the last term explicitly written as $\mu (H_u)_\alpha (H_d)_\beta \epsilon^{\alpha \beta}$. The superpotential is the superfield equivalent of the Yukawa terms in the Standard Model, and the interactions involving these terms can be derived by:

$$\mathcal{L}_{int} = -\sum_i \left| \frac{\partial W}{\partial z_i} \right|^2 - \frac{1}{2} \sum_{ij} (\bar{f}_i \epsilon^{\alpha} \frac{\partial^2 W}{\partial z_i \partial z_j} f_j + \text{h.c.})$$

(2.31)

Some of the interactions that are in the MSSM with the Yukawa term for the top quark ($y_t$) are shown in Fig. 2.2. The first term is the familiar Yukawa-like term for the top quark, with the Higgs for the up-type quark coupling to it. Then, the next two terms are the Higgs couplings to the quark and the supersymmetric scalar top quark, or the top squark. Finally, there are some quartic couplings of the top squark, either to the Higgs or to each other, not unlike the quartic Higgs scalar couplings seen in electroweak theory.

Figure 2.2: Some of the tree level diagrams for interactions involving the Yukawa term for the top quark [42].

The gauge field kinetic terms in the MSSM are similar to those in the Standard Model, other than a few extra terms to ensure supergauge invariance, and is given by:

$$\mathcal{L}_{gauge} = -\frac{1}{4} G_{\mu \nu}^a G^{a \mu \nu} + \tilde{G}^{i a} i \overline{\sigma}^\mu D_\mu \tilde{G}_a + f^{i a} i \overline{\sigma}^\mu D_\mu f_i - (D_{\mu} \phi_i)^i D^\mu \phi^i$$

(2.32)

where $G$ is a gauge boson field similar to $F_{\mu \nu}$. $\tilde{G}$ is the corresponding gaugino, $f$ is the chiral fermion, and $\phi$ is a scalar field.
There are two more types of terms in the MSSM Lagrangian, provided here for completeness:

$$\sqrt{2}g\tilde{G}\phi^*T^a f + h.c. + \frac{g^2}{2}|\phi^*T^a\phi|^2$$  

(2.33)

where $T^a$ are the generators of the gauge field associated with the superfield of the gaugino and its Standard Model gauge boson. These terms are the explicit terms for the fermion and boson interactions with the gauge bosons.

Having now introduced all the particles and interactions, one can now discuss some of the phenomenology of the MSSM. For one, the gauge hierarchy problem is now “solved” with the MSSM framework. Since there are equal number of bosonic and fermionic degrees of freedom in the MSSM, and since the supersymmetric partners of all the fundamental particles are coupled to the Higgs with the same couplings (i.e. the top squark and the top quark both have the same Yukawa to the up-type Higgs), their relative contributions to the Higgs mass squared correction cancel out. The issue, however, is that in the pure MSSM model, all of the supersymmetric partners are required to have the same mass as their Standard Model counterparts, and to date, none of these particles have been discovered even though these states should be accessible. This can only mean that there must be some mechanism that breaks supersymmetry in the universe in order for these particles to not have been discovered so far.

Unfortunately, the lack of evidence of supersymmetry also means that there has not been any evidence pointing to how supersymmetry is broken. There are a few different candidate theories describing how supersymmetry can be broken, including mediated by gravity or mediated by some extra gauge fields in hidden sectors [24,13], but in these cases, particles associated with these forms of supersymmetry breaking have also eluded detection. Thus, most times, to proceed despite not having a definitive mechanism, particle physics looks at the general supersymmetric Lagrangian and determines what form these supersymmetry breaking terms can take and still have the quadratic divergences associated with the Higgs mass squared corrections cancel, at least to a significant amount. This is known as soft
supersymmetry breaking, and the terms of soft supersymmetry breaking are given as:

\[ \mathcal{L}_{soft} = - \left( \frac{1}{2} M_a \lambda^a \lambda^a + \frac{1}{6} a^{ijk} \phi_i \phi_j \phi_k + \frac{1}{2} b^{ij} \phi_i \phi_j \right) + \text{c.c.} - (m^2)^i_j \phi^i \phi^j \]  

(2.34)

with the \( M_a \) matrix holding the gaugino masses for each gauge group, and the other couplings are either to quadratic or cubic terms with respect to scalar fields. All of these terms have been explicitly shown to ensure that the quadratic divergences still cancel [44], being positive in mass dimension and contributing at most logarithmic corrections to the Higgs mass correction.

Whatever the mechanism for supersymmetric breaking, the constraints for the MSSM as it is currently described in this section from experiments in the LHC is getting tighter and tighter. Some theorists may even argue that this version of supersymmetry is no longer viable without the existence of additional fields beyond any that could engender supersymmetry breaking. The next two sections discuss two such models which will also act as the signal models in the analysis that follows: \( R \)-parity violating SUSY and stealth SUSY.

### 2.4 \( R \)-Parity Violating Supersymmetry

There is an additional global symmetry of the MSSM that has not been discussed yet with the now familiar \( U(1) \) gauge group. This symmetry is known as R-symmetry [45], and is related to a phase change to the anti-commuting coordinates such that:

\[ \theta \rightarrow e^{i \alpha} \theta \quad \theta^\dagger \rightarrow e^{-i \alpha} \theta^\dagger \]  

(2.35)

Unfortunately, supersymmetry is not compatible with a global continuous R-symmetry, but it is compatible with the discrete version of the symmetry (\( Z_2 \)), dubbed \( R \)-parity symmetry with \( \theta \) having an \( R \) charge of +1 and \( \theta^\dagger \) having an \( R \) charge of -1. Propagating this to the different chiral and real superfields, the discrete symmetry \( R \)-parity is then mathematically equivalent to:

\[ \text{R-parity} = (-1)^{3(B-L)+2s} \]  

(2.36)
where $B$ is the baryon number, $L$ is the lepton number, and $s$ is the spin of the particle. All Standard Model particles have an $R$-parity of $+1$ and all supersymmetric particles have $R$-parity of $-1$.

The reasons behind introducing $R$-parity symmetry conservation is twofold: (1) it generates a particle that can be a dark matter candidate, and (2) it is able to help the MSSM avoid experimental constraints that $R$-parity violating terms have. For the first point, one can imagine that in the MSSM, if $R$-parity is conserved, the lightest supersymmetric partner would have a $R$-parity of $-1$ and not be able to decay. This particle would have mass and interact gravitationally since it has to be a part of the superfield, but could be neutral with respects to the other Standard Model forces. Under constraints on searches for anomalous heavy nuclei, this dark matter candidate cannot interact strongly or electromagnetically, but can interact weakly [46]. While such a particle may not seem likely in the MSSM given that the superpartners have the same charges as their Standard Model counterparts, there is actually a linear combination of the gauginos and Higgsinos that do meet this requirement. Particularly, these particles are termed the neutralinos, denoted by a $\chi^0$, and are defined as:

$$\chi^0 = \alpha \tilde{B} + \beta \tilde{W}^3 + \gamma \tilde{H}_u + \delta \tilde{H}_d$$

with $\tilde{B}$ being the bino, $\tilde{W}^3$ is the wino, and $\tilde{H}_u$ and $\tilde{H}_d$ are the higgsinos. The relative masses of the neutralinos depend on the exact mixing of the different gaugino and Higgsino fields, as well as the vacuum expectation values of the two Higgs superfields in the MSSM, but in general, the dark matter candidate is the lightest of these neutralinos.

As for the second point, one can imagine adding $R$-parity violating terms to the MSSM superpotential in Eq. 2.30, where all the terms currently conserve $R$-parity:

$$W_{RPV} = \lambda_{ijk} L_i L_j \tilde{E}_k + \lambda'_{ijk} Q_i L_j \tilde{D}_k + \lambda''_{ijk} \tilde{U}_i \tilde{D}_j \tilde{D}_k + \mu_{ij} L_i H_u$$

The issue with the majority of these terms is that they generate interactions for Standard Model particles that are not observed, and are in fact highly constrained. For one, each one of these terms violate lepton or baryon number, with the first two terms and the last term
violating lepton number by one and the third term violating baryon number by one. No lepton number or baryon number violating processes have been observed in the Standard Model.

One specific example of this is in the context of proton decay, with the Feynman diagram depicted in Fig. 2.3. In $R$-parity violating proton decay, the up and down quark interacts with a virtual down squark through a $\lambda''$ coupling, which then decays into a positron and a pion through an interaction with a $\lambda'$ coupling. This interaction ($p \to e^+ \pi^0$) violates both baryon and lepton number, and measurements conducted by different experiments, including SuperK, looking for this process places a limit on the lifetime of a proton to be around $10^{35}$ years [47, 48]. This measurement itself highly constrains the cross section of the $p \to e^+ \pi^0$ event occurring, which in turn forces the corresponding $\lambda''$ and $\lambda'$ coupling to be very small.

Figure 2.3: Proton decay through two $R$-parity violating couplings, where the first interaction is characterized by the $\lambda''$ coupling and the second interaction is characterized by a $\lambda'$ coupling [49].

This does not mean that $R$-parity violating versions of supersymmetry are not important and not sought after. In fact, one of the signal models in the analysis is based on $R$-parity violating supersymmetry through the $\lambda''$ couplings. To get around the experimental bounds on the lifetime of the proton decay or other experimental bounds, it is often the case to either set the couplings to be very small or have all but one coupling be non-zero. While a bit ad-hoc, this procedure is not necessarily an issue, since the superpotential can have one, some, or none of these terms, and there is no hard rule that all these couplings have to
be nonzero. The appeal of many $R$-parity violating models is that limits on these models are not as stringent and that they do still cancel out the quadratic divergences to the scalar mass squared corrections. The major downside is that these models normally also give up the possibility of a direct stable dark matter candidate in the model.

The $R$-parity violating model used for this analysis looks at top squark pair production, where each top squark decays into a neutralino and a top quark, potentially through a Higgsino-squark-quark coupling shown on the top row of Fig. 2.2 where the Higgsino acts as the neutralino. Subsequently, the neutralino decays via the $\lambda''$ coupling into three different quarks, denoted in the Feynman diagram shown in Fig. 2.4 as $j$ for jets. The neutralino is taken to be 100 GeV for the signal models in this analysis.

![Feynman diagram](image)

Figure 2.4: An example of a top squark decay into a top quark and a neutralino, and the neutralino decaying through an $R$-parity violating coupling to three jets.

### 2.5 Stealth Supersymmetry

Another set of theories tries to explain why supersymmetric partners have not been observed so far hypothesizes that maybe these particles are decaying to particles that are mostly invisible under the interactions of the Standard Model. In other words, one can imagine some sector of particles that are lighter than the supersymmetric partner particles, such that the supersymmetric partner particles decay to particles in this hidden sector preferentially.
The preferential decay to this hidden sector could be imposed, for example, by requiring that $R$-parity be conserved, and introducing $R$-parity odd particles in the hidden sector. In this case, even if supersymmetric partners are produced in abundant amounts, they could be decaying into hidden particles. This set of theories are collectively known as hidden sector theories, and there are many different flavors of these theories.

The main attractive quality of introducing a hidden sector is that it allows for supersymmetry in the MSSM to remain mostly unchanged, while at the same time, explain the lack of evidence for them in supersymmetric particle searches. Recall that in supersymmetry with $R$-parity conservation, the lightest of the supersymmetric partners is stable since they are the lightest $R$-parity odd particles. This lightest supersymmetric particle (LSP) can leave particle detectors without interacting with the rest of the detector. With this in mind, traditional supersymmetry searches try to search for these particles by reconstructing all particles from an interaction and imposing conservation of energy and momentum, hoping to find an anomalously large number of events with large missing energy that may indicate the existence of these particles. As of right now, these experiments have not discovered any excess in events with large missing energy, which places stringent limits on supersymmetry.

This, however, is also a problem with hidden sector theories, since if the supersymmetric particles decay only into particles that reside in an undetectable sector, then it would also manifest itself as missing transverse energy in collider experiments. To account for this, hidden sector theories often have a mechanism by which the hidden sector couples back to the visible sector and the Standard Model particles. One type of these theories is stealth supersymmetry (stealth SUSY) [50], which posits a very distinct hierarchal structure in this hidden sector such that the majority of the energy returns back to the visible sector.

The simplest version of this structure of the hidden sector in stealth SUSY, known as the stealth sector is shown in Fig. 2.5. In this version, a chiral superfield is introduced with the bosonic component of the superfield being $R$-parity even and the fermionic component of the superfield being $R$-parity odd., The boson and fermion components are known as the singlet and the singlino, respectively. A gaugino (represented in the diagram as a gluino)
produced through high energy collisions decays into the singlino since it is $R$-parity odd and a gluon. Because supersymmetry approximately holds in this stealth sector, the singlino and the singlet is approximately the same mass, but the singlino is a little bit more massive than the singlet. Therefore, the singlino decays to the singlet and a very light $R$-parity odd particle, which in this figure is the gravitino, but really could be any super light $R$-parity odd particle (or one of a set of superlight $R$-parity odd particles). The singlet, which is $R$-parity even, then decays back into the visible sector, in this case, chosen to be two gluons.

![Figure 2.5: The sequence of decays in stealth SUSY theories with the minimal number of particles](image)

The major difference between stealth SUSY and other SUSY theories, like compressed spectra SUSY, is that the lightest $R$-parity odd particle can be tuned to be very light. For this analysis, the mass of this particle was set to be 1 GeV, whereas the mass of the singlino was set to be 100 GeV. Because the mass splitting of the singlino and singlet is small and the mass of the lightest $R$-parity odd particle is basically negligible, the vast majority of the energy that enters the hidden sector returns to the visible sector through the decay of the singlet. In reconstruction of events, then, there would be no substantial missing energy detected.

From a mathematical perspective, the stealth sector superpotential is given by:

$$W = \frac{m^2}{2} S^2 + \lambda S Y \bar{Y} + m_Y Y \bar{Y}$$

(2.39)

where $S$ is the chiral superfield, and $Y$ and $\bar{Y}$ are two additional chiral supermultiplets.
added as vector like messenger fields that transform as $5 + \bar{5}$ under the Standard Model SU(5). The development of the mathematical properties of the SU(5)$_{GUT}$ group is beyond the scope of the dissertation, but the credit to a single group representation of the Standard Model is often given to Georgi and Glashow [51]. For the scope of this dissertation, it is sufficient to know that the addition of these two messenger fields is necessary to allow for the gaugino to decay into the stealth sector and the singlet to decay back into the visible sector, as shown in Fig. 2.6. With $m_Y$ at around the TeV scale, the decays of the gauginos and the singlet are prompt and should be observable in collisions at the LHC.

![Diagram](image)

Figure 2.6: The decay of the a bino to a singlino and a photon and the decay of a singlet to two gluons, both mediated by one-loop interactions of the $Y$ and $\bar{Y}$ vector-like messenger fields [50].
Chapter 3

The Experiment

Since the main analysis in this dissertation utilizes data collected from proton-proton (pp) collisions in the Large Hadron Collider and measured in the Compact Muon Solenoid experimental apparatus, this section will provide an overview of the collider in Sec. 3.1 and the CMS detector in Sec. 3.2. Then, the trigger system is described, including the triggers used in this analysis in Sec. 3.3.

3.1 Large Hadron Collider (LHC)

The Large Hadron Collider is currently the world’s largest particle accelerator and collider, reaching energies for its center-of-mass collisions at around 14 TeV. Located on the border of Switzerland and France, the maintenance and operation of the collider is conducted by the European Organization for Nuclear Research (CERN) and its affiliated worldwide partners. The Large Hadron Collider itself occupies a 26.7 kilometer long tunnel approximately 100 meters underground that used to house the Large Electron Positron (LEP) collider.

Consisting of a sequence of alternating straight and curved sets of superconducting magnets for most of its length, the Large Hadron Collider is constructed to guide two proton beams, one clockwise and one counterclockwise, around the ring, with the beams crossing at four different points. Situated at one of these points, at Point 5, is the Compact Muon Solenoid (CMS) detector.
3.1.1 Proton-Proton Collider

At design conditions, the LHC provides around 600 million $pp$ collisions per second with a luminosity ($\mathcal{L}$) on the order of $10^{34}\text{cm}^{-2}\text{s}^{-1}$. One of the defining aspects of the LHC is that it collides protons with protons, which deviates from both prior large collider experiments: Tevatron, which collided protons with antiprotons, and LEP, which collided electrons and positrons. This decision was made at the accelerator’s conception in order to achieve higher collision energies, trying to push the energy frontier as high as possible and to maximize the possibility of new discoveries. Given that electrons lose a lot of energy through bremsstrahlung because of their low masses and antiprotons are relatively unstable, limiting the amount of time able to accelerate and bunch them together before collisions, $pp$ collisions became the optimal choice for the collider. The LHC also operates by colliding lead ions, but these collisions tend to be much messier and more difficult to trigger on and reconstruct.

Having decided on $pp$ collisions, there were many technical hurdles to building an accelerator that could generate large equal and opposite magnetic fields (up to 9T) in close proximity to each other for the full length of the collider. This was necessary because the same tunnel had to be used for both beams and since all protons have the same charge, the magnetic fields for each beam must be opposite in order to get them to circulate in opposite directions. The engineering marvel that allowed for this are the now famous blue dipole magnets that are a symbol of CERN’s engineering capabilities, which is shown with its schematic in Fig. 3.1.
3.1.2 LHC Accelerator Complex

The overview of the entire expansive CERN accelerator complex as of the end of 2019 is shown Fig. 3.2. Besides the experiments focusing on the proton proton collisions of the Large Hadron Collider, this schematic makes it clear that there are many other experiments besides the LHC at CERN that utilizes the accelerator complex in different capacities. For example, the ISOLDE collaboration investigates different radioactive isotopes generated from a beam of high energy protons colliding with dense targets averaging about 7.5 MeV per nucleon. Similarly, the experiments associated with the Antiproton Decelerator (AD), including ASACUSA and ALPHA, studies low energy antiprotons and antimatter produced from directing a beam of protons at a block of metal.

For the $pp$ collisions, however, the first step that leads to the high energy collisions starts by passing hydrogen gas through a large electric potential in order to strip the gas of its electrons to generate protons. These protons are then accelerated by a linear accelerator, more specifically LINAC2, until they obtain an energy of around 50 MeV over a length of approximately 33m. After reaching those energies, the protons are ejected into the Proton Synchrotron Booster (labeled just as Booster in Fig. 3.2), where they are accelerated until they reach around 1.4 GeV. From there, the protons move to successively larger accelerator...
Figure 3.2: The entire CERN accelerator complex, including the Large Hadron Collider [54].
rings, the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS), where they reach energies of 25 GeV and 450 GeV, respectively.

Finally, the protons are then transferred to the Large Hadron Collider, where they are finally accelerated until they reach 6.5 to 7 GeV of energy while circulating around the ring many times. The protons obtain their energies through acceleration by the eight RF cavities, which are superconducting and made of niobium sputtered onto copper and cooled to 4.5K. The power of the RF cavity can be as high as 2 MV per cavity and is delivered by klystrons through waveguides, operating at the frequency of 400 MHz. After the protons reach their intended energies, the RF cavities then serve to concentrate the protons into bunches, with over 100,000 million protons in a bunch. The bunches circulate around until a maximum of 2556 bunches circulate around the LHC before the field is tuned so that collisions occur in one of four interaction points, corresponding to the four primary experiments with different physics programs: CMS, ATLAS, ALICE, and LHCb. Since there are two beams, each of 6.5 to 7 GeV, the overall center of mass energy of the collisions is on the order of 13 - 14 TeV.

The four different experiments are as follows:

- CMS (Compact Muon Solenoid): Located at Point 5 in Cessy, France, CMS is a general purpose experiment that conducts a broad range of measurements, from precision measurements of Standard Model processes to Higgs physics to searches for new physics. The defining feature of CMS is the very large solenoid magnet that can produce a 3.8 T magnetic field.

- ATLAS: Located at Point 1, ATLAS is the other general purpose experiment, weighing 7000 metric tons and having dimensions of 46 meters in length and 25 meters in diameter. Notable for having both a solenoidal magnet and a toroidal magnet, the physics agenda of ATLAS is similar to that of CMS.

- ALICE (A Large Ion Collider Experiment): Located at Point 2, this collaboration focuses on the physics of the quark-gluon plasma and heavy-ion physics.

- LHCb: An experiment focused on analyzing the processes involving the b quark
As the beams collide, each collision lowers the number of protons in the beam, and the beam intensity begins to fall off slowly as a function of time. At a certain point, the proton density is too low to maintain a high level of collisions, and the two proton beams have to be removed from the ring; however, this is not a simple process since the particles contain enough energy to heat up the magnets and cause a quench in the system. When a superconducting magnet quenches, it can no longer provide the needed magnetic field and can get so hot that the magnet is permanently destroyed, which can then lead to further damage in the system.

To deal with a safer way to dispose of the beam, the LHC developed a beam dump system \[55\], shown in Fig. 3.3. The safe disposal of the beam, known as a beam dump, starts with the activation of the septum magnet that acts to separate the two beams of

and rare $B$ meson decays, searching for tetraquark bound states and investigating CP-violation and the matter-antimatter asymmetry.
protons from each other so that one beam can be acted on without affecting the other beam. Then, a fast kicker magnet with an extremely quick rise time of around 3 \( \mu s \) guides the beam away from the ring and towards giant graphite beam dump blocks. Before the collision with the graphite blocks, the beam passes through a series of dilution magnets, which widen the beam both vertically and horizontally, so that the beam intensity will not vaporize the graphite blocks. Finally, the beams hits the graphite blocks, whose core may reach temperatures over 700\(^\circ\)C. After the remaining proton bunches are removed, the whole accelerating cycle can begin again by generating protons from hydrogen gas once more.

### 3.2 Compact Muon Solenoid (CMS)

The Compact Muon Solenoid (CMS) detector is one of two general purpose detectors that is a part of the Large Hadron Collider (LHC) accelerator ring. The entire detector is 28.7 meters in length, 15.0 meters in diameter, and about 14000 metric tonnes, with a maximum operating magnetic field of 3.8T. Conceptually, the detector is often easiest thought of as composed of three components: a large cylinder, often referred to as the barrel, with two smaller detectors on either end of the cylinder, often referred to as the end caps and labeled as plus and minus. In concentric rings around the axis of the cylinder where the beam pipe is located are a series of subdetectors, including the tracker, calorimeters, solenoid, and muon systems, each optimized for its specific function. This is all visible in the schematic shown in Fig. 3.4, and it is the purpose of this section to delineate each component and its corresponding function. As before, this is only a subset of all the information about the detector, and for a more detailed explanation, please refer to the following sources [56–58].

On a broader scale, the detector is optimized to measure the properties of the particles that are produced in collisions at the interaction point where the beams cross. For example, the superconducting solenoid magnet and tracker allow for distinguishing between neutral particles, which move in straight paths in a magnetic field, and charged particles, which move in helical paths in a magnetic field. Since the radius of curvature of charged particles is proportional to the magnetic field, the larger the magnetic field, the better the differenti-
Figure 3.4: A schematic of the CMS detector with the major subdetectors labeled [58].
ation between charged particles. Therefore, the CMS detector optimized the magnetic field strength by building a powerful magnet and a high granularity tracker.

Another requirement of the detector is to accurately measure the energy of all particles that interact with the detector in order to be able to use conservation of energy to potentially search for particles that do not interact with the detector. Thus, the calorimeters are hermetic, meaning they completely surround the collision and prevent particles from escaping without interacting, such that the energy measurements are as accurate as possible.

Before proceeding with the description of the subdetectors, the CMS coordinate system will first be described, since many of these variables are standardized and used in the description of different components of the subdetectors. First, when looking at the CMS detector as placed on the larger ring of the Large Hadron Collider, the x-axis points radially inward towards the center of the LHC. The y-axis points vertically upward against gravity, and from the viewpoint of P5 where CMS is located, the z-axis points towards the Jura mountains.

More commonly, instead of using the x,y,z coordinate system, two angles are often used to describe particle track position and momentum direction. First, the angle $\phi$ is the angle measured in the transverse (x-y) plane from the x-axis as is defined in Eq. 3.1.

$$\tan \phi = \frac{y}{x}$$

(3.1)

The $\phi$ angle is important because the beams are focused such that collisions should happen where the net transverse momentum is equal to zero. Since protons are composite particles and the exact momentum of the colliding quarks or gluons along the beam axis can be unknown, the momentum and energy in the transverse plane is often the quantity that can provide more information. Many quantities are defined in the transverse plane, including the transverse momentum ($p_T$), which is the momentum of a particle in the x-y plane; the transverse energy ($E_T$), which is the energy of a particle in the x-y plane represented as a vector in the x-y plane; $H_T$, which is the scalar sum of the energy in the transverse plane; and $E_{T\text{miss}}$, which is the opposite of the vector sum of transverse energy of
all particles.

The second angle of interest is $\theta$, which is an angle measured from the beam axis ($z$-axis), where $\theta = 0$ is along the positive $z$-axis and $\theta = 90$ is vertically upward. However, since this angle is not a Lorentz invariant quantity and collisions can have varying total momentum on the $z$-axis leading to a varying mapping of the lab frame angle $\theta$ with respect to the collision’s rest frame angle, it is often easier to use another variable $\eta$ that is defined in Eq. 3.2. Normally referred to as the pseudorapidity, this angle is Lorentz invariant with respects to longitudinal boosts. For a more intuitive feel of how this variable is used, an $\eta$ value of 0 points directly in the transverse plane, whereas an $\eta$ of infinity points towards the beam pipe.

$$\eta = -\ln \tan \frac{\theta}{2} \quad (3.2)$$

Finally, one last variable that is often used is $\Delta R$, which is defined in Eq. 3.3. This variable measures the distance between two particles in the 3-dimensional space of the detector, and is useful in determining how isolated a particle is from everything else produced in a collision.

$$\Delta R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2} \quad (3.3)$$

### 3.2.1 Tracker System

The primary roles of the tracker are to accurately measure the curved tracks of charged particles, make precise momentum measurements, and aid in the reconstruction of secondary vertices, such as for $b$ quark decays and tau lepton decays. To make the most accurate measurements possible, the tracker system is positioned closest to the interaction point and is made with minimal material in order to avoid increasing the probability of secondary interactions, such as multiple scattering, bremsstrahlung, photon conversion, and other nuclear interactions. Ideally, the tracker will be able to make its measurements as nondestructively as possible so that all energy of outgoing particles are measured in the calorimeters after
The CMS tracker system. In the center is the pixel detector, divided into the barrel and endcap regions, and right on the outside is the silicon strip detector system, composed of six different components [56].

Another major consideration is that the tracker system is handling the large number of high energy particles that can damage the electronics. Being so close to the interactions, this subdetector takes the highest radiation and radiation dose rates of any of the subdetectors, and so there was an additional challenge of making all components as radiation hard as possible. Furthermore, large fluxes of high energy particles result in considerable heat deposited in the detector, and so special cooling systems that keep the tracker system between around -10 and -20°C also needed to be installed.

The layout of the tracker is shown in Fig. 3.5, with an overall length of 5.8 meters, a diameter of 2.5 meters, and $|\eta|$ coverage up to 2.5. The tracker itself consists of two subsystems: the pixel detector, which has a total of 1440 modules, and the silicon strip detector, which has a total of 15,148 modules. Each module has its own set of fast front end electronics and are essentially in charge of recording when and where a particle passes through the active silicon area associated with that module. This location information is known as a hit, and these hits are later pieced together through algorithms to create particle tracks that are then important in trigger decisions and determining which particles...
Figure 3.6: Comparison of the pixel detector before and after the Phase I upgrade [59].

passed through the detector. The tracker system has about 75 million read-out channels, can measure the position of a particle within 10 µm, and can measure the $p_T$ of a particle with a 1 - 2% resolution for a particle with 100 GeV momentum.

The original pixel detector was comprised of three cylindrical layers for the barrel (BPIX) spaced 4.4 cm, 7.3 cm, and 10.2 cm away from the beampipe. On each endcap, there were two silicon disks, and the combination of the four disks is called FPIX because they measured particle tracks in the forward direction. However, part of the way through the second run of the LHC, between the end of 2016 and the spring of 2017, there was an upgrade to the pixel detector at CMS. The upgraded pixel detector added an additional layer to the BPIX, and the innermost layer became closer (about 3 cm) to the beam pipe. For the FPIX, each endcap had an extra ring added, and each ring is divided into two half rings, as shown in Fig. 3.6.

The silicon strip detector consists of 10 layers in the barrel region: 4 inner barrel layers (TIB) and 6 outer barrel layers (TOB). Each endcap has 2 inner silicon strip detectors of 3 small disks each (TID) and 2 outer silicon strip detectors with 9 disks each (TEC). In total, there are over 9.3 million strips in the silicon strip detector.
Hermeticity, compactness and high granularity

- Fast response (⇠25 ns) and particle id, energy and isolation measurement at trigger level
- Large dynamic range (5 GeV to 5 TeV) and excellent linearity (at the per-mill level)
- Radiation tolerance (ECAL was designed for 14 TeV and \( L = 10^{34} \) cm\(^2\) s\(^{-1}\), and for a total luminosity of 500 fb)

In the following we discuss the challenges of operating the CMS electromagnetic crystal calorimeter at a hadron collider, in particular in achieving and maintaining the required energy resolution in the harsh radiation environment of the LHC. We summarise the role of ECAL in the discovery of the Higgs boson. We also present the prospect for the LHC Run II starting in 2015 and the challenges that ECAL will face with the High Luminosity (HL) upgrade of LHC, based on the experience gained during Run I.

2. The CMS electromagnetic calorimeter

The CMS electromagnetic calorimeter (ECAL) is a hermetic, homogeneous, fine grained lead tungstate (PbWO\(_4\)) crystal calorimeter. The choice of an homogeneous medium was made to obtain a better energy resolution by minimizing sampling fluctuations [4]. Very dense crystals offer the potential to achieve the required excellent performance and compactness.

The CMS design enabled the electromagnetic calorimeter to fit within the volume of the CMS superconducting solenoid magnet.

The 75,848 crystals are arranged in a central barrel section (EB), with pseudorapidity coverage up to |\( \eta \)| = 1.48, closed by two endcaps (EE), extending coverage up to |\( \eta \)| = 3.0.

Crystals are projective and positioned slightly off-line-pointing (⇠30) relative to the interaction point (IP) to avoid cracks aligned with particle trajectories. The calorimeter has no longitudinal segmentation, the measurement of the photon angle relies on the primary vertex reconstruction from the silicon tracker.

The crystal length in EB is 230 mm (220 mm in EE) corresponding to ⇠26 (25) radiation lengths. The transverse size of the crystals at the front face is 2.2 \( \times \) 2.2 cm\(^2\) in EB (2.86 \( \times \) 2.86 cm\(^2\) in EE). The total crystal volume is 11 m\(^3\) and the weight is 92 t. The barrel calorimeter is organized into 36 supermodules each containing 1,700 crystals while the endcaps consist of two dees, with 3,662 crystals each.

A preshower detector (ES), based on lead absorber and silicon strips sensors (4,288 sensors, 137,216 strips, 1.90 \( \times \) 61 mm\(^2\) with x-y view), placed in front of the endcaps at 1.65 < |\( \eta \)| < 2.6, improves the photon-\( \pi^0 \) separation. The total thickness of the ES is ⇠3 radiation lengths.

Figure 1. Schematic view of the CMS electromagnetic calorimeter.

Figure 3.7: A schematic of the electromagnetic calorimeter showing the different components, including the barrel, the endcap, and the preshower.

3.2.2 Electromagnetic Calorimeter

Directly beyond the tracker system radially is the electromagnetic calorimeter (ECAL), a schematic of which is shown in Fig. 3.7. This subdetector is a hermetic homogenous calorimeter that has an |\( \eta \)| coverage up to 3.0, and also consists of a barrel section and two endcap sections. This subdetector is principally in charge of measuring the energy of electrons and photons, and is optimized for this with very fine granularity provided by fast and radiation resistant electronics.

The main component in the ECAL are the 75,000 transparent lead tungstate (PbWO\(_4\)) crystals, with around 61,000 in the barrel and 7,300 in each endcap. These 22 mm by 22 mm crystals have a depth of 23 cm and oriented next to each other so that their longitudinal axes point towards the interaction point. The crystals are optically clear, are extremely dense, have a small Molière radius (meaning the majority of the electromagnetic shower is well contained), have a short radiation length of \( .89 \) cm, and have a length corresponding to 25 radiation lengths; all properties which make them ideal in capturing the energy of the electrons and photons. In fact, the engineering success and resolution provided by these
Figure 4.4: Front view of a module equipped with the crystals.

crystals are some of the major reasons the decay channel where the Higgs boson discovery was most sensitive was the Higgs to two photons decay channel.

Functionally, when a photon or electron interacts with the lead tungstate crystal, the energy deposited causes the crystal to scintillate. About 80% of the light emitted in the crystal happens within the 25 ns between bunch crossings, and this scintillated light moves through the crystal until it is converted to an electrical signal either by avalanche photodiodes (APDs) in the barrel or vacuum phototriodes (VPTs) in the endcaps. The amount of light scintillated is proportional to the amount of energy deposited, which is about 30 photons per MeV of energy deposited, and so this information is extracted from the measured scintillated light.

At the endcaps, there are an extra two planes of lead and two planes of silicon sensors in front of the electromagnetic calorimeter. These extra layers are called the ECAL preshower. The purpose of these extra layers is to aid in the differentiation between two photons coming from a neutral pion that can sometimes mimic a high energy photon if they are close enough to each other and picked up in the same crystal. This is more of an issue at the endcaps than in the barrel region because the pions in this region tend to have a much higher forward momentum, and so the resulting two photons also have a higher forward momentum. The preshower layers have a higher granularity with its 2 mm wide detector strips when compared with the 2 to 3 cm wide ECAL crystals, giving it the ability to distinguish between clusters.
generated by two photons and clusters generated by one single highly energetic photon.

3.2.3 Hadronic Calorimeter

The hadronic calorimeter (HCAL) is the subdetector following the ECAL. As its name suggests, the HCAL is designed to detect and absorb hadrons produced in the collisions and measure the energy of these particles. In order to have the capability to measure missing transverse energy, the HCAL is fully hermetic, which means that the HCAL is able to capture and measure the energy of all Standard Model other than muons, which are measured in the outer muons systems, and neutrinos.

Geometrically, the HCAL is divided into four components: the barrel region (HB), the two endcap regions (HE), the forward region (HF), and the part of the calorimeter positioned outside the magnet (HO). The size of this subdetector is necessary because hadronic showers tend to be very large and interaction lengths tend to be long, even for very dense material. The overall HCAL has between 7 to 11 interaction lengths, depending on $\eta$, but ideally this would be even higher.

The HCAL is known as a sampling calorimeter because it interleaves dense absorber brass and steel layers that increase the chance of nuclear interactions with the outgoing hadrons with fluorescent plastic scintillator tiles that produce light when particles move through them proportional to their energy deposited. When a particle moves through the scintillator, light gets emitted proportional to the energy deposited, and this light is captured by specific wavelength shifting fibers, which then guide the light towards specific devices that convert the optical signal to an electrical signal. The particular nature of the device depends on the year of data taking, as the HCAL also underwent a significant upgrade during the extended year end technical stops in 2016 and 2017. In the barrel (HB) and endcaps (HE), the optical to electrical conversion devices were hybrid photodiodes during early operation, but since these devices were damaged more significantly by the radiation than expected, they were replaced with silicon photomultipliers during the shutdowns.

In the forward region (HF), instead of scintillator, they utilize over 1000 km of quartz
fibers with steel absorber. When particles interact with the quartz fibers, they slow down and release Cherenkov light, which are then collected by the photomultiplier tubes. These devices were chosen because they are more radiation resistant and the forward region is in a more severe radiation environment.

Finally, right outside the magnet is the hadronic outer calorimeter, which has about one meter of absorber material. Since this material is beyond the magnet, it is hard to differentiate between contributions between charged and uncharged particles in this region. Therefore, this part of the detector acts more as a tail catcher for very energetic hadronic showers and are paired with corresponding energy clusters in the main part of the sub detector.

Throughout the three years of data taking, the HCAL underwent successive upgrades that improved the resolution of the subsystems. During the year end technical stop (YETS) in 2016, the HF system was upgraded by replacing all of the photomultiplier tubes with a version with thinner windows. An issue with the original photomultiplier tubes were that stray muons from cosmic rays were interacting with the glass windows of the photomultiplier tubes and producing Cherenkov light, mimicking very high energy deposits. These hits arrived to the backend earlier than the information from the rest of the interaction, and so in Run I, proper event vetoes could be developed upon observation of the timing of these hits, especially since there were empty bunch crossings between each collision. In Run II, however, the increased luminosity and the fact that every bunch crossing contained a collision made it more of a challenge to distinguish between stray signals and those arising from particles created at the interaction point, so the photomultipliers, as well as all of the front-end electronics associated with the old photomultiplier tubes were replaced. The electronics had to be upgraded partially because the new photomultiplier tubes had an anode that was divided into four sectors, which allowed further differentiation between stray signals where only one of the four sectors would produce a hit, and real signals, where most of the anodes would produce a hit.

During the YETS in 2016, one sector of the HE system was upgraded, followed by the
The laser light was injected during the run at times without collisions, and a light distribution system that delivers the UV light to the scintillator tiles was inserted into the brass absorber of the HE. The tiles were megatiles, with sizes ranging from 0.2 cm to 20 cm in the rest of the layers. The sizes range from ro 0 and 0.37 cm in layer 1 and ro 0.15 cm to about 20 cm in layers to L7. A large signal loss, especially at high ieta, is seen both in L1 (35% vs 13%), better performance under a high magnetic field, and a light distribution system that delivers the UV light to the scintillator tiles with the largest damage. The data are not corrected for integrated luminosity. After the end of the run, a few percent recovery was observed for the tiles with the largest damage. The data show that the degradation is actually a stronger function of the dose rate rather than the absolute dose of radiation, but in either case, there is strong indication that a replacement was needed as soon as possible.

