ABSTRACT

First Search for Pair Production of Scalar Top Quarks Decaying to Top Quarks and Light-Flavor Jets with Low Missing Transverse Momentum

Christopher Michael Madrid, Ph.D.

Co-Advisor: Jay R. Dittmann, Ph.D.
Co-Advisor: Kenichi Hatakeyama, Ph.D.

After the discovery of the Higgs boson in 2012, the current best theoretical model that describes all observed particles and their interactions, the standard model (SM), was considered complete. However, there are a plethora of physical phenomena that the SM does not accurately describe, which motivates particle physicists to search for evidence of new particles produced by high energy collisions at the Large Hadron Collider (LHC). In traditional searches for physics beyond the standard model, a requirement of high missing transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$) is often used to identify potential new particles. However, without any signs of significant deviations from standard model expectations, a search where this requirement is removed has been performed. There are many well motivated and previously unexplored physics models, including versions of supersymmetry (SUSY) characterized by R-parity violation or with additional hidden sectors, that predict the production of events with low $p_{\mathrm{T}}^{\mathrm{miss}}$, many jets, and top quarks. In particular, a general search is performed to look for the pair production of scalar top quarks that would decay to two top quarks and six additional light flavor jets. The search is performed using events with at least seven jets and exactly one electron or muon. No requirement on $p_{\mathrm{T}}^{\mathrm{miss}}$ is imposed. With the use of a neural-network-based signal-to-background discriminator, a background estimation has been achieved where more traditional techniques would not be possible. The study is based on a sample of proton-proton collisions at $\sqrt{s} = 13$ TeV corresponding to 137.2 fb^{-1} of integrated luminosity collected with the Compact Muon Solenoid (CMS) detector at the LHC in 2016, 2017, and 2018. Results of the search are interpreted for stealth SUSY and SUSY with R-parity violation, resulting in a lower limit exclusion of scalar top production of 900 and 700 GeV, respectively.

First Search for Pair Production of Scalar Top Quarks Decaying to Top Quarks and Light-Flavor Jets with Low Missing Transverse Momentum

by

Christopher Michael Madrid, B.S., M.A.

A Dissertation

Approved by the Department of Physics

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

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

Approved by the Dissertation Committee

Jay R. Dittmann, Ph.D., Co-Chairperson

Kenichi Hatakeyama, Ph.D., Co-Chairperson

Lorin S. Matthews, Ph.D.

Andrew Brinkerhoff, Ph.D.

David Ryden, Ph.D.

Accepted by the Graduate School
August 2020

J. Larry Lyon, Ph.D., Dean

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

TABLE OF CONTENTS

LIST OF ACRONYMS

ALICE	A Large Ion Collider Experiment
APD	Avalanche Photodiode
ATLAS	A Toroidal LHC ApparatuS
BDT	Boosted Decision Tree
BPIX	Barrel Pixel
BSB	Baylor Sciences Building
BSM	Beyond the Standard Model
CERN	European Organization for Nuclear Research
CL	Confidence Level
CMS	Compact Muon Solenoid
CMSSW	CMS Software Framework
COM	Center-of-mass
CR	Control Region
CR	Color Reconnection
CSC	Cathode Strip Chamber
CTF	Combinatorial Track Finder
DM	Dark Matter
DNN	Dense Neural Network
DT	Drift Tube
ECAL	Electromagnetic Calorimeter

FNAL	Fermi National Accelerator Laboratory
FPGA	Field-Programmable Gate Array
FPIX	Forward Pixel
FPR	False Positive Rate
FSR	Final-State Radiation
GR	Gradient Reversal
GR	General Relativity
GSF	Gaussian Sum Filter
GUT	Grand Unified Theory
HB	Hadron Calorimeter Barrel
HCAL	Hadron Calorimeter
HE	Hadron Calorimeter Endcap
HEM	Hadron Calorimeter Endcap Minus
HF	Hadron Calorimeter Forward
HO	Hadron Calorimeter Outer
HLT	High Level Trigger
IP	Interaction Point
ISR	Initial State Radiation
L1T	Level–1 Trigger
LEP	Large Electron-Positron Collider
LHCb	LHC–beauty
LHC	Large Hadron Collider
LINAC	Linear Accelerator

LPC	LHC Physics Center
LSP	Lightest Supersymmetric Particle
MSSM	Minimal Supersymmetric Standard Model
NLL	Next-to-Leading Logarithmic
NLO	Next-to-Leading Order
PDF	Probability Distribution Function
PF	Particle Flow
POI	Parameter Of Interest
PS	Proton Synchrotron
PS	Parton Shower
PU	Pileup
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics
QIE	Charge Integrator and Encoder
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
RPC	Resistive Plate Chamber
RPV	R-Parity Violating
SiPM	Silicon Photomultiplier
SM	Standard Model
SPS	Super Proton Synchrotron
SST	Silicon Strip Tracker
SUSY	Supersymmetry

SYY	Stealth SUSY with SYY coupling
TEC	Tracker Endcap
TIB	Tracker Inner Barrel
TID	Tracker Inner Disks
TOB	Tracker Outer Barrel
TPR	True Positive Rate
UE	Underlying Event
VEV	Vacuum Expectation Value

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Like many people before me, I am fascinated by the big questions about our universe. I have had the pleasure and honor of joining the massive community of scientists all pursuing answers to those big questions. Over the course of my life I have interacted with countless individuals that have helped me reach this point. While it would be impossible to name them all, I would like to specifically acknowledge the ones below.

First, I would like to express my great appreciation to my advisors, Professors Jay Dittmann and Kenichi Hatakeyama, for their time and guidance. Without them I would still be stuck in the BSB scratching my head. I would also like to thank my pseudo third advisor Dr. James Hirschauer, for inviting me to join the LPC stealth stop analysis team, and for his prompt and helpful advice. All three have shown me that it is possible to have an infinite to-do list while still managing to stay sane with grace, and most importantly how to think like a physicist.

Second, I want to thank postdocs, Joe Pastika and Nadja Strobbe. We were stuck in the day-to-day trenches of my ignorance together, and I could never thank you both enough for "holding my hand" through all of my studies.

Third, I would like to thank Drs. Owen Long, Aron Soha, Scarlet Norberg, Kevin Pedro, Alexx Perloff, Messrs. Caleb Smith, Bryan Caraway, and everyone else that has crossed paths with me. Everyone listed above has helped forge me into the physicist that I am today, and I am lucky to have met all of you.

Finally, I express my greatest appreciation for the grant funding from the Department of Energy (DE-SC0007861), and to Fermilab's LPC for the graduate scholar position that both financially made this possible.

To my loving wife and parents,
without them this would not have been possible.

CHAPTER ONE

Introduction

> "We are at the very beginning of time for the human race. It is not unreasonable that we grapple with problems. But there are tens of thousands of years in the future. Our responsibility is to do what we can, learn what we can, improve the solutions, and pass them on."
>
> —Richard P. Feynman (1918–1988)

From the dawn of civilization, progress has always been correlated with an increased understanding of the world around us. With this in mind, it is vital that we as a society push the boundaries of understanding. The study of physics has a rich history, and although it would take far more than this thesis to properly describe it, we can highlight specific key developments.

It might seem obvious now, but using mathematics as a means to describe physical processes was a revolutionary concept. Who did this first is a subject of debate, but it is undeniable that the physical laws written down by Sir Isaac Newton in the 17$^{\text{th}}$ century using mathematics were a huge step toward the goal of a complete understanding of the universe. Newton's success with his models of motion and gravity effectively changed the way we approach physics. During the 19$^{\text{th}}$ century, hundreds of brilliant minds later, there were two more key developments: Dmitri Mendeleev's periodic table of elements and James Maxwell's equations of classical electromagnetism. The periodic table of elements established a set of building blocks for all matter, while the theories of Newton and Maxwell together provided a way to describe all known forces (at that time) and the motion of objects in the universe. It was thought at the time that putting these monumental models together, along with a few others, would conclude all of the fundamental work of physicists, leaving only more precise measurements of fundamental constants. Luckily, that was not the case, because the universe is much more interesting than these well-defined but "boring" models make it out to be.

One interesting consequence of Maxwell's equations was that they predicted a value of the speed of light in a vacuum. However, Maxwell's equations are not invariant under Galilean transformations, which implies that they are only valid for one inertial reference frame. This goes against the assumption that the laws of electricity and magnetism must be the same regardless of the frame of reference in which they are observed. This led Albert Einstein, in 1905, to solve this problem with his groundbreaking theory of special relativity. About a decade later he generalized this to include gravity with his theory of general relativity. This changed our concept of space by promoting it from a 3-dimensional space to a 4-dimensional spacetime. If you are keeping score at this point, we now have two models (Newton's and Maxwell's) that describe how objects in the universe interact, another model (Einstein's relativity) that generalize these to include an understanding of spacetime itself, and Mendeleev's periodic table of elements, which describes all matter in the universe. However, at the time it was questioned whether or not these models would hold for very small objects, or in other words, would these models hold for all length scales? It could be said that almost all breakthroughs in physics during the last century were a consequence of attempting to answer that question.

As it turns out, the laws of physics for extremely small objects differ greatly from the theories previously formulated. In fact, one of the first pieces of evidence of something interesting was J.J. Thompson's discovery of the first elementary particle in 1897, the electron. This forced physicists to consider the possibility that the elements in the periodic table were not fundamental, but were composed of something more elementary. The next piece of the puzzle was the discovery of the nucleus by Ernest Rutherford in 1911, which eventually led to the idea that all matter in the universe is composed of protons, neutrons, and electrons. Also, as the tried and true theories were inadequate to describe physics at the size scale of the electron, the formulation of quantum mechanics by Erwin Schrödinger and others during the 1920's was necessary. During the 1920's through 1950's, new theories of nature were becoming less intuitive but more accurate as we probed smaller and smaller objects. This was the first time in human history that probing physics many orders of magnitude smaller than a drop of water was possible. Then, in the late 1950's and 1960's something extraordinary began to happen. There was an explosion of newly observed particles, which was made possible by the construction of particle accelerators. There was then a race to reformulate our theories of the universe to include these new particles. Eventually, these developments led to the discovery of many fundamental particles and it became clear that the proton and neutron were composed of some of them. We ended up naming these particles quarks and leptons, and theories describing their nature began to emerge. At first, there were two separate theories for leptons and quarks formulated by Sheldon Glashow in 1961 and Murray Gell-Mann et al. in 1964, respectively. One can see the size of matter that was beginning to be probed in Fig. 1.1, and as scientists pushed to smaller and smaller scales, a reformulation of physical laws and the way we think about the composition of matter was necessary.

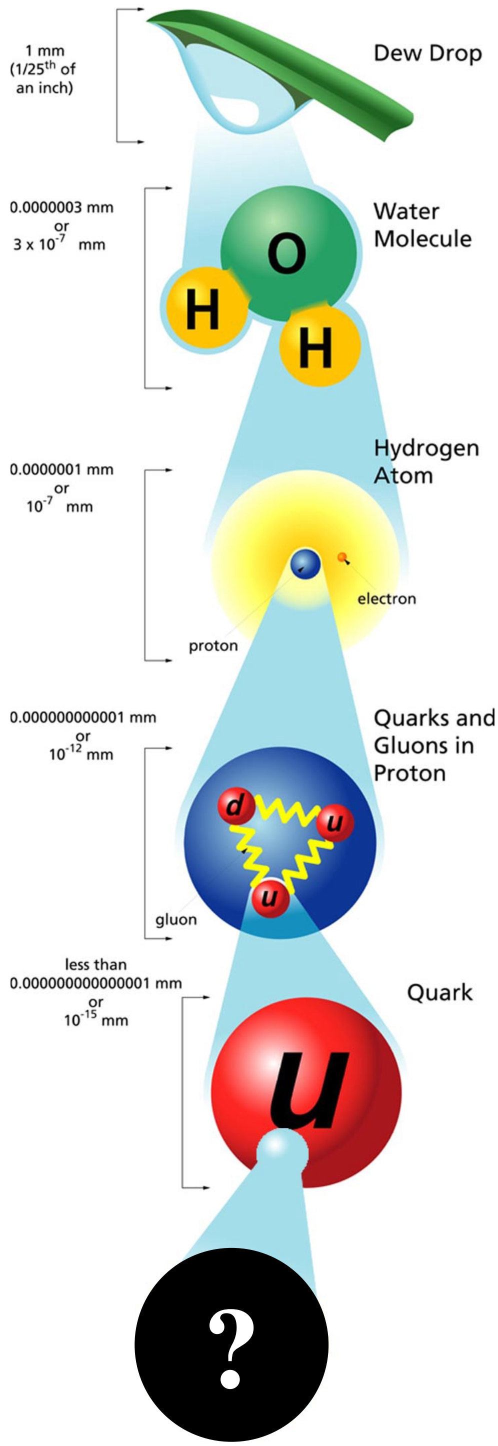


Figure 1.1: The composition of a drop of water as we zoom in. At one scale, we can say that it is composed of the elements oxygen and hydrogen. As we zoom in further, we can see that hydrogen is composed of a proton and an electron. Then we can see that a proton is composed of two up quarks and one down quark. Finally, we are at the limit of our zooming abilities and currently we believe that the up quark is fundamental. However, it could be composed of something more fundamental.

By the late 1960's, models existed that could describe the large number of newly discovered particles, but there was not a single, unified theory that described them all. Also, the origin of mass became an interesting question. In 1967, Steven Weinberg and Abdus Salam, using an interesting idea by Peter Higgs et al., were able to answer the latter question of mass by extending Glashow's theory of leptons. There was only one "problem" with this new theory of leptons: it predicted particles that had not been observed at the time. These particles would later be known as the massive gauge bosons (the W's and Z), and the Higgs boson. It was not obvious at this time that this extended lepton theory could be extended again to include the quarks, but by the 1970's, adding the quarks to this theory of leptons became the "standard." At last, this new combination of quarks and leptons summarized all known particles and their interactions, with the exception of gravity. This theory eventually was called the "standard model" of particle physics (SM). Einstein's general theory of relativity (GR) was then, and is now, the most generally accepted theory of gravity. However, merging GR with the standard model is extremely difficult, and to this day no one has managed to do it successfully. Since the strength of gravity is only relevant for extremely massive objects, it is generally safe to ignore it when considering elementary particles. With the formulation of the SM, it became clear that not all of the fundamental particles that it predicted had been discovered, so it was far from being fully verified.

During the 50 years after the initial formulation of the SM, particle physicists built larger and more complex particle accelerators, which began observing new particles predicted by the SM. The nature of these complex particle accelerators moved experiments from university labs to large-scale research facilities around the world, where thousands of physicists are needed to build, maintain, and analyze the data. It was no longer possible for a small group of physicists at a university to discover a new fundamental particle of nature. This massive effort eventually led to the discovery of all SM particles. With a huge announcement on July 4, 2012, the final particle, the Higgs boson, was discovered at CERN's Large Hadron Collider (LHC). The Higgs boson is the particle associated with the mechanism that gives all matter in the universe mass. Figure 1.2 shows the latest version of the "golden" channels used to discover the Higgs boson. This kind of approach is typically called a "bump hunt," where one looks for an excess or "bump" on top of a background. In these two independent plots, one can observe the presence of a bump at 125 GeV, and it was obvious that only a new particle with that mass could produce it.

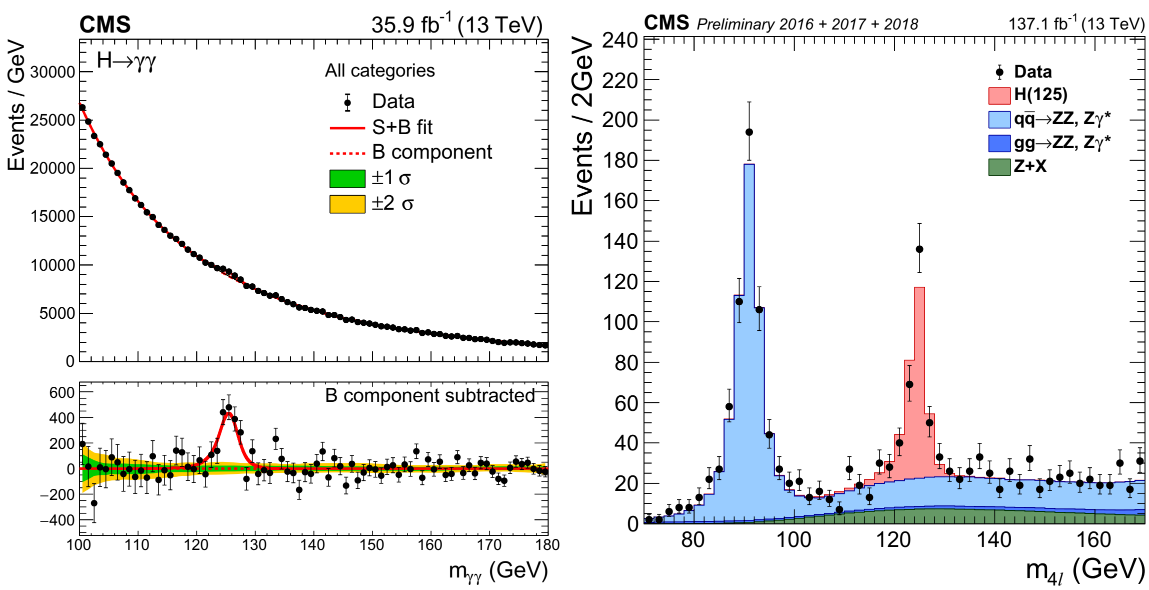


Figure 1.2: Higgs boson "golden" observation channels, which both show an excess of events at 125 GeV above the background based on the "no Higgs" hypothesis. (Left) Invariant mass of the diphoton system, where the Higgs decays to two photons. (Right) Invariant mass of a four-lepton system, in which the Higgs decays to four leptons [1, 2].

Today, physicists have discovered all fundamental particles that the SM predicts. A new "periodic table" of elementary particles is shown in Fig. 1.3 along with some of their properties. The current state of affairs is that physicists have a well-verified model with a few unobserved predictions, and are now in a situation where there are no more obvious particles left to be discovered. We now look to theoretical physicists to produce new models of nature that predict more particles or interesting phenomena for experimental physicists to investigate. One of these theories is called supersymmetry (SUSY), where the number of particles in Fig. 1.3 could effectively be doubled. This new unproven theory is the main motivation for this thesis, where we search for one of these new supersymmetric particles. In this case, we seek the scalar top quark, a particle that is associated with the top quark. I will go into more detail on the SM and this possible new theory, which is yet to be verified, in the following chapter. At this interesting point in physics we all wait with bated breath for any signs of new phenomena, because it will shape the field of particle physics for the foreseeable future.

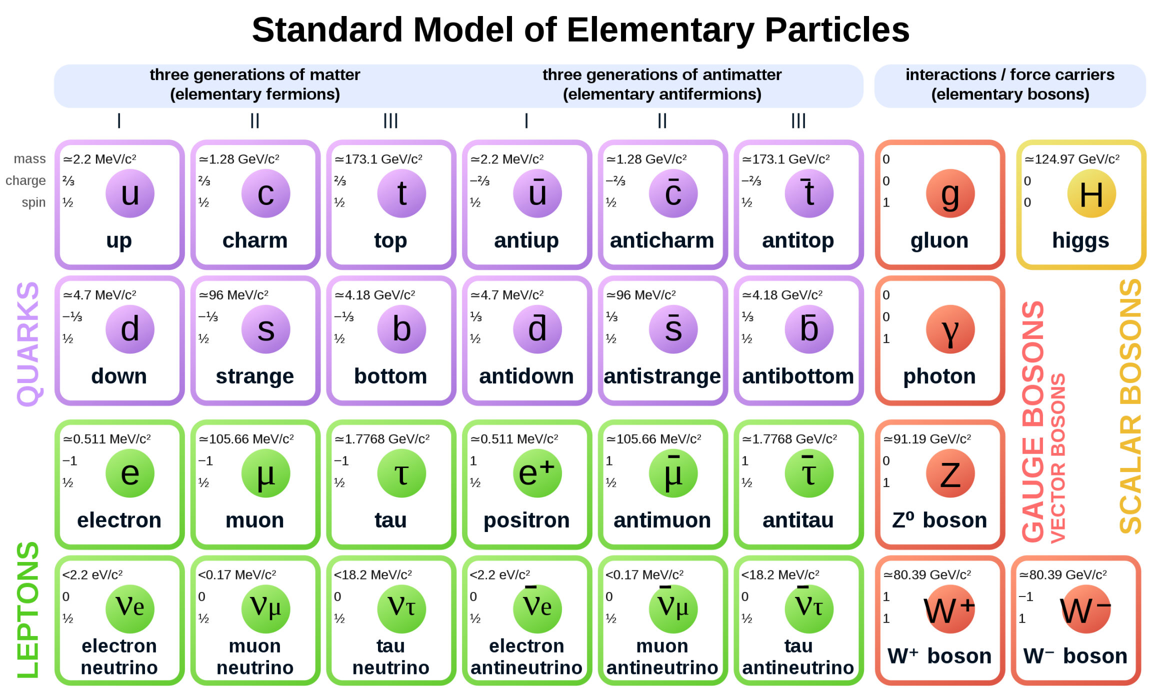


Figure 1.3: The particles of the standard model. The standard model includes three generations of quarks and leptons. The gauge bosons mediate the electromagnetic, weak, and strong forces, and the Higgs boson is associated with the mechanism that gives particles their mass. Image source [3].

CHAPTER TWO

The Standard Model and Supersymmetry

2.1 Introduction

What is commonly known as the standard model (SM) of particle physics is currently the best theory describing the particles in our universe and their interactions. The derivation of the SM is a conglomerate of brilliant ideas from many different physicists, with the most important concepts being the use of quantum field theory, Yang-Mills theory, the Brout-Englert-Higgs mechanism, and renormalization. It can be derived from a few assumptions, but one issue is that these assumptions can be viewed as "ad-hoc." We still need to put in a good amount of particle content by hand, assume a specific form for the Higgs potential, and require a specific local gauge invariance ($SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$). Also, there are a good number of free parameters that need to be measured. At the end of the day, it also does not describe all observed phenomena in the universe, and we are still actively searching for a so-called "theory of everything." I will go into more detail on the shortcomings of the SM in Section 2.2.5. However, it is undeniably the most successful theory to date for particle physics. One potential extension to the SM is Supersymmetry (SUSY), which effectively doubles the number of particles that the SM predicts. So far, no evidence of these extra particles has been observed, but a good candidate for discovery is the scalar top quark. I will go into more detail on SUSY in Section 2.3.

2.2 The Standard Model

I will go over a quick pedagogical derivation of the SM, highlighting important details from an experimental physicist's point of view. There are some aspects of this theory that go beyond the scope of this thesis, and for these I refer the reader to other dedicated sources. The summary I give below came mostly from these great sources [4–8].

2.2.1 Observed Particles and Interactions

First let us list out explicitly all of the known particles and their observed interactions. The fundamental particles can be broken up into two main categories, fermions (half-integer spin) and bosons (whole-integer spin), where spin is an intrinsic form of angular momentum carried by a particle. The fermions are considered the "matter" particles, while the bosons are considered the "force carrier" particles.

We have so far observed four distinct forces of nature. These forces are the electromagnetic, strong, weak, and gravitational forces. All four vary in strength and the reason why is still an open question. However, the SM can accurately describe all forces, except gravity. Three out of four is not bad, but so far gravity is best explained by general relativity, which is not a quantum theory. Unifying the description of the four forces is another open question in physics. For now we will ignore gravity and focus on the other three forces. We can do this because gravity can be safely ignored at a particle accelerator due to the small masses of the particles and the relatively tiny strength of gravity compared to the other forces. With these three forces in mind, we say the particles that mediate interactions for these forces are the bosons. The boson associated with the electromagnetic force is the photon (γ). The Z and

W bosons (Z^0, W^+, W^-) are associated with the weak force, and the eight gluons (g) are associated with the strong force. Note that the number of particles is as follows: one for the electromagnetic force, three for the weak force, and eight for the strong force.

We further divide the fermions into two categories, the quarks and the leptons, because the strong force only acts on the quarks. We can also group the fermions into different generations, where the up and down quarks (u, d), along with the electron neutrino and electron (ν_e, e), form the first generation. Similarly, the second generation consists of the charm and strange quarks (c, s) along with the muon neutrino and muon (ν_μ, μ). The third and final generation consists of the top and bottom quarks (t, b) along with the tau neutrino and tau (ν_τ, τ). The number of these particles increases by a factor of two when you include the antimatter version of each. We can look at Fig. 2.1 and see a full list of all observed particles. Note that the quarks all have different colors, which is to say they all possess a "color" charge related to the strong force. Because of this, the number of quarks and leptons differ by a factor of three.

You may be wondering where all of these particles come from, and how exactly they interact with each other. To answer most of this question we must first derive the Lagrangian for the SM. The SM is a quantum field theory, so we must start viewing particles as an excitation of a field, and it is these fields that we must spend our time thinking about.

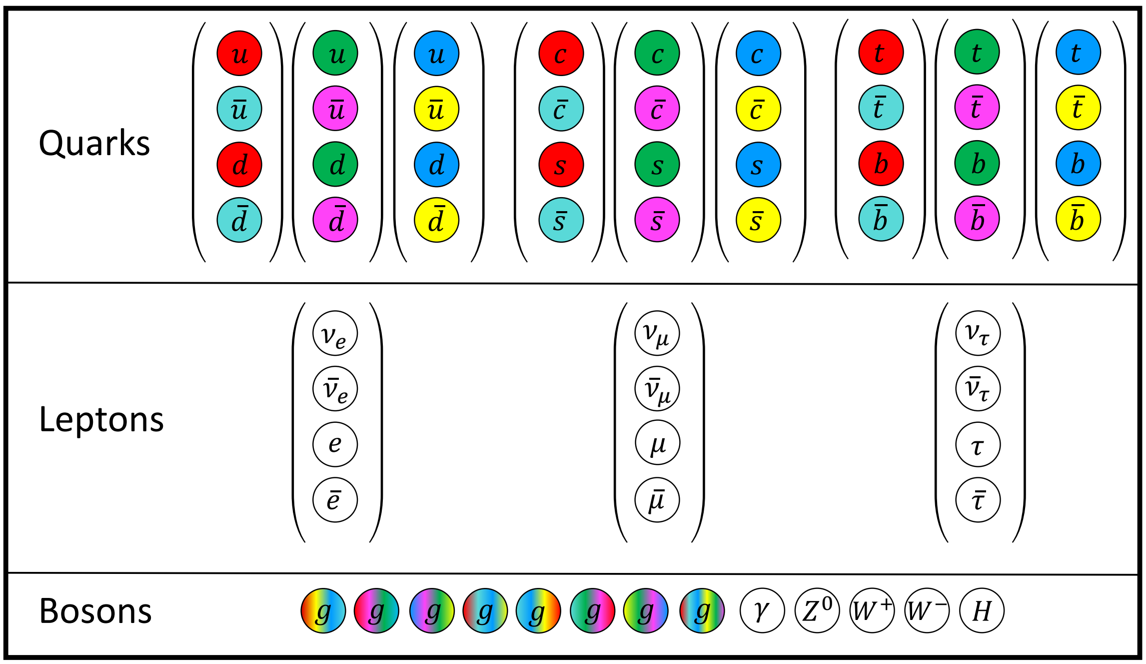


Figure 2.1: Summary of all observed particles, where a labeled circle corresponds to a unique particle. There are six types of quarks each having color charge, which means we have a red, green, blue version of each quark, and similarly we have an antired, antigreen, and antiblue version of each antiquark. This brings the total number of quarks to 36. There are six types of leptons, but unlike the quarks they only have an antilepton version. This brings the total number of leptons to 12. For the bosons, only the gluons (g) get multiple copies, each with a different double color charge for a total of 8. Then there are 5 more unique bosons, making the total number of bosons 13. This bring the total number of observed particles in the universe to $36 + 12 + 13 = 61$.

Table 2.1: Summary of fermion fields explicitly placed in the SM, including their quantum numbers: spin, electric charge (Q), third component of weak isospin (I_3), hypercharge Y, and color (r, g,b).

Fermions	Generation I	Generation II	Generation III	Spin	Q	I_3	Y	Color
Quarks	$\begin{pmatrix} u_L \\ d_L \end{pmatrix}$	$\begin{pmatrix} c_L \\ s_L \end{pmatrix}$	$\begin{pmatrix} t_L \\ b_L \end{pmatrix}$	$\frac{1}{2}$	$\frac{2}{3}$ $-\frac{1}{3}$	$\frac{1}{2}$ $-\frac{1}{2}$	$\frac{1}{3}$	r, g,b
	u_R	c_R	t_R	$\frac{1}{2}$	$\frac{2}{3}$	0	$\frac{4}{3}$	r, g,b
	d_R	s_R	b_R	$\frac{1}{2}$	$-\frac{1}{3}$	0	$-\frac{2}{3}$	r, g,b
Leptons	$\begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}$	$\begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}$	$\begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix}$	$\frac{1}{2}$	0 -1	$\frac{1}{2}$ $-\frac{1}{2}$	-1	-
	e_R	μ_R	τ_R	$\frac{1}{2}$	-1	0	-2	-

2.2.2 The Lagrangian of the Standard Model

When putting together the SM Lagrangian, we must first write down the kinetic terms for all of the fermions explicitly, and since they are spin-$\frac{1}{2}$ fields, we must treat them as spinors. For simplicity we will define the leptons as normal 4-component Dirac spinors. For example, the electron field is defined as

$$e = \begin{pmatrix} e_0 \\ e_1 \\ e_2 \\ e_3 \end{pmatrix} \tag{2.1}$$

such that e_i for $i \in \{0, 1, 2, 3\}$ are complex functions. However, we will define the quark fields as a triplet of Dirac spinors, one for each color. For example, the up quark field is defined as

$$u = \begin{pmatrix} u_{\text{red}} \\ u_{\text{green}} \\ u_{\text{blue}} \end{pmatrix} \tag{2.2}$$

such that u_{color} for $color \in \{\text{red}, \text{green}, \text{blue}\}$ are the 4-component Dirac spinors. We will also be using chiral representation, so we can consider the left- and right-handed components of each field separately. This is done because the weak force only interacts with the left-handed components of the fields. By using the left-handed $\left(\frac{1-\gamma^5}{2}\right)$ and right-handed $\left(\frac{1+\gamma^5}{2}\right)$ projection operators, any field can be separated into their left- and right-handed components. For example, the electron field becomes

$$e = e_L + e_R \tag{2.3}$$

$$e_L = \left(\frac{1-\gamma^5}{2}\right) e \tag{2.4}$$

$$e_R = \left(\frac{1+\gamma^5}{2}\right) e. \tag{2.5}$$

Next, it is convenient to group some of the left-handed fields into doublets (array of 4-component Dirac spinors), and leave the right-handed fields as singlets (single 4-component Dirac spinor). For example, the components of the first generation of left-handed fermions are grouped as

$$Q_{ud} = \begin{pmatrix} u_L \\ d_L \end{pmatrix} \tag{2.6}$$

$$L_e = \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}. \tag{2.7}$$

The last thing we need to consider is the right-handed component of the neutrino fields. We know in general that neutrinos do not have electric charge or color charge, and only interact with other particles through the weak force. Due to the weak force only allowing the left-handed component of fields to interact, no interaction will occur with the right-handed component of neutrino fields in the SM. The question of whether they exist is still an open question in physics, but since the SM does not need them, we will not consider right-handed neutrinos. Putting this altogether and writing out each field, we get the kinematic terms for the SM Lagrangian of the fermion fields:

$$\begin{aligned}\mathcal{L}_{\text{Kin,Fermions}} = & \underbrace{i\overline{Q}_{ud}\not{D}Q_{ud} + i\overline{Q}_{cs}\not{D}Q_{cs} + i\overline{Q}_{tb}\not{D}Q_{tb}}_{\text{Left-handed quarks}} \\ & + \underbrace{i\overline{u}_R\not{D}u_R + i\overline{d}_R\not{D}d_R + i\overline{c}_R\not{D}c_R + i\overline{s}_R\not{D}s_R + i\overline{t}_R\not{D}t_R + i\overline{b}_R\not{D}b_R}_{\text{Right-handed quarks}} \\ & + \underbrace{i\overline{L}_e\not{D}L_e + i\overline{L}_\mu\not{D}L_\mu + i\overline{L}_\tau\not{D}L_\tau}_{\text{Left-handed leptons}} \\ & + \underbrace{i\overline{e}_R\not{D}e_R + i\overline{\mu}_R\not{D}\mu_R + i\overline{\tau}_R\not{D}\tau_R}_{\text{Right-handed leptons (no right-handed neutrinos)}}\end{aligned} \tag{2.8}$$

such that $\not{D} \equiv \gamma^\mu D_\mu$,[1] γ^μ are the usual gamma matrices, and D_μ is the covariant derivative. Note that this is just the Dirac Lagrangian for each field, but without a mass term. The mass term will be handled later by the Brout-Englert-Higgs mechanism. The representations of the fermion fields that are added to the SM Lagrangian are listed in Table 2.1. Following Yang-Mills theory, we need to look closely at the covariant derivative, and think about the local symmetry that we will impose and what it means logistically.

The SM requires a $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ local gauge invariance, which in turn requires us to add some gauge fields to preserve the local gauge invariance. This can be done by using a particular covariant derivative that depends on how we

[1] We will use Einstein summation notation, where two identical indeces implies a sum over that index. If the index is the Greek μ or ν, then we add the Minkowski metric ($g_{\mu\nu}$):

$$\gamma^\mu D_\mu = \sum_{\mu=0}^{3}\sum_{\nu=0}^{3} g_{\mu\nu}\gamma^\mu D^\mu$$

expect the fermion to interact, and can be different for each field:

$$D_\mu = \partial_\mu - \frac{1}{2} i g_s \Gamma_s \lambda_\alpha G^\alpha_\mu - i g |I_3| \sigma_j W^j_\mu - \frac{1}{2} i g' Y B_\mu \tag{2.9}$$

where g_s is the coupling strength for $SU(3)$, Γ_s is 1 for quarks and 0 for other fields, λ_α for $\alpha \in \{1, 2, 3, 4, 5, 6, 7, 8\}$ are the eight generators for $SU(3)$ (also known as the Gell-Mann matrices), G^α_μ are the eight $SU(3)$ gauge fields, g is the coupling strength for $SU(2)$, I_3 is the third component of weak isospin, σ_j for $j \in \{1, 2, 3\}$ are the three generators for $SU(2)$ (also known as the Pauli matrices), W^j_μ are the three $SU(2)$ gauge fields, g' is the coupling strength for $U(1)$, Y is hypercharge, and B_μ is the $U(1)$ gauge field. The observed I_3 and Y values for each fermion field are listed in Table 2.1.

We will see later that the gauge fields that we were forced to add, due to requiring local gauge invariance, will turn into the vector boson fields of the SM. Unlike the fermions, the bosons (except for the Higgs) are a consequence of local gauge invariance. Although the covariant derivative adds in interacting terms between the fermions and the vector bosons, it does not add in the kinematic terms. Therefore, we must add them in by hand:

$$\mathcal{L}_{\text{Kin,Bosons}} = \underbrace{-\frac{1}{4} G^\alpha_{\mu\nu} G^{\mu\nu}_\alpha}_{SU(3)\text{ gauge field's term}} + \underbrace{-\frac{1}{4} W^j_{\mu\nu} W^{\mu\nu}_j}_{SU(2)\text{ gauge field's term}} + \underbrace{-\frac{1}{4} B_{\mu\nu} B^{\mu\nu}}_{U(1)\text{ gauge field's term}} \tag{2.10}$$

where we have defined tensors using the gauge fields, the $SU(3)$ structure constant tensor (f^{abc}), and the $SU(2)$ structure constants tensor (ϵ^{abc}):

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu \tag{2.11}$$

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g \epsilon^{abc} W^b_\mu W^c_\nu \tag{2.12}$$

$$B_{\mu\nu} = \partial_\mu B^a_\nu - \partial_\nu B^a_\mu. \tag{2.13}$$

If we add these two Lagrangians together we will have all SM fields plus their interactions that correspond to all observed particles (except for the Higgs boson), but they are all massless. If we were to add the mass terms now by hand, the local gauge invariance we imposed would not hold. There is a solution to this problem, which was alluded to earlier, called the Brout-Englert-Higgs mechanism.

First we must introduce the Higgs field, which is a complex scalar $SU(2)$ doublet with $Y = 1$:

$$\phi = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix} = \begin{pmatrix} r_+ e^{i\theta_+} \\ r_0 e^{i\theta_0} \end{pmatrix} \tag{2.14}$$

where $\phi_{+,0}$ are complex, $r_{+,0}$ are positive real-valued, and $\theta_{+,0}$ are the complex phases. The complex conjugate is defined as:

$$\tilde{\phi} = \begin{pmatrix} \phi_0^* \\ -\phi_+^* \end{pmatrix}. \tag{2.15}$$

The Higgs field can now contribute to the SM Lagrangian with a kinetic term, a cleverly chosen potential, and Yukawa terms that connect the Higgs field to the fermions:

$$\begin{aligned} \mathcal{L}_\phi = & (D^\mu\phi)^\dagger(D_\mu\phi) - V(\phi^\dagger\phi) \\ & - \left[\lambda_e \overline{L}_e \phi e_R + \lambda_u \overline{Q}_{ud} \tilde{\phi} u_R + \lambda_d \overline{Q}_{ud} \phi d_R + \text{h.c.}\right] \\ & - \left[\lambda_\mu \overline{L}_\mu \phi \mu_R + \lambda_c \overline{Q}_{cs} \tilde{\phi} c_R + \lambda_s \overline{Q}_{cs} \phi s_R + \text{h.c.}\right] \\ & - \left[\lambda_\tau \overline{L}_\tau \phi \tau_R + \lambda_t \overline{Q}_{tb} \tilde{\phi} t_R + \lambda_b \overline{Q}_{tb} \phi b_R + \text{h.c.}\right] \end{aligned} \tag{2.16}$$

where λ_f corresponds to the Yukawa coupling for a particular fermion (f). Note the use of "h.c." implies that the terms listed previously are repeated after taking the hermitian conjugate. The potential term is defined as

$$V(\phi^\dagger\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2. \tag{2.17}$$

This form of the potential was chosen because it gives us the results we are looking for, as we will see. It can be viewed as "ad-hoc" since there is no fundamental reason for this exact form. In fact, there are other, more complicated forms that can give us the same result.

Looking at the potential term in more detail, we can rewrite it in terms of the real numbers r_+ and r_0:

$$V(r_+, r_0) = \mu^2(r_+^2 + r_0^2) + \lambda(r_+^2 + r_0^2)^2. \tag{2.18}$$

This potential is plotted as a function of r_+ and r_0 in Fig. 2.2 for the experimentally observed values of the parameters. The sign of λ has to be positive to prevent the potential from becoming unstable, but there are two choices for the sign of μ^2. For $\mu^2 > 0$, the potential is always positive with a minimum or vacuum expectation value (vev) of zero:

$$\langle 0|\phi|0\rangle = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{2.19}$$

However, for $\mu^2 < 0$, the true minimum potential becomes $\sqrt{-\frac{\mu^2}{2\lambda}}$, and results in the famous "Mexican hat" potential.

Having the minima no longer occur at the origin gives the neutral component of ϕ a non-zero vev, which leads to spontaneous symmetry breaking as the Higgs field "rolls" from the origin down toward the minima:

$$\langle 0|\phi|0\rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \text{ where } v = \sqrt{-\frac{\mu^2}{\lambda}}. \tag{2.20}$$

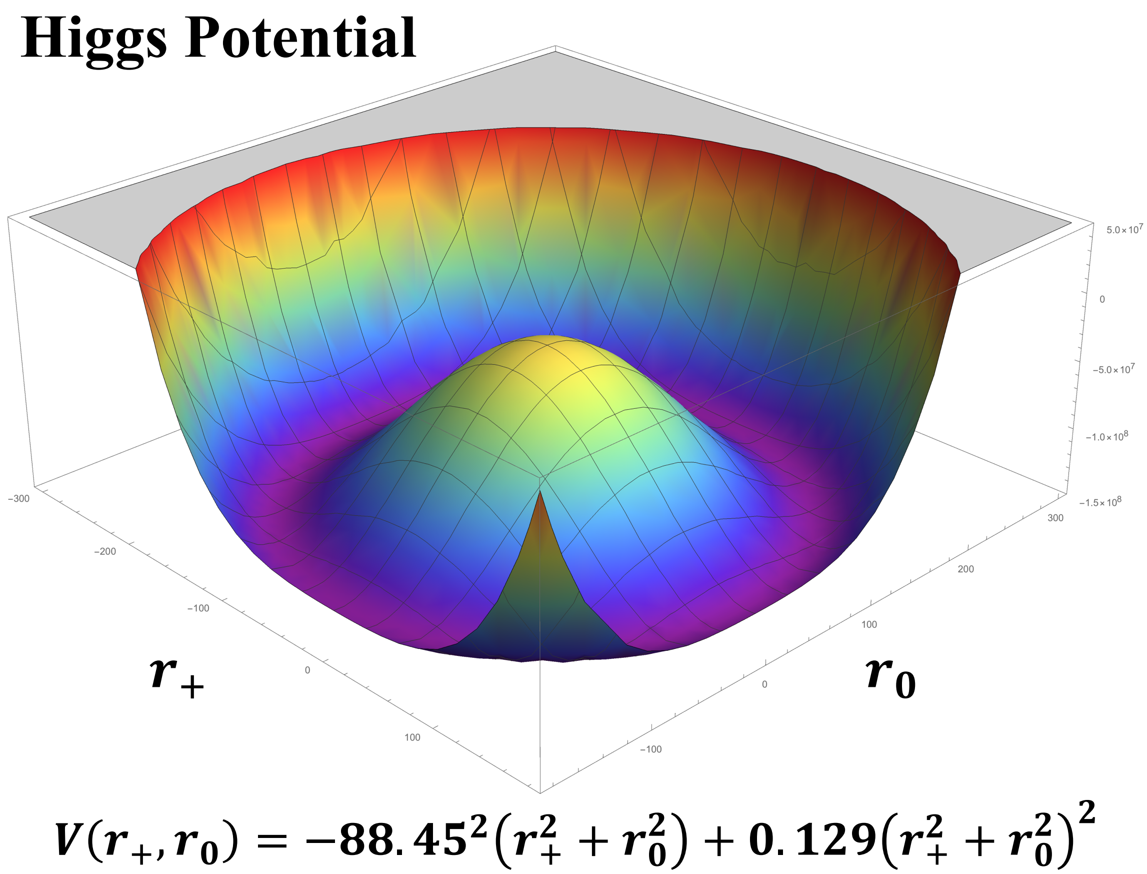


Figure 2.2: Higgs potential plotted with the experimentally observed parameter values.

This allows us to now redefine the Higgs field as an expansion around its minimum:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix} \tag{2.21}$$

where H becomes the Higgs boson field. After updating the Higgs field with this expansion, the Higgs covariant derivative leads to some of the bosons obtaining mass (including the Higgs boson), and the Yukawa terms lead to the fermions obtaining mass. This becomes apparent for the vector bosons after redefining the gauge fields as

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \tag{2.22}$$

$$Z_\mu = \frac{1}{\sqrt{g^2 + g'^2}}(gW_\mu^3 - g'B_\mu) \tag{2.23}$$

$$A_\mu = \frac{1}{\sqrt{g^2 + g'^2}}(gW_\mu^3 + g'B_\mu) \tag{2.24}$$

where we can now say $W_\mu^\pm$, Z_μ, and A_μ correspond to the two W's, Z, and photon fields, respectively. After some plugging and chugging for the Higgs portion of the Lagrangian, we can work out the mass of each particle, and its relation to the Higgs boson as

$$m_\mathrm{W} = \frac{1}{2}vg \tag{2.25}$$

$$m_\mathrm{Z} = \frac{1}{2}v\sqrt{g^2 + g'^2} \tag{2.26}$$

$$m_A = 0 \tag{2.27}$$

$$m_H = \sqrt{2\lambda v^2} \tag{2.28}$$

$$m_f = \frac{\lambda_f v}{\sqrt{2}} \text{ , for f being any massive fermion.} \tag{2.29}$$

Notice for all of the masses, except for the Higgs boson, there is an explicit coupling value that must be known. The mass of the Higgs can be determined from the free parameters of the Higgs potential.

Another interesting feature of the SM is that couplings depend on the energy scale, also known as the "running" of the couplings. In fact the running of the $SU(3)$ coupling explains confinement and the asymptotic-free nature of the strong force. At low energy scales g_s goes to infinity (confinement), while at high energy it gets weaker (asymptotic freedom). This can be seen in Fig. 2.3, which shows the measured running of α_s, which is related to g_s by

$$\alpha_s = \frac{g_s^2}{4\pi}. \tag{2.30}$$

The running of couplings also implies that the masses of all particles should change depending on the running of their Yukawa couplings to the Higgs field. In fact, it would be an interesting analysis to measure the running of the Higgs boson to see if it changes depending on the energy scale. The value of the Higgs mass depends on the value of the vev and parameters of the Higgs potential, so before quantum corrections are made it should be a constant. It is also interesting to note that the couplings of the electromagnetic, weak, and strong forces seem to have an energy scale where their values almost converge to the same value. This "almost" unification of the forces motivates people to think that a theory beyond the SM (BSM) should have a unification scale for all of the forces. This is actually one motivation for SUSY, which will be explained in Section 2.3. I will not go into more detail here about why the coupling runs as a function of energy scale, but the interested reader should look up the beta function from the renormalization group for more information.

With all of these pieces now established, we can finally write out the full SM Lagrangian as

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{Kin,Fermions}} + \mathcal{L}_{\text{Kin,Bosons}} + \mathcal{L}_{\phi}. \tag{2.31}$$

With the discovery of the Higgs boson in 2012, all particles expected from the SM have been observed, and there are only a few very rare interactions not yet observed.

2.2.3 Interactions of Particles

Looking at the Lagrangian, we can see how all particles in the SM interact with each other. Figure 2.4 shows how each particle interacts with the others, indicated by connecting lines. Note that there are some self interactions. In general, we see all fermions can only interact by means of a boson, while bosons can interact directly.

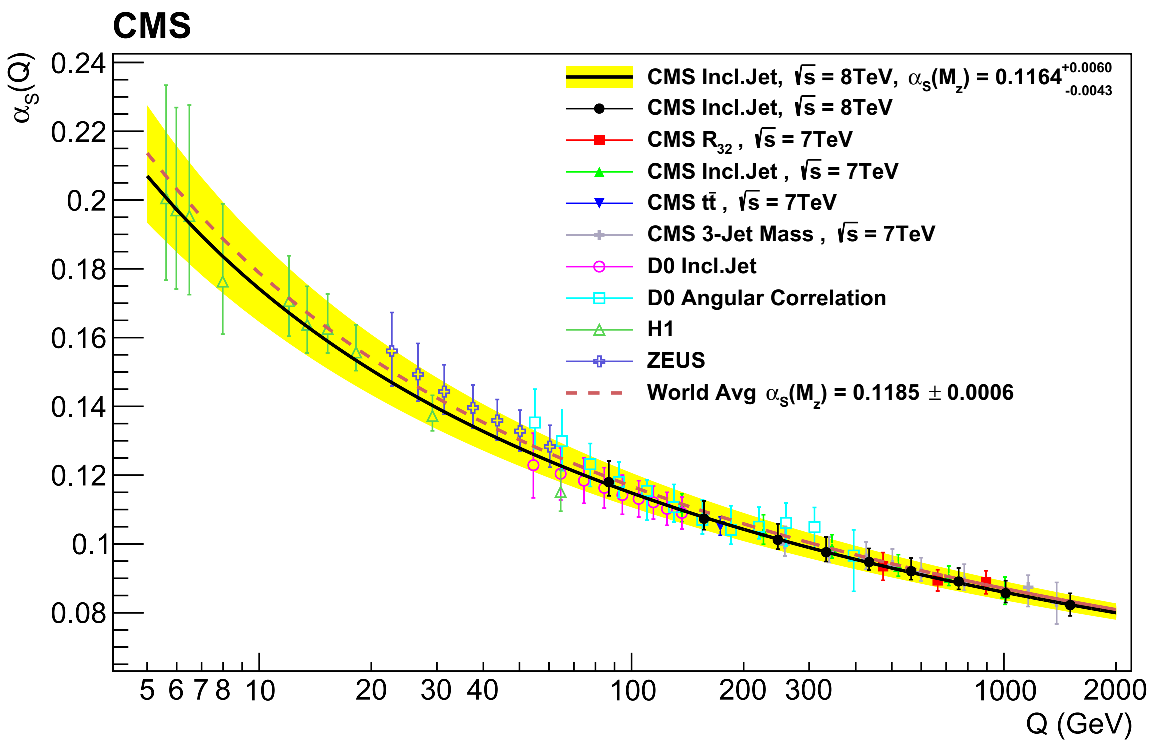


Figure 2.3: Running of α_s measured by various collaborations as a function of energy scale Q.

This leads us to say that the bosons are the force mediators for the fundamental forces of nature. We break interactions into two categories, ones that come from gauge bosons, and other interactions that include the Higgs boson. Figure 2.5 shows all vertices that come from gauge interactions. The gauge interactions are predominantly what is seen in nature, and at a hadron collider interactions that include a gluon (strong interactions) are very common. Interactions including a Higgs boson are very rare compared to gauge interactions, so we will not consider them for this thesis.

For this thesis, there are two strong interactions that I would like to focus on. Figure 2.6 shows the possible production of two top quarks, and production of a QCD multijet. A jet is defined as the signature seen in a detector for either a quark or gluon. Looking closer at the top quark production diagram, we see that two gluons interact

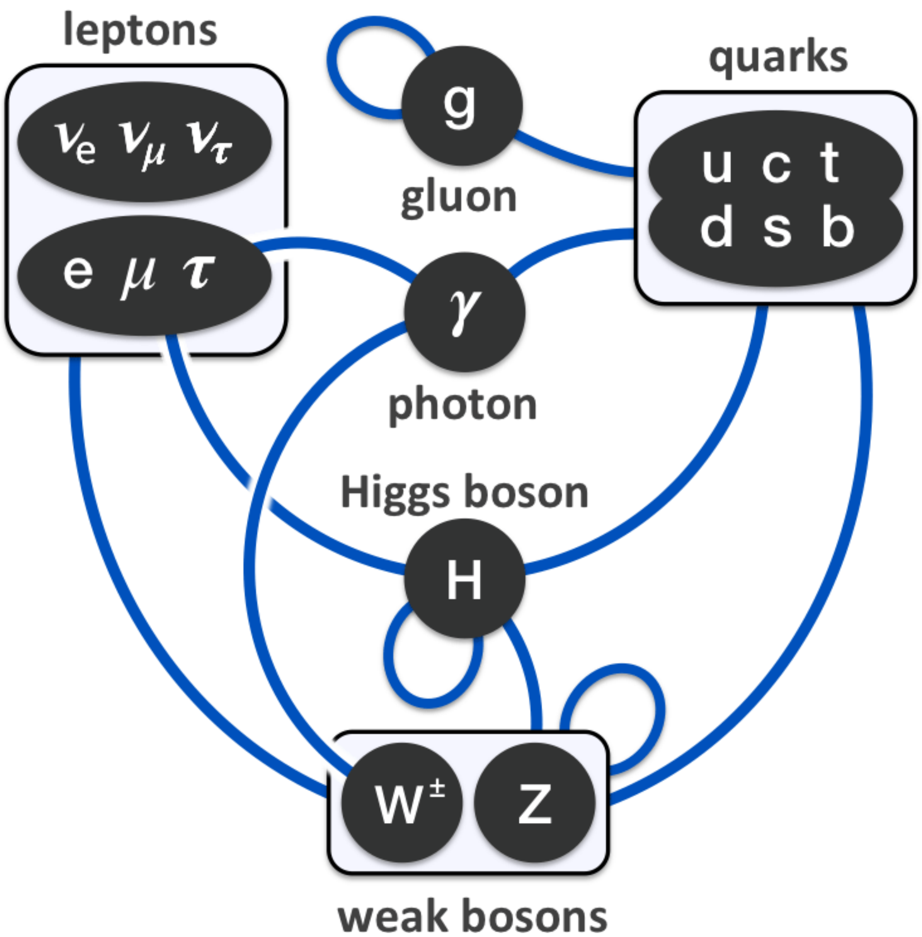


Figure 2.4: Diagram with lines showing the interactions between different particles including some that self interact [3].

and produce a $t\bar{t}$ pair. Since the top quark has a very large mass, it decays almost instantaneously into a W boson and a bottom quark. For this thesis we will consider the case where one of the W bosons decays into two low-mass quarks, and the other decays into a lepton and its neutrino. The QCD multijet Feynman diagram can look similar to top quark production except that it keeps using the gluon self-interaction vertex until it has produced the same number of objects as the $t\bar{t}$ production diagram. It is important to see that all vertices for these diagrams come from a combination of strong and weak gauge interactions, assuming two initial gluons. We will see that these two diagrams lead to the largest SM background for our search for new physics.

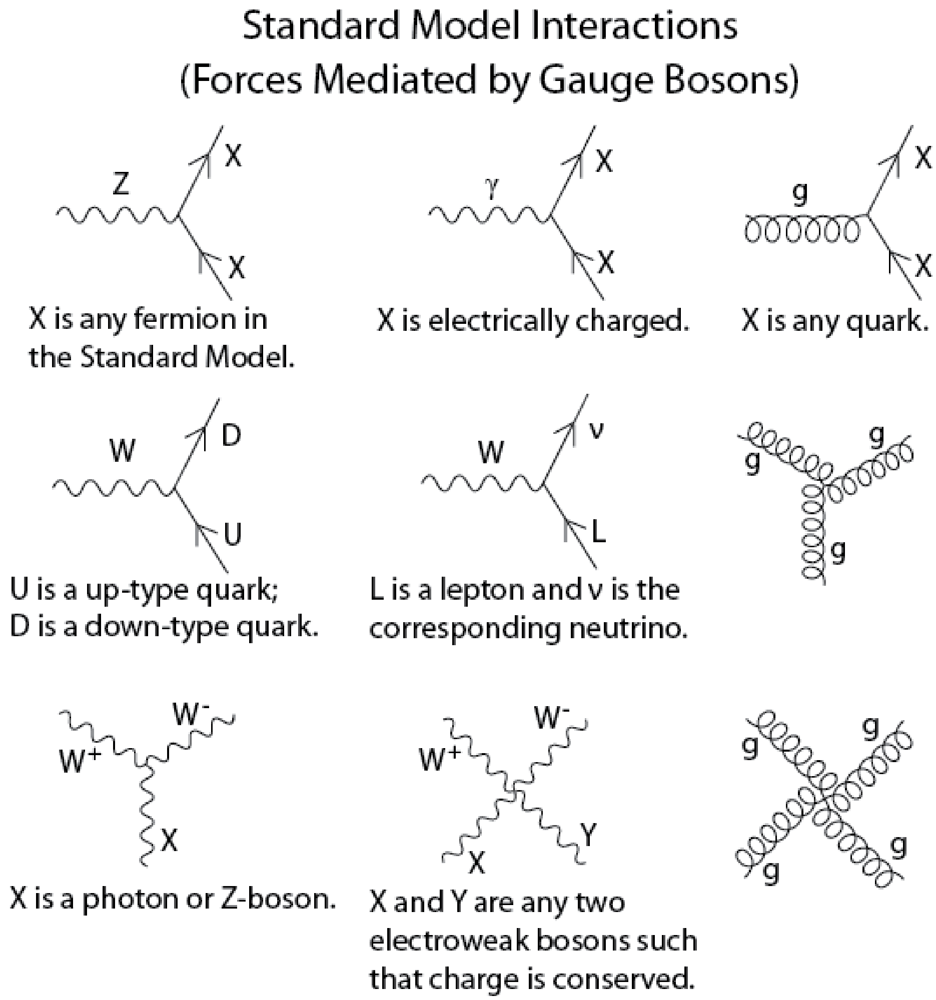


Figure 2.5: Summary of vertices of all gauge-boson-mediated interactions [3].

2.2.4 *Symmetries of the Standard Model*

The $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ gauge symmetries that we explicitly placed in the SM, following Neother's theorem, says that there must be some conserved quantities. This leads to the conservation of color charge from $SU(3)_C$, weak isospin (I_3) from $SU(2)_L$, and hypercharge (Y) from $U(1)_Y$. The observed electric charge can be obtained by combining these values, using the Gell-Mann-Nishijima relation

$$Q = I_3 + \frac{Y}{2}. \tag{2.32}$$

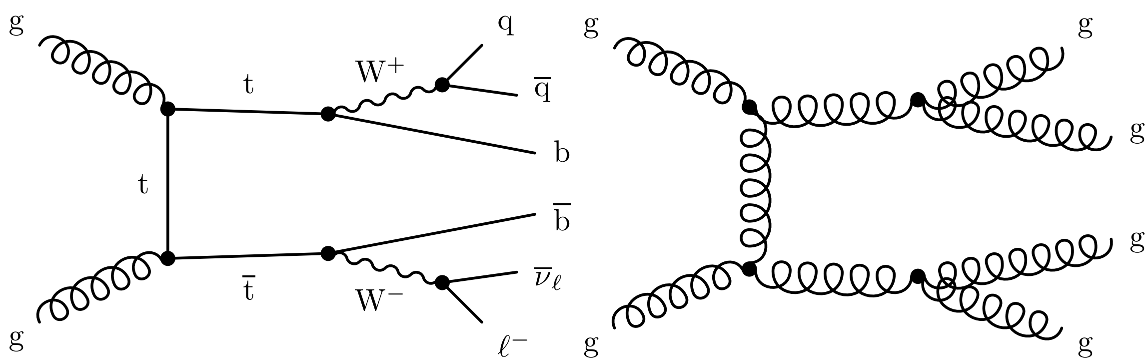


Figure 2.6: Diagrams from a hadron collider that produce four total gluons or quarks. (Left) Example for semileptonic $\mathrm{t\bar{t}}$ production. (Right) Example of QCD production, which produces the same number of jets as semileptonic $\mathrm{t\bar{t}}$.

Another interesting feature of the SM is the "accidental" symmetries of the Lagrangian. There are some transformations that leave the Lagrangian invariant that were not explicitly required, and for every symmetry there exists a conservation law. For example, we have four $U(1)$ global symmetries that lead to baryon and lepton number conservation:

- $Q \to e^{\alpha/3}Q \longleftrightarrow$ Conservation of baryon number
  (acting on all quark fields)
- $E \to e^{\beta}E \longleftrightarrow$ Conservation of electron number
  (acting on electron and electron neutrino fields)
- $M \to e^{\beta}M \longleftrightarrow$ Conservation of muon number
  (acting on muon and muon neutrino fields)
- $T \to e^{\beta}T \longleftrightarrow$ Conservation of tau number
  (acting on tau and tau neutrino fields)

For baryon number, all quarks are assigned a value of $1/3$, and all antiquarks a value of $-1/3$. For example, this leads to a baryon number of 1 for the proton and 0 for the pions. We define lepton number as the sum of electron, muon, and tau numbers.

The conservation of baryon and lepton numbers has so far been consistent with all experimental observations. This means that any extension to the SM should also have these symmetries or must weakly break them.

2.2.5 Issues with the Standard Model

Although the SM can successfully describe an extensive amount of physical phenomena, we know that it is not the final story. In fact we can view the SM as an effective theory of some larger more inclusive theory, such that when the appropriate approximations are made it becomes equivalent to the SM. This has been the litmus test of extensions to well established models ever since Isaac Newton wrote his famous laws of motion. The open questions of the SM can be thought of as two separate categories, experimental observations that it does not describe, and fundamental assumptions or consequences that are unknown. At the end of the day it could be argued that any theory will philosophically have some open questions, regardless of how successful it is. For brevity, I will leave those questions to the theorists and philosophers, and only list a subset of the issues, focusing on what the community views as the most important. Some of the biggest experimentally observed issues are:

- *Gravity:* This is a famous issue with the SM. I will not go into the reasons why, but so far there have been no successful quantum theories of gravity.
- *Dark Matter and Dark Energy:* We know that the energy and matter content of the universe is only composed of $\sim$5% ordinary matter. As of today there has been no confirmed theory that accurately describes what dark matter or dark energy is composed of.

- *Matter-Antimatter Asymmetry:* Considering the ~5% of matter and energy that we know about, the universe is composed of mostly ordinary matter, and there is almost no antimatter. Assuming equal amounts of matter and antimatter were produced at the beginning of the universe, there is no obvious mechanism for this asymmetry.
- *Origin of Neutrino Mass:* With the observation of neutrino oscillations, we know that the neutrinos should possess a non-zero mass. However, presently this is considered a minor change to the SM, and in fact some people define the SM by the most popular method for giving neutrino masses. I list it here because the exact mechanism is still in question, so until it is definitive I will consider the SM to have massless neutrinos.

The issues listed above, with the exception of neutrino mass, will require a major change to the SM or will come from a theory with completely different assumptions. Some of the fundamental assumptions or consequences of the SM that are still open questions include:

- *Higgs Mass Hierarchy Problem:* The value of the Higgs boson mass, when we take into account quantum corrections, should be many orders of magnitude higher than it is measured to be without some extraordinary fine tuning. I will go over this in more detail in the SUSY section (Section 2.3.1) because it can offer a solution to this problem.
- *Particles' Mass Values:* This might be more of a philosophical question, but is there a reason for the particular value of each particle's mass? This can be reformulated as: why does each particle have its particular value of the

Yukawa coupling to the Higgs? Perhaps it is a sign of some symmetry that has yet to be seen.

- *Form of the Higgs Potential:* The exact form of the Higgs potential is unknown, and the form that is used in the SM was assumed.
- *Baryon and Lepton Number Conservation:* These conservation laws are from accidental symmetries in the SM. Up to the point that they have been experimentally verified, they could be violated.
- *Three Families of Fermions:* Nature has chosen to have three copies of each family of fermions with different mass values. Is this a sign of some underlying symmetry that has yet to be formulated?
- *Force Unification:* Electromagnetism and the weak force unify in strength at a particular energy scale, but at no energy scale do all three forces unify. Should we be looking for a Grand Unified Theory (GUT)?
- *Right-handed Neutrinos:* The SM does not have right-handed neutrinos (left-handed antineutrinos) in its current form. Perhaps, this can be solved when we explain the origin of neutrino mass.

Finding a model that answers these questions and is consistent with the SM would lead the Nobel committee to make an easy decision. Many great physicists are working on this, and in time we will find solutions to many of these problems. SUSY offers solutions to some of these problems, which is the reason why we are considering it for this thesis.

2.3 Supersymmetry

A possible extension to the standard model is SUSY. The general idea behind SUSY is that all particles in the SM have a partner particle. The main motivation for assuming that there are almost double the number of particles that have been observed is that SUSY offers a solution to the Higgs mass hierarchy problem. This is based on an argument that nature has chosen fundamental values of some physical parameters such that their effective values are not "unnaturally" different—in this case, the bare mass of the Higgs boson vs. the observed mass. The summary I give below came mostly from these great sources [9–13].

2.3.1 Higgs Mass Hierarchy Problem

At the heart of the SM is the energy scale for which it is valid. It works well at the electroweak energy scale of $\mathcal{O}(10^2)$ GeV, but as energy increases to the Planck scale of $\mathcal{O}(10^{19})$ GeV, gravity is no longer negligible. This issue manifests itself when calculating the observed mass of each particle by including the bare mass, tree-level, and radiative loop corrections. When evaluating the momentum integrals of the loop corrections it is necessary to stop at the scale at which we believe the theory is no longer valid, in this case the Planck scale. When doing this for all massive fermions and vector gauge bosons, the loop correction terms have a reasonable dependence on this cut-off scale. However, the Higgs boson is a different story since it is a scalar which gives its mass calculation a quadratic dependence on the cut-off scale from the loop correction terms. As Fig. 2.7 illustrates, including the loop correction terms from the heaviest particle that couples to the Higgs, the top quark with $m_{top} \approx 173$ GeV, results in a correction term of the form

$$\Delta m_H^2 = N_{top}\frac{m_{top}^2}{4\pi^2 v^2}\left[-\Lambda^2 + 6m_{top}^2 \log\left(\frac{\Lambda}{m_{top}}\right) - 2m_{top}^2\right] + \mathcal{O}\left(\frac{1}{\Lambda^2}\right) \tag{2.33}$$

where Λ is the cut-off scale—in this case the Planck scale—and N_{top} is a repetition number. For the top quark, $N_{top} = 3$ due to its color charge. This leads to the calculation of the observed Higgs mass (m_H) as

$$m_H^2 = m_{Bare}^2 - N_{top}\frac{m_{top}^2}{4\pi^2 v^2}\Lambda^2 + \mathcal{O}(\Lambda) \approx (125\text{ GeV})^2 \tag{2.34}$$

which implies the Higgs bare mass (m_{Bare}) is

$$m_{Bare}^2 \approx 125^2 + N_{top}\frac{m_{top}^2}{4\pi^2 v^2}\Lambda^2 + \mathcal{O}(\Lambda) \approx (\mathcal{O}(10^{19})\text{ GeV})^2. \tag{2.35}$$

Looking at the size of m_H^2 vs. m_{Bare}^2 implies that in the SM there is fine tuning—an extremely contrived cancellation between the Higgs bare mass and the corrections needed to calculate the observed Higgs mass.

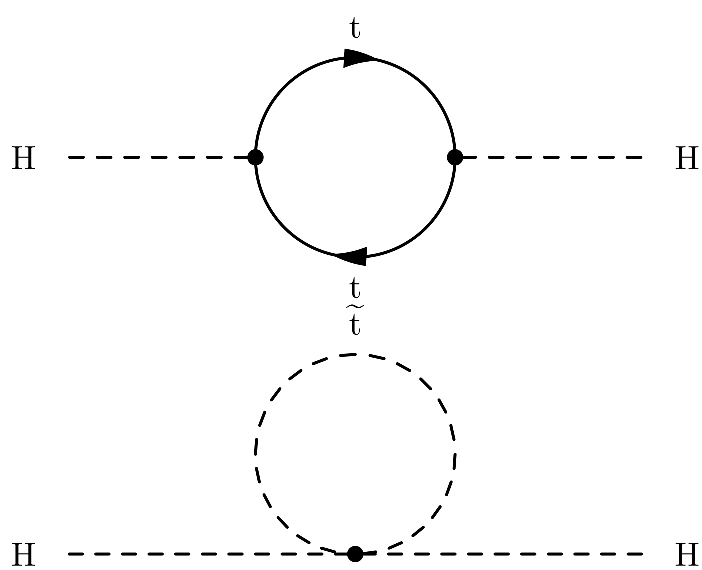


Figure 2.7: The quadratically divergent loop corrections to the Higgs boson mass. (Top) Corrections from the top quark. (Bottom) Corrections from the scalar partner of the top quark.

One idea for rectifying the Higgs mass hierarchy problem is to assume that there are some other corrections to the Higgs mass that are on the same order that can cancel out these extremely large fermion loop corrections. This can be achieved for the top quark by assuming there are two new complex color-triplet scalar fields ϕ_s, one for the right-handed top field and another for the left-handed top field, with masses m_s. We then assume that they couple to the Higgs field with trilinear and quadrilinear terms with $v\lambda_s$ and λ_s couplings. The correction terms to the Higgs mass, as seen in Fig. 2.7, for this scalar field have the form

$$\Delta m_H^2, s = -\frac{N_s \lambda_s}{16\pi^2}\left[-\Lambda^2 + 2m_s^2 \log\left(\frac{\Lambda}{m_s}\right)\right] - \frac{N_s \lambda_s^2}{16\pi^2} v^2 \left[-1 + 2 \log\left(\frac{\Lambda}{m_s}\right)\right] + \mathcal{O}\left(\frac{1}{\Lambda^2}\right) \tag{2.36}$$

where N_s is the repetition number. For this case $N_s = 2N_{top}$ since we are adding two color-triplet scalar fields. If we assume that $\lambda_s^2 = \frac{2m_{top}^2}{v^2}$, adding the Higgs' mass contribution terms of the top quark and this new scalar yields

$$\Delta m_H^2, tot = \frac{N_{top} m_{top}^2}{2\pi^2 v^2}\left[\left(m_{top}^2 - m_s^2\right) \log\left(\frac{\Lambda}{m_s}\right) + 3m_{top}^2 \log\left(\frac{m_s}{m_{top}}\right)\right] + \mathcal{O}\left(\frac{1}{\Lambda^2}\right). \tag{2.37}$$

Now all quadratically divergent terms have been canceled out, and if it turns out that the $m_{top} = m_s$, the logarithmically divergent terms will vanish too.

With an extra scalar field for every fermion, all divergent terms for the SM fermions can be canceled. Following a similar argument, the addition of a fermion for every boson in the SM that has the right couplings to the Higgs boson will remove their associated divergent terms as well. Thus, it appears the hierarchy problem can be solved by adding a boson for every fermion, and a fermion for every boson. Noticing that there seems to be a symmetry between fermions and bosons, we can now start using the nomenclature of SUSY: for every SM particle there is a supersymmetric partner particle. The name given to the scalar (boson) partner for each fermion in the SM is often given by adding a "s" to the SM fermion's name, and the name given to the fermion partner for each boson is found by adding "ino" to the end of the SM boson's name. For example, the top quark's SUSY partner is often called the scalar top quark, top squark, or stop, and the SUSY partner for the gluon is often called the gluino. By following this idea of a symmetry between bosons and fermions through, SUSY can be formulated.

2.3.2 Supersymmetry Formalism

I will now provide a very brief overview of the SUSY formalism by trying to highlight important differences from the SM, but it is in no way complete. As an extension to special relativity, the fields in the SM satisfy the Poincaré space-time symmetry, such that the fields are invariant under Lorentz ($\Lambda^{\mu}{}_{\nu}$) and translation (ξ^{μ}) transformations

$$x^{\mu} \rightarrow x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu} + \xi^{\mu} = e^{i\omega_{ij}(M^{ij})_{\mu\nu}} x^{\nu} + e^{ib^{\mu}P_{\mu}}, \tag{2.38}$$

where ω_{ij} and b^{μ} are arbitrary parameters. $M_{\mu\nu}$ and P_{μ} are the generators of the group that satisfy the Poincaré algebra

$$\begin{aligned} [P_{\mu}, P_{\nu}] &= 0 \\ [M_{\mu\nu}, P_{\rho}] &= i(g_{\mu\rho}P_{\nu} - g_{\nu\rho}P_{\mu}) \\ [M_{\mu\nu}, M_{\rho\sigma}] &= i(g_{\mu\rho}M_{\nu\sigma} - g_{\mu\sigma}M_{\nu\rho} - g_{\nu\rho}M_{\mu\sigma} + g_{\nu\sigma}M_{\mu\rho}), \end{aligned} \tag{2.39}$$

where $g_{\mu\nu}$ is the Minkowski metric.