The device chosen to replace the hybrid photodiode was the silicon photomultiplier (SiPM), of which an 8 by 10 array of them in a prototype readout module is shown in Fig. 3.9. These devices have the benefit of operating at a lower voltage (70V vs 10kV), higher photon detection efficiency (35% vs 13%), better performance under a high magnetic field, and sensitivity to distinguish between each photoelectron detected. To accommodate this better device, an update to the entire front end system was conducted, including the clock and control distribution that was integral in synchronizing signals and updating tunable parameters for each readout channel. With this upgrade, with a snapshot of a day during the upgrade shown on the right of Fig. 3.9, much of the measurable response loss improved.
The HB system also switched out their hybrid photodiodes for SiPMs at the end of 2018, and the hardware in the HB system is analogous to that of the HE system. The improvements to the system will be observable in Run III. Fig. 3.10 also shows the increased resolution in jet energy measurements through finer segmentation. Since the upgraded front end system can handle more information at any given time and the SiPMs are small enough to pack them more densely, the readout from each HCAL tower can be divided up into more channels. This is more easily visualized in the \( r - z \) map, shown in Fig. 3.10 for the entire HB-HE system, where the numbers at the top of each column (from 1 to 15) and the numbers on the left side (16 to 28) corresponds to a different \( i \eta \) tower in the HCAL subsystem. The \( i \eta \) numbering scheme is just a one to one mapping from \( i \eta \) of 1 corresponding to \( \eta = 0 \) all the way up to \( i \eta \) of 28 corresponding to the largest \( \eta \) covered by the HCAL system. For the endcap HCAL towers 18 to 26, the left mapping shows that during Run I, there were two channels that read out the energy measurement per tower: one for the front 5 layers (in yellow) and one for the remaining layers (in pink). On the right mapping, which shows the segmentation after the upgrade, there are now five channels reading out the energy measurement of each tower (green-yellow-blue-purple-orange), with the layers closest to the interaction point having their own channels.

Figure 3.10: The number of layers in each channel readout before the upgrade (left) and after the Phase 1 upgrade between 2017 and 2019 62.
Figure 2.1: General artistic view of the 5 modules composing the cold mass inside the cryostat, with details of the supporting system (vertical, radial and longitudinal tie rods).

magnetic pressure ($P = B^2 \mu_0 = 6.4 \text{ MPa}$), the elastic modulus of the material (mainly aluminium with $Y = 80 \text{ GPa}$) and the structural thickness ($D_{Rs} = 170 \text{ mm}$ i.e., about half of the total cold mass thickness), according to $PR_{Rs} = Ye$, giving $e = 1.5 \times 10^{-3}$. This value is high compared to the strain of previous existing detector magnets. This can be better viewed looking at a more significant figure of merit, i.e. the $E/M$ ratio directly proportional to the mechanical hoop strain according to $E/M = PR_{Rs}D_{Rs}d_{Rs}D_{Rs}Y e^2 d$, where $d$ is the mass density. Figure 2.3 shows the values of $E/M$ as function of stored energy for several detector magnets. The CMS coil is distinguishably far from other detector magnets when combining stored energy and $E/M$ ratio (i.e. mechanical deformation). In order to provide the necessary hoop strength, a large fraction of the CMS coil must have a structural function. To limit the shear stress level inside the winding and prevent cracking the insulation, especially at the border defined by the winding and the external mandrel, the structural material cannot be too far from the current-carrying elements (the turns). On the basis of these considerations, the innovative design of the CMS magnet uses a self-supporting conductor, by including in it the structural material. The magnetic hoop stress (130 MPa) is shared between the layers (70%) and the support cylindrical mandrel (30%) rather than being taken by the outer mandrel only, as was the case in the previous generation of thin detector solenoids. A cross section of the cold mass is shown in figure 2.4.

The construction of a winding using a reinforced conductor required technological developments for both the conductor [11] and the winding. In particular, for the winding many problems had to be faced mainly related to the mandrel construction [12], the winding method [13], and the module-to-module mechanical coupling. The modular concept of the cold mass had to face the problem of the module-to-module mechanical connection. These interfaces (figure 2.5) are critical—7–7–

Figure 3.11: Picture of the solenoid during installation (left) [63] and a schematic showing the five modules that make up the cold mass inside the cryostat and some of the supporting system features [56].

segmentation for the entire upgraded HB system as well

3.2.4 Magnet

The role of the CMS magnet, shown in Fig. 3.11, is to generate as strong a magnetic field as possible in order to distinguish between different particles and to calculate the transverse momentum of the particles created at the interaction point [64]. Since the path of charged particles bend within a magnetic field, this magnetic field is integral in distinguishing between two particles that have similar energy deposit pattern but with different charges, i.e. electrons and photons. Similarly, given the radius curvature of the track of the particle moving through the magnetic field, the particle’s transverse momentum can be calculated using the Lorentz force equation, where the transverse is proportional to the magnetic field strength and the radius of curvature. Therefore, the stronger the magnetic field and the more consistent the magnetic field within the volume in which the particles traverse, the more precise charge measurements and transverse momentum measurements are.

In pursuit of this ideal benchmark of the strongest and most consistent magnetic field, the CMS magnet system consists of a superconducting solenoid that can produce a uniform magnetic field of 3.8 T. The size of the magnet and the magnetic field it provides was at the limits of engineering during the construction of CMS, with the solenoid itself measuring 6.3
m in diameter and 13 m in length. The steel return yokes on the periphery have an outer diameter of 14 m. The yoke itself consists of three layers in order to have enough windings to generate the magnetic field, and the whole system weighs over 12,000 metric tons, which contributes to the magnet system playing a secondary role of being the structural support for all of the other subdetectors within the solenoid. The magnet and yoke system also acts to provide a big buffer of material that blocks any Standard Model particles produced in the interaction point other than muons and neutrinos from passing through.

During operation, the solenoid is cooled to 4.5 K and as a benchmark for the momentum resolution, for a 1 TeV particle, the momentum resolution, defined as $\Delta p/p \approx 10\%$ [64].

### 3.2.5 Muon Systems

Beyond the magnet in the CMS detector are the muon systems. The muon systems are primarily composed of three different types of devices, which all play complementary roles in identification, position and energy measurement, and fast triggering on muons. This special attention to muons in the construction of the experimental apparatus is primarily because muons have a unique signature in terms of particle detection in accelerator physics. Since muons are leptons, muons do not interact via the strong force, and so muons can penetrate through several meters of iron and the hadronic calorimeter with minimal interactions, whereas pions and neutrons are hindered. However, unlike electrons that interact with the ECAL and deposit the majority of its energy in that sub detector, muons are about 400 times heavier than an electron and do not interact as much, allowing muons to pass through the electromagnetic calorimeter relatively unscathed. Thus, at the distance from the detector where these muon systems are located, all hits and tracks are almost certainly from muons, providing a very clean signature and a great handle for searching for any process that have muons in the final decay products.

Similar to the previous subdetectors, the muon system can be subdivided by $\eta$ range in the barrel muon system ($|\eta| < 1.2$) and the endcap muon system ($1.2 < |\eta| < 2.4$), with some overlap. The barrel muon system consists of four stations, with one station mounted
on the inner face of the return yoke of the magnet, one station mounted on the outer face of the return yoke, and two stations located inside the iron yoke. The endcap muon system is similar, with the first layer mounted right after the HCAL, and then successive layers separated by steel, as shown in the schematic in Fig. 3.12. In the barrel muon system, the main active elements are the drift tubes (DT) and the resistive plate chambers (RPCs), whereas in the endcap muon system, the primary devices are the cathode strip chambers (CSCs) and RPCs. These three devices provide different, but complementary, information for the position and momentum of the muons that pass through it, allowing for quick triggering without a loss of information.

The drift tubes utilize the drift speed of an electron that is created when gas is ionized as a muon passes through it to measure the position of the muon passing through the device. The drift tubes in the muon chambers at CMS are on average 2 m by 2.5 m and have a positively charged stretched wire within a dense gas. The drift tubes themselves are arranged into three superlayers, and are placed orthogonally to each other so that the combined information can give accurate position information by triangulating all the hits.
Figure 3.13: The cathode strip chamber modules placed where they will be installed in ME1 (left) [65]. The installation of an RPC module into the endcap (right) [66].

from the different layers together to reconstruct the muons’ trajectory.

The trapezoidal cathode strip chambers, shown on the left in Fig. 3.13 are used in the endcaps in place of the drift tubes because they are more radiation resistant. These cathode strip chambers work similarly to that of the drift tubes, but they have multiple positively-charged anode wires crossed with negatively-charged copper cathode strips in a gas volume. When a muon passes through a cathode strip chamber, the gas ionizes, with the positive ions moving towards the copper and the electrons moving towards the wires. Since the copper and wires are orthogonal to each other, they once again give accurate position information of the muon. One of the differences between the cathode strip chambers and the drift tubes is that the wires in the cathode strip chambers are more closely spaced together, and so the response is a little faster than that of the drift tubes.

Finally, the last component is the resistive plate chambers, which are found in both the barrel and endcap. These devices are the quickest of the three devices, having a time resolution of just around one nanosecond, and thus is integral to the muon trigger system. These devices are generally made up of two plates that are highly resistive separated by a gas volume called the gap; however, the version of this device used in the muon systems have two gaps. As a muon traverses the resistive plate chambers, the gas is ionized and initiates a cascade of electrons that move towards the anode and detected by readout strips
made of metal. The shape of the cascade on the metal strips then provides a rough estimate of the muon’s momentum.

3.3 Trigger

Having given an overview of the hardware of the CMS detector that generates the hits, the next section focuses on how those hits are used to decide whether an event is saved to disc, which is the job of the trigger system. The trigger system for the CMS experiment holds an integral role in determining what analyses are possible in the collaboration. Since there can be over one hundred million collisions per second at design luminosity [67], there is no physical nor technological capability to store all this information. Furthermore, many of the collisions arise from well understood processes, so it is not important to keep every single event.

Thus, the trigger system effectively pares down the number of events recorded from around 1 GHz to around 500 Hz, about a six order of magnitude reduction. This is possible because of the timing resolution and synchronization of all the readout channels in the calorimeters and muons systems, as well as custom-made technologies that allow decisions to be made on the order of microseconds. Furthermore, the trigger decision is divided up into two tiers: the Level-1 (L1) Trigger System that is in charge of the first round of decisions that lower the rate to about 100 kHz in about 4 µs, and the High Level Trigger (HLT) System that is in charge of the remaining reduction.

For further in-depth details on the trigger system in CMS, please refer to [67–69].

3.3.1 L1 Trigger

The L1 trigger is the first level of triggering that has to process each and every event. Since collisions happen on the order of once per 25 ns, the electronics for the L1 trigger system are placed in a cavern next to the detector underground in order to minimize the latency and transit time between data acquisition and triggering.

Even with the physical proximity, given the wealth of information, the high resolution
information can only be kept in memory buffers for about 4 µs, giving the electronics slightly less than 1 µs to make a decision. This means that only a subset of all the information can be used in trigger and that full object reconstruction is not possible. Only information from the calorimeters and the muon systems are included in the L1, as shown in Fig. 3.14. Plans to add information from the outer tracker to the L1 Trigger for future runs are expected [70].

Figure 3.14: The flow of information in the Level 1 Trigger [71].

From the global calorimeter trigger, the trigger primitives from the ECAL and HCAL are aggregated to form basic Level 1 objects, such as electrons, tau jets, and jets, as well as calculate some higher level kinematic quantities, such as total transverse hadronic energy \( H_T \) and missing transverse energy \( E_{T\text{miss}} \). From the global muon trigger, the hits representing location and timing information of the muons from the resistive plate chambers (RPCs), cathode strip chambers (CSCs), and drift tubes (DTs), are combined to create basic muon objects. The final decision is then determined by the global trigger, which makes the decision to keep or throw away the event based on the presence of one or more of these objects meeting certain angular and momentum thresholds.

The final decision is propagated back to the front end electronics, as well as forward to the high level trigger/data acquisition.
3.3.2 HLT

The HLT system is different than the L1 trigger in that there is no set hardware close to the detector nor a set list of thresholds for objects. In general, however, the HLT object requirements are more stringent than those of the L1 trigger, as it has more time to make decisions and has access to the full event information, even if only part of the event is used in reconstruction. For example, an HLT muon in the isolated muon trigger requires a higher $p_T$ threshold, associated tracks in the tracker, and isolation in addition to the L1 requirements for a muon. Furthermore, in the process of decision making, only the parts of the muon chambers indicated by the L1 trigger results are considered for the validation of the muon.

On average, each event takes about a tenth of a second to be evaluated by the hundreds of different HLT paths, but no longer than about one second per event. The algorithm is ordered such that more computing intensive processes are completed last, in case an event is vetoed in a less computationally intensive step, optimizing for CPU time. Furthermore, HLT paths with the same object have the reconstruction of the object done only once, and this information is shared amongst the different paths, also minimizing CPU computations. Overall, the goal of the HLT trigger is to make decisions based on physics objects that are as close to the objects used in offline analysis as possible while keeping the rate manageable.

3.3.3 Analysis Triggers

Before embarking on the next section on event reconstruction and simulation, since triggers are being discussed in this section, the choice of triggers used in the analysis will be discussed here. As will be described in the next chapter, one of the choices made in this analysis was to select for physics events where a top quark decays leptonically; i.e. $t \rightarrow bW; W \rightarrow e\bar{\nu}$ or $m\bar{\nu}$. To take advantage of the existence of the lepton in the selected events, the triggers used for this analysis are the isolated and non-isolated single muon and isolated single electron triggers, shown in Tab 3.1 and 3.2.

The reason for the choice for these triggers are that they tend to have very high efficiency...
and are well tested and used by many different analyses in the collaboration. Leptonic
decays of the top quark to the tau lepton were not studied in this analysis, as the efficiency
of triggering on single isolated electron and muon events are significantly higher than that
of tau lepton triggers, and a significant proportion of tau lepton events are still taken into
account when the tau lepton decays leptonically.

<table>
<thead>
<tr>
<th>Year</th>
<th>Trigger</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>HLT_IsoMu24 OR HLT_IsoTkMu24 OR HLT_Mu50 OR HLT_TkMu50</td>
</tr>
<tr>
<td>2017</td>
<td>HLT_IsoMu24 OR HLT_IsoMu27 OR HLT_Mu50</td>
</tr>
<tr>
<td>2018</td>
<td>HLT_IsoMu24 OR HLT_IsoMu27 OR HLT_Mu50</td>
</tr>
</tbody>
</table>

Table 3.1: Muon triggers used by year

<table>
<thead>
<tr>
<th>Year</th>
<th>Trigger</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>HLT_Ele27_WPTight_Gsf OR HLT_Ele115_CaloIdVT_GsfTrkIdT OR HLT_Photon175</td>
</tr>
<tr>
<td>2017</td>
<td>HLT_Ele35_WPTight OR HLT_Ele115_CaloIdVT_GsfTrkIdT OR HLT_Photon200</td>
</tr>
<tr>
<td>2018</td>
<td>HLT_Ele35_WPTight OR HLT_Ele115_CaloIdVT_GsfTrkIdT OR HLT_Photon200</td>
</tr>
</tbody>
</table>

Table 3.2: Electron triggers used by year

The trigger paths chosen for each year were the ones with the lowest $p_T$ threshold that
were not pre-scaled. Pre-scaled triggers are triggers in which the rate at which events that
pass the trigger requirements are produced at the LHC is too high for the trigger system
to process, and so for each event that passes the trigger requirements, a random number
is thrown between 0 and 1. If the random number is lower than some set threshold, then
the event is not saved. This effectively lowers the rate of the process being considered by
whatever factor the threshold is set to; however, it also means that an analysis using this
trigger is using a total data set size that is the threshold times the full dataset, which may
have a significant impact on the statistical uncertainties. It is generally recommended to
use only non pre-scaled triggers if possible.

The trigger paths differ slightly between years, as the $p_T$ threshold for the triggers for
electrons increased in 2017 and 2018. To recover efficiency at higher $p_T$, where isolation is
less of an issue, the non isolated not pre-scaled triggers were also included. For electrons specifically, it is recommended at very high $p_T$ to also include the photon trigger, as electrons with large momenta can be mistaken for photons.

### 3.4 Event Simulation

After the data has been collected by the CMS detector, it is important to be able to compare what is observed in data with what is expected from theoretical predictions. To do the latter, the physics community has built a set of tools, including code that generates events based on the hard physics process, such as MadGraph5_aMC@NLO \cite{72,74} and POWHEG \cite{75,79}, and code that takes the physics process and simulates how the output particles would evolve after the interaction, such as PYTHIA \cite{80,81}. To describe all of these packages would take a dissertation within itself, but the general process itself will be elucidated here.

First, in any given bunch crossing that results in a collision at the LHC, often times there are hundreds of particles produced, as shown in a data event in 2016 in Fig. 3.15. Not all of these particles may arise from the same collision; in fact, it is likely that there are multiple collisions happening at the energies at the LHC. However, normally only one of these interactions is likely to be interesting, which also tends to be the interaction that produces the highest $p_T$ particles. This collision is known as the hard interaction, and other collisions that do not produce high $p_T$ particles are called soft interactions.

In simulation, the first step then is to define the hard process or the collision of interest. When a typical hard process is used as an input to either MadGraph5_aMC@NLO or POWHEG, the program checks if there are any Standard Model vertices that start with all the input particles and end with all the output particles. If not, it adds intermediate particles and interactions, until it can determine at least one Feynman diagram. The program then cycles through each pairwise combination of particles in the diagram, and determines whether a similar vertex can be substituted with a particle that generates the pairwise combination. For example, if the input has an intermediate vertex with an $e^+$ and $e^-$ combination, the generator checks if it can be replaced with a photon. Depending
on the generator, it then takes all the Feynman diagram combinations up to the order the
generator has been designed to calculate up to, and calculates the matrix element associated
with the hard process using helicity wavefunctions and amplitudes. The calculated matrix
elements then determine the probability in which the hard process will undergo each specific
decay, and events are generated by sampling from the phase space and dictated by the
probabilities in these matrix elements, with extra input from parton distribution functions
(PDFs), which describe the initial momentum of the partons, and other parameters, such
as the renormalization and regularization scale. The output of this step is then a series
of events with output particles that have a variety of momenta corresponding to the hard
process, taking into account all of the coherence and interference in Feynman diagrams that
may arise.

After this hard interaction is defined, how the output particles evolve over time in the
detector is important to characterize, and this is often done using PYTHIA. First, the
underlying event, which is a characterization of the other constituents in the protons that
may be in the vicinity of the particles in the hard process, is determined based on the energy
scale used to model the event. Since the underlying event is really just the modeling of the
other quarks and gluons that may be near the interaction, it must be factored into the event in order for the evolution of the event to be realistic. Following this, radiation from high $p_T$ particles both before and after the interaction is modeled as initial state and final state radiation. Special care is taken to make sure that showering and radiation diagrams do not overlap. Next, all the quarks and gluons in the interaction undergo a simulation of hadronization, which is the modeling of how colored particles interact with the vacuum and the material around it.

Finally, having the particles more or less fully simulated, the next step is simulate how these particles would interact with the CMS detector. This is done using Geant4 [82], which has a catalogue of many of the typical electromagnetic and hadronic physics interactions particles have with material, such as bremsstrahlung, photon conversion, and multiple scattering. This step is where the detector effects, such as geometry, number of layers, and magnetic field can all be factored in to the time evolution of the particles as they propagate through the detector. The detector simulation is calibrated using both test beam and cosmic muon data. After the detector simulation is calibrated, the pileup profile is added to the event, which is a characterization of the number of extra interactions from particles not in the hard process determined from minimum bias data collected beforehand. The culmination of the simulation of these processes is then the best estimate of the number and types of particles that will be produced in the detector, and the propagation of all of these particles to the hits in the tracker, the energy deposits in the calorimeters, and the hits in the muon chambers is done in the digitization step.

### 3.5 Simulation Samples

For this analysis, simulated event samples produced by the CMS collaboration are used to study the contributions of the Standard Model background and signals to the data samples. The MadGraph5_aMC@NLO v2.2.2 generator is used in leading-order (LO) mode to simulate events originating from various processes including production of a W boson plus jets, Z boson plus jets, and QCD multijet production. Top quark pair production and
single top quark events produced in the \( t \) channel are generated with the next-to-LO (NLO) POWHEG v2.0 generator, whereas single top quark events in the \( tW \) channel are generated with POWHEG v1.0 \cite{78}. Single top quark production in the \( s \) channel, as well as rare SM processes such as \( t\bar{t}Z, t\bar{t}W, \) and triboson production, are generated at NLO accuracy with the MadGraph5_aMC@NLO v2.2.2 program. The generation of these processes is based on either LO or NLO parton distribution functions (PDFs) using NNPDF3.0 \cite{83} for the simulated samples corresponding to the detector conditions in 2016, and using the next-to-NLO (NNLO) PDF sets from NNPDF3.1 \cite{84} for the 2017 and 2018 simulated samples. Parton showering and hadronization is simulated with PYTHIA v8.212 using underlying-event tune CUETP8M1 \cite{85} for 2016 samples, except for \( t\bar{t} + \text{jets} \) production which used tune CUETP8M2T4 \cite{86}, or PYTHIA v8.226 with tune CP5 \cite{87} for all 2017 and 2018 samples. The CMS detector response is simulated using a Geant4-based model, and event reconstruction is performed in the same manner as for collision data, which will be delineated in the next section. The most precise cross section calculations available are used to normalize the Standard Model simulated samples, corresponding to NLO or NNLO accuracy in most cases \cite{88-95}.

For the signal, top squark pair production events are generated using MadGraph5_aMC@NLO in LO mode, including up to two additional partons in the matrix element calculation. The
top squarks are decayed using \textsc{pythia} according to the $R$-parity violating models and the stealth SUSY models mentioned in Chapter 2, with the Feynman diagrams shown in Fig. 3.16. For the benchmark for the simulated $R$-parity violating models, the neutralino mass is assumed to be at 100 GeV, and the considered mass of the scalar top quark is from 300-1200 GeV in steps of 50 GeV. For the stealth SUSY models, the $\tilde{S} - S$ splitting is held constant at 10 GeV, the $\tilde{S}$ is kept constant at 100 GeV, and the LSP is the gravitino with a mass of 1 GeV. The considered scalar top quark mass for the stealth SUSY models is also from the range of 300-1200 GeV in steps of 50 GeV. The signal production cross sections are calculated using approximate NNLO plus next-to-next-to-leading-logarithm (NNLL) calculations \cite{96,97}.
Chapter 4

Event Selection

Having described how the collisions are generated and read out as hits and energy deposits in the detector, as well as how theoretical predictions for physics events are realistically simulated as hits and energy deposits in the detector, the next step is to reconstruct the particles. To do this, CMS uses the Particle Flow (PF) algorithm [98], which combines the information from the tracker, the ECAL, the HCAL, and the muon systems to create an event wide picture with muons, electrons, photons, charged hadrons, and neutral hadrons. This chapter starts with a brief description of the PF algorithm, followed by the definitions for electrons, muons, and jets, which have specific offline selection criteria that may differ from other analyses. Next, the search strategy is described, followed by the motivation behind the baseline selection used to define the signal region. In this section, the implementation and fundamental assumptions for this search strategy are described, followed by how the search was conducted. Afterwards, studies related to the neural network, an integral part of this analysis, and how this variable is used are explained.

4.1 Object Definition and Reconstruction

This PF algorithm’s main advantage is that it utilizes all the information simultaneously from all of the subdetectors. This approach makes it easier to not have overlapping objects and lower the chance of reconstructing fake objects. One example of this is that charged
hadrons carry about two-thirds of the energy of the jet and also leave tracks in the tracker; therefore, combining the momentum information from the tracker improves energy resolution of jets. In decays of b quarks into a muon, information from the muon systems can actually help in identifying a secondary vertex of the jet, as well as improve on the overall energy resolution of the jet.

PF is best described as a sequence of well-defined operations on all the tracks and energy deposit information, as follows:

- Track reconstruction: all the tracks in the tracker are reconstructed using an interactive, Kalman filter based algorithm with the goal of optimizing track efficiency and minimizing fake tracks [99]. The tracker provides the highest quality information for PF with good momentum resolution and a measurement for the direction of all charged particles.

- Calorimeter clustering: the energy deposits in the preshower, ECAL barrel, ECAL endcap, HCAL barrel, and HCAL endcap are then individually clustered, using energy deposits above a certain threshold as seeds. Neighboring energy deposits are lumped into the seeds if they pass a minimum energy threshold. If an energy deposit is close to multiple seeds, then it is divided up based on energy and distance.

- Link algorithm: the tracks in the tracker are then extrapolated to the calorimeters, and tracks that pass near or directly through energy deposits are associated with them with a link parameter value, which describes how well the energy clusters and the tracks match. Brehmsstrahlung from electrons has its separate link algorithm using tangents from the electron’s path. Tracks are also extrapolated to the muon system and tracks from the muon system and the tracker are matched. If there are multiple matches, the tracks with the lowest $\chi^2$ value are kept.

- Muons: all muons that have a track in the inner tracker linked to a track in the muon system are classified as PF muon objects. All tracks associated with muons are then removed from the next steps.
Electrons: PF Electrons are reconstructed using the Gaussian Sum Filter algorithm, with the additional requirement of a good track linked to an ECAL cluster with extra tracking and calorimeter quality requirements. After classification, all tracks and bremsstrahlung tracks related to the new PF electron objects are removed from the next steps.

Track quality check: all tracks now have extra quality criteria applied to them, with the main requirement that the uncertainty on the transverse momentum measurement of the track must be smaller than the ECAL or HCAL calorimeter energy resolution. These tracks are paired with ECAL and HCAL energy deposits based on track momenta, location, and calorimeter energies and become tentative photons and neutral/charged hadrons, respectively.

Charged hadrons: To differentiate the three, tracks that traverse through both ECAL clusters and HCAL clusters are grouped together as charged hadron candidates. If after all the energy deposits are associated with a charged track:

- the energy in the calorimeters is smaller than the energy measured by the track, a looser muon selection is applied to check if this track can be associated with a muon object, even with little muon system information. If it passes the looser muon selection, it is considered a muon; if not, the whole track is removed based on poor track quality.

- the energy in the calorimeters is equal to or only slightly larger than the energy measured by the track, then the track and energy deposits are grouped together as a charged hadron object. If the two are very similar, the charged hadron energy and momentum are refit to get the best resolution possible for a high $p_T$ hadron and treated as a charged hadron object.

- the energy in the calorimeters is much larger than the energy measured by the track, then ECAL energy clusters associated with the track and closest to the tracker are separated from the charged hadron object, and first reconstructed as
photons. This is done iteratively until the energy measurements are about the same. If after all ECAL clusters are classified as photons, the energy is still much much larger than the energy measured by the track, the HCAL clusters associated with the track closest to the tracker are reconstructed as neutral hadrons.

- Photons and neutral hadrons: all remaining clusters in the ECAL and HCAL with no track association are classified as photons and neutral hadrons, respectively.

For all of the objects in the analysis, a more detailed description of the reconstruction and selection criteria is provided in the following subsections.

### 4.1.1 Electrons

Electrons are reconstructed from hits in the tracker and energy deposits in the ECAL, utilizing information from both sub detectors in tandem before determining whether a set of hits is an electron [100]. The process begins in the ECAL, where hits and energy deposits in the ECAL are grouped together into “clusters,” which are of different size depending on location of the cluster. The cluster size is dependent on the $|\eta|$ of the cluster, and are used preferentially over raw hits since electrons moving through the detector radiate energy in the form of bremsstrahlung and photon conversions. Thus, the energy deposited by an electron in the ECAL is generally more diffuse and dispersed than, for example, the energy deposited by a muon, which tends to ionize very little as it traverses the ECAL. These individual clusters are then further combined into “superclusters,” and these objects, along with the collection of primary vertices, are used to start fitting and matching with a track.

From here, there are two complementary algorithms that determine the track that is associated with this supercluster. The first utilizes the location of the supercluster, calculated as the weighted average of the location of the clusters by the energy deposited, and propagates the location back to the tracker in hopes of intersecting a track segment, and this is known as ECAL-based seeding. The second algorithm starts in the tracker, creating tracks based on hits in consecutive tracker layers and propagated out to the ECAL. This is known as tracker-based seeding. In both cases, the propagation of the track utilizes a
Gaussian Sum Filter algorithm, which specifically addresses how electron loses energy in material via the Bethe-Heitler model \cite{101}. Eventually, with the application of these algorithms, the tracks with the best fit with the superclusters are then paired together and considered a reconstructed electron. This pairing of track and energy cluster is done by the Particle Flow algorithm, which is also to reconstruct and associate bremsstrahlung photons with the electron object \cite{98}.

Figure 4.1: The reconstruction of an electron in CMS, including clusters in the ECAL and the hits in the tracker \cite{102}.

Finally, the charge of the electron is determined by taking the majority of three different estimates. The first estimate is to look at the Gaussian Sum Filter track and its radius of curvature, which given the direction of the magnetic field, will determine the charge of the electron. The second estimate is to compare a Kalman Filter based track with the Gaussian Sum Filter track, and then determining the charge based on the direction of the deviation. The third estimate is to determine the sign of the difference between the vector joining the beam spot to the supercluster position and the vector joining the beam spot to the first hit associated with the track. This method of determining charge has an overall 1.5% misidentification rate in Run 1 for $Z \rightarrow e^+ e^-$ events, and has improved slightly by improved
tracking during Run 2 [103].

From the collection of PF electrons, the collaboration wide “tight” identification was also applied in order to increase the purity of the sample of electrons. This results in a collection of electrons that is over 99% pure in the barrel region and slight under 99% pure in the endcap region when looking at $Z \rightarrow e^+ e^-$ events with an overall efficiency of around 70% [104]. Additional selection requirements for electrons in the tight ID that are geometrical in nature are listed in Tab. 4.1.

<table>
<thead>
<tr>
<th>Electron Tight Identification Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td>$\sigma_{\eta\eta}$</td>
</tr>
<tr>
<td>$</td>
</tr>
<tr>
<td>$</td>
</tr>
<tr>
<td>$H/E$</td>
</tr>
<tr>
<td>$I_{iso,corr}$</td>
</tr>
<tr>
<td>$</td>
</tr>
<tr>
<td>$n_{miss}$</td>
</tr>
<tr>
<td>Conversion Veto</td>
</tr>
</tbody>
</table>

Table 4.1: Electron tight identification selection criteria

The first variable cut on is $\sigma_{\eta\eta}$, which is defined as $\sigma_{\eta\eta}^2 = \frac{\Sigma(\eta_i \eta_j)^2 w_i}{\Sigma w_i}$. The sum is over all the ECAL crystals in a 5 by 5 matrix weighted by the logarithm of the deposited energy. This variable measures the shower shape and the lateral extension of the shower in the $\eta$ direction, which is useful in determining the likelihood that this shower originated from an electron.

The next two variables are purely geometrical in nature, and are a measure of how well a track and the supercluster align. $|\Delta\eta|$ measures the distance in $\eta$ of the super cluster position from the innermost track position, and $|\Delta\phi|$ is the corresponding variable for $\phi$.

The next few variables are used to distinguish electrons produced from interactions near the collision point with electrons that may be produced in the ECAL as other particles traverse that detector, like charged hadrons. One useful variable is $H/E$, which can be
considered a variable that defines a hadronic veto. $H$ is the sum of the HCAL tower energies in a cone of radius $\Delta R < 0.15$ and $E$ is the energy of the super cluster in the ECAL. If the supercluster in the ECAL really is a result of a charged hadron, then, this ratio will be high, as the charged hadron will deposit more energy in the HCAL than the ECAL.

The variable $Iso_{corr}$ is the relative isolation of the cone, which is a measure of how close the reconstructed electron is from other reconstructed particles. For Run 2, this relative isolation is calculated using the formula $Iso_{corr} = \Sigma p_T^{charged} + \max(0.0, Iso_{PF} - \rho \times A_{eff})$, where $Iso_{PF}$ is the sum of the transverse momentum of the neutral hadrons and photons originating from the primary vertex of interest, $\rho$ is a variable that scales with the number of pileup vertices [105], and $A_{eff}$ is the effective area, which is the area corrected by $\eta$ with values shown in Tab. 4.2. The effective area values are obtained by conducting linear fits to the 90% isolation efficiency point in bins of $\rho$.

<table>
<thead>
<tr>
<th>Eta Range</th>
<th>$A_{eff}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 &lt; $</td>
<td>\eta</td>
</tr>
<tr>
<td>1 &lt; $</td>
<td>\eta</td>
</tr>
<tr>
<td>1.479 &lt; $</td>
<td>\eta</td>
</tr>
<tr>
<td>2 &lt; $</td>
<td>\eta</td>
</tr>
<tr>
<td>2.2 &lt; $</td>
<td>\eta</td>
</tr>
<tr>
<td>2.3 &lt; $</td>
<td>\eta</td>
</tr>
</tbody>
</table>

Table 4.2: Effective area values used for isolation calculations [104]

The next variable $|1/E - 1/p|$ is the difference between the inverse of the energy of the supercluster versus the inverse of the track momentum at the point of closest approach to the vertex, and has been shown to be good at picking out misidentified leptons in $Z$ enriched data samples. $n_{miss}$ is the number of missing hits in the track in the inner layers, and helps differentiate between electrons originating from the interaction point and electrons arising from converted photons that often do not have hits in the innermost layer. Finally, a conversion veto is applied, which looks at electron impact parameters in order to further
separate out electrons from photon conversions.

In addition to the “tight” identification requirements, electrons in this analysis are also required to pass a more stringent transverse momentum cut. For 2016, this cut is set to be $p_T \geq 30$ GeV since the trigger threshold was at 27 GeV for the lowest un-prescaled isolated electron trigger. For 2017 and 2018, this transverse momentum cut is set to be at $p_T \geq 37$ GeV since the trigger threshold was raised to 35 GeV for the single isolated electron triggers. Studies have been conducted that show that this has no impact on the shapes of the major variables that are utilized in this analysis, and the fact that 2016 has a lower $p_T$ cut is purely a relic of the evolution of the analysis.

Further requirements are that all electrons are required to have $|\eta| < 2.4$, as this region has the highest efficiency and resolution for electrons. Additional impact parameter cuts are imposed to increase confidence that the electron arises from the associated vertex, with cuts of $|d_0| < 0.05(0.10)$ cm and $|d_z| < 0.10(0.20)$ cm for the barrel (endcap) region electrons, where $d_0$ is the transverse distance and $d_z$ is the longitudinal distance. Finally, a more stringent isolation requirement is applied to the electron, calculating isolation using an algorithm known as mini-isolation [106]. This way of calculating isolation increases the chances that the lepton originated from a heavy quark decay, by narrowing the size of the isolation cone with increasing lepton $p_T$. The size of the isolation cone is given in Eq. (4.1) and electrons are required to have a mini-isolation value less than 0.1 [107].

$$R_{\text{mini-iso}} = \begin{cases} 
0.2 & p_T \leq 50 \text{ GeV} \\
10 \text{ GeV}/p_T & 50 \text{ GeV} < p_T < 200 \text{ GeV} \\
0.05 & p_T \geq 200 \text{ GeV}
\end{cases} \quad (4.1)$$

Electrons that pass all these requirements are considered “good” electrons in the analysis and are candidates to be the lepton in the signal region. A summary of the requirements of the electrons are shown in Tab. [4.3].
Electron Selection Criteria

<table>
<thead>
<tr>
<th>Signal Region (Good)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tight ID</td>
</tr>
<tr>
<td>$p_T &gt; 30(37)\text{ GeV}$ for 2016 (2017/2018)</td>
</tr>
<tr>
<td>$</td>
</tr>
<tr>
<td>MiniIso &lt; 0.1</td>
</tr>
<tr>
<td>$</td>
</tr>
<tr>
<td>$</td>
</tr>
</tbody>
</table>

Table 4.3: Electron offline selection for the signal region

4.1.2 Muons

Muons are reconstructed through a process of matching tracks formed from the hits from the three components of the muon systems (RPCs, DTs, and CSCs) \[108\]. There are generally three types of muon built from three types of tracks: (1) standalone-muons, (2) tracker muons, and (3) global muons \[109\].

Standalone-muon and their tracks, as its name suggests, are built only using hit information from the muon chambers. The muon trajectory uses seeds arising from hits in the DT and CSC segments and propagated radially inwards using a Kalman filter algorithm \[110\]. Since standalone muons created from these tracks do not require matching with tracks in the tracker, they have a higher percentage of cosmic muons and lower momentum resolution.

The second category are tracker muons, which are sometimes described as having “inside-out” tracks. This means that the tracks originate from hits from the tracker system, and then propagated radially outward towards the muon systems through the Kalman filter algorithm. Here, to be considered muons, these tracks have to be matched to either a DT or a CSC segment, though not necessarily a full track in the muons systems. Tracker muons, therefore, have higher efficiencies for correctly identifying muons with low $p_T$ that do not produce many hits in the muon systems.

Finally, global muons are the opposite of tracker muons, being characterized with “outside-in” tracks. This means that global muons take standalone muon tracks and match
them with tracker tracks by utilizing a combined fit and a Kalman filter algorithm to propagate the standalone muon tracks. Global muons tend to have the highest transverse momentum resolution and the lowest mistag rate for muons with high $p_T$ out of the three types of muons.

Figure 4.2: Definition of a tracker and global muon based on how their tracks are reconstructed, shown on a cross section slide of the CMS detector [111].

Overall, the propagation of the tracks and final determination of whether a track constitutes a muon utilize a bunch of variables, such as track fit $\chi^2$, the number of hits per track, matching between tracks, as well as a kink-finding algorithm and compatibility with the primary vertex [109]. While this can be done independently, in CMS, this pairing of track to muon chamber hits is done by the PF algorithm [98]. For tracks which pass the selection, they are put into collections named standalone muons, tracker muons, and global muons. These muon collections are not mutually exclusive; in fact, all global muons are by default also standalone muons, and many tracker muons are also global muons. The combination of these three collections contain about 99% of all muons that pass through the muon systems.

In this analysis, the collaboration wide “medium” muon ID is further implemented [109]. This identification uses only either tracker and global muons and have cuts to optimize for a low misidentification rate of charged hadrons and high likelihood of the muon originating
from heavy flavor decays. Since the lepton from this analysis originates from a top quark decay, this selection optimization works well with the analysis. This selection further requires that the track from the tracker have hits from 80% of the layers. If the muon is a tracker muon, the segment compatibility must be greater than a value of 0.451. If the muon is a global muon, the segment compatibility requirement is lower, at 0.303, but must also have a goodness-of-fit for the track less than 3, a goodness-of-fit in the kink finder algorithm less than 20, and the position match between the tracker and standalone muon track to be less than 12. With this additional selection and requirements, over 99.5% of the muons from simulated W and Z events are captured in the collection.

Finally, additional offline selection requirements are imposed on the muons for the signal selection. Since the primary muon triggers in this analysis are the isolated muon triggers that have a $p_T$ threshold of 24 GeV and 27 GeV, a muon $p_T$ selection cut of greater than 30 GeV was imposed to avoid turn-on trigger effects. The muon is also required to have an $|\eta| < 2.4$, as this region has the highest reconstruction efficiency and resolution. Impact parameter cuts are also applied such that the muon’s transverse distance from the primary vertex $d_0$ has to be less than 0.2 cm and the longitudinal distance $d_z$ has to be less than 0.5 cm. Finally, the muon is also required to pass an isolation cut like the electron utilizing a mini-isolation cut value less than 0.20.

To create a control region, but still with high jet multiplicity, another type of muon, deemed the anti-isolated or non-isolated muon, was defined. To ensure complete orthogonality, the mini-isolation requirement is opposite of that in the signal region, a mini-isolation cut value greater than or equal to 0.20. By inverting the isolation, a different set of triggers were necessary, as the triggers for the signal region have an online isolation cut. Thus, the lowest non pre-scaled non-isolated muon triggers were used, which happens to have a trigger selection cut of $p_T > 50$ GeV. To not be impacted by turn-on effects of the trigger, a $p_T$ cut of 55 GeV is imposed. The $\eta$ and “medium” identification points were unchanged.

Table 4.4 give a summary of the selection requirements of signal region muons, called good muons, and control region muons, called non-isolated muons.
Muon Selection Criteria

<table>
<thead>
<tr>
<th>Signal Region (Good)</th>
<th>Control Region (Non-isolated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium ID</td>
<td>Medium ID</td>
</tr>
<tr>
<td>$p_T &gt; 30 \text{ GeV}$</td>
<td>$p_T &gt; 55 \text{ GeV}$</td>
</tr>
<tr>
<td>$</td>
<td>\eta</td>
</tr>
<tr>
<td>MiniIso &lt; 2.0</td>
<td>MiniIso $\geq 2.0$</td>
</tr>
<tr>
<td>$</td>
<td>d_0</td>
</tr>
<tr>
<td>$</td>
<td>d_z</td>
</tr>
</tbody>
</table>

Table 4.4: Muon offline selection for the signal and control region

4.1.3 Jets

Since quarks and gluons are colored particles, immediately after they are produced, they begin to interact with the vacuum and nuclei nearby in processes collectively termed parton showering and hadronization. This flurry of hadronic activity results in a plethora of hadrons, predominantly charged and neutral pions, which then proceed to interact again with different parts of the detector, resulting in a shower of particles that grow in size as it exits the detector. This often collimated collection of particles are known collectively as a jet, and jets are the physical manifestations of quarks and gluons produced in collisions at the interaction point or from the decays of heavy particles, such as the top quark.

The exact definition of a jet in any specific context depends very heavily on the types of algorithms used to group tracks and energy deposits together to reconstruct them, also known as clustering [112]. This variety of algorithms are a result of the difficulty of distinguishing between energy deposits and hits arising from quarks and gluons produced close to each other, and also from partons coming from other interactions, known as pile-up [113]. Each algorithm has its own strengths and weaknesses, where some are better for separating higher energy jets in proximity, whereas other are better at distinguishing jet substructure and produce variables useful for flavor tagging.

The jets used in this analysis are the AK4 CHS jets, which stands for jets where clustering is done with the anti-$k_T$ algorithm [114], with an additional step of removing charged
hadrons as inputs to the algorithm.

The anti-$k_T$ algorithm can be described as an iterative algorithm that combines energy clusters based on distance parameters. More explicitly, the anti-$k_T$ begins by looking at the distances between all particles in an event, with the distance between particle $i$ and $j$ defined as:

$$d_{ij} = \min \left( \frac{1}{p_{T,i}}, \frac{1}{p_{T,j}} \right) * \frac{\Delta R_{ij}^2}{R^2}$$

(4.2)

where $p_T$ is the transverse momentum and $\Delta R_{ij}$ is a geometric invariant with respect to boosts along the beam axis, defined as:

$$\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$$

(4.3)

with $y$ being the rapidity and $\phi$ being the angle in the transverse plane. $R$ is a parameter that can be tuned based on the average size of the jet, which in this case is set to 0.4, thus making the collection of jets known as AK4 jets. These distances are then compared to the distances of the particles from the beam line, defined as:

$$d_{i,B} = \frac{1}{p_{T,i}}$$

(4.4)

and the algorithm is set to successively find the smallest distance in the collection of $d_{ij}$ and $d_{i,B}$. If the smallest distance is $d_{ij}$, then the two clusters are recombined and clustered together, and if the smallest distance is $d_{i,B}$, then particle $i$ is removed from the collection and added to the collection of jets. After this step, all distance are recalculated, and this iterative process continues until all input particles are either clustered together to form a jet or is a jet by itself.

Compared to other jet clustering algorithms, the anti-$k_T$ algorithm tends to have jets that are more conical in nature. This can be seen in the algorithm, where soft particles are recombined under one highly energetic parton in a cone of ever growing radius until the cone either reaches near size $R$ or approaches the conical area arising from a different highly energetic parton. When the jet boundaries collide, the more highly energetic jet maintains its conical shape, and the softer jet gets clipped. In this way, the resulting jets are very
similar to the shapes of early jets resulting from different “cone” algorithms \cite{112}. A sample of different jet clustering algorithms applied on an event generated using HERWIG and with about 1000 soft particles added is shown in Fig. 4.3.

Figure 4.3: Different jet algorithms applied to a generated HERWIG event with additional soft particles added in. These algorithms are the $k_T$ algorithm (top left), the Cambridge-Aachen algorithm (top right), the SISCone algorithm (bottom left), and the anti-$k_T$ algorithm (bottom right) \cite{114}.

The inputs to the anti-$k_T$ algorithm is the collection of PF candidates, with the removal of a set of charged hadrons that are associated with a pileup vertex. This is done by pairing each charged hadron with its corresponding track in the tracker. This track is then traced backed to a reconstructed vertex, and if this vertex is not the primary vertex, or the vertex with the highest $\Sigma p_T^2$, then the charged hadron is removed from the collection of PF candidates used for clustering. This is what the CHS in the nomenclature stands for: charged hadron subtraction.

The result of this procedure is a set of jet-like objects, known as reconstructed jets, whose energy and direction is a function of a combination of effects, like the number of
layers the particle is going through, the initial energy and direction of the particle, and stochastic processes regulating the probability the particle interacts with different layers of the detector. The quantity of interest in most cases, however, is the initial energy of the quark or gluon immediately after the collision, as this is the quantity that arises from theoretical predictions. Therefore, all reconstructed jets have to undergo a series of calibrations, collectively termed the jet energy corrections, which essentially map the reconstructed energy to the initial particle energy as best as can be done based on the particle’s trajectory and energy, which affects the energy resolution of the measurement. These corrections are derived generally in simulation, but validated and tested in data samples, and there are a different series of steps that are applied just to data. The series of steps in which the jet energy is corrected for are shown in Fig. 4.4 with the top row referring to steps applied to data and the bottom row referring to steps applied to simulation.

The first sets of corrections, known as L1 Residual Offset Correction, tries to correct for remaining energy impacts from pileup interactions. Recall that there is already pileup effects subtracted from using the charged hadron subtraction, but this L1 corrections focus on removing residual impacts not covered by charged hadron subtraction, namely particles that do not have reconstructed tracks. The offset corrections are determined from a simulation of QCD dijet events with and without pileup events overlaid, and then the overall increase in jet energy is calculated and parameterized as a function of offset energy density \( \rho \), jet area \( A \), jet \( \eta \), and jet \( p_T \) via the hybrid jet area method [116]. For simulation events,
a further correction is derived for differences between the data and detector simulation using the random cone method in Zero Bias data and compared to the Single Neutrino simulation sample [117]. Since the Zero Bias data contains no energy deposits from hard interactions, the average transverse momentum of a PF object in a random cone with a particular direction is highly correlated to the average energy offset resulting from pileup, and this can be directly compared to a simulation sample also with no hard interaction deposits (single neutrino), shown in the left plot of Fig. 4.5. This ratio is then used as a scale factor that maps reconstructed jet energy with pileup effects to reconstructed jet energy with no pileup effects, and the values of the scale factors are shown in the right plot in Fig. 4.5.

Figure 4.5: The pileup offset in both data and simulation, normalized by the average number of pileup interactions for AK4 CHS jets (left) and the resulting scale factor for the jets in different periods of data taking in 2016 (right) [118].

The second set of corrections is derived to correct for the jet detector response non-uniformity in $\eta$ and $p_T$, and are known as L2 Relative and L3 Absolute Corrections, respectively. These corrections used to be derived separately, but now they are combined into one set of corrections. These corrections are also known as MC-truth because they are in charge of corrections that match reconstructed jet energy with no pileup with the initial
jet energy, which in this case is the generated jet information in simulation (MC). This set of corrections are a function of $\eta$ and $p_T$, and focuses on the variable know as jet response ($R$) \cite{117}, which is defined as:

$$R(<p_{T,\text{reco}},\eta>, p_{T,\text{gen}}, \eta) = \frac{<p_{T,\text{reco}}>>}{<p_{T,\text{gen}}>>}[p_{T,\text{gen}}, \eta]$$ \hspace{1cm} (4.5)

The angle brackets in the equation indicate that the average is taken in each bin for which the jet response is calculated, the square brackets represent the value of the binning variables, $\text{reco}$ is the value for the reconstructed jet and $\text{gen}$ is the corresponding generated value for the jet. The value of jet response is plotted in Fig. 4.6, which shows a stable jet response for all jets in the barrel region above 30 GeV, but an increased $p_T$ dependence in the endcap regions. The drop of jet response under 30 GeV is a result of HCAL acceptance for lower $p_T$ jets. A set of corrections parameterized by $\eta$ and $p_T$ is then calculated using this response function to calibrate all jets to create a flat jet response, and maps the generated jet $p_T$ and $\eta$ with that of the final reconstructed value. This is the last step for corrections to jets in simulation outside of optional additional flavor tagging.

For data, there are a set of extra corrections, known as the L2L3 Residual jet energy scale corrections, that uses data samples to correct for small differences in data and simulation that are $p_T$ and $\eta$ dependent. Since the L2L3 MC Truth corrections are derived solely in simulation, extra care has to be taken to ensure that these corrections apply in data as well. For these relatively smaller corrections, QCD dijet samples and $\gamma/Z+$ jet samples are used since the dijet samples have very large statistics and can be very cleanly tagged.

In the dijet sample, the missing transverse energy projection fraction (MPF) is used to calculate this residual correction by comparing two back to back jets and using the recoil of one jet to calculate the response \cite{119}. This method assumes that the $E_T^{\text{miss}}$ arises solely from mis-measurements of the hadronic recoil to the reference object, and so the jet response can be determined solely by the $E_T^{\text{miss}}$ projection onto the $p_T$ of the probed jet. The visualization of the setup of the tagged jet and the probe jet, along with the $E_T^{\text{miss}}$ recoil, is shown on the left in Fig. 4.7. The relative correction is calculated as:
Figure 4.6: The jet response for jets of different transverse momenta. The jet response is relatively flat in the barrel region, but has a stronger $p_T$ dependence in the endcaps and forward regions [118].

\[
\text{Residual}(|\eta|) = \frac{1+ \langle B \rangle}{1- \langle B \rangle}, \quad B = \frac{p_{T,\text{miss}}}{2p_{T,\text{ave}}} \cdot \left( \frac{\vec{p}_{T,\text{tag}}}{p_{T,\text{tag}}} \right)
\]  

(4.6)

In this equation, the tagged jet is the higher $p_T$ jet, and $\text{ave}$ refers to the average magnitude of the $p_T$ of the two jets in the event, which helps to reduce biases from the jet energy resolution [117]. The resulting corrections, shown in the right of Fig. 4.7 shows that the relative $\eta$ correction is small for the barrel, with the largest impact in the endcap and the forward region.

In the $\gamma/Z+$ jet samples, a similar procedure is used to determine the final residual corrections with respects to $p_T$, by comparing the $p_T$ of the jet with the $\gamma/Z$ as the reference. For this specific correction calculation, the Z decays either into an electron position pair or a pair of muons, and these leptonic decays are easily identifiable such that the Z can be reconstructed with great fidelity. The correction is calculated using the same MPF method.
Jet Energy Scale: $p_T$-dependent Data/MC corrections

- Residual correction derived as a function of $p_T$ for $|\eta| < 1.3$.
- $Z/\gamma + \text{jet}$ events.

Determine the average absolute scale in $\frac{1}{1}$. The correction is correlated to the value of the $p_T$.

Figure 4.7: The $\eta$ dependent L2L3 Residual correction plotted as a function of $\eta$, showing that only near the boundary between the endcap and the forward detector is the $\eta$-dependence of the jet response mismodeled [118].

as with the dijet sample, with the correction calculated as:

$$\text{Residual}(p_T) = 1 + \frac{p_T^{\text{miss}}}{p_T^{\text{ref}}} \cdot \frac{p_T^{\text{ref}}}{(p_T^{\text{ref}})^2}$$  \tag{4.7}$$

where the reference is $\gamma/Z$. The overall value of the correction is shown on the right of Fig. 4.8, where the discrepancies between Run I and Run II are a result of different reconstruction algorithms in the two runs. The correction as a function of $p_T$ clearly shows the correction is correlated to the value of the $p_T$.

Figure 4.8: The $p_T$ dependent L2L3 Residual correction plotted as a function of $p_T$ for jets with $|\eta| < 1.3$ in 2016 for the data taking periods G and H (end of the year) [118].
After applying these residual corrections, the final collection of calibrated jets are the standard jet collection used in the collaboration for a majority of analyses. Uncertainties related to the jet energy scale and the jet energy resolution coming from the corrections above are also taken into account later in the analysis - this will be addressed in Chapter 6. Particular to this analysis are a few extra selection cuts to the jet collection to lower the number of jets with misconstructed properties from entering the analysis.

One set of these requirements are the Jet ID requirements, which are a set of criteria for jets tuned to remove jets arising from residual noise from the HCAL and ECAL, while still retaining 98-99% of all genuine jets. These jet identification (Jet ID) requirements are explicitly listed in Tab. 4.5 for the data taken in 2016 and 2017/2018.

<table>
<thead>
<tr>
<th>Eta Range</th>
<th>Variable</th>
<th>Selection</th>
<th>2016</th>
<th>2017 / 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\eta</td>
<td>&lt; 2.4$</td>
<td>Charged Hadron Energy Fraction</td>
<td>$&gt; 0$</td>
</tr>
<tr>
<td></td>
<td>Charged Multiplicity</td>
<td>$&gt; 0$</td>
<td>$0$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Charged EM Energy Fraction</td>
<td>$&lt; 0.99$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$2.4 \leq</td>
<td>\eta</td>
<td>&lt; 2.7$</td>
<td>Neutral Hadron Energy Fraction</td>
<td>$&lt; 0.99$</td>
</tr>
<tr>
<td></td>
<td>Neutral EM Energy Fraction</td>
<td>$&lt; 0.99$</td>
<td>$0.9$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of Constituents</td>
<td>$&gt; 1$</td>
<td>$1$</td>
<td></td>
</tr>
<tr>
<td>$2.7 \leq</td>
<td>\eta</td>
<td>&lt; 3.0$</td>
<td>Neutral EM Energy Fraction</td>
<td>$&gt; 0.01$</td>
</tr>
<tr>
<td></td>
<td>Neutral Hadron Energy Fraction</td>
<td>$&lt; 0.98$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neutral Multiplicity</td>
<td>$&gt; 2$</td>
<td>$2$</td>
<td></td>
</tr>
<tr>
<td>$</td>
<td>\eta</td>
<td>\geq 3.0$</td>
<td>Neutral EM Energy Fraction</td>
<td>$&lt; 0.9$</td>
</tr>
<tr>
<td></td>
<td>Neutral Hadron Energy Fraction</td>
<td>$-$</td>
<td>$0.02$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neutral Multiplicity</td>
<td>$&gt; 10$</td>
<td>$10$</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.5: Jet identification criteria that need to be passed by all jets to reject fake jets arising from HCAL and ECAL noise.

All jets in the jet collection have to pass these ID requirements. If even one jet fails these requirements, then the entire event is vetoed - this is to ensure that the one bad jet does not affect event wide quantities, like $H_T$, since these quantities are used in the derivation of scale factors and systematics. Corrections shown in Fig. 4.6 and Fig. 4.7 show that jets with
Less than 30 GeV or jets residing beyond the endcap have an overall worse jet response in the detector; thus, all jets in this analysis have to have a $p_T > 30$ GeV and $|\eta| < 2.4$.

Finally, jets within a cone of radius 0.4 ($\Delta R < 0.4$) of any good lepton are also removed from the analysis in order to not accidentally double count a lepton as a jet or vice versa. This arises in cases where a lepton has particularly high hadronic activity or, more likely, a hadronic shower has a relatively high energy lepton within it.

4.1.4 b tagged jets

The majority of jets used in this analysis are used as described in the last section, but special note has to be taken for jets arising from $b$ quark decays. Since hadrons consisting of $b$ quarks tend to have a finite lifetime longer than hadrons containing lighter quarks (on the order of $1.5 \times 10^{-12}$ s), they can often be distinguished by a secondary vertex displaced a short distance away from the primary vertex. This is particularly useful in this analysis since the signal models of interest have two top quarks that each decay into a $b$ quark and a $W$ boson, and the tagging of the jet arising from the $b$ quark can be useful in reducing the number of background events, particularly those arising from the very high cross section QCD multijet processes.

The process of tagging a jet as a $b$ jet can be done using a variety of algorithms [120], but the one used in this analysis utilizes machine learning techniques with a deep neural net, known as DeepCSV. The CSV standards for “Combined Secondary Vertex,” which is the name of the most often used $b$ algorithm prior to the application of machine learning techniques, and combines the information from secondary vertices and the impact parameters of tracks to come up with a discriminator. The DeepCSV algorithm takes all of the variables previously implemented in the CSV algorithm and uses them as inputs to a deep neural net with additional hidden layers, more nodes per layer, and a simultaneous training for all jet flavors and vertex categories. These machine learning techniques have been validated in many contexts and have been shown to be able to distinguish between $b$ jets, $c$ jets, and uds jets. [121]. The comparison of the new DeepCSV algorithm as opposed to the
CSV algorithm is shown in Fig. 1.9 and it is clear that the DeepCSV has a much better efficiency and discriminating power. The “Medium” working point corresponding to a 1% mistag rate is used for this analysis.

Figure 4.9: The distribution of the discriminator used for determining b jets using the CSV algorithm (left) and using the DeepCSV algorithm (right) [120].

4.2 Search Strategy

With all the objects defined, proceeding forward with defining the baseline selection and search strategy is relatively simple. The signal models have one main distinguishing feature: high jet multiplicity. The final states of signal model processes have two top quarks, which can produce as many as six jets if they both decay hadronically, and six additional jets arising from either the neutralinos or the decays to and originating from the stealth sector. Even with both top quarks decaying leptonically, the final state would have at least eight jets from the hard process alone (two b jets from each top quark decay and the six from the signal model itself).

When comparing this jet multiplicity to different Standard Model backgrounds, only a few processes produce as many jets as the signal models. One set of processes are those that have very high cross section and rely on low initial state or final state radiation probabilities of multiple gluons or quarks to reach the high jet multiplicity, such as QCD multijet, W boson + jets, Drell-Yan + jets, and single top + jets. The other set of processes are those arising from decays with multiple vector bosons or top quark pair production that naturally
have a lot of jets, such as diboson (WW, ZZ, WZ), triboson, $t\bar{t} + X$ (where X is one or more W/Z bosons), $t\bar{t} +$ jets. Among these Standard Model processes, the two most prominent are (1) QCD multijet processes and (2) $t\bar{t} +$ jets.

For QCD multijet processes, while the probability for such high jet multiplicity events are very low since it requires a lot of extra radiation from any hard process scatter, the enormous cross section of QCD multijet events provides enough events to be comparable with signal processes. Particularly, when looking at pure hadronic decays for the two top quarks of the signal model events, the QCD multijet background dominates purely from its large cross-section. On the other hand, for the semileptonic or purely leptonic decays of the two top quarks in signal model events, $t\bar{t} +$ jets becomes the dominant background because of its similar final event topology to the signal models and because of its high cross section at the LHC. All other backgrounds tend to have small total event yields at high jet multiplicity and are generally not a major concern. Thus, the first step of the analysis strategy is to develop a baseline selection that could either minimize or accurately predict both of these Standard Model process events.

One way to minimize the QCD multijet background is to focus on the semileptonic final state, i.e. having one of the top quarks decaying hadronically and the other top quark decaying leptonically. Since QCD multijet events generally do not have prompt leptons in their final states, this requirement drastically reduces the number of QCD multijet events, making it the sub dominant background. Early studies on the viability of focusing on final states where both top quarks decay hadronically show that the QCD multijet background had too large of a negative impact on signal sensitivity. As a side note, having both top quarks decay leptonically had too large of a negative impact on lowering signal event yield caused by the combination of (1) top quarks decay leptonically less often than hadronically and (2) signal cross sections are small to begin with. Given time and resource constraints, therefore, only the semi-leptonic channel was pursued in this analysis, with the requirement of having exactly one good lepton. A good lepton is a good electron or good muon that passes the additional selection that is defined in Tab. 4.3 and 4.4 respectively. For
bookkeeping, all events with a good electron must be from the Single Electron data set and all events with a good muon must be from the Single Muon data set.

To further reduce QCD multijet background, the analysis requires at least one medium b tagged jet that is defined in Sec. 4.1.4. Each b tagged jet is then paired with the good lepton and their invariant mass \(M_{b,l}\) is calculated. The event is required to have at least one \(M_{b,l}\) value that is between 50 and 250 GeV, constituting a very loose semileptonic top tag, and its impact can be seen in Fig. 4.10, which displays the \(M_{b,l}\) distribution of different Standard Model backgrounds compared to that of signal. In this figure, events with more than one b tagged jet in the event had the \(M_{b,l}\) value closest to 105 GeV plotted. It is clear that the QCD multijet background has more events in the lower part of this distribution and so the minimum \(M_{b,l}\) of 50 GeV removes some of the remaining QCD multijet background with minimal impact on signal. The combination of the good, isolated lepton and \(M_{b,l}\) cuts then reduces the QCD multijet background to around 5% of all events.

For \(t\bar{t} + \text{jets}\), the other dominant background, simple cuts seem to have negligible impact on signal sensitivity since most cuts that eliminate \(t\bar{t} + \text{jets}\) events also tend to proportionally eliminate signal model events. Furthermore, there were concerns that \(t\bar{t} + \text{jets}\) events at such high jet multiplicities may not be modeled as well in simulation, since hadronization processes in signal generation generally are validated in regions of lower jet multiplicity. Moreover, many jet energy corrections are done in scenarios where events with lower jet multiplicities dominate, and so this was also a minor concern. Thus, the decision was made to obtain the \(t\bar{t} + \text{jets}\) background shape from a fit to data, and \(t\bar{t} + \text{jets}\) background is treated as an irreducible background. The fit shape used for \(t\bar{t} + \text{jets}\) will be described in the background section of the dissertation in Sec. 5.2, but the fit function is motivated from theoretical considerations on jet scaling. Much of the remaining parts of the search strategy procedure is ensuring that the \(t\bar{t} + \text{jets}\) background estimation is predicted accurately.

For example, the choice to focus on the semi-leptonic final state and to get the \(t\bar{t} + \text{jets}\) background shape from data motivated the decision to apply a jet multiplicity \(N_j\)
Figure 4.10: The distribution of the invariant mass of the lepton and the b tagged jet, $M_{b,l}$, used in the loose semileptonic top tag for all four analysis time periods: 2016, 2017, 2018A, and 2018B. If there is more than one b tagged jet, the value closest to 105 GeV is used. Note that the signal cross sections are increased by a factor shown in the legend so that shapes can be prepared.
cut requiring 7 or more jets. By purposely making this $N_j$ selection and noting that the signal events should have nominally around ten jets, the lower $N_j$ multiplicity bins will be dominated by $t\bar{t} + \text{jets}$ background and can act as a background dominated region used to constrain the normalization of the data-driven background fit shape. At the same time, signal events where some of the jets may fall out of acceptance (for example, be under 30 GeV) can still be incorporated in the analysis in the lower $N_j$ bins.

Ideally, the $N_j$ shape of the $t\bar{t} + \text{jets}$ background would be derived in a dedicated control region and then applied to the signal region, perhaps with some transfer factor. Unfortunately, as mentioned above, the final states of the signal models and the $t\bar{t} + \text{jets}$ background are too similar, and any attempt to create a $t\bar{t} + \text{jets}$ dominant region is also highly signal contaminated. Thus, to create an effective control region, a neural network (NN) was designed to distinguish between signal and $t\bar{t} + \text{jets}$ events. The inputs to the NN are the jet four-momenta, the lepton four-momentum and additional event wide variables, and the NN is setup to ensure that there is minimal $N_j$ dependence in the NN discriminator score ($S_{NN}$). The NN will be discussed in Sec 4.3.

After developing this NN, the signal region is divided into four regions (referred to as NN bins), where NN bin 1 is highly dominated by the $t\bar{t} + \text{jets}$ background and NN bin 4 has the highest signal sensitivity. One can think of the $S_{NN}$ score as providing a second variable independent of $N_j$ such that low $N_j$ and low $S_{NN}$ is background enriched and high $S_{NN}$ score and high $N_j$ is signal enriched, like in Fig. 4.11. In each of the NN bins, the $N_j$ shape for the $t\bar{t} + \text{jets}$ background is ensured to be the same by (1) constructing the $S_{NN}$ output to be independent of $N_j$; (2) removing residual shape differences by deriving the divisions between the regions separately for each $N_j$ bin in the simulation; and (3) having systematic uncertainties to account for any potential differences between data and simulation. The first two steps will be presented in this chapter, with the list of systematic uncertainties and their derivations in Chapter 6. Given the above procedure and accounting for all types of mis-modeling in the systematics, the $t\bar{t} + \text{jets}$ shape is then simultaneously fit in all four NN bins, with the assumption now that the $t\bar{t} + \text{jets}$ shape should be the
same in all four NN bins. Any shape differences that are a result of statistical fluctuations or systematic uncertainties will have been taken into account, and so any further deviations in the $t\bar{t} +$ jets shape in the different $S_{NN}$ bins will be potentially new physics.

Figure 4.11: The signal region shown divided up by $N_j$ and $S_{NN}$ such that high $S_{NN}$ and high $N_j$ is signal enriched and low $S_{NN}$ and low $N_j$ is background enriched.

Given the sensitivity of the entire analysis to the $t\bar{t} +$ jets shape amongst the four bins, one thing that definitely must be taken into account is that events in this analysis have to be well reconstructed. Therefore, there are some event wide variable cuts that are necessary to make sure the analysis is robust against potential reconstruction errors. The first cut, as mentioned before, is that all jets in the event, even those that do not pass the special selection for the analysis, must pass the Jet ID requirements detailed in Tab. 4.5. Since this analysis relies very heavily on the jet multiplicity and utilizes quantities that may be impacted by even one bad jet, it is more conservative and safer to remove any events with even one jet arising from potential reconstruction errors. In this similar line of logic, an $H_T$ cut requiring all events to have at least 300 GeV is applied, where $H_T$ is defined as the scalar sum of the transverse momentum of jets with $|\eta| < 2.4$. The reasoning for this is twofold: (1) over 99% of signal events have a $H_T > 300$ GeV and (2) simulation samples, particularly of the QCD multijet sample, for events with less than 300 GeV have large event weights.
because event generation is computing intensive and the cross section for low $H_T$ events is very high. These highly weighted events can contribute unnecessarily to large statistical uncertainties.

Finally, a series of cuts to eliminate anomalous events that can arise from improper object reconstruction or detector noise, known as the missing transverse energy ($E_T^{\text{miss}}$) filters, were also implemented in this analysis. The naming of these filters is a relic of the fact that these events tend to have a larger than average impact on analyses utilizing a common variable known as $E_T^{\text{miss}}$, which is not used in this analysis. However, since these filters normally point to failures in the reconstruction algorithms and has a sub .1% effect on the signal event yield, they were implemented as a safety precaution. The complete list of $E_T^{\text{miss}}$ filters applied are:

- Primary Vertex Filter - removes all events that do not have a good primary vertex
- Beam Halo Filter - removes all events contaminated by the beam halo (particles created through beam-gas and beam-pipe interactions)
- HBHE Noise Filters - removes events that have large HCAL noise, from various sources such as ion feedback and the response of the hybrid photodiodes.
- ECAL TP Filter - removes events in which certain dead cells in the ECAL produce noise and irregular trigger primitives.
- Bad PF Muon Filter - removes events with muons that have high momentum and pass Particle Flow, but have a poor quality track and large momentum measurement uncertainty

4.2.1 2018 Event Veto

For 2018, there is an additional veto that is implemented because of a hardware failure that occurred in the latter part of the year. Two neighboring readout boxes in the front end system of the $-z$ endcap of HCAL—colloquially, named HEM 15 and HEM 16—lost
communication over the GBT link. This resulted in no hadronic energy deposit information for the \( \eta \phi \) region covered by these two RBXs, corresponding to \( \eta \in (-3.00, -1.30) \) and \( \phi \in (-1.57, -0.87) \).

The impact of the lack of hadronic energy information can impact this analysis in several ways:

- the H/E variable used to determine electron quality is now essentially zero for all electrons in this geometric area, leading to an increase in fake electrons.

- the isolation calculations for muons and electrons utilize the sum of \( p_T \) over the charged hadrons and neutral hadrons, which can be mis-measured or non-existent in this region.

- the energy of jets in this region can be mis-measured given the natural width of jets, and this can in turn affect variables that utilize jet four-momenta, like \( H_T \).

To understand how the communication failure affects event reconstruction and this analysis, data-data comparisons for different analysis variables are made. Since the communication failure occurred during Run #319077, data collected after and including Run #319077 is compared to data collected in all previous data runs of 2018, datasets termed 2018B and 2018A in the analysis, respectively. All variable distributions are normalized to unit area/volume before comparison in order to hone in on potential shape differences and trends and remove effects from the difference in integrated luminosity.

First, to understand the effects on leptons, the \((\eta, \phi)\) coordinates of the good leptons were plotted. The choice of binning was intentional such that each bins roughly correspond to regions of phase space covered by each readout box in the detector. Fig. 4.12 shows the ratio of the normalized number of events for the Single Muon data set for 2018B to the normalized number of events in 2018A on the left, with the corresponding plot for the Single Electron dataset on the right. In the Single Muon dataset, there exists a smaller excess of good muons found in the region with the communication failure—the isolation calculation skews lower when there is no hadronic information. In the Single Electron dataset, the
excess is much more significant, as there is a sharp increase in fake electrons because the H/E value for all objects in this region is approximately zero.

Figure 4.12: On the left is the ratio of events after and before the HEM failure for the $\eta$-$\phi$ of good muons and on the right is the equivalent plot for good electrons. Binning is roughly chosen to match the $\eta$-$\phi$ coverage of RBXs. Looking in the region of HEM 15/16, specifically for good electrons, there is an excess when comparing data taken after the HEM failure to data taken before.

With evidence that there are fake electrons/muons in the region of HEM 15/16 and also knowing the sensitivity of the analysis to jet multiplicity and reconstruction, an event veto that rejects events when an electron, muon or jet is within the region of HEM 15/16 is studied. If an electron or a muon is within the $\eta$-$\phi$ region associated with the two readout boxes where communication failed, then the event is vetoed. For jets, a slightly larger veto region around HEM 15/16 is used to avoid jets which partially overlap in that region. The exact definition of the HEM event veto is found in Table 4.6. With this veto applied $\sim$ 30\% of data events that pass the single lepton baseline selection are vetoed.

<table>
<thead>
<tr>
<th>Jet</th>
<th>$\eta \in (-3.20, -1.10)$ AND $\phi \in (-1.77, -0.67)$ AND $p_T &gt; 20.0$ OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muon</td>
<td>$\eta \in (-3.00, -1.30)$ AND $\phi \in (-1.57, -0.87)$ AND $p_T &gt; 20.0$ OR</td>
</tr>
<tr>
<td>Electron</td>
<td>$\eta \in (-3.00, -1.30)$ AND $\phi \in (-1.57, -0.87)$ AND $p_T &gt; 20.0$</td>
</tr>
</tbody>
</table>

Table 4.6: The exact $\eta$, $\phi$ and $p_T$ requirements for jets and leptons that are used to define the HEM event veto.
4.2.2 Baseline Selection

Given all the considerations above, a summary of the final baseline selection for the analysis is provided here:

- $H_T > 300 \text{ GeV}$
- Exactly 1 Good Muon or Good Electron
- $N_j \geq 7$
- $N_b \geq 1$
- $50 \text{ GeV} < M_{b,l} < 250 \text{ GeV}$
- Pass all $E_{\text{miss}}^T$ filters
- For 2018B, pass the extra event veto

After this set of selection cuts, the overall percentage of events in the background are shown in Fig. 4.13. As expected the $t\bar{t} + \text{jets}$ background is the dominant background, being at least 85% of the total background in any given $N_j$ bins, but reaching an overall percentage of at least 87%. The discrepancies between the years is a result of different tunes used in the generation of the simulation between 2016 and 2017/2018.

![Figure 4.13: The percentage of total background after the baseline selection is applied for 2016 (left), 2017 (middle), and 2018 (right). The category other consists of diboson, triboson, single top + jets, W + jets, and Drell-Yan + jets events.](image-url)
4.3 Neural Network

With no viable control region to get an estimate for the $t\bar{t} +$ jets background, a neural network was developed to separate between signal and background in order to create a pseudo $t\bar{t} +$ jets dominated region with little signal contamination.

An initial approach of using a Boosted Decision Tree (BDT) in the TMVA package utilizing some common event shape variables derived using the jets in the event failed when the resulting discriminant proved to be highly correlated with $N_j$. In essence, the BDT algorithm was figuring out that events with lower $N_j$ were more likely to be background and events with higher $N_j$ were more likely to be signal. Of course, this type of event discrimination is not useful for this analysis since the goal was to utilize the output to create bins independent of $N_j$. Even when special care was made to not provide information about $N_j$ explicitly, variables highly correlated with $N_j$, like $H_T$, were ranked highest among the variables used to discriminate events.

The current approach is to construct a Keras based neural network with adversarial gradient reversal (GR) training techniques to remove the $N_j$ dependence from the network during training [122]. Gradient reversal has been used by other NN based taggers in high energy physics before, such as DeepAK8, DeepFlavor, and several early stage top taggers to remove data versus simulation differences [120,123].

4.3.1 Input Variables

The full list of variables used for the neural net are listed in Tab. 4.7. Some of the variables are standard in general multivariate techniques, such as the four-momenta of the jets and the lepton. The choice to only provide seven jets to the neural net was a way to minimize any $N_j$ information given to the neural net. Similarly, the choice to not pass in the $H_T$ and $N_j$ itself were both to limit information about $N_j$ being utilized by the algorithm.

For events with more than seven jets, the jets are first ranked by their momentum in the center of mass frame. This requires that all objects in the event be boosted along the beam axis ($z$-axis) to the center of mass frame. This boost was calculated by taking the inverse
NN Input Variables

Four momentum of the top 7 jets: \((p_T, \eta, \phi, \text{mass})\)
Four momentum of the lepton: \((p_T, \eta, \phi, \text{mass})\)
Fox-Wolfram moments: \((2^{\text{nd}}, 3^{\text{rd}}, 4^{\text{th}}, 5^{\text{th}})\)
Jet-energy momentum tensor eigenvalues: \((\lambda_1, \lambda_2, \lambda_3)\)

Table 4.7: Input variables to the neural net, totaling 39 variables per event

of the sum of the four vectors of all jets with a \(|\eta| < 5.0\), which is a significantly larger range than the \(\eta\) range for the jets in the analysis. The calculation of this boost variable is why all jets in the event had to pass the jet identification requirements. After all jets are boosted, they are ranked by momentum, not \(p_T\), and the top seven jets with the highest momentum have their four vectors used as inputs, such that the \(p_T, \eta, \phi, \text{and mass of the first jet}\) are the first four variables, followed by the \(p_T, \eta, \phi, \text{and mass of the second jet}\), etc... For the lepton four vector, the lepton is also boosted into the center of mass frame.

The Fox-Wolfram moments (FWM) \([124]\) are defined in Eq. 4.8, where \(P_l\) are the Legendre polynomials, the sum over \(i\) and \(j\) run over all the hadrons produced in the event, and \(\theta_{ij}\) is the angle between the particles. These moments are used to define the geometric “shape” of an event, focusing on angular correlations between individual jets and particles. There have been studies showing that Fox-Wolfram moments are useful in determining the structure of QCD dijet events \([124]\) and also in discriminating between events with different topologies, such as weak boson fusion Higgs production from \(Z + \text{jets}\) and \(t\bar{t} + \text{jets}\) \([125]\). In this particular instance, the FWM were calculated only using the seven jets used as inputs to the neural net, and only the 2nd to 5th moments were used, as moments of higher order did not improve the training.

\[
H_l = \sum_{i,j} \frac{|p_i||p_j|P_l(\cos \theta_{ij})}{4}
\] (4.8)

Whereas the FWM defines the geometric “shape” of an event, the jet momentum tensor eigenvalues (JMTE) define more or less the “flow” of energy in an event. The JMTE...
are defined as the eigenvalues of the jet momentum tensor, defined in Eq. 4.9, and these variables show where the energy of the interaction is concentrated [126]. For example, the sum of $\lambda_2$ and $\lambda_3$ is proportional to the sphericity of the event, which measures how much energy in the event can be encapsulating as back to back jets. A dijet event will have a sphericity close to 0 and an isotropic event will have a sphericity close to 1. The JMTE are calculated using the seven input jets to the neural net.

$$ S^{\alpha,\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_j |p_j|^2} $$  \hspace{1cm} (4.9)

Before proceeding, it is important to give a brief side note about including the FWM and the JMTE in the NN. In theory, the neural network would be able to calculate the FWM and the JMTE from the inputs of the four-momenta of the jets, and so it seems unnecessary to also add these variables as inputs by hand. What was noticed during the training, however, is that adding these variables improved the overall discriminating power of the output. The reason for this is most likely that there are some correlations between the jets in the signal models that are captured by these more event wide variables that are harder to tease out of the raw values in the four-momenta. For example, one such correlation could be that the four momenta of three jets have an invariant mass close to the mass of the neutralino in the RPV SUSY model, whereas this is not the case in $t\bar{t}$ + jets. Given limited statistics in the training samples and computing time, these event wide variables are a way of providing more information to the neural network during the training to help the training reach its optimal discrimination faster.