SUSY is a hypothetical symmetry between boson fields and fermion fields. It manifests as an extension to the Poincaré group, often called the super-Poincaré group, that is a space-time symmetry that assumes there is some new operator Q that transforms fields as

$$
\begin{aligned}
Q|\text{Boson}\rangle &= |\text{Fermion}\rangle \\
Q|\text{Fermion}\rangle &= |\text{Boson}\rangle.
\end{aligned}
\tag{2.40}
$$

The generators for Q are usually represented as Weyl spinors with 1/2 spin,[2] where the generators are denoted by Q_α $(\alpha = 1, 2)$ and their conjugates $Q^\dagger_{\dot{\alpha}} \equiv (Q_\alpha)^\dagger$ must satisfy an algebra of this form:

$$
\begin{aligned}
\{Q_\alpha, Q^\dagger_{\dot{\beta}}\} &= 2\sigma^\mu_{\alpha\dot{\beta}} P_\mu \\
\{Q_\alpha, Q_\beta\} &= \{Q^\dagger_{\dot{\alpha}}, Q^\dagger_{\dot{\beta}}\} = 0 \\
[P_\mu, Q_\alpha] &= [P_\mu, Q^\dagger_{\dot{\alpha}}] = 0 \\
[Q_\alpha, M_{\mu\nu}] &= (s_{\mu\nu})_\alpha{}^\beta Q_\beta \\
[Q^{\dagger\dot{\alpha}}, M_{\mu\nu}] &= (\overline{s}_{\mu\nu})^{\dot{\alpha}}{}_{\dot{\beta}} Q^{\dagger\dot{\beta}},
\end{aligned}
\tag{2.41}
$$

where $s^{\mu\nu} = \frac{i}{4}(\sigma^\mu\overline{\sigma}^\nu - \sigma^\nu\overline{\sigma}^\mu)$ and $\overline{s}^{\mu\nu} = \frac{i}{4}(\overline{\sigma}^\mu\sigma^\nu - \overline{\sigma}^\nu\sigma^\mu)$. The main takeaway is that we extend the usual Poincaré invariance of the SM and impose the super-Poincaré invariance. We will see that this is not the only "super" extension that is done in order for SUSY to be formulated.

In the formulation of SUSY, we extend the definition of fields that are functions of spacetime to superfields that are functions on a "superspace." The space that the usual fields depend on, the four-dimensional spacetime, is extended into this larger "superspace" with four extra anticommuting variables.[3] They are normally denoted

[2] A Weyl spinor can be defined as the two nonzero components of a four component left-handed (right-handed) Dirac spinor. The representation of the spinors used for the SM formulation in Section 2.2.2 used Dirac spinors, but for SUSY it is usually chosen to represent them as Weyl spinors.

[3] Variables that anticommute are often called Grassmann variables and in this case behave like Weyl spinors.

by θ_α ($\alpha = 1,2$) and their conjugates $\theta^\dagger_{\dot\alpha} \equiv (\theta_\alpha)^\dagger$. They must satisfy an algebra of this form:

$$\begin{aligned} \{\theta^\alpha, \theta^\beta\} = \{\theta^\dagger_{\dot\alpha}, \theta^\dagger_{\dot\beta}\} = \{\theta^\alpha, \theta^\dagger_{\dot\alpha}\} = 0 \\ [x^\mu, \theta^\alpha] = [x^\mu, \theta^\dagger_{\dot\alpha}] = 0. \end{aligned} \tag{2.42}$$

By taking advantage of the fermionic nature of θ that any combination of three or more θ's is zero, we can now write a general superfield as an expansion of these new θ and $\theta^\dagger$ variables:

$$\begin{aligned} \Phi(x, \theta, \theta^\dagger) = \phi(x) + \sqrt{2}\theta\xi(x) + \sqrt{2}\theta^\dagger\chi^\dagger(x) + \theta\theta F(x) + \theta^\dagger\theta^\dagger G(x) \\ + \theta\sigma^\mu\theta^\dagger A_\mu(x) + \theta\theta\theta^\dagger\lambda^\dagger(x) + \theta^\dagger\theta^\dagger\theta\kappa(x) + \frac{1}{2}\theta\theta\theta^\dagger\theta^\dagger D(x) \end{aligned} \tag{2.43}$$

where ϕ, F, G, and D are Lorentz-scalar functions, ξ, $\chi^\dagger$, κ, and $\lambda^\dagger$ are Weyl spinors, and A_μ is a vector. We can see the three types of fields (scalar, spinor, and vector) that the SM utilizes explicitly in a general superfield. We can also imagine how the observed particles and their superpartners can be grouped together in a superfield. Taking this one step further, a chiral (left-handed) superfield can be defined by requiring:

$$\mathcal{D}_\alpha \Phi_L = \left(\frac{\partial}{\partial\theta^\alpha} + i\sigma^\mu_{\alpha\dot\alpha}\theta^{\dagger\alpha}\partial_\mu \right) \Phi_L = 0. \tag{2.44}$$

This reduces a large number of free parameters in the general superfields. Following a similar argument, the chiral and vector fields can be written as

$$\Phi_L(y, \theta, \theta^\dagger) = \phi(y) + \sqrt{2}\theta\xi(y) + \theta\theta F(y)\ , \qquad y^\mu = x^\mu - i\theta\sigma^\mu\theta^\dagger \tag{2.45}$$

$$V(x, \theta, \theta^\dagger) = \theta\sigma^\mu\theta^\dagger A_\mu(x) + \theta\theta\theta^\dagger\lambda^\dagger(x) + \theta^\dagger\theta^\dagger\theta\lambda(x) + \frac{1}{2}\theta\theta\theta^\dagger\theta^\dagger D(x). \tag{2.46}$$

All fermions, sfermions, and Higgs fields come from left-handed chiral superfields, while gauge fields come from supervector fields.

There is one final piece that is needed in order to develop a realistic SUSY theory. It turns out, by construction, that the masses of each SM particle and their SUSY partners will be identical. This is a problem because no SUSY particle has been observed in nature. Either the universe does not obey SUSY, or the masses of these new particles must be much heavier than their SM partners. We can increase the mass of the SUSY particles by adding extra terms to the Lagrangian, so-called "soft" terms, by hand, which breaks SUSY. One issue with this is that if the masses of the superpartners are too large then the Higgs' mass hierarchy problem will be reintroduced.

2.3.3 Minimal Supersymmetric Standard Model

The Minimal Supersymmetric Standard Model (MSSM) is a SUSY-based extension to the SM that assumes the minimum possible number of fields. Table 2.2 shows details of all superfields included in the MSSM. In short, the fields of the SM are promoted to be superfields. The components of these fields can be broken up into the SM fields and their SUSY partner fields. One exception, however, is that a second Higgs superfield must be added. Unlike the SM, we must introduce two Higgs doublets because in SUSY, triangular anomalies will appear and cause higher-order diagrams with left-handed fermions to be divergent. This divergence is avoided by introducing two Higgs doublets with opposite hypercharge.

We can write the Lagrangian of the MSSM as two parts. There is one that conserves SUSY ($\mathcal{L}_{\text{SUSY}}$), and one that breaks SUSY ($\mathcal{L}_{\text{Soft}}$), such that the mass of the superpartners are much higher than their SM counterparts. This gives the total

Table 2.2: Summary of MSSM fields. The subscripts i, a, and b used for different fields take the following values: $i \in \{1,2,3\}$, $a \in \{1,2,3,4,5,6,7,8\}$, and $b \in \{1,2,3\}$.

Superfield	Component Fields			Spin	Rep.	Name
Matter Fields	Generation					
	I	II	III			
$\hat{Q}_i$	$\begin{pmatrix} u_L \\ d_L \end{pmatrix}$	$\begin{pmatrix} c_L \\ s_L \end{pmatrix}$	$\begin{pmatrix} t_L \\ b_L \end{pmatrix}$	$\frac{1}{2}$	$(\mathbf{3},\mathbf{2},\frac{1}{3})$	Quarks
	$\begin{pmatrix} \tilde{u}_L \\ \tilde{d}_L \end{pmatrix}$	$\begin{pmatrix} \tilde{c}_L \\ \tilde{s}_L \end{pmatrix}$	$\begin{pmatrix} \tilde{t}_L \\ \tilde{b}_L \end{pmatrix}$	0		Squarks
$\hat{u}^c_i$	$u^\dagger_R$	$c^\dagger_R$	$t^\dagger_R$	$\frac{1}{2}$	$(\overline{\mathbf{3}},\mathbf{1},-\frac{4}{3})$	Quarks
	$\tilde{u}^*_R$	$\tilde{c}^*_R$	$\tilde{t}^*_R$	0		Squarks
$\hat{d}^c_i$	$d^\dagger_R$	$s^\dagger_R$	$b^\dagger_R$	$\frac{1}{2}$	$(\overline{\mathbf{3}},\mathbf{1},\frac{2}{3})$	Quarks
	$\tilde{d}^*_R$	$\tilde{s}^*_R$	$\tilde{b}^*_R$	0		Squarks
$\hat{L}_i$	$\begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}$	$\begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}$	$\begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix}$	$\frac{1}{2}$	$(\mathbf{1},\mathbf{2},-1)$	Leptons
	$\begin{pmatrix} \tilde{\nu}_{eL} \\ \tilde{e}_L \end{pmatrix}$	$\begin{pmatrix} \tilde{\nu}_{\mu L} \\ \tilde{\mu}_L \end{pmatrix}$	$\begin{pmatrix} \tilde{\nu}_{\tau L} \\ \tilde{\tau}_L \end{pmatrix}$	0		Sleptons
$\hat{e}^c_i$	$e^\dagger_R$	$\mu^\dagger_R$	$\tau^\dagger_R$	$\frac{1}{2}$	$(\mathbf{1},\mathbf{1},2)$	Antileptons
	$\tilde{e}^*_R$	$\tilde{\mu}^*_R$	$\tilde{\tau}^*_R$	0		Antisleptons
Gauge Fields						
$\hat{G}_a$		G^μ_a		1	$(\mathbf{8},\mathbf{1},0)$	Gluon
		$\tilde{G}^\mu_a$		$\frac{1}{2}$		Gluino
$\hat{W}_b$		W^μ_b		1	$(\mathbf{1},\mathbf{3},0)$	Bosons
		$\tilde{W}^\mu_b$		$\frac{1}{2}$		Winos
$\hat{B}$		B^μ		1	$(\mathbf{1},\mathbf{1},0)$	Bosons
		$\tilde{B}^\mu$		$\frac{1}{2}$		Bino
Higgs Fields						
$\hat{H}_u$		$\begin{pmatrix} H^+_u \\ H^0_u \end{pmatrix}$		0	$(\mathbf{1},\mathbf{2},1)$	Higgs
		$\begin{pmatrix} \tilde{H}^+_u \\ \tilde{H}^0_u \end{pmatrix}$		$\frac{1}{2}$		Higgsino
$\hat{H}_d$		$\begin{pmatrix} H^0_d \\ H^-_d \end{pmatrix}$		0	$(\mathbf{1},\mathbf{2},-1)$	Higgs
		$\begin{pmatrix} \tilde{H}^0_d \\ \tilde{H}^-_d \end{pmatrix}$		$\frac{1}{2}$		Higgsino

Lagrangian as:

$$\mathcal{L}_{\text{MSSM}} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{Soft}}. \tag{2.47}$$

As mentioned before, SUSY assumes an extension of spacetime to superspace where four Grassmann variables have been added. This also implies that we must integrate over these new variables when defining the SUSY action. We do this as we write out the Lagrangian to be able to compare the SUSY Lagrangian with the SM. I will not write out the appropriate rules of integration for Grassmann variables, because we will not perform the integration explicitly. The SUSY-conserving part can be written as

$$\begin{aligned}
\mathcal{L}_{\text{SUSY}} = & \underbrace{\int d^2\theta \frac{1}{4}\left(2Tr(\mathbf{W}_s^\alpha \mathbf{W}_{s\alpha} + 2Tr(\mathbf{W}^\alpha \mathbf{W}_\alpha) + W^\alpha W_\alpha)\right) + \text{h.c.}}_{\text{Gauge boson and gaugino kinetic terms}} \\
& + \underbrace{\sum_i \int d^2\theta d^2\theta^\dagger \; \hat{Q}_i^\dagger e^{\left(g_s \lambda^a \hat{G}_a + g\sigma^b \hat{W}_b + g' Y_Q \hat{B}\right)} \hat{Q}_i}_{\text{Left-handed quarks/squarks kinetic terms}} \\
& + \underbrace{\sum_i \int d^2\theta d^2\theta^\dagger \left(\hat{u}_i^{c\dagger} e^{\left(-g_s \lambda^{a*} \hat{G}_a + g' Y_u \hat{B}\right)} \hat{u}_i^c + \hat{d}_i^{c\dagger} e^{\left(-g_s \lambda^{a*} \hat{G}_a + g' Y_d \hat{B}\right)} \hat{d}_i^c\right)}_{\text{Right-handed quarks/squarks kinetic terms}} \\
& + \underbrace{\sum_i \int d^2\theta d^2\theta^\dagger \left(\hat{L}_i^\dagger e^{\left(g\sigma^b \hat{W}_b + g' Y_L \hat{B}\right)} \hat{L}_i + \hat{e}_i^{c\dagger} e^{\left(g' Y_e \hat{B}\right)} \hat{e}_i^c\right)}_{\text{Lepton/sleptons kinetic terms}} \\
& + \underbrace{\int d^2\theta d^2\theta^\dagger \left(\hat{H}_u^\dagger e^{\left(2g\sigma^b \hat{W}_b + g' Y_{H_u} \hat{B}\right)} \hat{H}_u + \hat{H}_d^\dagger e^{\left(2g\sigma^b \hat{W}_b + g' Y_{H_d} \hat{B}\right)} \hat{H}_d\right)}_{\text{Higgs kinetic terms}} \\
& + \underbrace{\int d^2\theta \mathcal{W} + \text{h.c.}}_{\text{Superpotential Term}}
\end{aligned} \tag{2.48}$$

where the terms that do not appear in Table 2.2 are defined as

$$\mathbf{W}_{s\alpha} = -\frac{1}{4}\mathcal{D}^\dagger \mathcal{D}^\dagger e^{-\frac{1}{2}\lambda^a \hat{G}_a} \mathcal{D}_\alpha e^{\frac{1}{2}\lambda^a \hat{G}_a} \tag{2.49}$$

$$\mathbf{W}_\alpha = -\frac{1}{4}\mathcal{D}^\dagger\mathcal{D}^\dagger e^{-\frac{1}{2}\sigma^b\hat{W}_b}\mathcal{D}_\alpha e^{\frac{1}{2}\sigma^b\hat{W}_b} \tag{2.50}$$

$$W_\alpha = -\frac{1}{4}\mathcal{D}^\dagger\mathcal{D}^\dagger\mathcal{D}_\alpha\hat{B} \tag{2.51}$$

$$\mathcal{W} = \sum_{ij}\left(\lambda^u_{ij}\hat{Q}_i\hat{H}_u\hat{u}^c_j - \lambda^d_{ij}\hat{Q}_i\hat{H}_d\hat{d}^c_j - \lambda^e_{ij}\hat{L}\hat{H}_d\hat{e}^c_j\right) + \mu\hat{H}_u\hat{H}_d. \tag{2.52}$$

Similarly to building the SM Lagrangian, we add in the SUSY equivalent kinetic terms for all superfields, Higgs Yukawa terms, and Higgs potential. Imposing local $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ gauge invariance is done by including the vector superfield terms in the exponential of the kinetic terms. The potential term is usually referred to as the "superpotential." The superpotential, Eq. (2.52), is where the SUSY-equivalent Yukawa and Higgs potential terms are defined. We will see that it is possible to add terms to the superpotential in Sections 2.3.4 and 2.3.5 to extend beyond the MSSM.

The soft SUSY breaking component of the Lagrangian is:

$$\begin{aligned}\mathcal{L}_{\text{Soft}} = & \underbrace{-\frac{1}{2}\left(M_3\sum_{a=1}^{8}\tilde{G}_a\tilde{G}_a + M_2\sum_{b=1}^{3}\tilde{W}_b\tilde{W}_b + M_1\tilde{B}\tilde{B} + \text{h.c.}\right)}_{\text{Gaugino Mass Terms}} \\ & \underbrace{-\sum_{i=3}^{3}\left(m^2_{\tilde{Q}_i}\tilde{Q}^\dagger_i\tilde{Q}_i + m^2_{\tilde{u}_{Ri}}|\tilde{u}_{Ri}|^2 + m^2_{\tilde{d}_{Ri}}|\tilde{d}_{Ri}|^2 + m^2_{\tilde{L}_i}\tilde{L}^\dagger_i\tilde{L}_i + m^2_{\tilde{e}_{Ri}}|\tilde{e}_{Ri}|^2\right)}_{\text{Sfermion Mass Terms}} \\ & \underbrace{-m^2_{H_u}H^\dagger_uH_u - m^2_{H_d}H^\dagger_dH_d - B\mu(H_uH_d + \text{h.c.})}_{\text{Higgs Mass and Bilinear Terms}} \\ & \underbrace{-\sum_{ij}\left(A^u_{ij}\lambda^u_{ij}\tilde{Q}_iH_u\tilde{u}_{Rj} - A^d_{ij}\lambda^d_{ij}\tilde{Q}_iH_d\tilde{d}_{Rj} - A^e_{ij}\lambda^e_{ij}\tilde{L}_iH_d\tilde{e}_{Rj} + \text{h.c.}\right)}_{\text{Higgs and Sfermions Trilinear Terms}}.\end{aligned} \tag{2.53}$$

This includes soft mass terms for all of the SUSY particles, an analog of the μ term for the Higgs potential (bilinear) with extra constant B, and an analog for the Higgs Yukawa terms (trilinear) with extra constants A_{ij}. Unlike the SM potential, we do not need to assume any sign or shape because imposing SUSY breaking triggers electroweak spontaneous symmetry breaking with the appropriate sign for μ^2. I will not go into detail on the Higgs mechanism for SUSY, which gives similar results as the Higgs mechanism in the SM, with the exception that there are now five Higgs bosons where two should have electric charge. The soft SUSY terms add $\mathcal{O}(100)$ extra free parameters on top of the SM, which makes it less than ideal. This adds to the list of parameters that must be measured. One bright side to this long list of free parameters is that measuring these parameters will keep particle physicists busy, most importantly employed, for decades if not hundreds of years to come.

It is important to point out that local gauge invariance allows for more terms to be added to the superpotential, but most physicists impose an extra discrete symmetry that removes them. This symmetry is called R-parity and its quantum number can be calculated as:

$$P_R = (-1)^{3B-L+2s} \tag{2.54}$$

where B corresponds to baryon number, L corresponds to lepton number, and s corresponds to spin. All SM particles have $P_R = 1$, while all SUSY particles have $P_R = -1$. All vertices must conserve this number, which implies that a SUSY particle can never decay to SM particles only. The motivation for assuming this symmetry is to stabilize the MSSM. Without it, these extra terms allowed by gauge invariance would introduce lepton and baryon number violation and lead to proton decay. Since lepton and baryon number violation has never been observed, the MSSM must either conserve it or break it softly. Similarly, no evidence for proton decay has been observed. For this thesis we will actually consider a model that does not conserve R-parity.

The cross sections for the production of various SUSY particles expected at the LHC are shown in Fig. 2.8. For the purpose of this thesis, we are only interested in the red curve, which corresponds to top squark production. Other groups look for the higher cross section options, such as gluino production, and they have placed stringent limits on their allowable cross sections. In general, SUSY groups have looked for gluino or top squark production first because of a combination of high cross section and low SM backgrounds.

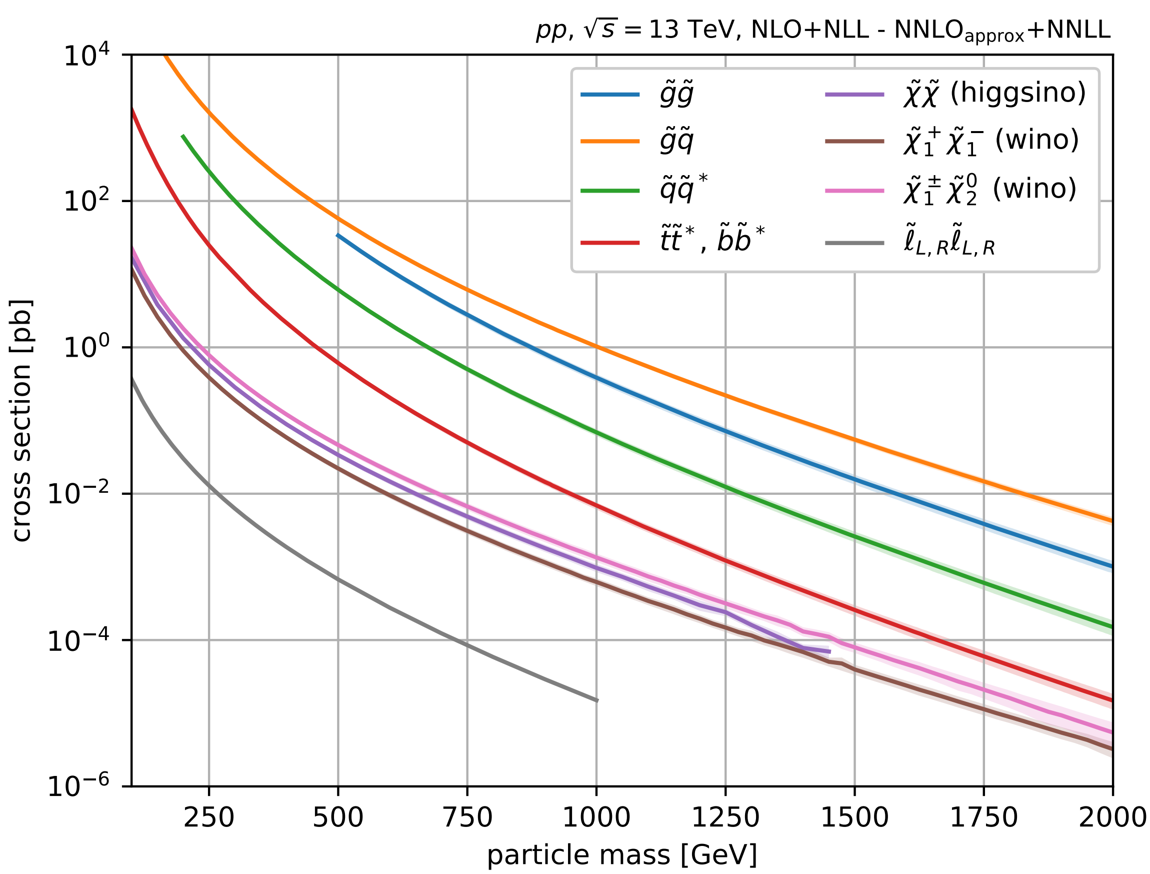


Figure 2.8: SUSY cross sections as a function of particle mass.

For simplicity, most searches for the MSSM assume a simplified version where only a few of the interactions are strong enough to be seen. For this thesis we only allow for terms that produce the partner of the top quark, the scalar top quark ("stop"), and its decay to the lightest neutral gaugino, the neutralino, which is often called the lightest supersymmetric particle (LSP). Figure 2.9 illustrates the assumed dominant interactions for this thesis.

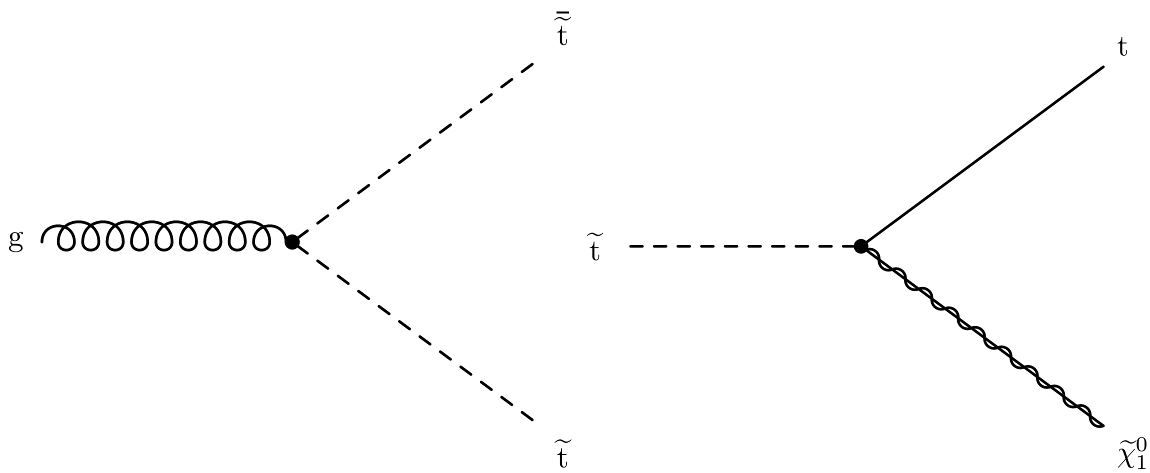


Figure 2.9: Relevant vertices from the MSSM for this thesis. (Left) Vertex needed for strong scalar top quark production at the LHC. (Right) Vertex needed for a scalar top quark decaying into a top quark and neutralino.

Some physicists make the argument that the theory of everything should not have hundreds of free parameters that need to be measured. However, the MSSM is still a very powerful theory, and this leads physicists to think that if it does exist it must manifest from some more inclusive theory. Many searches for the MSSM have been performed, but no evidence for it has been observed. This fact leaves us with three options: the MSSM is just out of reach and if we keep looking with more data we will find it, the MSSM is too simple and there are extra fields or interactions that we should be looking for, or all forms of SUSY are not physical. For this thesis we will consider the second option and make small extensions to the MSSM that will open up currently unexplored regions of phase space.

2.3.4 R-Parity Violating SUSY

One reasonable extension to the MSSM that can be performed is to relax the R-parity requirement that was added by hand. This type of SUSY is called R-parity violating (RPV) SUSY. This allows for the addition of extra terms to the superpotential that still allow for the usual local gauge invariance:

$$\mathcal{W}_{\Delta L=1} = \frac{1}{2}\lambda^{ijk}\hat{L}_i\hat{L}_j\hat{e}^c_k + \lambda'^{ijk}\hat{L}_i\hat{Q}_j\hat{d}^c_k + \mu'^i\hat{L}_i\hat{H}_u \tag{2.55}$$

$$\mathcal{W}_{\Delta B=1} = \frac{1}{2}\lambda''^{ijk}\hat{u}^c_i\hat{d}^c_j\hat{d}^c_k. \tag{2.56}$$

As stated before, these terms cause lepton and baryon number violation respectively. With this in mind, the couplings for these terms must be very small to match current observations. For the purpose of this thesis we will consider only the baryon-number-violation term. In general, both the baryon-number-violating terms and lepton-number-violating terms are needed in order to have proton decay, so allowing for the baryon-number-violating term only with an appropriate coupling is consistent with current measurements. We will refer to this RPV by the UDD term, shown in Eq. (2.56), since it allows for interactions with a right-handed up and two right-handed down type quarks or squarks. The UDD term introduces interactions of this type, as shown in Fig. 2.10.

With this UDD term, the MSSM requirement that all SUSY particle decays must produce a SUSY particle is no longer true. However, all interactions in RPV SUSY still conserve R-parity, except for when it comes to the UDD vertex, which is

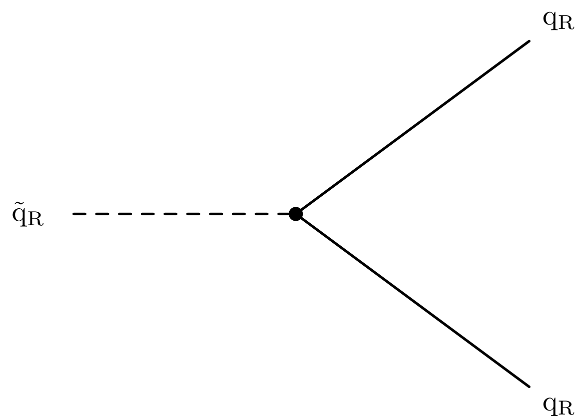


Figure 2.10: Vertex from the UDD superpotential that allows for a right-handed scalar quark to decay into two right-handed quarks.

assumed to have a weak coupling. This interaction only allows for the LSP to decay to SM particles. Normally in the MSSM, the LSP is stable and leaves detectors unseen, but now it has a means to decay. This means most physicists look for signatures that have a large amount of missing energy because the LSP is stable, but now we can look for SUSY with signatures that have small amounts of missing energy. These types of signatures have been mostly ignored in the past. This is why there are effectively no limits placed on the top squark mass when considering a SUSY theory with the MSSM plus this UDD term. For simplicity we will consider a simplified version of the MSSM described earlier. This means we will only allow for interactions summarized in Figs. 2.9 and 2.10. Finally, it is worth noting that this extra RPV term does not contribute significantly to top squark production, so the top squark production cross section is identical to what the MSSM predicts.

2.3.5 Stealth SYY SUSY

For this thesis, the other extension to the MSSM that we consider, stealth SYY, adds new superfields $\hat{S}$ and $\hat{Y}$ to the MSSM. These fields are considered to be in a "hidden" sector, separate from the SM and SUSY sectors described by the MSSM. This theory actually imposes the $SU(5)$ local gauge invariance rather than the usual $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. The SM or MSSM can be embedded in a $SU(5)$ local gauge invariance rather than the usual one. It also adds in a superfield that corresponds to gravity, which can be considered to be mediated by a spin-2 particle, the graviton, and its superpartner, the gravatino.

A chiral singlet superfield ($\hat{S}$) is added to this hidden sector that contains a scalar singlet (S) and its SUSY partner singlino ($\tilde{S}$). Unlike all the fields in the MSSM, we assume that the masses are almost degenerate, which means that SUSY is assumed to be mostly conserved (a nonzero mass difference is needed for the singlino to decay). This field is not charged, interacting only with the MSSM fields via superpotential couplings to vector-like fields. Physicists refer to interactions via an intermediate vector-like field to be the hidden "portal" to the SM and SUSY sectors. This nature of the singlet and singlino is why the name of this theory is stealth SUSY, and why it is considered to be in its own sector. The other chiral superfields that are added are the $\hat{Y}$ and $\hat{\overline{Y}}$ pair that transform as $5+\overline{5}$ under $SU(5)_{\text{GUT}}$. The main takeaway is that these fields are charged and interact with particles in the MSSM. The superpotential that allows for the singlet and singlinos to interact with MSSM particles via $\hat{Y}$ and $\hat{\overline{Y}}$ is defined as

$$\mathcal{W}_{SYY} = \frac{1}{2}m\hat{S}^2 + \lambda\hat{S}\hat{Y}\hat{\overline{Y}} + m_Y\hat{Y}\hat{\overline{Y}}. \tag{2.57}$$

We can now see why we consider the name of the theory to be stealth SYY, since there is a stealth sector that only interacts with the MSSM through the $\hat{S}\hat{Y}\hat{\overline{Y}}$ term in the superpotential. This is done through loop diagrams that are shown in Fig. 2.11.

This theory still conserves R-parity, and gives a signature with a small amount of missing energy since the gravatino is assumed to have a low mass.

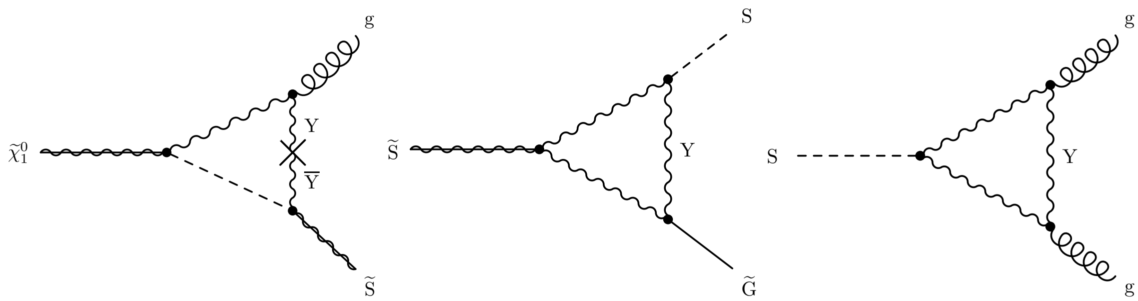


Figure 2.11: Effective vertices coming from Y and $\overline{Y}$ loops. (Left) Effective vertex allowing the neutralino to decay. (Middle) Effective vertex allowing the singlino to decay. (Right) Effective vertex allowing the singlet to decay.

These extra fields, or sector, that are added to the MSSM for this theory are not considered to be strong enough to contribute significantly to top squark production. This means the MSSM top squark cross section calculations can be used. These terms only contribute to the decay of the top squark, which effectively allows for the top squark to decay to particles that can be completely observed in a particle detector. The most important thing to consider is that similarly to the RPV SUSY model described earlier, this model will produce a signature that has not been fully explored by past searches, and effectively has no limits on the possible top squark masses.

CHAPTER THREE

Experimental Apparatus

> "An experiment is a question which science poses to Nature and a measurement is the recording of Nature's answer."
>
> —Max Planck (1858–1947)

3.1 Introduction

The experimental apparatus used for this thesis, in short, is one of the most complicated machines ever created. We process data from proton-proton collisions produced at the Large Hadron Collider (LHC) and recorded by the Compact Muon Solenoid (CMS) detector. The basic principle is to accelerate two sets of protons as fast as possible, traveling clockwise and counterclockwise around a ring, and then collide them together. It is the product of these collisions that we study, and where we hope to see signs of new and interesting physical phenomena.

3.2 The Large Hadron Collider

The LHC is the largest and most powerful particle accelerator ever created. The LHC is designed to collide protons at a center-of-mass energy (COM) of up to 14 TeV, the highest energy of any accelerator ever constructed. It is located at the European Organization for Nuclear Research (CERN) in Geneva, Switzerland. The accelerator itself spans the border between France and Switzerland as shown in Fig. 3.1, and it is located in the 26.7 km tunnel originally built for the Large Electron-Positron Collider (LEP) in the late 1980's. The LHC is made up of eight straight sections and eight arcs, where the straight sections facilitate the crossings of the clockwise and counterclockwise beams. It is these high energy beam crossings where proton-proton collisions occur. The LHC passes through four main detectors located around the ring. There are two general purpose detectors, the Compact Muon Solenoid (CMS) and A Toroidal LHC ApparatuS (ATLAS), and two detectors designed for more specific purposes, A Large Ion Collider Experiment (ALICE) and LHC-beauty (LHCb). Most of the details and figures in this section come from Ref. [14].

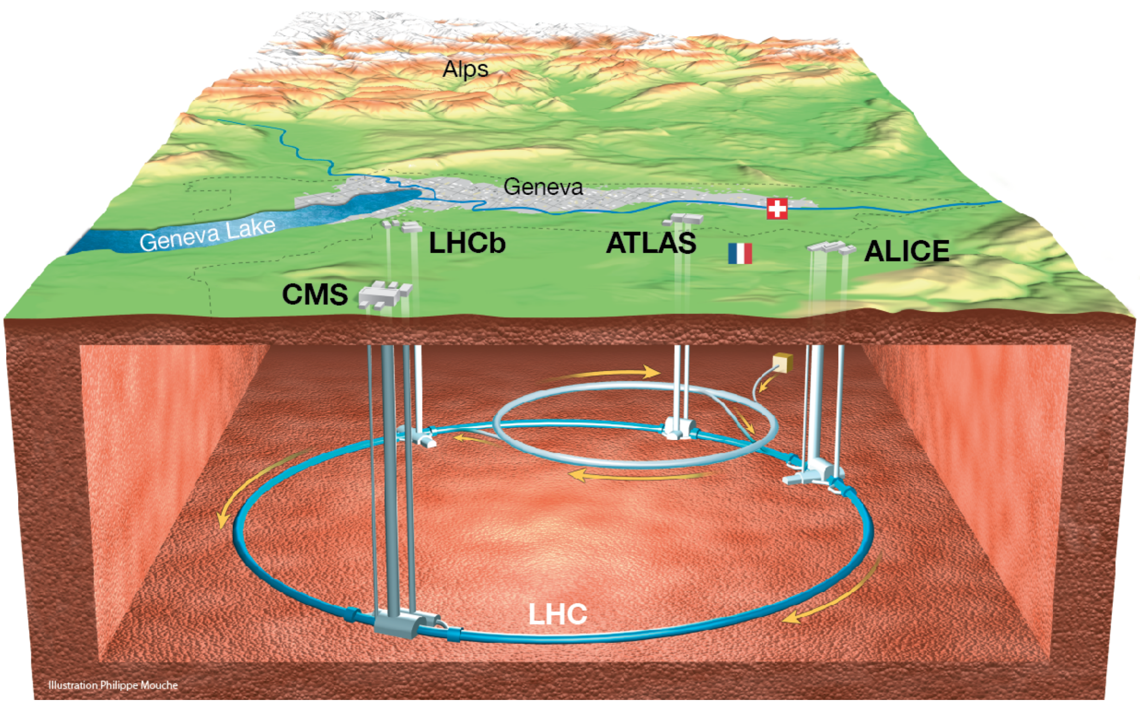


Figure 3.1: A diagram of the LHC, showing the location of the tunnel with respect to the French-Swiss border, Geneva, Switzerland, and the Swiss Alps.

3.2.1 *The Journey of a Proton into the LHC*

In order for a proton to generate collisions in the LHC, it must first pass through a few stages as shown in Fig. 3.2. The process begins in a bottle of hydrogen gas, where hydrogen atoms are ionized to produce free protons. These protons are then directed to the linear accelerator (LINAC). The LINAC is the first step in the accelerator chain, increasing the energy of the protons to 50 MeV. Next, the proton synchrotron booster, a small circular accelerator (relative to the LHC), boosts the protons to about 1.4 GeV. The booster then feeds the protons to the proton synchrotron (PS), which accelerates them to about 25 GeV and starts "bunching" them together.

It is important for the PS to bunch protons into densely packed groups, and it is these groups of protons that are steered to collide with each other in the LHC. When these bunches of protons collide it is called a bunch crossing. The denser the protons in each bunch, the higher the probability of two protons interacting when two bunches cross. These bunches are spaced every 25 nanosecond (ns) in the beam to ensure that there are no proton collisions at times other than during the bunch crossing. All detectors have very sensitive electronics that need to know exactly when two protons can collide in order to measure the protons' interaction properly.

After exiting the PS, the proton bunches are passed into the super proton synchrotron (SPS), which boosts the protons to about 450 GeV. Finally, after the SPS, the protons make their way into the LHC, where the protons circulate clockwise and counterclockwise with an energy of 7 TeV.

3.2.2 Performance of the LHC

While the main purpose of the LHC is to deliver proton-proton collisions, the LHC can be thought of as an event producing machine. An event is all information that can be saved during a proton-proton collision. However, not every event has

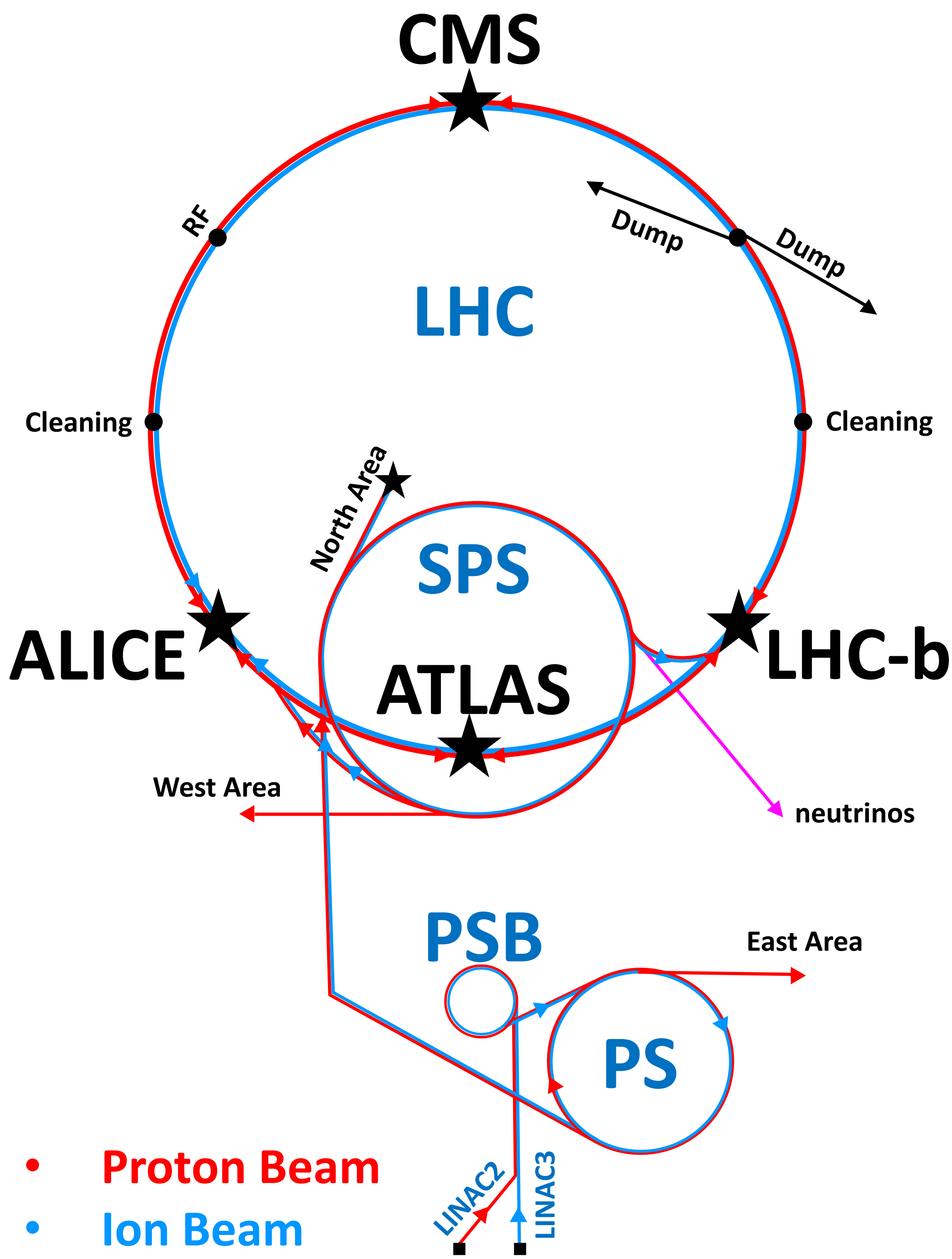


Figure 3.2: A schematic diagram of the CERN accelerator chain including the LINAC, PS Booster (PSB), PS, SPS, and LHC. The locations of the four major detectors positioned around the LHC are also shown.

something interesting—in fact, over 99.9975% of the events produced by the LHC are never saved to be processed later. The LHC was designed to maximize the number of interesting events by maximizing the COM and the instantaneous luminosity (L).

For any particular physical process produced in the LHC, we can calculate the number of events per second we expect for this process by using

$$\frac{dN_{event}}{dt} = L\,\sigma_{event}, \tag{3.1}$$

where L is the instantaneous luminosity and σ_{event} is the cross section for a particular process. The cross section is related to the probability of a specific process to occur. It can depend on the proton beam energy, and how complicated the physics of the particular event you are interested in. At the LHC, events that produce two gluons occur much more often than events that produce a Higgs boson, and one easy way to see this is to compare the cross sections of both processes. However, the instantaneous luminosity, Eq. (3.2), depends only on the proton beam parameters. For a Gaussian distribution of protons in each bunch of a proton beam, the instantaneous luminosity is given by

$$L = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi\epsilon_n \beta^*} F \tag{3.2}$$

where N_b is the number of particles per bunch, n_b is the number of bunches per proton beam, f_{rev} is the revolution frequency, γ_r is the relativistic gamma factor, ϵ_n is the normalized transverse proton beam emittance, β^* is the so-called beta function at the collision point, and F is the geometric luminosity reduction factor due to the crossing angle at the interaction point (IP). F is defined as

$$F = \left[1 + \left(\frac{\theta_c \sigma_z}{2\sigma_x^*}\right)^2\right]^{-\frac{1}{2}} \tag{3.3}$$

where θ_c is the full crossing angle at the IP, σ_z is the standard deviation of the bunch length, and σ_x^* is the transverse standard deviation of the bunch size at the IP. Some of these variables are shown schematically in Fig. 3.3, which shows a proton bunch crossing. It is assumed that the two bunches of protons are identical and spherical in shape.

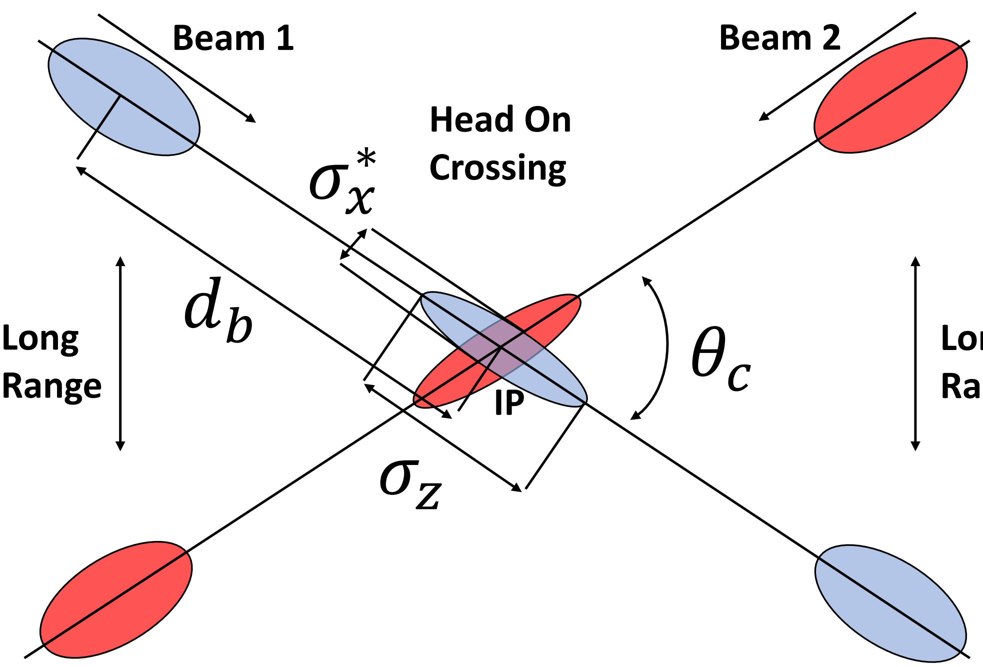


Figure 3.3: A diagram showing two bunches of protons crossing. They are modeled as two identical Gaussian distributions with σ_x^* and σ_z as the standard deviation along the transverse and longitudinal directions, and θ_c as the full crossing angle with respect to the IP.

Given the formula for the instantaneous luminosity, Eq. (3.1) can be used to calculate the total number of expected events over a certain amount of time

$$N = \int \frac{dN_{event}}{dt}\,dt = \sigma_{event} L_{tot}, \tag{3.4}$$

where the total integrated luminosity (L_{tot}) is defined as:

$$L_{tot} = \int L\,dt. \tag{3.5}$$

Thus, in general, the total number of events expected for a particular interesting physics process is just the cross section for that process times the total integrated luminosity of the corresponding data sample. Cross section has units of area, which implies that total integrated luminosity has units of inverse area. The total integrated luminosity is often quoted as a measure of the amount of data collected, since it is not specific to a particular analysis. When looking at the total integrated luminosity it is important to keep in mind that there are two possible numbers that can be quoted, the total "delivered" and "recorded" luminosity. The total delivered luminosity measures the amount of possible data delivered to CMS by the LHC, while recorded luminosity is the amount of data that CMS recorded while fully functional. The data used for this thesis were collected during 2016–2018 of LHC Run 2 at a center-of-mass energy of 13 TeV, and correspond to the total recorded integrated luminosity shown in Fig. 3.4.

3.3 The Compact Muon Solenoid

CMS is a general purpose detector located on the LHC near Cessy, France, between the Jura Mountains and Lac Léman (Lake Geneva), across the ring from the main CERN site in Meyrin, Switzerland. It fits snuggly in a massive cavern located 100 m underground. CMS is assembled from many subdetectors that each provide unique and useful information about the particles that are produced when protons are collided. It possesses, as its name suggests, a large superconducting solenoid magnet that produces a constant 3.8 Tesla (T) magnetic field within the solenoid and quickly drops off outside the solenoid. The magnetic field is critical for determining the momenta of charged particles, as the particles' trajectories bend in the field. The design of CMS and its subdetectors was based on the geometry of the magnet. As

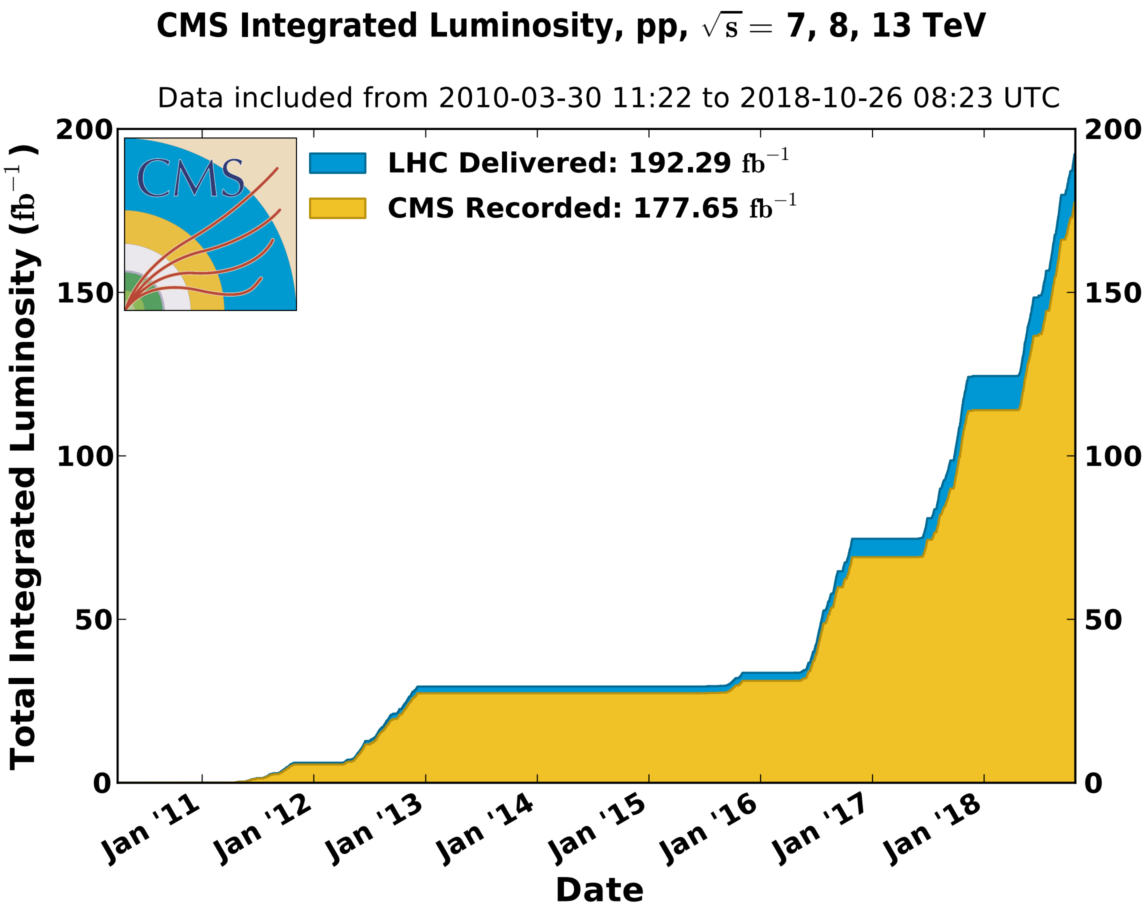


Figure 3.4: Total delivered and recorded luminosity for CMS from the turn-on of the LHC to the end of 2018. Most of the integrated luminosity at a center-of-mass energy of 13 TeV came during the 2016–2018 running period.

shown in Fig. 3.5, there are subdetectors inside and outside the magnet. Inside, and closest to the beam pipe, is the silicon tracker consisting of silicon pixels and strips, followed radially outward by the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). Outside the magnet there are three muon systems. The barrel region consists of aluminum drift tubes (DT) and resistive plate chambers (RPC). The endcaps are intermixed with cathode strips chambers (CSC) and more RPCs.

In general, CMS is designed to measure the position and energy of all particles that come from the collision. The interactions of particles are different for each subdetector, as shown in Fig. 3.6. For example, some subdetectors are designed to minimally interact with the particles, while some are designed to stop and absorb them completely. During the data-taking period, the CMS detector was undergoing the Phase 1 upgrade. After taking data in 2016, part of the tracking system was replaced, and after taking data in 2017, parts of the HCAL readout system were replaced. The original electronics are referred to as the legacy system, while the new electronics are referred to as the Phase 1 system. The subdetectors are discussed in their respective sections below. The information about CMS comes primarily from references [14, 15].

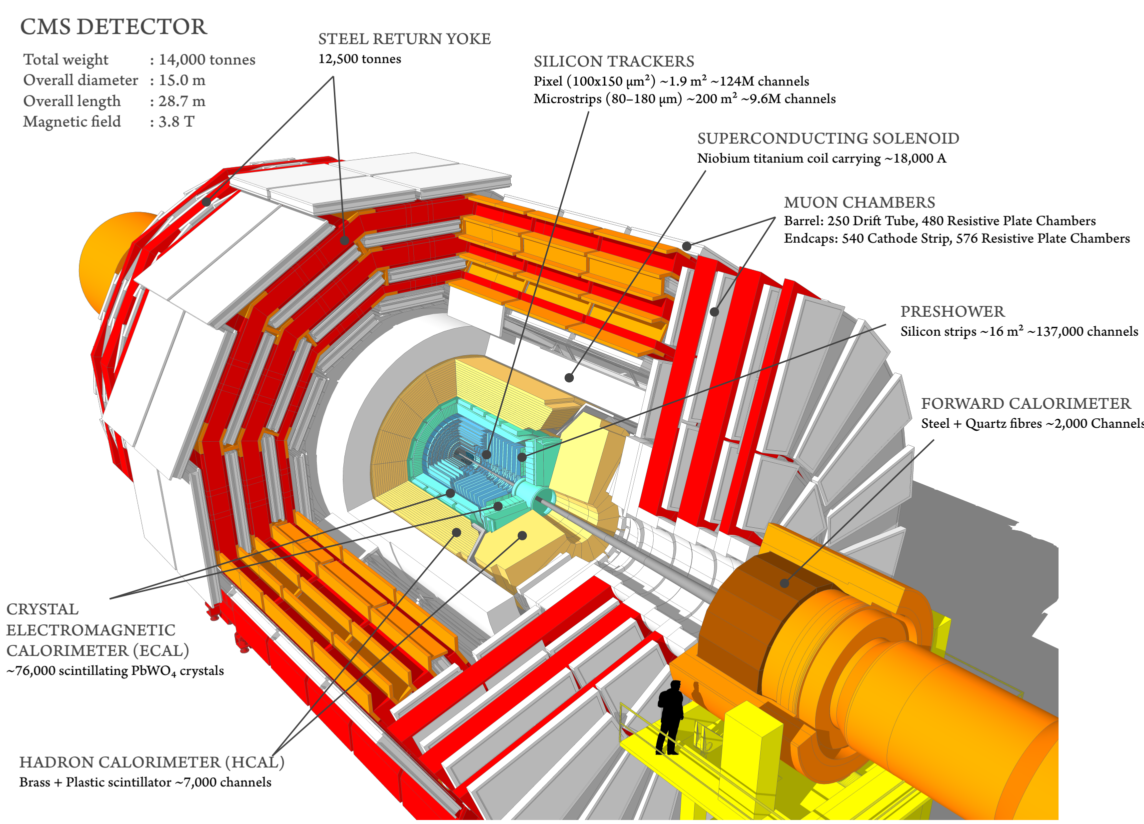


Figure 3.5: Cutaway view of the CMS detector. Each subdetector is listed in the figure, and there is a person near the front for scale.

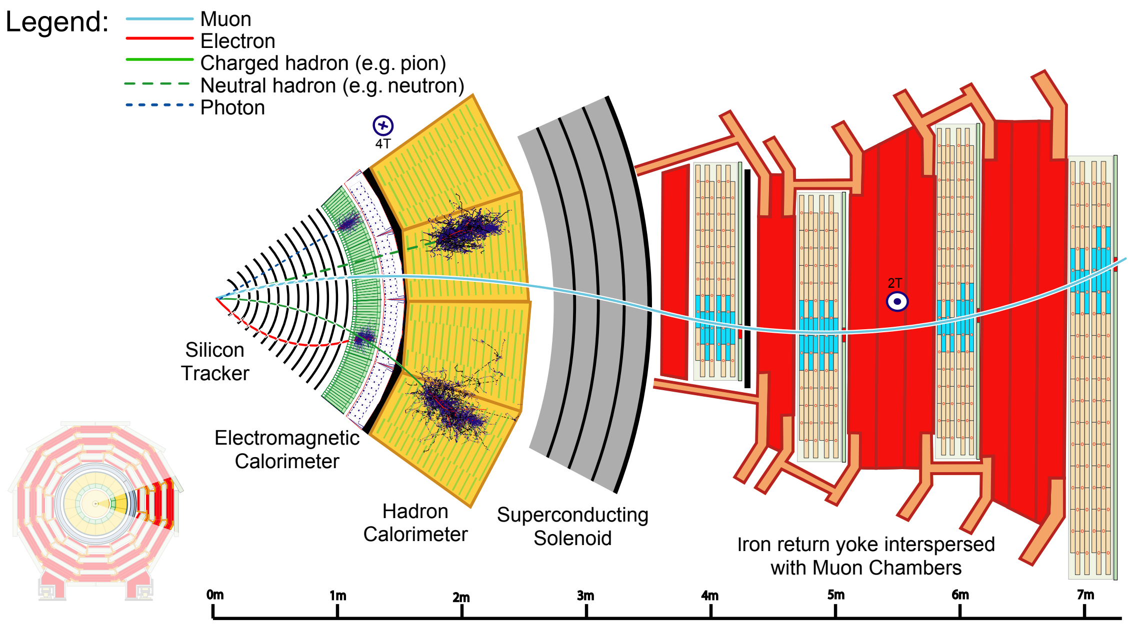


Figure 3.6: Transverse view of the CMS detector showing how different particles interact with different subdetectors. Muons leave a curved track in the tracker, interact minimally in the ECAL and HCAL, and then deposit hits in the muon systems before leaving the detector. Electrons leave a curved track in the tracker, and deposit their energy in the ECAL. Charged hadrons leave a curved track in the tracker, and deposit their energy in the HCAL. Neutral hadrons only deposit their energy in the HCAL. Photons only deposit their energy in the ECAL.

3.3.1 Detector Geometry and Basic Variables

Similar to the detectors used at the Fermilab Tevatron and many other collider experiments, CMS has a cylindrical shape consisting of a barrel piece and two endcaps. It has a length of 21 m, a diameter of 15 m, and weighs about 14,000 tonnes. We define the coordinate system such that the x-axis points radially toward the center of the LHC, the y-axis points vertically upward, and the z-axis points along the beam pipe. The center of the detector, also referred as the interaction point (IP), is at the origin of the coordinate system. The positive z direction points toward the Jura Mountains in the direction of the counterclockwise LHC beam. It is convenient to work in a coordinate system in which the usual azimuthal angle ϕ is measured from the x-axis in the (x, y) plane, and the polar angle θ is measured from the $+z$-axis. The polar angle θ can be mapped to a quantity known as the pseudorapidity (η) defined as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]. \tag{3.6}$$

For massless particles (e.g. photons), pseudorapidity is equal to rapidity, i.e. a quantity that defines a Lorentz boost along the z-axis. The geometry of CMS is designed such that particles up to $|\eta| \sim 5$ can be detected. From the mapping of η to θ, $\theta = \pi/2$ (for a particle with no z component of momentum) corresponds to $\eta = 0$. Similarly, as $\theta \to 0$ or π (closer to the beam pipe), $\eta \to \pm\infty$.

We also define useful variables in terms of η. One of them is the Lorentz-invariant angular separation (ΔR) between two particles:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{3.7}$$

Having a Lorentz-invariant component of the 4-vector is useful, especially at hadron colliders, since the boost along the z-axis for the interacting particles is not known exactly.

Proton-proton collisions can often be thought of in terms of collisions of partons inside protons. This then implies that when two partons interact they will have some fraction of the total energy of the proton. This fraction varies from event to event; thus initial momenta along the beam axis (z-axis) of the partons before the interaction is not known. However, the initial momentum of each parton in the transverse plane is nearly zero. Therefore, it is often convenient to think about each particle in terms of its transverse components, for example instead of the total momentum (P_{tot}) we consider the component along the (x,y) plane p_{T}. The particle's 4-momentum can be fully determined by measuring the mass or energy, p_{T}, η, and ϕ. They can then be related to the usual P_x, P_y, and P_z by

$$P_x = P_{tot} \sin\theta \cos\phi = p_{\mathrm{T}} \cos\phi \tag{3.8}$$

$$P_y = P_{tot} \sin\theta \sin\phi = p_{\mathrm{T}} \sin\phi \tag{3.9}$$

$$P_z = P_{tot} \cos\theta = p_{\mathrm{T}} \sinh\eta. \tag{3.10}$$

3.3.2 Tracker

Once a particle has been created from a proton-proton collision and passes through the beam pipe, the first subdetector that it can interact with is the tracker. It is important to keep in mind that particles produced parallel to the beam pipe cannot be measured by the CMS detector. The main purpose of the tracker is to measure the curvature of charged particles with $p_{\mathrm{T}} > 1$ GeV, so that we can accurately determine their momentum. The tracker consists of many small modules that are grouped in layers transverse to the beam pipe. In fact, CMS is pretty much an extremely complicated cylindrical onion of detectors. When a charged particle goes through the tracker it will leave a "hit" in many layers that can later be aligned to form the path of the particle.

The tracker is designed to measure the trajectory of charged particles with minimal interactions between incident particles and the tracker materials in order to reduce bremsstrahlung and nuclear interactions. With this in mind, the tracker is colloquially known as a non-destructive detector, and is constructed with a minimum amount of material. It is also the closest to the beam pipe and IP, which implies that it sees the highest flux of particles. Thus, it must be composed of radiation-hard material, which is why it is made of silicon. The tracker has a diameter of 2.5 m, a length of 5.8 m, and about 200 m^2 of active silicon. This makes the CMS tracker the largest silicon tracker ever constructed. It has about 75 million readout channels and provides a position measurement resolution of about 10 µm. It also has a p_{T} resolution of about 1–2% for a 100 GeV particle. The tracker consists of two components, the pixel tracker and strip tracker. A cross section view and pictorial view of the tracker are shown in Figs. 3.7 and 3.8, respectively.

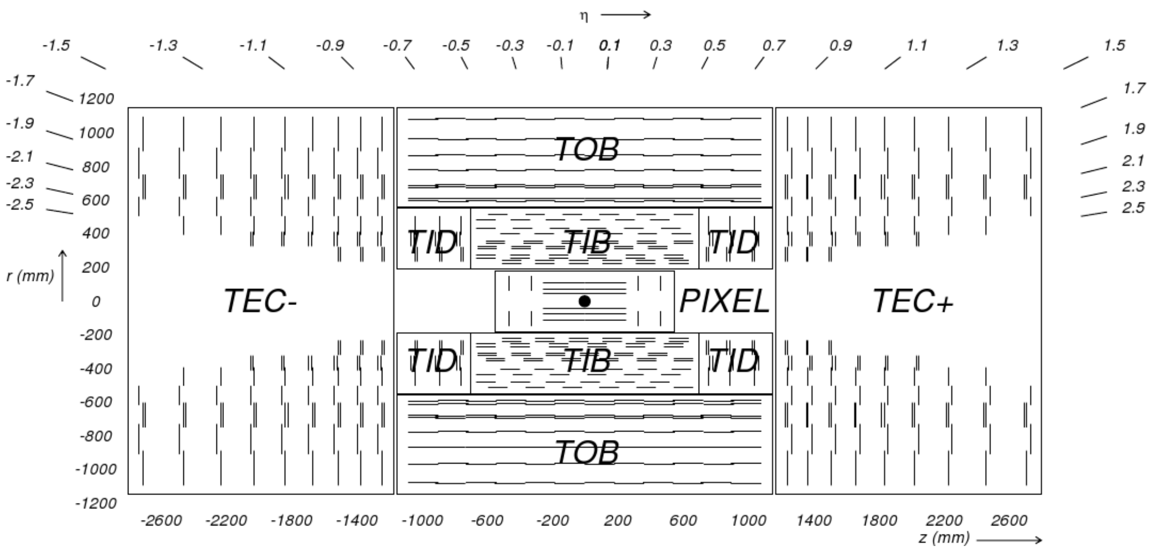


Figure 3.7: A cross section view of the tracker with its subsystems labeled. Closest to the IP is the pixel detector (BPIX and FPIX) followed by the inner (TID and TIB) and outer (TEC and TOB) strip tracker. The tracker has coverage out to $|\eta| = 2.4$.

The pixel tracker consists of over 1400 modules that comprise the segmented silicon sensor with a pixel pitch of $100 \times 150\,\mu\mathrm{m}^2$. The modules are aligned in multiple layers around the beam pipe. Like other subdetectors, the pixel tracker is further divided into a barrel region (BPIX) and two endcaps (plus or minus FPIX). In the legacy system there were three layers in BPIX, and two layers in FPIX. After the 2016 data-taking period, the pixel tracker was completely replaced and new electronics were

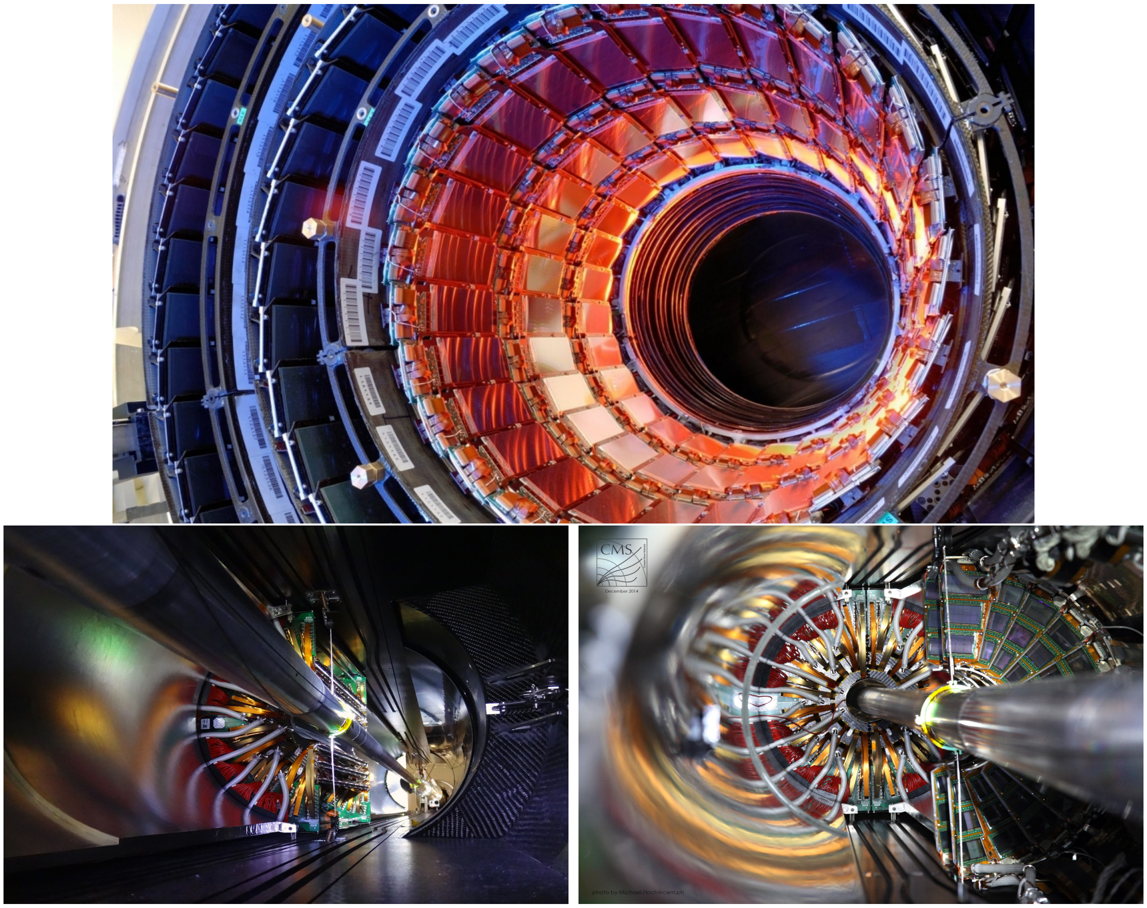


Figure 3.8: Images of the legacy tracker. (Top) The inner and outer strip tracker barrel. (Bottom left) Half of the BPIX. (Bottom right) Half of the FPIX, placed in front of the BPIX. Image source [16].

installed [17]. With this upgrade, the first layer was moved closer to the beam pipe, and an extra layer was added to both BPIX and FPIX. This can be seen in Fig. 3.9. This led to an increase in vertex reconstruction efficiency, while assuring that the pixel detector as a whole will continue to function well during the CMS data-taking periods between 2017–2023. If the pixel tracker had not been replaced, the radiation damage would eventually lead to a significant loss in data quality.