A selection of the input variables to the NN are shown in Fig. 4.14.

2018 Veto Validation - NN Input Variables

Since all objects in the event undergo a boost calculated from jets up to an $|\eta| < 5$, an additional study was done to check whether the veto for 2018B handling the RBX communication issue would impact these variables. The distributions of the boost variable ($\beta_z$), as well as the FWM and JMTE, were compared between datasets 2018A and 2018B with and
Figure 4.14: The normalized shape for $t\bar{t} + \text{jets}$ and some signal models for 2016 for input variable distributions. The top row shows leading jet $p_T$, leading jet $\eta$, leading jet $\phi$, leading jet mass. In the middle row is the FWM for moments 2, 3, 4 and 5, from left to right. The bottom row shows the shape of the three JMTE.
without the veto. These distributions were normalized to unit area to remove the impact from differing luminosities. A sample of these distributions are shown in Figs. 4.15, 4.16, and 4.17 for $\beta_z$, the 2nd Fox-Wolfram moment, and the 3rd jet-momentum tensor eigenvalue. The black points correspond to data collected in the 2018A dataset that pass the baseline selection and the red points correspond to data collected in the 2018B dataset that pass the baseline selection with the extra 2018 veto applied. On the left are the plots for events in the Single Muon dataset, and the right are the plots for events in the Single Electron dataset. In all cases, even those not shown below, there are no discrepancies between the two dataset shapes beyond those attributable to statistical fluctuation, prompting confidence that the extra 2018 event veto does not affect the applicability of the training to the 2018B dataset.

![Figure 4.15](image-url)

Figure 4.15: On the left is the event $\beta_z$ for events in the SingleMuon data set and on the right is the equivalent plot for the SingleElectron data set. The black distribution corresponds to all events recorded in 2018A and the red distribution corresponds to all events recorded in 2018B.

### 4.3.2 Neural Network Training

The NN itself was constructed using code based on Keras with a Tensorflow backend Python package. The goal of this NN is to assign a discriminant score, $S_{NN}$, to each event.
Figure 4.16: On the left is the 2nd Fox-Wolfram moment for events in the SingleMuon data set and on the right is the equivalent plot for the SingleElectron data set. The black distribution corresponds to all events recorded in 2018A and the red distribution corresponds to all events recorded in 2018B.

Figure 4.17: On the left is the 2nd eigenvalue of the jet momentum tensor for events in the SingleMuon data set and on the right is the equivalent plot for the SingleElectron data set. The black distribution corresponds to all events recorded in 2018A and the red distribution corresponds to all events recorded in 2018B.
such that more signal like events have a score closer to 1 and more background like events have a score closer to 0, also known as binary classification. An infographic of how binary classification works in a deep neural network is shown in Fig. 4.18.

![Infographic of binary classification in deep neural network](image)

Figure 4.18: Example of a fully connected neural network with arbitrary number of input variables, two hidden layers with arbitrary number of nodes, and an output layer that has two nodes corresponding to the prediction of the classifier (signal or background in this case) [129].

This $S_{NN}$ is determined through supervised learning, which means that the input events are labeled as signal or background prior to the training. For signal, the totality of all signal model events that pass the baseline selection were summed together into one sample, and then a random subset (25%) was chosen to be passed into the training, with an orthogonal 25% of the sample used for testing. For background, the $t\bar{t} + \text{jets}$ POWHEG sample and the $t\bar{t} + \text{jets}$ MadGraph sample that passed the baseline selection were summed together and underwent the same 25% random selection for training and testing. The choice to use both the POWHEG and the MadGraph samples were to allow the training to pick up on features that may be unique to each set of parameters used to generate the $t\bar{t} + \text{jets}$ samples, and since both samples are used in the collaboration as equally valid simulation samples, there was no particular reason to choose one over the other for the training. The remaining 50% of events for both signal and background were kept as an independent sample to be use later for validation of the training.

There is some critique that can be made for the signal sample choice. It can be argued,
for example, a dedicated training using only one signal model with a particular scalar top quark mass can result in a better separation for the signal and background. While this is a valid argument, the size of the training samples was limited by the computing constraints, which made it difficult to have a large enough sample size to conduct individual trainings. Many higher $m_{\tilde{t}}$ signal models have lower cross sections, and so these signal samples were generated with fewer events. To further cut these samples down by 25% for the training was too small and had issues with overtraining. Separately, a study was conducted on generating a specific training just for the RPV SUSY model with $m_{\tilde{t}} = 350$ GeV, which has one of the highest cross sections and number of generated events. In that study, the improvement in the signal efficiency was minimal, with an increase in the area of the ROC curve by less than 1% overall; therefore, this study showed no strong motivation to do an individual signal mass model based training. Finally, the overall goal was to have a broad search for new physics instead of a narrow model-independent search, so by defining the signal training sample in this way, the analysis is not being tuned to fit any particular signal model.

The initial attempt to use a NN as described above actually suffered from the same issues as the BDT did, as the output was correlated with the jet multiplicity. To remove the jet multiplicity dependence, a technique known as gradient reversal \cite{122} was applied, which is a technique that removes the dependence of the classifier on a secondary classification.

To understand how gradient reversal works, it is important to understand how a normal classification layer in a NN works. In a broad sense, a NN works by taking the vector of inputs for each event and weighting each input in a way such that it produces a value between 0 or 1 for each event, with values closer to 0 being more background like and values closer to 1 being more signal like. This is done for each event in the training sample, and then how well the set of weights does in classifying each event is characterized in a quantity known as the loss function. When the weights in the NN are such that all events are classified correctly, the loss function has a value close to zero, whereas if all events are classified in a way that is no better than random classification, the loss function has a maximal value,
which could differ depending on the algorithm. In each iteration of the training, the weights are tuned so that classification is improved by minimizing the loss function until maximal separation between background and signal events has been observed.

Figure 4.19: A diagram that shows a network with a gradient reversal layer [122].

In gradient reversal, a new classification layer is added in addition to the normal classifier. This is shown in Fig. 4.19, where the regular classification layer is shown in blue, and the gradient reversal layer is in pink. In this analysis, the gradient reversal layer is a five way $N_j$ classification, where $N_j = (7, 8, 9, 10, \text{ or } 11)$ and 11 is an inclusive classification for all events with 11 or more jets. Thus, a particular event could be a 7 jet background event, or a 10 jet signal event, or any mix of these two categories. What differs for the gradient reversal layer is that the contribution to the loss function has an opposite sign, as shown in Eq. 4.10. In Eq. 4.10, the contribution to the loss function of the normal classification layer has a subscript $y$, the contribution from the gradient reversal layer has a subscript $d$, and there is a loss function weight factor $\lambda$ such that $\lambda > 0$.

$$L = L_y - \lambda L_d$$  \hspace{1cm} (4.10)

With the minus sign in front of the contribution to the loss function of the gradient reversal layer, the training actually works actively against classifying events in the classification associated with the gradient reversal layer. In other words, in our specific example, when the training finds a set of weights that clusters 7 jet events together, 8 jet events together, etc..., but separates these clusters away from each other, the loss function gets larger instead
of smaller. The relative impact on the overall loss function is determined by the value of the penalty variable $\lambda$, which is a tunable parameter. This parameter is set to around three in this analysis, though studies have shown that the overall efficiency of the training in separating signal and background has only a very minor dependence on the value of $\lambda$ as long as $0.5 < \lambda < 5.0$. It is with this configuration then, that the training is set up to classify the events as signal or background, but not able to classify the events as having 7 jets, 8 jets, 9 jets, etc...

Figure 4.20: For each $N_j$ region: (top left) shape of $S_{NN}$ for $t\bar{t}$ + jets, (top right) shape of MVA output of all signal models together, and (bottom) ROC curve of these distributions for the no gradient reversal case.

Figure 4.21: For each $N_j$ region: (top left) shape of $S_{NN}$ for $t\bar{t}$ + jets, (top right) shape of MVA output of all signal models together, and (bottom) ROC curve of these distributions with gradient reversal applied.

The effect of gradient reversal on the training can be seen in the difference between the $S_{NN}$ shapes for the training without gradient reversal in Fig. 4.20 and the training with
gradient reversal in Fig. 4.21 broken down by $N_j$. In increasing $N_j$, the plots show the $S_{NN}$ shapes for $N_j = 7$ (blue), $N_j = 8$ (orange), $N_j = 9$ (green), $N_j = 10$ (red), and $N_j \geq 11$ (purple). From Fig. 4.20, the training was clearly able to pick up on specific $N_j$ information in the $t\bar{t} +$ jets samples before gradient reversal, as the $S_{NN}$ distribution for all $N_j \geq 11$ events peak at one and the $S_{NN}$ distribution for $N_j = 7$ peaks at zero. This has been mostly removed by gradient reversal, as shown in Fig. 4.21, where the $t\bar{t} +$ jets $N_j \geq 11$ $S_{NN}$ distribution peaks at the same value as all the other $S_{NN}$ distributions. It is clear that even with gradient reversal, there is still some residual $N_j$ dependence, most visible in the tails in the upper left plot of Fig. 4.21 but this is much smaller than before gradient reversal. The overall efficiency of signal and background classification, defined as the area of the ROC curve, is also shown in the right plot of both Fig. 4.20 and 4.21, and this metric actually improves with the implementation of gradient reversal. This is believed to be the impact of the training honing in on correlations between the jets more heavily now that it is being penalized for utilizing $N_j$ information, and the fact that signal events may have correlations between the jets arising from the neutralino or stealth sector decays.

The overall structure of the network is shown in Fig. 4.22, which shows the two dense layers used in the binary classification network and the extra dense layer for gradient reversal. There are extra lambda and batch normalization layers, which just normalize all variables so that the variable distributions have a mean of zero and a variance of one before being used as inputs to the next layer. This process is standard practice in machine learning and neural nets since this eliminates any potential bias on the training from the raw values of the inputs. The dropout layers can be understood as masks on a random selection of nodes at each iteration such that only a subset of nodes are used when evaluating the final classification. This lowers the impact any one specific node on the overall training of the system and minimizes the chance for overfitting. The drop out rate for the trainings used in this analysis is 70%. To compare with more standard multivariate techniques used in other analyses in the CMS collaboration, the neural network used in this analysis only utilizes one extra dense layer beyond the standard neural network options in the TMVA package.
in the classification process.

Figure 4.22: The architecture of the neural network. Starting from the input layer, the network moves to a lambda layer that scales all inputs to have a mean of 0 and variance of 1, followed by a hidden layer that is then batch normalized, a common technique used in neural networks. Next, there is another hidden layer that has dropout performed and is connected to an output layer that is supervised for the signal vs background classification. Finally, to remove $N_j$ dependence, this output layer is connected to a gradient reversal layer followed by a hidden layer, and finally another output layer that is supervised for classifying $N_j$.

To conclude, the final $S_{NN}$ shape for all backgrounds stacked together, a few representative signal models, and the data overlaid are shown in Fig. 4.23. On the left is the simulation and training for the dataset collected in 2016, whereas on the right is the simulation and training for the dataset collected in 2017. There were two separate trainings conducted - (1) with the simulation in 2016 and (2) with the simulation in 2017. The simulation in 2016 was generated using a different underlying Tune (i.e. different parameters
and parton distribution function set versions) than in 2017, so the signal and $t\bar{t}$ + jets had slightly different properties. The simulation generated in 2018 used the same parameters as 2017, so the training in 2017 was used for 2018 after checks were made to confirm that the distributions of all important variables in this analysis were similar between the two sets of simulations samples. The uncertainty band plotted is the proper quadrature sum of the statistical uncertainty of the simulation, the statistical uncertainty of the data, and the two largest systematic uncertainties in the fit procedure.

Figure 4.23: The overall $S_{NN}$ score distribution for all the estimated backgrounds and the data for 2016 (left) and 2017 (right). A few representative signal samples’ $S_{NN}$ distributions are also overlaid with the cross section increased so that the shapes are visible. The uncertainty band incorporates the statistical uncertainty and the two largest systematic uncertainties added in quadrature.

4.3.3 Removing Residual $N_j$ Correlation

As mentioned in Sec. 4.2 and worth emphasizing here, the most important assumption of this analysis is that the $N_j$ distribution in the final NN bins be the same, meaning that $N_j$ and $S_{NN}$ must be uncorrelated. The majority of the correlation has already been removed by gradient reversal, but some small correlation undoubtedly remains. Thus, an extra step has to be done to explicitly remove the last of the correlation between $N_j$ and $S_{NN}$.
A pictorial representation of how this procedure was done is shown in Fig. 4.24. In general, the procedure of deriving the bin boundaries between the NN bins is such that instead of having one set of bin edges that is independent of $N_j$, the bin boundaries between each NN bin is a function of $N_j$. By tuning the bin boundaries per $N_j$, the $t\bar{t} + \text{jets}$ fraction ($f_{t\bar{t}+\text{jets},D_i}$, where $i \in [1,4]$ representing the NN bin) can be kept constant in each NN bin. For example, $f_{t\bar{t}+\text{jets},D_1}$ can be 55%, $f_{t\bar{t}+\text{jets},D_2}$ can be 30%, $f_{t\bar{t}+\text{jets},D_3}$ can be 10%, and $f_{t\bar{t}+\text{jets},D_4}$ can be 5%. In this example, the $N_j = 7$ bin of NN bin D1 will have a bin boundary, let’s say at a $S_{NN}$ of .40, such that 55% of all $t\bar{t} + \text{jets}$ $N_j = 7$ events are in that bin. The $N_j = 8$ bin of the NN bin D1 will also have a cut such that 55% of all $t\bar{t} + \text{jets}$ $N_j = 8$ events are in that bin, but the $S_{NN}$ boundary value may be .41. This procedure then guarantees in simulation that the $t\bar{t} + \text{jets}$ shape is the same in all four NN bins. The total $t\bar{t} + \text{jets}$ simulation events can be divided up in many different ways between the four bins, so long as the fractions sum up to one.

To determine the actual $f_{t\bar{t}+\text{jets},D_i}$, an optimization procedure was concocted using the simulation samples for $t\bar{t} + \text{jets}$, QCD multijet, and the RPV SUSY signal model with $m_{\tilde{t}} = 550$ GeV. The signal model with $m_{\tilde{t}} = 550$ GeV was chosen because it was an intermediate mass out of all the simulation samples at the time. The significance metric per $N_j$ bin per

Figure 4.24: Bin division with respect to $N_j$ and MVA value, in the ideal scenario (left), and the actual scenario (right). The discrepancy is a result of the residual dependence on $N_j$ in the training.
NN bin is defined in Eq. 4.11 where \( N_{\text{sig}} \) is the number of events for the RPV SUSY signal model, \( N_{t\tau} \) is the number of \( t\tau + \text{jets} \) events, and \( N_{QCD} \) is the number of QCD multijet events in a given \( N_j \) bin after the NN bin edges (corresponding to a given \( t\tau + \text{jets} \) fraction) are applied. The first two terms of the denominator represent the statistical uncertainty for the \( t\tau + \text{jets} \) and QCD multijet background, respectively, whereas the latter two terms of the denominator represent a 20% systematic uncertainty for both samples. The overall optimized value, then is the sum of the significance metric over all \( N_j \) bin and all NN bins for a total of \( 6 \times 4 = 24 \) bins.

\[
\frac{S}{\sigma_B} = \frac{N_{\text{sig}}}{\sqrt{N_{t\tau} + N_{QCD} + (0.20 \times N_{t\tau})^2 + (0.20 \times N_{QCD})^2}} \quad (4.11)
\]

While this significance metric is not exactly the significance associated with the analysis, this metric is correlated to the significance metric and was the best approximate estimate for the size of the systematic uncertainties at the time. The optimization was done utilizing the 2016 simulation scaled to \( 80 \text{ fb}^{-1} \), and the final \( t\tau + \text{jets} \) background fractions were: \( f_{t\tau+\text{jets},D1} = 47.8\% \), \( f_{t\tau+\text{jets},D2} = 38.9\% \), \( f_{t\tau+\text{jets},D3} = 6.5\% \), and \( f_{t\tau+\text{jets},D4} = 2.4\% \). The actual values of the bin boundaries for all four years in terms of the \( S_{NN} \) are shown in Tab. 4.8.

With these bin edges, the number of signal events for different RPV SUSY signal models in each of the NN bins are shown on the left in Fig. 4.25 with the sum in quadrature of the significance metric overall \( N_j \) bin shown on the right. While the plot seems to indicate that the number of signal events do not necessarily increase with the NN bin, the important thing to note is that higher NN bins also have lower background event yields, and so the fraction of signal events in higher NN bins is higher. This is exemplified in the significance metric plot on the right, which show that at higher mass models, the higher NN bins have a much higher significance than the lower NN bins. For lower mass models, such as RPV SUSY with \( m_{\tilde{t}} = 350 \text{ GeV} \) shown in red, the higher NN bins are about as sensitive as the lower NN bins because the overall \( S_{NN} \) distribution is more similar to that of \( t\tau + \text{jets} \), and so more signal events lie in the D2 and D3 bins.
<table>
<thead>
<tr>
<th></th>
<th>$N_j = 7$</th>
<th>$N_j = 8$</th>
<th>$N_j = 9$</th>
<th>$N_j = 10$</th>
<th>$N_j = 11$</th>
<th>$N_j \geq 12$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1 $\rightarrow$ D2</td>
<td>.349</td>
<td>.358</td>
<td>.371</td>
<td>.386</td>
<td>.410</td>
<td>.410</td>
</tr>
<tr>
<td>D2 $\rightarrow$ D3</td>
<td>.680</td>
<td>.740</td>
<td>.787</td>
<td>.816</td>
<td>.848</td>
<td>.853</td>
</tr>
<tr>
<td>D3 $\rightarrow$ D4</td>
<td>.835</td>
<td>.878</td>
<td>.901</td>
<td>.917</td>
<td>.936</td>
<td>.936</td>
</tr>
<tr>
<td>2017</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1 $\rightarrow$ D2</td>
<td>.346</td>
<td>.363</td>
<td>.376</td>
<td>.384</td>
<td>.396</td>
<td>.396</td>
</tr>
<tr>
<td>D2 $\rightarrow$ D3</td>
<td>.714</td>
<td>.752</td>
<td>.780</td>
<td>.794</td>
<td>.801</td>
<td>.801</td>
</tr>
<tr>
<td>D3 $\rightarrow$ D4</td>
<td>.833</td>
<td>.864</td>
<td>.885</td>
<td>.895</td>
<td>.901</td>
<td>.901</td>
</tr>
<tr>
<td>2018A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1 $\rightarrow$ D2</td>
<td>.345</td>
<td>.363</td>
<td>.374</td>
<td>.386</td>
<td>.386</td>
<td>.386</td>
</tr>
<tr>
<td>D2 $\rightarrow$ D3</td>
<td>.713</td>
<td>.752</td>
<td>.779</td>
<td>.805</td>
<td>.805</td>
<td>.805</td>
</tr>
<tr>
<td>D3 $\rightarrow$ D4</td>
<td>.832</td>
<td>.863</td>
<td>.883</td>
<td>.900</td>
<td>.911</td>
<td>.911</td>
</tr>
<tr>
<td>2018B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1 $\rightarrow$ D2</td>
<td>.364</td>
<td>.386</td>
<td>.405</td>
<td>.411</td>
<td>.411</td>
<td>.411</td>
</tr>
<tr>
<td>D2 $\rightarrow$ D3</td>
<td>.729</td>
<td>.771</td>
<td>.798</td>
<td>.826</td>
<td>.826</td>
<td>.826</td>
</tr>
<tr>
<td>D3 $\rightarrow$ D4</td>
<td>.842</td>
<td>.874</td>
<td>.895</td>
<td>.914</td>
<td>.914</td>
<td>.914</td>
</tr>
</tbody>
</table>

Table 4.8: MVA bin edges per $N_j$ bin for all four data sets
Figure 4.25: (Left) Number of signal events for different RPV signal models in each NN bin. (Right) The sum in quadrature over each $N_j$ bin of the significance metric as defined in Equation 4.11 per NN bin.

Finally, the impact of the per $N_j$ bin boundaries on the $t\bar{t} + \text{jets}$ simulation shapes in the four NN bins are shown in Fig. 4.26 for 2016. On the left are the $t\bar{t} + \text{jets}$ shapes for all 4 NN bins, each normalized to unit area, using only the bin boundaries defined for $N_j = 7$, whereas on the right are the same shapes using the per $N_j$ bin boundaries. The need for this per $N_j$ bin boundary correction is clear from the improvement in the shape agreement. The residual shape disagreements are a result of finite size of the simulation samples and the use of weighted events. This shape disagreement will subsequently be taken care of in its own systematic.
Figure 4.26: Dividing up the $t\bar{t} +$ jets simulation into the four NN bins using bin boundaries defined only using the $N_j = 7$ bin (left) and defined by each individual $N_j$ bin (right).
Chapter 5

Background

After deriving the $S_{NN}$ and the bin boundaries, the next part of the search strategy is to then accurately predict the background processes in each of the NN bins. First, the scale factors used to correct the simulation for known discrepancies between data and simulation are addressed; these scale factors were all applied to the simulation in the derivation of the bin boundaries, but it was more natural to discuss the bin boundaries immediately after defining the neural network, and so the scale factors are discussed here. Then, the strategy to obtain the $t\bar{t} + \text{jets}$ estimate using a fit function focusing on jet scaling is described, followed by the QCD multijet estimate that is derived from a control region. Finally, preliminary fits to pseudodata are presented to validate the background estimation and analysis strategy.

5.1 Correction of the simulation: scale factors

For simulation, there are a set of corrections needed since there are known discrepancies arising from different aspects of the baseline selection affecting data differently than the simulation. In a very simple example, the trigger efficiency in data may be 10%, but when the trigger is simulated, since it is impossible to send the simulation events through the hardware trigger, the efficiency may be 11%. These discrepancies need to be taken into account, and the standard way to do this is by scaling the simulation by a scale factor derived
in a region where the impact of a selection cut can be isolated from all other effects. In this analysis, these scale factors are even more important since these discrepancies can result in differences in the $t\bar{t}$ + jets $N_j$ distributions amongst the four NN bins and potentially fake a signal. This section aims to address the derivation and utilization of these different scale factors for all the selections in the baseline.

### 5.1.1 Lepton Scale Factors

For leptons, the overall scale factor is calculated as a product of scale factors addressing discrepancies at each stage of the object reconstruction. Each scale factor itself is defined as the ratio of the efficiency ($\epsilon$) of a specific part of the reconstruction in data to the efficiency of the same step in the simulation. In this way, the scale factors, when applied to simulation will result in a distribution that matches the data.

More explicitly, the efficiency of constructing a lepton can be broken down as

$$\epsilon_{lep} = \epsilon_{track} \times \epsilon_{reco+1D} \times \epsilon_{isolation} \times \epsilon_{trigger}$$  \hspace{1cm} (5.1)

where the four individual components are the track reconstruction efficiency, the object reconstruction and identification efficiency, the isolation efficiency, and the trigger efficiency. Depending on whether the lepton is a muon or an electron, one or more of these efficiencies may be calculated at the same time. For electrons, the track reconstruction efficiency is taken into account in the object reconstruction and identification efficiency. The final scale factor is then:

$$w = \frac{\epsilon_{track, data}}{\epsilon_{track, sim}} \times \frac{\epsilon_{reco+1D, data}}{\epsilon_{reco+1D, sim}} \times \frac{\epsilon_{isolation, data}}{\epsilon_{isolation, sim}} \times \frac{\epsilon_{trigger, data}}{\epsilon_{trigger, sim}}$$  \hspace{1cm} (5.2)

#### Muon Scale Factors

For the tracking efficiency of muons in data, a tag-and-probe method is utilized by the tracking physics object group [108]. In this method, Drell-Yan events with $\gamma/Z$ decaying to two muons are selected for by requiring oppositely charged muons to be identified in the
muon chambers. One of the muons is considered the “tag” muon if it has a reconstructed track in the muon chambers and the tracker. The other “probe” muon is required to have a reconstructed track in the muon chambers, but not the tracker. The efficiency is then the number of “probe” muons that also have a reconstructed track in the tracker. A few extra requirements and steps have to be taken, such as ensuring the invariant mass of the di-muon pair to be between 50-130 GeV (loosely the Z boson mass) and subtracting out the non-resonant background, but this effectively gives the track reconstruction efficiency. This is then compared to applying the same procedure to a Drell-Yan simulation sample, and the ratio of the two gives the track reconstruction scale factor. The efficiencies in data and in simulation for 2016, 2017, and 2018, binned by $p_T$ are shown in Fig. 5.1 which show great efficiency in the data across all $p_T$.

![Figure 5.1: The muon track reconstruction efficiency as a function of $p_T$ for 2016 (left), 2017 (middle), and 2018 (right) derived using the tag-and-probe method in Drell-Yan simulation samples and data [131].](image)

The muon $\epsilon_{reco+ID}$ and $\epsilon_{isolation}$ are computed in tandem in two consecutive tag-and-probe methods by the muon physics object group, with the requirements on the “tag” and “probe” muon getting more stringent with each new efficiency. These efficiencies check to ensure that the “Medium” ID and mini-isolation cuts that were applied to the good muon objects act the same between data and simulation. In the calculation of the muon $\epsilon_{reco+ID}$ efficiency, the “tag” muon has to pass the “Tight” muon requirements [132], have a $p_T > 29$ GeV, and be isolated in a relative combined isolation cone of 0.2. In turn, the
“probe” muon has to pass either the global or tracker track requirements, and the efficiency is calculated as the number of “probe” muons that pass the “Medium” identification over total number of “probe” muons.

Similarly, the muon $\epsilon_{\text{isolation}}$ requires the “tag” muon to pass the mini-Isolation requirements (<0.2) in addition to the previous requirements, whereas the “probe” muon has to now pass both the “Medium” identification requirements and the track requirements. The efficiency is then the number of “probe” muons that pass the isolation cut over the total number of “probe” muons.

In this way, the “probe” muon always incorporates the selection of each previous step before calculating the next efficiency, setting up these efficiencies and scale factors in a way that the overall muon efficiency is just the product of each individual step. Fig. 5.2 shows the reconstruction and identification efficiency for data and simulation for low $|\eta|$ on the left and high $|\eta|$ on the right. Fig. 5.3 shows the isolation scale factor binned by $p_T$ and $|\eta|$ for muons, where the ratio of the efficiencies have already been calculated. In all these plots, the efficiencies and scale factors are very close to unity for the vast majority of the muons.

Figure 5.2: The muon reconstruction and “Medium” identification efficiency for muons with $|\eta| < 0.9$ on the left and for muons with $2.1 < |\eta| < 2.4$ on the right. 

\[132\]
Finally, the last muon efficiency/scale factor that needs to be calculated is the $\epsilon_{\text{trigger}}$ because the emulation of the hardware trigger was utilized for all simulation samples. Since the triggers used in this analysis uses either the isolated muon or electron as its handle, the trigger efficiency is a subset of the overall lepton efficiency. The $\epsilon_{\text{trigger}}$ was specifically for this analysis, and was derived in a region with high jet multiplicity in order to have it be as applicable as possible to the signal region. With that in mind, the baseline selection for the trigger efficiency calculation was:

- $H_T > 300$ GeV
- $N_j \geq 5$
- $N_b \geq 1$
- $50$ GeV $< M_{b,l} < 250$ GeV
- 1 Good Muon
- 1 Good Electron ($p_T > 40$ GeV)
The requirement of two good leptons ensures that this region is completely orthogonal to the signal region. In addition, the data set used to derive the muon efficiency is the Single Electron data set. Recall that in the signal region, all events with good muons must originate from the Single Muon data set. For simulation, the $t\bar{t} + $ jets simulation sample was used.

To make sure that trigger turn-on effects of the electron do not affect the calculation, the $p_T$ requirement of the electron is set past the trigger threshold at 40 GeV. The $\epsilon_{\text{trigger}}$ was then calculated by looking at the number of events that pass the above selection and the muon trigger set defined in Tab. 3.1 over the total number of events that just pass the above selection. The trigger efficiency scale factor is the ratio of $\epsilon_{\text{trigger}}$ in the Single Electron data set over that of the $t\bar{t} + $ jets simulation sample, and parameterized as a function of $p_T$ and $\eta$. The trigger efficiencies binned in $p_T$ and $\eta$, individually, are shown on the left and right in Fig. 5.4 respectively, and the overall scale factor binned in both $p_T$ and $\eta$ simultaneously is shown in Fig. 5.5.

Figure 5.4: Trigger efficiency for the isolated muon triggers binned in $p_T$ (left) and $\eta$ (right) in both data and the $t\bar{t} + $ jets simulation for 2016.
Figure 5.5: Trigger efficiency scale factor for events that pass the isolated muon triggers binned by $p_T$ and $\eta$. 
**Electron Scale Factors**

For electrons, there is no equivalent tracking efficiency calculation to the one utilizing the muon tracks in the muon systems. Therefore, a tag-and-probe method that combines the track reconstruction, the object reconstruction, and the electron identification efficiencies is used. For the efficiency calculation in data, Drell-Yan events where the $\gamma/Z$ decays into an electron positron pair are used, whereas in simulation, an analogous Drell-Yan sample is simulated.

The “tag” electron is then required to pass a minimum $p_T$ threshold of 25 GeV and the “Tight” working point, whereas the only requirement of the “probe” electron candidate is that it has an invariant mass with the “tag” electron between 60 and 120 GeV. The dielectron mass distributions for candidates that pass the “Tight” identification and for candidates that fail the “Tight” identification are then fit with a Breit-Wigner function with a falling exponential in order to correct for the background component before comparing to simulation [100]. The $\epsilon_{\text{track,reco+ID}}$ is then calculated as normal, using the number of electron candidates that pass the “Tight” identification criteria over the total number of electron candidates, and binned by $p_T$ and $\eta$. $\epsilon_{\text{isolation}}$ is then calculated in a similar way, with the “probe” electron candidate also having to pass the mini-isolation cut of $< 0.2$. The overall efficiency in data and the scale factors are shown in Fig. 5.6.

The trigger efficiency is then calculated in a way that is analogous to its derivation for the muon triggers, except this time, the good muon is required to have a $p_T > 40$ GeV and the data set used is the Single Muon data set. The corresponding efficiencies binned by $p_T$ and $|\eta|$ individually are in Fig. 5.7, and the corresponding scale factors binned by both $p_T$ and $|\eta|$ are in Fig. 5.8.

**5.1.2 b tag scale factor**

The $b$ tag algorithm can also be more or less efficient in simulation samples than in data. Therefore, a $b$ tag scale factor that utilizes the truth information in the simulation samples was derived [133].
Figure 5.6: The electron $\epsilon_{\text{track, reco+ID}}$ binned by $p_T$ for different $|\eta|$ ranges on the left and $\epsilon_{\text{isolation}}$ binned by $p_T$ and different $|\eta|$ ranges on the right \cite{132}.

Figure 5.7: Trigger efficiency for the isolated electron triggers binned in $p_T$ (left) and $\eta$ (right) in both data and the $t\bar{t}$ + jets simulation for 2016.
Figure 5.8: Trigger efficiency scale factor for events that pass the isolated electron triggers binned by $p_T$ and $\eta$. 
For each simulation sample, the b tagging efficiencies (ε) for truth b, c, and udsg jets are first calculated. Then, the probability that any event in a sample has the given b tagged jet multiplicity can be calculated as:

\[
P(\text{sim}) = \prod_{i=\text{tagged}} \epsilon_i \prod_{j=\text{not tagged}} (1 - \epsilon_j) \quad (5.3)
\]

Similarly, efficiencies for a particular working point can be calculated in data such that:

\[
P(\text{data}) = \prod_{i=\text{tagged}} SF_i \epsilon_i \prod_{j=\text{not tagged}} (1 - SF_j \epsilon_j) \quad (5.4)
\]

where SF are data-determined factors that are functions of the jet flavor, jet \( p_T \), and jet \( \eta \) and calculated by the CMS BTV (b tagging and vertex-ing) physics object group. With these, two probabilities, an event weight scale factor per sample (\( w \)) can be defined as:

\[
w = \frac{P(\text{data})}{P(\text{sim})} \quad (5.5)
\]

### 5.1.3 Prefiring scale factor

During operation between 2016 and 2017, there was a gradual shift in the timing of the L1 trigger primitives in the ECAL [134]. This affected primarily events with large electromagnetic activity at very high \( \eta \), particularly at \( 2.0 < |\eta| < 3.0 \), and the timing delay was such that this activity was mistakenly associated with the previous bunch crossing. Since the L1 forbids two consecutive bunch crossings from firing, events with abnormally high activity can cause a self veto. This effect was not fully understood until the end of data taking in 2017, so it was fixed in 2018 and primarily had an impact on data collected in 2017, where the drift was maximal.

Since this effect was not foreseen, it was not factored into any of the simulation samples. As a result, a specific scale factor had to be derived to re-scale the simulation to account
for this, such that:

\[ w = 1 - P(\text{prefiring}) = \prod_{i=\text{photons,jets}} (1 - \epsilon_i^{\text{prefiring}}(\eta, p_T^E)) \]  

(5.6)

where \( \epsilon_i^{\text{prefiring}}(\eta, p_T^E) \) is a function of the photon or jet \( \eta \) and the electromagnetic component of the \( p_T \). Upon applying the recommended recipe of reweighting the events that pass the signal selection due to the ECAL prefiring, it was discovered that the overall impact to the analysis was very small, mostly because the analysis does not use photons and requires that all jets have an \( |\eta| < 2.4 \). Furthermore, the majority of the jets in this analysis are softer jets and do not produce large electromagnetic activity. The scale factor distribution for the \( t\bar{t} + \text{jets} \) sample and a signal model sample of RPV SUSY with \( m_{\tilde{t}} \) of 550 GeV in 2017 is shown in Fig. 5.9.

Figure 5.9: ECAL pre-firing scale factor distribution for the \( t\bar{t} + \text{jets} \) sample (left) and the RPV SUSY model with \( m_{\tilde{t}} \) of 550 GeV (right). The y-axis refers to the binning by a multivariate neural net output that will be described in Sec. 4.3.

5.1.4 QCD Renormalization and Factorization Scale Factor

This scale factor comes from the determination of two parameters used in calculations for perturbative QCD. Since QCD has both infrared and ultraviolet divergences in perturbation calculations, it is unavoidable that there is the introduction of two scales in order to remove these divergences from calculations.
One example of an ultraviolet divergent calculation in perturbative QCD is in a loop integral corresponding to two virtual massive scalars with momentum $p$ (like in the gluon self energy calculations) \cite{135}.

\[ B(p^2; m, m) = \int \frac{d^4q}{16\pi^2} \frac{1}{q^2 - m^2} \frac{1}{(q + p)^2 - m^2} \]  

(5.7)

The standard procedure for dealing with these divergences is to regularize the integral (either through the introduction of a cutoff scale or through dimensional regularization), take note of the divergence in a concrete manner, and then renormalize the integral. The most consequential result from this procedure for experimental particle physics is that it introduces a renormalization scale ($\mu_R$), which manifests itself in the running strong coupling $\alpha_s$ in the renormalization group equation (RGE), such that:

\[ \frac{1}{\alpha_s(p^2)} = \frac{1}{\alpha_s(\mu_R^2)} + \frac{1}{4\pi} \beta_0 \ln \frac{p^2}{\mu_R^2} \]  

(5.8)

where $\beta$ is the standard beta function calculated first by Wilczek, Politzer, and Gross that is characteristic of QCD \cite{136}. Introducing a QCD scale parameter $\Lambda_{QCD}$ and taking the $\beta_0$ order, this simplifies to

\[ \alpha_s(p^2) = \frac{4\pi}{\beta_0 \ln(p^2/\Lambda_{QCD}^2)} \]  

(5.9)

Meanwhile, an example of infrared divergences is in calculations for the probability of soft gluon radiation from a hard quark \cite{137}. The expression for soft gluon emission is proportional to $1/(E \times \sin(\theta))$ where $\theta$ is the angle between the hard parton and the quark and $E$ is the energy of the gluon. As a result, if the gluon has very little energy or if it is emitted collinear to the quark, then the probability for a gluon emission can be infinite.

To deal with this, a factorization scale ($\mu_F$) is introduced in calculations, which separates the physics of the hard process with that of the soft process \cite{138}. This is possible since the processes occurring in the proton and in hadronization occur on time scales orders of magnitude slower than that of hard processes, like Higgs and Z production. The infinities in the infrared divergences are then absorbed into the parton densities, and the physics of
the hard process can be calculated perturbatively. The infrared divergences are then taken care of through the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi equations:

\[
\frac{df_i(x, \mu_F)}{d \log \mu_F^2} = \frac{\alpha_s}{2\pi} \sum_j \int_x^1 \frac{dx'}{x} P_{i \rightarrow j} \frac{x}{x'} f_j(x', \mu_F) \tag{5.10}
\]

where \(P_{i \rightarrow j}\) are the splitting kernels for different partons involved (quark or gluon) and \(x\) is the momentum fraction that the incoming parton transfers to the parton entering the hard process.

While the exact value of these two scales do not have an impact on the underlying physics if taken into account properly, they do have an impact on how large the theoretical uncertainties in next to leading order corrections for physics observables are. Particularly, they can factor into the calculation of cross sections; for example, in a calculation of the cross section of \(e^+e^- \rightarrow \text{hadrons}\), shown in Fig. 5.10.