During the data-taking period of 2017, the first year of running with the Phase 1 pixel detector, there were issues with the electronics that made some of

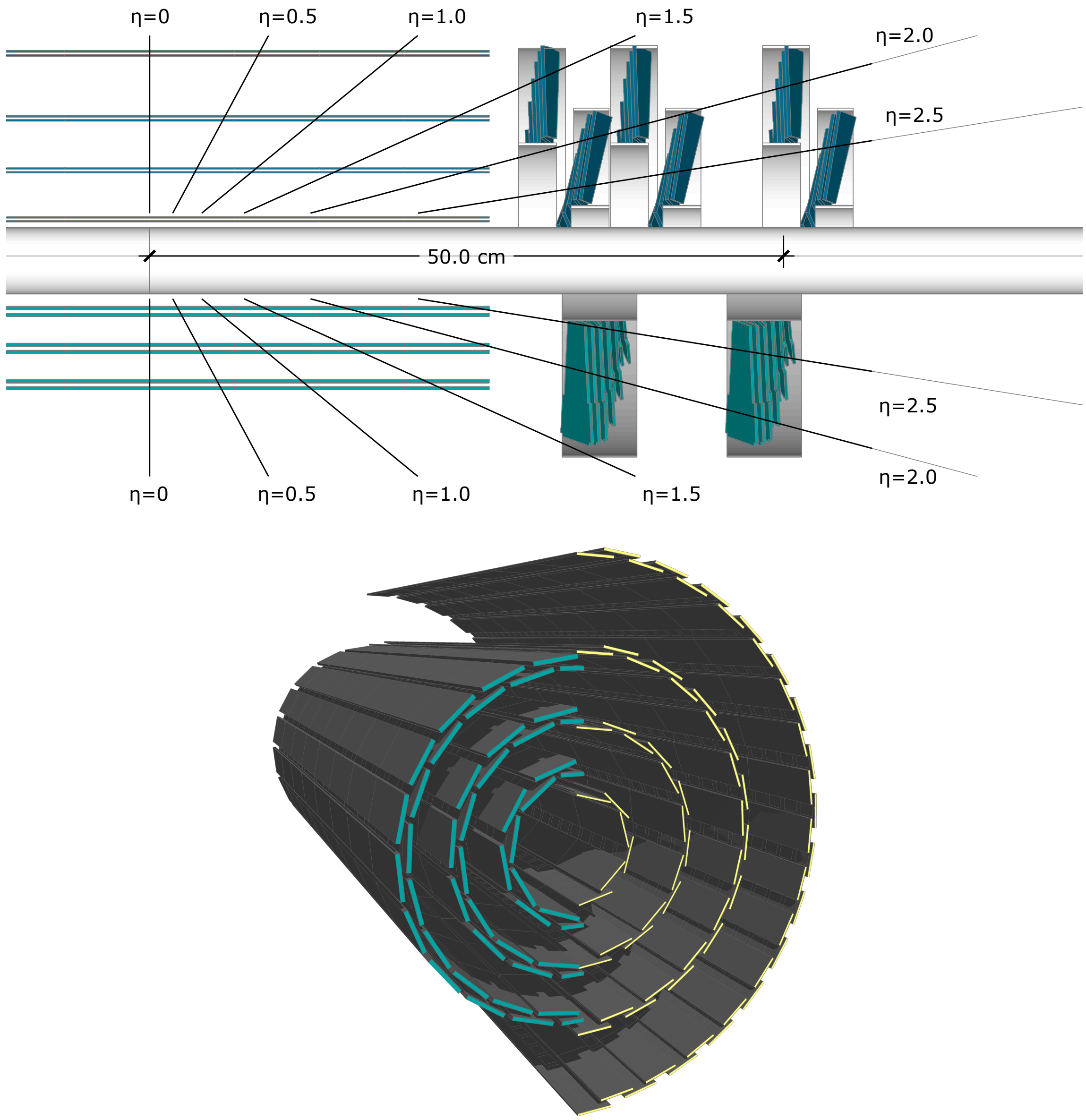


Figure 3.9: Diagrams comparing the legacy pixel detector and new Phase 1 upgrade pixel detector. (Top) A cross sectional view of the pixel detector where top is the new Phase 1 pixel detector and the bottom is the legacy pixel detector. (Bottom) A view of the pixel barrel where the left is the legacy barrel pixel detector and the right is the Phase 1 barrel pixel detector.

the modules stop functioning. It was later determined that a component related to powering the module was failing in the unique high magnetic field, high radiation flux environment. The impact on physics analysis was minimal, and the detector simulation was adjusted to include the missing modules. For the analysis presented in this thesis, the failing pixel tracker electronics had only a small impact on the reconstruction of electrons, muons, and bottom quarks. All the damaged modules were replaced during the year-end technical stop after the data-taking period of 2017.

The Silicon Strip Tracker (SST) consists of over 15,000 silicon strip modules with varying sizes. The Tracker Inner Barrel (TIB) has four layers of modules along with the Tracker Inner Disks (TID), which have two discs of modules. The Tracker Outer Barrel (TOB) has six layers of modules along with the Tracker Endcaps (TEC), which have nine discs of modules. The overall layout of the silicon strip tracker is shown together with the silicon pixel tracker in Fig. 3.7.

3.3.3 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is almost completely made of lead tungstate ($PbWO_4$) crystals, with the exception of two layers of the preshower detector in front of the endcaps, which are composed of silicon strips. A diagram of the layout and coverage is shown in Fig. 3.10. The main purpose of the ECAL is to measure the energy and position of any electromagnetically interacting particle that makes it past the tracker, for example, electrons and photons. Unlike the tracker, calorimeters such as the ECAL are considered destructive instruments, designed to completely stop and absorb the particles. When a particle hits the ECAL it produces an electromagnetic shower by interacting with the $PbWO_4$ crystals. For particles above a specific threshold energy, about an MeV, the dominant processes are electron-positron pair creation for photons and bremsstrahlung for electrons. These two processes will continue until the energy of the photons (electron) produced fall below the 1 MeV threshold, after which the photoelectric effect and Compton scattering dominate. This cascade from a single high-energy electromagnetically interacting particle can produce many orders of magnitude more low energy particles. This can be seen in the bottom image of Fig. 3.11, which shows an example of the $PbWO_4$ crystals with an artistic rendering of an electromagnetic shower.

The ECAL is a homogeneous calorimeter. The $PbWO_4$ crystals fill most of the ECAL volume, serving as the active medium and producing light signals. The ECAL barrel (EB) is composed of about 61200 crystals that have a front-face cross section of about 22×22 mm^2 and a length of 230 mm. The geometry of the crystals is designed such that they can almost completely contain a single electromagnetic shower. The ECAL endcap (EE) is composed of about 7324 crystals that have a front-face cross section of 28.6×28.6 mm^2 and a length of 220 mm. At the end of each crystal, a silicon avalanche photodiode (APD) collects the light produced by the crystal. The APD then converts the light it collects to an electrical signal, which is amplified for further processing. The energy of the particle that produces the shower is related to the amount of light the APD collects. The resolution of the ECAL energy measurements can be described by a function of the energy, and a few experimentally measured constants

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{3.11}$$

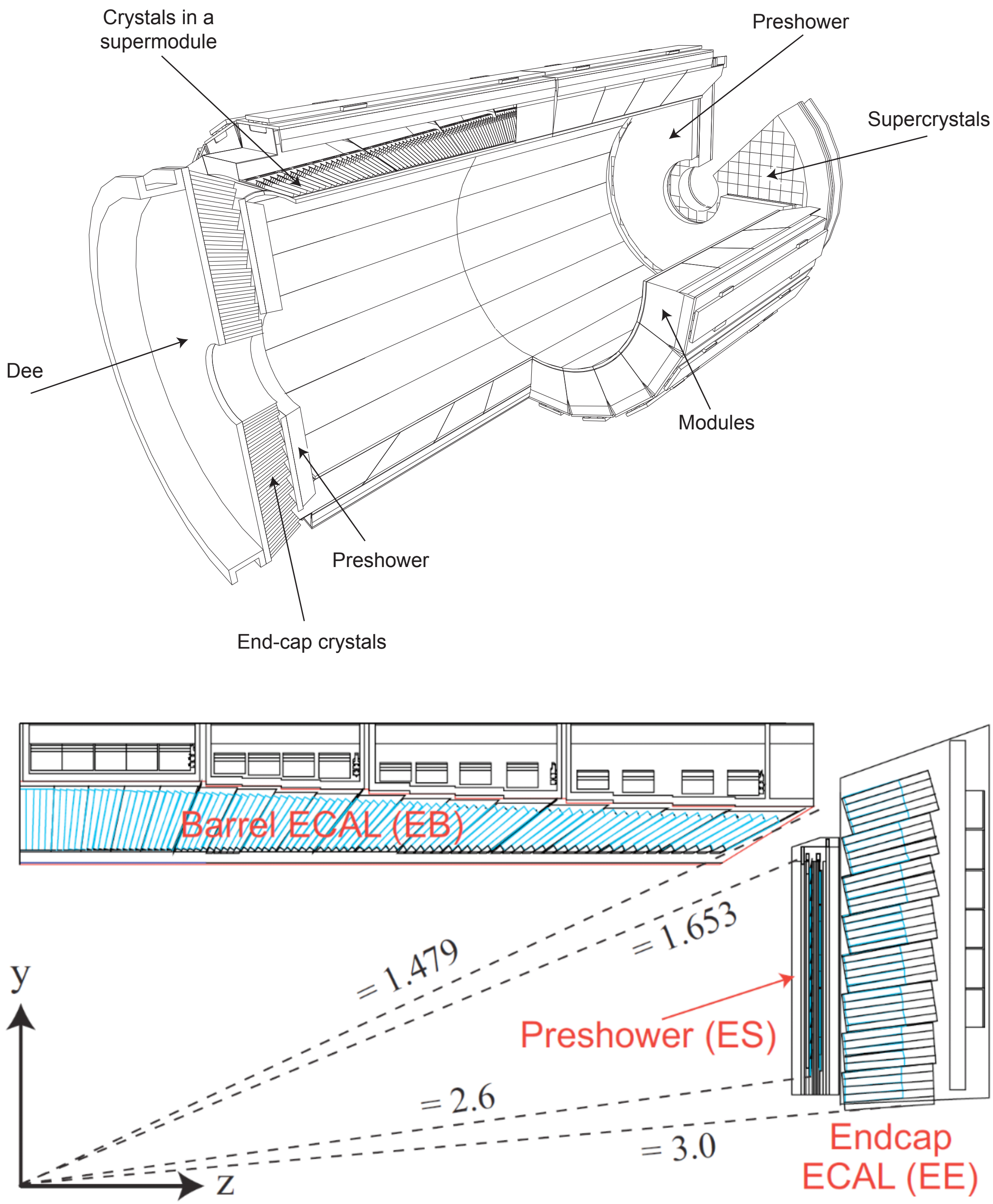


Figure 3.10: Diagrams of ECAL. (Top) A sketch of the ECAL showing the barrel modules that contain the crystals, the endcaps, and the preshower detectors. (Bottom) Diagram showing the coverage of the barrel and endcap of the ECAL subdetector.

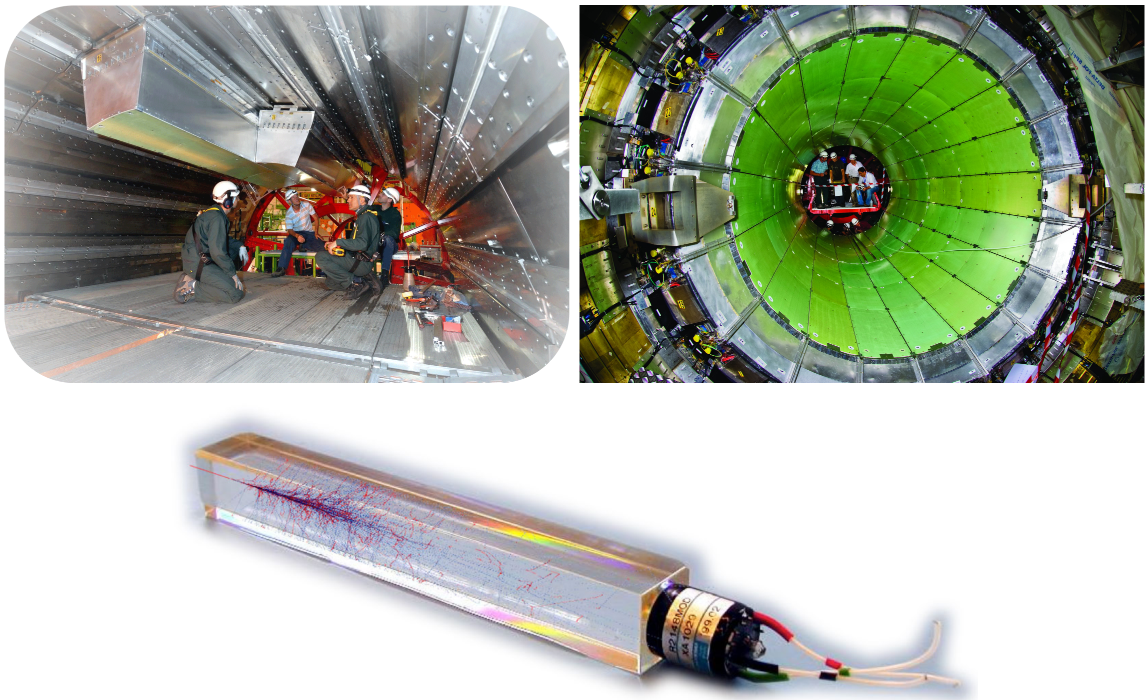


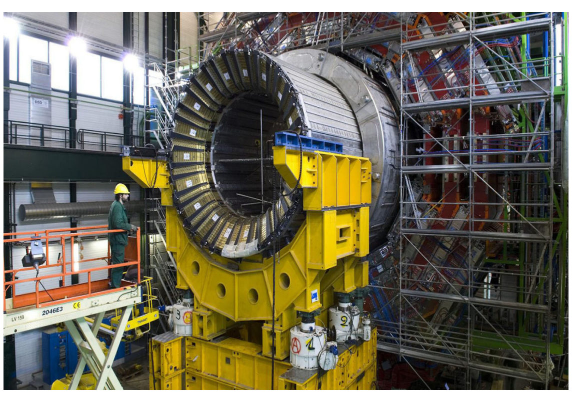
Figure 3.11: Images of ECAL. (Top left) Image of a EB module sliding into place. (Top right) Image of the fulling installed EB (reflecting green laser light). (Bottom) Image of a lead tungstate ($PbWO_4$) crystal with a electromagnetic shower shown inside.

where S is the stochastic term, N is related to the noise, and C is a constant offset. One interesting feature of this equation is that the measured energy resolution goes down (improves) for higher energy particles.

3.3.4 Hadron Calorimeter

The hadron calorimeter (HCAL) surrounds the ECAL system and is the last of the subdetectors inside the solenoid. Like the ECAL subdetector, the main purpose of the HCAL is to measure the energy and position of hadrons that pass through the ECAL and enter it, for example charged pions and neutrons. It is a destructive detector in that the particles interact with the calorimeter material, and the interactions cause a cascade of secondary particles called a hadronic shower. In the HCAL, incident hadrons produce a hadronic shower where one high energy hadron produces a large number of low energy particles. The main difference between a hadronic shower and an electromagnetic shower is that the hadrons interact with the nucleus of the material in the calorimeter which releases high energy charged particles, which then produces electromagnetic showers.

The HCAL consists of several subsystems. The HCAL barrel (HB) and two endcaps (HE) are located inside the solenoid magnet. These two subsystems contribute the most to the physics analysis discussed in this thesis. The other systems are a very far forward component (HF), and another barrel layer (HO), which is outside the solenoid magnet. The HF is mostly used by analyses that look at events with physics objects at large $|\eta|$, for example, heavy ion physics. The HO was added in hopes of measuring particles that make it past the HB and the solenoid, but it turns out that HO does not contribute significantly to physics object reconstruction. Pictures of HB and HF are included in Fig. 3.12.



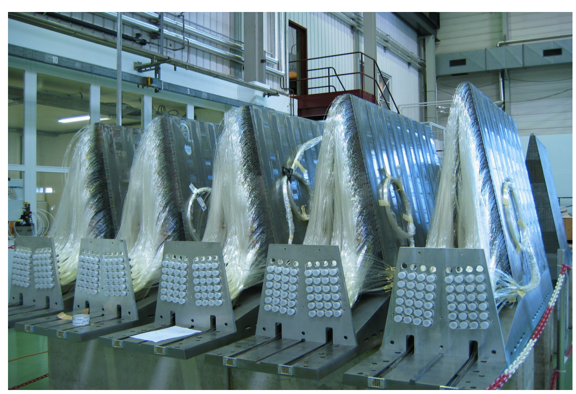


Figure 3.12: (Left) Image of the HB brass and scintillator being installed inside the solenoid. (Right) Five segments of the HF during assembly, showing the quartz fibers and steel. Image source [18].

The HCAL was another subdetector that was enhanced during the Phase 1 upgrade, and is detailed in the Phase 1 upgrade CMS technical design report [19]. This was done by replacing all readout electronics of each subsystem. HO was the first subsystem to be upgraded before the 2016 data-taking period, followed by HF after the 2016 data-taking period. The two most important subsystems were purposely chosen to go last to build on experience gained during the HO and HF upgrades. After the 2017 data-taking period, HE was upgraded, followed by HB after the 2018 data-taking period. Due to the staged upgrading of HCAL, the HCAL system was effectively different for all three data-taking periods considered in this thesis. These transitions are accounted for in the simulation of events and modifications to the event reconstruction software. The HB subsystem used the legacy electronics for all three years. Figure 3.13 includes two diagrams that show the legacy and Phase 1 HCAL systems, detailing the depth segmentation. Individual channels of the HCAL are defined as a function of η, ϕ, and depth, and they are read out separately. The readout electronics, as well as more compact photosensors adopted for the Phase 1 upgrade, enable a larger number of readout channels, which is reflected in an increase in the depth segmentation.

The exact means of collecting the energy and position information vary over different subsystems of the HCAL. We will focus on the HE and HB subsystems, because they are the only part of HCAL used for this thesis. Unlike the ECAL, the HE and HB are not composed of a homogeneous material. They are sampling calorimeters, consisting of alternating layers of plastic scintillator and brass. The brass, as the absorber, causes incident hadrons to produce cascade showers of secondary particles well contained in the HCAL volume. The scintillator then receives showered

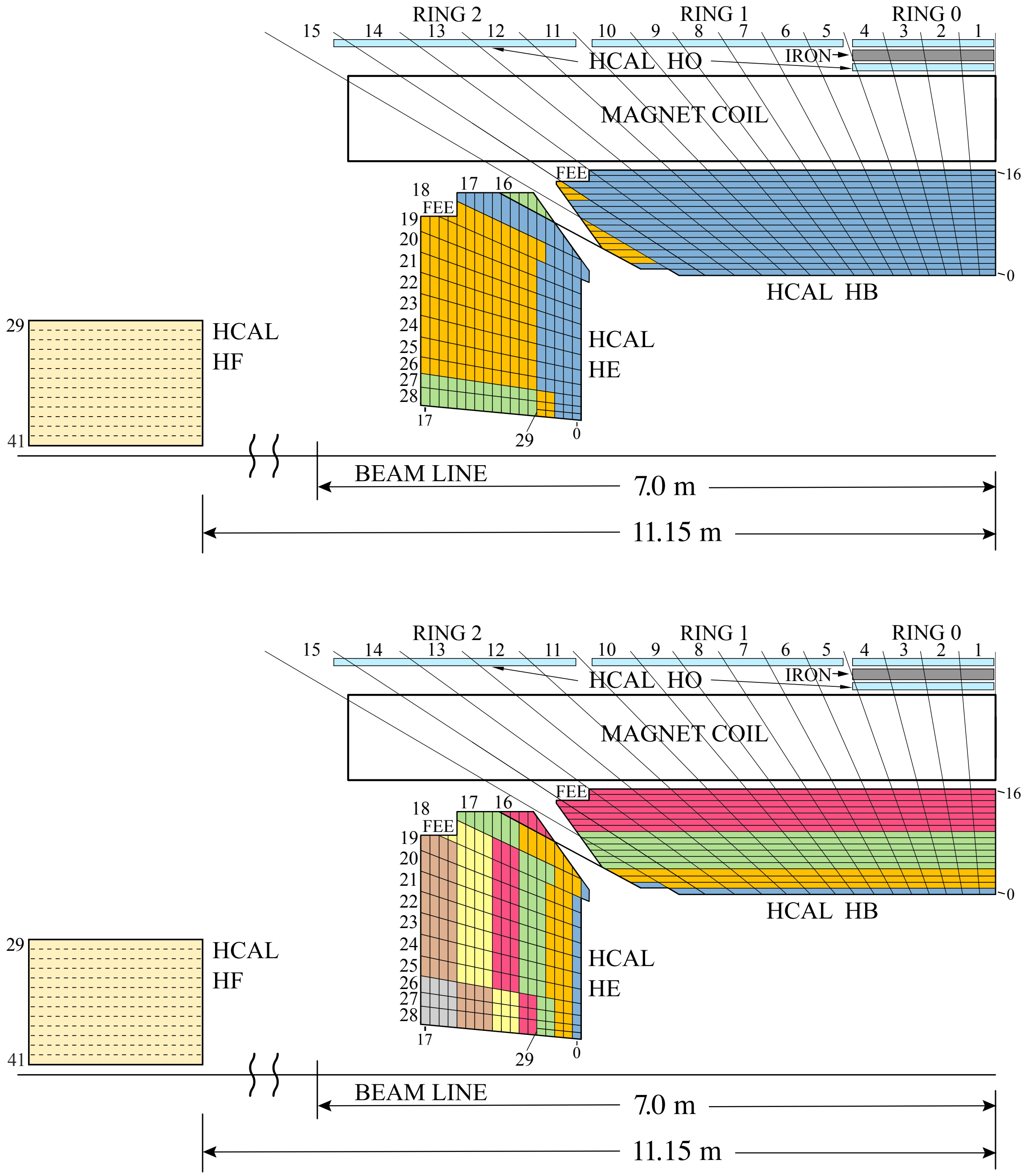


Figure 3.13: Diagrams of HCAL. (Top) A diagram detailing the depth segmentation for the legacy HCAL. (Bottom) A diagram detailing the depth segmentation of the upgraded Phase 1 HCAL.

secondary particles produced in the brass absorbers, and emits light in proportion to the number of particles that interact with the scintillator. The energy information of the hadron is then converted to an analog light signal inside the plastic scintillator. The frequency of scintillating light corresponds to blue. It is shifted to the range of green in optical fibers, which are often referred to as wavelength shifting fibers (WSF). However, if light can get into this fiber it can come out, which is why it needs to be shifted to a different frequency, so that it will not leave the fiber in transit. The scintillator and readout optical fibers were not replaced during the Phase 1 HCAL upgrade. The part that was upgraded turns this light into an analog signal, then eventually into a digitized signal. This digitized signal is then sent off-detector to be processed further. Figure 3.14 shows some photos from the HCAL Phase 1 installation effort.

3.3.5 Magnet

The magnet used by CMS is a large solenoid that has properties detailed in Table 3.1. It is a superconducting magnet that is cooled to 4.5 K by means of an extensive cryogenic system. Figure 3.15 shows a magnetic field map around the CMS detector. The main purpose of the magnet is to bend the trajectories of charged particles to measure their momentum, and to identify the sign of their charge. The resolution of the momentum decreases as the p_{T} of charged particles increases. This is the main motivation for maximizing the field strength. The magnetic field is a constant 3.8 T inside the magnet, and falls rapidly outside, by means of a steel return yoke that has an outer diameter of 14 m. The yoke is made of three layers and provides most of the structural support for CMS. In general, only muons and

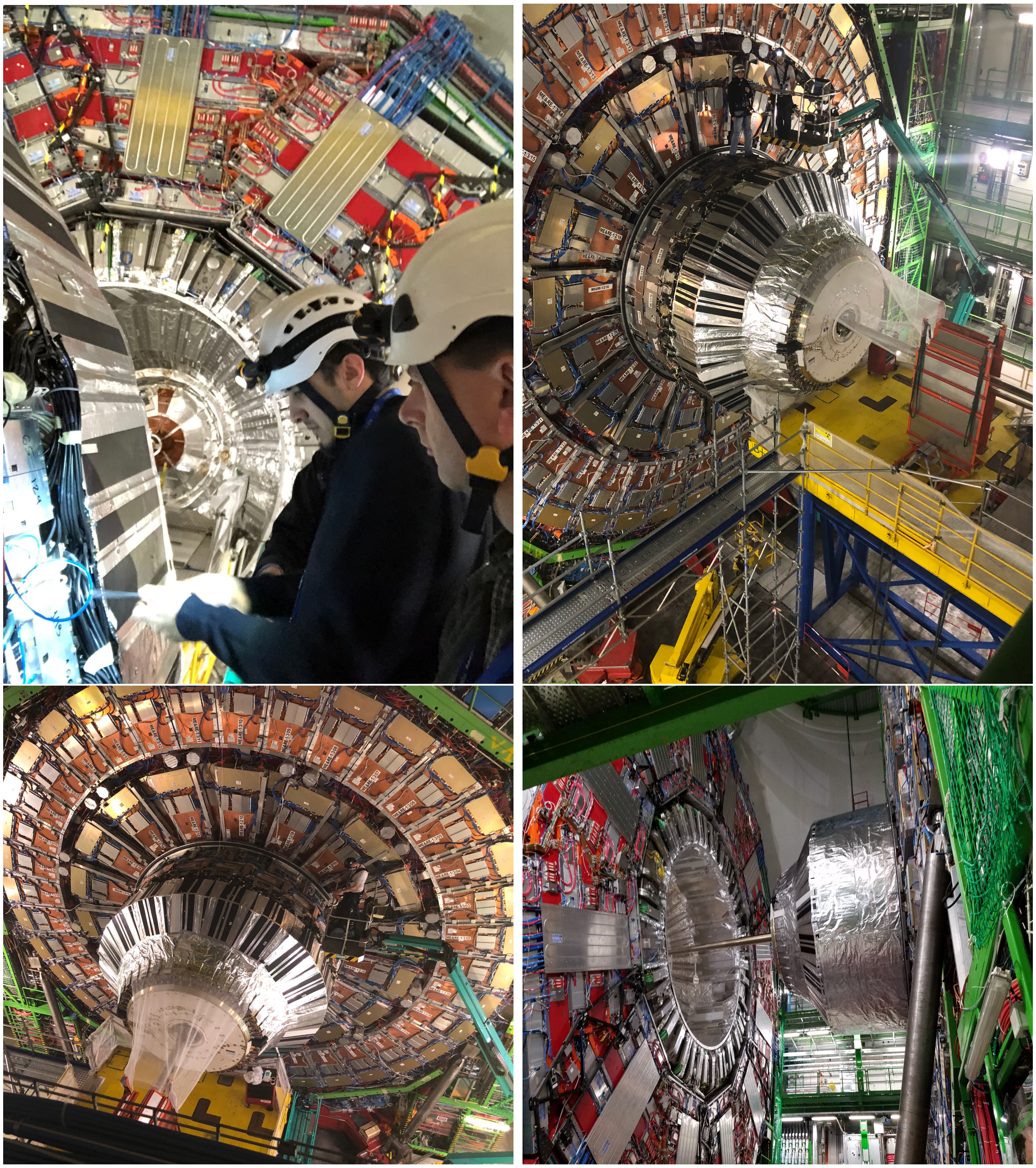


Figure 3.14: (Top) Images of the HCAL Phase 1 upgrade. (Bottom left) View of the endcap showing HE and the CSC muon systems. (Bottom right) The closing of the endcaps after the HE Phase 1 upgrade, showing the intermixed muon DTs and the solenoid.

neutrinos are expected to make it past the magnetic field, and of course new weakly interacting particles could make it this far, too. Although, there are some hadrons that make it past the HCAL, it is a small fraction. The tracker, ECAL, and HCAL fit inside the solenoid, and the muon systems are interleaved with the return yoke. It has a momentum resolution of $\Delta p/p \approx 10\%$ for momenta at 1 TeV, which is sufficient to determine the sign of a muon up to $p_{\mathrm{T}} \approx 1$ TeV. For the 2016–2018 data-taking periods used in this thesis, the magnetic system was fully functional; in previous years it had not operated fully.

Table 3.1: Properties of the CMS solenoid magnet.

Property	Value
Field	3.8 T
Inner Diameter	5.9 m
Length	12.9 m
Current	19.5 kA
Stored Energy	2.7 GJ
Number of Turns	2168
Weight	12,000 tonnes
Hoop Stress	64 atm

3.3.6 Muon Systems

Last but not least of the CMS subdetectors are the muon systems. The purpose of the muon systems, as the name suggests, is to measure the position of muons. Muons are the heavier counterpart of the electron, having a lifetime long enough to pass through CMS before decaying. Unlike electrons, muons do not interact much with material, so they can pass through several meters of iron. This means that muons leave a very clean signature, and are therefore one of the namesakes of the Compact *Muon* Solenoid. In fact, CMS was designed, and named, with the goal of detecting

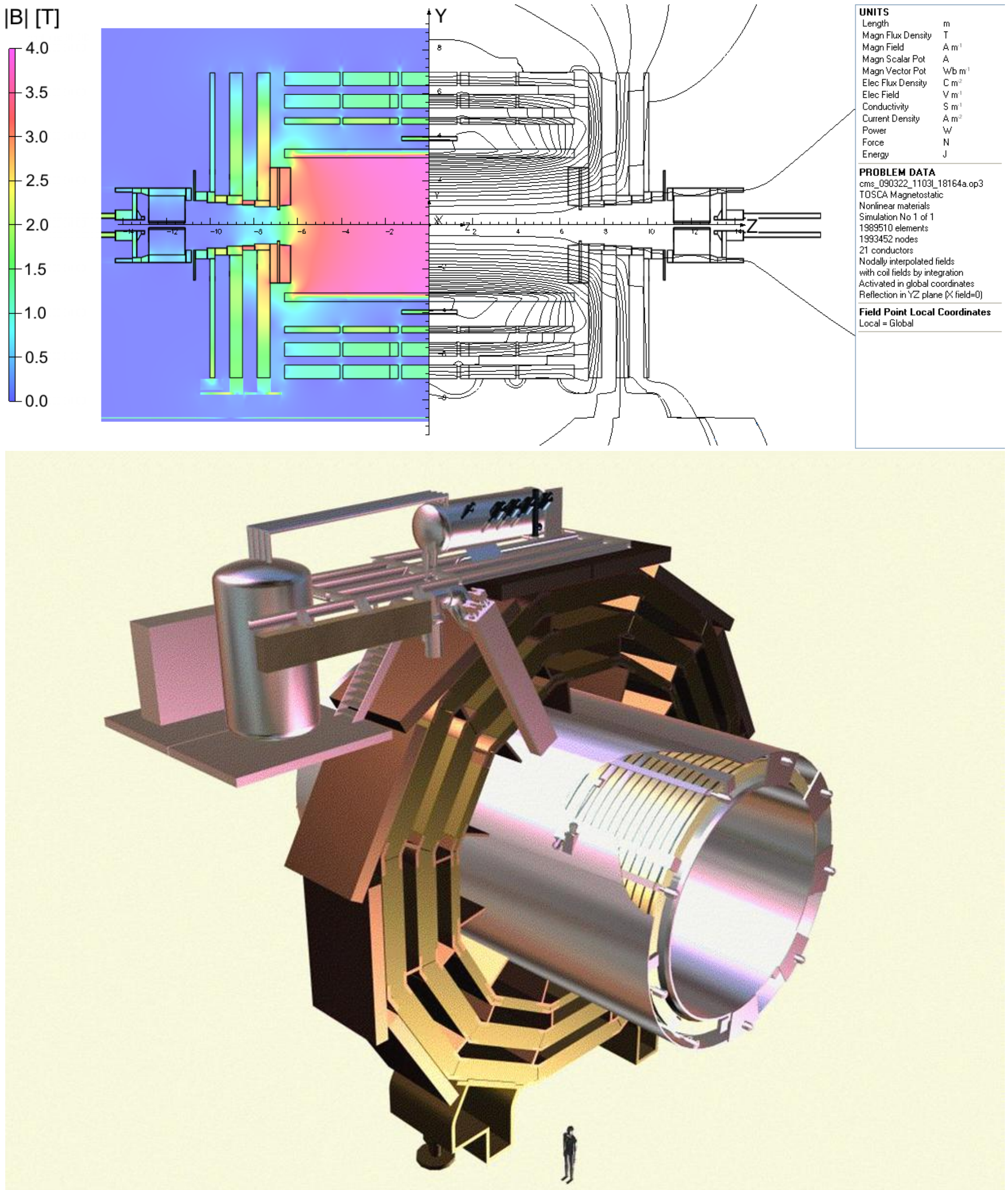


Figure 3.15: (Top) Diagram showing the magnetic field strength inside and outside of the solenoid. (Bottom) A schematic of the solenoid and its cryogenic system. Image source [20].

muons from Higgs boson decays. It turns out, with no surprise, that detecting muons was critical for the discovery of the Higgs boson. As shown in Fig. 3.16, most of the CMS volume is filled with the muon systems. The muon systems are composed of three separate subsystems, the Drift Tubes, Resistive Plate Chambers, and Cathode Strip Chambers.

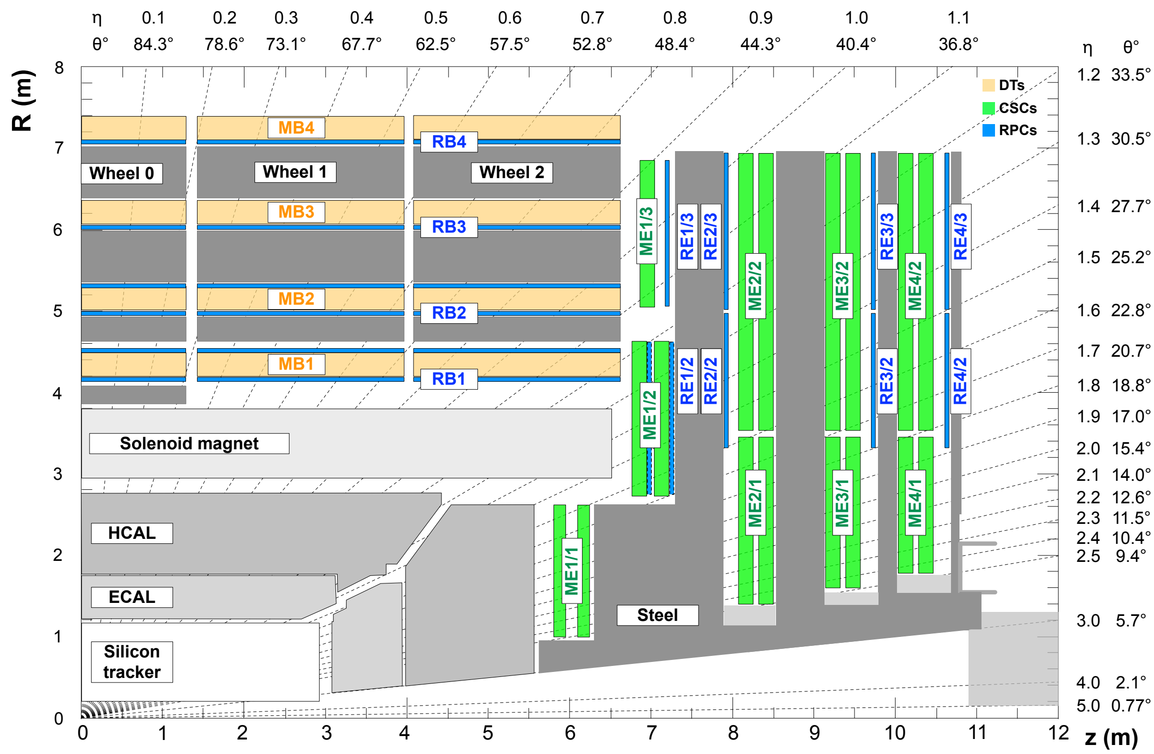


Figure 3.16: Diagram showing the layout and coverage of the CMS subdetectors, highlighting the coverage of the muon detector DTs, CSCs, and RPCs.

The *Drift Tubes* (DT) are wire gas detectors in which a charged particle (muon) passing through creates free electrons in the gas volume that are then attracted to positively charged wires. Each DT is composed of many tube cells that are about 4 cm wide and are filled with gas, together with the positively charged wire. Based on the position of the wires that registered the free electrons, the position of the particle that went through the DT can be measured. The (r, ϕ) coordinate is measured by wires parallel to the beampipe, while the z coordinate is obtained from wires perpendicular to the beampipe. The electrons' drift speed and arrival time is used to determine the muon's distance away from the wire. There are about 250 DTs located in CMS. Each DT is about 2 m by 2.5 m in size, and they are arranged in four layers parallel to the beam pipe in the barrel of CMS. Each DT has 1 or 2 RPCs attached to it, as shown in Fig. 3.16, to provide redundancy because the timing of the DT can be a problem. It takes longer than 25 ns for the DT to readout a muon hit. The timing and radiation sensitivity is the reason why DTs are only used in the barrel, where the number of particles produced and radiation damage is much lower compared to the endcaps. A high-p_{T} muon can cross up to 6 RPCs and 4 DTs, which produces about 44 measured points that can be used to construct a muon track. The single point resolution of a DT is about 200 μm.

The *Resistive Plate Chambers* (RPC) consist of two high resistive parallel plates. One plate is positively charged and the other is negatively charged, with a gas volume separating them. When a charged particle (muon) passes through the gas it produces free electrons, causing an avalanche that moves towards the positively charged plate. The electrons are then collected by metallic readout strips. This gives position information, but the resolution is not as good as the DT or CSC. However, the timing resolution is about 1 ns, which is very fast and is ideal for event triggering. There are 610 RPCs in the barrel and endcap sections.

The *Cathode Strip Chambers* (CSC) have a similar design to the DTs. The CSCs also use free electrons produced in a gas by charged particles (muons) that are then collected by a positivity charged wire. However, unlike the DTs, they also have a negatively charged wire that is meant to collect the positively charge ions that form when the free electrons are produced. The CSCs are more radiation hard compared to the DTs, which is why they are used in the endcaps. There are 468 CSCs intermixed with the RPCs to cover the endcaps, as shown in Fig. 3.16.

3.3.7 Trigger

As described in Section 3.2, proton bunches cross every 25 ns, which means that there are 40 million proton collisions every second. One may want to record the data for every bunch crossing, but it would result in an impossible amount of information to store. Also, not every bunch crossing has an "interesting" event, where an interesting event is subjectively defined as possibly having information worth studying. Logistically, we can only store about a few hundred events per second. This is where the trigger system comes in. The trigger system, aptly named, will "trigger" on interesting events by flagging them for storage and further processing. This system will perform a quick reconstruction of every event, and if it passes a predetermined threshold, the event will be passed to long-term storage. The time it takes to fully collect the data for a particular event, and the time it takes the trigger system to reconstruct that event in order to make a decision, forces each subdetector to independently buffer every event before dumping it or passing it along to long-term storage. The trigger system consists of two main parts, the Level-1 trigger and the High Level Trigger.

Passing the *Level-1 Trigger* (L1T) is the first step before an event is saved for future analysis. The L1T is designed to make extremely fast decisions with a small amount of information. It reduces the number of events which may be saved down to about 100,000 per second, which means that it dumps about 99.75% of all events. The L1T electronics are stored in the CMS service cavern adjacent to the experimental cavern. The electronics must be specially designed and constructed. A total time of 3.2 μs is budgeted for the partial data to be transfered to the service cavern, a decision to be made, and that decision to be transfered back to each subdetector, which are meanwhile still buffering the full event information. This short time window makes it impossible to use the full information from every subdetector to make this decision. Thus, only the ECAL, HCAL, and the muon systems are used. Objects such as photons, electrons, muons, and jets are quickly constructed, and events in which these objects pass a p_{T}, E_{T}, and other simple variable thresholds are kept and passed along to the High Level Trigger.

The *High Level Trigger* (HLT) is a computer farm stored near the control room, and it has access to all the information available for every event. This means that a more complex event reconstruction can be performed for each event, similar to the reconstruction performed for final analysis. The HLT filters the number of events saved per second to a few hundred. This means it dumps about 99.9% of the events that it receives from the L1T. The HLT was designed to accept different sets of object requirements, also known as trigger paths, to help filter events into data sets. For example, the single lepton triggers collect events with at least one lepton candidate passing certain criteria. The HLT allows for analyses to use a trigger path specific to each analysis, assuming the trigger requirements keep the number of events stored below the storing threshold. For the analysis in this thesis, the generic single-lepton trigger paths commonly used by a large number of CMS analyses were sufficient.

CHAPTER FOUR

Event Simulation and Reconstruction

4.1 Introduction

Chapter Three described how the LHC delivers proton-proton collisions and how CMS was constructed to measure the collisions. This chapter describes in greater detail how an "event" is processed. First, what exactly is an event? In short, an event is everything that happens during one proton-proton bunch crossing inside of CMS. Each event contains all the physics information relevant for that particular bunch crossing, and events are independent of one another. The data for each event are later processed to reconstruct the 4-vectors of the physics objects that came from the collision. This step of analyzing the data collected by CMS is called reconstruction, and it is performed centrally for all analyses using the CMS software framework (CMSSW), which was written by thousands of CMS physicists.

Figure 4.1 shows an example of a $\mathrm{t\bar{t}} + \mathrm{jets}$ candidate event. It depicts how each subdetector responds to particles coming from top quark decays. As described in Chapter Two, a top (antitop) quark decays to a bottom quark and a W boson before it can travel a measurable distance in the detector. In this event, the top quark and antitop quark each decay to a W boson, and the W bosons also decay before they travel a measurable distance. One of the W bosons decays to two quarks, and the other produces a muon and missing energy (neutrino). As Chapter Five describes, the $\mathrm{t\bar{t}} + \mathrm{jets}$ process is the largest SM background that needs to be estimated. This reconstructed event shows the detectable particles produced in top quark decays.

Different kinds of physics objects are reconstructed in different ways, as each object interacts with the CMS detector differently as discussed in Chapter Three. The reconstruction of physics objects will be described later in this chapter.

The event reconstruction described above applies to real proton-proton collision events collected by CMS, but we also need to reconstruct events that are simulated. Simulated events are generated by a series of computer programs based on the SM or physics beyond the SM, and they are passed through the CMS detector simulation. The simulation programs produce signals as expected from each subdetector, and simulated events can be treated in the same way as real collision data events collected by CMS. In general, there are many applications of simulated events, but for the purpose of this thesis we use them to predict some SM backgrounds and the SUSY signals. The summary below will go into more detail on event simulation and reconstruction. It also comes primarily from Refs. [21–23].

4.2 Event Simulation

Event simulation is performed using Monte Carlo (MC) simulation software. Simulated events, colloquially called MC events or MC data, are used to predict SM backgrounds as well as the particular BSM signal events that we are searching for. Before describing how an event is reconstructed, we discuss how they are simulated, since the reconstruction process for MC data and collision data events is the same.

4.2.1 Background Prediction with Simulated Events

Assuming we can successfully produce simulated events, there is a question that immediately arises: how many events should be produced? The answer depends on a few factors, but ignoring computational constraints, events are simulated until

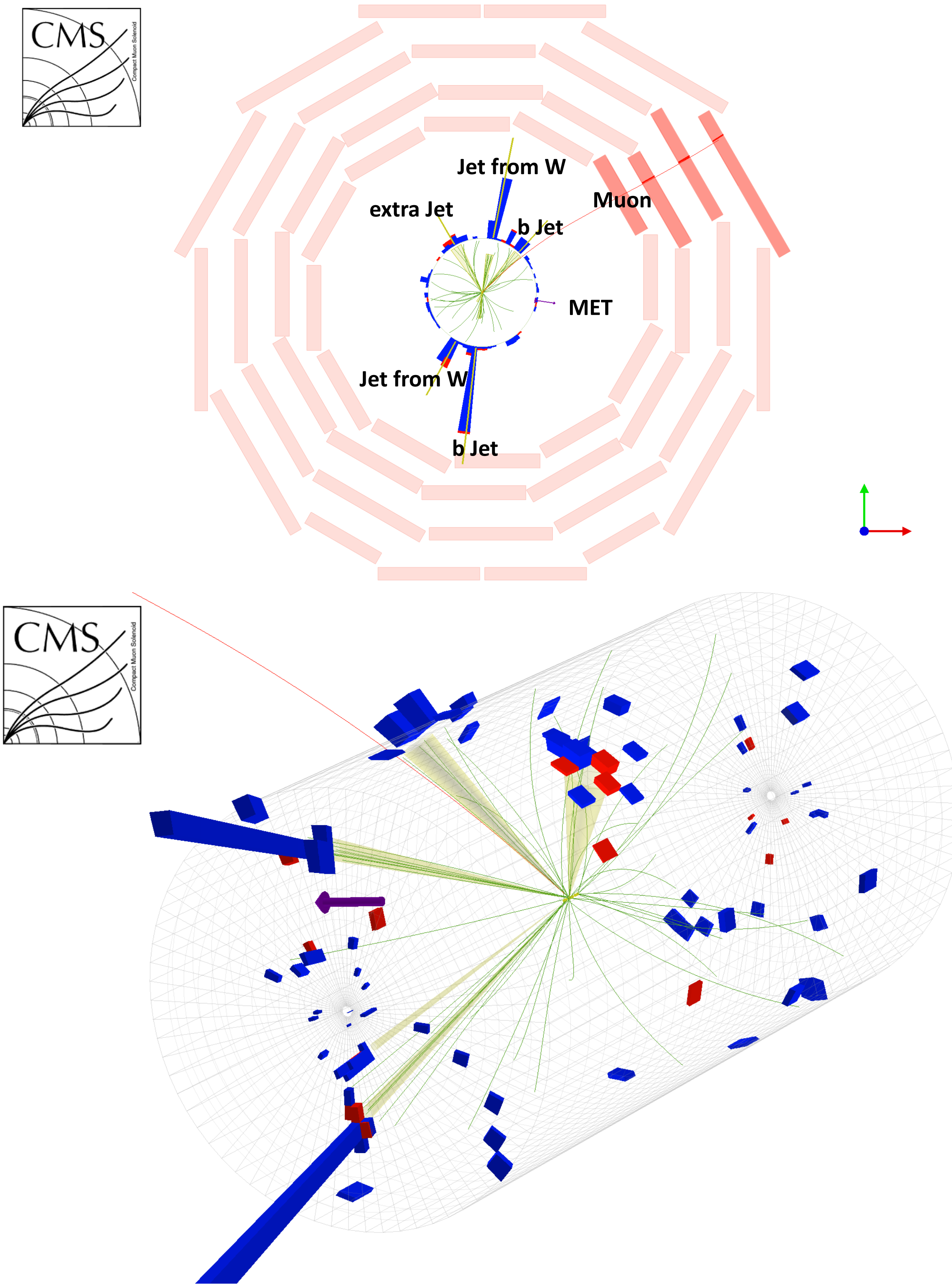


Figure 4.1: Semileptonic $t\bar{t}$ event display from different perspectives.

a signal or background prediction is considered accurate. This means the statistical uncertainty is sufficiently small. This constraint might imply that the number of generated events should match the number of events that the physics model predicts. However, this is either not enough or too many events depending on the process. What is done instead is to weight each event such that we get an effective number of predicted events corresponding to the amount of real data we are analyzing:

$$N_{\text{eff}} = \sum_{i}^{N_{\text{gen}}} w_i, \tag{4.1}$$

where N_{gen} is number of events that are generated. We can define w_i and calculate the uncertainty on the number of predicted events using

$$w_i = \frac{\sigma L}{N_{\text{gen}}} \tag{4.2}$$

$$\sigma_{N_{\text{eff}}} = \sqrt{\sum_{i}^{N_{\text{gen}}} w_i^2}, \tag{4.3}$$

where σ is the cross section of the process being simulated and L is the integrated luminosity or the amount of collision data collected by CMS. This allows us to generate an arbitrary number of events and still be able to produce a signal or background prediction and its uncertainty. These equations work not only for generated events but for data events as well. Data is just a special case where $w_i = 1$ and $N_{\text{gen}} = N_{\text{data}}$. Another interesting point is that we can reduce $\sigma_{N_{\text{eff}}}$ by generating more simulated events.

Predicting the backgrounds for any SM (or BSM) processes requires that we know the cross section for that process. There are many groups of theorists who perform state-of-the-art theoretical predictions for various SM processes and experimentalists who measure these processes as accurately as possible. A summary of the latest greatest cross section measurements is shown in Fig. 4.2 for various SM processes. For the signal models that we are searching for, we use the top squark pair production cross section values shown in Fig. 2.8.

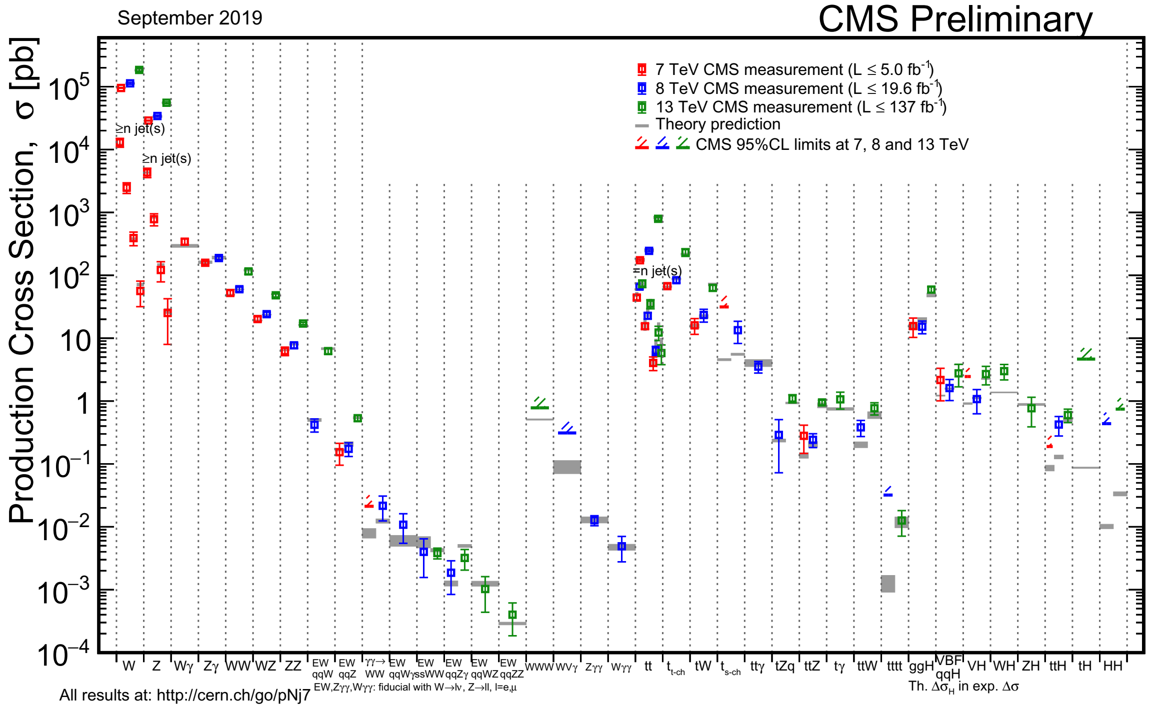


Figure 4.2: Theoretical and CMS measured production cross sections of various processes.

Now that we know how to predict the effective number of events from a set of simulated events, we still need to actually produce and reconstruct them to calculate interesting quantities for our studies.

4.2.2 Event Generation and Parton Shower

The first step in simulating events is event generation and the parton shower (PS) calculation, where the interactions of all initial particles and their byproducts are calculated. Figure 4.3 shows an example event where each blob and line represent parts of the event that need to be simulated. In general, there is more than one proton-proton interaction occurring within each event, which is called in-time pileup. This requires us to define one proton-proton interaction as the interaction of interest, which is called the hard process. All other proton-proton interactions in the event are then considered to be pileup. The existence of in-time pileup means that each event contains particles that leave tracks and deposit energy in the detector that are not relevant to the part of the event of interest.

Once we determine the process that we want to simulate, for example $t\bar{t}$ production, most of the relevant calculations are done when we simulate the hard process in the event generation. For proton-proton scattering we write Feynman diagrams that show either the free quarks or gluons, often called partons, within the protons. These partons are the particles that are actually interacting since protons are composite particles. The initial energy of each parton is not simply the energy of the incoming proton, because this energy will be shared with the other partons not interacting within the proton. In fact, the initial energy of each parton ends up being described probabilistically by a distribution, the parton distribution function (PDFs), that needs to be measured and describes the structure of the proton. Using the PDFs we can determine the initial kinematics of the partons involved in the scattering. And then, using the SM and relevant Feynman rules, we can predict the initial particles produced, for example, the two top quarks and their decay products for $t\bar{t}$ production.

Performing these calculations for the wide variety of processes that are needed by CMS is a huge undertaking. There are a few groups of theoretical physicists that write and maintain dedicated software for generating events. For the MC data

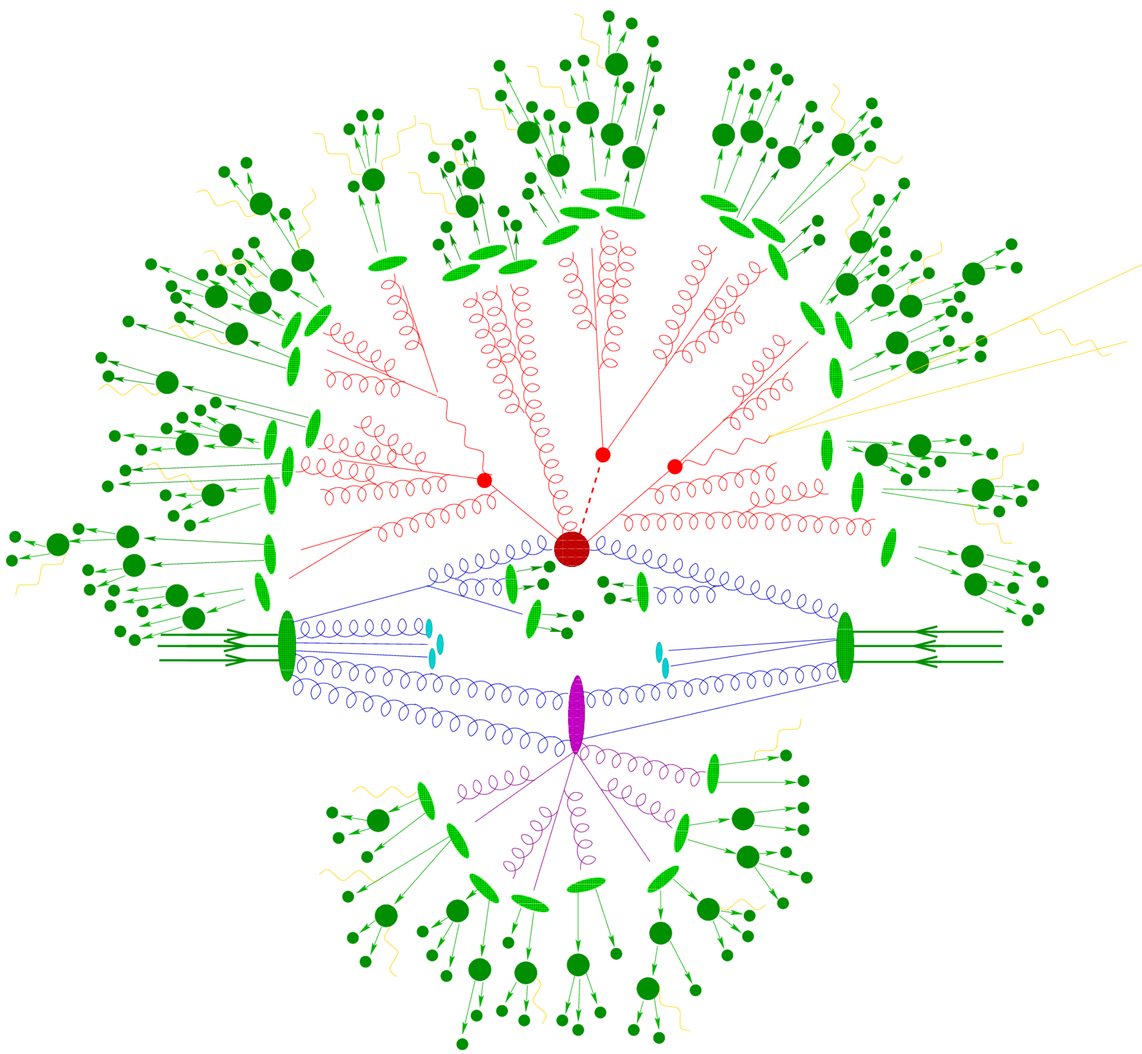


Figure 4.3: Monte Carlo-simulated event of a hadron-hadron collision. The red blob in the center represents the hard process, surrounded by lines representing bremsstrahlung as simulated by the parton showers. The purple blob indicates a second hard-scatter event. Parton-to-hadron transitions are represented by light green blobs. Dark green blobs indicate hadron decays, and yellow lines represent low energy photon radiation. Image source [24].

used in this thesis, the POWHEG and MadGraph event generator programs are used primarily.

After generating the hard process, the next step in event simulation is to simulate the rest of the particles produced in the event, the parton shower. This includes the hadronization and showering of the particles originating from the hard process. When quarks and gluons are produced, they cannot stay free in the detector due to the nature of QCD. In fact, when produced, they have enough energy to pull particles from the vacuum that then do the same until they run out of energy. This process is called hadronization. Hadronization is done such that the overall color charge is conserved, which is handled by the PS during the color reconnection (CR) step. The PS step is also where the underlying event is calculated but without the full rigor used for the hard processes calculation. Another important step processed during the PS calculation is to model extra jets that are produced from QCD radiation. When a strongly interacting particle is produced in the event and has sufficient energy, it can radiate a gluon jet. This is usually referred to as initial state radiation (ISR) or final state radiation (FSR) depending on whether it comes from the initial partons or the produced particles. For the purpose of this thesis, the modeling of extra jets in the event is important because we look at $\mathrm{t\bar{t}}$ events with at least seven jets, and we only expect four of them to come from the hard process: two b-jets plus two jets from the W boson decay.

Just like for event generation, several groups of people are dedicated to writing and maintaining software for the parton shower simulation that is used by everyone in CMS. The PYTHIA program is used for the parton shower almost exclusively in CMS and for all MC data used in this thesis. After event generation and the PS, the next step in the process is to simulate the interactions of the particles coming from the proton-proton interaction and the detector material.

4.2.3 *Detector Simulation*

The event generation and PS calculations are in general detector-independent. The detector simulation step uses an accurate representation of each subdetector and its materials for simulating the interactions of the particles coming from the hard process and PS, and is implemented by software based on GEANT4. This is where the magnetic field and the electromagnetic and hadronic showers are simulated, as well as the interactions of the particles with the detector material.

Now that we have the actual particles that each subdetector observes, dedicated software written for each subdetector is used to simulate the readout electronics. The detector simulation simulates all information that we know about the detector. Not only does it simulate all active layers and channels, it uses calibration information to output signals and energy deposit information that we would expect from real data. We even simulate dead channels in various subdetectors. The detector simulation configuration changes based on which year of data taking we want to simulate, since the CMS detector was undergoing upgrades between each year. We effectively had a different detector for each of the three years of the data-taking period.

For example, the simulation of the HCAL readout electronics needed to be updated. One important aspect of this was the modeling of the distribution of light from the scintillators and wavelength shifting fibers (Y11) that carry the light from the scintillators to the readout electronics, and the analog charge distribution coming from the photomultiplier (SiPMs). This is detailed in CMS internal note [25] for the

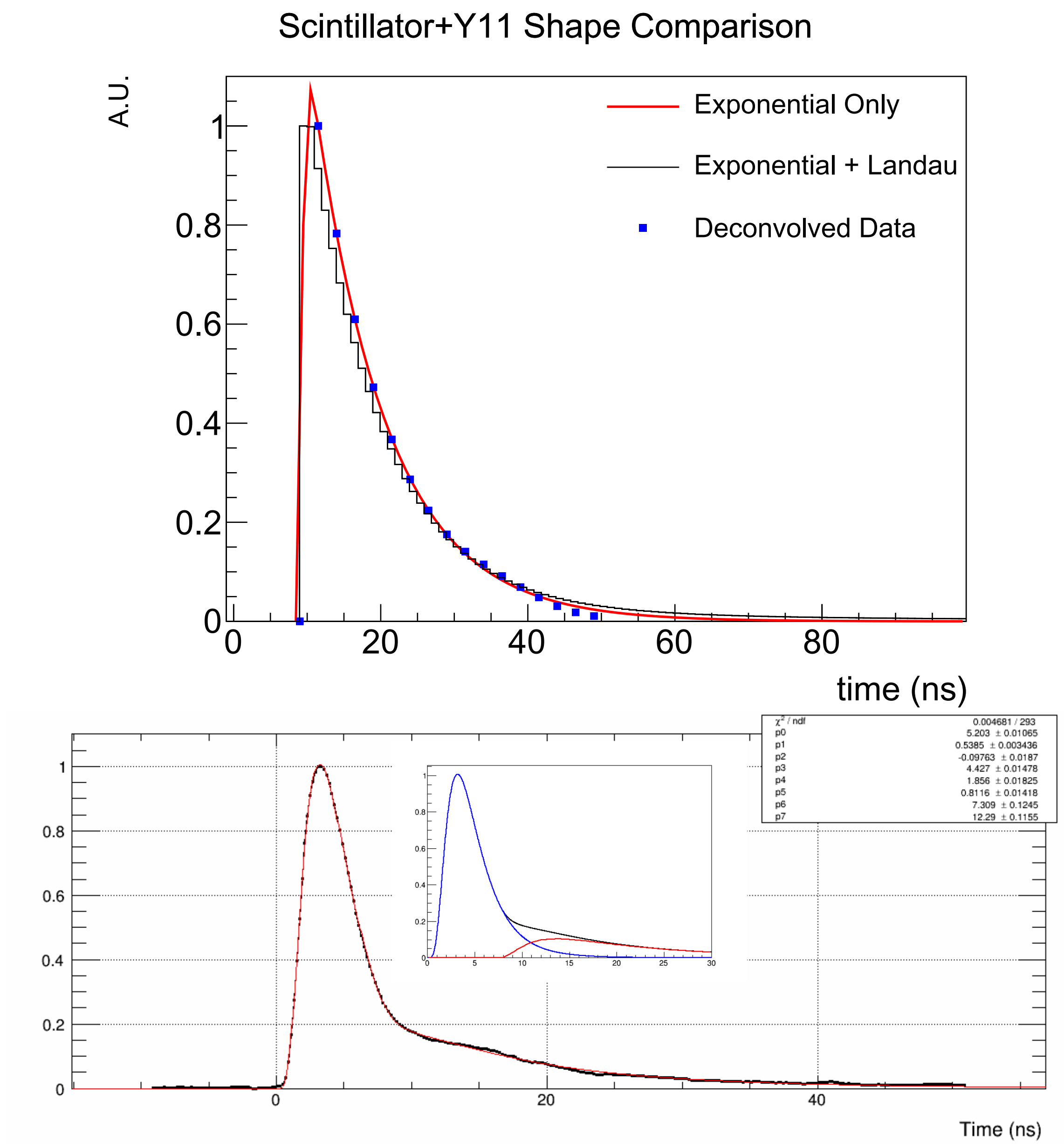
Phase 1 upgrade, which describes the electronics used for the 2018 data-taking period. As described in Section 3.3.4, the particles that make it to HCAL will interact with the brass absorber and scintillators, and will produce light proportional to the energy of the initial incident particles. This light is then passed to the readout electronics by the Y11 fibers for further processing, until we have a measurement of the amount of energy deposited into the HCAL system. It is important for both reconstruction and simulation to understand these two distributions in time. Figure 4.4 shows the bench measurements of the distribution of light from the scintillator + Y11 system and the charge distribution from the SiPMs. These two distributions can be convolved together to give us the pulse shape or charge distribution coming out of the SiPMs from data measurements. This pulse shape is integrated by the on-detector electronics every 25 ns, denoted by time step, and is sent off to the reconstruction algorithm for further processing. Figure 4.5 shows a comparison of this pulse shape for data, the updated detector simulation, and an older detector simulation that used an incorrect scintillator + Y11 shape. One important thing to note is that it takes about four time steps or $\sim$100 ns to read out all the charge coming from the readout electronics for one hadronic hit in the HCAL. That means if two successive events have hadrons that enter the HCAL, we will read out a pulse shape that is effectively two independent pulse shapes added together but shifted by 25 ns. This is handled during the reconstruction stage, but we can take advantage of the fact that the pulse shape is the same for all hadronic hits and only the amplitude of the shape changes depending on the energy of the hadron. This means if we know the pulse shape expected from a hadronic hit, we can simulate the electronics by adding multiple pulses with varying amplitudes that are shifted in time depending on whether successive events have hadronic particles striking the HCAL. This pulse shape is the lowest level of saved information for HCAL and everything necessary for reconstruction is obtained from it.

A similar process was achieved for all subdetectors. After the detector simulation step, the output of simulated data looks as if it was collected by the CMS detector.

4.3 Event Reconstruction

Once CMS has either collision data or simulated data, the next step is to reconstruct and identify all stable particles in the event. CMS has a dedicated algorithm, called Particle Flow (PF), that achieves this by reconstructing candidates (PF candidates) of electrons, muons, photons, charged hadrons, and neutral hadrons by determining their direction, energy, and type. Once we have this list of PF candidates, we can use them to reconstruct more complicated objects, such as jets.

Information from the tracker and calorimeter systems are critical for PF reconstruction. Hits in the tracking subdetector are used to reconstruct the initial path of particles from the primary vertex through CMS by the *iterative tracking algorithm*. The ECAL and HCAL measure the energy of particles that pass through the tracking system by absorbing them. These hits in the calorimeter systems are then clustered together to fully reconstruct the energy deposited in multiple channels by a dedicated *clustering algorithm*. The information from these two independent algorithms is then linked together, along with hits measured by the muon systems, by the *linking algorithm* to fully reconstruct the path, energy, and type of particles that came from the primary vertex. The PF algorithm takes advantage of the fact that each particle



Typical (averaged over 10 SiPMs x100 waveforms) laser (20 psec, FWHM) response of 2.8 mm HE SiPM. It was measured using fast 500 MHz DSO + fast amplifier (>500MHz BW, 15 Ohm input resistor, gain=10, designed by S.Los).

Figure 4.4: (Top) Fits to deconvolved pulse shape data from the scintillator + Y11 system. (Bottom) Time response of a SiPM from laser-injected bench data.

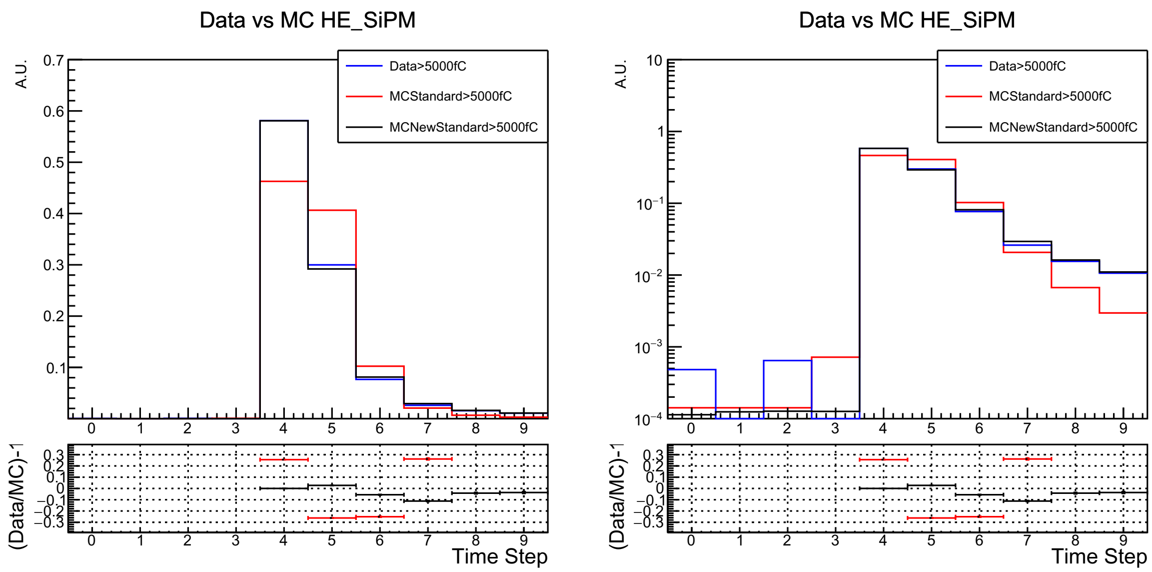


Figure 4.5: Comparison of integrated pulse shape per 25 ns for 2017 collision data and two simulation options using a linear scale (left) and log scale (right).

type leaves a unique signature in the detector as shown in Fig. 3.6. For example, a neutral hadron will not leave a track in the tracking system or a large energy deposit in the ECAL, but will leave hits in the HCAL. If we see a cluster of hits in HCAL that cannot be linked to a track or ECAL cluster, it is likely to be from a neutral hadron.

Once the PF algorithm creates the initial list of reconstructed particles for each event, there is usually further processing of the objects to filter out undesirable or misidentified objects. This stage is analysis dependent, and details of the requirements that are unique to this thesis are described in Section 4.3.1. There will also be some events that suffer from a large amount of detector noise or reconstruction failures that must be filtered out. For simulated events, the event weight described in Section 4.2.1 can be modified to account for discrepancies between collision data and MC data. This modification of the event weight w_i is often referred to as event reweighting or applying a scale factor. Most MC data corrections can be applied by deriving a set of scale factors that change the weighting of simulated events such that it matches the observed data prediction.

4.3.1 Physics Objects

4.3.1.1 Electrons The selection for electrons starts with the collection of electron candidates determined using the Gaussian Sum Filter algorithm. From this collection, electrons are further selected based on their transverse momentum ($p_{\mathrm{T}} >$ 30 GeV for 2016 and $p_{\mathrm{T}} >$ 37 GeV for 2017 and 2018) and on their pseudorapidity ($|\eta| < 2.4$). The different p_{T} requirements in 2016 and in 2017–2018 are a result of the increase in the p_{T} threshold for the electron trigger. This analysis uses the "Tight" working point for the cut-based electron ID.

The electron isolation is computed using the mini-isolation (miniIso) algorithm, and electrons are required to pass miniIso < 0.1. In this algorithm, the size of the ΔR cone used to calculate isolation is changed based on the transverse momentum of the particle, which means that depending on the p_{T} of the lepton, a different cone size is used. The radius of the cone used is computed as $\max\left(0.05,\ \min(0.2,\ 10/p_{\mathrm{T}})\right)$. In addition to the ID, the recommended impact parameter requirements are also imposed on the electron: $|d_0| < 0.05$ (0.10) cm and $|d_z| < 0.10$ (0.20) cm for the barrel (endcap) region. The kinematic distributions for electrons that pass the above criteria are shown in Fig. 4.6. No issues of concern are observed in comparisons of collision data and MC data for 2016.

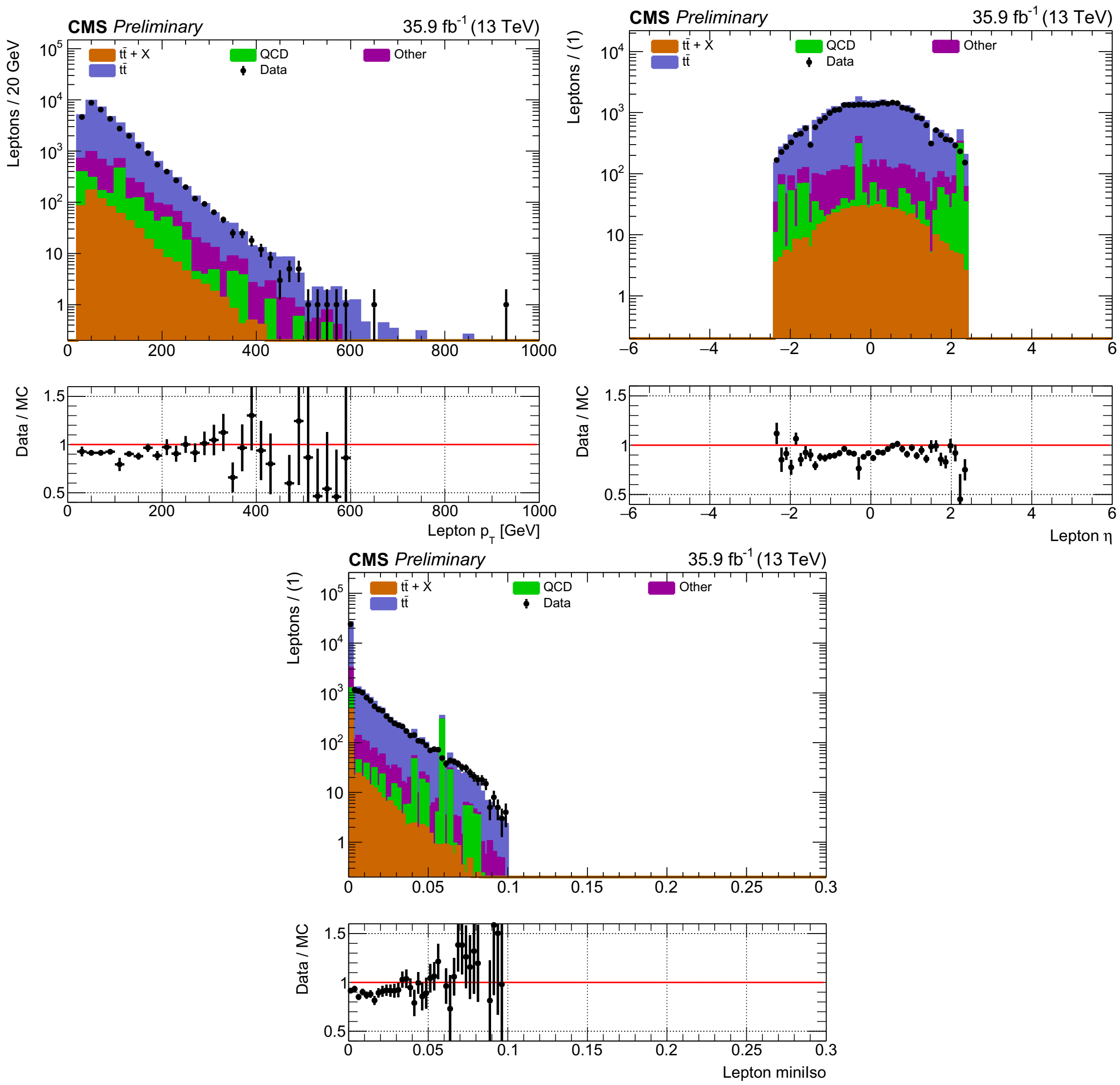


Figure 4.6: The p_{T} spectrum (top left), η (top right), and miniIso (bottom) of electrons that pass the electron selection.

4.3.1.2 Muons For muons, the baseline selection is similar to that of electrons. Muons identified by the "Medium" working point are used. Additional requirements that the muon's transverse distance from the primary vertex be less than 0.2 cm ($|d_B| < 0.2$ cm) and the longitudinal distance from the primary vertex be less than 0.5 cm ($|d_z| < 0.5$ cm) are also imposed. Muons must have $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$. Mini-isolation for muons is defined like that for electrons, with the requirement on muons being miniIso < 0.2. The kinematic distributions for muons that pass the above criteria are shown in Fig. 4.7. No issues of concern are observed in comparisons of collision data and MC data for 2016.

4.3.1.3 Jets When a quark or gluon is produced with a sufficient amount of energy and p_{T}, it produces a shower of particles through hadronization. When reconstructing the event, this cascade of particles that come from the quark or gluon hadronization can be grouped together into a single 4-vector that is called a jet. There are a few standard algorithms that take the list of particle flow candidates and cluster them into a group of jets. For this analysis, jets are clustered using the anti-k_{T} algorithm with a distance parameter of $R = 0.4$ (AK4) and uses PF candidates that pass the charge hadron subtraction (CHS). This complex algorithm is detailed in Refs. [26–31].

The main idea is to group particle flow candidates based on two distance values, d_{ij} between particles i and j, and d_{iB} between particle i and the beam (B). They are defined as:

$$d_{ij} = \min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right)\frac{\Delta R_{ij}^2}{R^2} \tag{4.4}$$

$$d_{iB} = \frac{1}{p_{T,i}^2} \tag{4.5}$$

where R is a tunable parameter that determines the size of the jets, in this case $R = 0.4$ for AK4 jets. The variable ΔR_{ij}^2 using the ϕ and rapidity y of a particle is given by:

$$\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \tag{4.6}$$

$$y = \frac{1}{2}\ln\frac{E + p_z}{E - p_z}. \tag{4.7}$$

The algorithm then starts by finding the smallest distance in the set of distances, and if it is a d_{ij} (rather than a d_{iB}) it combines them and uses the combined 4-vector to recompute all distances again. If the smallest distance found is a d_{iB}, that particle is removed and added to the list of final jets. This procedure is repeated until all particles have either been clustered or labeled as a final jet.

The jet collection that comes from the AK4 algorithm is further processed. Jets are required to satisfy $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$. These requirements are used to ensure the high quality of the jets, since jets can be more accurately reconstructed once they have a sufficient amount of p_{T} and are composed of PF candidates that have tracks in the tracking system. To avoid overlap between jets and leptons, a simple cleaning procedure is used, since the AK4 algorithm does not remove particles that have been identified as well-defined leptons. Any jet that is within $\Delta R = 0.4$ of an isolated lepton and whose transverse momentum is within 100% of the p_{T} of the lepton is removed from consideration. We further require jets to pass the jetID, which is defined by a combination of requirements on a list of variables related to each jet. This list of requirements is detailed in Table 4.1. The jetID variable is calculated for each jet and we veto the event if there is a single jet that fails its jetID.

4.3.1.4 b-tagging When looking at the objects in the jet collection there is no label that identifies what kind of jet it is, i.e., what type of particle produced it. We know that all quarks, other than the top quark, and all gluons produced by the hard processes will produce a jet in the event. However, it becomes difficult to classify the exact type of particle that produced a given jet.

When a bottom (b) quark is produced, it tends to hadronize into B mesons that have a lifetime long enough to travel a measurable distance. This means that if we see a jet that was produced at a different vertex than the primary vertex, there is a good chance that the jet contains a B meson that decayed after moving away from the primary vertex where it was produced by the bottom quark. Dedicated b-taggers try to take advantage of this property of bottom quarks by looking at the PF constituents within each jet to determine if any PF candidate has a secondary vertex. This information along with other properties of the jet is processed by a neural network to calculate a discriminator value which can be used to classify the jet as a b-jet [32,33]. For this analysis, we use a relatively new b-tagging algorithm based on a neural network developed for use in any CMS analysis called DeepCSV [34].

Figure 4.8 shows the receiver operating characteristics (ROC) curves for various b-tagging algorithms, which parametrically shows the efficiency of tagging a b-jet (probability of tagging a b-jet given that it actually is a b-jet) vs. the misidentification probability (probability of misidentifying a jet that does not originate from a b quark as a b-jet—also known as the fake rate). When using a tagger, we adopt a particular threshold value such that if a jet passes this threshold it is considered a b-jet and if it is below that value it is not considered a b-jet. The value that defines this threshold is called the working point (WP). Table 4.2 shows the values of the WP used for each year of data together with the efficiency and fake rate. The WP is chosen such that the fake rate is approximately 1%.

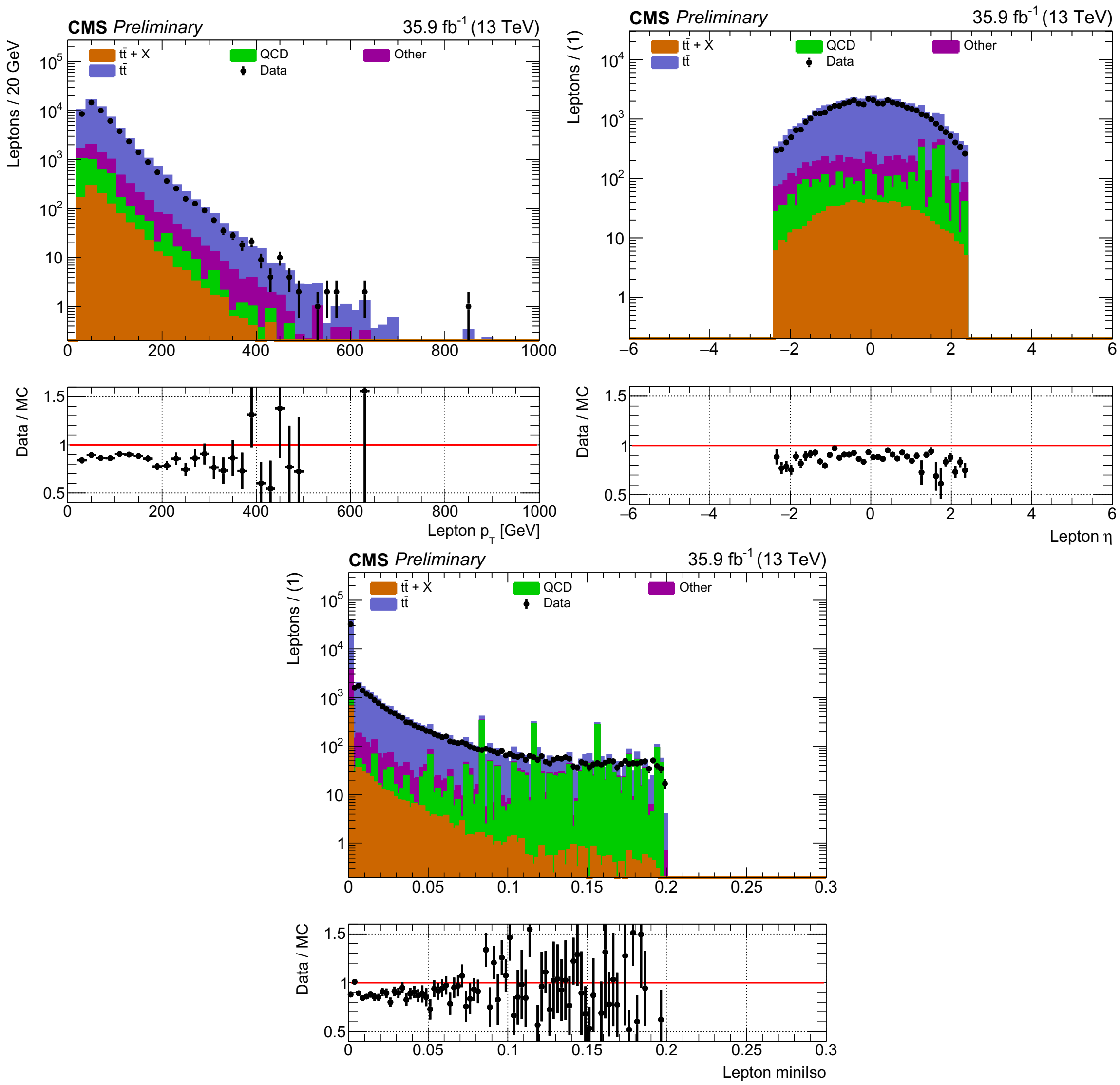


Figure 4.7: For 2016, the p_{T} spectrum (top left), η (top right), and miniIso (bottom) of muons in the lab frame that pass the muon selection are shown. Only events with exactly one of these muons are used, as these events pass the full baseline selection described in the next chapter. There is reasonable agreement between collision data and MC data with no obvious issues.

Table 4.1: Criteria for the AK4 CHS Jet ID variable

Region of $\|\eta\|$	Variable	Requirement 2016	Requirement 2017	Requirement 2018
$\|\eta\| < 2.4$	Charged hadron energy fraction	> 0	> 0	–
	Charged multiplicity	> 0	> 0	–
	Charged EM energy fraction	< 0.99	–	–
$\|\eta\| < 2.6$	Neutral hadron energy fraction	–	–	< 0.9
	Neutral EM energy fraction	–	–	< 0.9
	Number of constituents	–	–	> 1
	Charged hadron energy fraction	–	–	> 0
	Charged multiplicity	–	–	> 0
$2.6 < \|\eta\| < 2.7$	Neutral hadron energy fraction	–	–	< 0.9
	Neutral EM energy fraction	–	–	< 0.99
	Charged multiplicity	–	–	> 0
$\|\eta\| < 2.7$	Neutral hadron energy fraction	< 0.99	< 0.9	–
	Neutral EM energy fraction	< 0.99	< 0.9	–
	Number of constituents	> 1	> 1	–
$2.7 < \|\eta\| < 3.0$	Neutral EM energy fraction	> 0.01	> 0.02 and < 0.99	> 0.02 and < 0.99
	Neutral hadron energy fraction	< 0.98	–	–
	Neutral multiplicity	> 2	> 2	> 2
$\|\eta\| > 3.0$	Neutral EM energy fraction	< 0.9	< 0.9	< 0.9
	Neutral hadron energy fraction	–	> 0.02	> 0.02
	Neutral multiplicity	> 10	> 10	> 10

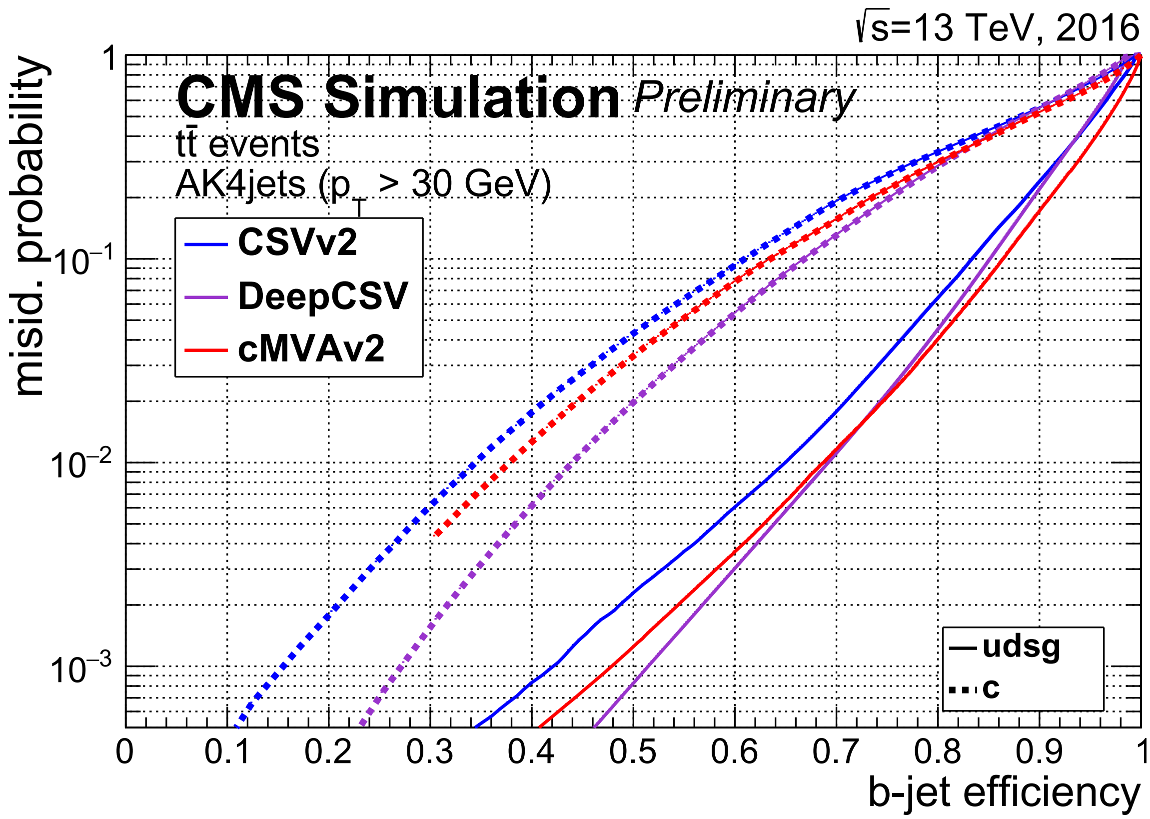


Figure 4.8: Performance of the widely accepted b taggers used in CMS compared to tagging jets from charm quarks and jets from other sources. The lines shown are for CSVv2, DeepCSV, and cMVAv2.

Table 4.2: Working points used for DeepCSV with the corresponding efficiency and fake rate.

Year	WP	Efficiency (%)	Fake rate (%)
2016	0.6321	≈68	≈1
2017	0.4941	≈68	≈1
2018	0.4184	≈68	≈1

CHAPTER FIVE

The Search for Scalar Top Quarks

"What do **you** think you are doing? ... Just do the right thing."

—Nathaniel "Joe" Pastika (Postdoc Extraordinaire)

5.1 Introduction

Although many searches for supersymmetry have been done at the LHC, no significant deviation from the Standard Model background has been observed, and strong limits on squark and gluino masses have been set for many simplified models. Nearly all traditional SUSY searches require the presence of substantial missing transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$) caused by undetected SUSY particles, usually the lightest neutralinos. Therefore, they would not be sensitive to well-motivated models of new physics that do not produce large $p_{\mathrm{T}}^{\mathrm{miss}}$, such as R-parity violating (RPV) [35] or Stealth SUSY [36–38], as discussed in Chapter Two.

This thesis is based on a novel search for top squark pair production in the unexplored final state with two top quarks, many extra light-flavor jets, and no $p_{\mathrm{T}}^{\mathrm{miss}}$. The search utilized a neural network (NN) and a data-driven background estimation method to determine the largest standard model background. The NN was trained using a novel adversarial technique to remove undesirable correlations between the NN score and the number of jets in the event (N_{j}). The specific signal models considered in the search are the following:

- RPV with UDD coupling, resulting in a final state with two top quarks and six light-flavor (quark) jets

- Stealth with SYY coupling, resulting in a final state with two top quarks and six gluon jets

These SUSY models are illustrated in Fig. 5.1. The search uses the data sets collected in 2016, 2017, and 2018, and covers the listed signal topologies with a single analysis.

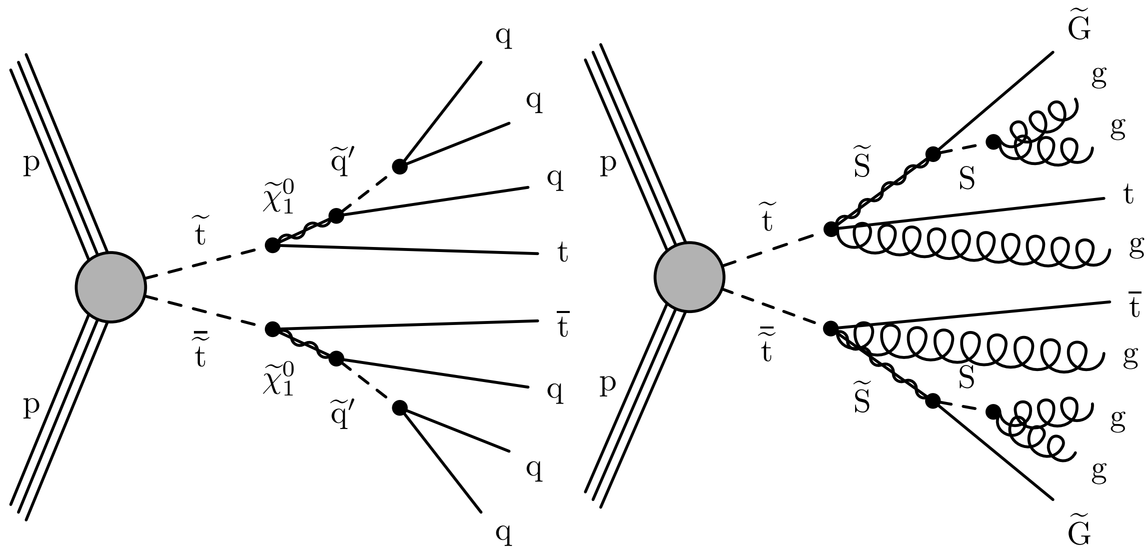


Figure 5.1: Top squark decay modes considered in this analysis. (Left) The RPV UDD model where the neutralino decays via an off-shell squark that then decays using the UDD interaction. (Right) The Stealth SYY model that has the top squark decay through effective stealth vertices to stealth particles that decay to gluons and a soft gravitino.

5.2 Analysis Strategy

The main distinguishing feature of the considered signal models from the SM backgrounds is their high jet multiplicity (N_{j}). The two-jet or three-jet resonances present in the signals are hard to exploit because of the large combinatorics and the probability of jets going outside of the detector acceptance. Therefore, the final state is essentially $\mathrm{t\bar{t}}+\mathrm{jets}$, without $p_{\mathrm{T}}^{\mathrm{miss}}$. Events are required to have at least seven jets, at least one b-tagged jet, and exactly one lepton. The presence of the lepton strongly reduces the background from QCD multijet production. In addition, the presence of the lepton is useful as a handle for selecting the triggers that were used in this analysis.

Since the jet multiplicity distribution is hard to model correctly with MC simulation at very large number of jets ($N_{\mathrm{j}} \geq 10$), we rely on data to fit the shape of the N_{j} distribution for the $\mathrm{t\bar{t}}$ background. At large jet multiplicity the ratio of the N_{j+1} distribution to the N_j can be described by a well-behaved function that is used to fit the N_{j} distribution directly from data. A signal region (SR) is the selection of data that is used to maximize the amount of possible signal compared to the SM backgrounds, while a control region (CR) is a particular selection of data that is similar to the SR but is not expected to have signal present. Ideally, we would fit the N_{j} shape in a dedicated CR, and then use it in our SR to estimate the $\mathrm{t\bar{t}}$ background; however, a signal-free control region that is also dominated by the $\mathrm{t\bar{t}}$ background is hard to construct. Therefore, we train a NN to distinguish signal from background using jet 4-momenta and event shape information. The network is explicitly set up to minimize dependence on N_{j}.

We construct four regions based on the output of the NN, where region 1 (D1) is highly dominated by background, and region 4 (D4) is the most signal-enriched. These four regions are then defined as D1, D2, D3, and D4. We then proceed to fit all four regions simultaneously, with D1 effectively acting as the sought-after control region, as illustrated in Fig. 5.2. The N_{j} shape is ensured to be the same in each region by the following procedure: (1) the NN is constructed to be independent of N_{j}; (2) residual shape differences are removed by deriving the divisions between the regions separately for every N_{j} bin, such that the background fraction is the same for every N_{j} bin, at least in simulation; (3) systematic uncertainties are assigned to cover for differences between collision data and simulation data resulting in a different N_{j} shape between the four regions.

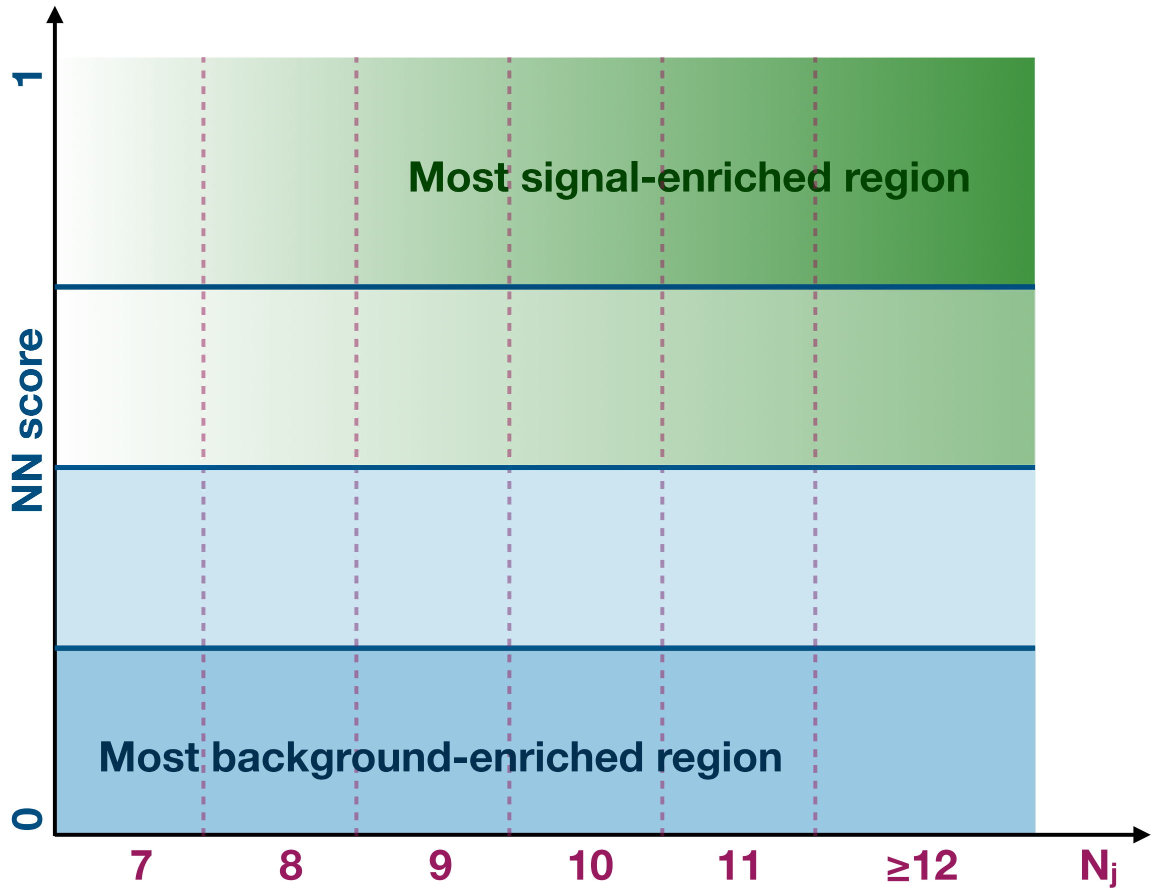


Figure 5.2: Signal region bin definition as a function of N_{j} and NN bins.

5.3 Event Selection

The event selection is performed in multiple stages. The first stage is at the trigger level. Due to the overwhelming number of proton-proton collisions delivered by the LHC to CMS, data are analyzed online by the trigger system and only "interesting" events are selected and saved for offline analyses. Offline, we then apply a baseline event selection that further filters out events not necessary for this analysis. Finally, all events are then analyzed for the 2016, 2017, and 2018 data-taking periods separately in great detail as discussed in the rest of this chapter. At the middle of the 2018 data-taking period, part of HE minus (HEM) readout electronics stopped functioning, which forces us to treat 2018 differently. The special treatment of 2018 results in the separation of the 2018 data into two data sets: (1) 2018 pre-HEM issue (2018pre) which consists of the data collected prior to the HEM issue, and (2) 2018 post-HEM issue (2018post) which consists of the data collected after the HEM issue.

5.3.1 Online Event Selection

There is an extensive number of trigger paths that were used by CMS in order to accommodate needs from a broad range of physics analyses. For this analysis, we use a data set that was created by taking the "or" of a few triggers designed to record events with an electron and another combination of triggers designed for events with a muon. These combinations of triggers define our single electron and single muon data sets as shown in Table 5.1.

For each single lepton data set, we need to measure the efficiency of the combination of triggers that we used so we can correct the MC data to represent this effect. When the MC data is simulated there is no explicit cut performed to remove events in the same manner that collision data would have, due to the inefficiencies of triggering. To model the trigger inefficiencies of collision data, we use a trigger emulation that does a reasonable job of removing events that would have been missed due to trigger inefficiencies. We then proceed to measure a simulation-to-data scale factor to correct for residual differences in trigger efficiencies between collision data

Table 5.1: Luminosities of CMS collected data sets used in this analysis, and the combinations of triggers used to collect the data. Note that the 2018 data are broken into two separate data sets as discussed in the text.

Data set	Triggers	Integrated luminosity (pb^{-1})
	2016	
Single electron	HLT Ele27 WPTight Gsf OR HLT Ele115 CaloIdVT GsfTrkIdT OR HLT Photon175	35908.886
Single muon	HLT IsoMu24 OR HLT IsoTkMu24 OR HLT Mu50 OR HLT TkMu50	35917.149
	2017	
Single electron	HLT Ele35 WPTight OR HLT Ele115 CaloIdVT GsfTrkIdT OR HLT Photon200	41525.529
Single muon	HLT IsoMu24 OR HLT IsoMu27 OR HLT Mu50	41525.250
	2018(pre and post)	
Single electron	HLT Ele35 WPTight OR HLT Ele115 CaloIdVT GsfTrkIdT OR HLT Photon200	20994.864 + 38648.875
Single muon	HLT IsoMu24 OR HLT IsoMu27 OR HLT Mu50	21071.298 + 38654.122

and simulated events. For the single electron, the efficiency is measured with the denominator containing events with exactly one electron, one muon with p_{T} greater than 40 GeV, and a selection very close to our baseline:

- All jets in the event passing JetID,
- H_{T} greater than or equal to 300 GeV,
- Number of jets in the event greater than or equal to 5,
- At least one b-tagged jet,
- The invariant mass between the lepton and b-tagged jet in the event between 50 and 250 GeV. If there is more than one b-tagged jet in the event, the invariant mass of each b-tagged jet and the lepton is considered, and only one combination has to meet this criterion.

For the numerator, we require the same selection as the denominator in addition to passing the electron trigger. After making the above efficiency plot binned by electron

p_{T} for both the single muon data set and the simulated samples, we obtain a scale factor by dividing the former by the latter. The trigger efficiencies for the simulated data binned in electron p_{T} and η are shown in Figs. 5.3 and 5.4. The scale factor between data and simulation is determined as a function of electron p_{T} and η as shown in Fig. 5.5.

We then repeat the same steps but for the single muon data set. We use the same selection except that we look at the single electron data and require one electron with p_{T} greater than 40 GeV and one muon with the same selection. In this case, we bin by the muon p_{T} and η; the plots are shown in Figs. 5.6 and 5.7. The scale factor between collision data and simulation was determined as a function of muon p_{T} and η, and is shown in Fig. 5.8.

5.3.2 Baseline Selection

The baseline event selection is motivated by both the requirements of the single lepton triggers that are used for the online data selection and also to reduce backgrounds, in particular those from QCD multijet production. The selection is not very complicated compared to those used in other SUSY analyses. There is only one selection used to define the signal region (SR). The baseline selection is as follows:

- Exactly one electron or muon: the lepton comes from the $\mathrm{t\bar{t}}$ decay for our signal, where where one top quark yields a W boson that decays into quarks and the other top quark's W boson decays into a lepton plus a neutrino. This is not the largest branching fraction for $\mathrm{t\bar{t}}$, but this choice was motivated to greatly reduce the QCD multijet background. QCD multijet events in general do not have high-p_{T} isolated leptons.

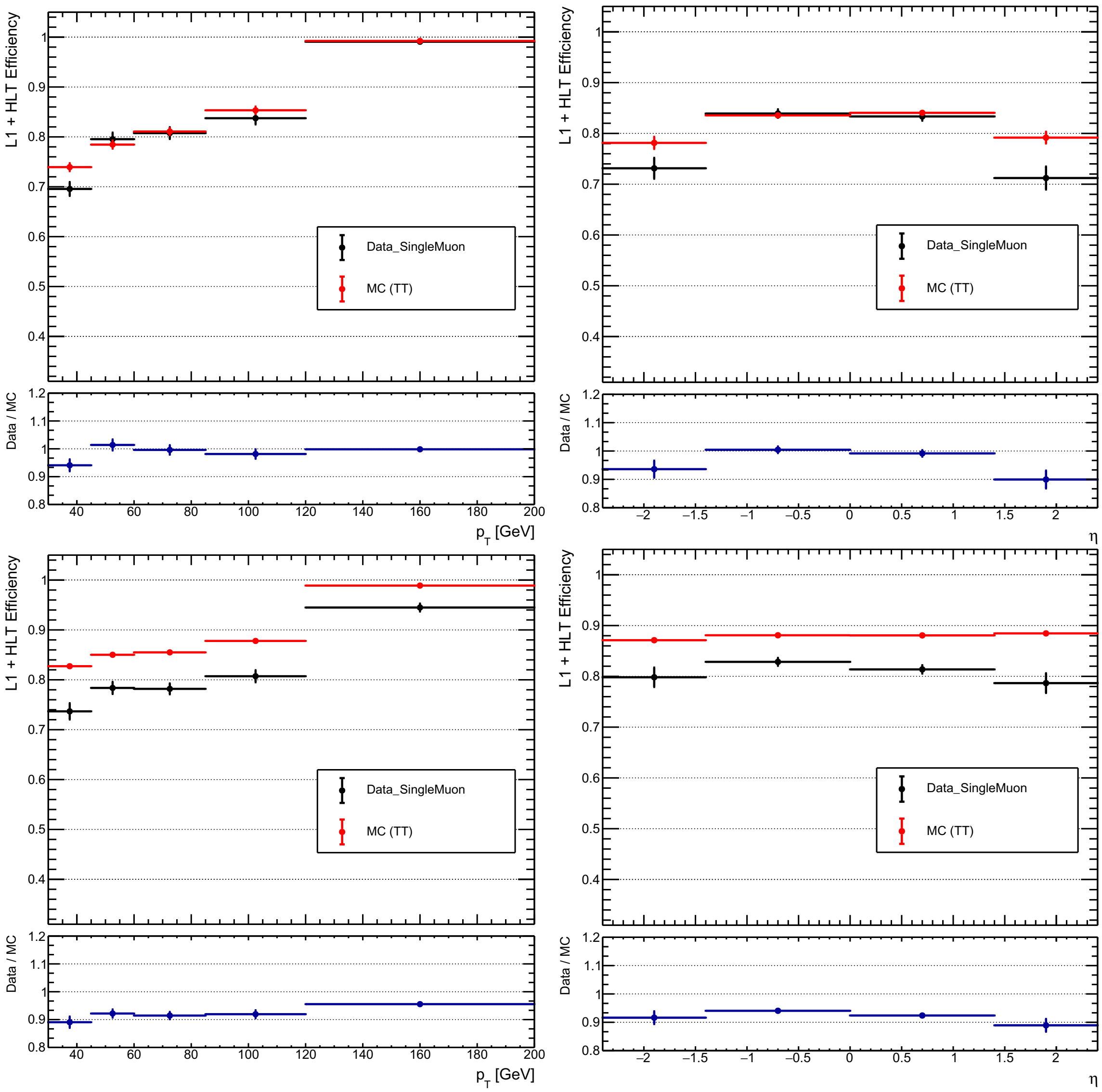


Figure 5.3: Electron trigger efficiency in data and simulation for 2016 (top) and 2017 (bottom) as a function of electron p_{T} (left) and η (right).

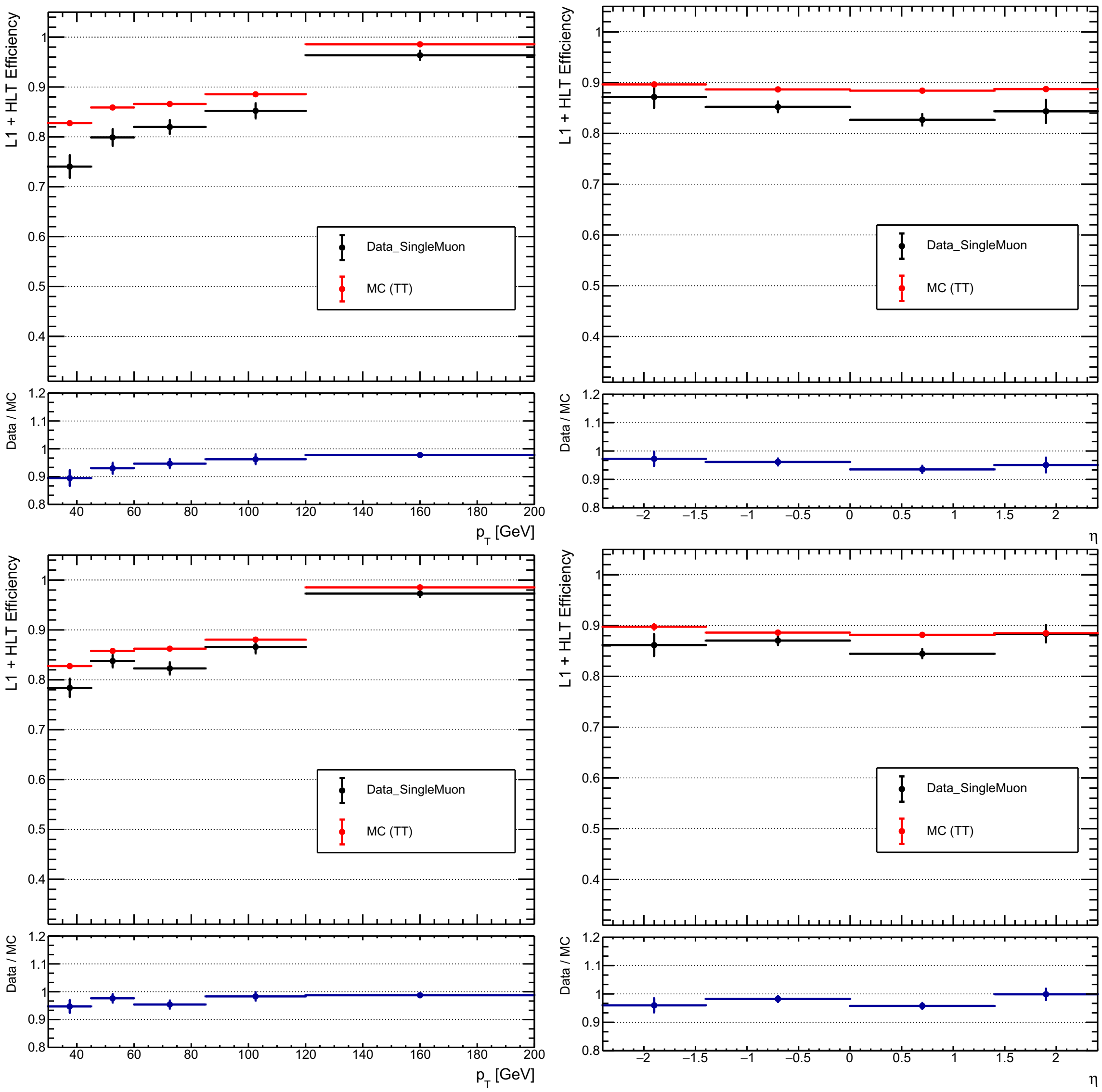


Figure 5.4: Electron trigger efficiency in data and simulation for 2018pre (top) and 2018post (bottom) as a function of electron p_{T} (left) and η (right).

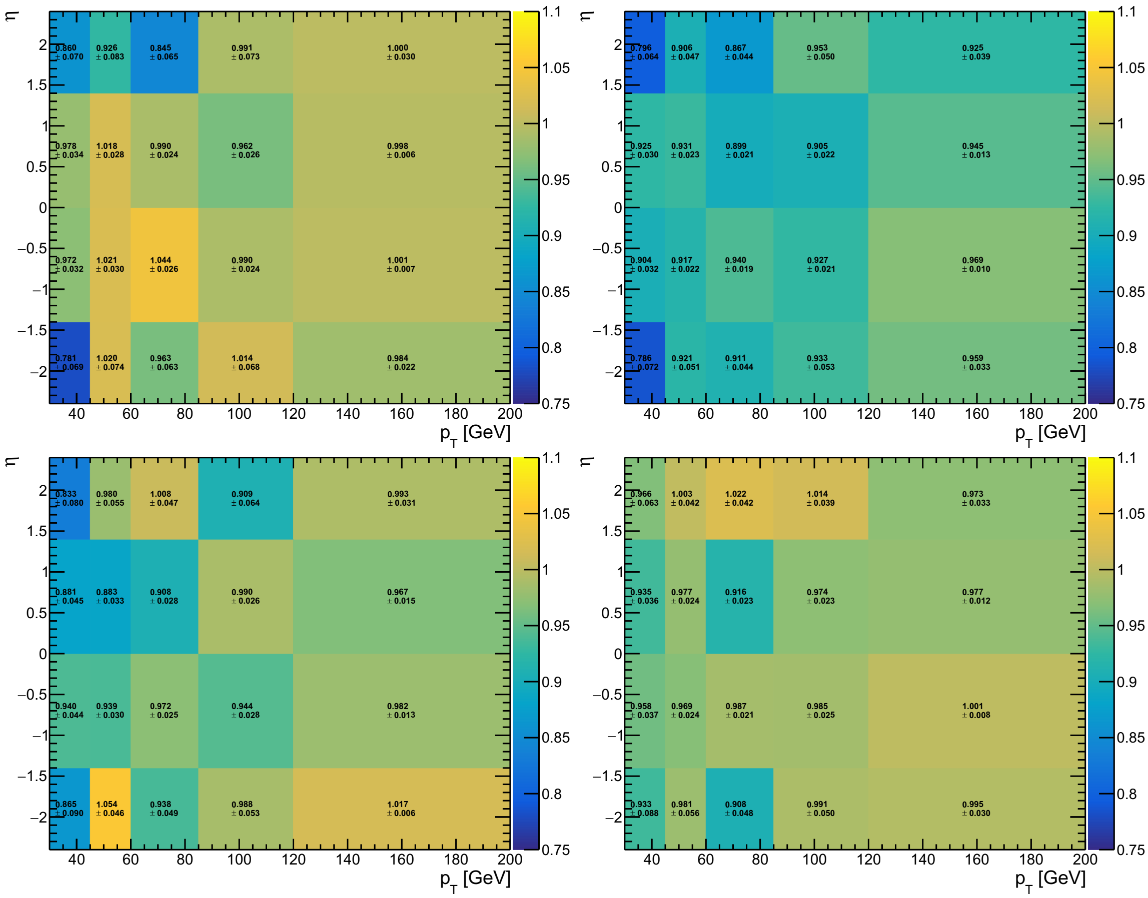


Figure 5.5: The electron trigger efficiency scale factor as a function of p_{T} and η is shown for 2016 (top left), 2017 (top right), 2018pre (bottom left), and 2018post (bottom right).

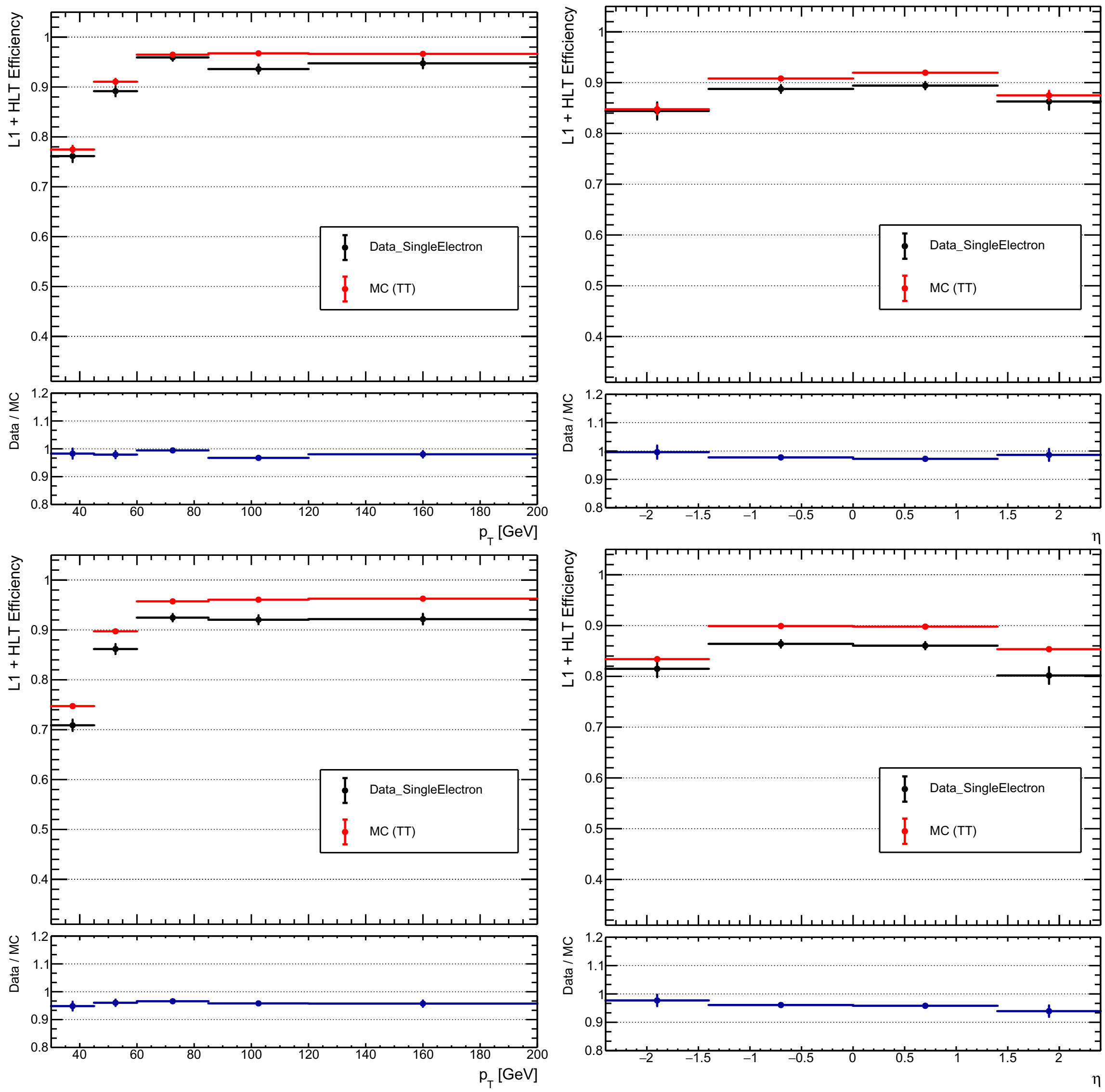


Figure 5.6: Muon trigger efficiency in data and simulation for 2016 (top) and 2017 (bottom) as a function of muon p_T (left) and η (right).

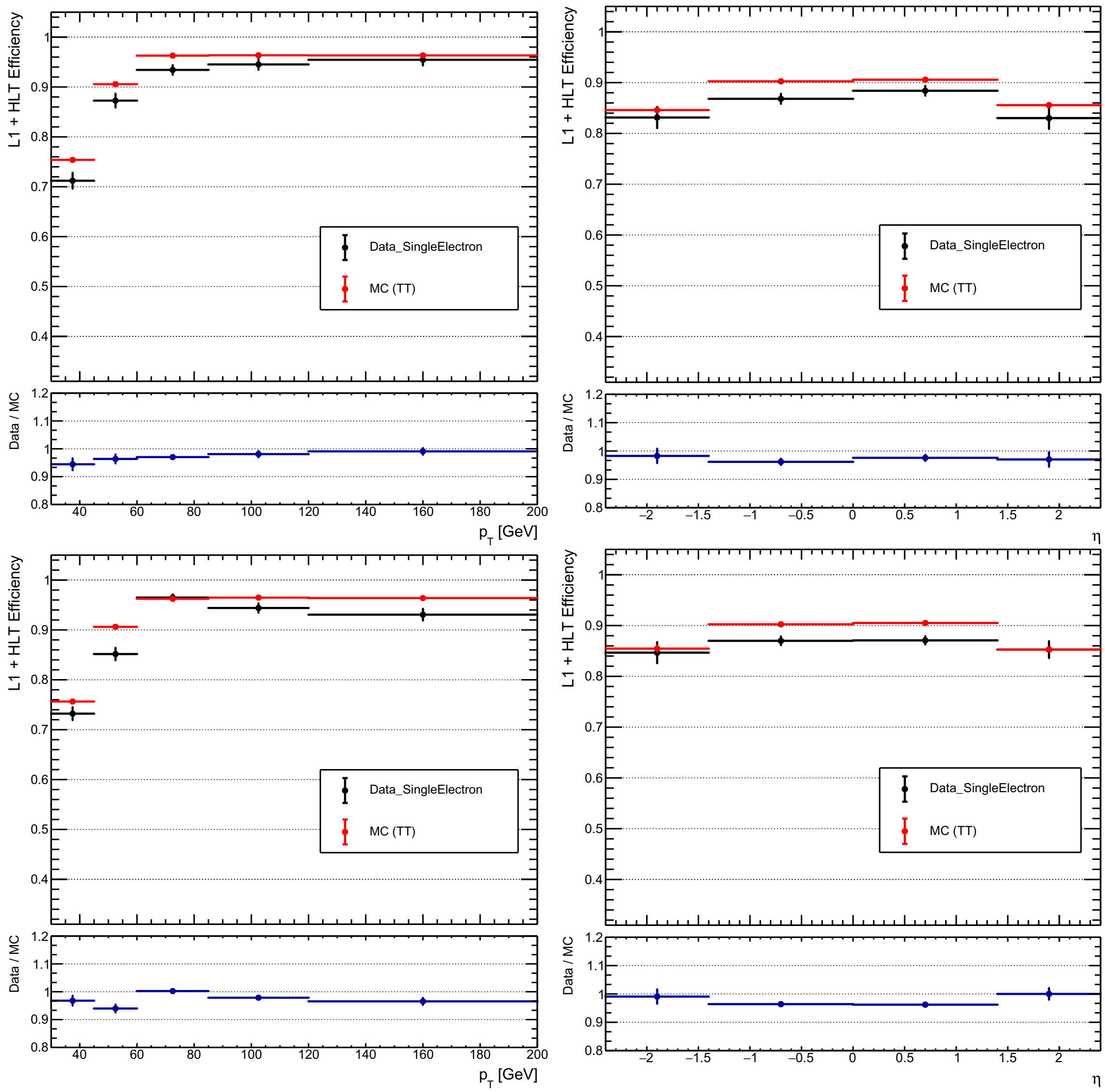


Figure 5.7: Muon trigger efficiency in data and simulation for 2018pre (top) and 2018post (bottom) as a function of muon p_{T} (left) and η (right).

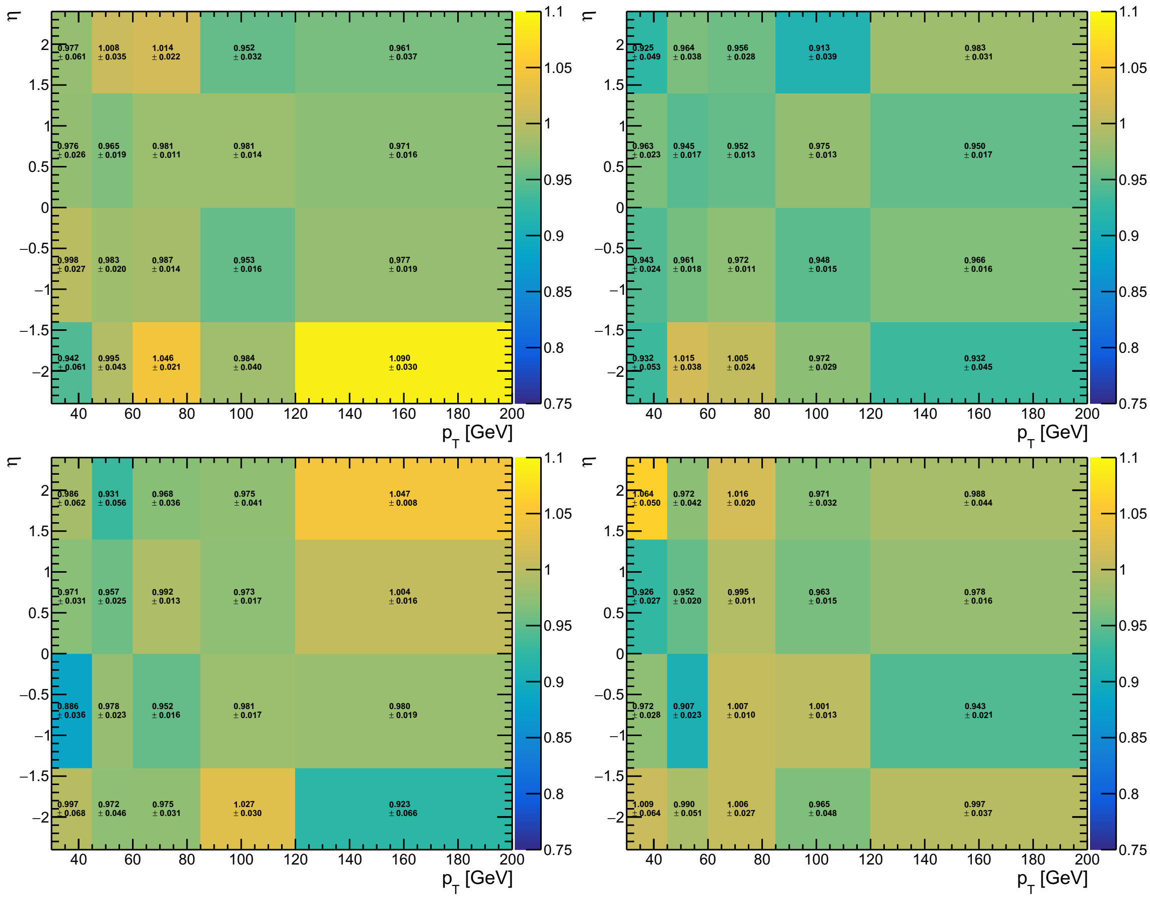


Figure 5.8: The muon trigger efficiency scale factor as a function of p_{T} and η is shown for 2016 (top left), 2017 (top right), 2018pre (bottom left), and 2018post (bottom right).

- $N_{\mathrm{j}} \geq 7$: this choice is motivated by our signal models having a large number of produced jets. The background estimation for $\mathrm{t\bar{t}}$ also requires that the N_{j} distribution be a falling spectrum. If events with less than seven jets are included, the N_{j} spectrum would no longer match the functional form used for the $\mathrm{t\bar{t}}$ background estimation. More details on this are discussed later in this chapter.
- $N_{\mathrm{b}} \geq 1$: for $\mathrm{t\bar{t}}$ decay we expect exactly two b-jets. However, tagging all the b-jets in an event can be difficult. Instead of requiring exactly two b-jets, a looser requirement of one or more b-jets is used which leads to good acceptance of signal events while still rejecting reducible backgrounds. The requirement of at least one b-jet also allows for the derivation of other useful search variables.
- $50 < M(l, b) < 250$ GeV, where $M(l, b)$ is defined as the invariant mass of the lepton and a b-jet. This requirement is based on the expectation that the b-jet and the lepton in the event come from the same top quark. If there are multiple b-jets, at least one combination in the event is required to meet this requirement.
- $H_{\mathrm{T}} > 300$ GeV: this is motivated to reduce some QCD multijet backgrounds and to ensure that we are not susceptible to missing low energy MC data. H_{T} is a falling distribution, which means we can allow for orders of magnitude more events the lower the H_{T} cut is. This also means we would need an order of magnitude more simulated events to describe these events. Since the signals do not produce events with low H_{T} it becomes logistically favorable to have an H_{T} cut.

The $M(l, b)$ distributions are shown in Fig. 5.9. The idea behind the $M(l, b)$ cut is to require a loose leptonic top quark tag so that selected events would have the lepton and the b-jet coming from the same top quark decay. Figure 5.10 shows the N_{j} distribution for the total background and a few selected signal models. From this figure, it is clear that the signal populates the tail of the jet multiplicity distribution, and that the main background process after the baseline selection is $\mathrm{t\bar{t}}$+jets. All other backgrounds make up around 10% of the total background. To further enhance the signal component and provide a $\mathrm{t\bar{t}}$ background estimation method, a neural network (NN) is trained. Details of this NN are given in Section 5.4.

During 2018 data taking in late June, two neighboring readout boxes (RBXs) in the $-z$ endcap, HEM15 and HEM16, were lost due to a commercial power supply failure. The HEM failure motivates the splitting of the 2018 data set into two parts, the pre-HEM failure (2018pre) and post-HEM failure (HEM). The specific η-ϕ region covered by these two RBXs corresponds to $\eta \in (-3.00, -1.30)$ and $\phi \in (-1.57, -0.87)$. Having no input from HCAL in this region can result in an increase to the number of fake leptons, and the mismeasurement of jet energy, among other issues. The mismodeling of events motivated an additional requirement to be applied to the 2018post data set. In general, if an object is reconstructed to be in the vicinity of the HEM failure region, the entire event is vetoed. Table 5.2 shows the definition of this veto.

5.3.3 H_{T} Reweighting

Upon investigating the agreement between collision data and MC data for the low jet multiplicity bins ($\leq$7), a trend in the ratio of collision data to MC data for

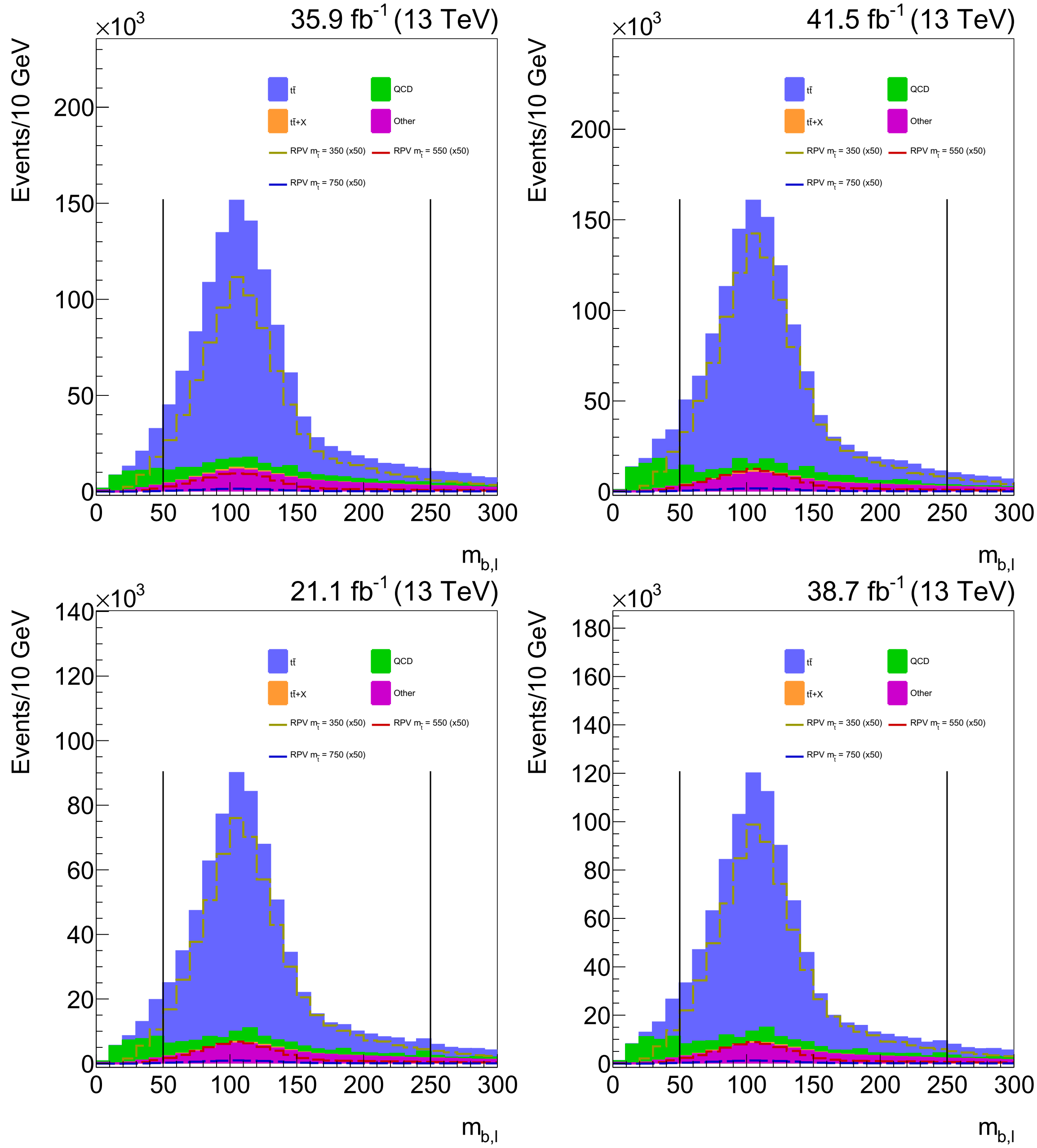


Figure 5.9: Distributions of $M(l, b)$ in simulated backgrounds for 2016 (top left), 2017 (top right), 2018pre (bottom left) and 2018post (bottom right). The distributions for the signal have been scaled by 50× for comparison.

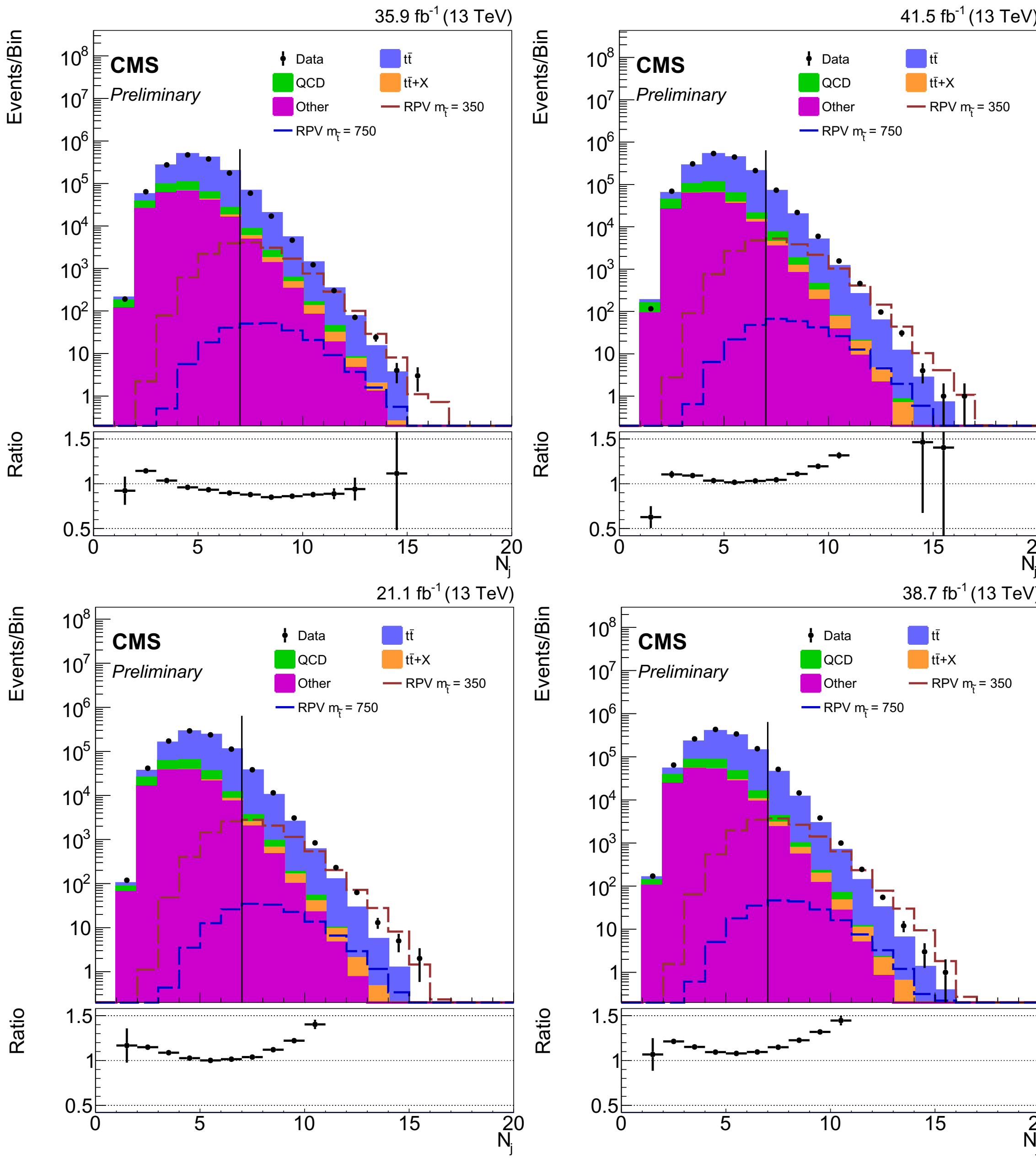


Figure 5.10: The N_{j} distribution for the four different background processes, overlaid with a few selected signal models. The simulation is normalized to 35.9 fb^{-1} for 2016 (top left), to 41.5 fb^{-1} for 2017 (top right), to 21.1 fb^{-1} for the 2018 data set prior to the HEM issue, and to 38.7 fb^{-1} for the 2018 data set after the HEM issue.

Table 5.2: The exact η, ϕ and p_{T} requirements for jets and leptons that are used to define the HEM event veto.

Jet	$\eta \in (-3.20, -1.10)$	AND	$\phi \in (-1.77, -0.67)$	AND	$p_{\mathrm{T}} > 20.0$	OR
Muon	$\eta \in (-3.00, -1.30)$	AND	$\phi \in (-1.57, -0.87)$	AND	$p_{\mathrm{T}} > 20.0$	OR
Electron	$\eta \in (-3.00, -1.30)$	AND	$\phi \in (-1.57, -0.87)$	AND	$p_{\mathrm{T}} > 20.0$	

the H_{T} distribution is apparent as shown in Fig. 5.11. The N_{j} variable is correlated with H_{T}. It is expected then that any H_{T} disagreement will induce disagreements in the N_{j} shape. This motivates the need for a scale factor to correct for the H_{T} ratio of collision data to MC data disagreement. An H_{T}-dependent scale factor is derived by performing a fit with an exponential function

$$SF_{H_{\mathrm{T}}}(H_{\mathrm{T}}, N_{\mathrm{j}}) = Ae^{kH_{\mathrm{T}}} \tag{5.1}$$

to the ratio of the H_{T} distributions for collision data and MC data. A and k are free parameters that are functions of N_{j}. The fit is repeated for multiple N_{j} bins in order to find a trend in the fit parameters as a function of N_{j} as shown in Fig. 5.12. Using this fitted trend, the scale factor can be extrapolated to higher N_{j} bins. A comparison between collision data and MC data for the H_{T} distributions is shown before and after the H_{T} reweighting in Fig. 5.13 for events with exactly seven jets.

5.4 $\mathrm{t\bar{t}}$ *Background Estimation using a Neural Network*

With the absence of an obvious control region for the dominant background ($\mathrm{t\bar{t}}$), an effective control region using the discriminant of a signal vs. background NN, as discussed in Section 5.2, is constructed. The NN discriminant value should not be correlated with N_{j} because it is important for the N_{j} shape to remain the same for every NN bin. If the N_{j} shape changes between the different NN bins then

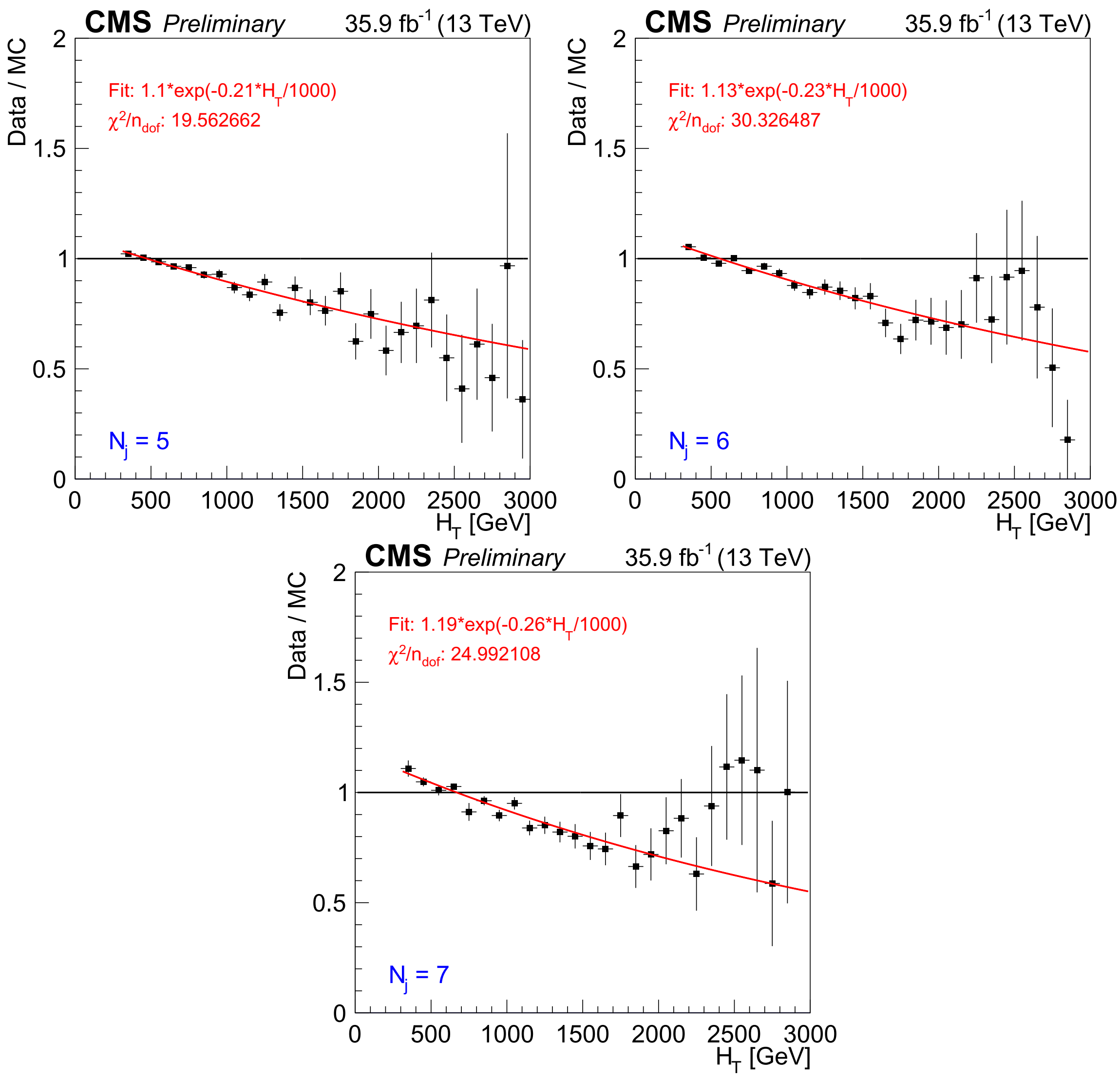


Figure 5.11: Ratio of data over simulation as a function of H_{T} for a selection requiring exactly five (top left), six (top right), and seven (bottom) jets.