Figure 5.10: The dependence on the cross section of \(e^+e^- \rightarrow \text{hadrons}\) relative to the hard process of \(e^+e^- \rightarrow qq\) on the QCD renormalization scale. Here Q is taken to be the reference scale of the mass of the Z boson \[137\].

Thus, the choice of the value of the two scales (\(\mu_R\) and \(\mu_F\)) used in sample generation must be compared with the values obtained from a fit to data. The discrepancy between the two are used to calculate a scale factor for all events in the simulation.
This scale factor is, functionally, very close to unity for most events, but variations of these scale parameters are useful in defining a systematic utilized in this analysis.

### 5.1.5 Parton distribution function scale factor

Similar to the QCD factorization and renormalization scale factors, there is an additional scale factor related to the parton distribution functions (PDFs) [139]. The PDFs describe the momentum distribution functions for all the partons within the proton, and are a function of the square of the energy scale \( (Q^2) \) and particle type, as shown in Fig. 5.11.

The PDF shapes factor heavily in the simulation of parton-parton collisions and the overall likelihood of a certain hard process occurring.

![Graph](image)

**Figure 5.11:** The PDF set (NNPDF3.1) used for the generation of the simulation for a scale of \( \mu^2 = 10 \text{ GeV}^2 \) on the left and \( \mu^2 = 10^4 \text{ GeV}^2 \) [140].

The particular PDF sets used (NNPDF 3.0 and NNPDF 3.1) were derived from fits to data obtained from deep inelastic scattering experiments. These include fixed-target lepton-nucleon scattering experiments at SLAC, FNAL, and CERN, along with electron-proton scattering experiment at the HERA collider at DESY [141]. The difference between v3.0 and v3.1 is that NNPDF 3.1 included more recent fits from 7 and 8 TeV fits from CERN and modeled the charm quark independently for higher precision and accuracy [140].
The PDF scale factor itself is calculated by deriving the PDF set using the most recent collision data and comparing it with that of the generating set. Given that NNPDF 3.1 contained many of the most up to date data fits, the scale factor approached unity for those samples, while the scale factor for samples generated with NNPDF 3.0 varied a little more, even though they were still close to unity. The choice of using NNPDF 3.0 or NNPDF 3.1 for sample generation was only dependent on which samples were available at the time of the generation of the simulation.

5.1.6 Pileup scale factor

In order to have all the simulation samples ready as soon as data has been collected, many of the simulation samples were generated with an estimated pileup profile. Most times during operation, however, the actual pileup profile will differ as a result of different running conditions or just standard operational fluctuations. Thus, the pileup profile at the end of data collection needs to be checked with the pileup profile used during generation. This pileup profile (\( \mu \)) can be characterized by the inelastic proton-proton scattering cross section (\( \sigma_{\text{inel}} \)), the instantaneous luminosity (\( L_{\text{inst}} \)), and the LHC orbit frequency \( f_{\text{rev}} \) of 11246 Hz as:

\[
\mu = \frac{L_{\text{inst}} \sigma_{\text{inel}}}{f_{\text{rev}}}
\]  \hspace{1cm} (5.11)

The LHC orbit frequency is necessary to convert the instantaneous luminosity to a value that can be interpreted per lumi section as opposed to per unit of time. This gives the average pileup for that lumi section, which can then be modeled for any given event by using a Poisson distribution around this average value. In PYTHIA, the default generation cross section is \( \sigma_{\text{inel}} = 80.0 \text{ mb} \). For Run 2, the CMS collaboration LUMI POG has recommended the use of \( \sigma_{\text{inel}} = 69.2 \text{ mb} \) with a 4.6\% uncertainty \cite{42}. The difference in the shapes of the pileup profiles between the standard simulation for PYTHIA and the recommendation based on data for 2016 are shown in Fig. 5.12. The weight for each simulation event is then a function of the true number of interactions of an event (determined at generator level),
and is the ratio of these two histograms at that value.

Figure 5.12: Number of events, measured in luminosity, as a function of the mean number of interactions in 2016 for simulation (left) and for the recommendation derived in data (right) [143].

To check how the pileup weights affected the distribution, the recommended re-weighting was applied to the region for a selection similar to the signal selection, but with the jet multiplicity relaxed to \( N_j \geq 3 \). The relaxation of the \( N_j \) requirement was motivated in an effort to understand better the general pileup profile with the bulk where distributions are better defined, and not to focus on discrepancies in the high jet multiplicity tail that this analysis focuses on. The improvement at high \( \rho \) is marked when applying the pileup re-weighting is significant, reigning in the high \( \rho \) events, which makes sense given the higher pileup profile used during generation. The drop at low \( \rho \) is a mix of several factors, with \( \rho < 10 \) being most likely a result of low statistics and \( 10 < \rho < 20 \) showing a similar downward trend that the LUMI POG observes as well.

### 5.1.7 \( H_T \) scale factor

Another scale factor unique to this analysis was derived to address discrepancies in the \( H_T \) distribution in the signal region. After applying all other scale factors, the \( H_T \) distribution for both simulation and data were plotted like in Fig. 5.14. Recall that the \( H_T \) of the analysis is calculated as the scalar sum of the \( p_T \) for all jets with \( p_T > 30 \text{ GeV} \) and \(|\eta| < 2.4\). From these plots, it is clear that the data to simulation ratio has a clear downward trend that
is slightly more pronounced in 2016 and 2018. This discrepancy is worrying since \( N_j \) is correlated with \( H_T \), and so mis-modeling of \( H_T \) could indicate mis-modeling of \( N_j \); thus, a scale factor was developed to address this.

One important caveat to the previous conclusion is that the \( H_T \) distributions above are shown in the signal region, and so conclusions about data to simulation discrepancies can be skewed by signal contamination. It is important, therefore, that any \( H_T \) scale factor be developed in a region that is not signal contaminated; one way of doing so is to look at a region with lower \( N_j \). The data to simulation ratio for the \( H_T \) distribution was checked for the \( N_j = 5, 6, \) and 7 bins, as shown in Fig. 5.15. While the \( N_j = 7 \) bin is used in the signal
region and does have some signal contamination, in general, the signal contamination is less than 1% in this bin. To minimize the number of plots presented here, only the distributions in 2016 are presented.

Figure 5.15: The $H_T$ distribution in the signal region selection of all simulation backgrounds and data for the $N_j = 5$ (left), $N_j = 6$ (middle), and $N_j = 7$ (right) events in 2016. The data to simulation ratio is shown in the bottom, with the normalization in the simulation done to the predicted theoretical cross section.

From Fig. 5.15 it is clear that the downward trend in the signal region is reproducible in the $H_T$ distributions for the lower $N_j$ bin as well. This gives us confidence that a scale factor can be derived in the background dominated low $N_j$ region and extrapolated to the higher $N_j$ signal region. To fit the trend in the data to simulation ratio, an exponential was used so that the scale factor at high $H_T$ where statistics are low would not be too sensitive to an early downward trend. These fits are shown in Fig. 5.16.

These fits show that the exponential fit differs in each $N_j$ bin, with the exponential parameter increasing in absolute value with increasing $N_j$. Therefore, to extrapolate to higher $N_j$, a scale factor parametrized by $N_j$ and $H_T$ was used of the form:

$$w(H_T, N_j) = a(N_j) \times \exp^{b(N_j) \times H_T}$$

such that both parameters of the exponential function are functions of $N_j$ themselves. To determine $a(N_j)$ and $b(N_j)$, a linear fit was done to the fit parameters of the exponentials of the $N_j = 5$, 6, and 7 bins, as shown in Fig. 5.17. These linear extrapolations are relatively flat, but when extrapolated to the higher $N_j$ bin, it can change the exponential shape quite
Figure 5.16: The data over simulation ratio in the $H_T$ distribution for $N_j = 8$ in 2016 (left). A direct fit to this distribution is compared to the extrapolated $H_T$ scale factor (middle) and then, the ratio of one over the other is used to determine an $H_T$ scale factor error used to calculate a systematic (right).

Figure 5.17: The linear fit to extrapolate the exponential fit parameters, $a(N_j)$ (left) and $b(N_j)$ (middle), of the data to simulation ratio in the $N_j = 5$, 6, and 7 bins to higher $N_j$ bins. The resulting $N_j$ dependent $H_T$ correction is shown on the right.

These extrapolations were done independently for the data set in each year, and the overall $H_T$ distribution agreement is shown in Fig. 5.18. Special care was done to make sure this new $H_T$ scale factor does not change the overall normalization of each simulation sample, and the overall $H_T$ agreement improves drastically for all three years.

Despite the improved agreement, there might be general concerns that this $H_T$ scale factor is rather ad-hoc and does not claim to address any underlying physics effects that is the motivation for all the other scale factors. The criticism is well founded, but upon
Figure 5.18: The $H_T$ distribution in the signal region selection of all simulation backgrounds and data in 2016 (left), 2017 (middle), and 2018 (right). The data to simulation ratio is shown in the bottom, with the normalization in the simulation done to the predicted theoretical cross section.

Further review and discussions with relevant experts, there is a general feel that at such high jet multiplicities, some of the parton showering and hadronization in the generation of simulation may not be sufficiently well modeled. This $H_T$ scale factor, in some sense, tries to address these issues post sample generation. Special care was made to comprehensively cover for any potential mis-modeling caused by the $H_T$ scale factor, resulting in four different $H_T$ related systematics directly related to this scale factor, including a conservative systematic that addresses the overall $N_j$ shape if no $H_T$ scale factor is applied at all. More details about these systematics can be found in Sec. 6.1.5.

5.2 $t\bar{t}$ estimation: fit function

Having described all the scale factors to the simulation, the next step is to describe the procedure to obtain the actual background estimates for $t\bar{t}$ + jets, QCD multijet, and the other backgrounds. As noted in the analysis strategy, the $t\bar{t}$ + jets estimate will be obtained through a fit function over all for NN bins. The fit function and its motivation will be described in this section.
5.2.1 Fit function motivation

The $t\bar{t} +$ jets estimate is obtained using a fit function motivated by theoretical predictions on jet scaling [144–146]. Generally, jet scaling in collider physics have been demonstrated to have two different patterns of scaling: Poisson scaling and staircase scaling.

For Poisson scaling, one takes the approximation where each gluon emitted is much softer than the particle that is emitting it, normally either a quark or another gluon produced from a hard process. In this approximation, emission of the gluon does not affect the trajectory or the momenta of the particle arising from the hard interaction, and so the probability of a gluon being emitted at any given point of the particle's trajectory is the same. Mathematically, in this approximation, one can imagine that soft gluon emission is very similar to soft photon radiation off of an electron, and in this case, as demonstrated by Peskin and Schroder, the probability of finding $n$ photons within energies between $E_-$ and $E_+$ is given by a Poisson expression [2]:

\[ P(n) = \frac{1}{n!} \lambda^n e^{-\lambda} \]
\[ \lambda \propto \frac{\alpha}{\pi} \log \frac{E_+}{E_-} \]

This makes sense from an intuitive point of view, as the numerator is the exponentiation of $n$ emission probabilities, determined by the average number of emissions, whereas the $n!$ is a result of the the combinatorics associated with bosons, which both photons and gluons are. Thus, in the eikonal approximation, it follows that the ratio $(R_{N_j+1/N_j})$ of the number of events with $N_j + 1$ jets to the number of events with $N_j$ jets follows the classical Poisson pattern:

\[ R_{N_j+1/N_j} = \frac{\lambda}{N_j + 1} \]

Of course, given that the above extrapolation is for a calculation from an Abelian process with photons to a non-Abelian process with gluons, this Poisson relationship is not perfect. However, studies have shown that this Poisson jet scaling nature does indeed manifest in
collider physics; particularly in cases where the jet acceptance cut is much lower than the energy of the hard process, such that the jet emitted only has a small impact on the hard process quark or jet. This is shown on the right in Fig. 5.19, where the leading jet is required to have $p_T > 150 \text{ GeV}$, but the jet acceptance cut is set at $p_T > 30 \text{ GeV}$.

Moving away from the soft gluon regime to a regime where the hard process jet energies and the emitted jet energies are of similar scale, the jet scaling behaves as:

$$R_{N_j+1/N_j} = c_0$$  \hspace{1cm} (5.16)

where $c_0$ is a constant. This is known as staircase scaling, which is the characteristic shape of the $N_j$ distribution in a plot where the y-axis is in log scale. The derivation of this jet scaling is a bit involved and can be found in Ref. [144]. This type of scaling has been experimentally observed in QCD jets production [146], in W production with jets [147], and in Z production with jets [148], with the latter shown on the left plot of Fig. 5.19.

Figure 5.19: The two different jet scaling patterns in a study measuring the cross section of jets in addition to a Z boson, with the staircase scaling (left) and Poisson scaling (right) [148].
Thus, the overall fit function, taking into account both types of jet scaling is:

\[ R_j = c_0 + \frac{c_1}{N_j + c_2} \]  \hspace{1cm} (5.17)

with \( c_0 \) being the staircase scaling and \( \frac{c_1}{N_j + c_2} \) being the Poisson scaling. A graphical representation of this fit function from simulated \( e^+e^- \rightarrow jets \) is shown in Fig. 5.20 where the x-axis denotes the number of jets in addition to the hard process. At low additional \( N_j \), the hard process jet energy is larger than the jet energy of the emitted jets, which is where the Poisson scaling regime dominates. At high additional \( N_j \), the hard process jet energy is more similar to the jet energy of the emitted jets, which is where the staircase scaling regime dominates.

![Figure 5.20](image)

Figure 5.20: The overall \( N_j \) distribution where the Poisson scaling dominates in the lower additional jet multiplicity regime and the staircase scaling for the high additional jet multiplicity regime. [144].

### 5.2.2 Fit function implementation

Early studies of the fit function in Eq. 5.17 found that all three parameters were highly correlated, and so an alternate functional form was derived in the form of a falling exponential,
which matches the shape shown in Fig. 5.20.

\[ R'_j = p_1 + (p_0 - p_1) \exp(p_2 \times (N_j - 7)) \]  

where \( p_0 \) is the normalization constant determined by the value at \( N_j = 7 \), \( p_1 \) is the asymptotic value, or the ratio in the staircase scaling, and \( p_2 \) is the exponential parameter that determines the rate of the staircase scaling. This parametrization, while better, still suffers from large correlations between the latter two parameters, so the function is recast once again in terms of the values evaluated at two different points, such that:

\[ f(x) = a_2 + \left[ \frac{(a_1 - a_2)^x}{(a_0 - a_2)^{x-2}} \right]^{\frac{1}{2}} \]  

\( a_0 = N(8)/N(7) \)
\( a_1 = N(10)/N(9) \)
\( a_2 = \text{asymptotic value as } x \to \infty \)
\( x = N_j - 7 \)

The first parameter \( a_0 \) is constrained by the statistics dominated, low signal-contamination \( N_j = 7 \) and \( N_j = 8 \) bin, whereas the second parameter \( a_1 \) determines how steep the falling function is. The third parameter \( a_2 \) is actually itself parametrized as:

\[ a_2 = \begin{cases} 
  a_1 - \frac{1}{d} & d \geq 1 \\
  a_1 - (2 - d) & d < 1 
\end{cases} \]  

but is kept as above in the equation to demonstrate its role as the asymptotic value. In this final parametrization of \((a_0, a_1, d)\), the three parameters are no longer correlated in fits, and the final shape with the staircase and Poisson scaling, though somewhat obscured, is maintained.

As a final step, a transformation is made to express \( f(x) \) in terms of \( N_j \) itself—denoted \( F(x) \)—rather than the ratio \( N(j + 1)/N(j) \). This decorrelates the \( N_j \) bins from each other and allows for the proper handling of the Poisson uncertainties. A recursive expression of
the form:

\[ F(0) = N_7 \]  
\[ F(x) = F(x-1)f(x-1) \] for \( x > 0 \)

where

\[ N_7 = N(7) \]
\[ f() = \text{The function defined in Equation 5.19} \]

is used. Note that \( x = 0 \) corresponds to the 7 jets bin and \( x = 5 \) corresponds to the inclusive \( \geq 12 \) jets bin. Expanding out the expression for each \( N_j \) bin, the following is obtained:

\( N_j = 7 : \) \[ F(0) = N_7 \]
\( N_j = 8 : \) \[ F(1) = F(0)f(0) = N_7f(0) \]
\( N_j = 9 : \) \[ F(2) = F(1)f(1) = N_7f(0)f(1) \]
\( N_j = 10 : \) \[ F(3) = F(2)f(2) = N_7f(0)f(1)f(2) \]
\( N_j = 11 : \) \[ F(4) = F(3)f(3) = N_7f(0)f(1)f(2)f(3)f(4) \]
\( N_j \geq 12 : \) \[ F(x \geq 5) = F(4)f(4) + F(5)f(5) + \ldots + F(9)f(9) + F(10)f(10) \]

For the last term, it is sufficient to include \( N_j \) bins from 12 through 18, as there were no data events recorded with more than 18 jets that pass the selection. Checks on how the fit function performs using pseudodata is presented in Chapter 7.

\section*{5.3 QCD multijet estimation: control region}

The second largest background is from QCD multijet events at around 5 to 6\% according to simulation. This background, however, poses a specific problem to this analysis because the simulation sample sizes are low in raw number of events. This means that for a select set of events passing the signal selection, the events can have a very large weight and large uncertainties. Since these statistical uncertainties are used as inputs to the fit as an
independent per NN bin per $N_j$ bin systematic, these large weights can easily mask shape differences in the $N_j$ shape between the different NN bins.

To avoid factoring in these large statistical fluctuations that have a disproportionate impact on the fit because of its immense flexibility, the QCD multijet estimate is instead taken directly from a QCD multijet dominated control region. The full selection for the control region was set to be as close to the signal region as possible, with the exception of the lepton, which was taken to have the opposite isolation. This selects for a certain subset of QCD multijet events, like semileptonic $B$ hadron decays, where the $b$ quark decays via the weak force to an $s$ quark and a lepton. In this case, the lepton will not be isolated and may be within the radius of the jet also arising from the decay. The baseline selection for the control region is:

- $H_T > 300\text{ GeV}$
- Exactly 1 non-isolated muon ($\text{miniIso} < 0.2$)
- No good electrons
- No good muons
- $N_j \geq 7$
- Pass all $E_T^{\text{miss}}$ filters
- For 2018B, pass the extra event veto

This control region was defined using only muons and not electrons, since the $p_T$ thresholds for the lowest un-prescaled muon trigger is at 50 GeV, while the $p_T$ threshold for the lowest un-prescaled electron trigger is at 105 GeV or 115 GeV for 2016 and 2017, respectively. Since the number of remaining events with at least seven jets and a lepton of over 105 GeV was too low to make an adequate estimate, only the muon control region is used. Furthermore, the requirement of a $b$ jet was removed, as that was linked mostly to the decay of the $t$ quark in the signal region and of which there are none in this control region.
The overall shape of the $S_{NN}$ and $N_j$ for the control region are shown in Fig. 5.21, and the data to simulation agreement in this control region is good. For the calculation of the $S_{NN}$, the four vector of the non-isolated muon was used in place of the isolated lepton, and an additional jet cleaning procedure was applied that removed all jets within a cone of radius 0.4 or less from the non-isolated muon.

Figure 5.21: The $N_j$ distribution and the $S_{NN}$ distribution of the QCD multijet dominated control region selection.

To ascertain whether this control region was a good place to obtain the shape of the QCD multijet estimate, the shapes of the $S_{NN}$, the $H_T$, and $N_j$ distributions in QCD multijet events in the signal region containing an electron and containing a muon were compared with the shape derived from the control region. All three shapes more or less agree within statistical uncertainty, though these statistical uncertainties are large. An example of this is the $N_j$ shapes in the four NN bins of the QCD multijet simulation sample, shown in Fig. 5.22 for both the signal region and the control region. In these plots, the events with the good muon or good electron are required to pass the signal region baseline, and the events with the non-isolated muon are required to pass the control region baseline. The shapes are normalized to one in order to compare the signal region and control region
At first glance, the agreement between the shapes may be difficult to see because the uncertainties on the signal region shapes are large. However, with prior knowledge that the $N_j$ distribution should be a smooth function and that the bins with the most statistics ($N_j = 7$ or 8) tend to have good agreement, these shapes give confidence that the control region is a good estimate for the QCD multijet estimate in the signal region. These plots also emphasize how large the uncertainties really are for the QCD multijet simulation in
the signal region.

To translate the event yield from the control region to the signal region, an overall transfer factor (TF) was used. Given the low number of raw events, one transfer factor was used for the overall normalization and luminosity calculated as:

\[ TF = \frac{N(SR)}{N(CR)} \]  

(5.23)

where \( N(SR) \) is the total weighted number of simulated QCD multijet events in the signal region and \( N(CR) \) is the total weighted number of simulated QCD multijet events in the control region. This transfer factor is then applied to all the shapes as one normalization factor. In terms of uncertainties associated with this prediction, the statistical uncertainties from the control region shapes are kept on a per \( N_j \) and per NN bin level, and an additional nuisance parameter associated with the overall normalization on the transfer factor is added to the overall fit. The size of the transfer factor systematic uncertainty is the sum in quadrature of the total percent error in the signal region and the control region simulation, which each are calculated as the sum in quadrature of the error in each \( N_j \) and NN bin.

All these values are shown for the four years in Tab. 5.1.

<table>
<thead>
<tr>
<th>Year</th>
<th>QCD SR MC</th>
<th>QCD CR MC</th>
<th>Transfer Factor ((\times10^{-2}))</th>
<th>% Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>3917.22</td>
<td>133694.90</td>
<td>2.93</td>
<td>16.6%</td>
</tr>
<tr>
<td>2017</td>
<td>3033.92</td>
<td>115726.86</td>
<td>2.62</td>
<td>22.4%</td>
</tr>
<tr>
<td>2018A</td>
<td>1423.75</td>
<td>57642.84</td>
<td>2.47</td>
<td>15.3%</td>
</tr>
<tr>
<td>2018B</td>
<td>1586.10</td>
<td>72428.21</td>
<td>2.19</td>
<td>17.5%</td>
</tr>
</tbody>
</table>

Table 5.1: Total number of weighted QCD multijet simulation events in the SR and the CR, along with the transfer factor and its uncertainty for each year.

5.4 Other Backgrounds and Signal Simulation

All other backgrounds and the signal samples have their estimates taken directly from simulation.
For the other backgrounds, they are broken down into two categories: those that had on average more jets than the $t\bar{t} + \text{jets}$ backgrounds ($t\bar{t} + X$, where $X$ is one or more vector bosons), and those that had on average less jets than the $t\bar{t} + \text{jets}$ background (Other), as shown in Fig. 5.23. Triboson events also had on average more jets than $t\bar{t} + \text{jets}$ event, but the statistical uncertainties were large enough such that event yield is consistent with the other backgrounds. This division between $t\bar{t} + X$ and other backgrounds was made because using only one nuisance parameter for all these backgrounds seemed to be an overgeneralization of the actual physics, but giving every single sub 5% background their own nuisance parameter also seemed unwarranted and slowed the fit down significantly. This choice of separation seemed to strike a nice balance without adding too many nuisance parameters. Studies on the impact to fits show that the separation of these backgrounds had little to no impact on the sensitivity of the analysis or fits to pseudodata. The Other backgrounds contain simulation from simulated $W$ boson + jets, Drell-Yan + jets, single top + jets, diboson, and triboson processes, and all of these processes have the same scale factors as listed in 5.1 applied.

![Normalized $N_j$ distributions for different backgrounds](image)

Figure 5.23: The normalized $N_j$ distributions for the different components of the Other background compared to $t\bar{t} + \text{jets}$ and $t\bar{t} + X$ in 2016 (left) and 2017 (right).

For signal, the treatment with respects to scale factors are very similar to the Other and $t\bar{t} + X$ backgrounds. The only major difference is that the $H_T$ scale factor was not
applied to the signal events, mostly because the $H_T$ scale factor was derived from data to simulation discrepancies in the low $N_j$ bins that have no signal. It is not clear if the $H_T$ scale factor has any applicability to the signal model shapes, and so the choice was made to not implement it. Overall, the $N_j$ shapes with and without the scale factor are not too different given the lower cross section and event yield of the signal models.
Chapter 6

Systematic Uncertainties

As expected from the procedure in this analysis, since the shapes of the jet multiplicity distributions have to be the same in all four NN bins for it to work, any discrepancies in the shapes that can arise from any mis-modeling of the simulation must be taken into account properly. This section will detail the different systematics used in this analysis and their derivation. Plots in this section are for the systematics generated for the year 2016 in order to conserve space.

6.1 tt + jets Systematics

The main background in the analysis is t\bar{t} + jets, comprising of over 85\% of the background in all bins. Being the dominant background, and also a background that is difficult to model, deriving t\bar{t} + jets systematics proved an extra challenge since there is no signal contamination free control region for t\bar{t} + jets. Therefore, extra care was given to ensuring that as many potential physics-motivated deviations from the nominal t\bar{t} + jets simulation N_j distribution shape were taken into account. In deriving these systematics, there are some systematics that may address similar issues; for example, the H_T systematics all cover for different aspects of the similar effect of the mis-modeling of H_T. However, without a control region that can be utilized to disentangle these effects and correlations and the knowledge that each systematic was derived to target a different potential cause for mis-modeling, the
conservative approach was taken with these systematics taken as independent.

The complete list of $t \bar{t} + \text{jets}$ systematics and what they account for are:

- $t \bar{t} + \text{jets}$ “nominal” systematic: statistical fluctuations in the simulation
- $b$ tag: efficiencies in the $b$ tag algorithm (in this case, DeepCSV algorithm)
- Lepton identification/isolation/trigger: the lepton working point efficiency, the isolation efficiency, and the efficiency in the emulation of the trigger in simulation
- Parton distribution function (PDF): uncertainties arising from approximations and theoretical calculations in the initial momentum distribution of the colliding particles [149]
- QCD renormalization and the factorization scale: theoretical uncertainties in defining the QCD scale for renormalization in perturbative calculations, particularly changing parameters $\mu_R$ and $\mu_F$ [135][150]
- Prefiring: correction for the timing shift in the ECAL that was not propagated to the trigger primitives, impacting mostly high $|\eta|$ jets
- Pileup: uncertainties related to discrepancies in the pileup profile between the simulation and the collected data
- Jet energy corrections (JEC): uncertainties in the corrections for the measured energy of jets in simulation with those of the detector
- Jet energy resolution corrections (JER): uncertainties arising from the resolution of the components in the hardware of the detector
- Initial/Final state radiation (ISR/FSR): uncertainties with respects to modeling of initial state and final state radiation in the generator and parton showering simulation
- $t \bar{t} + \text{jets}$ modeling: uncertainties in the matrix element and parton shower matching, color reconnection parameters, and underlying event matching in the generator and simulation
• $H_T$ reweighting: mis-modeling of $H_T$ in the simulation and potential mis-modeling of the $H_T$ scale factor

• $N_j$–$S_{NN}$ residual correlation: uncertainty in any residual correlation between $N_j$ and $S_{NN}$ derived using data in the QCD multijet dominated control region

• Input variable correction: uncertainty in the $N_j$ shape derived from discrepancies in the data and simulation for the jet $p_T$ and jet mass input variables to the NN.

Before delving too deeply into how each systematic is derived, it is important to note a special feature of this analysis. Because the $t\bar{t} +$ jets background is derived directly from a fit in the data, any changes in the $t\bar{t} +$ jets shape that affects all the NN bins equally will have no impact. For example, suppose there was some hardware issue that diminished the yield of all events greater than 10 jets by 10%. While the final fit result in this scenario would be different than the actual distribution for $t\bar{t} +$ jets in the universe, the fit function is flexible enough that the fit over all four NN bins would adjust for this change, and this change would not generate a false signal. Additional studies have validated this by inducing a 20% discrepancy in the overall yield of $t\bar{t} +$ jets events in simulation, and this was absorbed by the robust fit function. Therefore, the systematics focus only on any shape differences induced amongst the four NN bins as a result of each of the above uncertainties.

All systematics for $t\bar{t} +$ jets are implemented as multiplicative factors from the nominal fit value. For example, if there were two systematics with a value of $R_{1,Di,N_j=j}$ and $R_{2,Di,N_j=j}$ for NN bin $Di$ and $N_j$ bin $j$, the final fit value for each bin would be:

$$N_{obs,Di,N_j=j} = N_{fit,Di,N_j=j} \times R_{1,Di,N_j=j}^{\theta_1} \times R_{2,Di,N_j=j}^{\theta_2} \tag{6.1}$$

where $\theta_1$ and $\theta_2$ are the nuisance parameters for the corresponding systematic. As a result, all NN bins and $N_j$ bins are 100% correlated for each systematic, and so the relative shapes of the systematic across all four NN bins is very important in this set up.
6.1.1 Nominal systematic

One of the first concerns about using the $t\bar{t} + \text{jets}$ simulation in the analysis is that the sample size of the simulation sample is finite, and so there are small shape discrepancies in the four NN bins solely from statistical fluctuations from event generation. While the size of this discrepancy is small, as shown in Fig. 6.1, the differences in the $N_j$ distribution in the higher $N_j$ bins are visible and apparent. This is reasonable as these bins tend to also have the lowest number of simulated events.

![Figure 6.1: The $t\bar{t} + \text{jets}$ shape in simulation for data collected in 2016, 2017, 2018A, and 2018B divided into the four NN bins. The finite sample contributed to deviations between the shapes at the highest $N_j$ bins](image)

Noting these differences, a systematic was derived to take into account these residual shape differences. The procedure for calculating the systematic is as follows:

- Take the $N_j$ distribution for the $t\bar{t} + \text{jets}$ background for each of the four NN bins ($D_i$, where $i = 1,2,3,4$).
- Conduct an individual fit to each $N_j$ distribution using the fit function, getting a fit value of $a_{0}^{D_i}$, $a_{1}^{D_i}$, $d^{D_i}$ for each NN bin.
- Conduct a simultaneous fit to all four $N_j$ distributions together using the fit function, getting a fit value of $a_{0}^{D_{\text{All}}}$, $a_{1}^{D_{\text{All}}}$, $d^{D_{\text{All}}}$
- Take the ratio of these fitted distributions ($R_{\text{nom}} = N_{j}^{D_i}/N_{j}^{D_{\text{All}}}$).

The results of conducting this procedure is a value around 1.0 for each $N_j$ and NN bin, which acts as a multiplicative factor that can be assigned a nuisance parameter in the final
fit. When the nuisance parameter is pulled, all four shapes move simultaneously, which means that the uncertainty derived is 100% correlated over all $N_j$ and NN bins. This is true for every single one of the $t\bar{t} +$ jets systematics.

The shape of the “nominal” systematic is shown in Fig. 6.2 for 2016. As expected, the systematic for the lowest $S_{NN}$ bin (D1) is essentially flat, as this bin has the most statistics and dominate the $t\bar{t} +$ jets background shape. The systematic is non zero for the higher $S_{NN}$ bins D3 and D4, and so this systematic mainly has an impact on these bins.

![Figure 6.2: The “nominal” shape systematic for the year 2016 derived as a ratio of the fitted $N_j$ distribution in each NN bin to the fitted total $N_j$ distribution.](image)

### 6.1.2 Event weight based systematics

The next set of systematics (b tag, lepton id/iso/trigger, PDF, QCD scale, and pileup) are all derived similarly because they are all implemented as scale factors/reweighting for the simulation events in the analysis.

For some of these systematics, such as b tag, lepton id/iso/trigger, and pileup, each event is weighted by a scale factor derived in data/MC comparisons, and so there are uncertainties associated with each of those procedures that can be used to vary the scale factor up and down. For the PDF uncertainties, during the generation of the events, different parameters of the PDF set were adjusted such that there were a total of 100 different variations with 100 different PDF weights. The median of these 100 different variations is taken as the nominal PDF weight, whereas the error is calculated as the PDF scale factor values in these 100
different variations such that it creates an envelope holds 68% of all PDF weights. For the QCD renormalization and factorization scale weights, \( \mu_F \) and \( \mu_S \) were either kept constant, increased by a factor of 2, or decreased by a factor of 2 in a total of 9 different variations, and the up/down error was taken as the difference between the largest/smallest value with the mean.

In each case, these errors are propagated to each of the \( N_j \) distributions shapes for the four NN bins by reweighting each event by the scale factor adjusted by 1 standard deviation up/down for the up/down variation. Once again, since any variation to the t\( t \) + jets \( N_j \) shape that affects all four bins equally is not important, these shapes are compared to the total t\( t \) + jets \( N_j \) shape in the corresponding up/down variation.

With these shapes, we follow the following procedure:

- Take the \( N_j \) distribution for the t\( t \) + jets background for each of the four NN bins for up/down (\( D_{i,\text{up/down}} \), where \( i = 1,2,3,4 \)) and normalize these shapes to unit area.
- Divide the \( N_j \) distribution by the total \( N_j \) distribution, also with the same fluctuation and also normalized to unit area (\( D_{\text{All,up/down}} \)).
- Take the raw ratio of these distributions (\( R_{i,\text{raw}} = N_j^{D_{i,\text{up/down}}}/N_j^{D_{\text{All,up/down}}} \)).
- Divide by the \( R_{\text{nom}} \), the nominal systematic in Sec 6.1.1 to get the final systematic \( (R_{i,\text{syst}} = R_{i,\text{raw}}/R_{i,\text{nom}}) \)

The first three steps are what is expected for deriving a standard multiplicative shape systematic, and is similar to the derivation of the nominal systematic without the fits. The final step of dividing by \( R_{\text{nom}} \) is done in order to avoid double counting the differences in shape arising from statistical fluctuations in the simulation that is all factored into these event weight based systematics inherently because they utilize the simulation. The shapes of these systematics are shown in Fig. 6.3 and the maximum size of the systematic is around 5%. This holds true across all four time eras, where the shape discrepancies in the four NN bins resulting from fluctuations in these scale factors are small in magnitude.
Figure 6.3: All event weight based shape systematics (up variation) for the year 2016 derived as a ratio of the $N_j$ distribution in each NN bin to the total $N_j$ distribution, and then further divided by $R_{i,\text{nom}}$.

### 6.1.3 Bin migration systematics

The next category of systematics are the standard systematics that may cause events to migrate between different bins: ISR, FSR, JEC, and JER. In the case of the first two, the mis-modeling of ISR and FSR can make an event have one more (less) jet arising from an additional (loss of a) gluon emission in the simulation. Similarly, the jet energy corrections and jet energy resolution corrections can cause a jet that passed the selection criteria in the baseline (for example, being over 30 GeV) to fall out of acceptance, placing the event in a different histogram bin. Finally, bin migrations across the different NN bins are also possible since these effects can even change the order of the ranking of the jets, and thus affect the NN score assigned to the event.

Therefore, for these systematics, the corrections and fluctuations have to be applied to the original jet collection before each event is passed into the neural net and propagated all the way through the analysis with its new $S_{NN}$ to the final $N_j$ distributions in the four NN bins. From there, the procedure is similar to deriving the nominal systematic, where each $N_j$ distribution in the four NN bins are fit using the fit function independently and simultaneously, getting four different ratios for the four NN bins, $R_{i,\text{raw}}$. Just like in the event based systematics, this raw ratio is further divided by the $R_{i,\text{nom}}$ to ensure not to double count those effects. The shapes of this set of systematics is shown in Fig. 6.4.
Figure 6.4: All bin migration based shape systematics for the year 2016 derived as a ratio of the $N_j$ distribution in each NN bin to the total $N_j$ distribution, and then further divided by $R_{i,\text{nom}}$.

It is immediately noticeable that these systematics are of a larger magnitude than the corresponding systematics for the nominal case and the event weight based uncertainties. This makes sense, since given the large jet multiplicity of this region of phase space, there are a number of different ways for these corrections to affect the overall event. The size of the systematics goes as high as almost 25% in one of the highest bins in D3 in 2016, but generally within the 10% range for bins of $N_j \leq 11$.

One notable point about the JEC/JER systematics is that sometimes the up variation and the down variation affects a particular $N_j$ bin in a particular NN bin such that the fluctuation is in the same direction; for example, all four variations cause a downward change in shape in the $N_j = 9$ bin in D3. While the exact mechanism is not fully traceable, it is likely a confounding of the sometimes competing effects of the change in the NN score of the event and the changing number of jets that pass acceptance. Since the systematics are fully correlated amongst all $N_j$ bins, it was decided to keep both shapes derived from the up and down variations of the JEC/JER as independent systematics, with both being able to be pulled up or down independently of each other. While conservative, since the shape of the systematic ultimately determines whether the fit will utilize a systematic, it was important to make sure all shapes were covered.
6.1.4 Parton shower modeling systematics

Under recommendation of the top quark modeling (TOP) and the jet and missing transverse energy (JetMET) groups in the collaboration, an extra set of five systematics were added that address mis-modeling that originate from parameters in both POWHEG and PYTHIA, known collectively as the parton shower systematics. To be exact, the ISR and FSR systematics from the last section are also considered systematics related to the parton showering, but the five mentioned here are specific to analyses where t quark modeling is important.

The first parton shower modeling systematic relates to color reconnection and how the quarks emerging from a particle collision may group together and interact with other color fields in the post-collision environment. Given the number of colored particles in the high instant luminosity environment of the LHC, these effects can account for around 20-40% of the top mass calculation theoretical uncertainties [151]. Currently, this is taken into account by turning on a parameter in the model in PYTHIA 8 known as the Early Resonance Decays, and generating an entirely new sample of t\bar{t} + jets with the ability for color reconnection to happen. The probability of color reconnection occurring for any particular quark-quark interaction is a function of the how soft the interaction is, as quarks with less energy are more likely to interact with surrounding color fields.

The second set of parton shower modeling systematics relates to the matrix element and parton shower matching scale [152]. This is specific to the generation of t\bar{t} + jets events using POWHEG and involves turning on a parameter known as \textit{hdamp}, which regulates the high \textit{p}_T radiation by dampening real emissions in the generator by a factor of \[ \frac{h_{damp}^2}{p_T^2 + h_{damp}^2} \] where \textit{h}_{damp} is set to the top mass of 172.5 GeV. This implementation of this parameter was integral in developing the special CUET8M2T4 Tune used specifically for t\bar{t} + jets simulation in 2016.