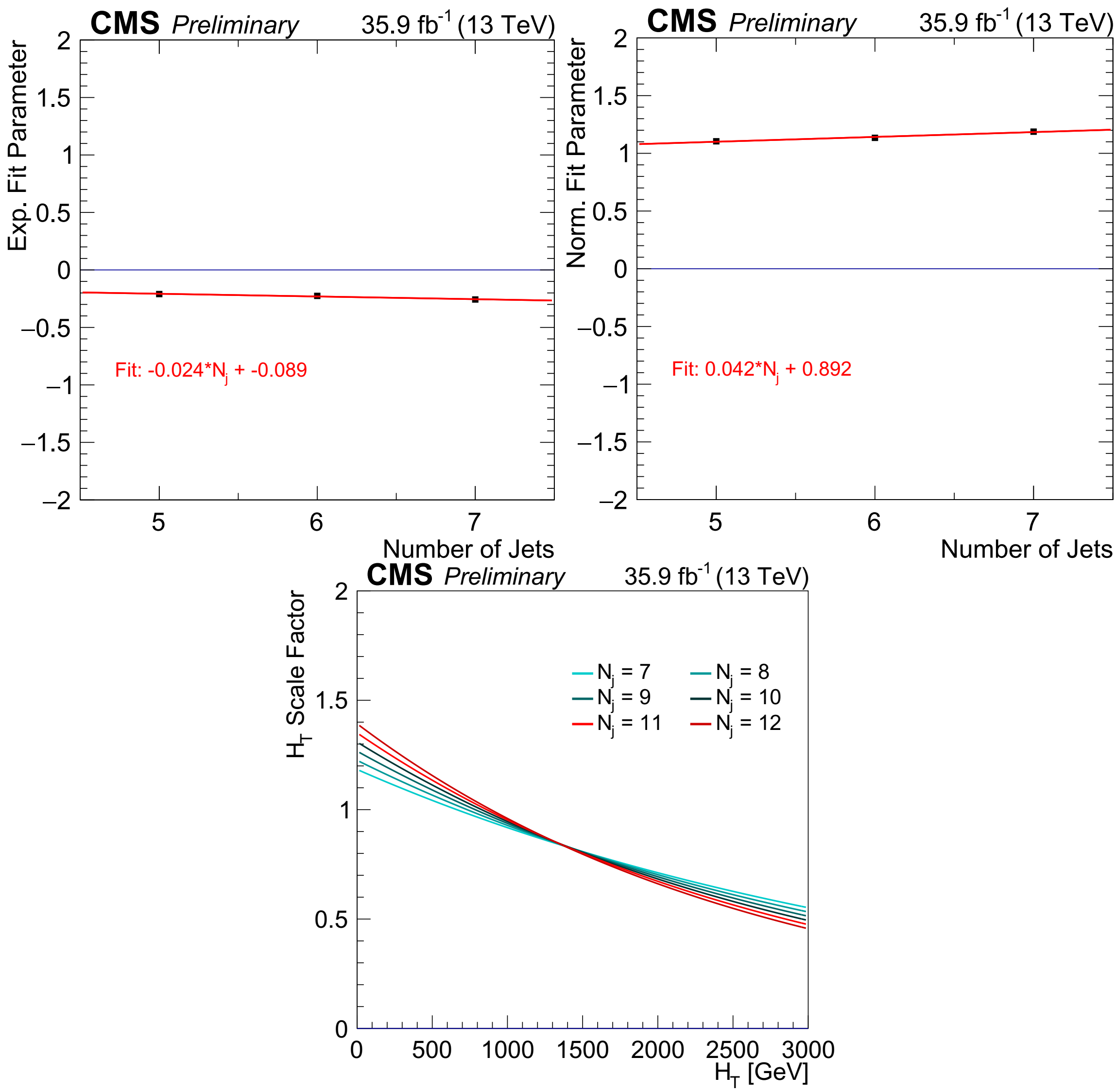


Figure 5.12: Derivation of the N_{j} dependent H_{T} scale factor. (Top left) Trend in the exponential parameter. (Top right) Trend in the normalization parameter. (Bottom) Resulting scale factor as a function of H_{T} and N_{j}.

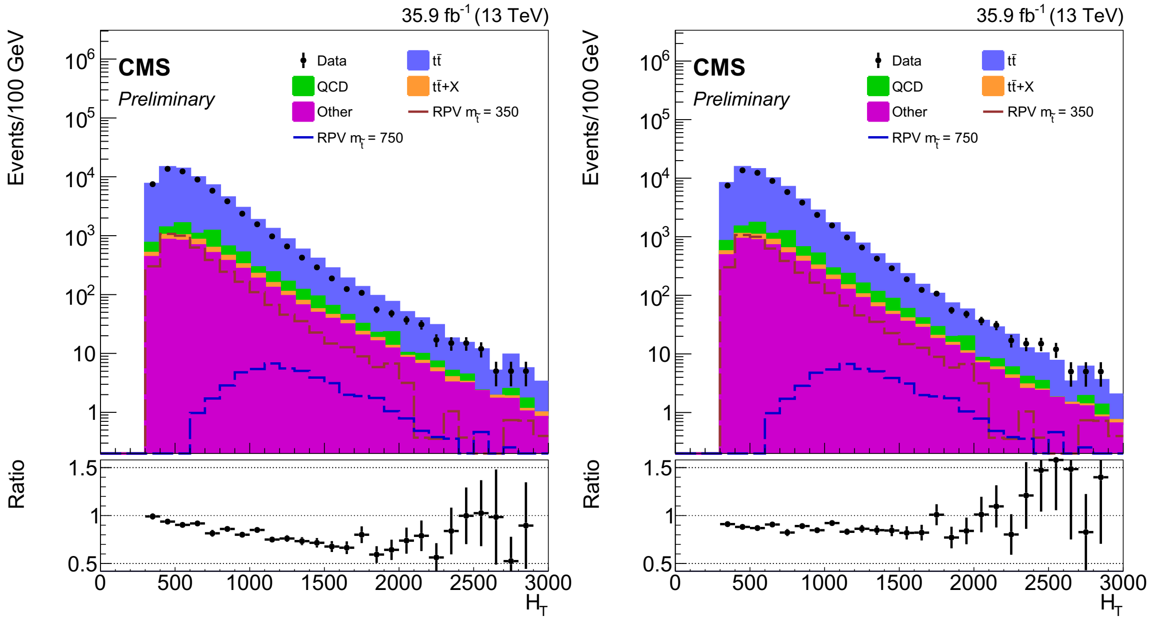


Figure 5.13: Data vs. MC data comparisons for our baseline with N_{j} exactly equal to 7 in 2016. (Left) H_{T} without the H_{T} scale factor. (Right) H_{T} with the H_{T} scale factor.

the shape obtained from a fit in the signal-depleted region is not applicable to the signal-enriched region.

Our first attempt to discriminate the signal from background events was to apply a boosted decision tree (BDT) using some common event shape variables derived using the jets in the event. However, this proved to be too strongly correlated with N_{j}, primarily because the best discriminating variable between signal and $\mathrm{t\bar{t}}$ production is N_{j}. We then proceeded to use a Keras-based NN [39] because we could repurpose the developments for adversarial gradient reversal (GR) training techniques [40] to remove N_{j} dependence from the model while it is being trained.

5.4.1 Background of Neural Network Approach

At the end of the day, a NN is just a function. When people say that they are going to train a NN, what they really mean is they are going to take a function with a large number of free parameters and perform a fit to some data. The function can have an arbitrary number of inputs, free parameters, and outputs, which makes it ideal for situations where writing an exact solution is difficult or impossible. The real art of training a NN is to find out how to build this arbitrary function and to identify a metric to optimize when determining the free parameters of the function.

Using calculus, if we would like to write a completely arbitrary function that takes one variable, has an arbitrary number of free parameters, and yields one output variable, a simple Taylor series such as

$$f(x) = \sum_{n=0}^{\infty} \frac{(x-a)^n}{n!} \frac{d^n f(a)}{dx^n} \tag{5.2}$$

can be used. We could simply stop after taking the first two terms and perform a linear fit, or keep one thousand terms and use a χ^2 minimization metric. It just boils down to what works for the given data and what is computationally feasible. A Taylor series works well for situations with one input and one output, but what about an arbitrary number of inputs and outputs? There are plenty of ways to build a function to do this, but at the end of the day one needs to be able to train it and use it. A NN is one computationally light and user friendly function that achieves this goal.

A NN is first built by taking a linear combination of the inputs and then passing that combination into an activation (simple non-linear) function. A common choice for the activation function is the rectified linear unit (ReLU) function, which is given by

$$\mathrm{relu}(x) = \max(0, x). \tag{5.3}$$

Given an input set $x_k^{(1)}$, where k is an index over inputs, and an output set $x_j^{(2)}$ where j is an index over outputs, a simple model can be written as

$$x_j^{(2)} = \sigma^{(1)} \left(\sum_k w_{jk}^{(1)} x_k^{(1)} + b_j^{(1)} \right) \tag{5.4}$$

where $\sigma^{(1)}(x)$ is the activation function, $w_{jk}^{(1)}$ is a matrix of the weights which are free parameters, and $b_j^{(1)}$ is an arbitrary bias that is added to the linear combination. By taking an arbitrary number of linear combinations of the set $x_j^{(2)}$, a new set $x_i^{(3)}$ can be formed with a similar transformation by

$$\begin{aligned} x_i^{(3)} &= \sigma^{(2)} \left(\sum_j w_{ij}^{(2)} x_j^{(2)} + b_i^{(2)} \right) \\ &= \sigma^{(2)} \left(\sum_j w_{ij}^{(2)} \left[\sigma^{(1)} \left(\sum_k w_{jk}^{(1)} x_k^{(1)} + b_j^{(1)} \right) \right] + b_i^{(2)} \right) . \end{aligned} \tag{5.5}$$

The addition of these linear combination transformations can be repeated until the NN is sufficiently complicated, usually referred to as making the NN "deep." A NN of this type can be written in a more general form for $n-1$ transformations and $n-2$ "hidden" layers as

$$x_i^{(n)} = \sigma^{(n-1)} \left(\sum_j w_{ij}^{(n-1)} \left[\sigma^{(n-2)} \left(\sum_k w_{jk}^{(n-2)} [...]_k + b_j^{(n-2)} \right) \right] + b_i^{(n-1)} \right) . \tag{5.6}$$

A NN that can be described by Eq. (5.6) is called a fully connected or dense NN. As the new and exciting field of machine learning continues to grow, many different ways of writing NN models will be developed and used.

The next step is to decide on a metric for goodness of fit, or in more commonly used terms, to develop an appropriate loss function. A loss function is a mapping of input data to a real value that, when minimized, optimizes the performance of the NN. A common loss function for a regression NN, which predicts a value that can be any real number in the open set $(-\infty, \infty)$, is the mean square error given by

$$L = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_i)^2 \tag{5.7}$$

where n is the number of data points, y_i is the predicted value for the i^{th} data point, and $\bar{y}_i$ is the true value for the i^{th} data point. As the difference between the predicted and true value gets smaller, the performance of the NN is improved. The NN used in this analysis is a classifier, which outputs a value from the closed set $[0, 1]$ to predict what class an item belongs to. For this NN a "0" indicates that the event is from background ($\mathrm{t\bar{t}}$) and a "1" indicates that the event is from signal. If the network is performing poorly it usually gives a value of 0.5, which means that it could not decide between background or signal. The loss function normally used for a classifier is the cross entropy

$$L = -\sum_{i=1}^{n}\bar{y}_i \log(y_i), \tag{5.8}$$

where n is the number of data points, y_i is the predicted value for the i^{th} data point, and $\bar{y}_i$ is the true value for the i^{th} data point.

5.4.2 Input Variable Selection

The input variables for the training are selected to maximize the information about the event, other than N_{j}, that could be used to discriminate between signal and background ($\mathrm{t\bar{t}}$). The input variables are derived after boosting the jets and lepton to the center-of-mass frame defined by the lepton and all jets with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 5$ in the event. Only the top seven jets are used as NN inputs even for events with more than seven jets because we did not want to introduce N_{j} dependence in these variables, as N_{j} is not constant from event to event. The input variables can be thought of as low- and high-level variables and are shown in Table 5.3. One set of high-level variables chosen for this analysis consists of the Fox-Wolfram moments (FWM) [41]

$$H_l = \sum_{i,j} |p_i||p_j| P_l(\cos\theta_{ij}) \tag{5.9}$$

where P_l is the l^{th} Legendre polynomial, p_i is the momentum of the i^{th} jet, and θ_{ij} is the angle between the i^{th} and j^{th} jet. Another set of high-level variables is the jet momentum tensor eigenvalues (JMTE) [42] defined using the JMT matrix

$$S^{\alpha,\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_j |p_j|^2} \tag{5.10}$$

where $\alpha, \beta = x, y, z$ and p_i^α is the momentum component of the i^{th} jet. The FWM variables have been used to define the event "shape," classifying the "flow" of energy in an event. The eigenvalues of Eq. (5.10) can be used to define the sphericity of the event. The low-level variables are simply the 4-vectors of the lepton and seven highest momentum jets in the event. Figures 5.14 and 5.15 show distributions of a subset of the NN input variables for the $\mathrm{t\bar{t}}$ background and signal events of the 2016 and 2017 scenarios, respectively.

Table 5.3: Summary of neural network input variables

Level	Name	Number of Variables
Low level	Jet p_{T} Jet η Jet ϕ Jet mass	Top momentum jets 1–7
	Lepton p_{T} Lepton η Lepton ϕ Lepton mass	One electron or muon
High level	Fox-Wolfram moments	2–5
	Jet momentum tensor eigenvalues	1–3

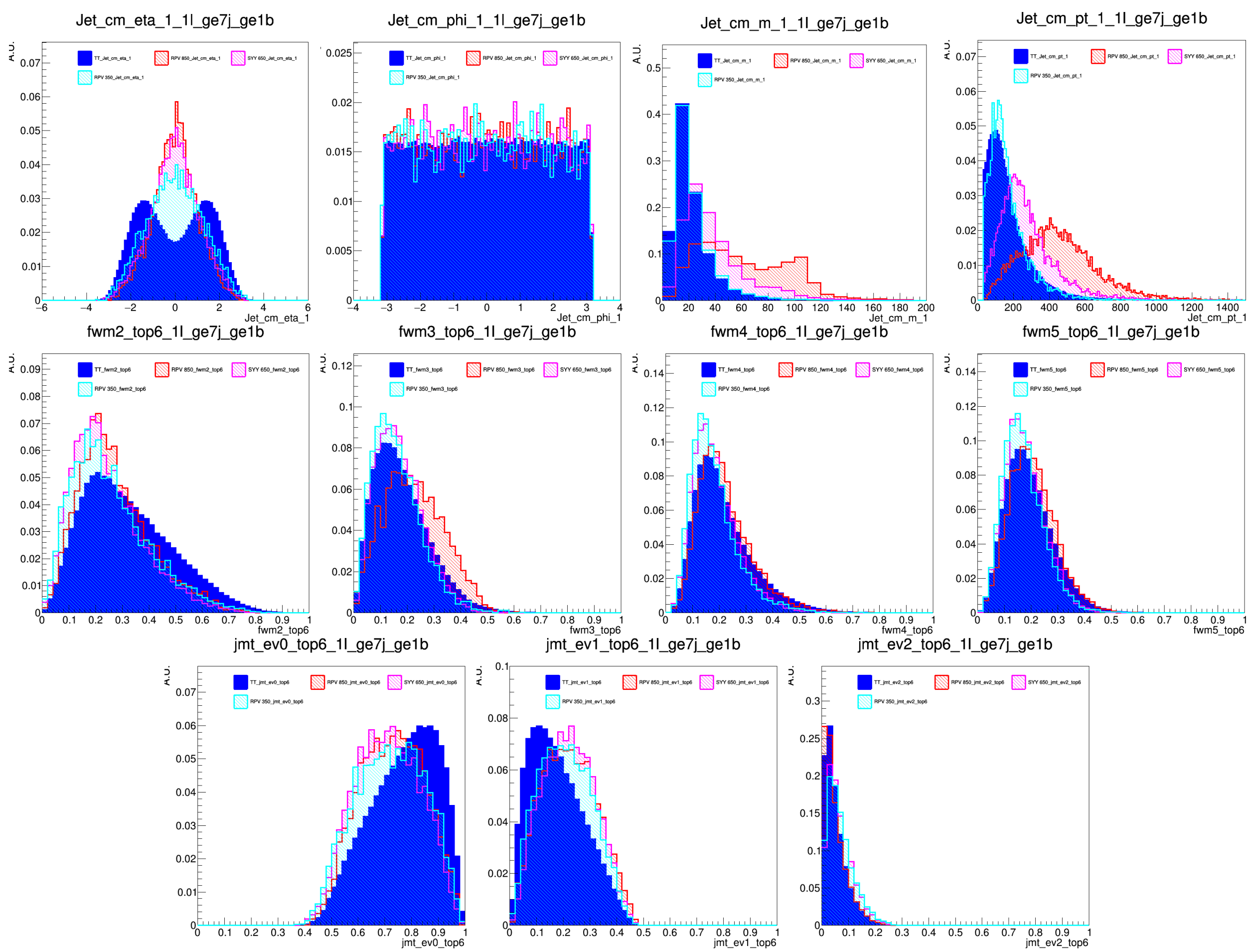


Figure 5.14: Distributions of normalized shape of NN input variables for $t\bar{t}$ and some signal models of the 2016 scenario: jet η (top left), jet ϕ (top middle left), jet mass (top middle right), jet p_{T} (top right). (Middle row) Fox-Wolfram moments for moments 2, 3, 4, and 5, respectively. Jet momentum tensor eigenvalue 1 (bottom left), jet momentum tensor eigenvalue 2 (bottom middle), jet momentum tensor eigenvalue 3 (bottom right).

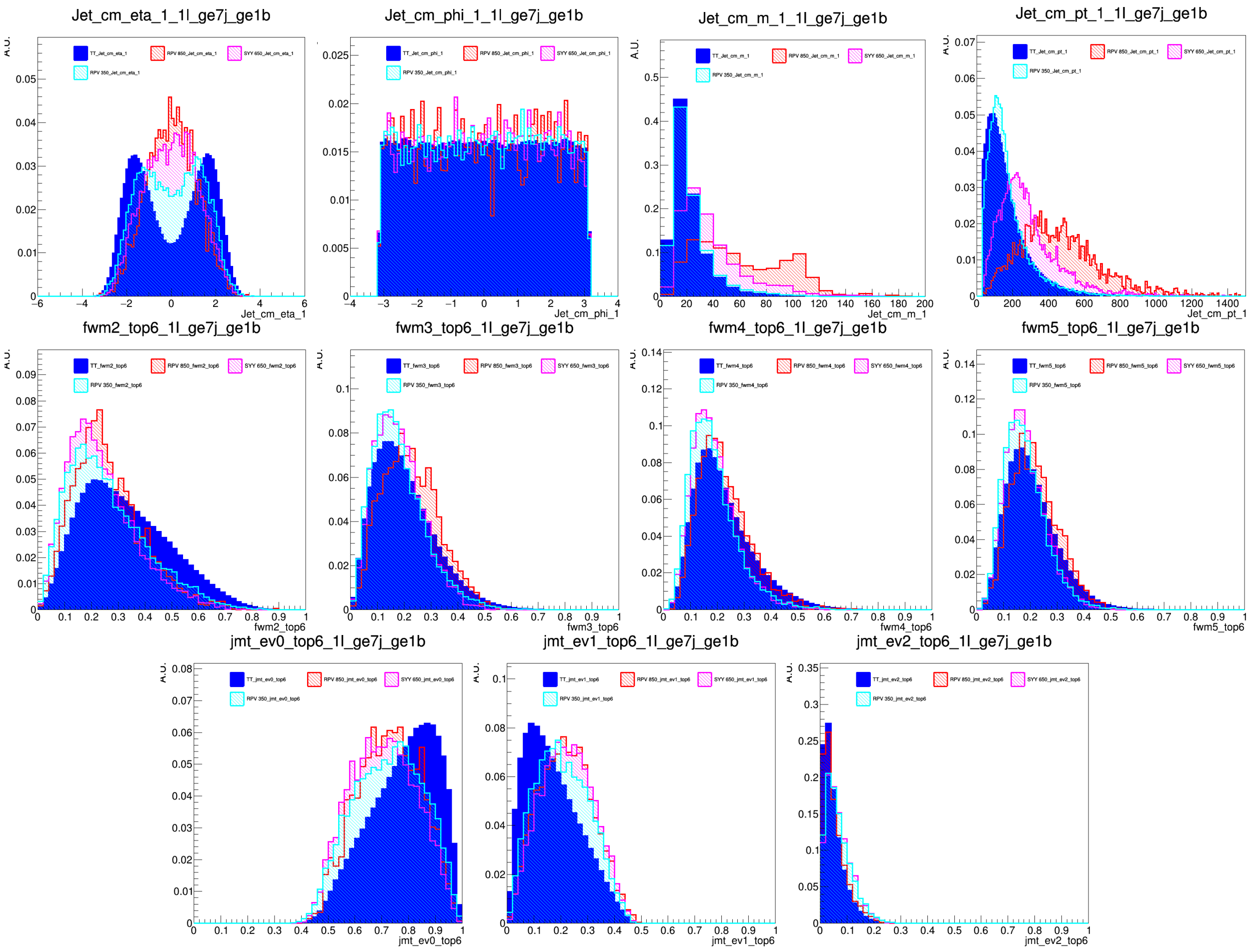


Figure 5.15: Distributions of normalized shape of NN input variables for $t\bar{t}$ and some signal models of the 2017 scenario: jet η (top left), jet ϕ (top middle left), jet mass (top middle right), jet p_{T} (top right). (Middle row) Fox-Wolfram moments for moments 2, 3, 4, and 5, respectively. Jet momentum tensor eigenvalue 1 (bottom left), jet momentum tensor eigenvalue 2 (bottom middle), jet momentum tensor eigenvalue 3 (bottom right).

5.4.3 Neural Network Training

The training is done using code based on the Keras with Tensorflow backend Python packages [39, 43]. We performed a supervised training on a binary classifier. Supervised training implies that the NN is informed about whether each event is background or signal. An example of a fully connected binary classifier is shown in Fig. 5.16.

The events used for training are required to pass the baseline selection defined in Section 5.3. $\mathrm{t\bar{t}}$ MC data samples generated using both POWHEG and MadGraph are used for the background, and all of the available signal MC data are used for the signal. During training, each event is labeled such that signal corresponds to (1) and background corresponds to (0). We also label each event based on how many jets are in the event, using five classes (7, 8, 9, 10, and 11 or more jets); for example an 8-jet event is labeled $(0, 1, 0, 0, 0)$. Each event is labeled as signal or background together with the number of jets in the event. The simplest possible network architecture is chosen as illustrated in Fig. 5.17, where a simple network only has the input, fully connected, dropout, and output layers. The NN is effectively two simple networks connected in series, where the second network punishes the first network for giving it useful information, in this case, to determine N_{j}. The mechanism for the second network to punish the first network comes from its contribution to the total loss function, where a negative hyperparameter (λ) is introduced as

$$L_{\mathrm{total}} = L_{\mathrm{first}} - \lambda L_{\mathrm{second}}. \tag{5.11}$$

When the training code minimizes the total loss function, it maximizes the part of the loss function that classifies N_{j}. The technique of having a part of the NN that punishes the whole network for having information needed to classify the number of jets in the event is called gradient reversal [40]. This technique was pioneered in CMS by NN-based object taggers that punished their NN for being able to distinguish the difference between collision data and MC data while training.

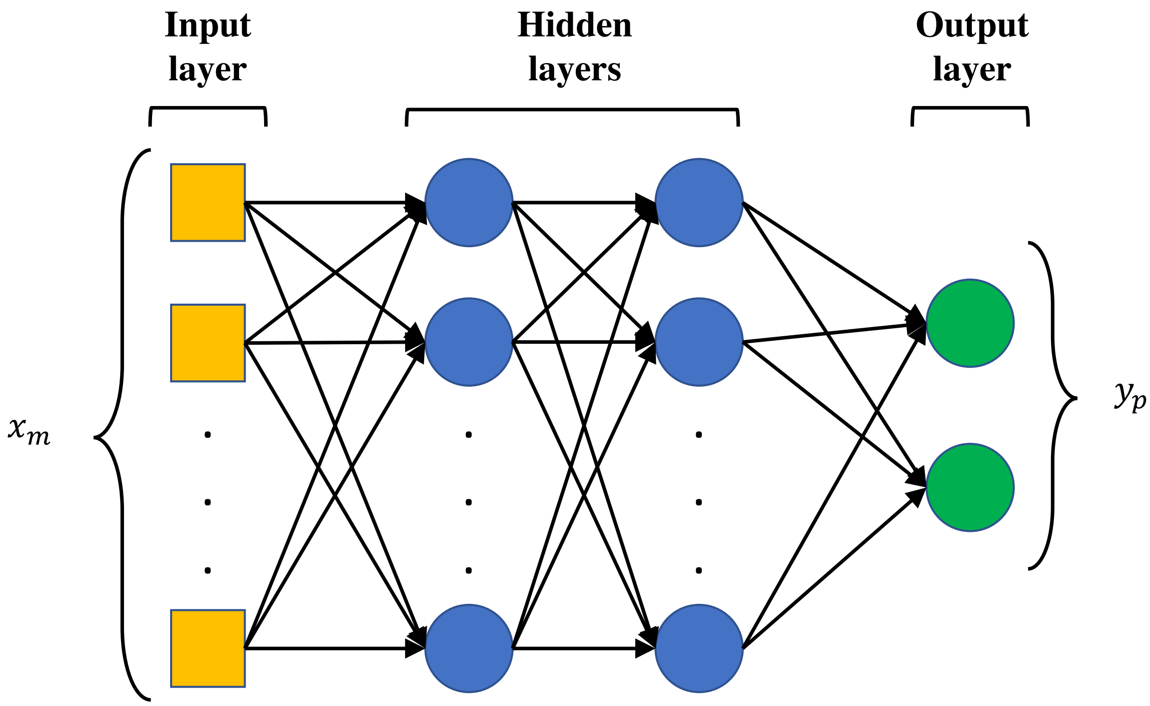


Figure 5.16: Example of a fully connected neural network with an arbitrary number of input variables, two hidden layers with an arbitrary number of nodes, and an output layer that has two nodes corresponding to the prediction of the classifiers [44].

5.4.4 Training Performance

The performance of the training is quantified by the shape of the receiver operating characteristic (ROC) curve and the value of the area under this curve, where a value of one would indicate perfect training. The ROC curve is a plot of the efficiency of tagging a signal event correctly as signal vs. the fake rate, in which a background event is incorrectly tagged as signal, as parameterized by the value of

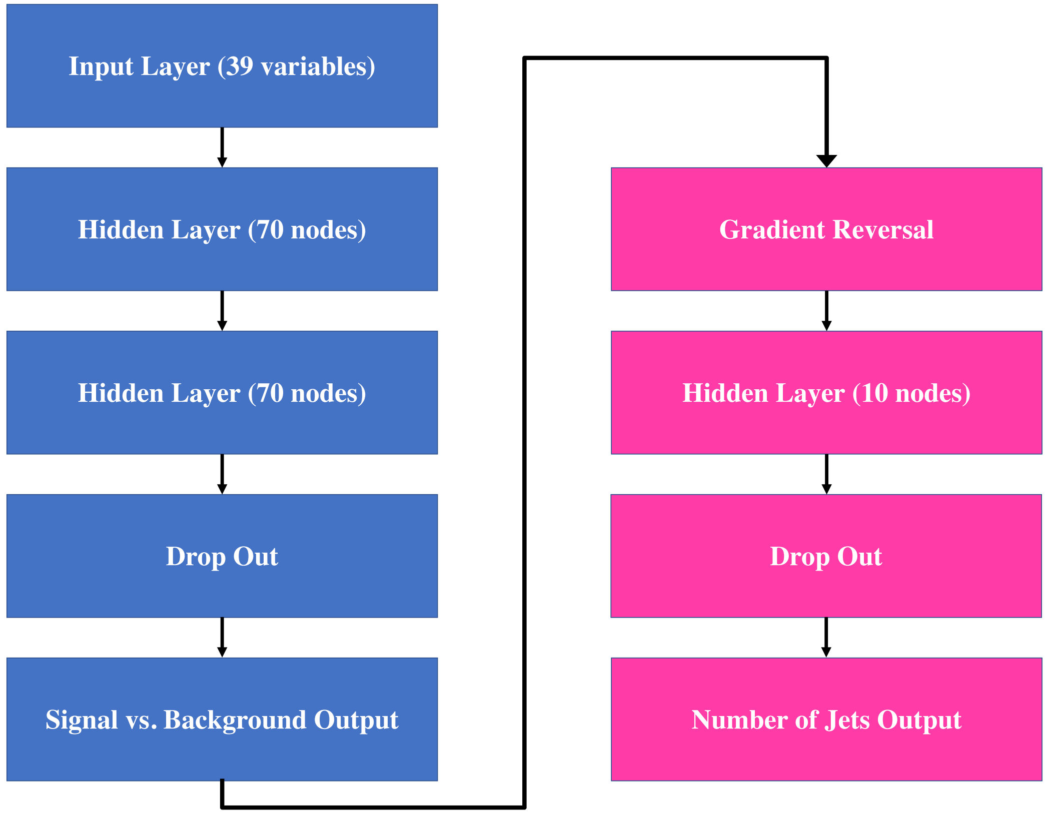


Figure 5.17: Architecture chosen for the NN. It includes an input layer to a hidden layer, to another hidden layer that then has dropout performed, to an output layer that is supervised for the signal vs. background classification, to a gradient reversal layer that is connected to a hidden layer, and finally to another output layer that is supervised for classifying N_{j}.

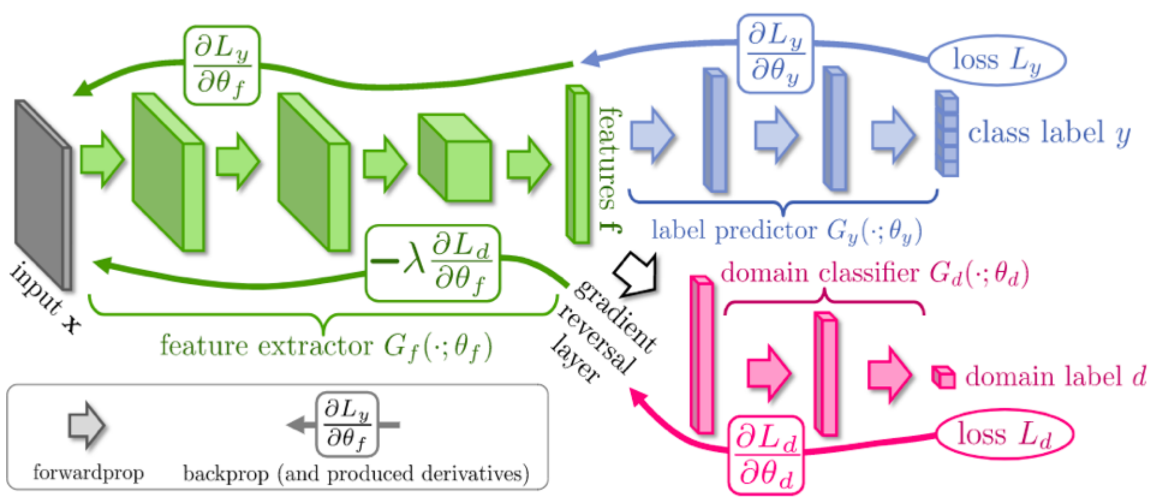


Figure 5.18: A diagram that shows a network with a gradient reversal layer [40].

the discriminant. Since different cut values are used to define the NN bins for each N_{j} bin, the ROC curve is evaluated for each N_{j} bin separately. Figure 5.19 shows the impact of gradient reversal, which forces the NN discriminator distribution shape for $\mathrm{t\bar{t}}$ to be almost the same for each N_{j} bin.

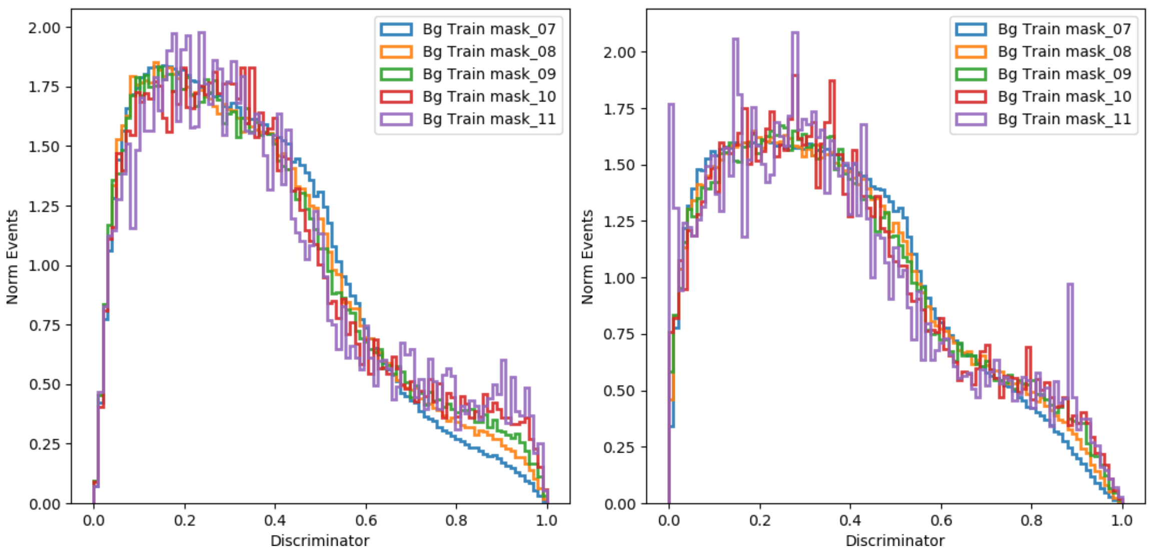


Figure 5.19: Discriminator distribution shape for each N_{j} that is classified in the $\mathrm{t\bar{t}}$ samples simulated for the 2016 (left) and 2017 (right) scenarios.

The NN output distributions for the $\mathrm{t\bar{t}}$ background and a few example signal points, as well as ROC curves, are shown in Fig. 5.20 for 2016 data and in Fig. 5.21 for 2017 data. These plots show that the performance of the training is different for each mass model, which is expected because the kinematics of the events change when the top squark mass is varied. For lower top squark masses, the kinematics of the event tend to look similar to a $\mathrm{t\bar{t}}$ decay (background).

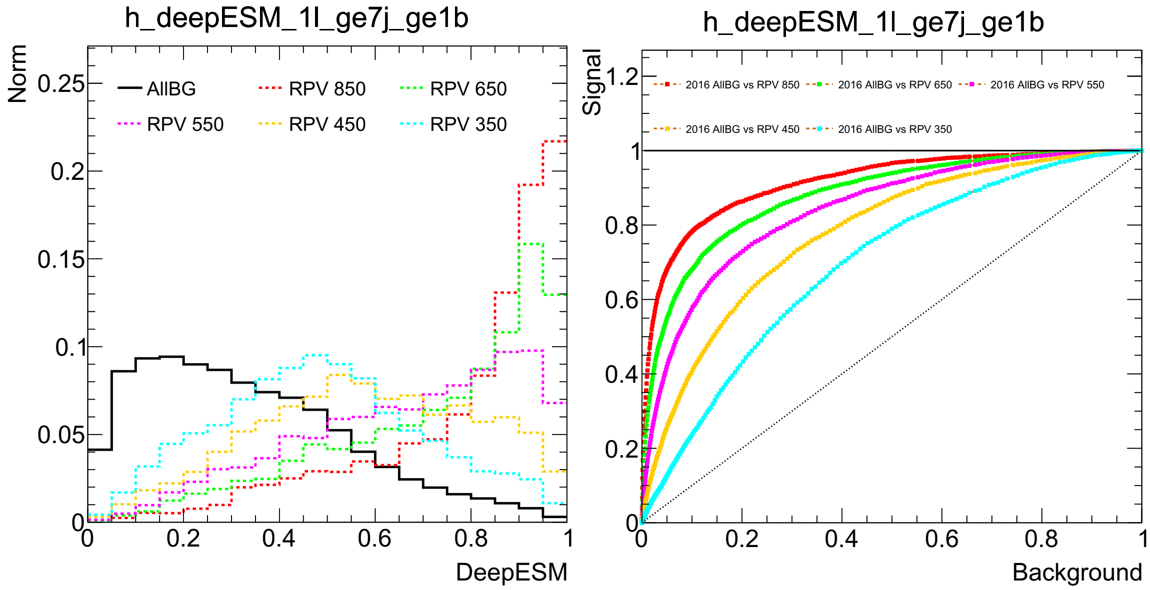


Figure 5.20: NN performance plots for the RPV model obtained from simulated events corresponding to the 2016 scenario. (Left) The normalized NN score (DeepESM) is shown for select signal models and all backgrounds summed together (AllBG). (Right) ROC curves for the corresponding signal models.

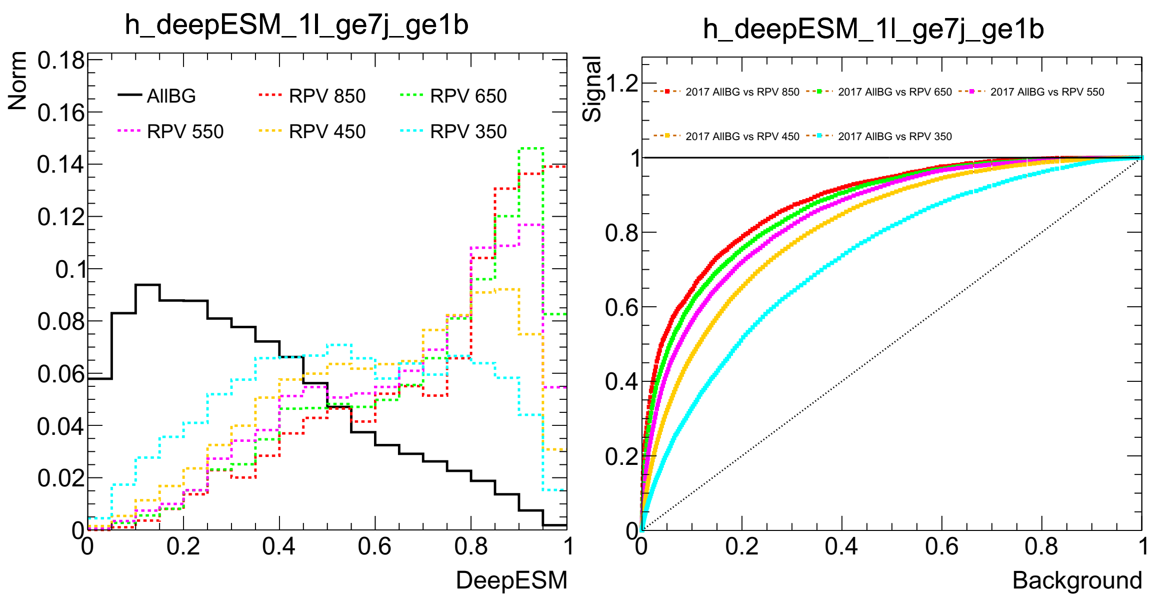


Figure 5.21: NN performance plots for the RPV model obtained from simulated events corresponding to the 2017 scenario. (Left) The normalized NN score (DeepESM) is shown for select signal models and all backgrounds summed together (AllBG). (Right) ROC curves for the corresponding signal models.

5.4.5 *Neural Network Bin Definition*

The background estimation method relies on the assumption that the N_{j} shape is the same between NN bins. After using gradient reversal in the training of the NN, the NN output is mostly independent of N_{j}. However, a small correlation remains. To ensure that the N_{j} shape remains the same in each NN bin, the NN bin edges defining NN bins 1, 2, 3, and 4 are derived separately for each N_{j} bin, such that each NN bin contains a given fraction of background events. This per-N_{j} binning is illustrated in Fig. 5.22. For example, NN bin 1 could have 55% of all $\mathrm{t\bar{t}}+\mathrm{jets}$ events, NN bin 2 could have 25%, NN bin 3 could have 15%, and NN bin 4 could have 5%. This means that while the NN bin edge values may differ from N_{j} bin to N_{j} bin (i.e. the value of the bin edge between NN bin 3 and NN bin 4 for $N_{\mathrm{j}} = 7$ may be 0.78 while the value of the bin edge between NN bin 3 and NN bin 4 for $N_{\mathrm{j}} = 8$ may be 0.8), the $\mathrm{t\bar{t}}$ fraction remains the same. This ensures that the N_{j} shape for $\mathrm{t\bar{t}}$ is the same for all four NN bins, a necessary requirement for the fit. A systematic uncertainty is derived to address the potential breakdown of this assumption for data, as described in later sections.

To decide on the exact value of the NN bin edges, a significance metric is maximized:

$$\frac{S}{\sigma_B} = \frac{N_{\mathrm{sig}}}{\sqrt{N_{\mathrm{t\bar{t}}} + N_{\mathrm{QCD}} + (0.20 \times N_{\mathrm{t\bar{t}}})^2 + (0.20 \times N_{\mathrm{QCD}})^2}} \tag{5.12}$$

where N_{sig} is the number of signal events for the RPV model with a top squark mass of 550 GeV, $N_{\mathrm{t\bar{t}}}$ is the number of $\mathrm{t\bar{t}}$ 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 $\mathrm{t\bar{t}}$ fraction) are adjusted. Note that the value of 0.20 is included in the equation to take a rough

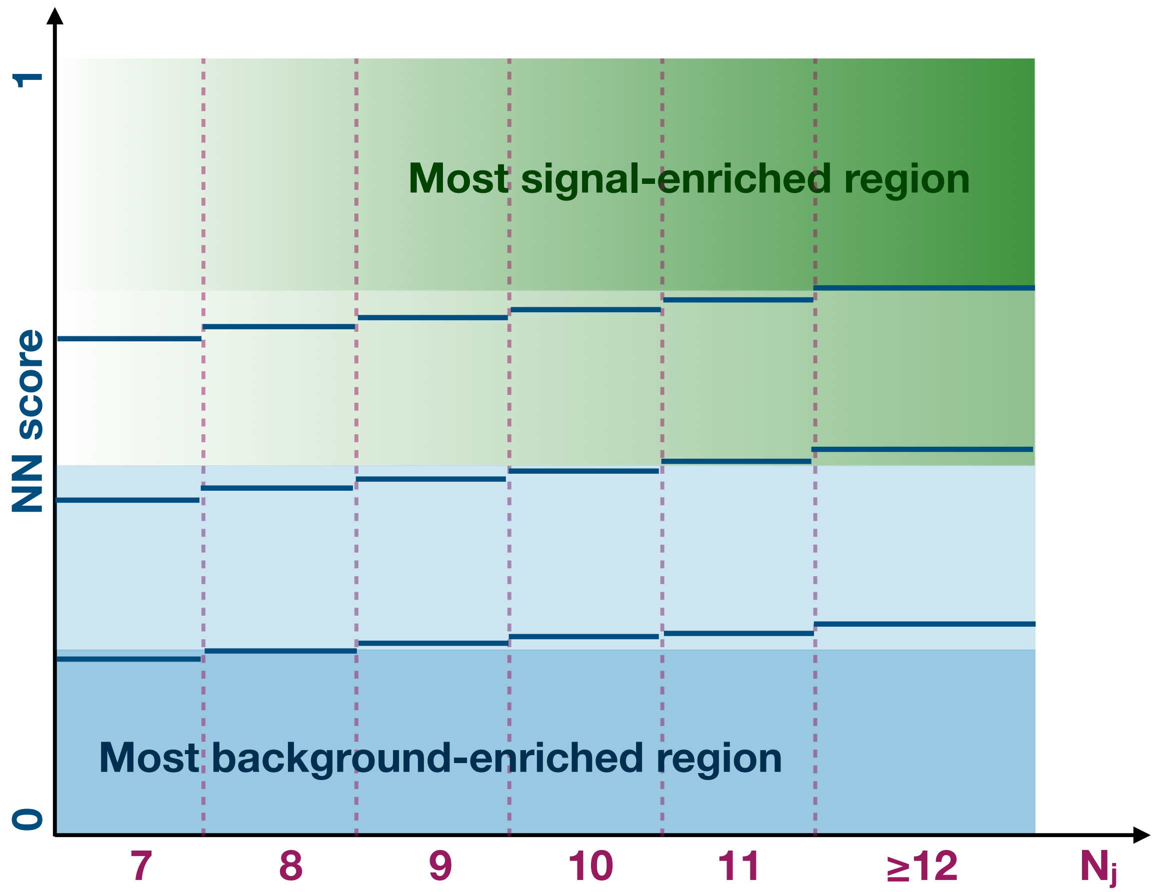


Figure 5.22: Signal region bin definition as a function of N_j and NN bins. The NN bins are defined as a function of N_j.

20% systematic uncertainty value of both $N_{\mathrm{t\bar{t}}}$ and N_{QCD}. The decision to optimize the analysis for the RPV model with top squark mass of 550 GeV was based on the expected sensitivity of the analysis. Models with top squark mass values less than 550 GeV have very large cross section values, which makes them easy to find if they are present in data. On the other hand, models with top squark mass values greater than 550 GeV have very small cross section values, which implies that they will be difficult to observe if present. However, the analysis remains sensitive to all signal models, because the signal N_j shapes do not change drastically as a function of top squark mass.

The significance value is calculated per N_j bin and then summed in quadrature across all N_j bins. The optimal set of background fractions is derived using 2016 MC data. The $\mathrm{t\bar{t}}$ + jets background fraction ends up being 47.8%, 38.9%, 6.5%, and 2.4% for NN bins 1, 2, 3, and 4, respectively. The same $\mathrm{t\bar{t}}$+jets MC data background fractions are applied to 2017, 2018pre, and 2018post, which yield unique NN bin edges for each era. The final sets of bin edges are shown in Table 5.4 for 2016, Table 5.5 for 2017, Table 5.6 for 2018pre, and Table 5.7 for 2018post data, respectively. As expected, the values are close to each other for all N_j, but staggered slightly as the NN distributions have a very minor N_j dependence, even after gradient reversal. The use of the N_j-dependent NN bin edge definition is shown in Fig. 5.23, where the left plot shows the normalized N_j shape for each NN bin if the N_j-dependent NN bin edges are not used and the right plot shows the normalized shapes if they are used.

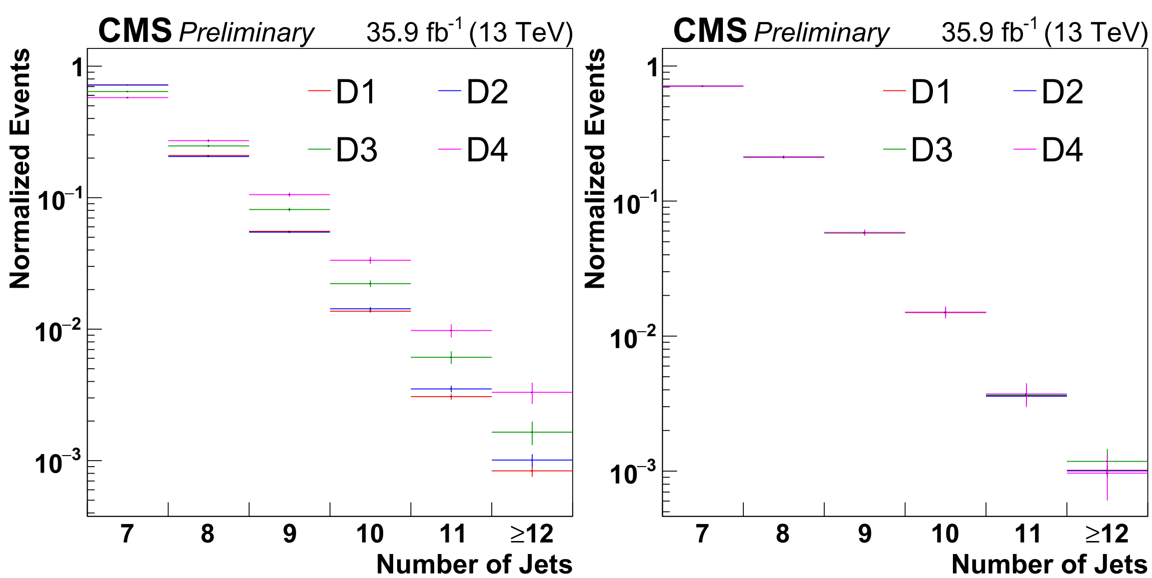


Figure 5.23: Applying the NN binning: using the same binning for all N_j bin (left), and using the per N_j defined binning (right).

Table 5.4: NN bin edges per N_j bin for 2016.

NN Bin Edge	$N_\mathrm{j}=7$	$N_\mathrm{j}=8$	$N_\mathrm{j}=9$	$N_\mathrm{j}=10$	$N_\mathrm{j}=11$	$N_\mathrm{j}\geq 12$
1 to 2	0.349	0.358	0.371	0.386	0.410	0.410
2 to 3	0.680	0.740	0.787	0.816	0.848	0.853
3 to 4	0.835	0.878	0.901	0.917	0.936	0.936

Table 5.5: NN bin edges per N_j bin for 2017.

NN Bin Edge	$N_\mathrm{j}=7$	$N_\mathrm{j}=8$	$N_\mathrm{j}=9$	$N_\mathrm{j}=10$	$N_\mathrm{j}=11$	$N_\mathrm{j}\geq 12$
1 to 2	0.346	0.363	0.376	0.384	0.396	0.396
2 to 3	0.714	0.752	0.780	0.794	0.801	0.801
3 to 4	0.833	0.864	0.885	0.895	0.901	0.901

Table 5.6: NN bin edges per N_j bin for 2018 pre HEM issue.

NN Bin Edge	$N_\mathrm{j}=7$	$N_\mathrm{j}=8$	$N_\mathrm{j}=9$	$N_\mathrm{j}=10$	$N_\mathrm{j}=11$	$N_\mathrm{j}\geq 12$
1 to 2	0.345	0.363	0.374	0.386	0.386	0.386
2 to 3	0.713	0.752	0.779	0.805	0.805	0.805
3 to 4	0.832	0.863	0.883	0.900	0.911	0.911

Table 5.7: NN bin edges per N_j bin for 2018post.

NN Bin Edge	$N_\mathrm{j}=7$	$N_\mathrm{j}=8$	$N_\mathrm{j}=9$	$N_\mathrm{j}=10$	$N_\mathrm{j}=11$	$N_\mathrm{j}\geq 12$
1 to 2	0.364	0.386	0.405	0.411	0.411	0.411
2 to 3	0.729	0.771	0.798	0.826	0.826	0.826
3 to 4	0.842	0.874	0.895	0.914	0.914	0.914

With these NN bin edges defined, the numbers of signal events for different RPV mass models in each of the NN bins are shown in the left plot of Fig. 5.24 for 2016. This plot indicates that the number of signal events does not increase with NN bin. The right plot of Fig. 5.24 shows the value of the sum in quadrature per N_{j} of the significance metric in Eq. (5.12) for each NN bin. It is clear from this plot that the higher NN bins are more sensitive to the signal models than the lower NN bins, as expected. Finally, the resulting N_{j} shapes per NN bin in the $\mathrm{t\bar{t}}+\mathrm{jets}$ MC data is shown in Fig. 5.25. These N_{j} distributions give confidence that if the $\mathrm{t\bar{t}}+\mathrm{jets}$ events in collision data and the $\mathrm{t\bar{t}}+\mathrm{jets}$ events in MC data have the same N_{j}-NN dependence, the four NN bins will have the same $\mathrm{t\bar{t}}+\mathrm{jets}$ N_{j} shape.

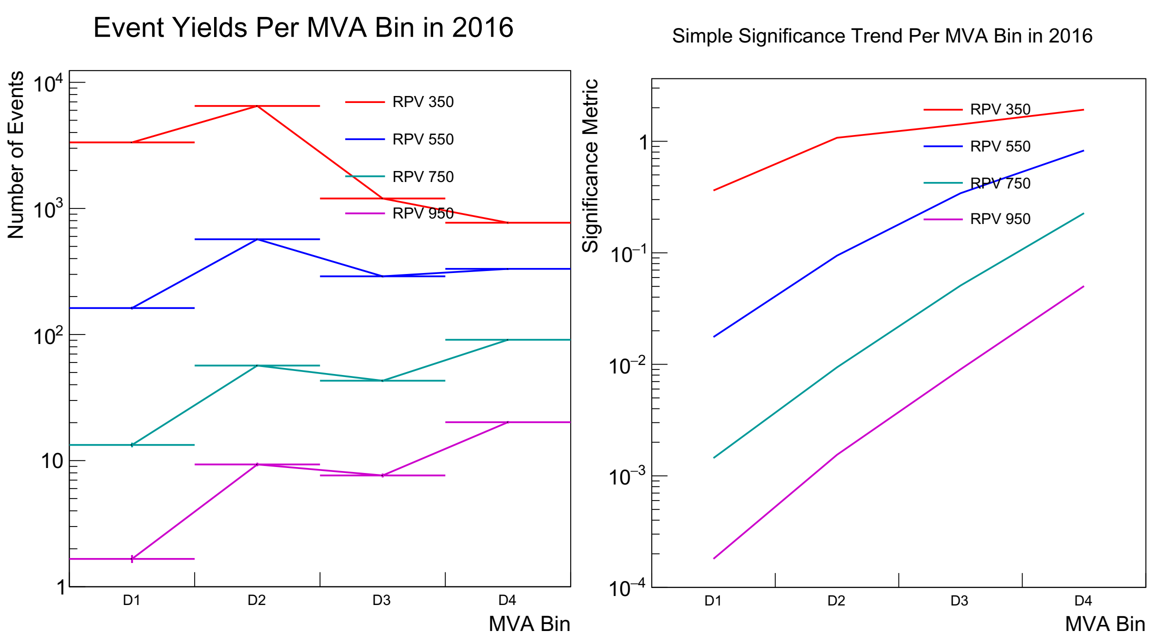


Figure 5.24: (Left) Number of signal events for different RPV signal models in each NN bin. (Right) The sum in quadrature over each N_{j} bin of the significance metric as defined in Eq. (5.12) per NN bin.

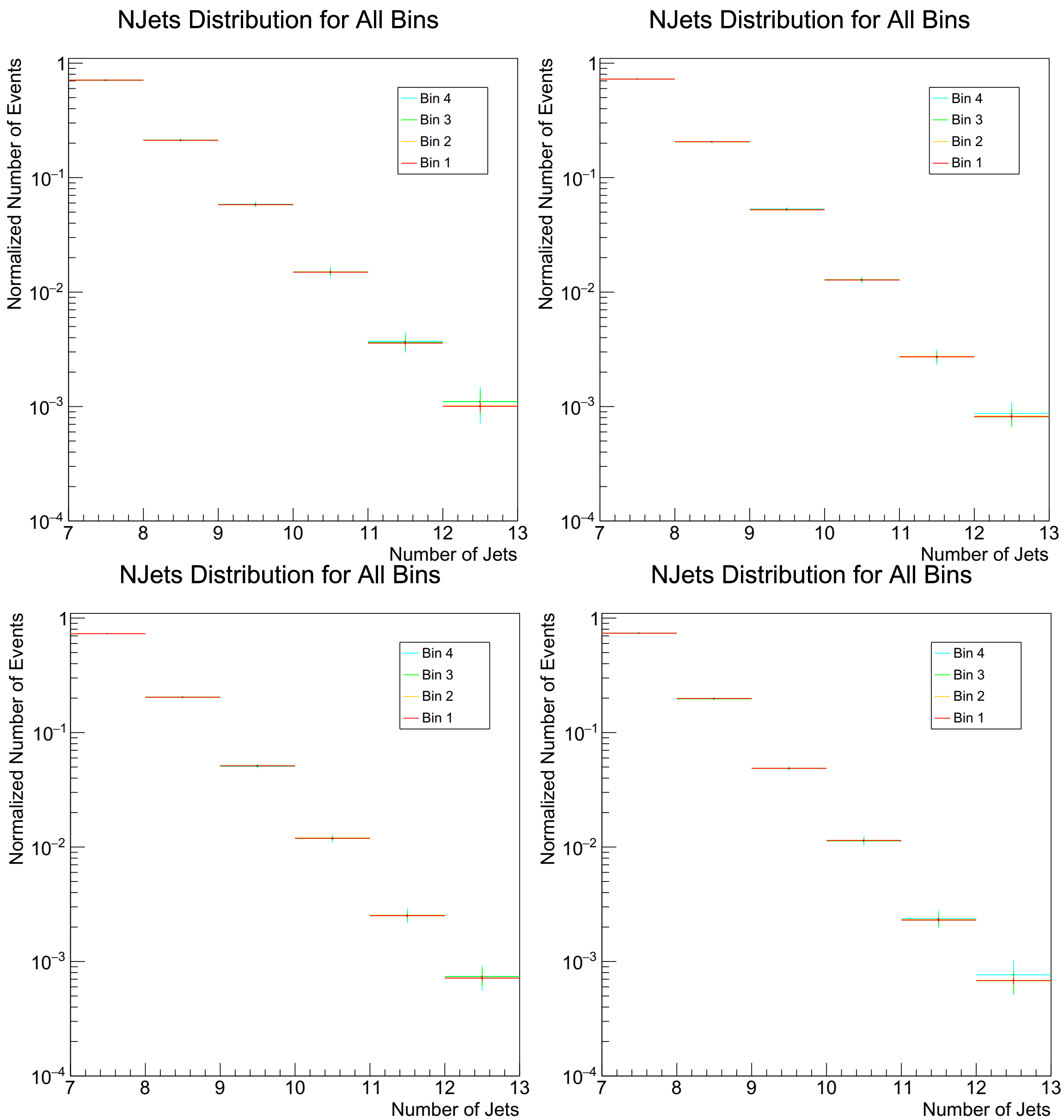


Figure 5.25: N_j distribution for 2016 (top left), 2017 (top right), 2018pre (bottom left), and 2018post (bottom right) for each of the NN bins for the MC data $\mathrm{t\bar{t}}$ background using the per N_j bin edges.

5.4.6 $t\bar{t}$ *Fit Function*

The N_j distribution is hard to model correctly at very high jet multiplicities ($N_j \geq 10$), so we rely on a fit of the shape of the N_j distribution in collision data for the $t\bar{t}$ background estimation. From theoretical considerations [45], the ratio of the N_{j+1} distribution to the N_j distribution can be described by two components: (1) a constant at high N_j, the so-called staircase behavior, and (2) a falling function at lower N_j, often called the Poisson regime. We rely on an assumption that, beyond a certain number of jets, the probability of producing an additional jet is independent of the number of jets. For perfect "staircase" scaling the ratio $R(j)$ is given by

$$R(j) = \frac{N(j+1)}{N(j)} = c_0 \text{ (constant)} \tag{5.13}$$

where $N(j)$ gives the number of events with j number of jets in the event. When adding lower jet multiplicities, an additional "Poisson" behavior is incorporated:

$$R(j) = c_0 + \frac{c_1}{j + c_2} \tag{5.14}$$

where c_0, c_1, and c_2 are free parameters that need to be profiled in the fit to data. Figure 5.26 shows the $R(j)$ ratio for data collected at an electron-positron collider. However, we found the parameters in Eq. (5.14) to be highly correlated. Having fit parameters that are highly correlated can confuse the fit optimization algorithm and result in unfavorable results. Therefore, we formulated an alternate functional form, $R'(j)$, which retained the desired property of "staircase" scaling at high jet multiplicity. In this case the "Poisson" behavior is modeled by a falling exponential

$$R'(j) = p_1 + (p_0 - p_1)e^{p_2(j-7)} \tag{5.15}$$

where p_0 is the value of the first point, p_1 is the asymptotic value (staircase ratio), and p_2 is the exponential parameter. The parameterization in Eq. (5.15) still suffers

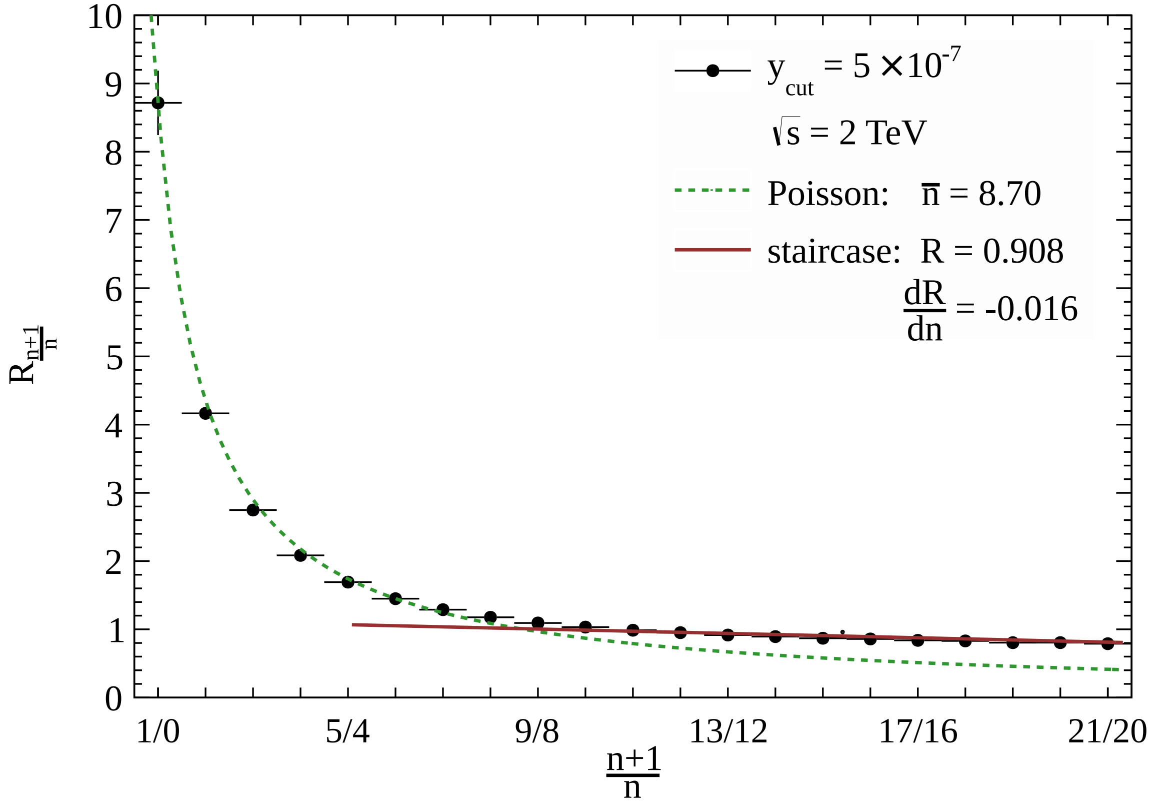


Figure 5.26: Jet ratios $R(n)$ in $e^+e^- \to$ jets production at 2 TeV center-of-mass energy [45].

from large correlations between p_1 and p_2, so the function was recast in terms of the values of the function evaluated at two points resulting in a new parameterization of a mathematically equivalent function

$$f(x) = a_2 + \left[\frac{(a_1 - a_2)^{x-m}}{(a_0 - a_2)^{x-n}}\right]^{\frac{1}{n-m}} \tag{5.16}$$

where $a_0 = f(m)$ is the function at $x = m$, $a_1 = f(n)$ is the function at $x = n$, and a_2 is the asymptotic value. Setting $m = 0$ corresponds to $f(0) = N(8)/N(7)$, and $n = 2$ corresponds to $f(2) = N(10)/N(9)$. The choice of $m = 0$ allows the fit to take advantage of the large number of events in a background-dominated region for constraining the shape. The choice of $n = 2$ gives the fit a greater lever arm in fitting the shape, and is still in the region of a large number of $t\bar{t}$ events and relatively low signal events. With $m = 0$ and $n = 2$, Eq. (5.16) simplifies to

$$f(x) = a_2 + \left[\frac{(a_1 - a_2)^x}{(a_0 - a_2)^{x-2}}\right]^{\frac{1}{2}} \tag{5.17}$$

where $a_0 = N(8)/N(7)$, $a_1 = N(10)/N(9)$, and a_2 is the asymptotic value as $x \to \infty$. The correlation behavior is as desired, with no large correlations among the three parameters. To further help with the behavior of the fit, we redefine a_2 to be

$$a_2 = \begin{cases} a_1 - 1/d, & d \geq 1 \\ \\ a_1 - (2-d), & d < 1. \end{cases} \tag{5.18}$$

Thus, the final three parameters used in each fit are a_0, a_1, and d.

As a final step, we transform from using the ratio of neighboring N_j bins, $f(x)$, to instead expressing a function, $F(x)$, in terms of N_j itself, by effectively multiplying out the recursively defined prediction of N_j. This removes correlations for the bins of the observables, and will allow for the proper handling of the Poisson uncertainties. A recursive expression for N_j can now be defined as

$$\begin{aligned} F(0) &= \text{Number of entries in the 7-jets bin} \\ F(x) &= F(x-1)f(x-1) \;\text{ for } x > 0 \end{aligned} \tag{5.19}$$

where $f(x)$ is the function defined in Eq. (5.17). Note that $x = 0$ corresponds to the 7-jets bin and $x = 7$ corresponds to the inclusive ≥12-jets bin. Expanding the expression for each N_j bin gives the following:

$$\begin{aligned} N_\mathrm{j} = 7: \qquad & F(0) = N_7 \\ N_\mathrm{j} = 8: \qquad & F(1) = F(0)f(0) = N_7 f(0) \\ N_\mathrm{j} = 9: \qquad & F(2) = F(1)f(1) = N_7 f(0)f(1) \\ N_\mathrm{j} = 10: \qquad & F(3) = F(2)f(2) = N_7 f(0)f(1)f(2) \end{aligned}$$

$$N_{\mathrm{j}} = 11: \qquad F(4) = F(3)f(3) = N_7 f(0)f(1)f(2)f(3)$$

$$N_{\mathrm{j}} \geq 12: \qquad F(5) = F(4)f(4) + F(5)f(5) + \cdots + F(10)f(10).$$

For the last term, it is sufficient to include N_{j} bins from 12 through 18. The analysis uses the assumption that the $\mathrm{t\bar{t}}$ background shape is the same for all four NN bins, since the NN training was designed to remove shape differences across NN bins, and the binning selection was designed to remove any residual differences. Therefore, the floating parameters a_0, a_1, and d are shared among the $\mathrm{t\bar{t}}$ background shape for the four NN bins. There is a separate floating parameter for the number of entries in the $N_{\mathrm{j}} = 7$ bin for each of the four NN bins, resulting in a total of seven free parameters for the $\mathrm{t\bar{t}}+\mathrm{jets}$ background fit: four normalization parameters and three shape-based parameters.

5.5 QCD Multijet Background Estimation

The second largest background arises from QCD multijet events, which amounts to around 5% of the total background (depending on the N_{j} bin). The number of MC data for this subdominant background is low and some of these simulated events have a relatively large weight, which causes the statistical uncertainties associated with these events to be large. These statistical uncertainties would be used in the statistical interpretation of the signals. In fact, these large statistical uncertainties can hide a signal, because they give the fit too much freedom, which decreases the overall sensitivity of the analysis.

To avoid factoring in these large statistical fluctuations, the QCD multijet background is estimated using collision data in a QCD-dominated control region. To ensure that this control region is orthogonal to the search region, a requirement of exactly one non-isolated muon and zero well-defined isolated leptons is used. A p_{T} requirement of 55 GeV is used for this non-isolated muon due to the online selection required by the non-isolated muon trigger.

The baseline selection for this control region is:

- A non-isolated muon with the following properties:
  - $|\eta| < 2.4$
  - $p_{\mathrm{T}} > 55$ GeV
  - Not isolated (using a 0.2 miniIso value as the requirement)
  - Medium ID
- $H_{\mathrm{T}} > 300$ GeV
- All jets pass the loose jet ID for 2016 and the tight jet ID for 2017 and 2018
- $N_{\mathrm{j}} \geq 7$

The N_{j} distribution obtained using this selection of the QCD-enriched control region is shown in Fig. 5.27. The NN score is calculated for the CR by using the non-isolated muon in place of the isolated lepton used for the SR. It is important to stress that the NN used in this control region (CR) is the exact same NN used to evaluate the signal region. Distributions of the NN discriminant are shown in Fig. 5.27.

To estimate the QCD multijet background, all the events that pass the above CR baseline selection are divided into the four NN bins utilizing the bin edges derived in the signal region. Next, the non-QCD multijet events, taken from simulation, are subtracted from the data in the control region. The QCD MC data is then used to define a transfer factor, which is the ratio of the QCD yield in the signal and control

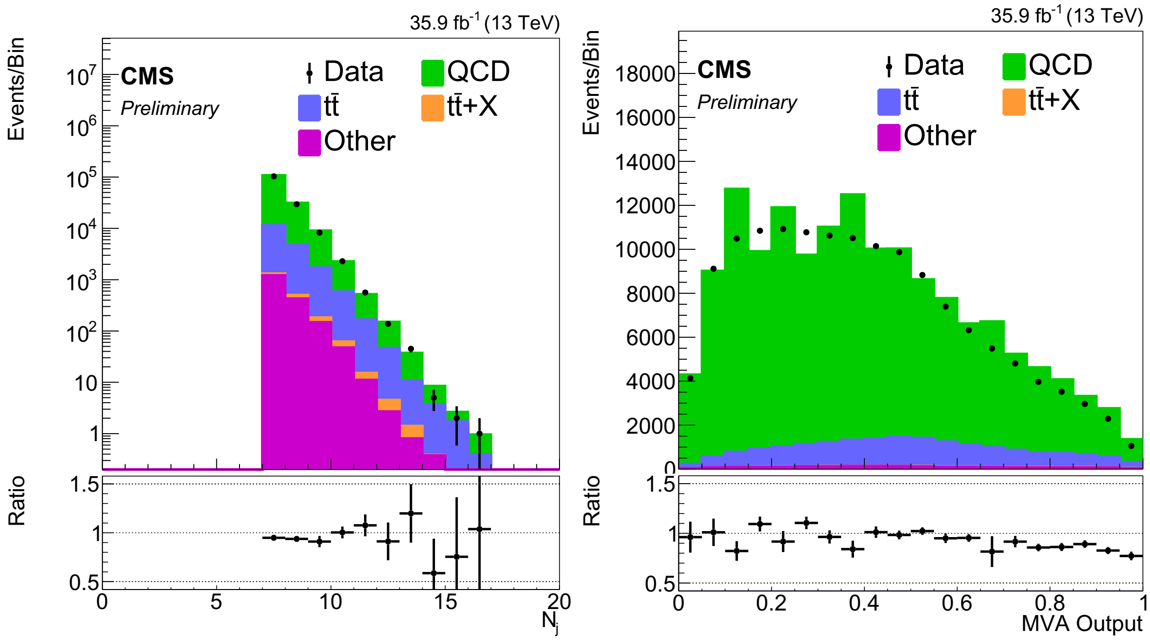


Figure 5.27: Distributions for the QCD-enriched control region for 2016: N_j in logarithmic scale (left) and NN score (right).

regions in simulation

$$TF = \frac{N^{SR}}{N^{\mathrm{CR}}}, \tag{5.20}$$

where N^{SR} is the number of total weighted QCD multijet events in the signal region simulation and N^{CR} is the number of total weighted QCD multijet events in the control region simulation. This transfer factor is used as an overall normalization for the data in the control region to match the total expected QCD yield in the signal region, shown for each era in Table 5.8. By having the transfer factor be an overall normalization, the N_j shapes are taken from the CR directly for the QCD multijet prediction. The prediction for a particular bin in the signal region is calculated from

$$N_{ij}^{\mathrm{SR}} = TF\ N_{ij}^{\mathrm{CR}} \tag{5.21}$$

where i, j runs over all jets and all NN bins.

The N_{j} distributions in each NN bin for the QCD multijet estimate in the four time periods (2016, 2017, 2018pre, and 2018post) are shown in Figs. 5.28, 5.29, 5.30, and 5.31, respectively.

Table 5.8: Transfer factor values used for each era for the QCD multijet background estimate as well as the event yields of QCD multijet simulated events in the search region and QCD control region.

Year	QCD SR	QCD CR	TF (%)
2016	3917.2	133694.9	2.9
2017	3033.9	115726.9	2.6
2018pre	1423.8	57642.8	2.5
2018post	1586.1	72428.2	2.2

5.6 Systematic Uncertainties

After outlining the estimation method for each SM background, the last piece needed before performing fits to the observed data is to determine all systematic uncertainties. A systematic uncertainty is effectively a source of potential error that is not purely statistical and can cause the overestimation or underestimation of a background. For example, although the p_{T} of the jets is carefully calibrated, the simulation could still overestimate the p_{T} of the jets in a given event. Since the estimation of all small background processes is taken directly from simulation, this overestimation of the jet p_{T} could lead to more simulated events passing the signal region selection than the observed data. This potential systematic shift in the jet p_{T} needs to be taken into account by a dedicated systematic uncertainty. In order to account for the systematic shift, a multiplicative R value is derived based on the uncertainty of the jet p_{T} correction, such that when the expected number of background events produced is multiplied by the R value, it is then corrected to the

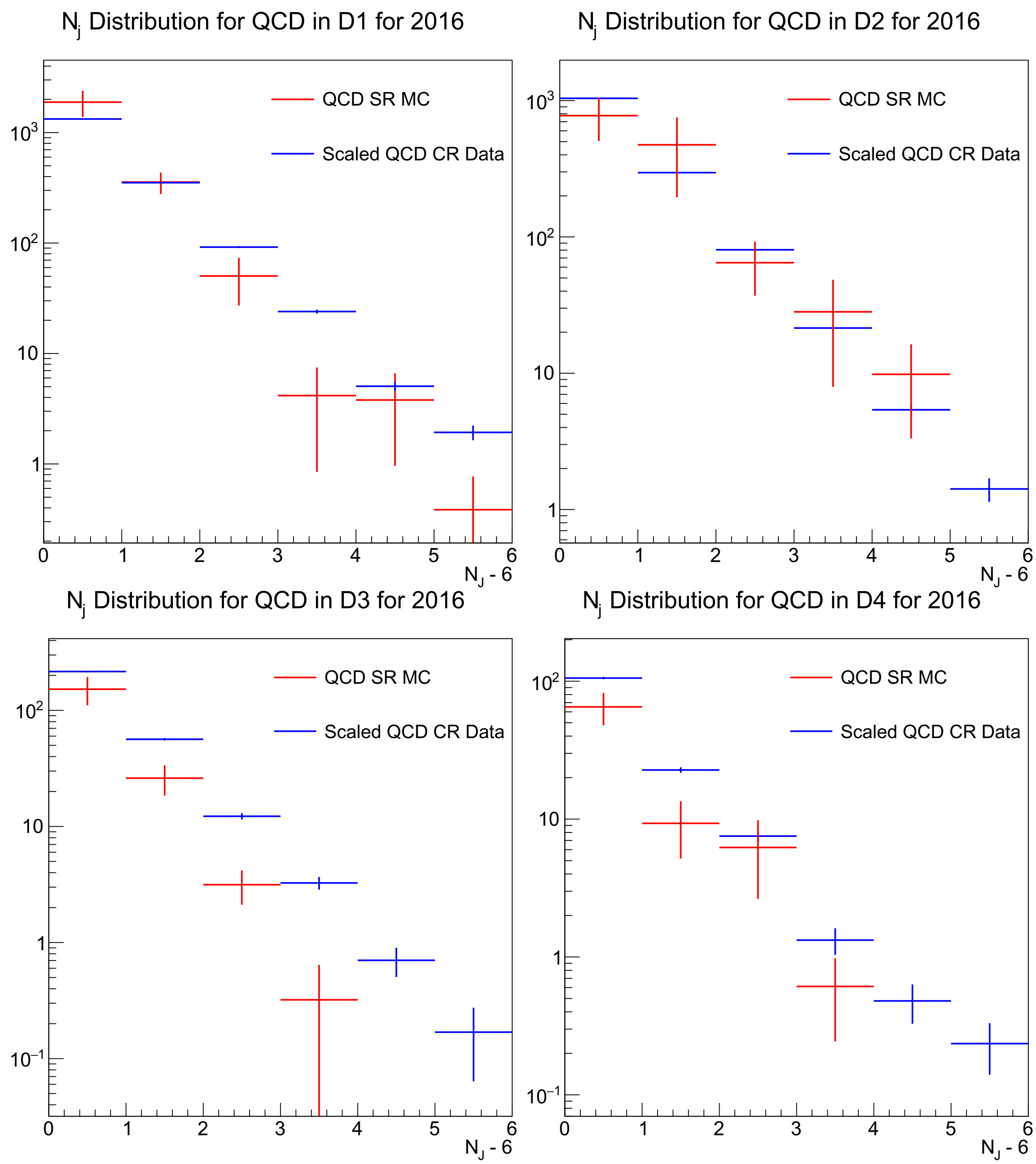


Figure 5.28: The N_j shape distribution comparison between the control region data-driven estimate (blue) and the estimate taken directly from the QCD simulation (red) in 2016 for four NN bins.

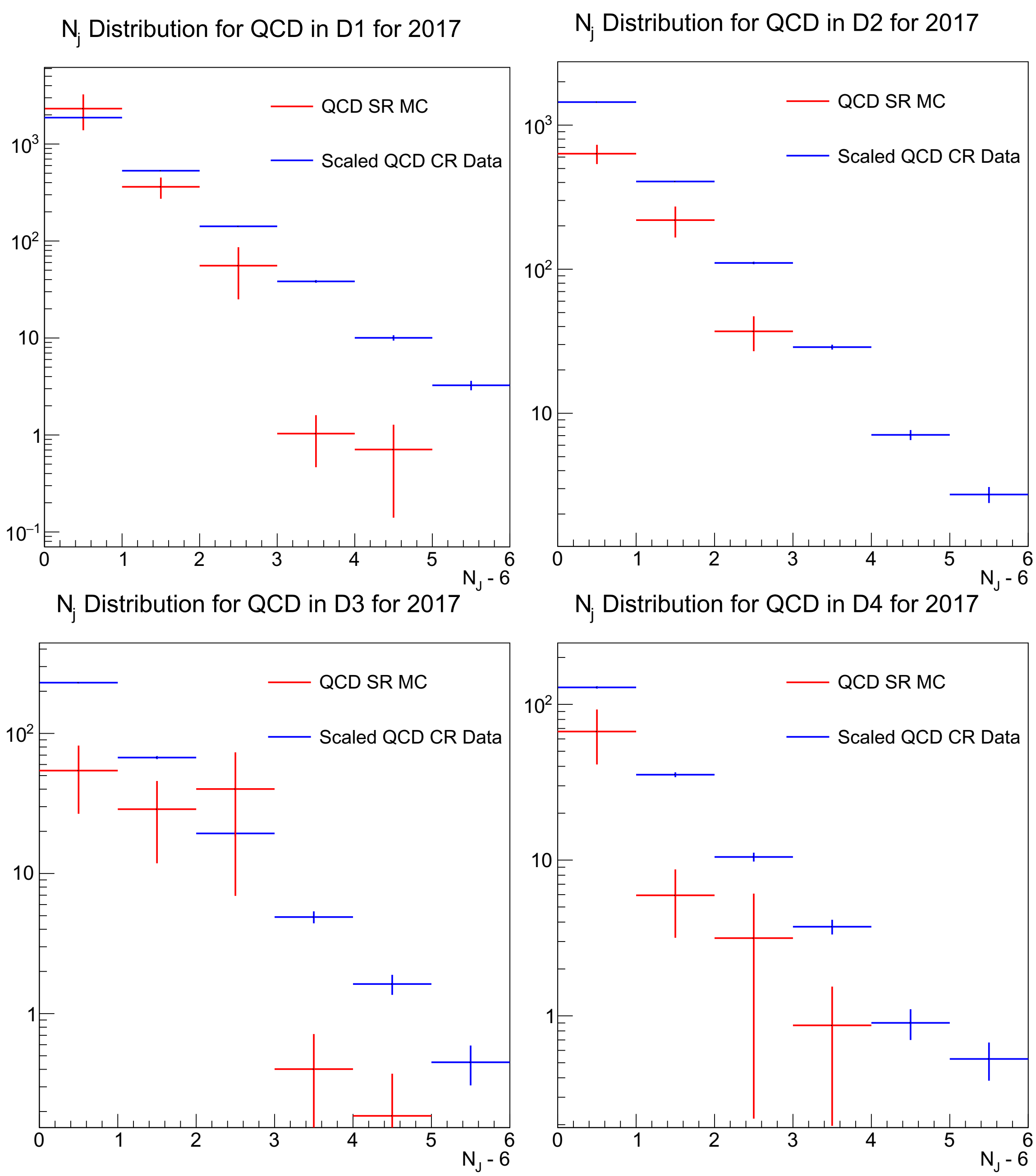


Figure 5.29: The N_{j} shape distribution comparison between the control region data-driven estimate (blue) and the estimate taken directly from the QCD simulation (red) in 2017 for four NN bins.

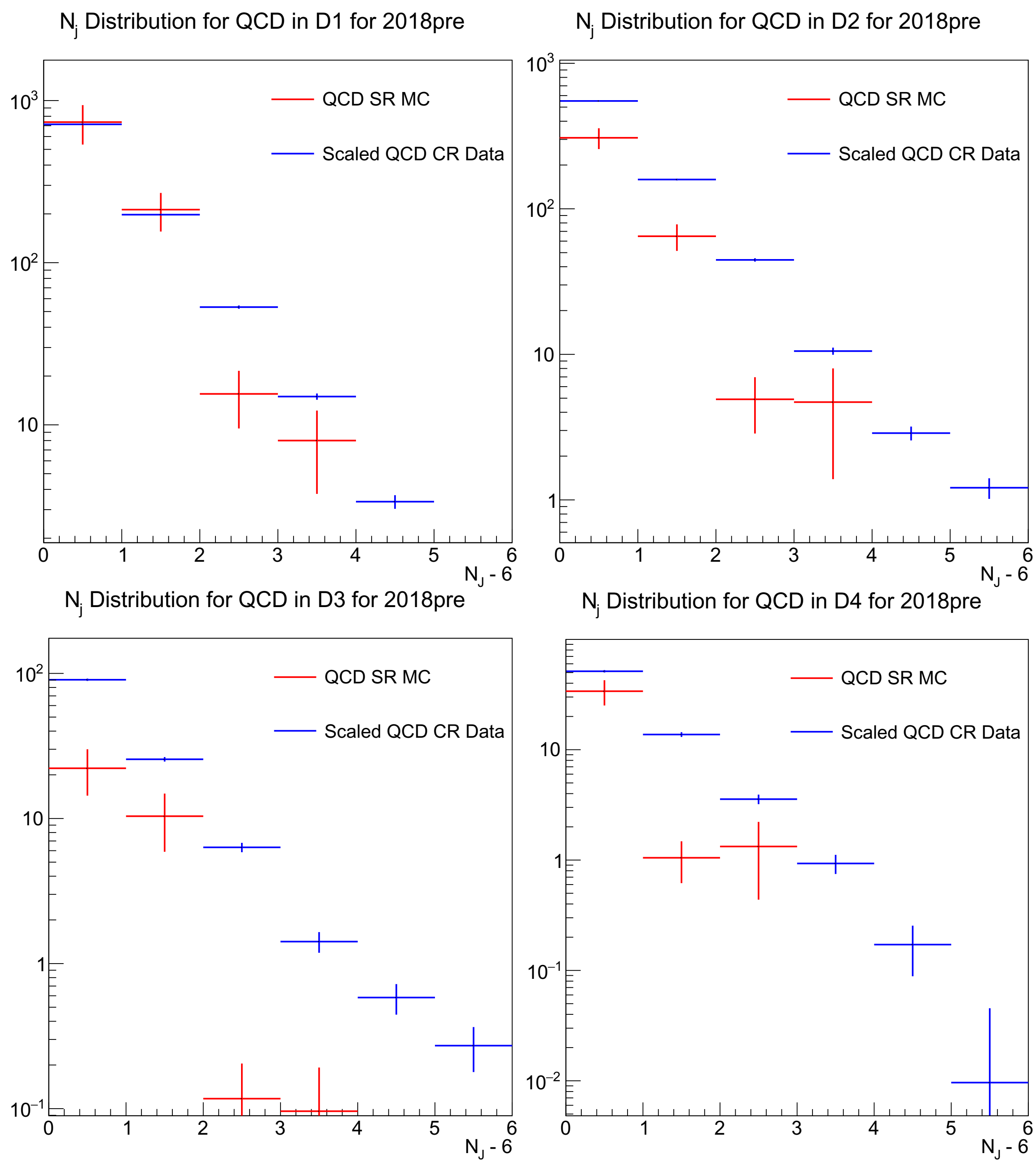


Figure 5.30: The N_{j} shape distribution comparison between the control region data-driven estimate (blue) and the estimate taken directly from the QCD simulation (red) in 2018pre for four NN bins.

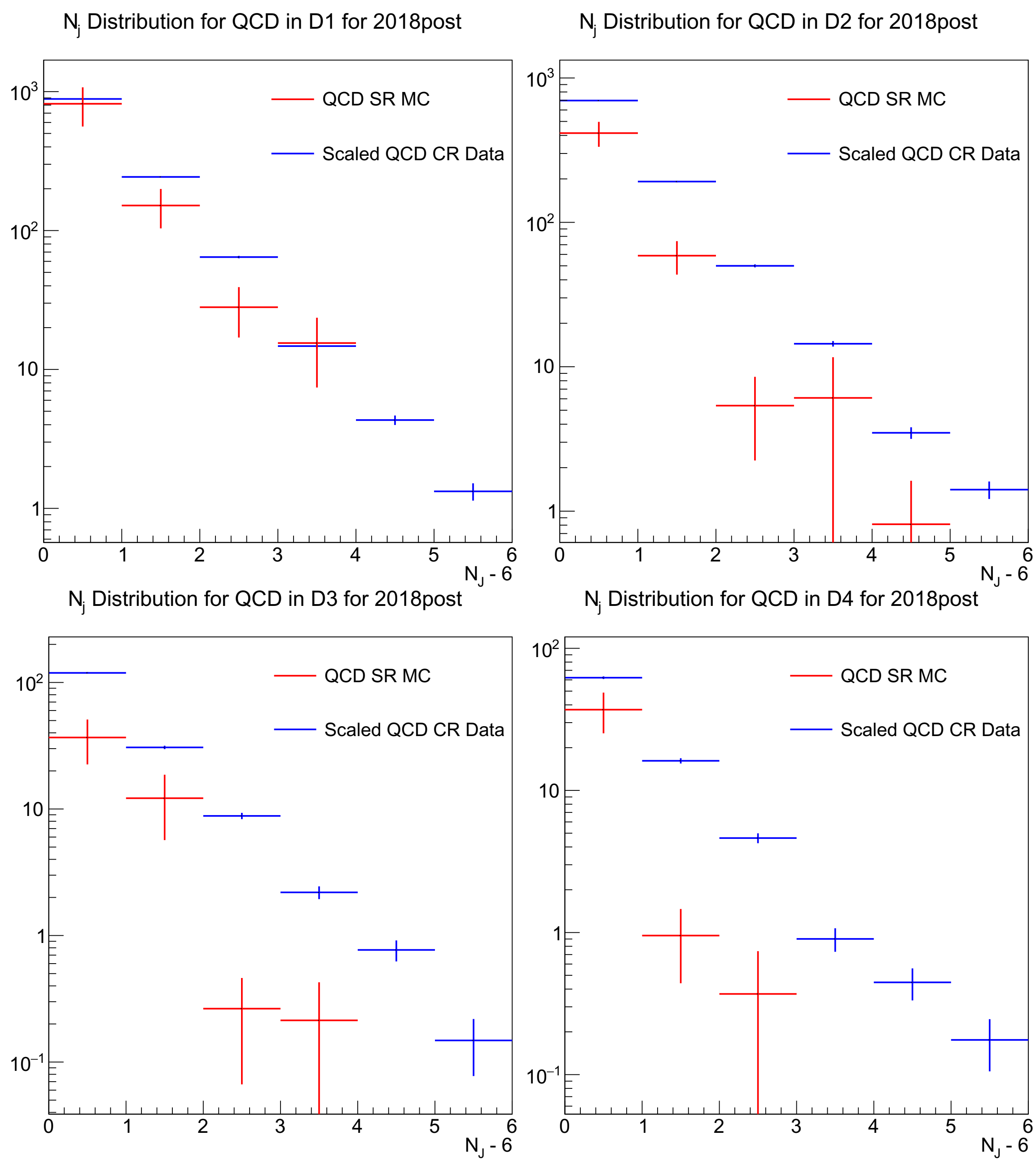


Figure 5.31: The N_{j} shape distribution comparison between the control region data-driven estimate (blue) and the estimate taken directly from the QCD simulation (red) in 2018post for four NN bins.

appropriate value based on the observed data. Continuing with our example, if the number of observed events is 10% lower than the expected prediction, given a 20% uncertainty on the expected number of events corresponding to the jet p_{T} uncertainty, then the R value is equal to 0.8 and will have an observed nuisance parameter θ value equal to 0.472. This systematic uncertainty is then included in a fit to the observed data with a correction to the background estimation:

$$
\begin{aligned}
N_f &= R^{\theta} N_i \\
&= 0.8^{0.472} N_i = 0.9 N_i
\end{aligned}
\tag{5.22}
$$

where N_f is the final background prediction, R is the systematic uncertainty correction value, θ is a nuisance parameter used to adjust the size of R, and N_i is the initial background prediction. The fit to data is performed by constructing a likelihood function necessary for a statistical evaluation of the observed data. Details on the statistical evaluation are given in Section 5.7. All systematic uncertainties applied in the analysis are summarized in Table 5.9. Representative values for the range of uncertainties across different NN and N_{j} bins are shown.

5.6.1 Signal and Non-$\mathrm{t\bar{t}}$ *Backgrounds*

The signal and non-$\mathrm{t\bar{t}}+\mathrm{jets}$ backgrounds are estimated directly using simulation, with the exception of the QCD multijet background estimation, which is derived in the control region using the procedure discussed in Section 5.5.

The remaining non-$\mathrm{t\bar{t}}+\mathrm{jets}$, non-QCD multijet background events are divided into two categories depending on whether the hard process produces more or fewer jets than $\mathrm{t\bar{t}}$. The first category is denoted as TTX, which includes mostly events with $\mathrm{t\bar{t}}+\mathrm{V}$ where V is a vector boson or the Higgs boson. The second category is denoted

Table 5.9: Summary of systematic uncertainties for the $t\bar{t}$ and non-$t\bar{t}$ backgrounds, as well as for the RPV signal model with top squark mass of 550 GeV.

Source	ttbar				non-ttbar				signal			
	2016	2017	2018pre	2018post	2016	2017	2018pre	2018post	2016	2017	2018pre	2018post
Luminosity	-	-	-	-	2.5	2.3	2.5	2.5	2.5	2.3	2.5	2.5
Jet energy scale	0–3	0–18	0–8	0–14	0–85	0–67	1–37	0–86	0–17	0–21	0–25	0–30
Jet energy resolution	0–6	0–10	0–3	0–8	0–92	0–54	0–53	1–46	0–9	0–8	0–7	0–13
btagging	0–2	0–2	0–1	0–1	0–3	0–11	0–3	0–3	0–2	0–1	0–1	0–1
Lepton id/iso/trigger	0–1	0–1	0–1	0–1	3–4	3–4	3–4	3–4	3–4	3–4	3–4	3–4
H_{T} (fit extrapolation)	0–2	0–1	0–1	0–3	2–7	0–2	0–3	2–10	-	-	-	-
H_{T} (high H_{T} tail corr.)	0–8	0–2	0–6	0–5	-	-	-	-	-	-	-	-
H_{T} (N_{j} dependence)	0–7	0–15	0–26	0–7	-	-	-	-	-	-	-	-
PDF	0–1	0–1	0–1	0–1	0–5	0–1	0–1	0–7	0–6	0–5	0–3	0–6
Prefiring	-	0–1	-	-	-	0–1	-	-	-	0–1	-	-
PU reweighting	0–5	0–6	0–4	0–7	0–6	0–24	0–27	0–20	0–3	0–4	0–2	0–2
Renorm./fact. scale	0–4	0–4	0–2	0–4	0–9	0–18	0–10	0–13	0–3	0–3	0–3	0–3
ISR	0–14	0–7	0–7	0–3	-	-	-	-	-	-	-	-
FSR	0–24	0–17	0–19	0–26	-	-	-	-	-	-	-	-
Nominal shape difference	0–12	0–16	0–27	0–26	-	-	-	-	-	-	-	-
CR shape difference	0–17	0–35	0–36	0–36	-	-	-	-	-	-	-	-
Jet mass-p_{T}	0–14	0–10	0–4	0–8	-	-	-	-	-	-	-	-
No H_{T} reweight	0–10	0–7	0–16	0–9	-	-	-	-	-	-	-	-
ME-PS matching	0–26	0–19	0–31	0–81	-	-	-	-	-	-	-	-
Color reconnection	0–14	0–39	0–44	0–23	-	-	-	-	-	-	-	-
Underlying event	0–37	0–56	0–40	0–36	-	-	-	-	-	-	-	-

as Other, and includes all other backgrounds, such as W + jets, Drell-Yan + jets, diboson, triboson, and single top. The latter two components are then put into the fit as separate components, each with its own normalization nuisance parameter, since the N_j shape distribution may be different. A range of systematic uncertainties is assigned to cover for potential mismodeling of these non-$\mathrm{t\bar{t}}$ + jets, non-QCD multijet background events, and are included as nuisance parameters in the fit.

- Luminosity uncertainty: 2.5% for 2016, 2.3% for 2017, and 2.5% for 2018
- Jet energy scale and resolution uncertainties as recommended by the JetMET POG
- Uncertainties in the b-tagging efficiency data vs. MC data scale factor
- Uncertainties in the lepton identification and isolation data vs. MC data scale factor
- Uncertainties in the trigger efficiency data vs. MC data scale factor
- Uncertainties in the H_T correction data vs. MC data scale factor (not used for signal)
- Uncertainties in the pileup reweighting
- Uncertainties in the renormalization and factorization scales
- Uncertainties in the parton distribution functions (PDF)
- Uncertainty in the cross section for the non-$\mathrm{t\bar{t}}$, non-QCD multijet backgrounds: 30%

A scale factor to correct for the mismodeling of H_T is applied, and so the uncertainty in this correction is also used as a systematic uncertainty as shown in Fig. 5.32. This scale factor is derived with the N_j = 5, 6, and 7 bins and is then extrapolated to the higher N_j bins. The uncertainty is derived by looking at the difference in the extrapolated scale factor function for the $N_{\mathrm{j}} = 8$ bin to a direct fit of the ratio of collision data to MC data in the $N_{\mathrm{j}} = 8$ bin. The systematic uncertainty related to the H_{T} reweighting for the $\mathrm{t\bar{t}}$+jets background is described in Section A.2.

As mentioned in the beginning of this section, all background components besides $\mathrm{t\bar{t}}$ and QCD multijet processes are grouped into TTX and Other. Systematic uncertainties for these two components are computed in a similar manner as for $\mathrm{t\bar{t}}$, except that the resulting uncertainty is the ratio of the number of events in the varied N_{j} shape to the nominal shape in each NN and N_{j} bin. Results for 2016 data are shown in Fig. 5.33 for the TTX component and in Fig. 5.34 for the Other component. The majority of variations are of the order $\lesssim$10% over all NN and N_{j} bins. Compounded by the fact that the TTX and Other background components make up $\sim$5% of the total contribution in the signal region, the relative size, and thus impact, of these uncertainties is small.

The same treatment of systematics uncertainties performed for the two previously mentioned background components is also performed for the simulated signal events. Figure 5.35 shows these sizes of the systematic uncertainties in the four NN bins for the RPV signal model with $m_{\tilde{\mathrm{t}}} = 350$, 550, and 850 GeV. In general, it is found that the size of any systematic uncertainty variation across the vast majority of N_{j} and NN bins is $\lesssim$10%.

5.6.2 $\mathrm{t\bar{t}}$ *Shape Systematic Uncertainties*

The $\mathrm{t\bar{t}}$ + jets background prediction comes directly from the fit to data. As explained previously in Section 5.4, the key assumption in the fit procedure is that the $\mathrm{t\bar{t}}$ + jets shape is the same in all four NN bins. In the MC data simulation, this

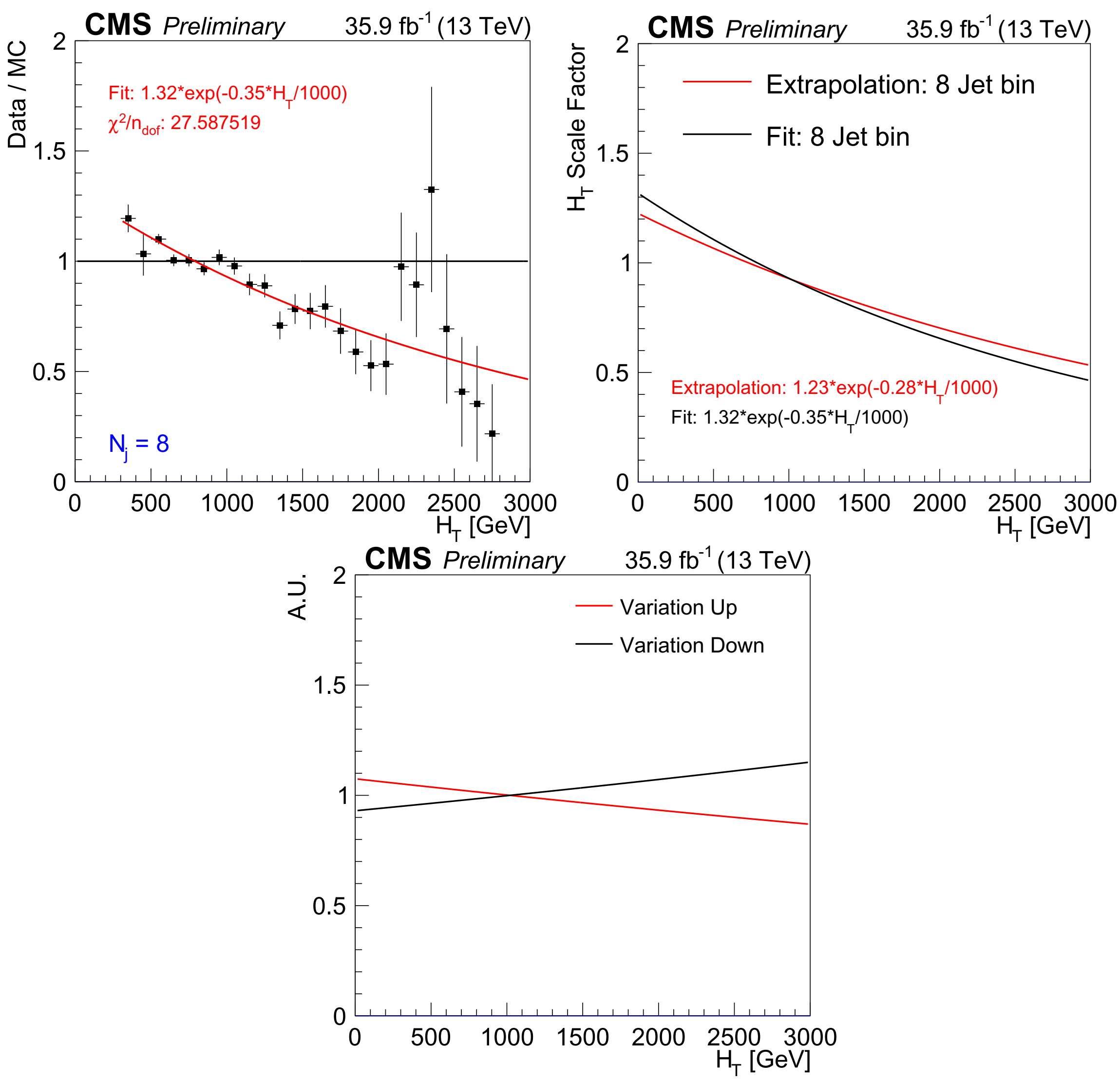


Figure 5.32: (Top left) Fit to the ratio of data and MC data for the $N_{\mathrm{j}} = 8$ bin. (Top right) comparison of the scale factor derived from extrapolating to the $N_{\mathrm{j}} = 8$ bin (red) and fitting the $N_{\mathrm{j}} = 8$ bin directly (black). (Bottom) multiplicative up and down variation to the H_{T} scale factor derived from the $N_{\mathrm{j}} = 8$ bin extrapolation vs. direct fit comparison for 2016.

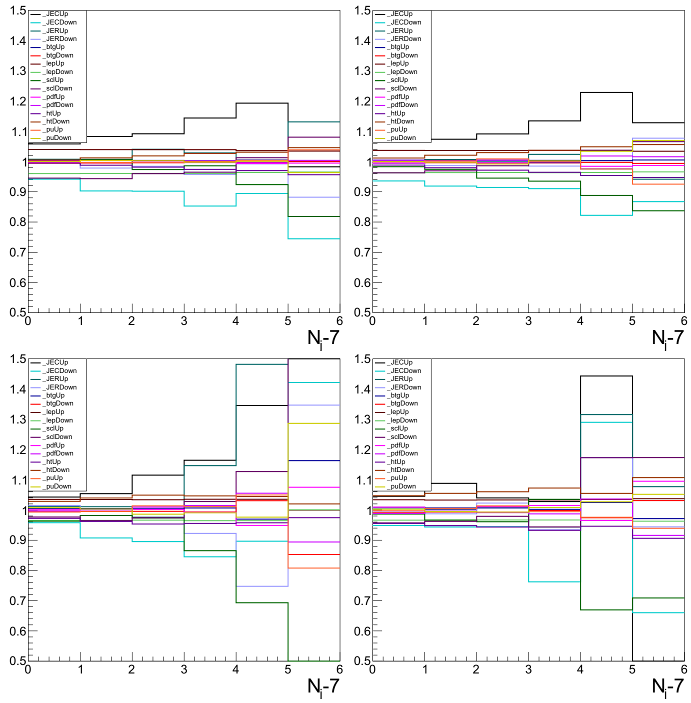


Figure 5.33: The ratio $N_{\mathrm{j}}(\mathrm{syst})/N_{\mathrm{j}}(\mathrm{nom})$ for the various sources of systematic uncertainty is shown for the TTX background component for 2016 data for the four NN bins: D1 (top left), D2 (top right), D3 (bottom left), and D4 (bottom right).

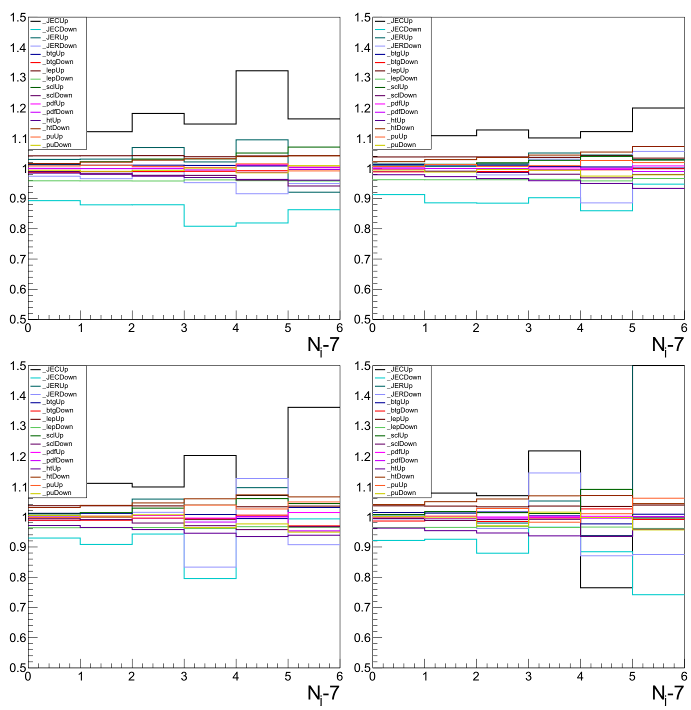


Figure 5.34: The ratio $N_{\mathrm{j}}(\mathrm{syst})/N_{\mathrm{j}}(\mathrm{nom})$ for the various sources of systematic uncertainty is shown for the Other background component for 2016 data for the four NN bins: D1 (top left), D2 (top right), D3 (bottom left), and D4 (bottom right).

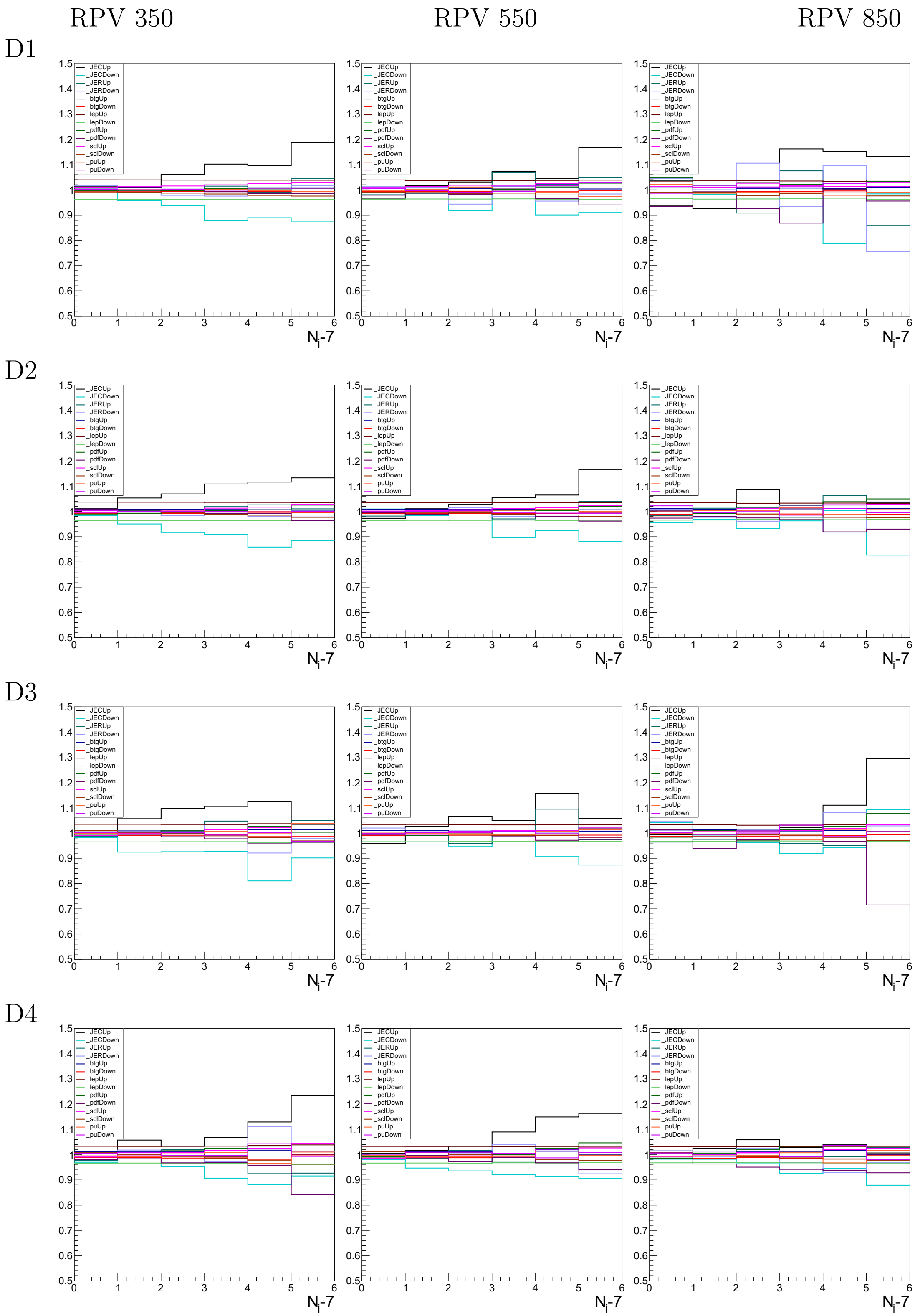


Figure 5.35: The ratio $N_{\mathrm{j}}(\mathrm{syst})/N_{\mathrm{j}}(\mathrm{nom})$ for the various sources of systematic uncertainty are shown for the RPV SUSY model with $m_{\tilde{t}}$ = 350 GeV (left), 550 GeV (middle), and 850 GeV (right) for 2016 data for the four NN bins: D1 (top), D2 (top middle), D3 (bottom middle), and D4 (bottom).

is guaranteed by first removing most of the N_{j} dependence during the NN training and then defining the NN bin edges such that each NN bin D_i contains a certain fraction of the $\mathrm{t\bar{t}}$ background for each N_{j} bin. Of course, differences between collision data and simulation would break this assumption, so it is important to address any of the potential differences as uncertainties in the fit. It is important to note that the only differences relevant here are those that would cause the N_{j} shape to be different between the NN bins. If the difference is such that the overall N_{j} shape is different in collision data and simulation, but the shapes in the different NN bins remain the same, then the fit will be able to adjust to this new shape automatically and there is no need for a systematic uncertainty.

Ideally, the systematic uncertainty related to differences in N_{j} shape between NN bins would be derived from a dedicated $\mathrm{t\bar{t}}+\mathrm{jets}$ control region. However, for this analysis an adequate control region that is also free of the signal has not been identified. Therefore, the analysis proceeds to evaluate many of the systematic uncertainties from variations in the simulation. Variations considered include those related to the b-tagging scale factor uncertainty, lepton identification/isolation/trigger scale factor uncertainty, JEC/JER uncertainties, parton shower uncertainties, PDF uncertainties, renormalization/factorization scale uncertainties, and H_{T} scale factor uncertainties.

The procedure to derive a given systematic uncertainty consists of comparing N_{j} shapes per NN bin before and after applying a given systematic uncertainty variation. Since there is already a small difference between the NN bins in the nominal case where no systematic uncertainty variation is done, only the additional differences are considered as a source of systematic uncertainty. For systematic uncertainty variations where only the event weight is changed, such as for the b-tagging scale factors, the simulated N_{j} distributions are directly compared. The exact procedure is as follows:

(1) Consider the N_{j} shape in bin D_i, and normalize it to unit area

(2) Consider the total N_{j} shape obtained by summing across all four D_i regions, and normalize it to unit area

(3) Take the ratio of these two shapes ($R = N_{\mathrm{j}}^{D_i}/N_{\mathrm{j}}^{\mathrm{total}}$) as a way to assess the difference in N_{j} shape between NN bins

The results of this procedure for 2016 are shown in Fig. 5.36 for the effect from uncertainties in the b-tagging efficiency scale factor, in Fig. 5.37 for the effect from uncertainties in the lepton ID/isolation scale factors, in Fig. 5.39 for the H_{T} scale factor uncertainty, in Fig. 5.41 for the effect from varying the PDFs, and in Fig. 5.40 for the factorization/normalization scale uncertainty. As can be seen, the effect is very small, less than 1–2% in most cases, with the exception of the scale variation for bin D3 which reaches 15% at the highest N_{j}, and the PDF uncertainty for D4 which reaches 5%. For cases where there are no events in the last bin, e.g. for $N_{\mathrm{j}} \geq 12$ in bin D4, the systematic uncertainty is taken to be same as that of the preceding N_{j} bin.

To assess the effect of varying the jet energy scale and resolution, it is important to note that these corrections can cause bin migrations. When the transverse momentum of a jet is varied, the total number of jets above threshold can change, and the NN output value will also change (the jet momenta are inputs to the NN). Therefore, a direct comparison of the N_{j} shape in the different NN bins cannot be simply done because it would cause the analysis to be strongly affected by statistical fluctuations in the tails of the N_{j} distribution. To counteract this, a modification to the procedure outlined above is made. Instead of directly using the simulated N_j shape, a background-only fit is performed instead to determine the N_j shape.

The procedure thus becomes:

(1) Fit the N_j shape in each bin D_i separately, i.e. each bin D_i has its own set of shape parameters $a_0^{D_i},\ a_1^{D_i},\ d^{D_i}$

(2) Fit the N_j shape in all NN bins together, using just a single set of parameters $a_0^{D_i},\ a_1^{D_i},\ d^{D_i}$ as is done in the default fit.

(3) Take ratio of these fitted distributions ($R = N_\mathrm{j}^{D_i}/N_\mathrm{j}^{\mathrm{total}}$) as a way to assess the difference in N_j shape between NN bins

The results for the jet energy scale uncertainty are shown in Fig. 5.42. Analogous plots for the jet energy resolution are shown in Fig. 5.43. The nominal shape difference itself is also taken as a systematic uncertainty, see Fig. 5.46, and its size is small (1–5%). The effects of varying the parton shower settings for ISR and FSR are estimated in the same way as the effects of JEC and JER. Figures. 5.44 and 5.45 show the ISR and FSR systematic uncertainties for 2016. Further derivations of systematic uncertainties are discussed in the Appendix.

5.6.3 Systematic Uncertainty Correlations between Years

To combine the data from four different data periods, also known as “eras”, assumptions need to be made on how to correlate systematic uncertainties between different data eras. Wherever possible, the analysis follows internal CMS recommendations.

$\mathrm{t\bar{t}}$ shape uncertainties

- *JEC*: Treated them as fully uncorrelated to be conservative.

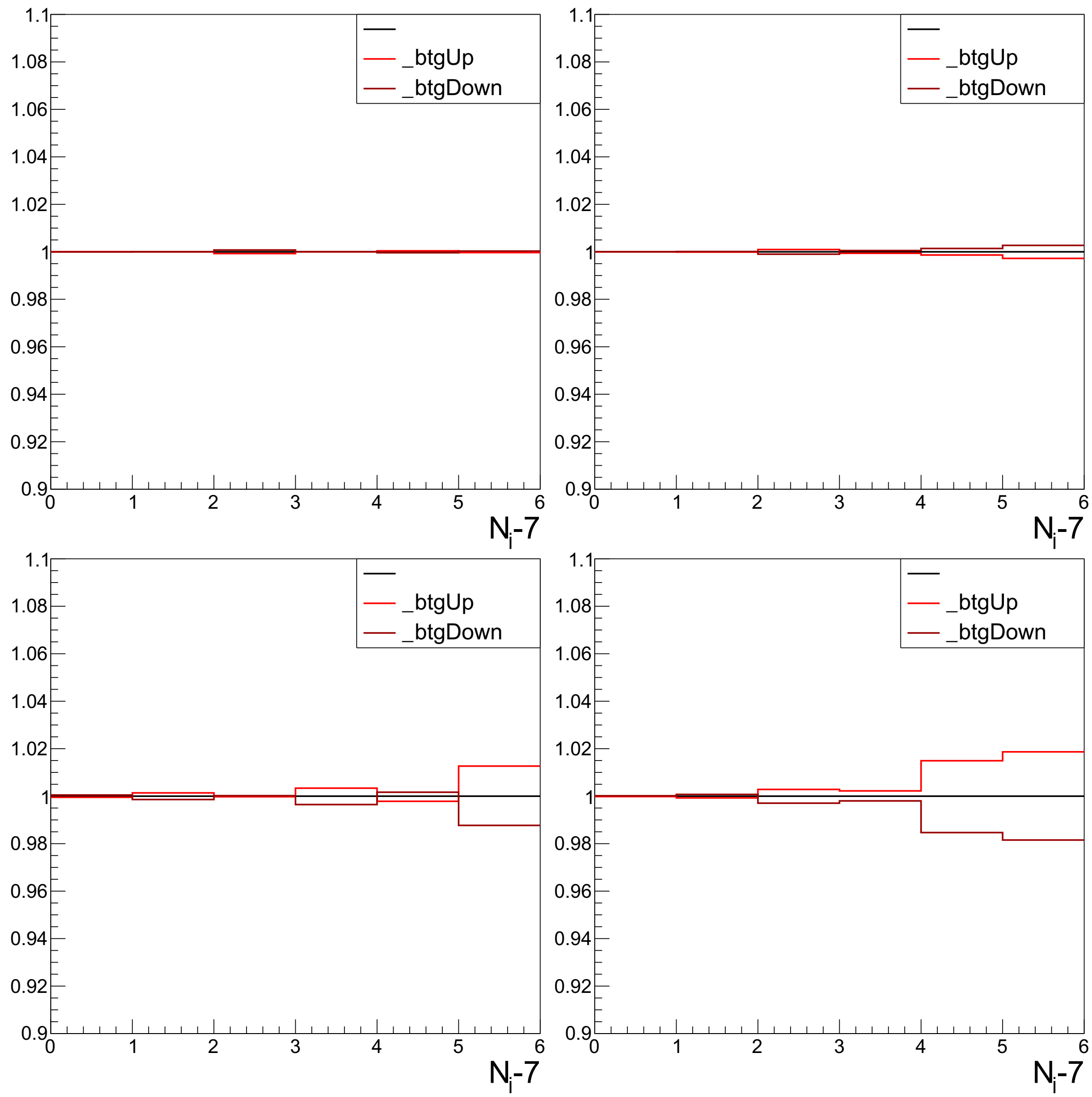


Figure 5.36: Ratio $R = R_{\text{syst}}/R_{\text{nom}}$ showing the introduced N_{j} shape differences by varying the b-tagging efficiency scale factor up and down by one standard deviation for the 4 NN bins: D_1 (top left), D_2 (top right), D_3 (bottom left), and D_4 (bottom right). The x axis shows N_{j} shifted down such that $N_{\text{j}} = 7$ is the first bin, here noted by 0.

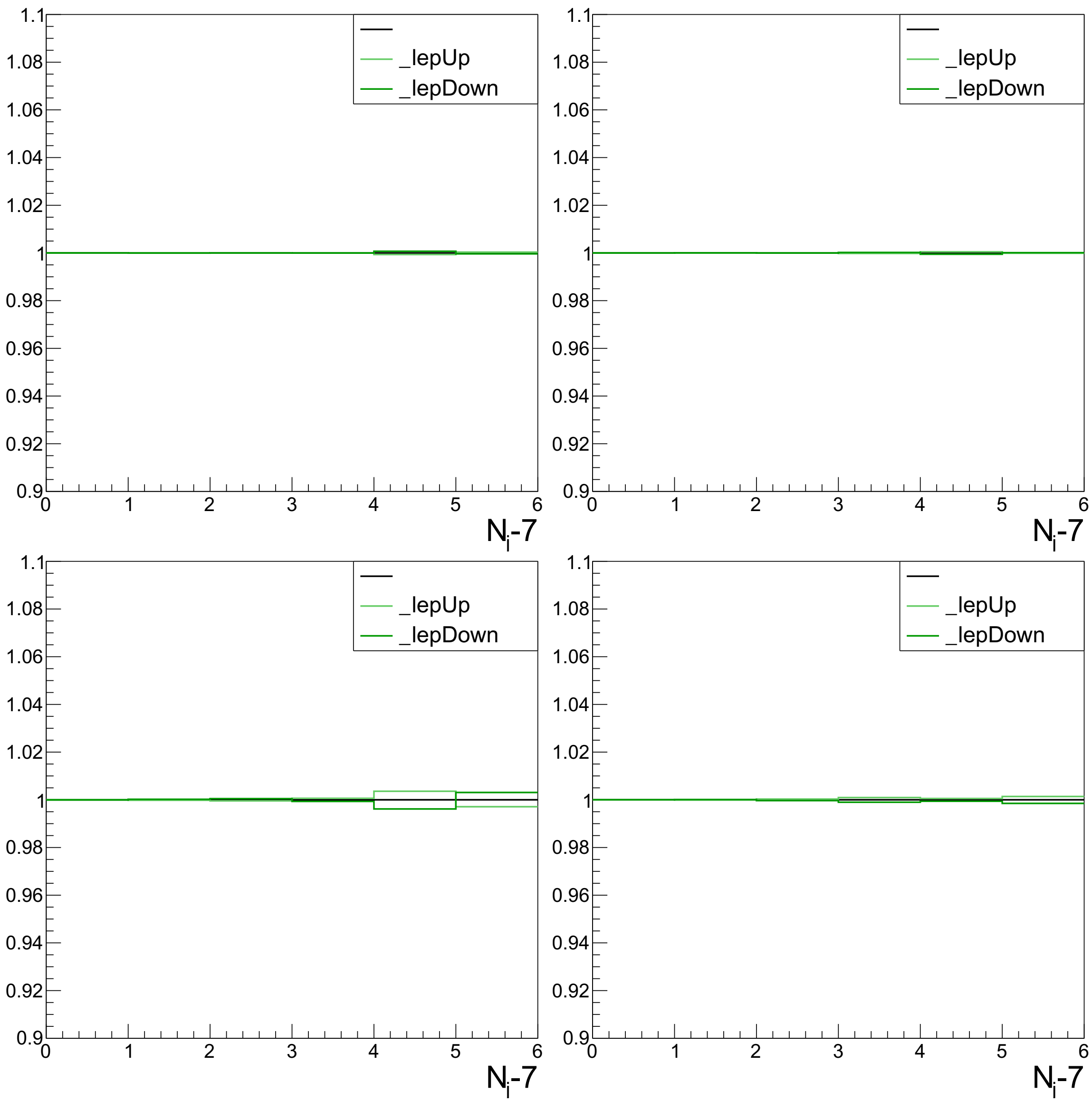


Figure 5.37: Ratio $R = R_{\text{syst}}/R_{\text{nom}}$ showing the introduced N_{j} shape differences by varying the lepton identification/isolation efficiency scale factor up and down by one standard deviation for the 4 NN bins: D_1 (top left), D_2 (top right), D_3 (bottom left), and D_4 (bottom right). The x axis shows N_{j} shifted down such that $N_{\text{j}} = 7$ is the first bin, here noted by 0.

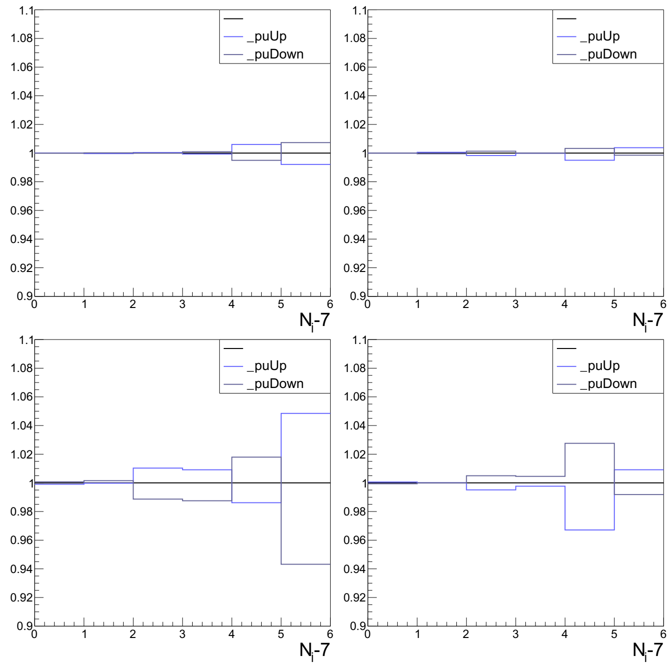


Figure 5.38: Ratio $R = R_{\mathrm{syst}}/R_{\mathrm{nom}}$ showing the introduced N_{j} shape differences by varying the pileup scale factor up and down by one standard deviation for the 4 NN bins: D_1 (top left), D_2 (top right), D_3 (bottom left), and D_4 (bottom right). The x axis shows N_{j} shifted down such that $N_{\mathrm{j}} = 7$ is the first bin, here noted by 0.

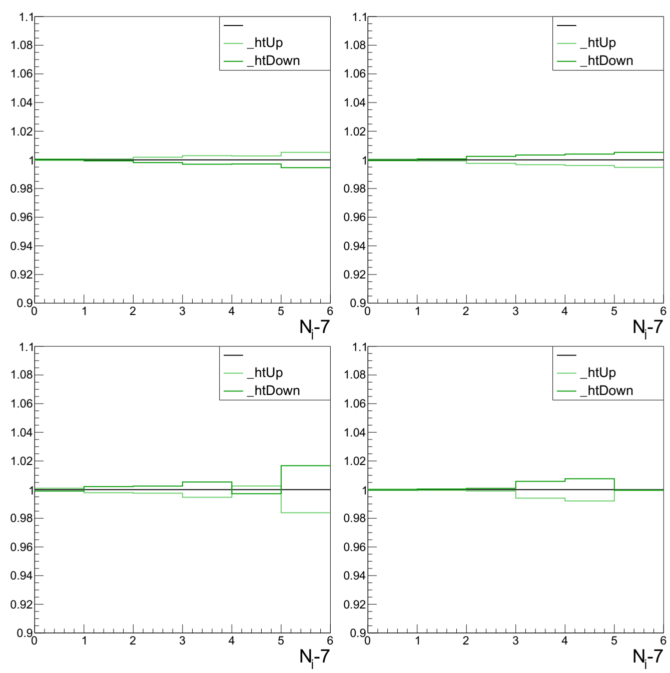


Figure 5.39: Ratio $R = R_{\mathrm{syst}}/R_{\mathrm{nom}}$ showing the introduced N_{j} shape differences by varying the H_{T} scale factor up and down by one standard deviation for the 4 NN bins: D_1 (top left), D_2 (top right), D_3 (bottom left), and D_4 (bottom right). The x axis shows N_{j} shifted down such that $N_{\mathrm{j}} = 7$ is the first bin, here noted by 0.

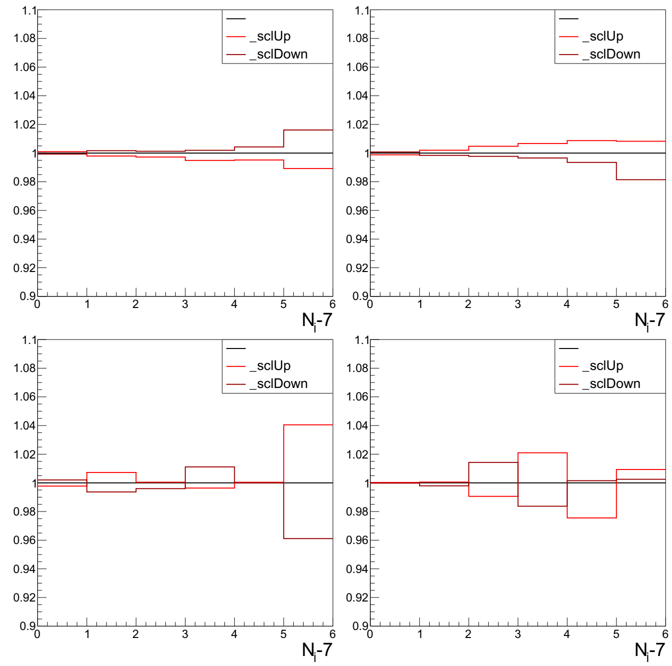


Figure 5.40: Ratio $R = R_{\mathrm{syst}}/R_{\mathrm{nom}}$ showing the introduced N_{j} shape differences by varying the factorization and normalization scales up and down by a factor of 2 for the 4 NN bins: D_1 (top left), D_2 (top right), D_3 (bottom left), and D_4 (bottom right). The x axis shows N_{j} shifted down such that $N_{\mathrm{j}} = 7$ is the first bin, here noted by 0.

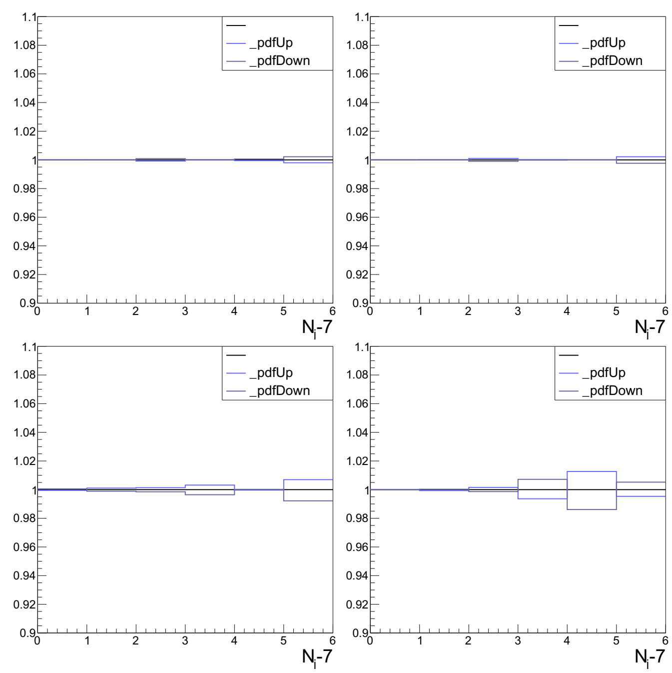


Figure 5.41: Ratio $R = R_{\mathrm{syst}}/R_{\mathrm{nom}}$ showing the introduced N_{j} shape differences by varying the parton distribution functions for the 4 NN bins: D_1 (top left), D_2 (top right), D_3 (bottom left), and D_4 (bottom right). The x axis shows N_{j} shifted down such that $N_{\mathrm{j}} = 7$ is the first bin, here noted by 0.

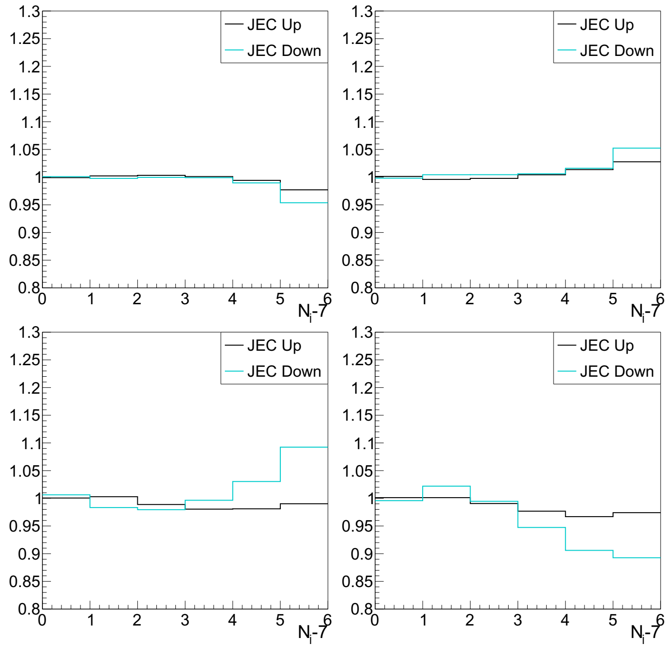


Figure 5.42: Ratio $R = R_{\mathrm{syst}}/R_{\mathrm{nom}}$ showing the introduced N_{j} shape differences by varying the jet energy scale up and down by one standard deviation for the 4 NN bins: D_1 (top left), D_2 (top right), D_3 (bottom left), and D_4 (bottom right). The x axis shows N_{j} shifted down such that $N_{\mathrm{j}} = 7$ is the first bin, here noted by 0.

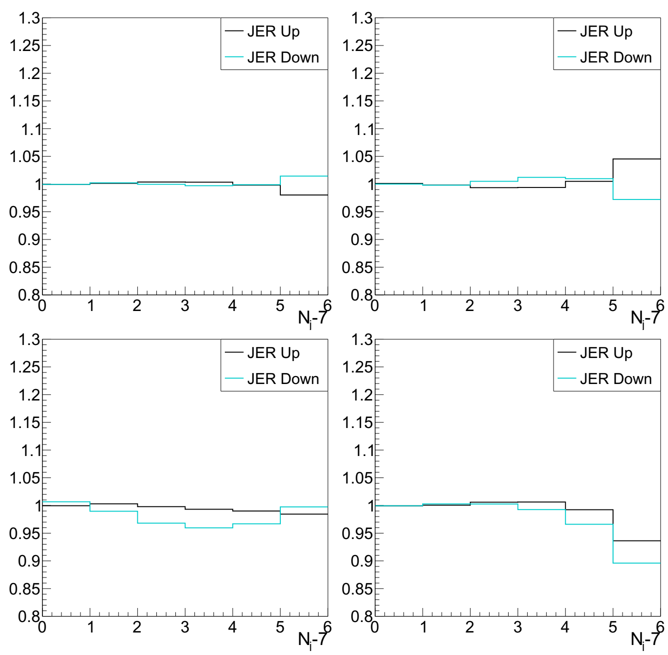


Figure 5.43: Ratio $R = R_{\mathrm{syst}}/R_{\mathrm{nom}}$ showing the introduced N_{j} shape differences by varying the jet energy resolution up and down by one standard deviation for the 4 NN bins: D_1 (top left), D_2 (top right), D_3 (bottom left), and D_4 (bottom right). The x axis shows N_{j} shifted down such that $N_{\mathrm{j}} = 7$ is the first bin, here noted by 0.

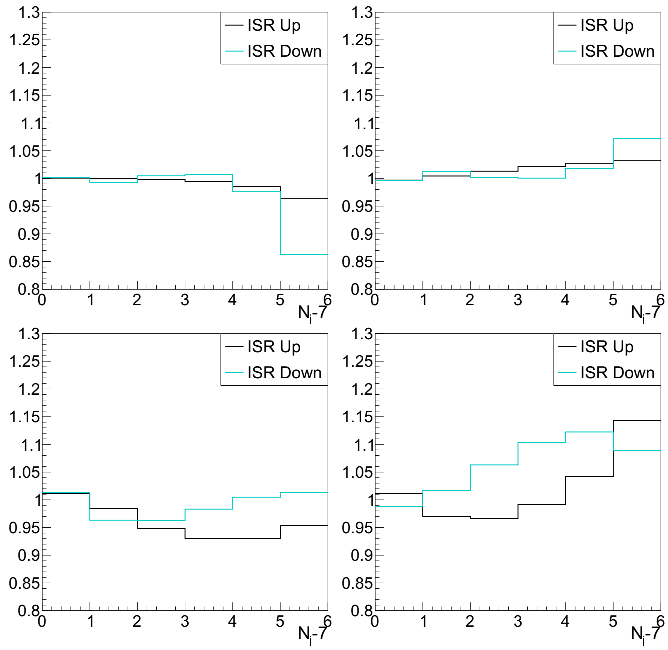


Figure 5.44: Ratio $R = R_{\mathrm{syst}}/R_{\mathrm{nom}}$ showing the introduced N_{j} shape differences by varying the ISR settings in the parton shower up and down by a factor of 2 for the 4 NN bins: D_1 (top left), D_2 (top right), D_3 (bottom left), and D_4 (bottom right). The x axis shows N_{j} shifted down such that $N_{\mathrm{j}} = 7$ is the first bin, here noted by 0.

- *JER*: Treated as uncorrelated between years
- *Lepton ID/isolation/trigger*: In this case, the uncertainty on the trigger efficiency scale factor is dominant, and it is of statistical nature. Therefore, the analysis treats this uncertainty as uncorrelated between the years.
- b-*tagging efficiency*: No official recommendation available so far. They are treated as uncorrelated for now because of the different pixel detector in 2016 and 2017.
- *Pileup*: Treated as uncorrelated between years.
- *PDF*: Treated as correlated for 2017, 2018pre, and 2018post, but uncorrelated for 2016, since a different PDF is used in 2016.
- *Scale*: Treated as correlated.

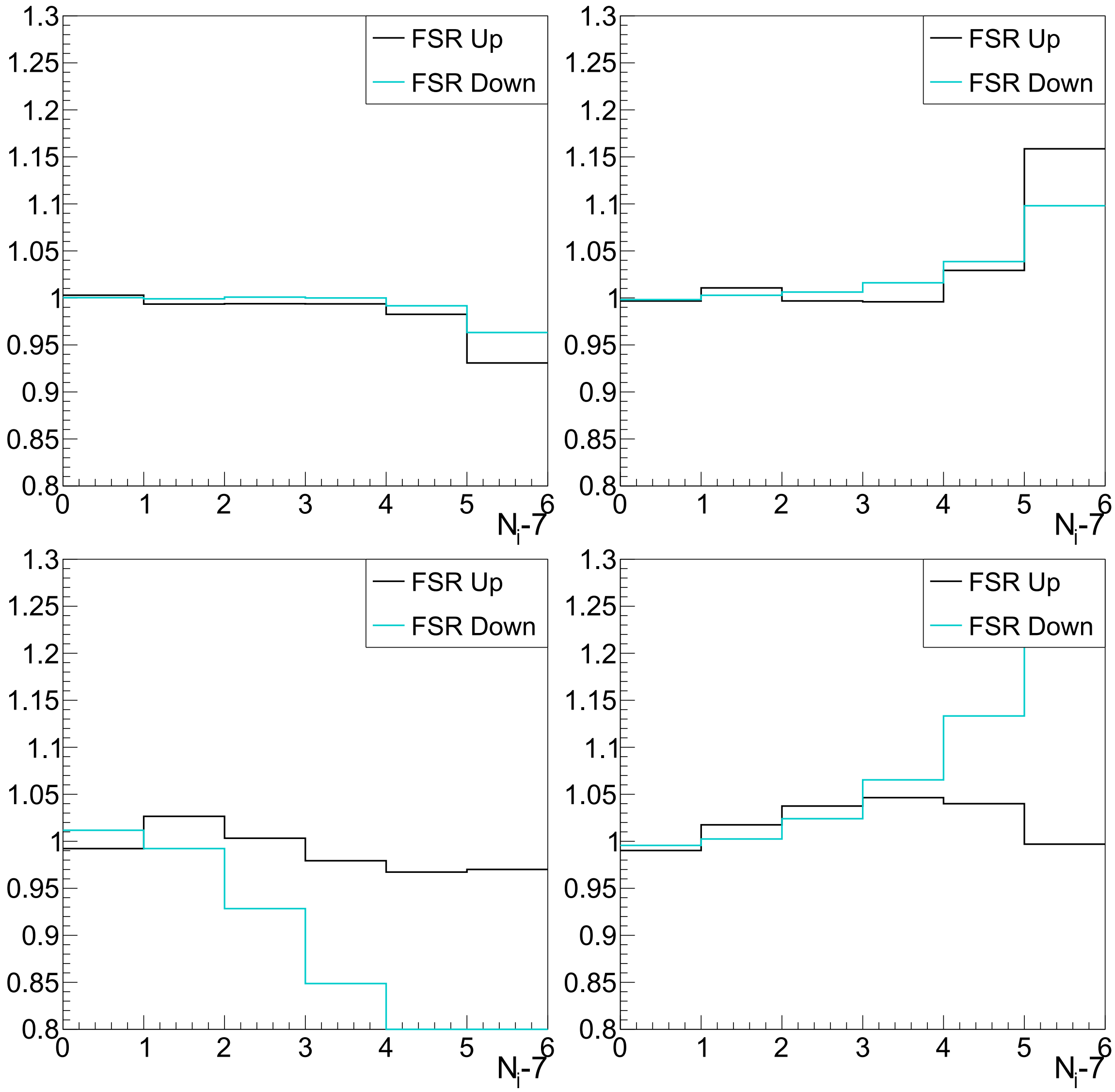


Figure 5.45: Ratio $R = R_{\text{syst}}/R_{\text{nom}}$ showing the introduced N_{j} shape differences by varying the FSR settings in the parton shower up and down by a factor of 2 for the 4 NN bins: D_1 (top left), D_2 (top right), D_3 (bottom left), and D_4 (bottom right). The x axis shows N_{j} shifted down such that $N_{\text{j}} = 7$ is the first bin, here noted by 0. For inclusion in the data card, the FSR systematic uncertainty has been scaled down by a factor of $\sqrt{2}$ in order to include only the recommended variations up and down by a factor of $\sqrt{2}$.

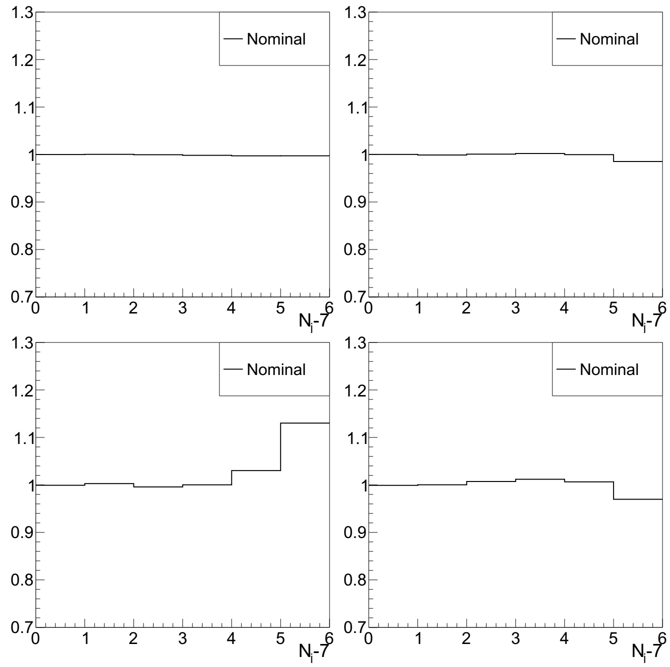


Figure 5.46: Ratio R_{nom} showing the N_{j} shape differences between the 4 NN bins without applying any systematic uncertainty variation: D_1 (top left), D_2 (top right), D_3 (bottom left), and D_4 (bottom right). The x axis shows N_{j} shifted down such that $N_{\mathrm{j}} = 7$ is the first bin, here noted by 0.