Finally, the last set of parton shower modeling systematics is derived from a concept known as “underlying event” [153][154]. Defined as the colored particles in an event not in the hard scatter, the underlying event consists of ISR and FSR arising from particles not from the t\bar{t} + jets interaction, other multiparton interactions, and beam beam remnants. Since
these particles recoil against the system and may interact with the hadronizing quarks, it is important to model this well and to take into account uncertainties in how the underlying event is modeled in PYTHIA.

In all the above cases, since these effects are induced at the generator or showering level, independent samples have to be generated with these parameters turned on in the case of color reconnection and fluctuated up and down in the case of hdamp and underlying event. From here, the procedure to get the systematic is very similar to the $t\bar{t} +$ jets nominal systematic, in which all events that pass the baseline selection are then divided amongst the NN bins. These four $N_j$ distribution shapes are fitted independently and simultaneously, and the ratio of the former to the latter provides the raw systematic size. Finally, as with the other systematics, these systematics are further divided by $R_{i,nom}$. The resultant shape for the systematics are shown in Fig. 6.5 and the systematics have a larger effect on the $t\bar{t} +$ jets shape, as expected. The magnitude of the systematic reaches 50% in the highest $N_j$ bin in D4, but remains within 20% for the majority of the bins.

Figure 6.5: All $t\bar{t} +$ jets modeling parton shower based shape systematics for the year 2016 derived as a ratio of the $N_j$ distribution in each NN bin to the total $N_j$ distribution, and then further divided by $R_{i,nom}$. CR is the color reconnection systematic, hdamp is the ME-PS matching systematic, and UE is the underlying event systematic.
6.1.5 HT reweighting systematics

Outside of $N_j$ and $S_{NN}$, the next most important physical quantity in this analysis is the $H_T$. The $H_T$ of the event is closely tied to all the input variables of the neural net, including the jet $p_T$, the Fox-Wolfram moments, and the jet-momentum energy tensor eigenvalues. Therefore, any mis-modeling of $H_T$ can be indicative of mis-modeling in simulation. In the analysis, this is already partially addressed by the $H_T$ scale factor, but since the $H_T$ correction is rather ad-hoc, a few systematics were designed to address potential failings of the $H_T$ scale factor.

Recall that the $H_T$ scale factor is derived from looking at the $H_T$ distribution in data for the $N_j = 5, 6, 7$ bins, separately, and comparing it to the simulation. An exponential fit is then conducted on the data to simulation ratio, and the two parameters in this fit are then linearly extrapolated to higher $N_j$ bins.

The first $H_T$ systematic directly probes how well this extrapolation works in the next lowest jet multiplicity bin ($N_j = 8$). First, the data to simulation ratio of the $H_T$ distribution of just the $N_j = 8$ bin is fit to the same exponential function directly and this fit result is compared to the extrapolated function, as shown in Fig. 6.6.

From here the ratio of the extrapolated fit over the actual fit result is used to generate a multiplicative factor that is dependent on $H_T$ and used to fluctuate the $H_T$ scale factor.
For example, an event with 8 jets and 500 GeV $H_T$ nominally would have a $H_T$ scale factor of 1.1. The $H_T$ up variation of the systematic, which is about an additional 3%, would then cause this event to have an $H_T$ scale factor of $1.1 \times 1.03$. This is done for all events, each with its individual $N_j$ dependent $H_T$ scale factor, but varied up or down by the percentage calculated here in the $N_j = 8$ bin. These events were then placed into the four NN bins, and the shape differences were utilized to create a systematic.

There is some criticism that can be garnered from this method. For example, the $N_j = 8$ bin is more likely to contain signal than the $N_j = 7$ bin, and so the variation may be a result of signal contamination. Furthermore, the $N_j = 8$ bin does not necessarily determine whether the extrapolation is adequate for $N_j \geq 8$. While these criticisms are accurate, this method is only used to gauge approximately how large the $H_T$ discrepancy can be. In lieu of finding a $t\bar{t} +$ jets control region with even less signal contamination, which was sought after and was not found, the extrapolation from low $N_j$ really is the only handle available to determine a good $H_T$ scale factor in the signal region.

The next $H_T$ systematic addresses the possibility for the $H_T$ scale factor to be inaccurate for events with high $H_T$. Looking at Fig. 6.6, the uncertainty on the ratio for events with $H_T \geq 2000$ GeV is relatively large. This is true also for the $H_T$ distributions for $N_j = 5, 6,$ and 7, as noted in Fig. 5.16. Since the number of events for high $H_T$ events are generally lower, the high $H_T$ tail is more susceptible to statistical fluctuations. Thus, an alternate $H_T$ scale factor is derived such that all events with an $H_T$ greater than 2000 GeV are weighted with the scale factor value at 2000 GeV, as shown in the left plot of Fig. 6.7. Using this alternate scale factor, the $H_T$ and the $N_j$ distributions for the $t\bar{t} +$ jets simulation are compared with the respective distributions using the nominal scale factor. The ratios of the $N_j$ distributions with the alternate scale factor over the $N_j$ distributions with the nominal scale factor, divided up by the different MVA bins per year, are added as a systematic for the $t\bar{t} +$ jets background shape.

Similarly, another $H_T$ systematic was derived to address potential mis-modeling introduced by the extrapolation to higher $N_j$ bin from the $N_j = 5, 6,$ and 7 bins. Since the
linear dependence for the extrapolation to higher $N_j$ was decided based on observation from the fits to the lower $N_j$ $H_T$ distributions, one can imagine an alternative $H_T$ scale factor using only the $N_j = 7$ bin $H_T$ distribution. This motivated utilizing a second alternative $H_T$ scale factor using only the $N_j = 7$ fit results, as shown in the right plot in Fig. 6.7. The ratio of the $N_j$ distribution scaled with this version of the $H_T$ scale factor compared to the nominal $H_T$ scale factor in the four NN bins is the third $H_T$ systematic that is used.

Finally, the last $H_T$ systematic derived came from the direct observation of the data to simulation ratio of input variables to the neural net. In 2017, there were some input variables in which application of the $H_T$ scale factor actually made the agreement worse. Of particular note are the jet mass variables, with the leading jet mass shown without the $H_T$ scale factor and with the $H_T$ scale factor on the left and right plots of Fig. 6.8 respectively.

The data to simulation agreement is actually much better without the $H_T$ scale factor, and this is true for the other jet mass variables in 2017. This, however, is not true for the other input variables or the $H_T$ distribution itself. Since this hints that the $H_T$ scale factor may potentially be a source of mis-modeling itself, a new systematic was derived by turning off the $H_T$ scale factor and propagating those changes to the four NN bins. From here,
Figure 6.8: The leading jet mass distribution without (left) and with (right) the $H_T$ scale factor in 2017. The jet mass is one of the input variables to the neural net.

the same procedure for deriving a systematic with fitting independently and simultaneously and then taking a ratio to the total shape is used. This systematic also addresses how the $H_T$ scale factor can induce shape changes because the binning is derived with the $H_T$ scale factor applied to the $t\bar{t} + \text{jets}$ simulation.

Figure 6.9: The “up” fluctuation for the four $H_T$ systematics for the year 2016. The first systematic (8j Extrapolation) utilizes the error from comparing the fit to the extrapolation in the $N_j = 8$ $H_T$ distribution, the high $H_T$ tail systematic addresses mis-modeling at high $H_T$, the constant $N_j$ systematic addresses the $N_j$ dependence of the scale factor, and the no $H_T$ SF addresses shape changes as a result of the scale factor.

None of these systematics are greater than around 10% in magnitude, but do have
distinct shapes in D1 and D2, having the largest fluctuation in the \( N_j = 9 \) and 10 bins.

### 6.1.6 Data driven simulation mis-modeling systematic

So far, most of the systematics were derived utilizing the simulation and applying different fluctuations to the simulation. The \( H_T \) systematics utilize data to some degree, but does not use the \( S_{NN} \) in any way. One particular concern that has yet to be addressed then is mis-modeling of the \( S_{NN} \) and \( N_j \) correlation in the simulation.

To address this, a new systematic is derived using the control region where the QCD multijet estimate is obtained, since this is the only non signal contaminated region that also has high enough jet multiplicity required by the neural net. The first step in doing this is to determine whether the QCD multijet simulation in the control region has a similar \( N_j - S_{NN} \) correlation to that of the data in the control region. Fig. 6.10 shows the data versus simulation agreement in the \( S_{NN} \) score, with the statistical uncertainty for the total simulation shape shown in green. Immediately, two things are evident in these plots: 1) the agreement is very good in general across all \( N_j \), and 2) the error bars for the MC simulation are very large. The former is a good indicator that the data and simulation agree, and the latter further reinforces the idea that the QCD multijet sample is lacking in statistics at the high jet multiplicity end.

From here, a direct comparison of the \( t\bar{t} + \) jets simulation \( S_{NN} \) shape for each \( N_j \) bin \( (S_{NN,N_j=i,\text{tt}}) \) with the QCD multijet simulation \( S_{NN,N_j=i,QCD} \) for each \( N_j \) bin would show that the two are not the same. Since the training utilized the \( t\bar{t} + \) jets simulation specifically and optimized for the separation of these events from signal events, there is no apriori reason why the corresponding \( S_{NN} \) shapes in a QCD multijet simulation should be the same. Thus, a more clever way to quantify the \( N_j - S_{NN} \) correlation was necessary: checking the ratio between the total \( S_{NN} \) shape \( S_{NN,\text{All}} \) and the per \( N_j \) \( S_{NN} \) shape \( S_{NN,N_j=i} \). By making judgements based on these ratios, the absolute shape of the \( S_{NN} \) distribution is not as important as how the shape changes with increasing \( N_j \). If these ratios agree for data, the QCD multijet simulation, and the \( t\bar{t} + \) jets simulation, it engenders confidence
Figure 6.10: The agreement in the data versus simulation in data collected in 2016 for the $S_{NN}$ distribution per $N_j$, with (from left to right) $N_j = 7, 8, \text{ and } 9$ on the top row and $N_j = 10$ and $N_j \geq 11$ on the bottom row. The $\geq 11$ bin was chosen as a result of simulation sample size.
that the simulation can accurately model the $N_j$-$S_{NN}$ correlation.

These ratios ($S_{NN,All}/S_{NN,N_j=i}$) are shown for the data in the control region, the QCD multijet simulation in the control region, and the $t\bar{t}$ + jets simulation in the signal region in Figs. 6.11, 6.12, and 6.13 respectively. To minimize the impact of the statistics of the QCD multijet, the $S_{NN}$ distributions were binned in 5 bins instead of the 10 bins shown in Fig. 6.10. The choice to look at the ratio of the total shape over the individual shape was decided upon over the inverse in order for the ratios at high $N_j$ to be within the range shown - there is no physical significance to this decision.

Figure 6.11: The shape differences in the $S_{NN}$ distribution per $N_j$ relative to the total $S_{NN}$ distribution for data from the control region collected in 2016, with (from left to right) $N_j = 7$, 8, and 9 on the top row and $N_j = 10$ and $N_j \geq 11$ on the bottom row.

The next step is to compare these shapes. One especially subtle point in comparing these ratios is how to adequately calculate the error bars. Since the events in the $S_{NN,N_j=i}$ shapes are a subset of the events in the $S_{NN,All}$ shapes, the error bars of this ratio must be carefully calculated. This was done utilizing the Clopper-Pearson confidence intervals, which are often used for binomial distributions [154]. This makes sense in this context since the $S_{NN,N_j=i}$ shape is directly obtained by determining whether an event has $N_j = i$ or
Figure 6.12: The shape differences in the $S_{NN}$ distribution per $N_j$ relative to the total $S_{NN}$ distribution for the QCD multijet simulation in the control region in 2016, with (from left to right) $N_j = 7, 8, $ and $9$ on the top row and $N_j = 10$ and $N_j \geq 11$ on the bottom row.

Figure 6.13: The shape differences in the $S_{NN}$ distribution per $N_j$ relative to the total $S_{NN}$ distribution for the $t\bar{t} +$ jets simulation in the signal region in 2016, with (from left to right) $N_j = 7, 8, $ and $9$ on the top row and $N_j = 10$ and $N_j \geq 11$ on the bottom row.
does not, making it binomial. In the case of having weighted events in the simulation, the error is calculated where the weighted simulated events were rounded down to the nearest integer, thus being conservative in the calculation of the Clopper-Pearson error bars.

Another subtle point that needs to be taken into account is the proper normalization when comparing these shapes. In the QCD multijet simulation, there are a few very highly weighted events arising particularly from the inclusive sample. These highly weighted events can pull the overall normalization of a shape and create seemingly large discrepancies between two shapes. To have a more fair comparison, the ratios were normalized using a weighted average that is proportional to the inverse square of the Clopper-Pearson error bars, so that bins with very large uncertainties have the smallest impact on the normalization factor. The final shape comparisons are shown in Fig. 6.14 and the size of the QCD multijet error bars relative to the $t\bar{t} +$ jets and the data error bars reinforce the decision to take special care of the errors in the comparison.

Figure 6.14: The $S_{NN,All}/S_{NN,N_{ij}}$ comparisons between data in the control region, QCD multijet simulation in the control region, and $t\bar{t} +$ jets simulation in the signal region for the year 2016.

Overall, the agreement between the simulations and the data for these ratios are good,
and thus, it is concluded that a comparison of the data in the control region with the \( t\bar{t} + \text{jet simulation} \) in the signal region is reasonable with respect to the \( N_j - S_{NN} \) correlation. Here, it should be stressed that in the next steps in the derivation of the systematic, the QCD multijet simulation is no longer used in order to not be impacted by the large statistical errors of that sample. While any systematic derived from the shape of the data in the control region and the \( t\bar{t} + \text{jets simulation} \) in the signal region will also inherently take into account any discrepancies arising from the applicability of a QCD multijet dominated control region to a \( t\bar{t} + \text{jets dominated signal region} \), given the lack of any \( t\bar{t} + \text{jets control region} \), this is the best way to get a data-driven systematic. Additional concerns about any residual \( N_j - S_{NN} \) correlation in data that may only impact \( t\bar{t} + \text{jets} \) are well founded; however, these concerns are expected to be covered by the \( t\bar{t} + \text{jets specific parton shower systematics} \), described in Sec. 6.1.4.

Proceeding forward, the ratios of the data \( S_{NN,\text{All}}/S_{NN,N_j=i} \) to the \( t\bar{t} + \text{jets simulation} \) \( S_{NN,\text{All}}/S_{NN,N_j=i} \) shapes, or the blue divided by the green in Fig. 6.14, create a 2D grid of data to simulation multiplicative scale factors dependent on both \( N_j \) and \( S_{NN} \) \((R(N_j,S_{NN}))\), where the \( S_{NN} \) are grouped into discrete bins 0.2 \( S_{NN} \) in width (as in the plots above). The error on this set of scale factors \((\delta R(N_j,S_{NN}))\) is calculated using the proper error propagation through the double ratio of two independent samples. From this set of scale factors and errors, a new \( t\bar{t} + \text{jets} \) systematic is derived through the following procedure:

- First, four new \( N_j \) histograms \((D_{i,j=0})\) were filled with all of the events in the \( t\bar{t} + \text{jets simulation} \) that pass the signal selection baseline. For each event, the total event weight was multiplied by a factor chosen randomly from a Gaussian with a mean of \( R(N_j,S_{NN}) \) and a sigma of \( \delta R(N_j,S_{NN}) \).

- These four \( N_j \) histograms are summed together to get the total histogram with this new multiplicative scale factor applied \((D_{\text{All},j=0})\).

- The four histograms and the total histograms are each normalized individually to
unit area, and then the $N_j$ distribution for each NN bin $i$ is divided by the total 
\( \frac{D_{i,j=0}}{D_{All,j=0}} \)

- The resulting four histograms are used to define $R_{syst,j=0}$, just like in the other $t\bar{t} +$ jets systematics.

- The ratio of $R_{syst,j=0}$ to $R_{nom}$ gives the shape for the systematic.

- Repeat this procedure 100 times ($j = 0$ to $j = 99$).

- For each $N_j$ bin and NN bin, take the mean value of the systematic in that bin for all
  100 toys as the nominal value.

- Take the error of the systematic as the combination of two components in quadrature:
  (1) the width of the distribution of the 100 means as the statistical component and
  (2) the deviation from 1 as an estimate of how well the shape is known.

The result of this procedure is the systematic shown in Fig. 6.15. The impact of this
systematic is small in the statistics rich bin D1, but it can be around 15 - 20% in the higher
$N_j$ bins.

![Figure 6.15: The nominal shapes for the $N_j$-$S_{NN}$ correlation systematic derived using the
data in the control region and the $t\bar{t} +$ jets simulation in the signal region.](image)

One particular concern about this final result is that amongst the 100 different toys that
were thrown, there were many different shapes for this systematic, and the average of all
those shapes is not necessarily consistent with any one particular shape. Since the shape of the systematic is very important and that this is the only data-driven systematic shape, ideally the systematic should have more flexibility to potentially cover for all the different shapes in the toy studies. In order to provide this flexibility, each one of the shapes in each NN bin was fit by a 2nd order polynomial, such that each NN bin has three parameters that represented the systematic shape. Each fit parameter was given its own nuisance parameter in the final fit such that a pull of one sigma of this nuisance parameter corresponds to the error in the fit parameter, and so this gives the fit some flexibility to alter the shape of this systematic. The results of these fits and the one standard deviation pull band overlaid is shown in Fig. 6.16.

Figure 6.16: The fits to the 2016 $N_j-S_{NN}$ correlation data-driven systematic with the one standard deviation band showing the flexibility of the systematic shape that the fit can pull. The shapes are for NN bin D1 (top left), D2 (top right), D3 (bottom left), and D4 (bottom right).

As expected, the range of shapes for the first NN bin (D1) is very small because of the
small error bars and general shape constraints from the high number of events, but the shapes can vary a lot more in the higher NN bins of D3 and D4. With this flexibility, there is confidence that the totality of these shapes will cover this hard to pin down correlation.

### 6.1.7 Input variable data-simulation discrepancy systematic

The last set of $t\bar{t} +$ jets systematics comes from another study particularly looking at the data and simulation agreement for the input variables to the neural net. Particularly, when comparing the data to simulation without any $H_T$ reweighting in the variables of jet $p_T$ and jet mass, there is a downward trend that is exacerbated with increasing jet rank, shown more explicitly for the jet $p_T$ in control region in Fig. 6.17. Of course, since this is the signal region and there is signal contamination, data to simulation agreement must be interpreted with caution. However, looking at the exact same set of plots in the QCD multijet dominated control region in Fig. 6.18, the trend is also apparent there.

![Figure 6.17](image-url)

Figure 6.17: The data compared to the simulation in the signal region for 2016 with the leading jet (ranked by $p_T$) in the top left, the subleading jet in the top middle, and so on until the sixth jet in the bottom right.
Figure 6.18: The data compared to the simulation in the control region for 2016 with the leading jet (ranked by $p_T$) in the top left, the subleading jet in the top middle, and so on until the sixth jet in the bottom right.

Thus, to address this trend in the input variable plots, another systematic was derived using the QCD multijet simulation in the control region and the data in the control region. Since the data in the control region is dominated by the QCD multijet component, only the QCD multijet simulation was used in the derivation of this statistic. Fig. 6.19 shows that these trends exist even without using the other backgrounds in the control region, and the trend line in the ratio is just used to direct the eye and is not used in the derivation of the systematic.

From these shapes, different manipulations were done to see whether better agreement can be obtained. One of these manipulations was to stretch the width of the shape in the QCD multijet distributions until it matched that of the data distributions. To find the value for the stretching, both sets of shapes were stretched by 1/RMS of the distribution, with the result in Fig. 6.20. With this manipulation, the shapes agree, with the fit in the ratio plot becoming much flatter.
Figure 6.19: The data compared to the simulation in the control region for 2016 with the leading jet (ranked by $p_T$) in the top left, the subleading jet in the top middle, and so on until the sixth jet in the bottom right.

The values of the RMS for each of the distributions are shown in Tab. 6.1. While the individual RMS values for each distribution are different based on the jet rank, the scaling ratio of data over simulation is consistent across all bins, arriving at a ratio around more or less 0.95.

<table>
<thead>
<tr>
<th>Jet Rank</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data RMS</td>
<td>144.47</td>
<td>106.53</td>
<td>63.06</td>
<td>41.85</td>
<td>27.56</td>
<td>18.05</td>
<td>11.28</td>
</tr>
<tr>
<td>QCD RMS</td>
<td>151.98</td>
<td>112.78</td>
<td>65.97</td>
<td>44.91</td>
<td>28.98</td>
<td>19.01</td>
<td>11.69</td>
</tr>
</tbody>
</table>

| Ratio    | 0.95  | 0.94  | 0.96  | 0.93  | 0.95  | 0.95  | 0.97  |

Table 6.1: The RMS values of the jet $p_T$ distributions for data in the control region and QCD multijet simulation by jet rank (ranked by $p_T$) in 2016, followed by the ratio of the two.
Figure 6.20: The data compared to the simulation in the control region for 2016 with the leading jet (ranked by $p_T$) in the top left, the subleading jet in the top middle, and so on until the sixth jet in the bottom right.

Having obtained this value, a similar study was conducted on the jet mass distributions, comparing the shapes of jet mass in the QCD multijet simulation and the data in the control region. The result of this study shows that the RMS ratios of data over simulation for the jet mass variables is also 0.95 for 2016. Therefore, to derive the systematic, the jet $p_T$ and jet mass collections per event from the $t\bar{t} +$ jets simulation were scaled by 95% of its total value before obtaining the neural net score. The main impact this scaling has on an event is twofold: (1) jets with a $p_T$ close to 30 GeV will fall out of acceptance, potentially causing events to migrate to a lower $N_j$ bin and (2) the NN score will change because the inputs are affected, but what the final change is depends on the relative weights within the NN. In any case, the impact of these changes were propagated to the $N_j$ distributions for the four NN bins and a systematic was derived from this, resulting in the shapes shown in Fig. 6.21. The average size of this systematic is around 5%, with a maximum size of around 15%.
Figure 6.21: The shape systematic for the year 2016 derived from scaling the jet $p_T$ and jet mass for each $t\bar{t} +$ jets simulation event based on a scale factor derived in the control region.

### 6.2 Other background systematics

For the non-$t\bar{t} +$ jets backgrounds, which are subdivided into the QCD multijet, $t\bar{t} + X$, and other categories, there are a different, but simpler set of systematics. Recall that the other category consists of Drell-Yan + jets, single top + jets, diboson ($WW, WZ, ZZ$), triboson, and $W +$ jets events.

For the QCD multijet estimate, since the shape is taken directly from the control region data, there is only one systematic implemented. This comes directly from the transfer factor used to normalize the data, where the transfer factor is set as the ratio between the total number of weighted QCD multijet simulated events that pass the baseline selection in the signal region and the total number of weighted QCD multijet simulated events that pass the control region selection. The error in this transfer factor is calculated as the quadrature sum of the percent error per $N_j$ bin in the signal region with the percent error per $N_j$ bin in the control region over all $N_j$ bins. This error value ranges from 15 - 23% depending on the year.

For the $t\bar{t} + X$ and the Other backgrounds, the shapes used in the fit are taken directly from simulation. The suite of systematics used for these backgrounds are similar to that of $t\bar{t} +$ jets, including a systematic for each of the scale factors and the jet energy related corrections. Two systematics that were not included in the $t\bar{t} +$ jets systematics that are
needed for the $t\bar{t} + X$ and Other backgrounds are the luminosity measurement uncertainty systematic and an overall cross section uncertainty systematic. The luminosity uncertainty is provided by measurements conducted by the collaboration and is around 2.3 to 2.5%. The overall cross section uncertainty is set to 30% to be conservative. Since these backgrounds constitute less than 10% of all events, a 30% normalization still has a minor impact on the overall fit. These two systematics were not needed for the $t\bar{t} +$ jets simulation because the final shape and normalization is determined by the fit to data and not directly from simulation.

More explicitly, the list of systematics for the $t\bar{t} + X$ and Other backgrounds are:

- Luminosity uncertainty: overall luminosity measurement uncertainties around 2.3 - 2.5% depending on the year.
- b tag: efficiencies in the b tag algorithm (in this case, DeepCSV algorithm)
- Lepton identification/isolation/trigger: the lepton working point efficiency, the isolation efficiency, and the efficiency in the emulation of the trigger in simulation
- Parton distribution function (PDF): uncertainties arising from approximations and theoretical calculations in the initial momentum distribution of the colliding particles
- QCD renormalization and the factorization scale: theoretical uncertainties in defining the quantum chromodynamics scale for renormalization in perturbative calculations, particularly changing parameters $\mu_R$ and $\mu_F$
- Pileup: uncertainties related to discrepancies in the pileup profile between the simulation and the collected data
- Jet energy corrections (JEC): corrections for the measured energy of jets in simulation with those of the detector
- Jet energy resolution corrections (JER): uncertainties arising from the resolution of the components in the hardware of the detector
• $H_T$ reweighting (8j extrapolation): mis-modeling of $H_T$ in the simulation and potential
mis-modeling of the $H_T$ scale factor

• Cross section: an additional normalization uncertainty in the cross section set to 30%
to be conservative.

For these systematics, instead of using a multiplicative factor as with $t\bar{t}$ + jets system-
atics, histograms were made with the appropriate uncertainty implemented and passed into
the fit setup. The fit is then able to scan from the nominal histogram to the one standard
deviation error histogram using a nuisance parameter that correlates all $N_j$ bin and NN
bins. In addition to these systematics, each bin is also individually allowed to fluctuate up
and down based on its independent weighted statistical error.

The up fluctuation histograms for the event weight based systematics for the $t\bar{t}$ + X
background and the Other backgrounds are shown in Fig. 6.22 and Fig. 6.23. The overall
percent deviation of these histograms with the nominal histogram for these backgrounds
is around 5% with a maximal difference around 10%. Given the low event yield of these
nominal histograms, the impact from these systematics are expected to be small.

Figure 6.22: All event weight based shape systematics (up variation) for the year 2016 for
the $t\bar{t}$ + X backgrounds.

The jet energy corrections, jet energy resolution, and the $H_T$ reweighting systematics
are shown in Fig. 6.24 and 6.25 for the $t\bar{t}$ + X and and the Other backgrounds. These
systematics are larger in size, averaging around 10 to 20%, with it reaching up to 40% in
Figure 6.23: All event weight based shape systematics (up variation) for the year 2016 for the Other backgrounds.

the $N_j \geq 12$ bin. These are expected to have more of an impact on the fit, as they allow the shape in NN bin D1 to have more flexibility – something that is more constrained in the $t\bar{t}$ + jets systematics where NN bin D1 has very little shape variations from all the different scale factors.

Figure 6.24: The jet energy resolution/correction and $H_T$ shape systematics for the year 2016 for the $t\bar{t} + X$ backgrounds.

6.3 Signal systematics

The implementation of systematics for the signal samples are analogous to those of the $t\bar{t} + X$ and Other backgrounds. In fact, it is the exact same list of systematics enumerated in Sec. 6.2 (luminosity, lepton ID/isolation/trigger efficiency, $b$ tag efficiency, QCD
Figure 6.25: The jet energy resolution/correction and $H_T$ shape systematics for the year 2016 for the Other backgrounds.

Factorization and renormalization scale, PDF, pileup reweighting, JEC, JER, and overall normalization), where there are input histograms with each systematic variation provided for each signal individually. Each of these histograms are given a nuisance parameter such that the fit can choose to pull the nuisance parameter corresponding to the systematic if the shape improves the overall fit. Of course, the signal systematics are only used as input to the fit when conducting signal + background fits, and only the systematic shapes for the specific signal model being fit are used in any given fit.

Fig. 6.26 and 6.27 show the systematic shape histograms for the RPV SUSY model with $m_{\tilde{t}} = 400$ GeV, while Fig. 6.28 and 6.29 for the SYY SUSY model with $m_{\tilde{t}} = 600$ GeV. Similar to the size of the systematics for the $t\bar{t} + X$ and Other backgrounds, the event weight based systematic fluctuations are sub 5% in size across all bins, whereas the JEC, JER, and $H_T$ correction systematics can be up to 30% for high $N_j$ bins, but average around 10 - 15% in size for the majority of the bins. Given that the total event yield is small in the high mass models, the impact of the signal systematics are expected to be small in the fit for signal models with the $m_{\tilde{t}}$ greater than around 550 GeV. For lower mass models, these systematics may play a bigger role.
Figure 6.26: All event weight based shape systematics (up variation) for the year 2016 for the RPV SUSY signal model with $m_{\tilde{t}} = 400$ GeV.

Figure 6.27: The jet energy resolution/correction and $H_T$ shape systematics for the year 2016 for the RPV SUSY signal model with $m_{\tilde{t}} = 400$ GeV.

Figure 6.28: All event weight based shape systematics (up variation) for the year 2016 for the SYY SUSY signal model with $m_{\tilde{t}} = 600$ GeV.
Figure 6.29: The jet energy resolution/correction and $H_T$ shape systematics for the year 2016 for the SYY SUSY signal model with $m_{\tilde{t}} = 600$ GeV.

### 6.4 Overall systematics summary

So far, the systematics have been discussed as individual entities per year. There is physics motivation, however, that certain nuisance parameters corresponding to the same systematic should be correlated over different years. For example, the parton shower systematics should be correlated in the simulation generated for 2017 and 2018 because they utilize the exact same input parameters and Tune, whereas these systematics should not be correlated with the corresponding systematic for 2016, which utilizes an entirely different Tune. Some of the decisions pertaining to whether or not to correlate systematics from different years were suggested by the respective internal collaboration groups that derive the correction factors used in this analysis. Functionally, to be “correlated” means that the systematics share one nuisance parameter together.

The correlations are as follows:

- **b tag**: the pixel detector was upgraded from 2016 and 2017, and the working points used to determine “medium” b tag is different for different years, so this is treated as **uncorrelated amongst the years**.

- **Lepton ID/isolation/trigger**: in the case of this uncertainty, the largest impact is statistical, and so this systematic is treated as **uncorrelated amongst the years**. Furthermore, the triggers used in 2016 and 2017/2018 were different, so at least these
two eras of data needed to be uncorrelated.

- Parton distribution function (PDF): This is treated as **uncorrelated between 2016 and 2017/2018, but correlated between 2017 and 2018**. The simulation from 2017 and 2018 used the same PDF for generation, whereas 2016 used a different PDF.

- QCD renormalization and factorization scale: This is treated as **fully correlated amongst all the years** because the scale parameters remained unchanged for the simulation generation of the three years.

- Pileup: treated as **fully uncorrelated** as the pileup profile differed amongst the years.

- JEC/JER: official recommendation was for these systematics to be treated as **uncorrelated amongst all three years**.

- ISR/FSR: since these systematics are derived from parameters in PYTHIA, these are **correlated only for 2017 and 2018, and uncorrelated with 2016**.

- Parton shower systematics (color reconnection, matrix element-parton shower matching, and underlying event): same as ISR/FSR - **correlated between 2017 and 2018, but not with 2016** because these systematics are tied to the generator and PYTHIA parameters.

- $H_T$ reweighting: derived independently for all years, so treated as **uncorrelated among all years**.

- $N_j$-$S_{NN}$ residual correlation and jet $p_T$ and jet mass input variable: derived from data in each year separately, so treated as **uncorrelated among all years**.

- Luminosity (signal and non-t$\bar{t}$ backgrounds): treated as **uncorrelated among all three years** under recommendation of the luminosity group.

- Cross section (signal and non-t$\bar{t}$ backgrounds): treated as **fully correlated among years** since all samples were weighted to the theory cross section.
Finally, to give a brief summary of the systematics, Table 6.2 is provided. The numbers listed in the table is the range of values that encompasses 68\% of the errors in all the bins, essentially a 1 sigma band that encompasses the median of the $R_{\text{syst}}$. In parentheses is the maximal value for the size of the systematic in any one bin, normally in the highest $N_j$ bins in either D3 and D4. For maximal values at 100\%, this is most often a result of bins with very low statistics and highly weighted events, especially for the other backgrounds, and so while ostensibly a big value, it amounts to very small fluctuations in the fit. For the $t\bar{t}$ + jets systematics in 2018B that have very large maximal values, this is a result of lower statistics at high $N_j$ on the samples with the special parton shower systematics, and so the fits to these distributions have large uncertainties and impacts in these high $N_j$ bins.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Luminosity</td>
<td>2.5</td>
<td>2.3</td>
<td>2.5</td>
<td>2.3</td>
<td>2.5</td>
<td>2.3</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>0–3 (8)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
</tr>
<tr>
<td>Jet energy resolution</td>
<td>0–2 (14)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
<td>0–2 (10)</td>
</tr>
<tr>
<td>Jet b tagging</td>
<td>0–2 (1)</td>
<td>0–2 (1)</td>
<td>0–2 (1)</td>
<td>0–2 (1)</td>
<td>0–2 (1)</td>
<td>0–2 (1)</td>
<td>0–2 (1)</td>
<td>0–2 (1)</td>
<td>0–2 (1)</td>
<td>0–2 (1)</td>
<td>0–2 (1)</td>
<td>0–2 (1)</td>
<td>0–2 (1)</td>
</tr>
<tr>
<td>Lepton id/iso/trigger</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
</tr>
<tr>
<td>$H_T$ (fit extrapolation)</td>
<td>0–1 (2)</td>
<td>0–1 (3)</td>
<td>0–1 (3)</td>
<td>0–1 (3)</td>
<td>0–1 (3)</td>
<td>0–1 (3)</td>
<td>0–1 (3)</td>
<td>0–1 (3)</td>
<td>0–1 (3)</td>
<td>0–1 (3)</td>
<td>0–1 (3)</td>
<td>0–1 (3)</td>
<td>0–1 (3)</td>
</tr>
<tr>
<td>$H_T$ (N dependence)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
</tr>
<tr>
<td>PDF</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
</tr>
<tr>
<td>Prefiring</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
</tr>
<tr>
<td>PU Reweighting</td>
<td>0–1 (2)</td>
<td>0–1 (2)</td>
<td>0–1 (2)</td>
<td>0–1 (2)</td>
<td>0–1 (2)</td>
<td>0–1 (2)</td>
<td>0–1 (2)</td>
<td>0–1 (2)</td>
<td>0–1 (2)</td>
<td>0–1 (2)</td>
<td>0–1 (2)</td>
<td>0–1 (2)</td>
<td>0–1 (2)</td>
</tr>
<tr>
<td>Fact/Rem. scale</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
</tr>
<tr>
<td>ISR</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
</tr>
<tr>
<td>FSR</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
</tr>
<tr>
<td>Nominal shape difference</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
</tr>
<tr>
<td>CR shape difference</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
</tr>
<tr>
<td>Jet mass-pt</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
</tr>
<tr>
<td>No H_T reweight</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
</tr>
<tr>
<td>ME-PS Matching</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
</tr>
<tr>
<td>Color Reconnection</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
</tr>
<tr>
<td>Underlying Event</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
<td>0–1 (1)</td>
</tr>
</tbody>
</table>

Table 6.2: The overall relative size of all systematics over all years for the t$\bar{t}$ + jets background, the non t$\bar{t}$ + jets background, and a representative signal model ($RPV \tilde{m}_t = 550$ GeV)
Chapter 7

Fit Validation

With all the non-\(t\bar{t} +\) jets backgrounds and the systematics defined, fit robustness tests and validation of the fit function using pseudodata were performed. This section aims to give a comprehensive overview of how the fit is conducted and the results of these above tests to demonstrate the sensitivity of the search. First, the statistical background necessary to understand the fit is described, followed by how the fit is implemented, including coverage on the two different types of fits used in this analysis: background-only and signal + background fits. Then, fits to pseudodata generated from the simulation \(t\bar{t} +\) jets and different signal models are presented, including fit attempts where the \(t\bar{t} +\) jets shape in the four NN bins were purposefully altered. Following all the simulation fit robustness tests, singular fits to just NN bin D1, which is \(t\bar{t} +\) jets background dominated, are presented as a data-driven test for the fit function. Finally, the expected sensitivity and limits are presented in preparation for comparison with the results section.

As a side note, this section will present mostly just the robustness tests for the 2016 dataset and simulation, just to not overwhelm the reader with an excessive number of plots. These robustness tests and validation studies were performed for all years individually and in tandem amongst all four data sets.
7.1 Statistical Background

Given the wealth of shapes in the systematics, obtaining the best $t\bar{t} + \text{jets}$ shape given the data is an extremely daunting task–there are over 300 different nuisance parameters in the final fit. Therefore, a specific statistics package, known as the Higgs Combine tool in the RooStats package \[155\] is used to take the data, conduct the fit, and produce the ultimate $p$ values and significances that will be quoted in the results. This section aims to give an overview of the underlying statistical methods implemented in these packages.

First, the concept of “fit” has to be defined, as there are many ways to conduct a fit, including $\chi^2$-squared and least squares methods. In this analysis, the maximum likelihood method (MLM) is implemented \[6\], which utilizes the concept of using a probability distribution function to determine how likely the data being observed is given a set of parameters. One normally denotes the likelihood in particle physics as $L(data| r \cdot s(\theta) + b(\theta))$, where $r$ is the signal strength, $\theta$ is the set of parameters used to define the probability distribution function, and $s$ and $b$ are the event yields for signal and background, respectively, given those parameters. In the case of this analysis, $\theta$ is the combination of all the fit parameters for the $t\bar{t} + \text{jets}$ parametric shape and all the nuisance parameters associated with each systematic, $r$ is the ratio of the observed signal cross section to the theoretical predicted cross section ($\sigma_{\text{obs}}/\sigma_0$), and the likelihood is calculated as a sum over all the $N_j$ bins in all of the NN bins.

Since the event yield in each $N_j$ bin in each NN bin is a count of the number of events, each bin can be considered a result of a Poisson process \[156\]. In that case, then, the likelihood function of binned data can be written as:

$$L(data| r \cdot s(\theta) + b(\theta)) = \left( \prod_n \frac{(r \cdot s_i + b_i)^{n_i}}{n_i!} e^{-r \cdot s_i - b_i} \right) \times p(\tilde{\theta}|\theta)$$ (7.1)

where the quantity in parentheses is just the Poisson probability to observe $n_i$ events in bin $i$ and $p(\tilde{\theta}|\theta)$ is the auxiliary “measurement” pdf representing the systematic errors.