- *ISR/FSR*: Treated as correlated for 2017, 2018pre, and 2018post, but uncorrelated for 2016, since a different tune is used in 2016.
- *Underlying event*: Treated as correlated for 2017, 2018pre, and 2018post, but uncorrelated for 2016, since a different tune is used in 2016.
- *Color reconnection*: Treated as correlated for 2017, 2018pre, and 2018post, but uncorrelated for 2016, since a different tune is used in 2016.
- *ME-PS matching scale (hdamp)*: Treated as correlated for 2017, 2018pre, and 2018post, but uncorrelated for 2016, since a different tune is used in 2016.
- *Jet mass-p_{T} rescaling*: Treated as uncorrelated.
- *No H_{T} reweighting*: Treated as uncorrelated.

- *Nominal* $t\bar{t}$ *shape difference*: Treat as uncorrelated between years because of the use of different generator settings resulting in different $t\bar{t}$ N_j shapes, as well as the use of a different NN training

Signal and non-$t\bar{t}$ background uncertainties

- *Luminosity*: Treated as uncorrelated between years
- *JEC*: They are treated as fully uncorrelated to be conservative.
- *JER*: Treated as uncorrelated between years
- *Lepton ID/isolation/trigger*: In this case, the uncertainty on the trigger efficiency scale factor is dominant, and it is of statistical nature. Therefore, the analysis treats this uncertainty as uncorrelated.
- b-*tagging efficiency*: No official recommendation available so far. They are treated as uncorrelated for now because of the different pixel detector in 2016 and 2017.
- *Pileup*: Treated as uncorrelated between years
- *PDF*: Treated as correlated for 2017, 2018pre, and 2018post, but uncorrelated for 2016, since a different PDF is used in 2016.
- *Scale*: Treated as correlated
- *Non-*$t\bar{t}$ *background cross section*: Fully correlated between years, but different for each category of backgrounds.

5.7 Results and Interpretation

At the end of the day, we are just counting the number of collision data events observed and comparing it to the expected number of background events. Determining whether or not the data deviates from the SM background is a question best answered using statistics. It is assumed that the number of collision data events follows "counting" or Poisson statistics, and that the statistical significance of an observed signal can be quantified by means of a p_0-value. If there is no statistically significant deviation from the background, the upper limit on the possible cross section of a SUSY signal is obtained using the $\mathrm{CL_s}$ method, which is used by all LHC experiments. This method is based on a frequentist significance test using a likelihood ratio as a test statistic. This section will cover the results of our statistical evaluation, but first it will start with a brief overview of the methods we use. A more detailed derivation can be found in [46, 47].

5.7.1 Formalism for Statistical Interpretation

In particle physics, when a claim is made that some new physics phenomenon has been observed, it is usually based on the statistical significance obtained from the data analysis. The significance Z is defined as the number of standard deviations away from the mean of a normal Gaussian distribution for an upper-tail probability equal to the p_0-value. It can be written as

$$Z = \Phi^{-1}(1 - p_0) = \sqrt{2}\sum_{k=0}^{\infty} \frac{c_k}{2k+1}\left(\frac{\sqrt{\pi}}{2}(1 - 2p_0)\right)^{2k+1} \tag{5.23}$$

for c_k defined as

$$c_k = \sum_{m=0}^{k-1} \frac{c_m c_{k-1-m}}{(m+1)(2m+1)} = \left\{1, 1, \frac{7}{6}, \frac{127}{90}, ...\right\} \tag{5.24}$$

where Φ^{-1} is the inverse of the cumulative distribution function of the standard Gaussian distribution. It is generally accepted, arbitrarily, that $Z = 5$ ($p_0 = 2.87 \times 10^{-7}$) is needed to claim that a deviation from the background is significant enough for a discovery.

In order to calculate the p_0-value, the likelihood function must first be built. The likelihood is related to the probability of the observed number of events given the expected number of events. It is used to perform a fit to the data where the value of the likelihood is maximized. It is a function of a parameter of interest (POI), in this case the signal strength μ defined as the ratio of the observed cross section divided by the nominal cross section, and *nuisance* parameters whose values are not taken as known *a priori* and must be obtained from fits to the data. For a binned analysis with i number of bins, j number of nuisance parameters θ, *prior* nuisance parameter values $\tilde{\theta}$, and signal strength μ, the likelihood function is defined as

$$\mathcal{L}(\text{data}|\,\mu,\theta_j) = \text{Poisson}(\text{data}\,|\,\mu s_i(\theta_j) + b_i(\theta_j))\;\pi(\tilde{\theta}_j,\theta_j) \tag{5.25}$$

For the analysis described in this thesis, the Poisson nuisance terms π, predicted signal s_i, and predicted background b_i are defined respectively as

$$\text{Poisson}(\text{data}\,|\,\mu s_i + b_i) = \prod_i \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)} \tag{5.26}$$

$$\pi(\tilde{\theta}_j,\theta_j) = \prod_j \frac{1}{\sigma(\tilde{\theta}_j)\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\theta_j-\tilde{\theta}_j}{\sigma(\tilde{\theta}_j)}\right)^2} \tag{5.27}$$

$$s_i(\theta_j) = s_{0,i} \prod_j R_{ij}^{\theta_j}(s) \tag{5.28}$$

$$b_i(\theta_j) = \sum_{x\,\in\,BG} \left(x_{0,i} \prod_j R_{ij}^{\theta_j}(x) \right) \tag{5.29}$$

where n_i is the observed number of data events per i^{th} bin, all of the nuisance parameters have a prior mean $\tilde{\theta}_j$ and sigma $\sigma(\tilde{\theta}_j)$, R_{ij} are the signal and background dependent systematic uncertainty correction values for each bin and nuisance parameter, BG is a set of all backgrounds for the analysis, and $x_{0,i}$ ($s_{0,i}$) are the nominal prediction for a background (signal). The values of $x_{0,i}$ often come from MC data. For the $\mathrm{t\bar{t}}$ background, the $x_{0,i}$ values come from the fit function defined in Section 5.4.

The $t\bar{t}$ fit function's parameters are free during the fit and are considered to be in the set of nuisance parameters θ_j. For the QCD background, the $x_{0,i}$ are taken directly from the prediction obtained based on the QCD control region data described in Section 5.5.

In general, to set limits on the signal strength μ, we are interested in the compatibility of the data with the *background-only* and the *signal + background* hypotheses, where the signal is allowed to scale by μ. With this in mind, we can construct a test statistic q_μ that depends on the value of μ, defined as

$$q_\mu = -2 \ \ln \frac{\mathcal{L}(\text{data} \,|\, \mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{data} \,|\, \hat{\mu}, \hat{\theta})}, \text{ with a constraint } 0 \leq \hat{\mu} \leq \mu \tag{5.30}$$

where a parameter with a hat indicates that it is profiled in its respective fit. Note that the values of the nuisance parameters in each likelihood function are not the same since they can change depending on whether signal is allowed to float in the fit or is fixed. When performing the fits to data there ends up being a value for the test statistic called the *observed* value q_μ^{obs} that depends on the fixed μ value. The value of q_0^{obs} is used to determine the significance Z as shown later. To extract useful information from the observed value, we need to determine how rare it is. This can be done by taking the data and best-fit values for the nuisance parameters, and producing a large amount of pseudo-data by varying each parameter within its range of uncertainty to construct probability distribution functions (PDFs) of the test statistics by performing fits to these pseudo-experiments.

To determine the p_0-value, a special case of these PDFs for the *background-only* hypothesis with μ fixed to zero, $f(q_0 | \, 0, \hat{\theta}_0^{\text{obs}})$, are produced. In this case, pseudo-experiments are generated from data without signal injected. The p_0-value is then defined as

$$p_0 = \int_{q_0^{\rm obs}}^{\infty} f(q_0|\,0, \hat{\theta}_0^{\rm obs})\, dq_0. \tag{5.31}$$

We can use the same formalism to define an upper limit on the SUSY cross section by calculating the $\mathrm{CL_s}$ values. The upper limit is calculated with the test statistics coming from the *signal + background* hypothesis $f(q_\mu|\,\mu, \hat{\theta}_\mu^{\rm obs})$ and the *background-only* hypothesis $f(q_\mu|\,0, \hat{\theta}_0^{\rm obs})$. Note that $f(q_\mu|\,\mu, \hat{\theta}_\mu^{\rm obs})$ means that pseudo-experiments were made by injecting signal at some μ value, and the same μ value is used in the likelihood calculation in the numerator of the test statistic. Similarly, $f(q_\mu|\,0, \hat{\theta}_0^{\rm obs})$ means that pseudo-experiments were made without injecting any signal, and μ is set to zero in the numerator of the test statistic. We can define two p-values from these PDFs that can be used to determine the $\mathrm{CL_s}$ values:

$$p_\mu = \int_{q_\mu^{\rm obs}}^{\infty} f(q_\mu|\,\mu, \hat{\theta}_\mu^{\rm obs})\, dq_\mu \tag{5.32}$$

$$1 - p_b = \int_{q_\mu^{\rm obs}}^{\infty} f(q_\mu|\,0, \hat{\theta}_0^{\rm obs})\, dq_\mu. \tag{5.33}$$

See Fig. 5.47 for an example of these distributions for fixed μ during the fits. These PDFs show the *background-only* and *signal + background* hypothesis distributions. $\mathrm{CL_s}$ is now defined by taking the ratio of these two values

$$\mathrm{CL_s}(\mu) = \frac{p_\mu}{1 - p_b} \leq 1 - \alpha \tag{5.34}$$

where α is the confidence level for which the results are quoted. For LHC experiments, results are often quoted at the 95% confidence level ($\alpha = 0.95$). One then can claim an upper limit for the value of μ at the 95% confidence level, defined as $\mu_{\rm up}$, that yields $\mathrm{CL_s} \leq 0.05$. Finally, it is customary to quote the 1σ and 2σ uncertainties on the upper limit $\mu_{\rm up}$ by examining the distribution of $\mu_{\rm up}$, which is often a Gaussian distribution. We find the values of $\mu_{\rm up}$ that yield 68% (1σ) and 95% (2σ) of the distribution, as shown in Fig. 5.48.

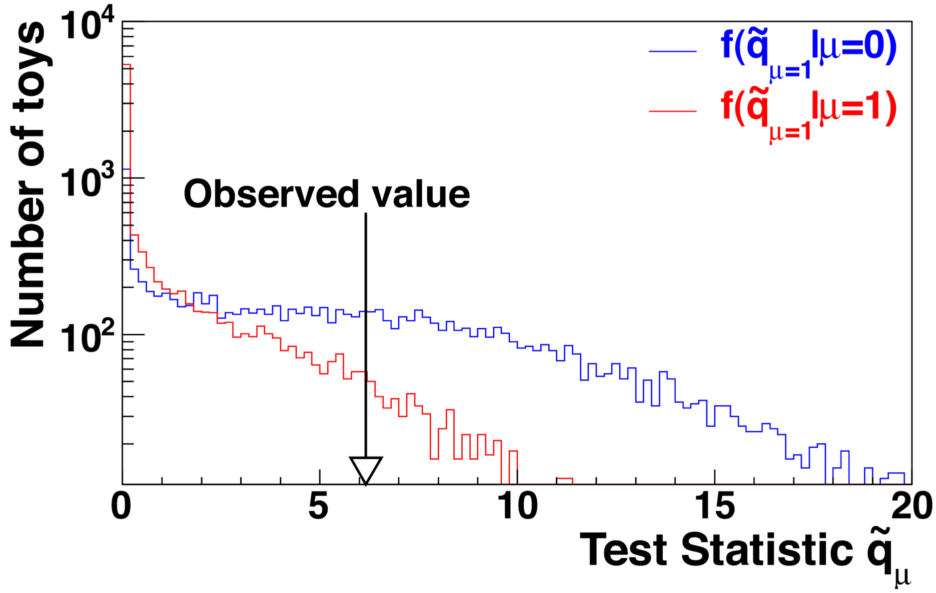


Figure 5.47: Distributions of the test statistic in pseudo-experiments for the *background-only* hypothesis (blue) and *signal + background* hypothesis (red) with μ set to 1 in the fit. The observed test statistic value is shown and the curves are integrated above this value to determine the needed p-values.

Finally, as you can imagine, producing these distributions of pseudo-data can be computationally expensive. The asymptotic approximation can be taken assuming a reasonable number of data events, and the observed significance Z and $\mathrm{CL_s}$ can be calculated using

$$Z = \sqrt{q_0^{\rm obs}} \tag{5.35}$$

$$\mathrm{CL_s}(\mu) = \frac{1-\Phi(\sqrt{q_\mu})}{\Phi(\sqrt{q_{\mu,A}}-\sqrt{q_\mu})} \tag{5.36}$$

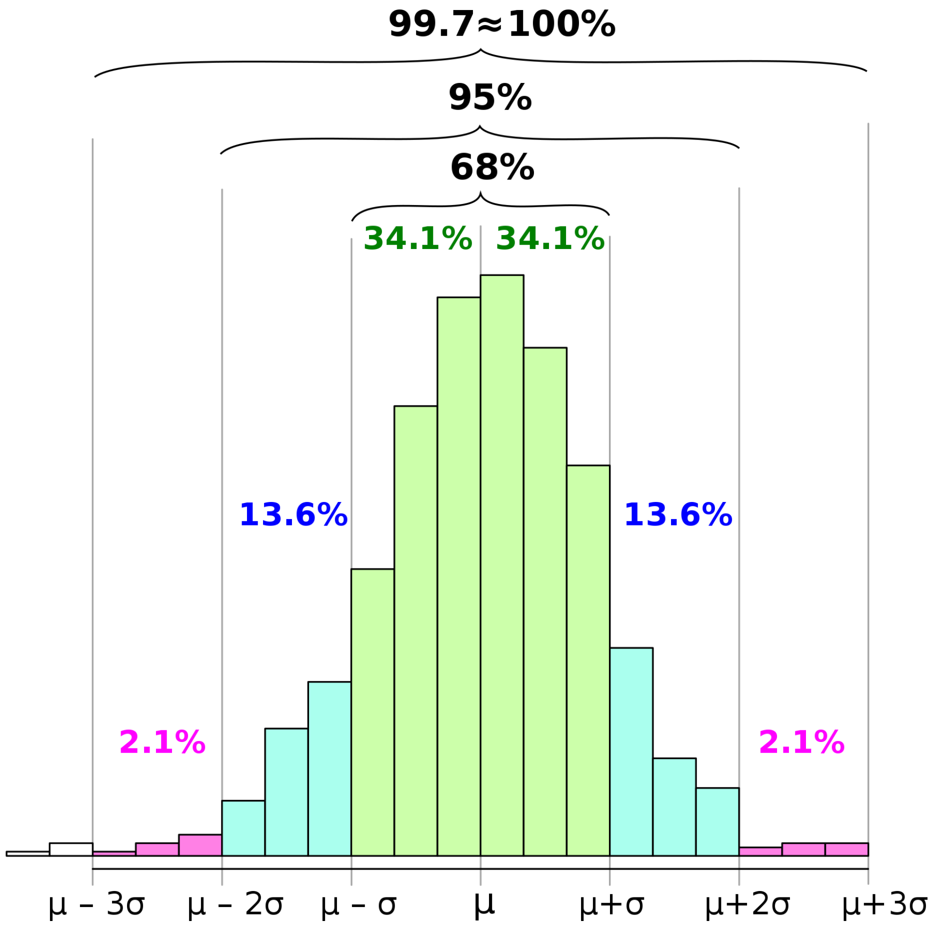


Figure 5.48: Distribution of μ_{up} measured for a set of fits using pseudo-data, showing the 1σ, 2σ, and 3σ bands that correspond to 68%, 95%, and 99.7% of the area of the distribution, respectively. Image source [48].

where Φ is the cumulative distribution function of the standard Gaussian distribution, and $q_{\mu,A}$ is the test statistic obtained from a fit to the Asimov data[1] rather than the observed data. The results quoted in the later sections utilize the asymptotic approximation.

5.7.2 Fits to Individual Years and Full Combined Data

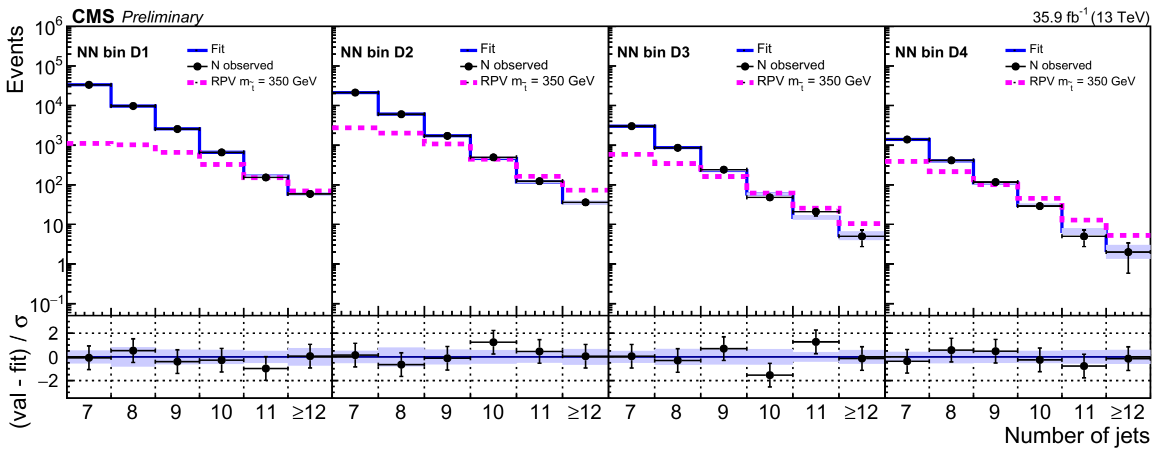
The fit results for both the background-only fit and the signal + background fit for the RPV SUSY model with top squark mass of 350 GeV for 2016, 2017, 2018pre, and 2018post data are shown in Figs. 5.49, 5.50, 5.51, and 5.52, respectively. The best-fit signal strengths and associated significances are shown in Tables 5.10, 5.11, 5.12, and 5.13. In addition to the best-fit signal strength, the corresponding observed significance is also shown.

The fit results for the combined fit of 2016, 2017, 2018pre, and 2018post data are shown in Figs. 5.53 and 5.54. A summary of the best-fit signal strengths and associated significances is shown in the lower section of Tables 5.10, 5.11, 5.12, and 5.13. The $\mathrm{t\bar{t}}$ fit shape parameters are listed in Table 5.14.

5.7.3 Observed Limits on SUSY Cross Sections

Observed significances, along with the best-fit signal strength, computed using combined 2016, 2017, 2018pre, and 2018post data are shown in Table 5.15. Since the 2018 data set is split into the pre-HEM and the post-HEM portions, 2017 data have the highest luminosity of the four eras. No signal yields a significance above 3σ indicating that the background-only hypothesis is a reasonable model for the

[1] The Asimov data set consists of the expected background with the nominal nuisance parameters, setting all fluctuations to be zero.



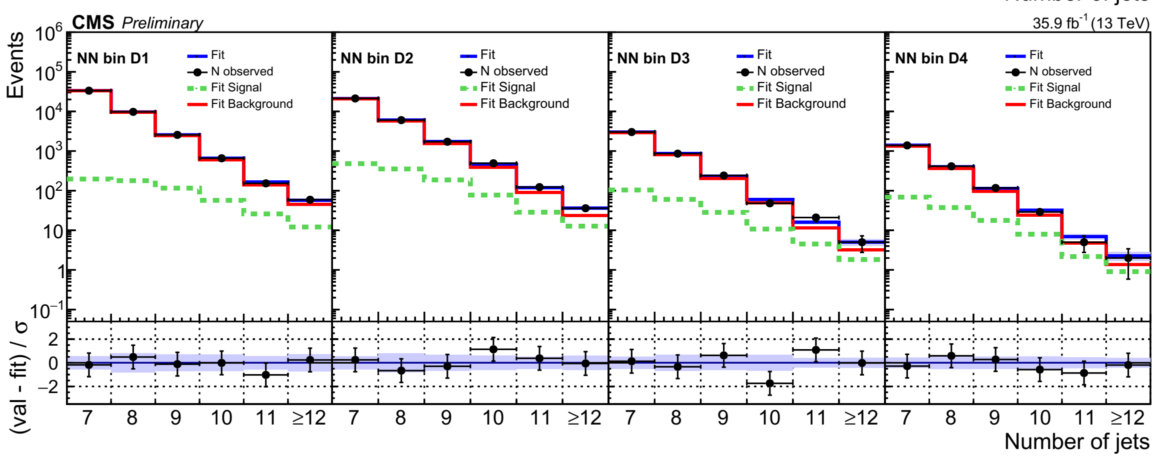


Figure 5.49: Post-fit predictions in logarithmic scale, along with the pulls, from the background-only (top) and signal + background (bottom) fits to 2016 data.

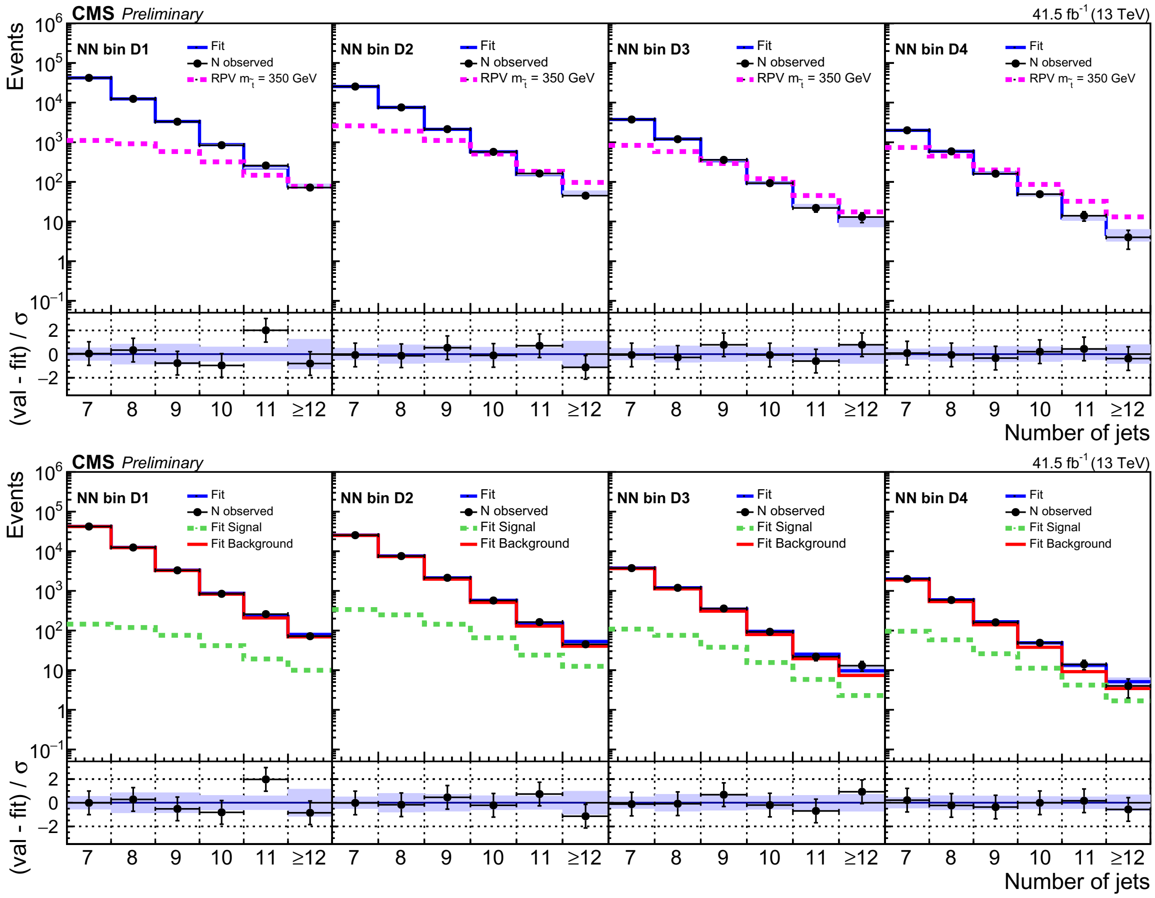


Figure 5.50: Post-fit predictions in logarithmic scale, along with the pulls, from the background-only (top) and signal + background (bottom) fits to 2017 data.

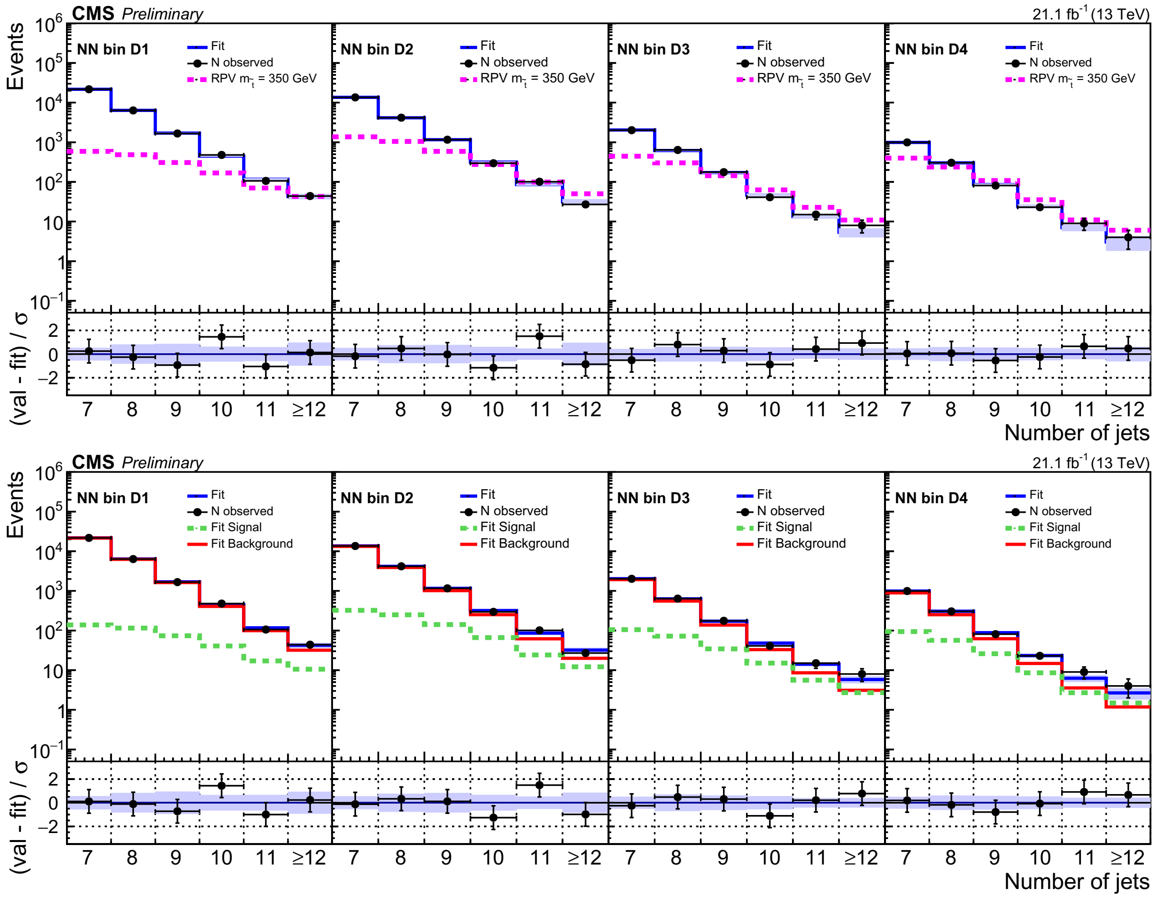


Figure 5.51: Post-fit predictions in logarithmic scale, along with the pulls, from the background-only (top) and signal + background (bottom) fits to 2018pre data.

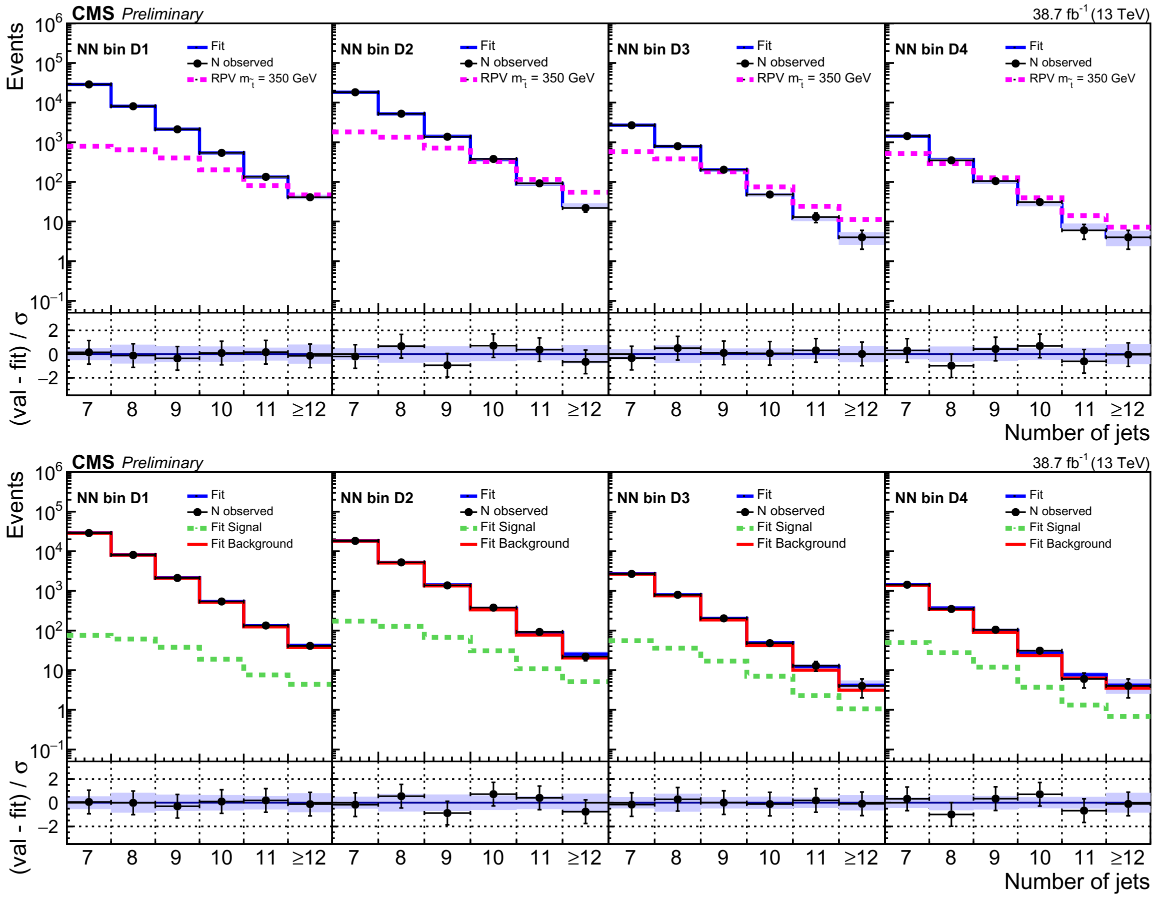


Figure 5.52: Post-fit predictions in logarithmic scale, along with the pulls, for the background-only (top) and signal + background (bottom) fits to 2018post data.

Table 5.10: Best-fit signal strength and corresponding observed significance for 2016 and 2017 data with different top squark mass points of the RPV SUSY model.

Mass	Best fit signal strength	Observed significance	p-value
		2016	
300	$0.17^{0.14}_{-0.17}$	0.81	0.209247
350	$0.18^{0.11}_{-0.14}$	1.20	0.115306
400	$0.19^{0.12}_{-0.14}$	1.33	0.0919423
450	$0.22^{0.14}_{-0.15}$	1.39	0.0818709
500	$0.29^{0.18}_{-0.19}$	1.48	0.0698057
550	$0.26^{0.23}_{-0.24}$	1.06	0.143891
600	$0.38^{0.30}_{-0.32}$	1.16	0.12223
650	$0.41^{0.39}_{-0.41}$	0.97	0.165662
700	$0.65^{0.58}_{-0.62}$	1.04	0.148284
750	$0.78^{0.79}_{-0.78}$	0.91	0.180499
800	$1.04^{1.16}_{-1.04}$	0.85	0.197531
850	$1.87^{1.83}_{-1.87}$	0.99	0.161917
900	$2.26^{2.65}_{-2.26}$	0.80	0.211634
950	$3.14^{3.98}_{-3.14}$	0.77	0.220715
1000	$6.24^{6.17}_{-6.16}$	1.01	0.157156
		2017	
300	$0.10^{0.12}_{-0.10}$	0.67	0.252949
350	$0.13^{0.10}_{-0.12}$	1.10	0.13555
400	$0.16^{0.13}_{-0.14}$	1.11	0.132814
450	$0.15^{0.17}_{-0.15}$	0.80	0.211727
500	$0.16^{0.22}_{-0.16}$	0.69	0.244811
550	$0.23^{0.31}_{-0.23}$	0.68	0.248567
600	$0.20^{0.44}_{-0.20}$	0.41	0.340254
650	$0.17^{0.65}_{-0.17}$	0.24	0.405109
700	$0.00^{0.94}_{-0.00}$	0.00	0.5
750	$0.82^{1.28}_{-0.82}$	0.58	0.280483
800	$0.66^{2.04}_{-0.66}$	0.29	0.384645
850	$0.60^{3.46}_{-0.60}$	0.16	0.437359
900	$2.04^{5.01}_{-2.04}$	0.39	0.347284
950	$2.57^{7.35}_{-2.57}$	0.33	0.370211
1000	$2.59^{9.87}_{-2.59}$	0.24	0.405917

Table 5.11: Best-fit signal strength and corresponding observed significance for 2018pre and 2018post data with different top squark mass points of the RPV SUSY model.

Mass	Best fit signal strength	Observed significance	p-value
		2018pre	
300	$0.27^{0.14}_{-0.16}$	1.42	0.0781714
350	$0.24^{0.15}_{-0.19}$	1.19	0.117299
400	$0.32^{0.15}_{-0.19}$	1.61	0.0535427
450	$0.42^{0.19}_{-0.24}$	1.58	0.0565996
500	$0.55^{0.25}_{-0.33}$	1.53	0.0632478
550	$0.75^{0.38}_{-0.52}$	1.38	0.0837839
600	$1.13^{0.52}_{-0.68}$	1.50	0.0666822
650	$1.55^{0.75}_{-1.05}$	1.38	0.0841509
700	$2.35^{1.10}_{-1.47}$	1.45	0.0734226
750	$3.43^{1.64}_{-2.14}$	1.46	0.0723764
800	$4.56^{2.51}_{-3.15}$	1.35	0.0877883
850	$7.80^{3.31}_{-3.99}$	1.68	0.0463929
900	$11.46^{5.08}_{-6.82}$	1.48	0.0697702
950	$17.80^{2.20}_{-9.69}$	1.61	0.0538171
1000	$20.00^{0.00}_{-11.34}$	1.48	0.0693613
		2018post	
300	$0.11^{0.12}_{-0.11}$	0.81	0.208396
350	$0.09^{0.11}_{-0.09}$	0.73	0.231885
400	$0.10^{0.13}_{-0.10}$	0.73	0.232014
450	$0.12^{0.17}_{-0.12}$	0.62	0.266474
500	$0.12^{0.23}_{-0.12}$	0.50	0.308055
550	$0.15^{0.32}_{-0.15}$	0.44	0.330671
600	$0.20^{0.49}_{-0.20}$	0.37	0.353904
650	$0.27^{0.66}_{-0.27}$	0.38	0.352319
700	$0.34^{1.01}_{-0.34}$	0.31	0.377204
750	$0.31^{1.60}_{-0.31}$	0.18	0.428874
800	$1.25^{2.11}_{-1.25}$	0.56	0.289366
850	$0.56^{3.29}_{-0.56}$	0.16	0.434948
900	$1.46^{4.80}_{-1.46}$	0.28	0.390584
950	$2.38^{8.27}_{-2.38}$	0.26	0.395665
1000	$4.93^{11.11}_{-4.93}$	0.42	0.338894

Table 5.12: Best-fit signal strength and corresponding observed significance for 2016 and 2017 data with different top squark mass points of the stealth SUSY model.

Mass	Best fit signal strength	Observed significance	p-value
		2016	
300	$0.00^{0.27}_{-0.00}$	0.00	0.5
350	$0.22^{0.15}_{-0.18}$	1.17	0.120041
400	$0.18^{0.12}_{-0.13}$	1.31	0.0955787
450	$0.18^{0.11}_{-0.13}$	1.42	0.0771406
500	$0.18^{0.13}_{-0.13}$	1.30	0.0968996
550	$0.16^{0.15}_{-0.16}$	1.01	0.155218
600	$0.19^{0.16}_{-0.18}$	1.09	0.138129
650	$0.16^{0.21}_{-0.16}$	0.71	0.239801
700	$0.17^{0.27}_{-0.17}$	0.58	0.279389
750	$0.16^{0.32}_{-0.16}$	0.47	0.318007
800	$0.17^{0.45}_{-0.17}$	0.36	0.360478
850	$0.12^{0.54}_{-0.12}$	0.22	0.414395
900	$0.19^{0.71}_{-0.19}$	0.25	0.399784
950	$0.01^{0.93}_{-0.01}$	0.02	0.49383
1000	$0.02^{1.19}_{-0.02}$	0.02	0.49299
		2017	
300	$0.30^{0.17}_{-0.20}$	1.43	0.076022
350	$0.15^{0.13}_{-0.14}$	1.04	0.149584
400	$0.09^{0.11}_{-0.09}$	0.76	0.224614
450	$0.09^{0.11}_{-0.09}$	0.70	0.240987
500	$0.07^{0.13}_{-0.07}$	0.51	0.30426
550	$0.04^{0.17}_{-0.04}$	0.25	0.402205
600	$0.01^{0.19}_{-0.01}$	0.01	0.494885
650	$0.03^{0.26}_{-0.03}$	0.10	0.458547
700	$0.02^{0.31}_{-0.02}$	0.06	0.477865
750	$0.01^{0.42}_{-0.01}$	0.00	0.5
800	$0.00^{0.44}_{-0.00}$	0.00	0.5
850	$0.00^{0.66}_{-0.00}$	0.00	0.5
900	$0.00^{0.89}_{-0.00}$	0.00	0.5
950	$0.10^{1.31}_{-0.10}$	0.08	0.469325
1000	$0.00^{1.49}_{-0.00}$	0.00	0.5

Table 5.13: Best-fit signal strength and corresponding observed significance for 2018pre and 2018post data with different top squark mass points of the stealth SUSY model.

Mass	Best fit signal strength	Observed significance	p-value
		2018pre	
300	$0.48^{0.16}_{-0.28}$	1.22	0.111325
350	$0.35^{0.12}_{-0.15}$	1.84	0.0327516
400	$0.28^{0.14}_{-0.17}$	1.51	0.0658724
450	$0.29^{0.13}_{-0.15}$	1.67	0.0474788
500	$0.31^{0.15}_{-0.20}$	1.43	0.0759924
550	$0.39^{0.18}_{-0.22}$	1.61	0.0531983
600	$0.48^{0.21}_{-0.23}$	1.83	0.0338282
650	$0.62^{0.29}_{-0.32}$	1.72	0.0431011
700	$0.81^{0.39}_{-0.43}$	1.69	0.0457666
750	$0.98^{0.47}_{-0.50}$	1.75	0.039826
800	$1.36^{0.65}_{-0.71}$	1.72	0.0428183
850	$1.76^{0.90}_{-0.95}$	1.71	0.0439532
900	$2.32^{1.20}_{-1.26}$	1.68	0.0460304
950	$3.12^{1.60}_{-1.68}$	1.70	0.0445289
1000	$4.36^{2.31}_{-2.37}$	1.71	0.0439408
		2018post	
300	$0.18^{0.21}_{-0.18}$	0.78	0.218975
350	$0.11^{0.13}_{-0.11}$	0.80	0.212731
400	$0.07^{0.11}_{-0.07}$	0.54	0.293415
450	$0.06^{0.12}_{-0.06}$	0.44	0.329579
500	$0.06^{0.14}_{-0.06}$	0.44	0.331469
550	$0.07^{0.17}_{-0.07}$	0.41	0.339326
600	$0.07^{0.20}_{-0.07}$	0.32	0.373177
650	$0.10^{0.27}_{-0.10}$	0.35	0.364211
700	$0.11^{0.36}_{-0.11}$	0.29	0.385345
750	$0.10^{0.46}_{-0.10}$	0.20	0.422023
800	$0.15^{0.66}_{-0.15}$	0.22	0.414241
850	$0.06^{0.89}_{-0.06}$	0.05	0.479287
900	$0.00^{1.20}_{-0.00}$	0.00	0.5
950	$0.17^{1.63}_{-0.17}$	0.10	0.461023
1000	$0.07^{2.35}_{-0.07}$	0.02	0.493415

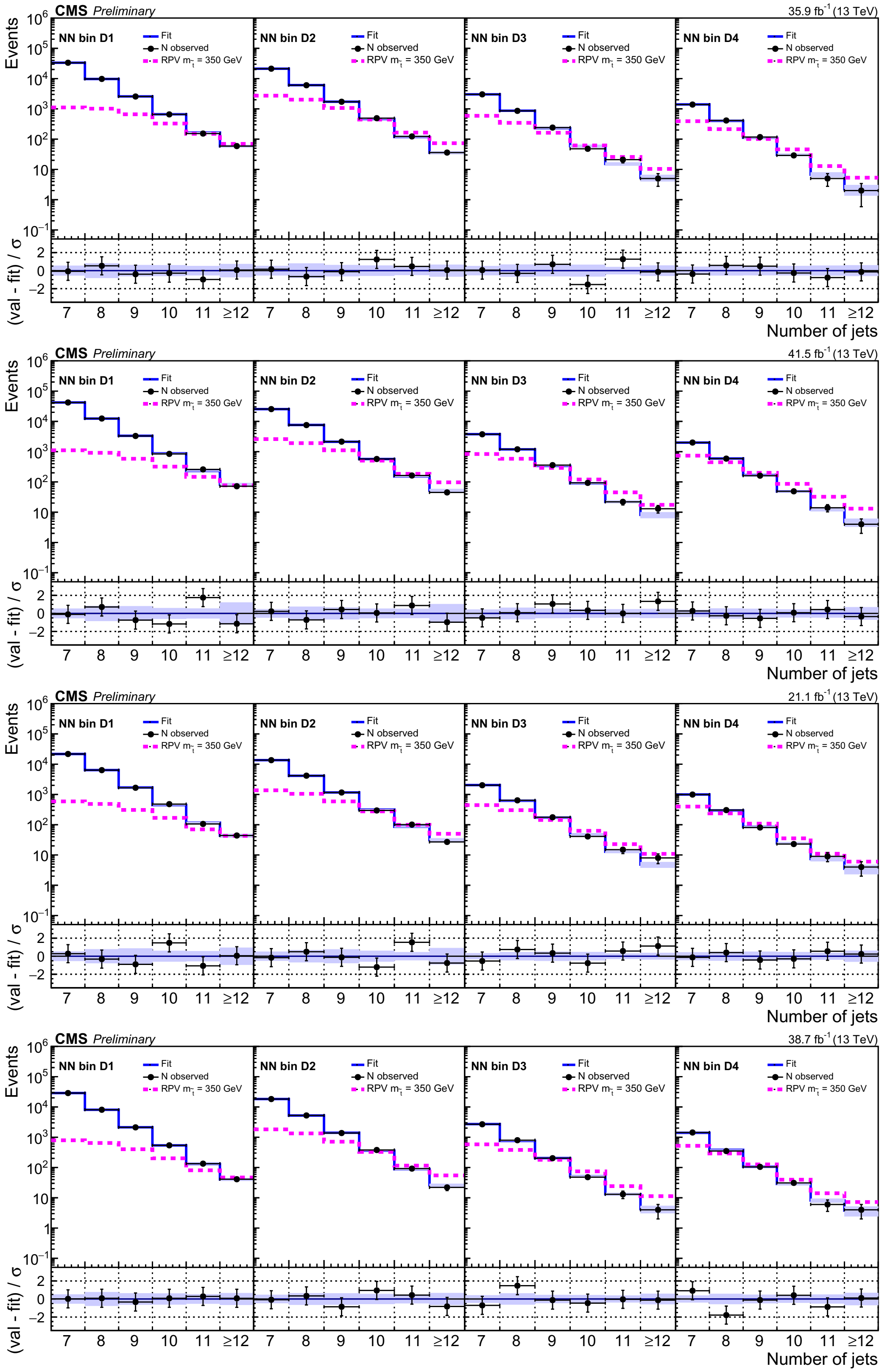


Figure 5.53: Post-fit predictions in logarithmic scale, along with the pulls, from the background-only fit to the combined full Run 2 data. The RPV SUSY models with $m_{\tilde{t}} = 350$ GeV is overlaid.

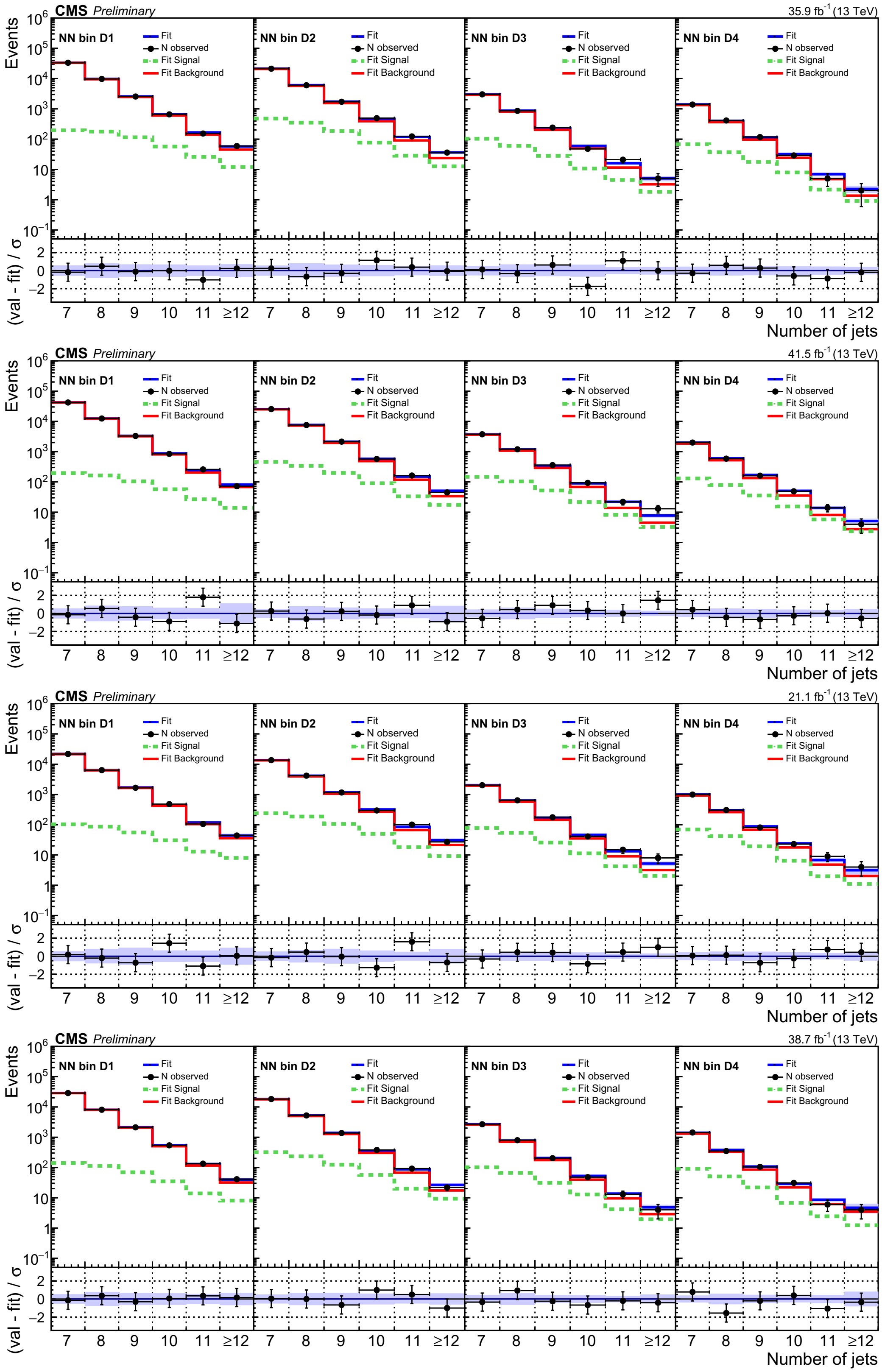


Figure 5.54: Post-fit predictions in logarithmic scale, along with the pulls, from the signal + background fit to the combined full Run 2 data using the RPV SUSY model with top squark mass of 350 GeV.

Table 5.14: Fit parameters for the $t\bar{t}$ shape in the fit to the data for the combined fit to 2016, 2017, 2018pre, and 2018post data.

Parameter	Background-only fit	Signal + background fit
	2016	
a_0	0.28938 ± 0.00578	0.27995 ± 0.00681
a_1	0.26195 ± 0.00401	0.24380 ± 0.00896
d	26.995 ± 58.2	17.596 ± 36.2
	2017	
a_0	0.30145 ± 0.00368	0.29333 ± 0.00480
a_1	0.26524 ± 0.00416	0.25086 ± 0.00724
d	118.10 ± 163	60.646 ± 82.5
	2018pre	
a_0	0.30338 ± 0.00418	0.29653 ± 0.00488
a_1	0.26581 ± 0.00560	0.25317 ± 0.00796
d	285.31 ± 297	259.95 ± 291
	2018post	
a_0	0.28405 ± 0.00364	0.27822 ± 0.00420
a_1	0.25502 ± 0.00412	0.24098 ± 0.00711
d	15.726 ± 45.1	17.291 ± 38.3

data. Figures 5.55 and 5.56 show the limits for the RPV and stealth SUSY models obtained from fits to data from 2016, 2017, 2018pre, 2018post, and all four data eras combined. It is important to remember that the expected limit depends on the observed data, which is especially relevant for our analysis because our $t\bar{t}$ signal shape is also extracted from the data in addition to the nuisance parameters.

The observed limit is just outside the 2σ band above the expected limit for the lower masses. This indicates that the fit prefers a non-zero amount of signal. However, the amount of signal used during the fit is not too significant. Using the combination of 2016, 2017, and 2018 data, we can exclude top squark masses up to 700 GeV for the RPV SUSY model, and masses up to 900 GeV for the stealth SUSY model with the SYY portal.

Table 5.15: Best-fit signal strength and observed significances as a function of top squark mass for the combination of 2016, 2017, 2018pre, and 2018post data.

Mass	Best-fit signal strength	significance	p-value
	RPV SUSY model		
300	$0.20^{0.07}_{-0.10}$	2.18	0.0144863
350	$0.18^{0.06}_{-0.07}$	2.39	0.00848925
400	$0.21^{0.07}_{-0.07}$	2.78	0.00272507
450	$0.24^{0.09}_{-0.09}$	2.41	0.00795862
500	$0.27^{0.12}_{-0.12}$	2.11	0.0172651
550	$0.28^{0.15}_{-0.16}$	1.68	0.0469655
600	$0.37^{0.21}_{-0.22}$	1.61	0.0534263
650	$0.37^{0.29}_{-0.30}$	1.21	0.113719
700	$0.46^{0.43}_{-0.45}$	1.02	0.154306
750	$0.80^{0.59}_{-0.62}$	1.28	0.100945
800	$0.97^{0.85}_{-0.88}$	1.10	0.134677
850	$1.46^{1.37}_{-1.42}$	1.02	0.153237
900	$1.90^{2.00}_{-1.90}$	0.91	0.182308
950	$2.78^{3.02}_{-2.78}$	0.90	0.184776
1000	$4.75^{4.38}_{-4.47}$	1.06	0.145448
	Stealth SUSY model		
300	$0.32^{0.12}_{-0.15}$	1.77	0.0382575
350	$0.21^{0.07}_{-0.08}$	2.48	0.0065171
400	$0.14^{0.06}_{-0.07}$	1.85	0.0319404
450	$0.14^{0.06}_{-0.07}$	2.04	0.0207175
500	$0.13^{0.07}_{-0.07}$	1.75	0.0399094
550	$0.13^{0.08}_{-0.09}$	1.49	0.0684942
600	$0.13^{0.09}_{-0.10}$	1.32	0.0932231
650	$0.13^{0.12}_{-0.13}$	1.03	0.150597
700	$0.16^{0.16}_{-0.16}$	0.97	0.164849
750	$0.16^{0.19}_{-0.16}$	0.83	0.203799
800	$0.18^{0.26}_{-0.18}$	0.67	0.252377
850	$0.19^{0.34}_{-0.19}$	0.55	0.290372
900	$0.27^{0.45}_{-0.27}$	0.59	0.278353
950	$0.33^{0.58}_{-0.33}$	0.56	0.2894
1000	$0.32^{0.79}_{-0.32}$	0.40	0.345508

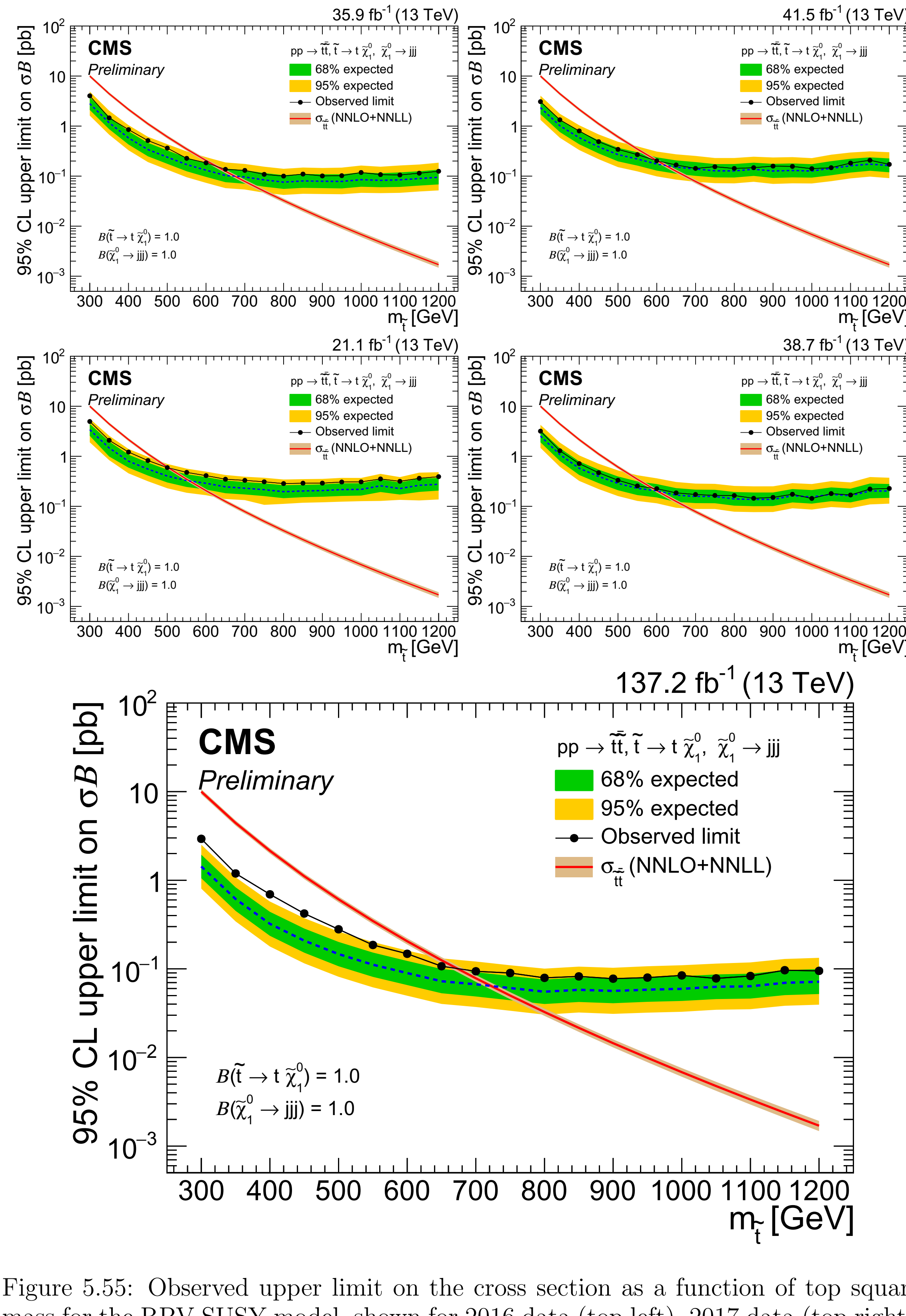


Figure 5.55: Observed upper limit on the cross section as a function of top squark mass for the RPV SUSY model, shown for 2016 data (top left), 2017 data (top right), 2018pre data (middle left), 2018post data (middle right), and the combination of all four eras (bottom). The green and yellow bands show the 1σ and 2σ uncertainty bands. The red line shows the top squark cross section as a function of its mass.

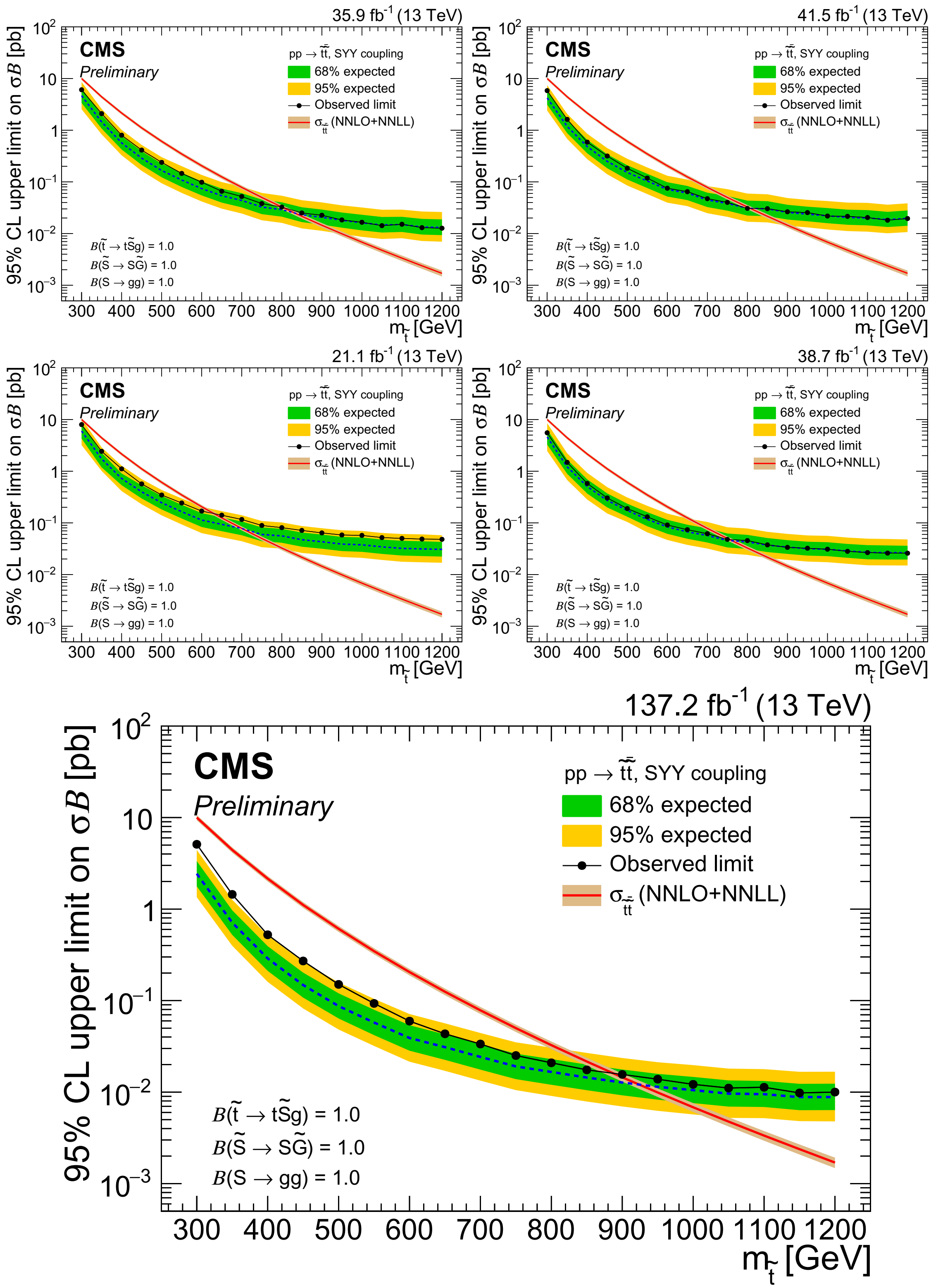


Figure 5.56: Observed upper limit on the cross section as a function of top squark mass for the stealth SUSY model with the SYY portal, shown for 2016 data (top left), 2017 data (top right), 2018pre data (middle left), 2018post data (middle right), and the combination of all four eras (bottom). The green and yellow bands show the 1σ and 2σ uncertainty bands. The red line shows the top squark cross section as a function of its mass.

CHAPTER SIX

Summary

The goal of the work detailed in this thesis is to look for any signs of new physical phenomena that had not been previously observed and that are not currently predicted by the standard model of particle physics (SM). This thesis presents a search for new physics featuring a final state with two top quarks, six additional light-flavor jets, and no requirement on the presence of missing transverse momentum. The search is performed in the context of supersymmetry (SUSY), a well-motivated extension of the SM, using data collected at CERN's Large Hadron Collider by the Compact Muon Solenoid (CMS) detector. Events are selected from data corresponding to an integrated luminosity of 137.2 fb^{-1} collected in proton-proton collisions at a center-of-mass energy of 13 TeV during 2016, 2017, and 2018.

The event selection requires exactly one electron or muon and at least seven jets, of which at least one should be b-tagged, and a loose leptonic top quark tag. The dominant $\mathrm{t\bar{t}}$ background is predicted from data using a simultaneous fit of the jet multiplicity distribution across four bins of a neural network score. The results are interpreted in terms of top squark pair production in the context of R-parity-violating and stealth supersymmetry models. Top squark masses up to 700 GeV can be excluded at 95% confidence level for the signal model in which the top squark decays to a top quark and the lightest neutralino, assuming the R-parity violating decay of the neutralino to three light-flavor jets through a λ'' coupling, as seen in Fig 5.55. Top squark masses up to 900 GeV can be excluded for the stealth supersymmetry model

in which the top squark decays to a top quark, a gluon, and a singlino, followed by the decay of the singlino into a singlet and a gravitino, and the decay of the singlet into two gluons, as seen in Fig. 5.56. This search is the first of its kind at the LHC, and provides further information about the viability of light top squarks.

One implication of these results is the following: assuming these models reflect reality, the mass of the top squark must be at least four times that of the SM's top quark. As discussed in Chapter Two, a major motivation for SUSY is that it offers a natural solution to the Higgs mass hierarchy problem, but this is only true if the mass of the top quark and the mass of the top squark are not too far apart. If their mass difference becomes large, the correction to the Higgs mass by radiative loop diagrams is dominated by logarithmically divergent terms, which results in the breakdown of the naturalness argument. This alone is not enough to rule out SUSY completely, even with these divergent terms in the Higgs mass calculation. However, it could move SUSY down the list of well-motivated theories and promote others.

It is also worth noting that the amount of signal found in the signal + background fits is nonzero in this search; however, the amount of signal preferred by the fit is not significant enough to claim a deviation from the background-only hypothesis. An optimistic physicist would be open to the possibility that a signal is present in the data, but is not produced at the expected rate. For example, the best-fit signal strength found for the RPV 400 GeV signal in the combined data set is 0.21 ± 0.07 at a significance of 2.8σ. Interestingly enough, this fit is very close to the boundary of declaring this result as evidence for a SUSY-like signal. In practice it has been accepted, arbitrarily, that a significance of 3σ is the boundary of a result being declared "evidence," and a significance of 5σ is used for a discovery. If we are to assume this result as evidence for this signal, the best-fit signal strength would indicate that the branching fraction for top squark decay is not 100% to this particular decay channel, but instead about 21%. As this analysis assumes a branching fraction of 100% for simplicity, the other 79% of top squarks could be decaying through some other modes that fall outside our event selection. However, another interpretation of this fit result is that there is no signal, and the fact that the fit prefers to include a nonzero amount could be due either to statistical fluctuations in the data or a missing or underestimated source of systematic uncertainty. For example, we rely on simulated $t\bar{t}$ events to estimate our systematic uncertainty, and if there is an unknown issue in modeling events with a large number of jets, our analysis would be uniquely impacted. The possibility of an unknown source of uncertainty in the simulated events is what motivated us to derive data-driven systematic uncertainties from the QCD control region. On the other hand, we chose to be as conservative as reasonably possible when deriving systematic uncertainties, which could imply that the uncertainties are somewhat overestimated. This could mean that the significance values reported earlier are underestimated.

Looking ahead, there are plans to repeat this search targeting a more inclusive decay mode of $t\bar{t}$ in signal events. The analysis reported in this thesis is based on events with exactly one lepton; however, $t\bar{t}$ can also produce events that are all hadronic (no leptons) and dileptonic (two leptons). With data included from these other decay modes of $t\bar{t}$, we could increase the number of events considered, which could push the observed significance of our signals to the evidence mark. Other plans for the future include interpreting our results using other well-motivated extensions of the SM. One interesting example is a search for vector-like quarks, which are predicted by several classes of theoretical models beyond the SM [49]. Vector-like quarks are non-chiral fermions in which their left- and right-handed components transform the same way under $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ gauge symmetry. The final-state signature for such particles can result in $\mathrm{t\bar{t}} + \mathrm{jets}$, which is similar to the SUSY models considered in this thesis. Pair production for these new vector-like quarks, for some models, would then have a predicted cross section that is on the same order of magnitude as the cross section for squark pair production.

With no obvious evidence for SUSY based on these results, or any other search, the future of the field of particle physics still remains in flux. However, there is still a lot of phase space that has not been fully explored. The motivations for some physics models may be changing, but the hunt for evidence of physics beyond the standard model at the LHC is far from over.

APPENDIX

APPENDIX

Additional Information about Systematic Uncertainties

A.1 $\mathrm{t\bar{t}}$ Parton Shower Variation Shape Systematic Uncertainties

Following the method described in Section 5.6.2, systematic uncertainties are derived by varying the parton shower settings in the MC data. Three different variations are used: "erdOn" which allows for color reconnection in the parton shower, "hdamp (Up/Down)" which changes the ME-PS matching scale, and "underlying event (Up/Down)" which varies internal PYTHIA parameters that affect the simulation of underlying event physics. Results for erdOn are shown in Fig. A.1, hdamp (Up/Down) in Fig. A.2, and underlyingEvt (Up/Down) in Fig. A.3.

A.2 H_{T} Shape Systematic Uncertainties

In addition to the systematic uncertainty derived for the H_{T} scale factor that is used for both $\mathrm{t\bar{t}}$ and the other backgrounds, there are also two more H_{T}-related uncertainties applied to the $\mathrm{t\bar{t}} + \mathrm{jets}$ background. The motivation for including these two uncertainties comes from the fact that both of the important variables in this analysis, N_{j} and the NN score, are correlated with H_{T}. Thus, any mismodeling in this variable, even if the mismodeling may be small, could have an impact on the final result. The approach to deriving these two systematic uncertainties is to create two alternative H_{T} scale factors (SF_1 and SF_2) and then use them on the $\mathrm{t\bar{t}}$ MC data following the procedure described in Section 5.6.2.

The first of these two uncertainties comes from the fact that the number of simulated events at high H_{T} diminishes, so the fit that is used to extrapolate to high

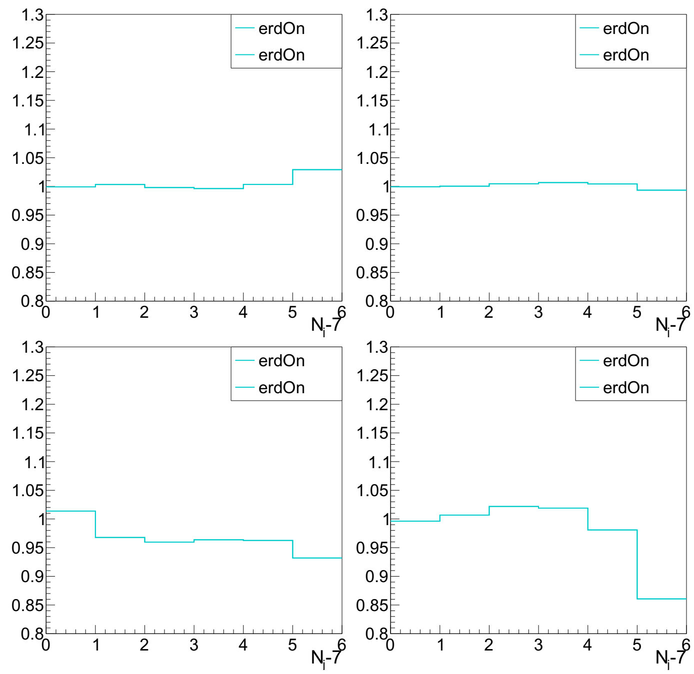


Figure A.1: Ratio $R = R_{\text{syst}}/R_{\text{nom}}$ showing the N_{j} shape differences introduced by allowing for color reconnection in the parton shower for the four NN bins: D_1 (top left), D_2 (top right), D_3 (bottom left), and D_4 (bottom right). The x axis shows N_{j} shifted down such that $N_{\text{j}} = 7$ is the first bin, here noted by 0.