With this function, the best “fit” is then a matter of varying all of the parameters $\theta$ until this value is maximal. For background only fits, $r$ is fixed at zero, whereas in signal
After obtaining the best fit, the natural question to ask is how well what is observed deviates from known processes \[157\]. For discovery purposes, the null hypothesis is generally taken as the event yield arises solely from all the Standard Model processes (background only), whereas the alternative hypothesis is that the event yield comes from both Standard Model and signal processes (signal+background). The measure of how well the observed matches either the null hypothesis or the alternative hypothesis is generally quantified in a number known as the p value \[158\]. This p value can be calculated in several different ways, but the generally accepted way is to use a test statistic \( q_r \) defined with the profile likelihood ratio, such that:

\[
q_r = -2 \ln \frac{L(data|r, \hat{\theta}_r)}{L(data|\bar{r}, \bar{\theta})}
\]

where the numerator is the likelihood of observing the data given a particular signal strength \( r \) and the optimized parameters for that signal strength \( \hat{\theta}_r \). Meanwhile, the denominator is the maximum likelihood with the global maximum parameters \( \bar{r} \) and \( \bar{\theta} \). For an intuitive feel for how this test statistic is interpreted, a high value of \( q_r \) corresponds to when data is incompatible with the signal strength \( r \) being tested, whereas a value of zero is when the numerator and the denominator have the same likelihood. In this definition of the test statistic, the limits of \( r \) needs to be well defined, particularly on the lower bound \[159\]. It is standard in searches for beyond Standard Model physics that signal strengths be positive both for technical reasons, so that all event yields remain positive, and for interpretation reasons, since a negative signal strength would need special treatment statistically.

Given this definition of the test statistic, for each \( r \) value, a test statistic can be calculated for the observed data \( q_r^{obs} \). This test statistic then needs to be compared to a distribution in the context of either the null hypothesis or the alternative hypothesis, depending on the goal of the analysis. If the question being answered is whether the observed data is indicative of new physics, i.e. a discovery analysis, then the important value is how often the background only processes (Standard Model only processes) give a test statistic at least
as large as the one calculated for the observed data.

In this case, since the focus is on whether the background model can produce any excess that is observed in data, the test statistic sets $r = 0$, such that:

$$q_0 = -2 \ln \frac{L(data|0, \hat{\theta}_0)}{L(data|\hat{r}, \hat{\theta})}$$

(7.3)

This is calculated for the observed data ($q_{0}^{\text{obs}}$), and the optimal parameter set $\hat{\theta}_0$ is set by finding the global maximum given that $r = 0$. Toy studies are then used to generate a set of pseudo data sets using this parameter set, and $q_0^{\text{pseudodata}}$ is calculated for each pseudo data set. Plotting all of these results into a histogram with the test statistic on the x-axis and the number of toys on the y-axis then provides a distribution in which the integral from the observed test statistic up to infinity gives the probability of observing a result at least as large as the test statistic corresponding to the data. An example of this is shown in Fig. 7.1, which shows the distribution of the test statistic $q_0$ for all of the pseudo data toy studies for the Higgs discovery analysis.

![Figure 7.1: Sample distribution for $q_0$ generated with toy studies for the Higgs discovery analysis. This is then used as the probability distribution function ($f(q_0|0, \hat{\theta}_0^{\text{obs}})dq_0$) [156].](image)

Given this probability distribution function from pseudo data studies, the p value can then
be calculated as:

\[ p_0 = P(q_0 \geq q_\text{obs}^0) = \int_{q_\text{obs}^0}^{\infty} f(q_0|0, \hat{\theta}_0^0) dq_0 \]  

(7.4)

This is then converted to a significance value, normally denoted \( Z \), using a “one-sided Gaussian tail,” which has the relationship:

\[ p_0 = \int_{z}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx \]  

(7.5)

\[ Z = \Phi^{-1}(1 - p_0) \]  

(7.6)

In particle physics, the metric is set such that a hint of new physics is observed at \( Z = 3 \), sometimes dubbed \( 3\sigma \) excess, and a discovery occurs at \( Z = 5 \)–the \( 5\sigma \) benchmark. This corresponds to a \( p \) value of \( 2.8 \times 10^{-7} \). In the plots that follow, when the sensitivity of analysis is discussed or when the discovery potential is being quoted, it is this \( p \) value and significance that will be cited.

In the absence of discovering new physics, the data can still be used to put constraints on the signal models that are being explored. In essence, the question that is being addressed in this case is how probable is it that the observed data is the result if there was signal. In looking at whether the observed data is consistent with a new model of physics, i.e. an exclusion analysis, the important value is now how often the signal + background processes predict a test statistic at least as large as the one calculated for the observed data relative.

The procedure to calculate this test statistic is very similar to that of the discovery analysis, except this time, the \( r \neq 0 \). For a given signal strength \( r \), one can define a test statistic such that:

\[ \tilde{q}_r = -2 \ln \frac{L(\text{data}|r, \hat{\theta}_r)}{L(\text{data}|\hat{r}, \hat{\theta})} \]  

(7.7)

and there is a nuisance and fit parameter set \( \hat{\theta}_r^\text{obs} \) that maximizes the likelihood of the observed data \([160]\). Toy studies are then used to generate pseudo data sets using \( \hat{\theta}_r^\text{obs} \), and a distribution of this test statistic is made using the test statistic calculated from all
of these pseudo data sets. The integral from the observed test statistic up to infinity for this distribution gives the probability of observing the data given the hypothesis, which is defined using $p_s$ as:

$$p_s = P(\bar{q}_r \geq q_r^{obs} | \text{signal + background}) = \int_{q_r^{obs}}^{\infty} f(\bar{q}_r | r; \theta_r^{obs}) d\bar{q}_r \quad \text{(7.8)}$$

This value, known as $CL_{s+b}$, is a good measure of how compatible the data is with a signal model when $r$ is high. When $r$ is low and the differences between signal + background and background only predictions are small, however, $CL_{s+b}$ can also be small. In the limit that the signal strength is almost zero and in the presence of downward fluctuations in data, in 5% of all studies, $CL_{s+b}$ will be less than 0.05, and thus produce a false significant result [6]. There are several ways to address this—for example, the Feldman-Cousins approach [161]—but in this analysis, the $CL_{s+b}$ method is modified to the $CL_s$ method, which requires calculating the pdf when $r = 0$, just like in the discovery analysis. Following a similar procedure, we get the definition that the probability of the data coming from the background-only hypothesis $(1 - p_b)$ as:

$$p_b = P(\bar{q}_r < q_r^{obs} | \text{background – only}) \quad \text{(7.9)}$$

$$1 - p_b = P(\bar{q}_r \geq q_r^{obs} | \text{background – only}) = \int_{q_r^{obs}}^{\infty} f(\bar{q}_r | r = 0, \theta_r^{obs}) d\bar{q}_r \quad \text{(7.10)}$$

The definition of $p_b$ as above is purely a result of convention. A representative plot for the distributions of the test statistic when calculated for the Higgs discovery analysis for signal strength $r = 1$, as well as for $r = 0$, is provided in Fig. 7.2.

The final value used in calculation of the observed limit is

$$CL_s(r) = \frac{p_s}{1 - p_b} \quad \text{(7.11)}$$

If $CL_s < 0.05$, then the particular signal model at that signal strength is excluded, meaning that the observed data is incompatible with the signal model with that specific signal
Figure 1: Test statistic distributions for ensembles of pseudo-data generated for signal+background and background-only hypotheses. See the text for definitions of the test statistic and methodology of generating pseudo-data.

\[ p_b = P(\tilde{q}_\mu < \tilde{q}_{\mu_{\text{obs}}}|\text{background-only}) = \int_{\tilde{q}_{\mu_{\text{obs}}}}^{\infty} f(\tilde{q}_\mu|0, \hat{\mu}_{\text{obs}}) d\tilde{q}_\mu \] (7)

and calculate \( CL_s(\mu) \) as the set of probabilities

\[ CL_s(\mu) = p + p_b \] (8)

7. If, for \( \mu = 1 \), \( CL_s \leq \alpha \), we would state that the SM Higgs boson is excluded with \( (1 - \alpha) \) confidence level (C.L.). It is known that the \( CL_s \) method gives conservative limits, i.e. the actual confidence level is higher than \( (1 - \alpha) \). See Appendix A for more details.

8. To quote the 95% Confidence Level upper limit on \( \mu \), to be further denoted as \( \mu_{95\%CL} \), we adjust \( \mu \) until we reach \( CL_s = 0.05 \).

2.2 Expected limits

The most straightforward way for defining the expected median upper-limit and \( \pm 1 \) and \( \pm 2 \) bands for the background-only hypothesis is to generate a large set of background-only pseudo-data sets with the background only hypothesis. Note that we define \( p_b \) as \( p_b = P(\tilde{q}_\mu < \tilde{q}_{\mu_{\text{obs}}}|\text{background-only}) \), excluding the point \( \tilde{q}_{\mu_{\text{obs}}} = \tilde{q}_{\mu_{\text{obs}}} \). With these definitions one can identify \( p_{\mu} \) with \( CL_s + b \) and \( p_b \) with \( 1CL_b \).

Figure 7.2: Sample \( \tilde{q}_{\mu} \) distributions for the Higgs discovery analysis for the background only hypothesis when \( r = 0 \) in blue and the signal + background hypothesis when \( r = 1 \) in red. Here \( \mu \) is the same as \( r \) in the body of the text [156].

strength at the 5% level. Normally, in calculating observed limits and determining whether a signal model is excluded, the signal strength is set to one.

Using just the background only pdf, one can also calculate the expected limits, which is a measure of the strength of the analysis in differentiating between the background only hypothesis and the signal + background hypothesis. From the expected limits, one can normally determine which signal models the analysis is most sensitive to and whether the analysis has the potential for discovering a new signal. The first step in calculating the expected limits is to generate toy pseudo data sets with the background only hypothesis. In each pseudo data, the signal strength \( r \) that corresponds to \( CL_s = 0.05 \) is calculated and plotted in a histogram until a distribution of the \( r^{0.95} \) is created. Once enough values of the signal strength is calculated, the cumulative probability distribution as a function of signal strength \( r^{0.95} \) is created and used to determine the 50% (median signal strength), 16% and 84% (\( \pm 1\sigma \)), and the 2.5% and 97.5% (\( \pm 2\sigma \)) percentiles. These values then define the expected limits, with the median value being the 50th percentile, followed by the \( \pm 1\sigma \) and \( \pm 2\sigma \) bands, respectively.
One subtle point here is that the above procedure requires the generation of many pseudo data toy simulations in order to calculate the signal strength. Conducting many different toy studies to get the pdf of the test statistic, especially if testing at multiple signal strength values, is very computationally intensive. A shorter procedure was developed to calculate the limits derived from the Wald theorem, known as taking the asymptotic limit [157]. Wald proved in his seminal paper [162] that when utilizing a version of the profile likelihood as the test statistic $t_r = -2 \ln(\mathcal{L}(\text{data}|r, \hat{\theta}_r)/\mathcal{L}(\text{data}|\hat{r}, \hat{\theta}))$, the pdf, $f(q_r|\sigma)$, where $r'$ is the signal strength in data, can be derived using a Gaussian distribution with a correction term that scales inversely proportional to the square root of the number of events:

$$\tilde{t}_r = \frac{(r - \hat{r})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N})$$ (7.12)

where $\hat{r}$ follows a Gaussian distribution defined by standard deviation $\sigma$ and a mean $r'$. The $\sigma$ is obtained from the covariance matrix of all the nuisance parameters, the fit parameters, and the signal strength.

If the sample size is large enough ($N \to \infty$), the test statistic can be shown to follow a non-central $\chi$-square distribution with one degree of freedom:

$$f(t_r; \Lambda) = \frac{1}{2\sqrt{t_r} \sqrt{2\pi}} \times \left[ \exp \left( -\frac{1}{2} (\sqrt{t_r} + \sqrt{\Lambda})^2 \right) + \exp \left( -\frac{1}{2} (\sqrt{t_r} - \sqrt{\Lambda})^2 \right) \right]$$ (7.13)

where $\Lambda$ is the noncentrality parameter defined as:

$$\Lambda = \frac{(r - r')^2}{\sigma^2}$$ (7.14)

Thus, all that is needed to calculate the p value and statistical significance in this asymptotic limit is to be able to find $\Lambda$ and $\sigma$. This is done by generating an Asimov dataset, where in every bin, the value is set to the expectation value given the optimal nuisance parameter set [157]. Fig. 7.3 shows an example of an Asimov dataset derived from a dataset that has a Gaussian pdf.

In this idealized dataset, taken directly from the pdf, the standard deviation of the dataset is related to the value of the test statistic in the Asimov dataset as:
where $L_A$ is the likelihood calculated using the Asimov data set. Given this relationship, the pdf of the test statistic can be fully defined using just one simulation instead of an ensemble of toy simulations, and so computation of the observed and expected limits are made much easier. This does require that the data set is large enough that the asymptotic limit can be used, which is the case for this analysis. Extra tests have been conducted to compare the asymptotic limit and limits derived from pseudo data toy studies, and they agree for this analysis.

### 7.2 Fit function technical implementation

The RooStats package, which is built to expand on the RooFit package [163], is responsible for calculating the maximum likelihood and profiling the different parameters to obtain the p values and the significances. Since there are over 300 different uncertainties and nuisance parameters corresponding to all the systematics and the statistical uncertainties in each bin for each background, it was a daunting task, even with a dedicated statistical package. The optimization for the maximum likelihood models in RooFit is done using the MINUIT
package [164], which uses the MIGRAD minimizer to profile all the nuisances provided to it at a reasonable rate.

With respects to the analysis, the fit is conducted over the four NN bins simultaneously, such that each NN bin is required to have the same fit parameters for the fit function \((a_0, a_1, d)\). The normalization for each NN bin is set as a floating parameter using the value of the \(N_j = 7\) bin \((N_{7,D1}, N_{7,D2}, N_{7,D3}, N_{7,D4})\) for a total of seven fit parameters. In combination fits with multiple years, each dataset has its own individual set of fit components, so each dataset (2016, 2017, 2018A, 2018B) has seven fit parameters, with a total of 28 fit parameters total for each full Run 2 combination fit. In addition to these 28 fit parameters, there are nuisance parameters for each systematic per year, except where systematics were correlated. Systematics for \(t\bar{t} + \text{jets}\) are included as shape systematics, whereas systematics for the \(t\bar{t} + X\) and Other backgrounds are included as up and down variations using the raw histograms with the corresponding fluctuation. The only non statistical uncertainty QCD multijet systematic attributed to the transfer factor has its own nuisance parameter, treated as a log normal systematic. The statistical uncertainty for each \(N_j\) bin and each NN bin for the QCD multijet, \(t\bar{t} + X\), and Other backgrounds, as well as for signal where appropriate, are also each given an independent nuisance parameter.

### 7.3 Background only pseudodata fits and validation

The first set of fits conducted were to test the ideal situation where the \(t\bar{t} + \text{jets}\) shape is modeled well by the simulation. In this case, one can imagine that all the correlations between \(N_j\) and \(S_{NN}\) were the same in data and simulation, and so the \(t\bar{t} + \text{jets}\) shape was the same in all NN bins. This ideal setup was created by taking the simulation event yield in each NN bin in each \(N_j\) bin, rounding that value up to the nearest integer bin \((N_{Di,j})\), and then randomly choosing a value in a Poisson distribution with the rounded up estimated value as the most likely parameter. In the limit of high event yield, i.e. in most of the bins, this is equivalent to a random pull from a Gaussian with the mean \(N_{Di,j}\) and the standard deviation as the \(\sqrt{N_{Di,j}}\).
The result of the fit to this pseudo data is presented in Fig. 7.4, which shows the fit result to NN bin 1 to 4, from left to right, respectively. The plots on the bottom of each fit plot are the pulls in the corresponding bin, defined as the difference between the observed (pseudo) data and the fit, divided by the statistical uncertainty in the bin. The fit performs well, and the agreement is visible both in the fit shapes agreeing with the observed pseudo data, and also where the pulls are all sub 1%. What is missing from this plot are the pulls as a result of systematic uncertainties, as the blue bands only show the impact of the statistical uncertainty. Given the fact that the systematics are very shape and yield dependent, it was difficult to show them in one plot, but the pulls on different systematic parameters will be shown in later fits where they are more significant. For now, it is sufficient to say that the pulls on the systematics were very small since the shapes agree very well amongst all the four NN bins. Finally, two signal model shapes, the one for the RPV SUSY model where \( m_{\tilde{t}} = 400 \text{ GeV} \) and the one for the stealth SUSY model where \( m_{\tilde{t}} = 600 \text{ GeV} \) are overlaid at full signal strength for comparison. This side by side comparison shows how the “staircase” scaling for the \( t\bar{t} + \text{jets} \) background is the same amongst all four NN bins, but the signal is divided up differently and have different shapes.

The next validation check is to check whether the fit can produce a false signal in this ideal case; i.e., if a signal + background fit is performed in a pseudo data set where the \( t\bar{t} + \text{jets} \)
+ jets shape is the same, would there be a measured signal strength. To allow for this, a fit was conducted to the same pseudo data generated above, but where the signal strength \( r \) was allowed to fluctuate. This was done for each individual signal model shape that is available, and the resulting fit was compared to the background only fit, with the signal strength also being observed.

![Figure 7.5: Signal + background fit to the pseudo data generated from the t\( \overline{t} \) + jets simulation for 2016. The signal model event yield is scaled to the measured signal strength in this plot.](image)

For the stealth SYY model with \( m_{\tilde{t}} = 600 \text{ GeV} \) case, the most compatible best fit signal strength for the above generated pseudo data was \( 0.02^{+0.49}_{-0.02} \) at 68% CL. The total fit shape is shown in Fig. 7.5, which shows that the pulls once again are very low, and that the fitted signal strength is very low and completely compatible with zero given the \( 1\sigma \) error bands. The large upper error limit is a sign of the number of systematics that is utilized in this analysis and the comparatively low event yield in signal.

As a summary, the fit parameters corresponding to the fits shown in Fig. 7.4 and 7.5 are in Tab. 7.1. The fit parameters from the background only and signal plus background agree within one standard deviation and are overall consistent with each other.
Table 7.1: Fit parameters for the $t\bar{t}$ shape for 2016 simulation.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Background-only fit</th>
<th>Signal+background fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>$0.29832 \pm 0.00769$</td>
<td>$0.29808 \pm 0.00598$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$0.25612 \pm 0.00399$</td>
<td>$0.25562 \pm 0.00603$</td>
</tr>
<tr>
<td>$d$</td>
<td>$10.691 \pm 22.5$</td>
<td>$9.6544 \pm 19.6$</td>
</tr>
</tbody>
</table>

7.4  Signal + background pseudo data fits and validation

The next test on the robustness of the fit function is whether in the presence of signal, the fit function is able to tell the difference between signal and background. To do this, pseudo data was generated in the same way as with the background only fits, but with the signal event yields added to the background event yield prior to pulling a random Poisson value for each $N_j$ bin in each NN bin. The resulting distribution is then fit using only the background histograms with all the systematics, and the result is shown in Fig. 7.6.

Figure 7.6: Background only fit to the pseudo data generated from the $t\bar{t} + \text{jets}$ simulation with the stealth SUSY model with $m_{\tilde{t}} = 600\text{GeV}$ injected at full theoretical cross section for 2016.

This background only fit shows a lot of the features that are unique to the fit procedure that was not observed in the background only pseudo data fits. First, unlike many other analyses, the appearance of signal in this analysis is rarely limited to one search bin or one $N_j$ bin. In fact, the signal often appears as trends and “wiggles” in the pulls in the different NN bins because the signal model shapes are different amongst the four NN bins and can
affect the overall shape such that there are excesses in some $N_j$ bins and deficits in adjacent $N_j$ bins. Therefore, this analysis is not the typical bump hunt search in many other particle physics analyses. The second poignant feature of these fit plots is that from the fit plots alone, all the pulls seem consistent with the statistical errors, and so the analysis seems to not have sensitivity to this signal models. However, upon inspecting the pulls of the nuisance parameters, which are shown in Fig. 7.7 ranked by absolute value, there are pulls of over 1σ for the no-$H_T$ systematic and a pull of approximately 1σ for one of the parton shower systematics, as well as pulls of more than half a sigma for four other systematics. Only pulls greater than 0.20 for all parameters are shown here. These pulls factor into the overall value of the negative log likelihood such that the analysis has a p value of $8.26 \times 10^{-5}$ corresponding to an expected significance of 3.77 for 2016 alone. The table also shows that the shape combination of the no-$H_T$ systematic and the parton shower systematics must be very similar across all four NN bins to that of the signal model since they are pulled so strongly in the signal injection test.

Next, the signal plus background fits for the stealth SUSY model on the signal injected pseudo data are shown in Fig. 7.8 with the corresponding pulls in Fig. 7.9. Given that the signal is injected at full signal strength, the measured signal strength in the signal + background fit of $1.02 \pm 0.26$ corresponds well to the expectation. Furthermore, the “wiggles” observed in the pulls in the background only fit are no longer in the signal + background fit, and the large pulls in the nuisance parameters in the background only fit are also much diminished in the pulls in the nuisance parameters observed in Fig. 7.9.

The overall fit parameters for the background only fit and the signal + background fit are shown in Tab. 7.2. Comparing the fit parameter values of the background only fit to the fit parameter values with pseudo data with no signal injection, the fit parameter $a_1$ differs by at least one standard deviation. This is characteristic of this analysis, where the fit function has some flexibility to absorb differences from the simulation $t\bar{t} +$ jets shape, as long as the discrepancy is the same amongst all four NN bins. When providing the right signal model shape though, the fit parameters return to within 1σ of the simulation $t\bar{t} +$
Figure 7.7: Table of pulls for the background only fit to the pseudo data generated from the $t\bar{t} + \text{jets}$ simulation with the stealth SUSY model with $m_{\tilde{t}} = 600$ GeV injected at full theoretical cross section for 2016, where the value for any non-fit parameters are in standard deviations. The uncertainty presented here is the larger value if the errors are asymmetrical.

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Unc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N7_tt_D1_2016</td>
<td>32474.0</td>
<td>868.0</td>
</tr>
<tr>
<td>N7_tt_D2_2016</td>
<td>21458.0</td>
<td>572.0</td>
</tr>
<tr>
<td>N7_tt_D3_2016</td>
<td>3098.5</td>
<td>119.0</td>
</tr>
<tr>
<td>N7_tt_D4_2016</td>
<td>1563.4</td>
<td>74.0</td>
</tr>
<tr>
<td>d_tt_2016</td>
<td>117.03</td>
<td>174.0</td>
</tr>
<tr>
<td>np_tt_noHT_2016</td>
<td>1.6354</td>
<td>0.72</td>
</tr>
<tr>
<td>np_tt_hdampUp_2016</td>
<td>1.0255</td>
<td>0.773</td>
</tr>
<tr>
<td>np_tt_htjet_2016</td>
<td>0.92059</td>
<td>0.93</td>
</tr>
<tr>
<td>np_tt_qcdCR_2016</td>
<td>0.82459</td>
<td>0.682</td>
</tr>
<tr>
<td>np_tt_underlyingEvtDown_2016</td>
<td>0.6602</td>
<td>0.563</td>
</tr>
<tr>
<td>np_tt_JECDown_2016</td>
<td>0.52045</td>
<td>0.807</td>
</tr>
<tr>
<td>np_tt_erdoN_2016</td>
<td>0.4417</td>
<td>0.814</td>
</tr>
<tr>
<td>np_tt_underlyingEvtUp_2016</td>
<td>0.41806</td>
<td>0.518</td>
</tr>
<tr>
<td>np_tt_nom_2016</td>
<td>0.37729</td>
<td>0.949</td>
</tr>
<tr>
<td>np_tt_hdampDown_2016</td>
<td>0.31283</td>
<td>0.813</td>
</tr>
<tr>
<td>a0_tt_2016</td>
<td>0.30647</td>
<td>0.00609</td>
</tr>
<tr>
<td>np_tt_JERUp_2016</td>
<td>0.29795</td>
<td>0.933</td>
</tr>
<tr>
<td>np_tt_lltail_2016</td>
<td>0.28897</td>
<td>0.972</td>
</tr>
<tr>
<td>a1_tt_2016</td>
<td>0.27668</td>
<td>0.00337</td>
</tr>
<tr>
<td>np_tt_qcdCRD4Coef1_2016</td>
<td>-0.27406</td>
<td>0.976</td>
</tr>
<tr>
<td>np_tt_qcdCRD3Coef3_2016</td>
<td>0.2377</td>
<td>0.97</td>
</tr>
<tr>
<td>np_tt_qcdCRD2Coef3_2016</td>
<td>0.21252</td>
<td>1.0</td>
</tr>
<tr>
<td>np_tt_qcdCRD2Coef1_2016</td>
<td>0.21061</td>
<td>0.93</td>
</tr>
</tbody>
</table>
Figure 7.8: Signal + background fit to the pseudo data generated from the t\(t\bar{t}\) + jets simulation with the stealth SUSY model with \(m_{\tilde{t}} = 600\) GeV injected at full theoretical cross section for 2016.

Figure 7.9: Table of pulls for the signal + background fit to the pseudo data generated from the t\(t\bar{t}\) + jets simulation with the stealth SUSY model with \(m_{\tilde{t}} = 600\) GeV injected at full theoretical cross section for 2016, where the value for any non-fit parameters are in standard deviations. The uncertainty presented here is the larger value if the errors are asymmetrical.
jets shape value.

Table 7.2: Fit parameters for the $t\bar{t}$ shape for 2016 simulation with the stealth SUSY SYY model with $m_{\tilde{t}} = 600$ GeV injected at full cross section.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Background only fit</th>
<th>Signal+background fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>$0.30647 \pm 0.00609$</td>
<td>$0.29802 \pm 0.00620$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$0.27668 \pm 0.00337$</td>
<td>$0.25555 \pm 0.00661$</td>
</tr>
<tr>
<td>$d$</td>
<td>$117.03 \pm 174.0$</td>
<td>$9.6119 \pm 21.3$</td>
</tr>
</tbody>
</table>

The conclusions presented above for the stealth SUSY SYY model with $m_{\tilde{t}} = 600$ GeV is the same for the signal injection tests with other signal models. Namely, when conducting a background only fit, the pulls in both the nuisance parameters and the individual $N_j$ bins are large, whereas when conducting a signal + background fit, the pulls become much smaller across the board.

Extra studies using both the 2016 and 2017 datasets have also been conducted where one signal model is injected, while the fit is done with a different signal model. In these studies, the mass resolution of this analysis is around $\pm 200$ GeV. This means that when generating pseudo data where one signal model is injected, i.e. RPV SUSY with $m_{\tilde{t}} = 550$ GeV, the search is still sensitive when using histogram shapes from signal models of different masses, in this case, between $m_{\tilde{t}} = 350$ GeV to $m_{\tilde{t}} = 750$ GeV. This is because the signal model shapes are only weakly dependent on the $m_{\tilde{t}}$, with the major difference between different mass models being the cross section and not the $N_j$ shapes. This is exemplified in the p value plot shown in Fig. 7.10, which shows the fit results with different signal model shapes with the stealth SUSY models with pseudo data generated with signal injection from the RPV SUSY $m_{\tilde{t}} = 550$ GeV model. This plot shows clearly the broad mass “hump” as opposed to the narrow mass resonance “bump” in most other searches.

Finally, other robustness studies with the fit setup using different fluctuations on the $t\bar{t}$ + jets shapes were conducted. One of the robustness tests was purposefully taking the $t\bar{t}$ + jets shape with the $1\sigma$ fluctuation of jet energy corrections, the jet energy resolution, and the final state radiation scale factors applied, and conducting the fit to see if a signal can
Figure 7.10: Local p value measured for pseudo data with the RPV SUSY model with $m_{\tilde{t}} = 550$ GeV injected with fits conducted using the stealth SUSY signal shapes. The term "observed" is used to express that the pseudo data is generated and input to the fit in the same way as real data would be.

be faked with systematic fluctuations. In a second test, a linear function of $N_j$ that differs in each NN bin was used to attempt to fake a signal. For example, in NN bin 1, the $N_j = 7$ bin would be changed by 0%, the $N_j = 8$ bin would be changed by 1%, and so on, until the $N_j \geq 12$ bin would be changed by 5%. In NN bin 4, the $N_j = 7$ bin would be changed by 0%, the $N_j = 8$ bin would be changed by 4%, and the $N_j \geq 12$ bin would be changed by 20%. The results of the first test is shown for the RPV SUSY model in the left plot and the results of the second test is shown for the stealth SUSY model in the right plot in Fig. 7.11. The left plot shows that the systematic shapes cover for any discrepancies in the $t\bar{t} + \text{jets}$ shape corresponding to any of the known disparities already taken into account by the systematics. The right plot shows that in order to get an approximately $2\sigma$ significant result, a discrepancy averaging 10% in the $t\bar{t} + \text{jets}$ shape is needed. For studies where all the $N_j$ bins in the NN bins were fluctuated by the same amount, the fit was able to absorb all those discrepancies.
Figure 7.11: Local p value measured for pseudo data with the 1σ JEC, JER, and FSR scale factors applied to the t\(\bar{t}\) + jets simulation with RPV SUSY signal models (left) and for pseudo data with an artificial signal injected using the stealth SUSY signal shapes (right). The term “observed” is used to express that the pseudo data is generated and input to the fit in the same way as real data would be.

7.5 Fits to data

One of the last fit robustness checks conducted was to do fits in the statistics dominated D1 bin for all four years. While this bin is not necessarily signal depleted, especially in high \(N_j\) bins, this is as close to a fit in a background dominated region as possible in data. Since the fit is only in one NN bin, all systematics were not utilized in these fits. The goal of these fits were just to get a general shape for the \(N_j\) distribution of the t\(\bar{t}\) + jets shape in data. The fit to the t\(\bar{t}\) + jets shape D1 bin individually for the 2016, 2017, 2018A, and 2018B datasets are shown in Fig. 7.12 with the fit parameters in Tab. 7.3. All the pulls are within systematic uncertainties, and the fit parameters for the four years agree within one standard deviation.
Table 7.3: Fit parameters for the $t\bar{t} + \text{jets}$ shape for the D1 bin in 2016, 2017, 2018A, and 2018B.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>2016 D1</th>
<th>2017 D1</th>
<th>2018A D1</th>
<th>2018B D1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>$0.29672 \pm 0.0149$</td>
<td>$0.30266 \pm 0.0143$</td>
<td>$0.29429 \pm 0.00868$</td>
<td>$0.28558 \pm 0.00769$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$0.25490 \pm 0.00701$</td>
<td>$0.26285 \pm 0.00916$</td>
<td>$0.26762 \pm 0.0117$</td>
<td>$0.25325 \pm 0.0103$</td>
</tr>
<tr>
<td>$d$</td>
<td>$83.420 \pm 163$</td>
<td>$249.32 \pm 338$</td>
<td>$111.85 \pm 301$</td>
<td>$66.978 \pm 227$</td>
</tr>
</tbody>
</table>

7.6 Expected signal sensitivity and limits

Having validated the fit procedure and its robustness, the next step is to calculate the expected sensitivity of the analysis to the different signal and mass models. This is done by calculating the significance and $p$ values using the pseudo data with the signal injected at full theoretical cross section. To get an accurate representation of the sensitivity, the full combination fit using the 2016, 2017, 2018A, and 2018B simulation samples were used. The results are shown in Tab. 7.4 and 7.5, with the same information conveyed in local $p$ value plots shown in Fig. 7.13.

From Tab. 7.4 and 7.5, the analysis has discovery potential ($\geq 5\sigma$) for RPV SUSY models less than or equal to 550 GeV and stealth SUSY SYY models up to around 700 GeV. The measured signal strength is close to one for all models where there is discovery capability, but the measured signal strength drops off at higher $m_{\tilde{t}}$ models because the cross section of those signal models are too low compared to the size of the systematics and background.
These models, particularly RPV SUSY models with $m_{\tilde{t}}$ greater than 700 GeV and stealth SUSY SYY models with $m_{\tilde{t}}$ greater than 950 GeV, have corresponding low signal sensitivity, as demonstrated by the low significance, large p values, and large error bars.

| Mass | Best fit signal strength $|\Delta|_{1}$ | sign. | p value |
|------|---------------------------------|-------|---------|
| 300  | $1.00_{-0.09}^{+0.09}$          | 8.02  | 5.24309e-16 |
| 350  | $1.00_{-0.09}^{+0.09}$          | 9.29  | 7.99866e-21 |
| 400  | $0.99_{-0.09}^{+0.09}$          | 9.58  | 5.0471e-22  |
| 450  | $0.99_{-0.10}^{+0.10}$          | 8.46  | 1.36475e-17 |
| 500  | $0.99_{-0.12}^{+0.12}$          | 7.07  | 7.91402e-13 |
| 550  | $0.98_{-0.16}^{+0.16}$          | 5.69  | 6.2112e-09  |
| 600  | $0.95_{-0.21}^{+0.20}$          | 4.36  | 6.43069e-06 |
| 650  | $0.96_{-0.27}^{+0.27}$          | 3.41  | 0.000321254 |
| 700  | $0.91_{-0.37}^{+0.37}$          | 2.42  | 0.00779883  |
| 750  | $0.91_{-0.52}^{+0.51}$          | 1.74  | 0.0406741   |
| 800  | $0.78_{-0.71}^{+0.71}$          | 1.10  | 0.136174    |
| 850  | $0.51_{-0.51}^{+1.09}$          | 0.46  | 0.321594    |
| 900  | $0.39_{-0.39}^{+1.57}$          | 0.25  | 0.403023    |
| 950  | $0.47_{-0.47}^{+2.35}$          | 0.20  | 0.420686    |
| 1000 | $0.61_{-0.61}^{+3.45}$          | 0.18  | 0.430414    |
| 1050 | $0.26_{-0.26}^{+5.32}$          | 0.04  | 0.48561     |
| 1100 | $0.00_{-0.00}^{+6.41}$          | 0.00  | 0.5         |
| 1150 | $0.00_{-0.00}^{+10.00}$         | 0.00  | 0.5         |
| 1200 | $0.00_{-0.00}^{+10.00}$         | 0.00  | 0.5         |

Table 7.4: Best fit signal strength and expected significances for the RPV SUSY model as a function of $m_{\tilde{t}}$, for the combination of 2016, 2017, 2018A, and 2018B.

The breakdown of the signal sensitivity across the many years are shown by the relative contributions to the local p value plots in Fig. 7.13. The contribution to the significance per data set is determined primarily by the total integrated luminosity contributed by the data set. The 2017 data set has the total integrated luminosity followed by 2018B, and their relative contributions to the significance attest to this. The p value plots also have the predicted signal strength in the bottom of the plot, with the $1\sigma$ bands shown in green.
<table>
<thead>
<tr>
<th>Mass</th>
<th>Best fit signal strength</th>
<th>significance</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYY</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>$0.98^{+0.16}_{-0.15}$</td>
<td>5.37</td>
<td>3.95509e-08</td>
</tr>
<tr>
<td>350</td>
<td>$1.00^{+0.10}_{-0.10}$</td>
<td>8.13</td>
<td>2.11756e-16</td>
</tr>
<tr>
<td>400</td>
<td>$1.00^{+0.09}_{-0.09}$</td>
<td>9.87</td>
<td>2.81321e-23</td>
</tr>
<tr>
<td>450</td>
<td>$0.99^{+0.09}_{-0.08}$</td>
<td>10.37</td>
<td>1.78286e-25</td>
</tr>
<tr>
<td>500</td>
<td>$0.99^{+0.09}_{-0.09}$</td>
<td>9.89</td>
<td>2.38637e-23</td>
</tr>
<tr>
<td>550</td>
<td>$0.99^{+0.10}_{-0.10}$</td>
<td>9.14</td>
<td>3.26659e-20</td>
</tr>
<tr>
<td>600</td>
<td>$0.99^{+0.11}_{-0.11}$</td>
<td>7.39</td>
<td>7.12237e-14</td>
</tr>
<tr>
<td>650</td>
<td>$0.98^{+0.13}_{-0.13}$</td>
<td>6.91</td>
<td>2.40758e-12</td>
</tr>
<tr>
<td>700</td>
<td>$0.97^{+0.15}_{-0.15}$</td>
<td>5.71</td>
<td>5.79617e-09</td>
</tr>
<tr>
<td>750</td>
<td>$0.95^{+0.18}_{-0.18}$</td>
<td>4.91</td>
<td>4.63562e-07</td>
</tr>
<tr>
<td>800</td>
<td>$0.96^{+0.23}_{-0.22}$</td>
<td>3.99</td>
<td>3.33043e-05</td>
</tr>
<tr>
<td>850</td>
<td>$0.96^{+0.29}_{-0.28}$</td>
<td>3.29</td>
<td>0.000493924</td>
</tr>
<tr>
<td>900</td>
<td>$0.90^{+0.37}_{-0.36}$</td>
<td>2.48</td>
<td>0.00651571</td>
</tr>
<tr>
<td>950</td>
<td>$0.92^{+0.46}_{-0.45}$</td>
<td>2.05</td>
<td>0.0200775</td>
</tr>
<tr>
<td>1000</td>
<td>$0.98^{+0.61}_{-0.58}$</td>
<td>1.68</td>
<td>0.0465547</td>
</tr>
<tr>
<td>1050</td>
<td>$0.60^{+0.78}_{-0.60}$</td>
<td>0.80</td>
<td>0.210621</td>
</tr>
<tr>
<td>1100</td>
<td>$0.66^{+1.03}_{-0.66}$</td>
<td>0.67</td>
<td>0.250526</td>
</tr>
<tr>
<td>1150</td>
<td>$0.65^{+1.33}_{-0.65}$</td>
<td>0.51</td>
<td>0.306545</td>
</tr>
<tr>
<td>1200</td>
<td>$0.24^{+1.85}_{-0.24}$</td>
<td>0.13</td>
<td>0.446532</td>
</tr>
</tbody>
</table>

Table 7.5: Best fit signal strength and expected significances for the stealth SUSY SYY model as a function of $m_{\tilde{t}}$, for the combination of 2016, 2017, 2018A, and 2018B.

The large uncertainties at high $m_{\tilde{t}}$ is more poignantly shown here by the blowing up of the green bands as $m_{\tilde{t}}$ is increased.

Another way to show the sensitivity of the analysis is to look at the expected limits, which are derived using the background only hypothesis from the simulation. Here, the expected limits are also only shown for the combination of the four data sets in Fig. 7.14 for the RPV SUSY model on the left and for the stealth SUSY SYY model on the right. The $1\sigma$ and the $2\sigma$ bands are shown in green and yellow, respectively, with the expected limit shown in black. Here, the expected limit is generated from pseudo data using the background only hypothesis.
Figure 7.13: The local p value for different $m_{\tilde{t}}$ values for the RPV SUSY model (left) and the stealth SUSY SYY model (right) for pseudo data generated with the simulation in 2016, 2017, 2018A, and 2018B with the signal injected at full theoretical cross section.

The expected limit plots are different from the expected sensitivity shown previously because the expected sensitivity looks at the signal + background hypothesis as the null hypothesis and answers the question: “Can the analysis pick out excesses from a signal model relative to the background model,” whereas the expected limits here uses the background only hypothesis as the null hypothesis and answers the question: “If no signal is observed, what can the data tell us about a particular signal model?” The expected limit is often compared to the theoretical cross section, drawn in red, and where the expected limit is below the theoretical cross section are signal models that can be expected to be excluded at a predetermined confidence level - here at 95%. These limits only speak to full theoretical cross section or signal strength $r = 1$. For signal models at higher $m_{\tilde{t}}$ where the expected limit is above the theoretical cross section line, no statistically significant conclusions can be drawn about these models. From the expected limit plots shown in Fig. 7.14, exclusion for RPV SUSY models are expected to approach between 750 and 800 GeV and the exclusion for the stealth SUSY SYY models are expected to approach around 1 TeV.
Figure 7.14: Expected upper limit on the cross section as a function of $m_{\tilde{t}}$ for the RPV SUSY model and the stealth SUSY SYY model on the left and right, respectively. The green and yellow bands show the 1 and 2σ uncertainty bands. The red line shows the theoretical top squark cross section as a function of its mass.
Chapter 8

Results

For the final chapter, the final background only and signal + background fit results will be shown for a subset of the signal models. While fits were conducted for all mass models and all signal models, only fit results for one signal model with low $m_{\tilde{t}}$ and one signal model with intermediate to high $m_{\tilde{t}}$ will be provided here for space considerations. The representative low $m_{\tilde{t}}$ model will be the RPV SUSY model with $m_{\tilde{t}} = 400$ GeV and the representative intermediate $m_{\tilde{t}}$ model will be the stealth SUSY SYY model with $m_{\tilde{t}} = 600$ GeV. Next, the observed significances will be provided, as well as the table of pulls, followed by discussion of the results. Finally, the limits that this analysis sets for the RPV SUSY and stealth SUSY models will be shown.

8.1 Background Only Fits

The first set of results are the background only fits, shown in Fig. 8.1, where the fit is not allowed to pull any signal model shape. From left to right are the respective plots for each of the NN bins in increasing order, with the sub plots on the bottom showing the pulls in each of the respective $N_j$ bin. As a reminder, the pulls shown in the sub plots are calculated as the difference between the number of observed events in the data and the fit value divided by the statistical uncertainty of the data, and so do not include any of the systematic uncertainties. While the signal is not pulled in these plots, the signal $N_j$
distributions in the four NN bins are overlaid on top of the fit in order to show relative size and yields that could be in each of the NN bins. These signal distributions are normalized to the predicted theoretical cross sections for that model at 100% branching fraction.

Analyzing all 96 bins, the pulls deviating from zero indicate that the data observed do not seem to agree fully with the fit function shape, and so it indicates that the data does not have the same $t\bar{t} +$ jets shapes in the four NN bins. To narrow down where the discrepancies are the largest, the general agreement of the fitted shape and the observed data can be broken down by year. Since NN bin 1 has the most events by construction, this bin has the largest impact on the fit parameters for the overall $t\bar{t} +$ jets shape, and so the general agreement of the fit shape and the observed data can be inferred from the general agreement in those bins. The pulls range from -2 to 2, and based on the agreement in NN bin 1, the largest disagreements in the fit shape and the observed data appear in the 2017 and 2018A dataset.

The fit parameters for this background only fit are shown in Tab. 8.1, with the corresponding pulls on the systematics for the background only fit shown in Fig. 8.2. Comparing the values of $a_0$, $a_1$, and $d$ for the four years, the parameters are within 1$\sigma$ of each other for 2016 and 2018B and for 2017 and 2018A, with the agreement between these two groups beyond 1$\sigma$ of each other. This may address directly that the conditions in the data collected in 2017 and 2018A are most similar, since 2018B has the extra veto for the HCAL applied and the data set in 2016 was collected prior to the upgrade to the pixel sub detector.

Comparing the fit parameters to the background only fits in the simulation, the values of $a_0$ and $a_1$ are larger than those of simulation. This discrepancy has no impact on the analysis, as the $t\bar{t} +$ jets shape is taken solely from data in the analysis, but it is interesting to note that this indicates that the $N_j$ scaling is less steep in the data than the simulation predicts given that there are no other effects. It is also interesting to note that the background only fits for simulation with signal injected also had larger fit parameter values.

Next, it is important to look at the pulls, ranked in order of magnitude of the pull, in
cross section for the RPV SUSY model with 

Figure 8.1: Fitted background prediction and observed data counts for 2016, 2017, 2018A and 2018B in each of the four NN bins. The signal distributions normalized to the theoretical cross section for the RPV SUSY model with \( m_{\tilde{t}} = 400 \text{GeV} \) and the stealth SUSY model with \( m_{\tilde{t}} = 600 \text{GeV} \) are overlaid.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Background-only fit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2016</td>
</tr>
<tr>
<td>$a_0$</td>
<td>$0.28938 \pm 0.00578$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$0.26195 \pm 0.00401$</td>
</tr>
<tr>
<td>$d$</td>
<td>$26.995 \pm 58.2$</td>
</tr>
<tr>
<td></td>
<td>2017</td>
</tr>
<tr>
<td>$a_0$</td>
<td>$0.30145 \pm 0.00368$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$0.26524 \pm 0.00416$</td>
</tr>
<tr>
<td>$d$</td>
<td>$118.10 \pm 163$</td>
</tr>
<tr>
<td></td>
<td>2018pre</td>
</tr>
<tr>
<td>$a_0$</td>
<td>$0.30338 \pm 0.00418$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$0.26581 \pm 0.00560$</td>
</tr>
<tr>
<td>$d$</td>
<td>$285.31 \pm 297$</td>
</tr>
<tr>
<td></td>
<td>2018post</td>
</tr>
<tr>
<td>$a_0$</td>
<td>$0.28405 \pm 0.00364$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$0.25502 \pm 0.00412$</td>
</tr>
<tr>
<td>$d$</td>
<td>$15.726 \pm 45.1$</td>
</tr>
</tbody>
</table>

Table 8.1: Fit parameters for the $t\bar{t} + \text{jets}$ shape in the fit to the data, for the combined background only fit for all four time eras.

Fig. 8.2. This ranked list shows that there were various systematics with a large impact, particularly the $H_T$ based systematics (no-$H_T$ and $H_T$ $N_j$ dependent systematic) and the bin migration systematics (ISR/FSR/JEC/JER). From a physics perspective, the $H_T$ based systematics having a large impact makes sense since the $H_T$ modeling is highly correlated to the $N_j$ modeling in the simulation, and the poor $N_j$ modeling was a primary motivation to take the $t\bar{t} + \text{jets}$ shape in this data driven approach. Since the $H_T$ reweighting was derived rather ad-hoc, it is possible that there is some unknown $H_T$ based effect that is not being taken into account. Furthermore, from the signal injection tests, the no-$H_T$ systematic was pulled a significant amount when there was signal injection. This pull here indicates that at the very least, the $N_j$ distribution in data might have more events at high $H_T$.

With respects to the bin migration systematics being pulled a large amount, this can be understood by pointing to the overall poor $N_j$ agreement between data and simulation.
As the category name suggests, these systematics are derived allowing for an event to move from one $N_j$ bin to another, which can affect the overall $N_j$ shape much more than just the event weight based systematics. Therefore, these systematics are preferentially pulled if the $N_j$ shape in data differ from simulation significantly. Overall, the strong pulls on these systematics just point to the need for better modeling of these bin migration effects, like initial and final state radiation in events with high jet multiplicity, which is expected.

The background only fits and pulls point to some discrepancies between the observed data and the expectation that the $t\bar{t} + \text{jets}$ shape is the same in all four $N_n$ bins, some of which are accounted for through the systematics and some of which are still unaccounted for, as observed in the pulls in the data. Thus, the next logical step is to check whether these discrepancies amongst the $N_n$ bins can be accounted for by conducting signal + background fits.

### 8.2 Signal + background fits

The fit parameters for the signal + background fit with the stealth SUSY SYY model with $m_{\tilde{t}} = 600$ GeV and the RPV SUSY model with $m_{\tilde{t}} = 400$ GeV are shown in Tab. 8.2, with the corresponding fit shapes in Fig. 8.4 and Fig. 8.5 respectively. The pulls to the signal + background fit with the stealth SUSY SYY model with $m_{\tilde{t}} = 600$ GeV is shown in Fig. 8.6, with the corresponding pulls for the RPV SUSY model with $m_{\tilde{t}} = 400$ GeV in Fig. 8.7. The fitted signal strength for the stealth SUSY SYY model with $m_{\tilde{t}} = 600$ GeV is $0.09^{+0.10}_{-0.09}$, corresponding to a local p value of 0.09, or a significance of 1.32. The fitted signal strength for the RPV SUSY model with $m_{\tilde{t}} = 400$ GeV is $0.21^{+0.07}_{-0.07}$, which corresponds to a local p value of $2.7 \times 10^{-3}$, or a significance of 2.78. The largest significance of any mass and signal model fitted in this analysis is the RPV SUSY model with $m_{\tilde{t}} = 400$ GeV at 2.78.

The measured signal strengths for the rest of the signal models with different masses is summarized in Tab. 8.3 and 8.4 and graphically in the local p value plots in Fig. 8.8.

Comparing the fit parameters of the stealth SUSY SYY model to the background only fit, the fit parameters are within $1\sigma$ of each other, and so this indicates that the discrepancy...
<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Unc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>np_tt_noHT_2016</td>
<td>0.90474</td>
<td>0.742</td>
</tr>
<tr>
<td>np_tt_htnjet_2018pre</td>
<td>0.75163</td>
<td>0.861</td>
</tr>
<tr>
<td>np_tt_JERDown_2018post</td>
<td>0.71032</td>
<td>0.814</td>
</tr>
<tr>
<td>np_tt_noHT_2018pre</td>
<td>0.70062</td>
<td>0.834</td>
</tr>
<tr>
<td>np_tt_ISR_2016</td>
<td>-0.68426</td>
<td>0.634</td>
</tr>
<tr>
<td>np_tt_JERDown_2016</td>
<td>0.68271</td>
<td>0.925</td>
</tr>
<tr>
<td>np_tt_underlyingEvtUp_2016</td>
<td>-0.67372</td>
<td>0.574</td>
</tr>
<tr>
<td>np_tt_nom_2017</td>
<td>0.66672</td>
<td>0.782</td>
</tr>
<tr>
<td>np_tt_erdOn_2016</td>
<td>0.61088</td>
<td>0.866</td>
</tr>
<tr>
<td>np_tt_JECDown_2016</td>
<td>0.53954</td>
<td>0.861</td>
</tr>
<tr>
<td>np_tt_qcdCR_2017</td>
<td>-0.53626</td>
<td>0.763</td>
</tr>
<tr>
<td>np_tt_mpTScaled_2018pre</td>
<td>0.52275</td>
<td>0.838</td>
</tr>
<tr>
<td>np_tt_noHT_2017</td>
<td>0.52116</td>
<td>0.867</td>
</tr>
<tr>
<td>np_tt_hdampDown_2017_2018</td>
<td>0.49743</td>
<td>0.408</td>
</tr>
<tr>
<td>np_tt_JECUp_2016</td>
<td>0.49634</td>
<td>0.844</td>
</tr>
<tr>
<td>np_tt_JECUp_2018pre</td>
<td>0.48847</td>
<td>0.923</td>
</tr>
<tr>
<td>np_tt_mpTScaled_2017</td>
<td>-0.45488</td>
<td>0.944</td>
</tr>
<tr>
<td>np_tt_mpTScaled_2017</td>
<td>0.43248</td>
<td>0.709</td>
</tr>
<tr>
<td>np_tt_JECDown_2018post</td>
<td>0.42709</td>
<td>0.928</td>
</tr>
<tr>
<td>np_tt_qcdCR_2018pre</td>
<td>-0.41633</td>
<td>0.604</td>
</tr>
<tr>
<td>np_tt_JECUp_2016</td>
<td>-0.40436</td>
<td>0.797</td>
</tr>
<tr>
<td>np_tt_JECUp_2018pre</td>
<td>0.34235</td>
<td>0.937</td>
</tr>
<tr>
<td>np_tt_JECDown_2018post</td>
<td>-0.33626</td>
<td>0.964</td>
</tr>
<tr>
<td>np_tt_ISR_2017_2018</td>
<td>0.31517</td>
<td>0.878</td>
</tr>
<tr>
<td>np_tt_nom_2018pre</td>
<td>-0.2951</td>
<td>0.955</td>
</tr>
<tr>
<td>np_tt_underlyingEvtUp_2017_2018</td>
<td>0.27101</td>
<td>0.33</td>
</tr>
<tr>
<td>np_tt_hdampUp_2017_2018</td>
<td>-0.26847</td>
<td>0.327</td>
</tr>
<tr>
<td>OTHER_bkg_xsec</td>
<td>0.2622</td>
<td>0.958</td>
</tr>
<tr>
<td>np_tt_JECUp_2017</td>
<td>-0.25211</td>
<td>0.929</td>
</tr>
<tr>
<td>JER_2017</td>
<td>-0.2292</td>
<td>1.24</td>
</tr>
<tr>
<td>np_tt_noHT_2018post</td>
<td>-0.22138</td>
<td>0.936</td>
</tr>
</tbody>
</table>

Figure 8.2: List of nuisance parameters that contribute to the likelihood fit in the background only fit. Only nuisance parameters that have a pull with an absolute value greater than 0.2 are listed here.
in the $N_j$ shapes between the four NN bins observed in data do not look like the $N_j$ shapes that correspond to this signal model. The lists of systematics pulled are also very similar to the background only model, except for the highest ranked pull for the nominal $t\bar{t} +$ jets systematic for 2017, which is shown in Fig. 8.3. The systematic has minimal impact on the first four $N_j$ bins, but a rather large impact on the $N_j \geq 12$ inclusive bin, where the uncertainty is already quite large; therefore, the large pull on this systematic is not surprising given the excess of events with high $N_j$ observed in data.

Figure 8.3: The “nominal” shape systematic for the year 2017 derived as a ratio of the fitted $N_j$ distribution in each NN bin to the fitted total $N_j$ distribution in simulation.

The story is slightly different in the fit to the RPV SUSY with the $m_{\tilde{t}} = 400$ GeV. Immediately, the fit parameters are different from the stealth SUSY SYY model with $m_{\tilde{t}} = 600$ GeV and the background only fit, being smaller for all four years and closer to the parameters derived in the simulation background only fit. Analyzing the statistical pulls in the fit plots, the pulls for the RPV SUSY fit here are smaller across the board than either the stealth SUSY SYY signal + background fit and the background only fit. This indicates that the signal model shapes associated with the RPV SUSY model with $m_{\tilde{t}} = 400$ GeV matches much better the discrepancies in the $t\bar{t} +$ jets shapes in data amongst the four NN bins. For the RPV SUSY signal + background fit in Fig. 8.7, there are fewer nuisance parameter pulled $0.2\sigma$ or more, and each nuisance parameter is pulled on average less. For example, some of the more sharply pulled systematics in the background only fit, like the $H_T$ systematics, are pulled in this fit about $0.5\sigma$ less. All of these observations summed
together indicate that the fit with RPV SUSY with $m_{\tilde{t}} = 400$ GeV corresponds to data better, and agree well with the overall significance of 2.78.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Signal (SYY 600) + background fit</th>
<th>Signal (RPV 400) + background fit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2016</td>
<td>2016</td>
</tr>
<tr>
<td>$a_0$</td>
<td>$0.28800 \pm 0.00590$</td>
<td>$0.28165 \pm 0.00631$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$0.25849 \pm 0.00494$</td>
<td>$0.24584 \pm 0.00727$</td>
</tr>
<tr>
<td>$d$</td>
<td>$13.951 \pm 38.4$</td>
<td>$16.077 \pm 34.0$</td>
</tr>
<tr>
<td></td>
<td>2017</td>
<td>2017</td>
</tr>
<tr>
<td>$a_0$</td>
<td>$0.30034 \pm 0.00381$</td>
<td>$0.29467 \pm 0.00432$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$0.26301 \pm 0.00449$</td>
<td>$0.2529 \pm 0.00611$</td>
</tr>
<tr>
<td>$d$</td>
<td>$78.357 \pm 112$</td>
<td>$46.83 \pm 65.0$</td>
</tr>
<tr>
<td></td>
<td>2018pre</td>
<td>2018pre</td>
</tr>
<tr>
<td>$a_0$</td>
<td>$0.30293 \pm 0.00426$</td>
<td>$0.29824 \pm 0.00450$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$0.26420 \pm 0.00584$</td>
<td>$0.25525 \pm 0.00698$</td>
</tr>
<tr>
<td>$d$</td>
<td>$259.64 \pm 292$</td>
<td>$256.1 \pm 289$</td>
</tr>
<tr>
<td></td>
<td>2018post</td>
<td>2018post</td>
</tr>
<tr>
<td>$a_0$</td>
<td>$0.28353 \pm 0.00364$</td>
<td>$0.27986 \pm 0.00385$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$0.25272 \pm 0.00412$</td>
<td>$0.24307 \pm 0.00589$</td>
</tr>
<tr>
<td>$d$</td>
<td>$9.2769 \pm 33.8$</td>
<td>$20.797 \pm 42.0$</td>
</tr>
</tbody>
</table>

Table 8.2: Fit parameters for the $t\bar{t} +$ jets shape for the combined signal + background fits to the data sets in 2016, 2017, 2018A, and 2018B, using the signal model of stealth SUSY SYY with $m_{\tilde{t}} = 600$ GeV and RPV SUSY with $m_{\tilde{t}} = 400$ GeV.
<table>
<thead>
<tr>
<th>Mass</th>
<th>Best fit signal strength</th>
<th>significance</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>$0.20^{+0.07}_{-0.10}$</td>
<td>2.18</td>
<td>0.0144863</td>
</tr>
<tr>
<td>350</td>
<td>$0.18^{+0.06}_{-0.07}$</td>
<td>2.39</td>
<td>0.00848925</td>
</tr>
<tr>
<td>400</td>
<td>$0.21^{+0.07}_{-0.07}$</td>
<td>2.78</td>
<td>0.00272507</td>
</tr>
<tr>
<td>450</td>
<td>$0.24^{+0.09}_{-0.09}$</td>
<td>2.41</td>
<td>0.00795862</td>
</tr>
<tr>
<td>500</td>
<td>$0.27^{+0.12}_{-0.12}$</td>
<td>2.11</td>
<td>0.0172651</td>
</tr>
<tr>
<td>550</td>
<td>$0.28^{+0.15}_{-0.16}$</td>
<td>1.68</td>
<td>0.0469655</td>
</tr>
<tr>
<td>600</td>
<td>$0.37^{+0.21}_{-0.22}$</td>
<td>1.61</td>
<td>0.0534263</td>
</tr>
<tr>
<td>650</td>
<td>$0.37^{+0.29}_{-0.30}$</td>
<td>1.21</td>
<td>0.113719</td>
</tr>
<tr>
<td>700</td>
<td>$0.46^{+0.43}_{-0.45}$</td>
<td>1.02</td>
<td>0.154306</td>
</tr>
<tr>
<td>750</td>
<td>$0.86^{+0.59}_{-0.62}$</td>
<td>1.28</td>
<td>0.100945</td>
</tr>
<tr>
<td>800</td>
<td>$0.97^{+0.85}_{-0.88}$</td>
<td>1.10</td>
<td>0.134677</td>
</tr>
<tr>
<td>850</td>
<td>$1.46^{+1.37}_{-1.42}$</td>
<td>1.02</td>
<td>0.153237</td>
</tr>
<tr>
<td>900</td>
<td>$1.90^{+2.00}_{-1.90}$</td>
<td>0.91</td>
<td>0.182308</td>
</tr>
<tr>
<td>950</td>
<td>$2.78^{+3.02}_{-2.78}$</td>
<td>0.90</td>
<td>0.184776</td>
</tr>
<tr>
<td>1000</td>
<td>$4.75^{+4.38}_{-4.47}$</td>
<td>1.06</td>
<td>0.145448</td>
</tr>
<tr>
<td>1050</td>
<td>$4.22^{+5.78}_{-4.22}$</td>
<td>0.60</td>
<td>0.275781</td>
</tr>
<tr>
<td>1100</td>
<td>$7.32^{+2.68}_{-7.32}$</td>
<td>0.71</td>
<td>0.237533</td>
</tr>
<tr>
<td>1150</td>
<td>$10.00^{+3.90}_{-10.00}$</td>
<td>0.93</td>
<td>0.177182</td>
</tr>
<tr>
<td>1200</td>
<td>$10.00^{+3.90}_{-10.00}$</td>
<td>0.77</td>
<td>0.21994</td>
</tr>
</tbody>
</table>

Table 8.3: Best fit signal strength and observed significances for the RPV SUSY model as a function of top squark mass, for the combination of 2016, 2017, 2018A, and 2018B.
<table>
<thead>
<tr>
<th>Mass</th>
<th>Best fit signal strength</th>
<th>significance</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SYY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>$0.32^{+0.12}_{-0.15}$</td>
<td>1.77</td>
<td>0.0382575</td>
</tr>
<tr>
<td>350</td>
<td>$0.21^{+0.07}_{-0.08}$</td>
<td>2.48</td>
<td>0.0065171</td>
</tr>
<tr>
<td>400</td>
<td>$0.14^{+0.06}_{-0.07}$</td>
<td>1.85</td>
<td>0.0319404</td>
</tr>
<tr>
<td>450</td>
<td>$0.14^{+0.06}_{-0.07}$</td>
<td>2.04</td>
<td>0.0207175</td>
</tr>
<tr>
<td>500</td>
<td>$0.19^{+0.07}_{-0.07}$</td>
<td>1.75</td>
<td>0.0399094</td>
</tr>
<tr>
<td>550</td>
<td>$0.15^{+0.08}_{-0.09}$</td>
<td>1.49</td>
<td>0.0684942</td>
</tr>
<tr>
<td>600</td>
<td>$0.13^{+0.09}_{-0.10}$</td>
<td>1.32</td>
<td>0.0932231</td>
</tr>
<tr>
<td>650</td>
<td>$0.12^{+0.12}_{-0.13}$</td>
<td>1.03</td>
<td>0.150597</td>
</tr>
<tr>
<td>700</td>
<td>$0.10^{+0.16}_{-0.16}$</td>
<td>0.97</td>
<td>0.164849</td>
</tr>
<tr>
<td>750</td>
<td>$0.16^{+0.19}_{-0.16}$</td>
<td>0.83</td>
<td>0.203799</td>
</tr>
<tr>
<td>800</td>
<td>$0.18^{+0.26}_{-0.18}$</td>
<td>0.67</td>
<td>0.252377</td>
</tr>
<tr>
<td>850</td>
<td>$0.19^{+0.34}_{-0.19}$</td>
<td>0.55</td>
<td>0.290372</td>
</tr>
<tr>
<td>900</td>
<td>$0.27^{+0.45}_{-0.27}$</td>
<td>0.59</td>
<td>0.278353</td>
</tr>
<tr>
<td>950</td>
<td>$0.39^{+0.58}_{-0.33}$</td>
<td>0.56</td>
<td>0.2894</td>
</tr>
<tr>
<td>1000</td>
<td>$0.32^{+0.79}_{-0.32}$</td>
<td>0.40</td>
<td>0.345508</td>
</tr>
<tr>
<td>1050</td>
<td>$0.39^{+1.04}_{-0.39}$</td>
<td>0.38</td>
<td>0.353795</td>
</tr>
<tr>
<td>1100</td>
<td>$0.72^{+1.44}_{-0.72}$</td>
<td>0.49</td>
<td>0.311235</td>
</tr>
<tr>
<td>1150</td>
<td>$0.54^{+1.89}_{-0.54}$</td>
<td>0.28</td>
<td>0.38883</td>
</tr>
<tr>
<td>1200</td>
<td>$0.91^{+2.65}_{-0.91}$</td>
<td>0.34</td>
<td>0.367769</td>
</tr>
</tbody>
</table>

Table 8.4: Best fit signal strength and observed significances for the stealth SUSY SYY model as a function of top squark mass, for the combination of 2016, 2017, 2018A, and 2018B.
Figure 8.4: Fitted background prediction and observed data counts for 2016, 2017, 2018A and 2018B in each of the four NN bins. In this particular fit, the $N_j$ shapes for the stealth SUSY model with $m_\ell = 600$ GeV were allowed to be pulled, with its theoretically predicted $N_j$ distribution at full cross section shown in pink.
Figure 8.5: Fitted background prediction and observed data counts for 2016, 2017, 2018A and 2018B in each of the four NN bins. In this particular fit, the $N_j$ shapes RPV SUSY model with $m_f = 400$ GeV were allowed to be pulled, with its theoretically predicted $N_j$ distribution at full cross section shown in pink.
<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Unc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>np_tt_nom_2017</td>
<td>0.74146</td>
<td>0.796</td>
</tr>
<tr>
<td>np_tt_JERDown_2016</td>
<td>0.71338</td>
<td>0.928</td>
</tr>
<tr>
<td>np_tt_ISR_2016</td>
<td>-0.65016</td>
<td>0.642</td>
</tr>
<tr>
<td>np_tt_noHT_2016</td>
<td>0.61929</td>
<td>0.778</td>
</tr>
<tr>
<td>np_tt_JERDown_2018post</td>
<td>0.6176</td>
<td>0.823</td>
</tr>
<tr>
<td>np_tt_erOn_2016</td>
<td>0.60094</td>
<td>0.871</td>
</tr>
<tr>
<td>np_tt_noHT_2018pre</td>
<td>0.59088</td>
<td>0.845</td>
</tr>
<tr>
<td>np_tt_hnjet_2018pre</td>
<td>0.58198</td>
<td>0.882</td>
</tr>
<tr>
<td>np_tt_hdampDown_2016</td>
<td>-0.57865</td>
<td>0.852</td>
</tr>
<tr>
<td>np_tt_underlyingEvtUp_2016</td>
<td>-0.56789</td>
<td>0.594</td>
</tr>
<tr>
<td>np_tt_hdampDown_2017_2018</td>
<td>0.54349</td>
<td>0.424</td>
</tr>
<tr>
<td>np_tt_qcdCR_2016</td>
<td>-0.53971</td>
<td>0.782</td>
</tr>
<tr>
<td>np_tt_mpTScaled_2018pre</td>
<td>0.5062</td>
<td>0.843</td>
</tr>
<tr>
<td>np_tt_qcdCR_2018pre</td>
<td>-0.47072</td>
<td>0.607</td>
</tr>
<tr>
<td>np_tt_JECUp_2016</td>
<td>0.44955</td>
<td>0.928</td>
</tr>
<tr>
<td>np_tt_JECDown_2016</td>
<td>0.427</td>
<td>0.877</td>
</tr>
<tr>
<td>np_tt_qcdCR_2017</td>
<td>-0.4039</td>
<td>0.875</td>
</tr>
<tr>
<td>np_tt_noHT_2017</td>
<td>0.37543</td>
<td>0.877</td>
</tr>
<tr>
<td>np_tt_hdampUp_2016</td>
<td>-0.37004</td>
<td>0.833</td>
</tr>
<tr>
<td>np_tt_underlyingEvtUp_2017_2018</td>
<td>0.36024</td>
<td>0.358</td>
</tr>
<tr>
<td>np_tt_FSR_2017_2018</td>
<td>0.35798</td>
<td>0.593</td>
</tr>
<tr>
<td>np_tt_JECUp_2018pre</td>
<td>-0.32754</td>
<td>0.955</td>
</tr>
<tr>
<td>np_tt_hdampUp_2017_2018</td>
<td>-0.31617</td>
<td>0.346</td>
</tr>
<tr>
<td>OTHER_bkg_xsec</td>
<td>0.29552</td>
<td>1.05</td>
</tr>
<tr>
<td>np_tt_nom_2018post</td>
<td>0.29352</td>
<td>0.886</td>
</tr>
<tr>
<td>np_tt_noHT_2018post</td>
<td>-0.28712</td>
<td>0.939</td>
</tr>
<tr>
<td>np_tt_hnjet_2017</td>
<td>0.26176</td>
<td>0.949</td>
</tr>
<tr>
<td>JEC_2016</td>
<td>-0.26141</td>
<td>0.914</td>
</tr>
<tr>
<td>np_tt_nom_2018pre</td>
<td>-0.25013</td>
<td>0.87</td>
</tr>
<tr>
<td>JER_2017</td>
<td>-0.22591</td>
<td>1.25</td>
</tr>
<tr>
<td>np_tt_ISR_2017_2018</td>
<td>-0.22119</td>
<td>0.962</td>
</tr>
<tr>
<td>np_tt_JECUp_2017</td>
<td>-0.21762</td>
<td>0.934</td>
</tr>
</tbody>
</table>

Figure 8.6: List of nuisance parameters that contribute to the likelihood fit in the signal + background fit using the stealth SUSY SYY signal model with $m_{\tilde{t}} = 600$ GeV. Only nuisance parameters that have a pull with an absolute value greater than 0.2 are listed here.
<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Unc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>np_tt_JERDown_2016</td>
<td>0.64362</td>
<td>0.935</td>
</tr>
<tr>
<td>np_tt_hdampUp_2016</td>
<td>-0.62663</td>
<td>0.851</td>
</tr>
<tr>
<td>np_tt_hdampDown_2016</td>
<td>-0.59652</td>
<td>0.871</td>
</tr>
<tr>
<td>np_tt_ISR_2016</td>
<td>-0.59236</td>
<td>0.673</td>
</tr>
<tr>
<td>np_tt_nom_2017</td>
<td>0.55282</td>
<td>0.836</td>
</tr>
<tr>
<td>JEC_2018post</td>
<td>-0.49075</td>
<td>0.972</td>
</tr>
<tr>
<td>np_tt_JECUp_2016</td>
<td>0.48177</td>
<td>0.937</td>
</tr>
<tr>
<td>np_tt_hdampUp_2017_2018</td>
<td>-0.48019</td>
<td>0.394</td>
</tr>
<tr>
<td>np_tt_nom_2018post</td>
<td>0.47531</td>
<td>0.893</td>
</tr>
<tr>
<td>JEC_2016</td>
<td>-0.41086</td>
<td>1.06</td>
</tr>
<tr>
<td>np_tt_hdampDown_2017_2018</td>
<td>0.40336</td>
<td>0.465</td>
</tr>
<tr>
<td>np_tt_noHT_2018post</td>
<td>-0.39741</td>
<td>0.945</td>
</tr>
<tr>
<td>JER_2017</td>
<td>-0.38994</td>
<td>1.09</td>
</tr>
<tr>
<td>np_tt_mpTScaled_2017</td>
<td>-0.36957</td>
<td>0.782</td>
</tr>
<tr>
<td>np_tt_qcdCR_2016</td>
<td>-0.35122</td>
<td>0.8</td>
</tr>
<tr>
<td>OTHER_bkg_xsec</td>
<td>0.28086</td>
<td>1.02</td>
</tr>
<tr>
<td>np_tt_JECDown_2016</td>
<td>0.2673</td>
<td>0.902</td>
</tr>
<tr>
<td>np_tt_JECDown_2017</td>
<td>-0.22719</td>
<td>0.949</td>
</tr>
<tr>
<td>np_tt_mpTScaled_2018post</td>
<td>-0.21521</td>
<td>0.776</td>
</tr>
<tr>
<td>np_tt_htnjet_2018post</td>
<td>-0.21168</td>
<td>0.98</td>
</tr>
<tr>
<td>np_tt_qcdCR_2017</td>
<td>-0.21147</td>
<td>0.868</td>
</tr>
</tbody>
</table>

Figure 8.7: List of nuisance parameters that contribute to the likelihood fit in the background only fit using the RPV SUSY signal model with $m_{\tilde{t}} = 400$ GeV. Only nuisance parameters that have a pull with an absolute value greater than 0.2 are listed here.

Figure 8.8: The local p-value for different $m_{\tilde{t}}$ values for the RPV SUSY model (left) and the stealth SUSY SYY model (right) for the observed data in 2016, 2017, 2018A, and 2018B.
The local p value plots shown in Fig. 8.8 show the relative contribution to the significance of the different years. There are a few salient features of the local p value plot that differ significantly from the expected p value plots calculated from the pseudo data. First, the data set that contributes the most to the p value and significance is 2018A and not 2017 or 2018B, as it was in the expected p value plot. This indicates that the discrepancy in the data set in 2018A looks more signal-like than the discrepancy in the other three years.

To see this more clearly, the fit to data was done to each data set individually and the signal strength parameter was profiled, as shown in Fig. 8.9 for the RPV SUSY model with $m_{\tilde{t}} = 350$ GeV. The measured signal strength is around 0.18 for the combined fit to this signal model, which corresponds well to the value of the fitted signal strength to the dataset in 2016 alone. Meanwhile, the data sets from 2017 and 2018B prefer a signal model strength on the lower end, between 0.1 to 0.15, and the data set from 2018A prefers a signal strength of 0.26. From these fitted signal strengths, it is clear that the 2018A discrepancy is more signal-like than either of the discrepancies in the other data sets. However, since the data set for 2018A is the smallest of the four data sets, the overall measured signal strength is pulled lower, which exacerbates the discrepancies observed in 2018A, which manifests itself in the statistical pulls in the fit plots.

The other major feature of the local p value plots is that the significance of signal injection is relatively constant as a function of $m_{\tilde{t}}$. This is different from the expected local p value plots that show the significance peak at intermediate $m_{\tilde{t}}$ around 400–600 GeV, and then diminish to zero at high $m_{\tilde{t}}$. This feature of the fit can be described by the combination of two effects: 1) the signal model shapes are weakly dependent on $m_{\tilde{t}}$ and 2) the fit is unable to account for discrepancies at high $N_j$ and need to pull the signal to partially cover for this discrepancy. The former effect was discussed before, but deserves to be emphasized again: the signal model shapes are similar across the different NN bins since the training is done with all signal model samples simultaneously, and so the mass resolution of this analysis is not very sharp. The analysis is better at pinpointing whether there is a discrepancy in the $N_j$ shapes amongst the many bins than distinguishing between a range of signal models of
Figure 8.9: The negative log likelihood scan as a function of the signal strength $r$ shown for the 2016 data (top left), the 2017 data (top right), 2018S data (bottom left), 2018B data (bottom right). All other nuisance parameters are
different masses, as long as these signal models predict an excess of events at high $N_j$. The second issue is most clearly observed in the statistical pulls in the fit plots: the analysis cannot account for all the discrepancies in the $N_j$ shape from the fit function in the four NN bins from the systematics alone, even when the shapes provided by the signal models are included. This points to either a potentially missing systematic, though painstaking care was done to include all of the well motivated systematics; a genuine mis-modeling in the simulation that is not known; and/or that a different signal model can account for these discrepancies better.

### 8.3 Observed Limits and Exclusion

Given that the largest observed significance is below $3\sigma$, there is not strong evidence of new physics in this analysis, and so exclusion limits are set for the two signal models in Fig. 8.10. For the RPV SUSY models, any models that predict the $m_{\tilde{t}} \leq 675$ GeV are excluded at full cross section, which corresponds to a cross section times branching fraction between $0.2 - 3$ pb. For the stealth SUSY SYY models, any models that predict the $m_{\tilde{t}} \leq 975$ GeV are excluded at full cross section, which corresponds to a cross section times branching fraction between $0.02 - 5$ pb. In these limit plots, the observed deviates from the expected (in the dotted line) since the data does show some discrepancy from the background only model. These deviations are higher for lower masses, where the signal looks more like the discrepancy, with the observed deviating from the expected larger than $2\sigma$. While these significances are not the same as those calculated for the local $p$ values in the previous section, the deviations from the expected here are a result of the same effects.
Figure 8.10: Observed upper limit on the cross section as a function of $m_{\tilde{t}}$ for the RPV SUSY model and the stealth SUSY SYY model on the top and bottom, respectively. The green and yellow bands show the 1 and 2σ uncertainty bands. The red line shows the theoretical top squark cross section as a function of its mass.
Chapter 9

Conclusions

An analysis was conducted on 137.2 fb$^{-1}$ of data collected from the CMS detector at the LHC during Run II from 2016 to 2018, searching for the pair production of scalar top quarks in RPV SUSY and stealth SUSY models decaying into many light-flavored jets and a lepton with low missing transverse energy. This analysis was novel because of its implementation of new machine learning techniques, such as gradient reversal, and for searching in a region of phase space often vetoed by other similar SUSY searches. With regards to the particular RPV SUSY model, where the $\tilde{t}$ quark decays into a top quark and a neutralino, and the neutralino decaying via the $\lambda''_{112}$ coupling, a limit was set on the $m_{\tilde{t}}$ up to 700 GeV with a 0.2–3 pb cross section, depending on $m_{\tilde{t}}$, at the 95% confidence level. For the stealth SUSY SYY model, where the $\tilde{t}$ quark decays into a top quark, a gluon, and a singlino, the singlino decays into a singlet and a gravatino, and the singlet decays into two gluons, a limit was set on the $m_{\tilde{t}}$ up to 900 GeV with a cross section between 0.02–5 pb, depending on the $m_{\tilde{t}}$, at the 95% confidence level. While no significant deviation from the Standard Model was observed, the analysis is systematics limited and can be improved with a more robust background estimation method, as well as further improvements in top quark reconstruction using more modern machine learning techniques. With new additional techniques, it is possible that the 2 standard deviation excesses observed in this analysis could be proven to be new physics or just statistical fluctuations.
References


[86] CMS Collaboration, “Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of tt at sqrt(s)=8 and 13 TeV,” CMS