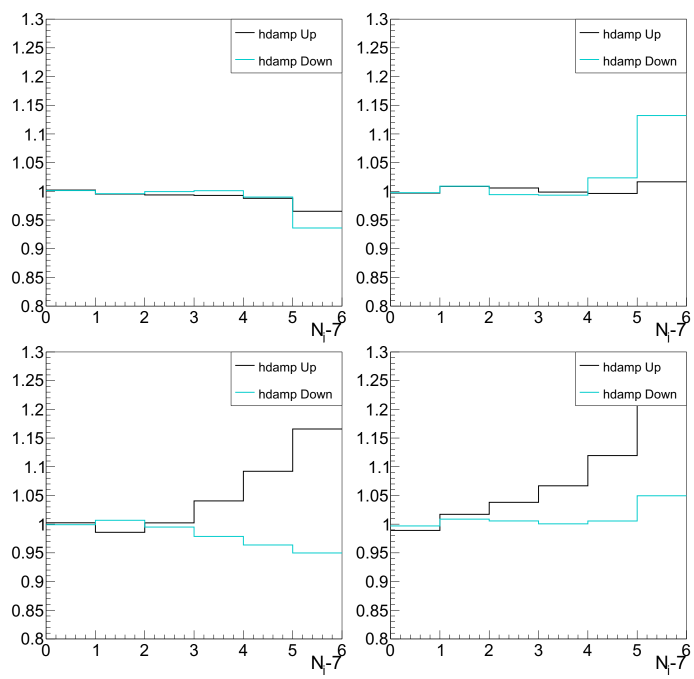


Figure A.2: Ratio $R = R_{\text{syst}}/R_{\text{nom}}$ showing the N_{j} shape differences introduced by varying the ME-PS scale for the four NN bins: D_1 (top left), D_2 (top right), D_3 (bottom left), and D_4 (bottom right). The x axis shows N_{j} shifted down such that $N_{\text{j}} = 7$ is the first bin, here noted by 0.

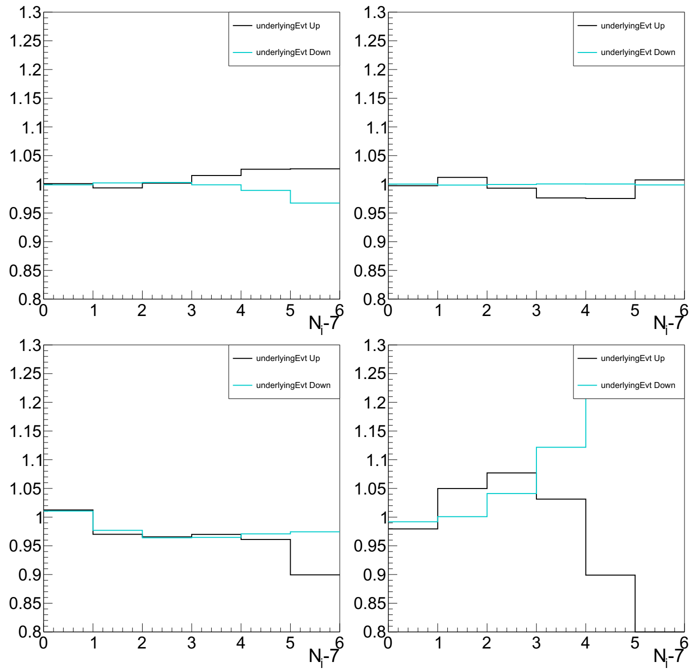


Figure A.3: Ratio $R = R_{\mathrm{syst}}/R_{\mathrm{nom}}$, showing the N_{j} shape differences introduced by varying the underlying event model parameters for the four NN bins: D_1 (top left), D_2 (top right), D_3 (bottom left), and D_4 (bottom right). The x axis shows N_{j} shifted down such that $N_{\mathrm{j}} = 7$ is the first bin, here noted by 0.

H_{T} may be inaccurate. By looking at some sample fits, like the ones shown in Fig. A.4 for the H_{T} distribution in the $N_{\mathrm{j}} = 7$ bin, the statistical errors for the ratio above 2000 GeV are large in comparison to the statistical errors in bins less than 2000 GeV.

Having observed this trend, the first alternative H_{T} scale factor (SF_1) is derived such that all events with H_{T} greater than 2000 GeV are weighted with the scale factor value used at 2000 GeV, as shown in Fig. A.5. Using this alternative scale factor, the H_{T} and the N_{j} distributions for the $\mathrm{t\bar{t}}$ simulation are compared with the respective distributions using the nominal H_{T} scale factor. The ratios of the N_{j} distributions with the alternative scale factor over the N_{j} distributions with the nominal scale factor, divided up by the different NN bins per year, are added as an uncertainty for the $\mathrm{t\bar{t}} + \mathrm{jets}$ background shape. These ratios are shown in Fig. A.6 for 2016.

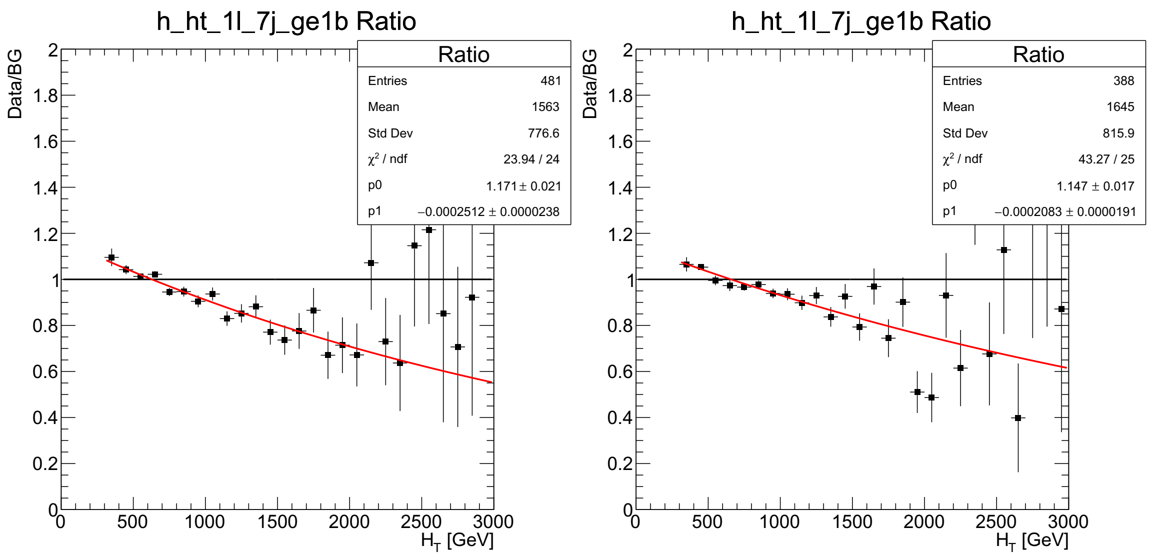


Figure A.4: The ratio of data to simulation and the corresponding fit used to derive the H_{T} scale factor in the $N_{\mathrm{j}} = 7$ bin for 2016 (left) and 2017 (right).

The second H_{T} systematic uncertainty is motivated by the fact that the extrapolation to higher N_{j} bin from the $N_{\mathrm{j}} = 5$, 6, and 7 bins may be affected by statistical fluctuations in these lower N_{j} bins. To derive an uncertainty that can account for this, a second alternative H_{T} scale factor using only the values derived in the $N_{\mathrm{j}} = 7$ bin was applied to the $\mathrm{t\bar{t}}$ simulation. The second alternative H_{T} scale factor (SF_2) for both years is shown in Fig. A.7, and the respective ratios of the N_{j} distributions for the four NN bins for 2016 are shown in Fig. A.8.

A.3 QCD Control Region Derived Systematic Uncertainties

So far, all of the systematic uncertainties have been derived directly from the simulation and variations on the simulation. In an attempt to have a data-driven uncertainty that addresses differences in the N_{j}-NN correlation in both collision data and MC data, a new shape systematic uncertainty from the QCD control region is evaluated.

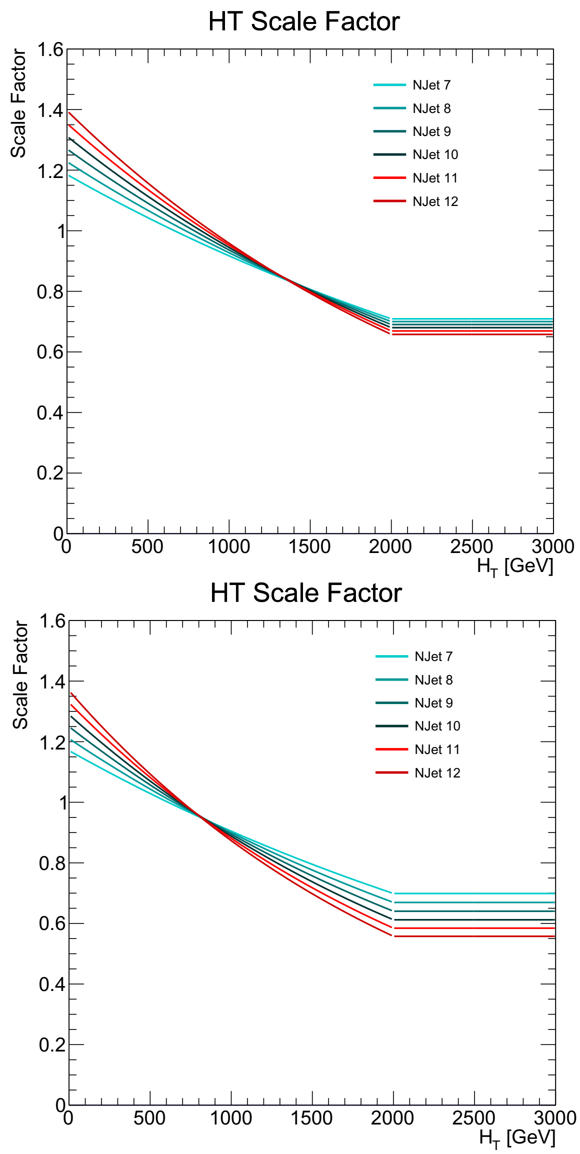


Figure A.5: The shape of alternative scale factor (SF_1) with the H_T scale factor constant after 2000 GeV for 2016 (top left), 2017 (top right), 2018pre (bottom left), and 2018post (bottom right).

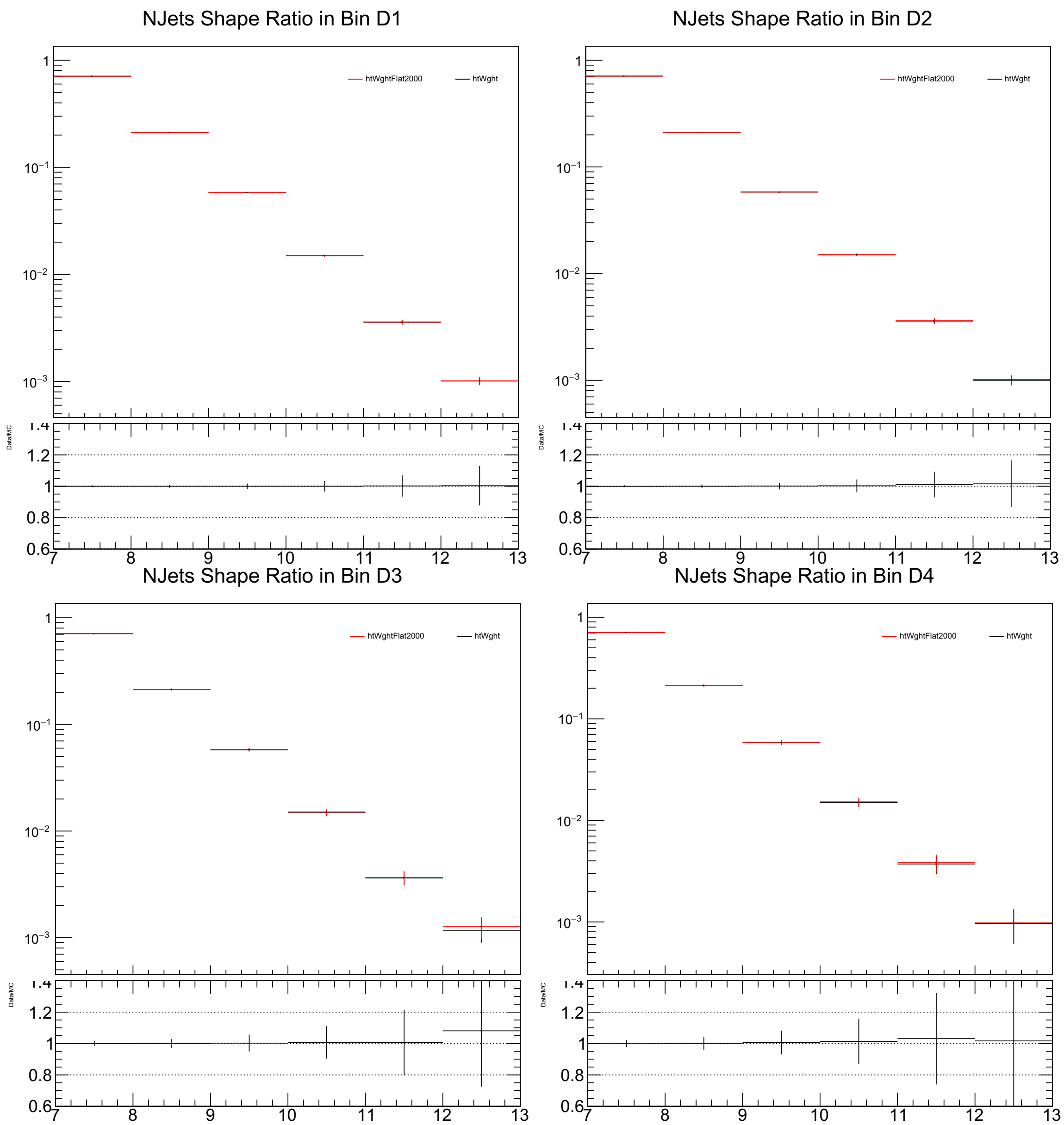


Figure A.6: The ratio of the N_{j} distribution, weighted by alternative H_{T} SF_1 that is constant beyond 2000 GeV, to the N_{j} distribution weighted with the nominal H_{T} scale factor for 2016 in four NN bins: D1 (top left), D2 (top right), D3 (bottom left), and D4 (bottom right).

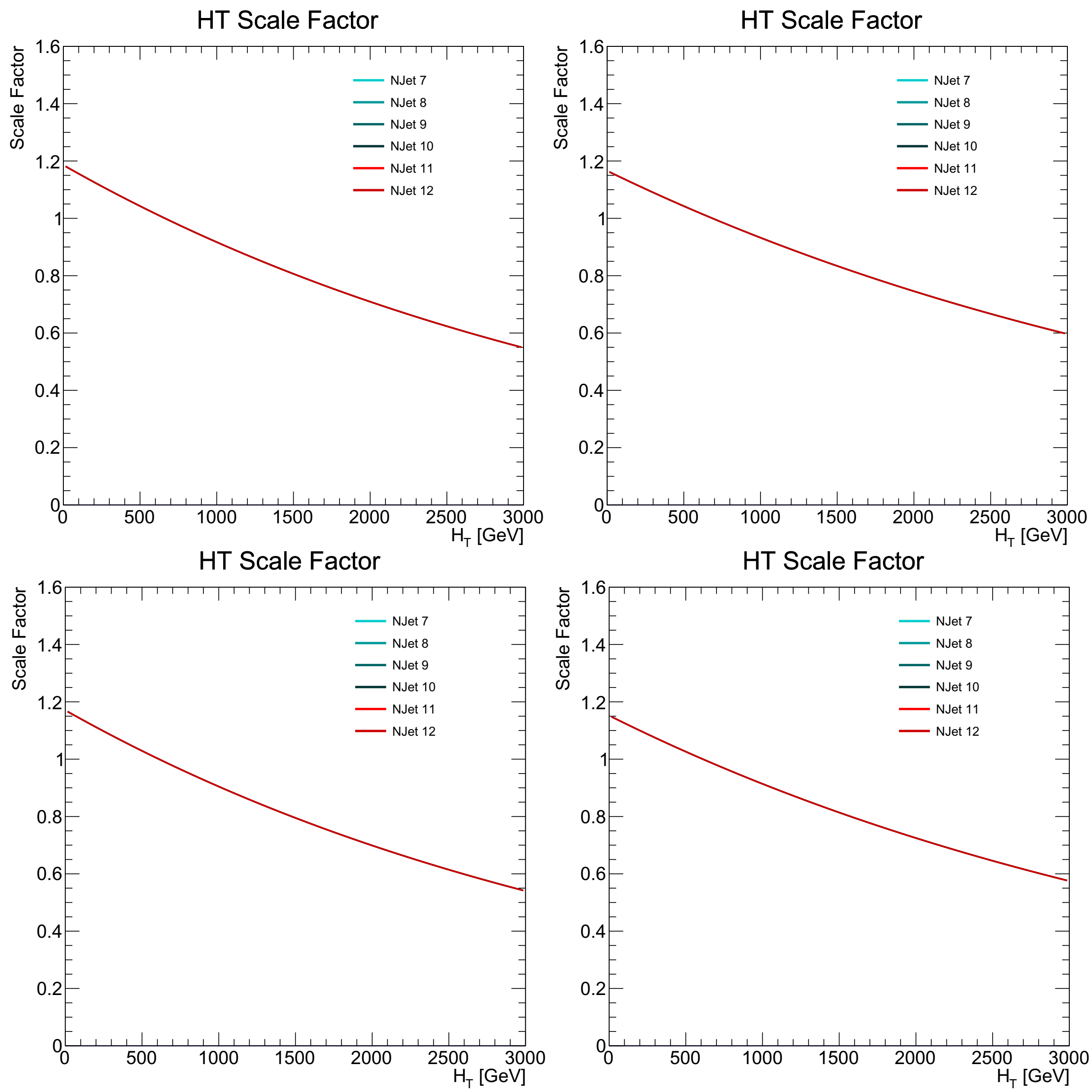


Figure A.7: The shape of alternative scale factor (SF_2) using only values derived for N_j = 7 for 2016 (top left), 2017 (top right), 2018pre (bottom left), and 2018post (bottom right).

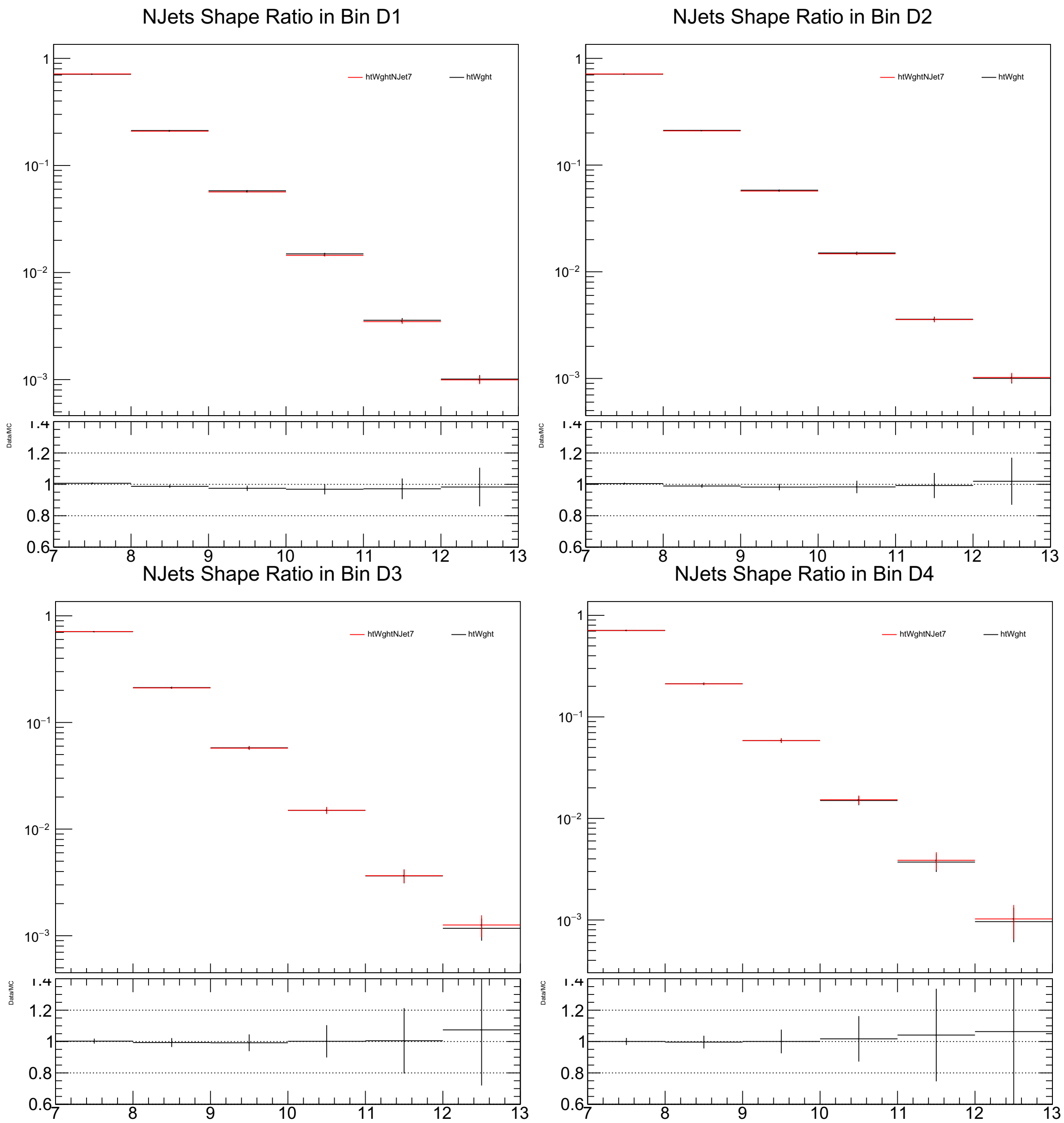


Figure A.8: The ratio of the N_j distribution, weighted with alternative scale factor (SF_2) using the $N_\mathrm{j} = 7$ value in higher N_j bins, to the N_j distribution weighted with the nominal H_T scale factor for 2016 in four NN bins: D1 (top left), D2 (top right), D3 (bottom left), and D4 (bottom right).

The first step is to determine whether the QCD MC data in the control region has a similar N_{j}-NN correlation to that of the data in the control region. The data vs. MC data agreement in the control region of the NN shape per N_{j} bin is shown in Fig. A.9. Although the NN shapes between data and MC data appear to be different in higher N_{j} bins, when taking into account the size of the error bars that result from some of the highly weighted QCD samples, there is agreement between the two.

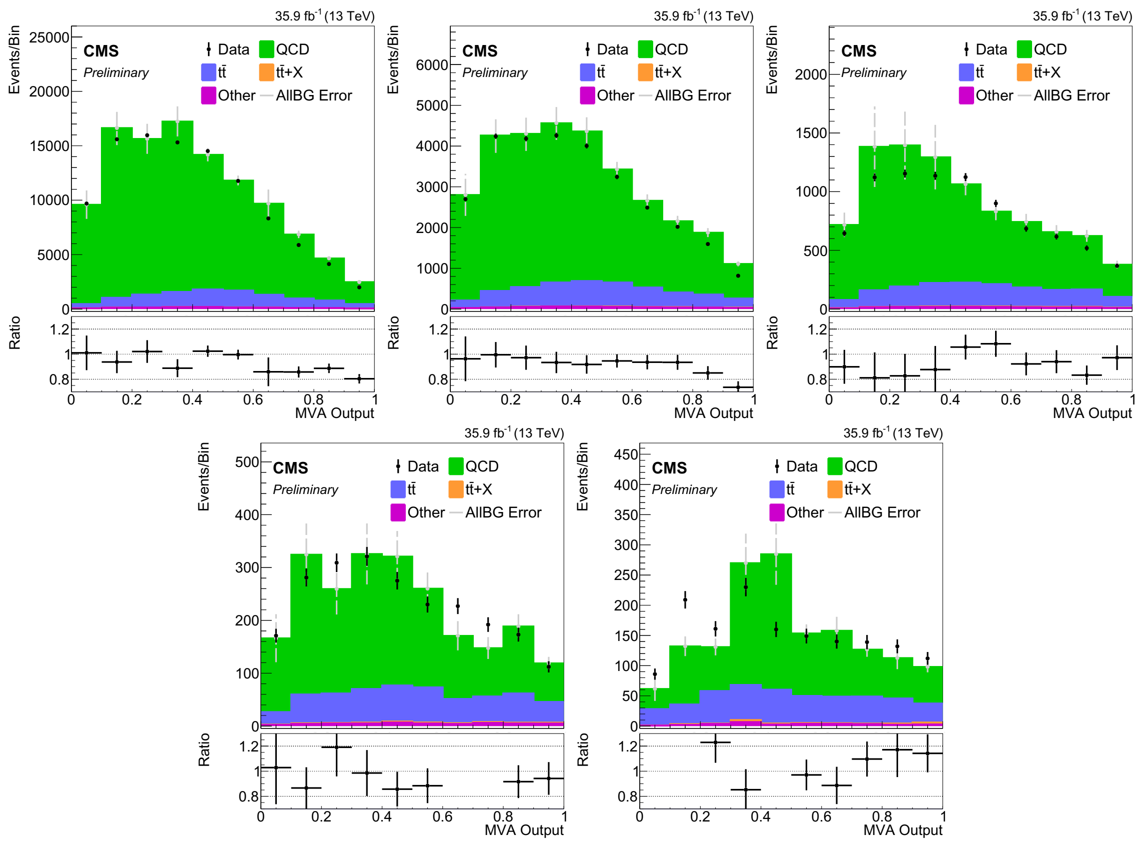


Figure A.9: Data vs. MC data agreement in the NN shape distribution for (top) N_{j} bins 7–9, and (bottom) N_{j} bins 10 and 11.

Next, the NN distributions for each N_{j} bin in the control region are shown for both QCD MC data and QCD data in Figs. A.10 and A.11, respectively. From Figs. A.10 and A.11, the NN shape per N_{j} seems to be slightly different, particularly at high N_j. For example, the 0.4 to 0.6 bin in the $N_\mathrm{j} = 11$ NN shape for the QCD MC data has the most events, whereas the 0.2 to 0.4 bin in the $N_\mathrm{j} = 11$ NN shape for data has the most events. However, when looking at NN shape relative to the total shape, shown in red, the trends in the ratio plots below are the same.

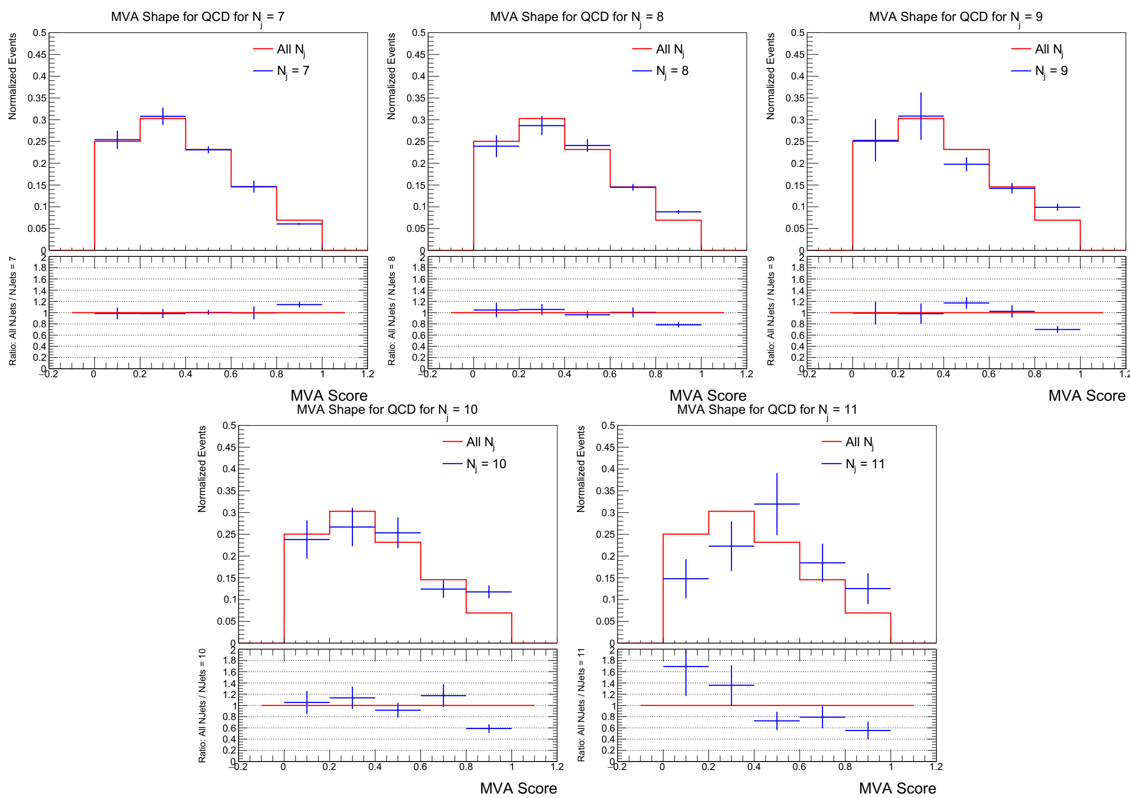


Figure A.10: The NN shape distribution for (top) N_j bins 7–9 and (bottom) N_j bins 10 and 11 relative to the overall NN shape distribution for all N_j for the QCD multijet MC data in the control region.

This can be generalized to the $\mathrm{t\bar{t}} + \mathrm{jets}$ MC data in the signal region as well, shown in Fig. A.12. It is clear that the overall shape for $\mathrm{t\bar{t}} + \mathrm{jets}$ MC data in the signal region is different from that in the QCD MC data in the control region, but the ratio of the total NN shape to the per N_j NN shape remains the same. This similarity in shapes point to a more general correlation between N_j and the NN score

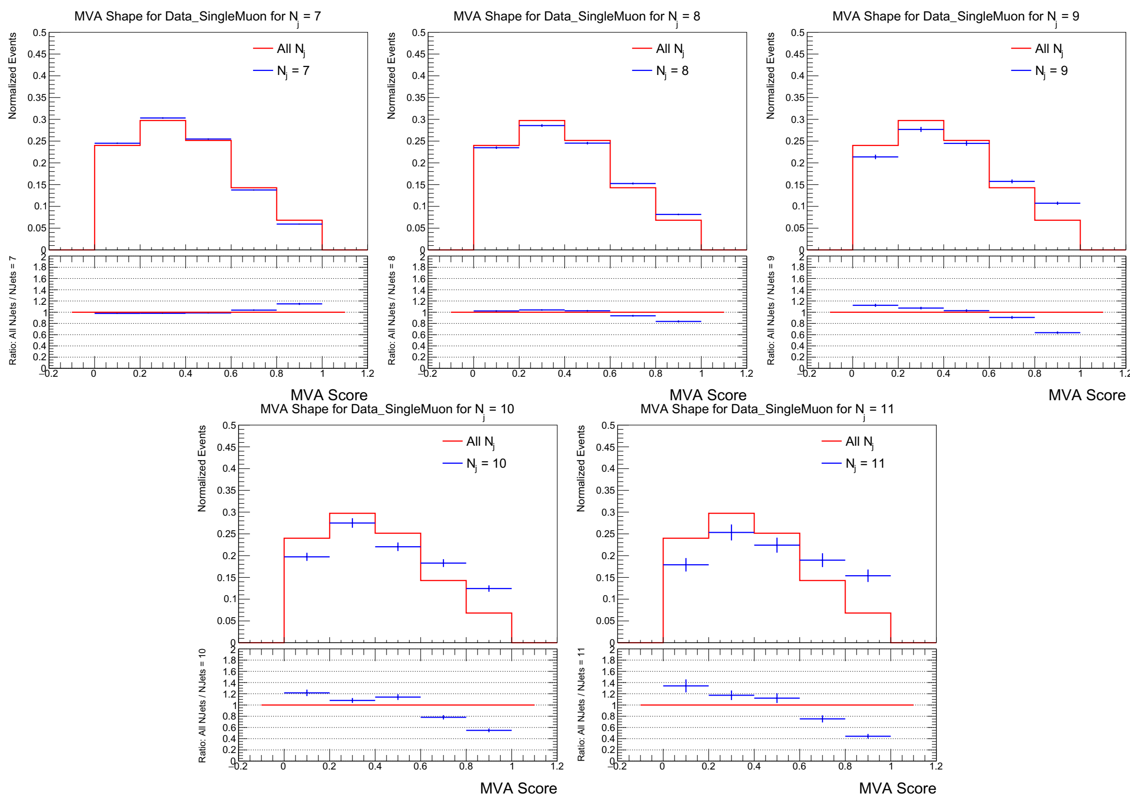


Figure A.11: The NN shape distribution for (top) N_j bins 7–9 and (bottom) N_j bins 10 and 11 relative to the overall NN shape distribution for all N_j for the collision data in the QCD control region.

that is present in both the $\mathrm{t\bar{t}}+\mathrm{jets}$ MC data in the signal region and the data in the QCD enriched control region. To further emphasize this point, the ratios of the full NN shape for all N_j over the NN shape per N_j for $\mathrm{t\bar{t}}$ MC data in the signal region are overlaid with those for the data in the QCD enriched control region in Fig. A.13.

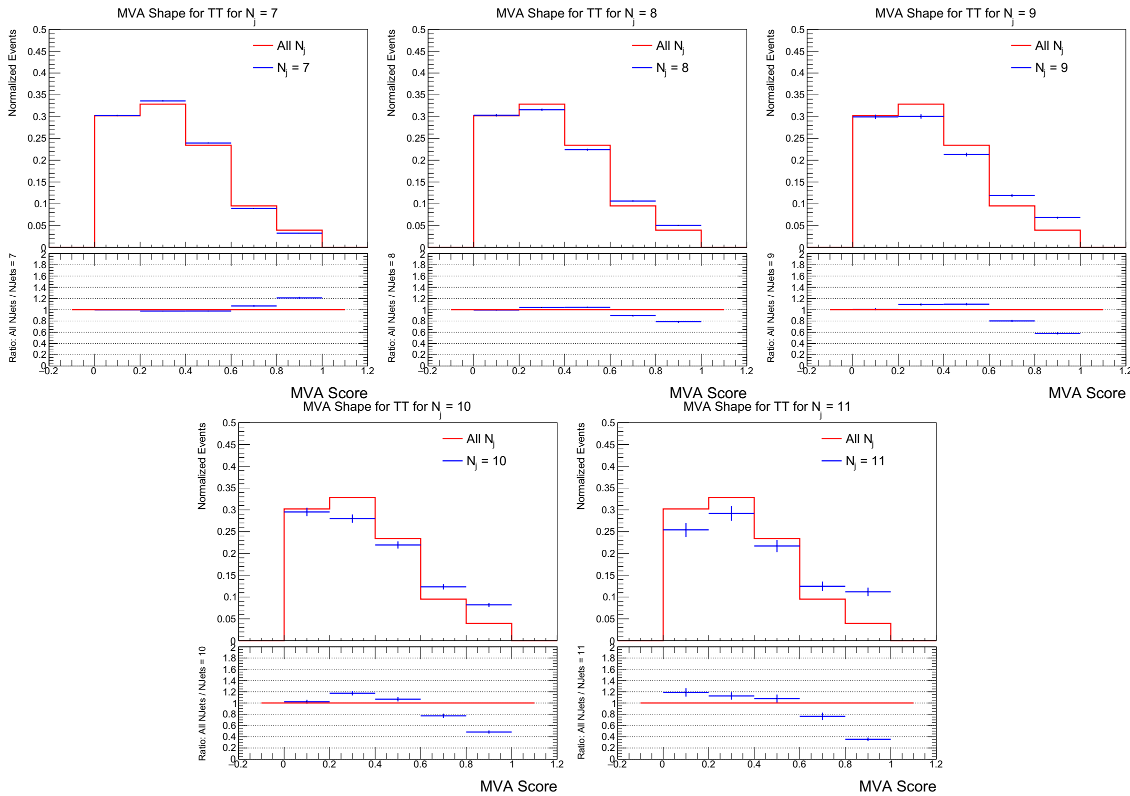


Figure A.12: The NN shape distribution for (top) N_j bins 7–9 and (bottom) N_j bins 10 and 11 relative to the overall NN shape distribution for all N_j for the $\mathrm{t\bar{t}}$ MC data in the signal region.

From the comparison of collision data in the control region vs. $\mathrm{t\bar{t}}$ MC data shown in Fig. A.13, a new $\mathrm{t\bar{t}}$ systematic uncertainty is derived using the following procedure:

- First, four new N_j histograms ($D_{i,j=0}$) are filled with all of the events in the $\mathrm{t\bar{t}}$ MC data that pass the signal selection. For each event, the total event weight

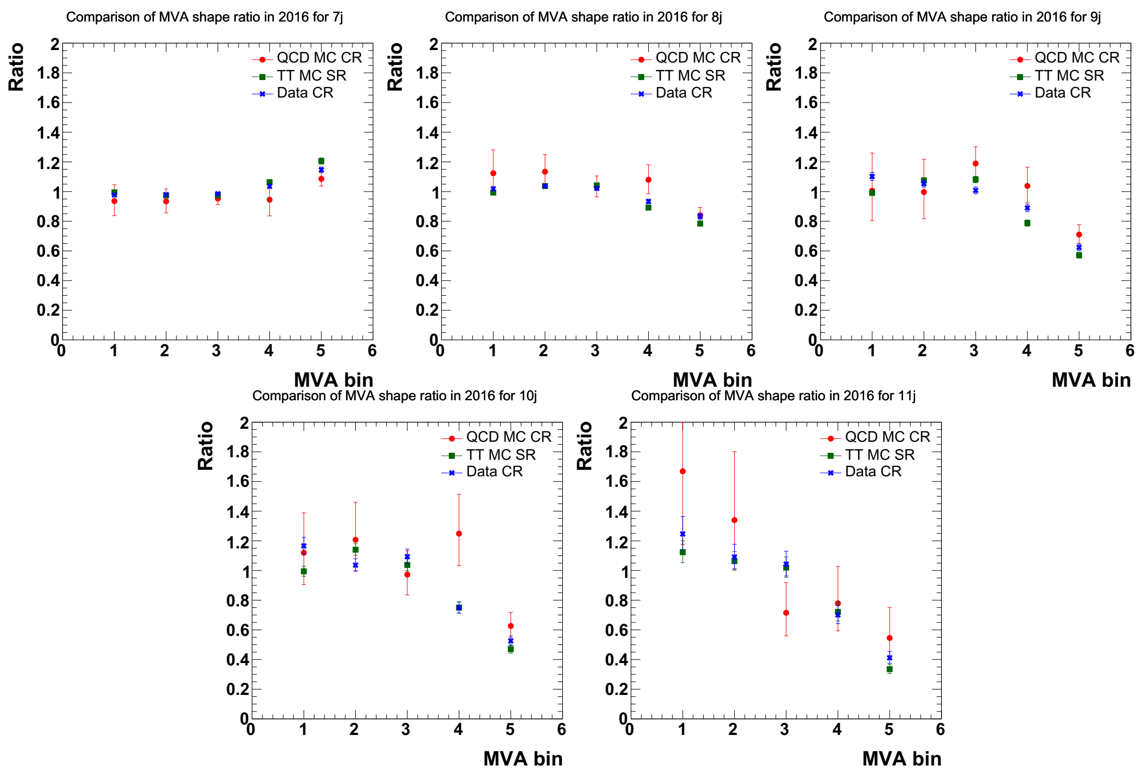


Figure A.13: The ratio of the full NN shape distribution, i.e. for all N_{j}, to the NN shape distribution for (top) N_{j} bins 7–9 and (bottom) N_{j} bins 10 and 11 for the data in the QCD control region, QCD MC data in the control region, and $\mathrm{t\bar{t}}$ in the signal region. The five MVA bins are defined for the NN score from 0.0 to 1.0 with an interval of 0.2.

is multiplied by a factor chosen randomly from a Gaussian distribution with a mean set from the value taken from Fig. A.13 and its sigma set from the uncertainty shown in Fig. A.13.

- These four N_j histograms are summed together to get the total histogram ($D_{total,j=0}$).
- The four histograms are normalized so that the area under each histogram is 1, and the N_j distribution for each NN bin is divided by the total ($\frac{D_{i,j=0}}{D_{tot,j=0}}$).
- The resulting four histograms (one for each $D_{i,j=0}$) are used to define $R_{\mathrm{syst},j=0}$, as done for the other $\mathrm{t\bar{t}}$ systematic uncertainties.
- The ratio of $R_{\mathrm{syst},j=0}$ to R_{nom} gives the shape of the systematic uncertainty.

This procedure is repeated 100 times ($j = 0$ to 99) and the final shape uncertainty uses the value in each N_j, NN bin as the mean of the 100 trials. The error bar for each bin consists of two components added in quadrature: (1) the width of the distribution of the 100 trials, which constitutes the statistical component and (2) the full deviation from one of the systematic uncertainties ($|1 - R|$) as an estimate for how well the shape is known. The nominal value and the quadrature sum of these two error components are shown in Fig. A.14.

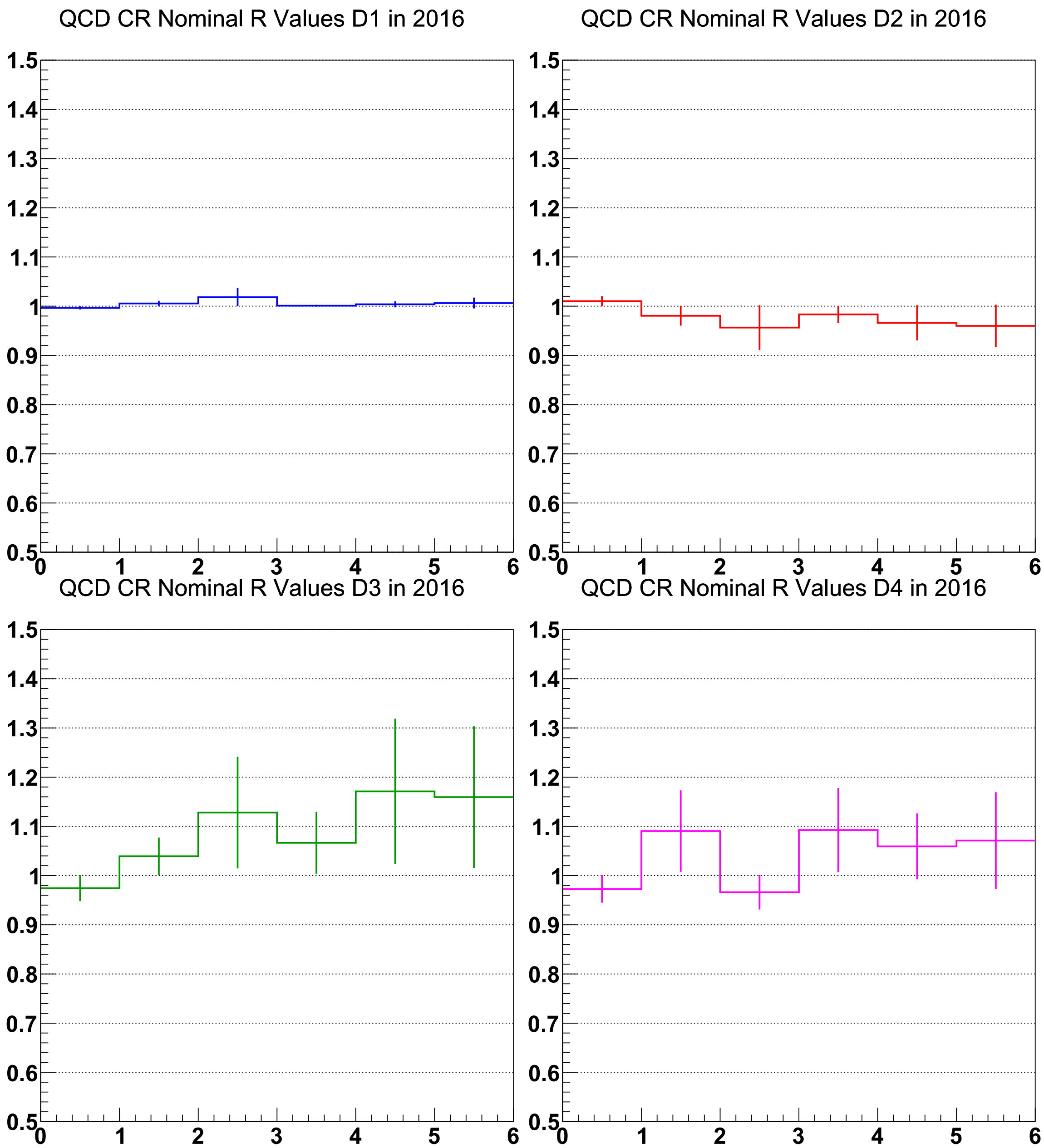


Figure A.14: Systematic uncertainty that gives an estimate on the impact of differences between the NN shape vs. N_{j} between data and MC data in 2016. The x axis shows N_{j} shifted down such that $N_{\mathrm{j}} = 7$ is the first bin noted by 0. The (top left) is for NN bin D1, the (top right) is for NN bin D2, the (bottom left) is for NN bin D3, and the (bottom right) is for NN bin D4. This systematic uncertainty is for the $\mathrm{t\bar{t}}$ background yields.

A.4 Jet Mass-p_{T} Rescaling

A.4.1 Jet p_{T} Rescaling

Figure A.15 shows distributions of jet p_{T} for the signal region and QCD control region for the two highest-p_{T} jets in the event. The ratio of data to MC data slopes downward as a function of jet p_{T}, suggesting that the jet p_{T} spectrum in the MC data is harder than in the data. It is interesting to note that the pattern is the same for the signal region and the QCD control region, suggesting that the discrepancy may have a common source in the generation and simulation of events with high jet multiplicities for both the QCD and $\mathrm{t\bar{t}}$ event simulation. Under this assumption, this section describes a systematic variation of the jet p_{T} scale in the event simulation that is derived from the data vs. MC data comparison in the QCD control region and then applied to the $\mathrm{t\bar{t}}$ in the signal region. By observing that scaling jet p_{T} in MC data by 0.95 yields a much flatter ratio of data to MC data, a $\mathrm{t\bar{t}}$ shape uncertainty is derived when scaling jet p_{T} (and jet mass).

To simplify things, the data are compared to QCD MC data in the QCD control region ignoring all other subleading MC data components. It is found that the ratio of data to MC data can be substantially flattened simply by rescaling the MC data p_{T} by the RMS ratio of the data and the MC data ($\mathrm{RMS_{data}/RMS_{QCD}}$). It is also found that the RMS ratio is about 0.95 and that the ratio is independent of jet p_{T} rank. The ratio of 0.95 works well for the 2016, 2017, and 2018 data sets. Figure A.16 shows the data and QCD MC data before and after this simple jet p_{T} scaling is applied. The RMS ratio is given at the top of each plot in the first column. The post-scaling plots in the third column show that the histogram mean and RMS

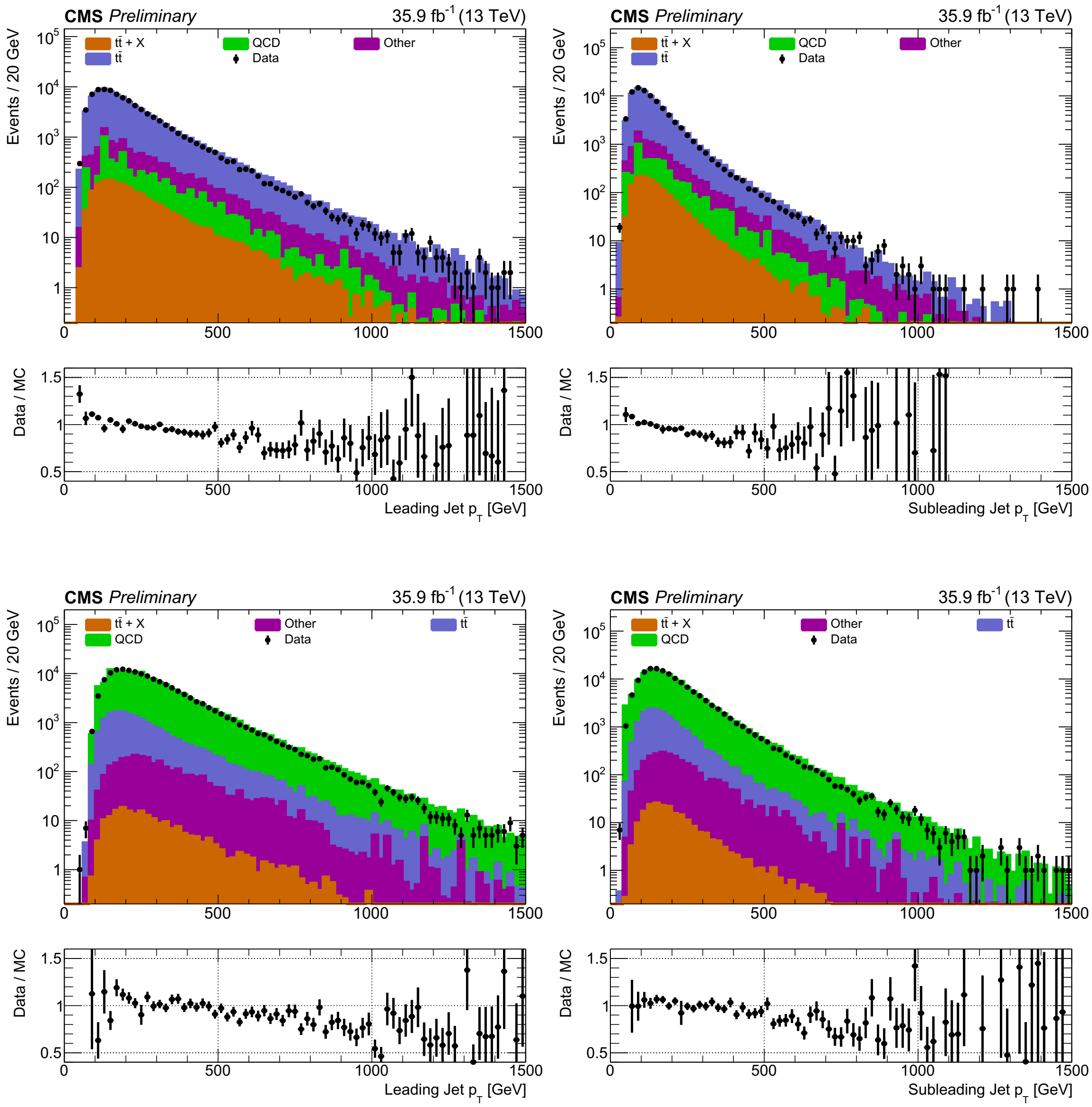


Figure A.15: Distributions of jet p_{T} for the signal region (top row) and QCD control region (bottom row) for the two highest-p_{T} jets in the event for the 2016 data set.

agree much better after the scaling. The fourth column shows the post-correction ratio of data to MC data as a function of jet p_{T}, which is much flatter.

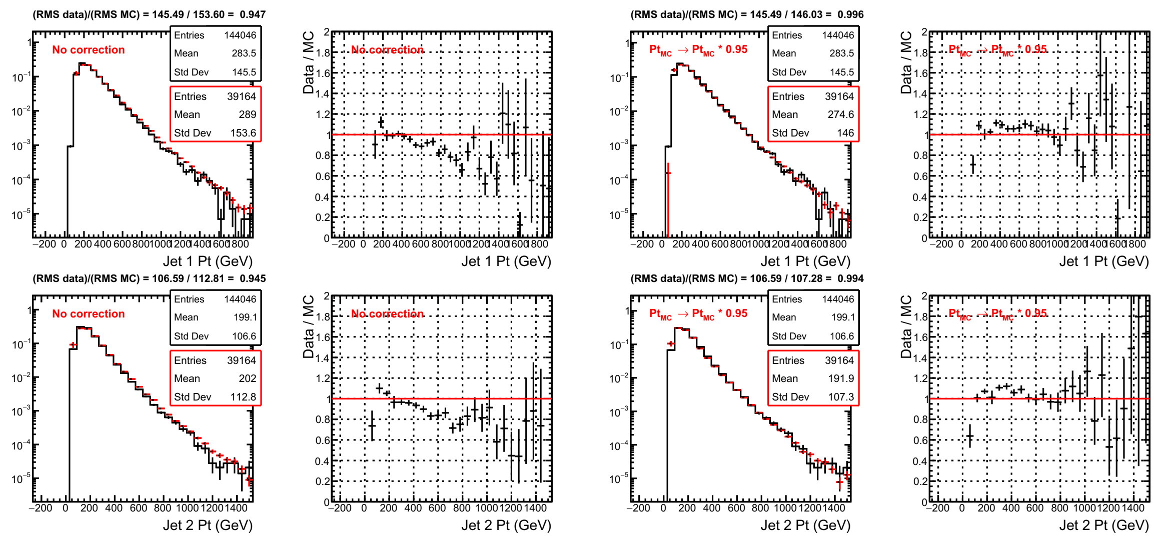


Figure A.16: Jet p_{T} distributions in the QCD control region for the data (black) and the QCD MC data (red) for the 2016 data set. The rows are sorted by jet p_{T} rank in the event, highest at the top. The first column shows the jet p_{T} distributions without any MC data correction. The second column shows the ratio of data to MC data of the first column. The third column shows the distributions after scaling the jet p_{T} by a factor of 0.95 for the QCD MC data. The fourth column shows the ratio of data to MC data of the third column. The distributions are normalized to unit area for $p_{\mathrm{T}} > 30$ GeV.

A.4.2 Jet Mass Rescaling

Figure A.17 shows distributions of the jet mass for the signal region and QCD control region for the two highest-p_{T} jets in the event. The ration of data to MC data demonstrates a clear downward trend. Seeing that the behavior in the control region and signal region is similar, this section describes a systematic variation of the jet mass scale in the event simulation that is derived from the data vs. MC data comparison in the QCD control region. By observing that scaling jet mass in MC data yields a much flatter ratio of data to MC data, a $t\bar{t}$ shape uncertainty is derived while simultaneously scaling jet mass and p_{T}.

Following the same procedure performed in the jet p_{T} scaling study in Section A.4.1 but instead using the jet mass spectrum, it is found that scaling jet mass by about 0.95 flattens the ratio of data to MC data. The value of 0.95 works well for 2016, while 1.01, 0.98, and 0.98 are the derived scale factors for 2017, 2018pre, and 2018post, respectively. Figure A.18 shows the data and QCD MC data before and after this simple jet mass scaling is applied. The post scaling plots are shown in the even columns and results for the first two jets are shown. The post-correction ratio of data to MC data as a function of jet mass is much flatter.

A.4.3 Deriving a $t\bar{t}$ *Shape Systematic Uncertainty*

As shown in Fig. A.19, since jet p_{T} and mass are correlated, a single uncertainty is derived by scaling jet mass and p_{T} at the same time. The systematic uncertainty is derived using the procedure described in Section 5.6.2. The magnitude and shape of the systematic uncertainty in each NN bin are shown for 2016 in Fig. A.21.

When this scaling of jet mass and p_{T} is performed to produce the varied N_{j} distribution, H_{T} reweighting is not applied. This is because the scaling of jet p_{T} and the reweighting of event H_{T} yield similar effects. In order to decouple the change in N_{j} shape simply by not applying H_{T} reweighting, an independent shape uncertainty is derived by removing the nominal H_{T} scale factor. This "noHT" shape variation is ultimately "divided out" from the systematic uncertainty shape derived for mass-p_{T} scaling, i.e. the systematic uncertainty shape from scaling mass and p_{T} (no H_{T}

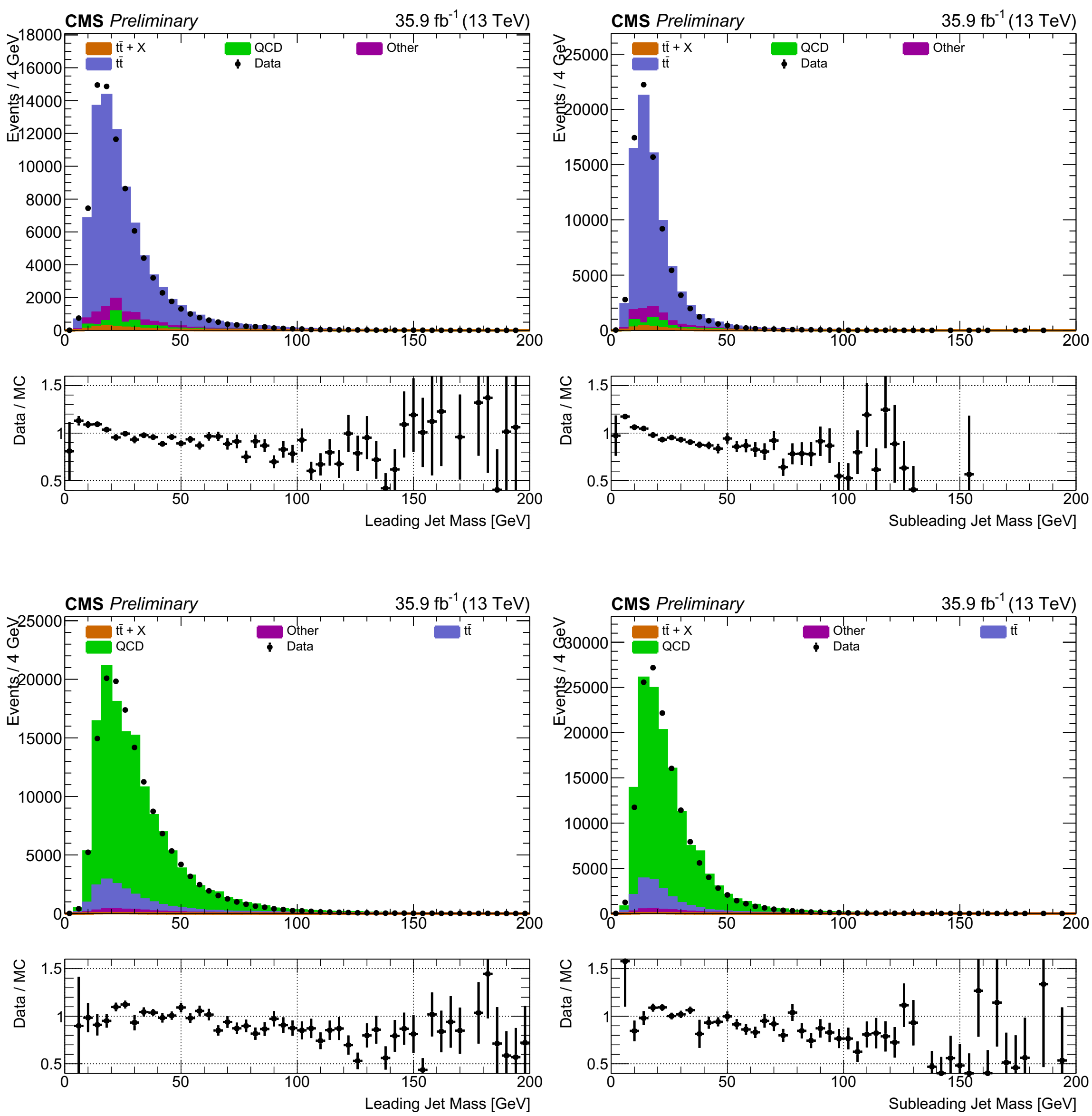


Figure A.17: Distributions of jet mass for the signal region (top row) and QCD control region (bottom row) for the two highest-p_{T} jets in the event for the 2016 data set. Here, MC data has been normalized to data to focus on shape differences.

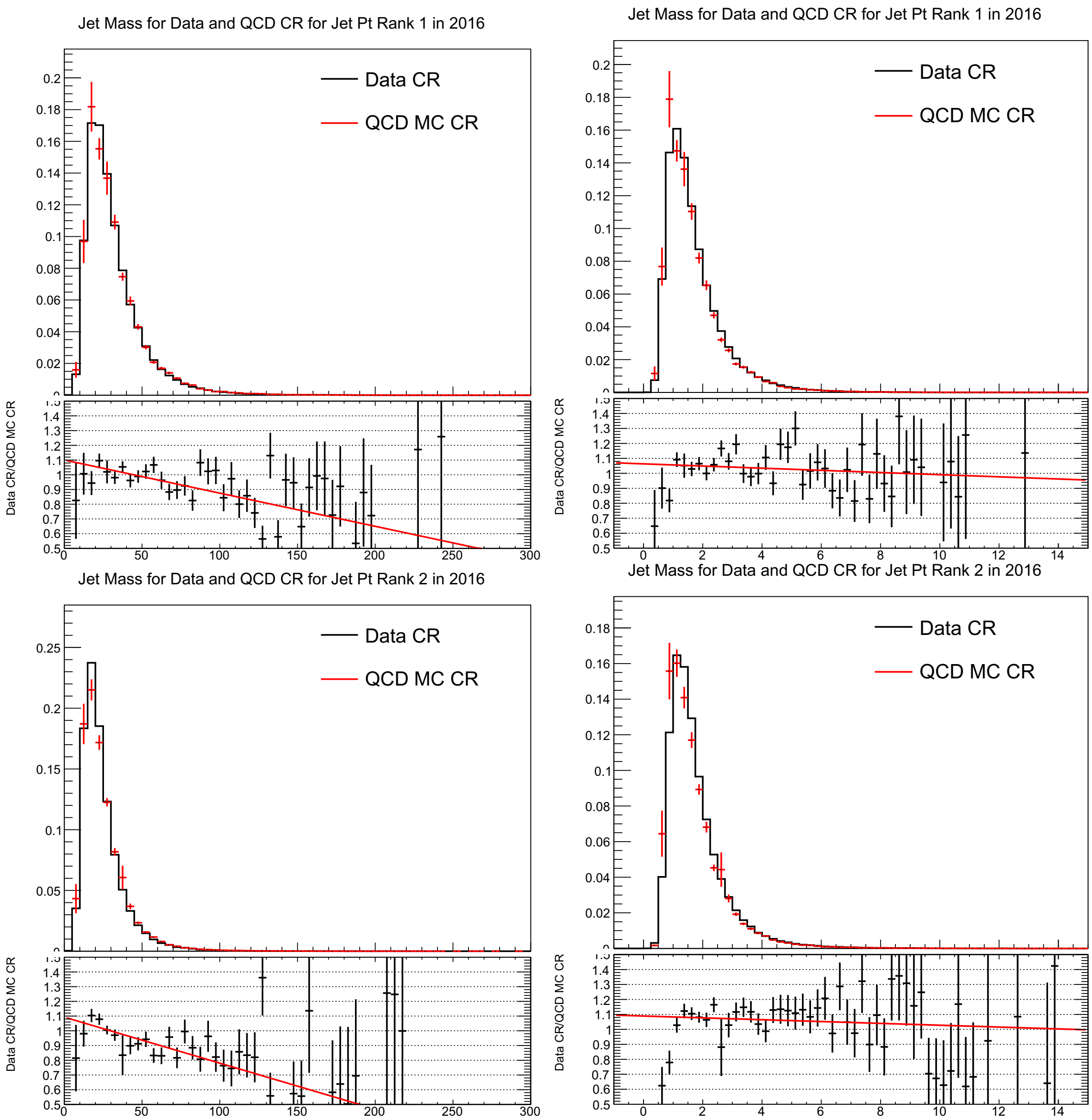


Figure A.18: Jet mass distributions in the QCD control region for the data (black) and the QCD MC data (red) for the 2016 data set. The plots are grouped in pairs: jet 1 is shown in the top row, and jet 2 is in the bottom row. The left plot for each pairing is before rescaling and the right plot is after rescaling.

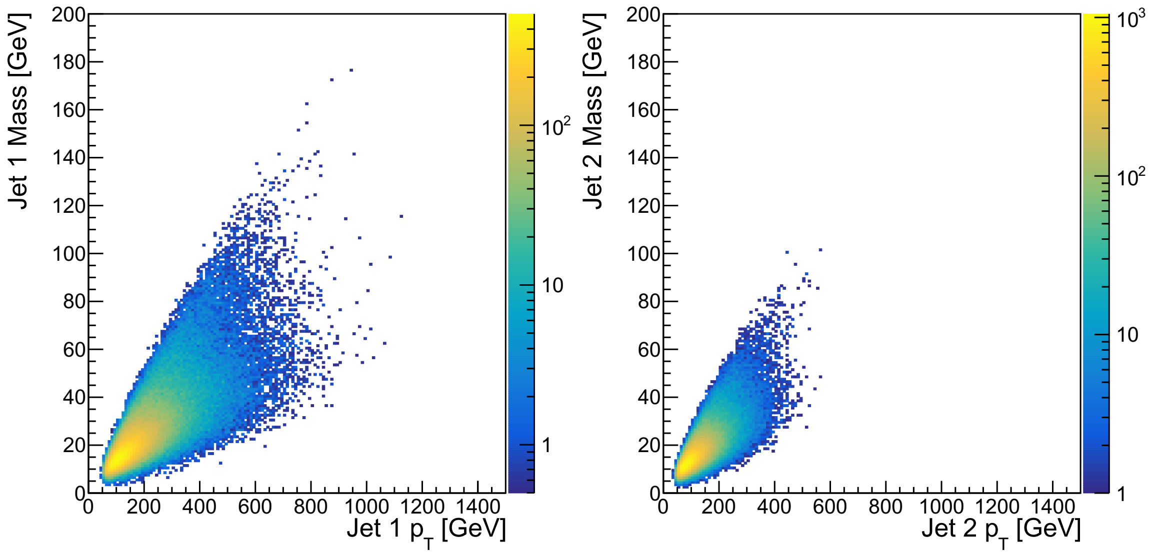


Figure A.19: Jet mass and p_{T} for $\mathrm{t\bar{t}}$ in the signal region. The two jet properties are clearly correlated across jet rank.

reweighting) is divided by the shape from only removing H_{T} reweighting, for each NN bin.

After performing this extra division, two orthogonal systematic uncertainty shapes are left. There is a mass-p_{T} scaled systematic uncertainty, which solely accommodates N_{j} shape changes from the scaling of jet mass and p_{T}. Then there is a new "noHT" systematic uncertainty, which gives a measure of how the H_{T} reweighting affects the $\mathrm{t\bar{t}}$ N_{j} shapes per NN bin.

Although the mass-p_{T} scaled systematic uncertainty is the primary one derived here, a conservative approach is chosen to retain the "noHT" systematic uncertainty. It is seen—at least for 2017, in Fig. A.20—that although H_{T} reweighting resolves the disagreement between data and MC data for H_{T}, it can degrade the agreement between data and MC data for jet mass. Thus, the result of removing H_{T} reweighting is treated as a valid variation and the corresponding uncertainty is propagated through the analysis.

Figure A.21 shows the three aforementioned systematic uncertainty shapes in the four NN bins. In purple is the shape from scaling mass and p_{T} by 0.95 and also removing H_{T} reweighting. In red is the shape obtained by simply removing H_{T} reweighting, and in blue is the final mass-p_{T} scaled shape when dividing the purple shape by the red shape.

To understand how significant these uncertainties are, the induced shape differences in the same systematic uncertainty ratio are computed when the signal N_{j} shape is added to the $\mathrm{t\bar{t}}$ shape, i.e. treating signal as a systematic uncertainty. The effect is shown in Fig. A.22 for three masses for the RPV SUSY model. Overall, both the jet mass-p_{T} rescaling and removal of H_{T} reweighting induce changes similar to how signal induces changes across NN bins, suggesting that these are important uncertainties to incorporate.

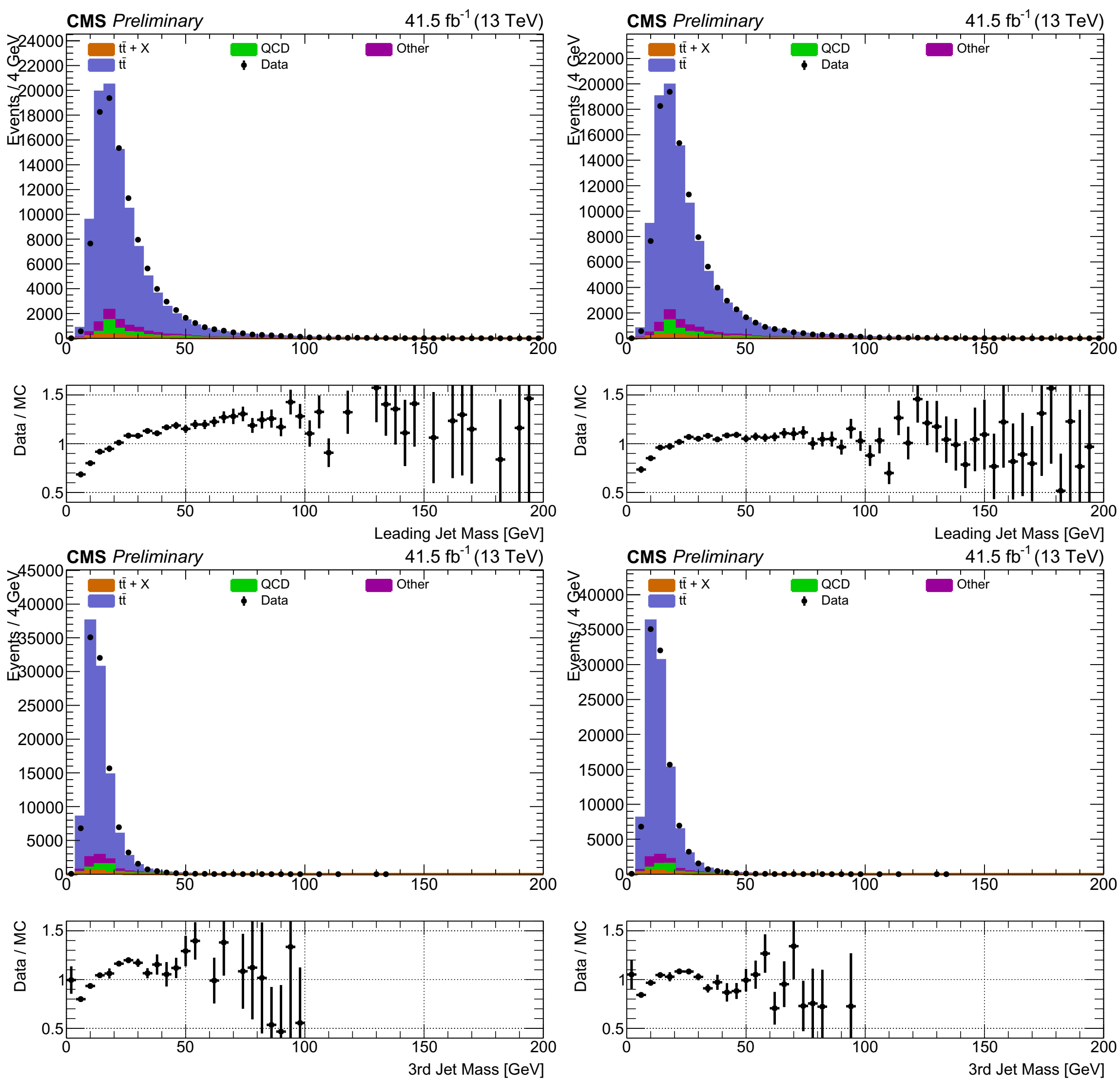


Figure A.20: Data vs. MC data comparisons for jet mass for 2017. The top two plots are for the leading jet in the event, and the bottom two for the third jet in the event. Progressing from left to right shows the result of removing H_{T} reweighting and it is seen that the data vs. MC data agreement improves by doing so.

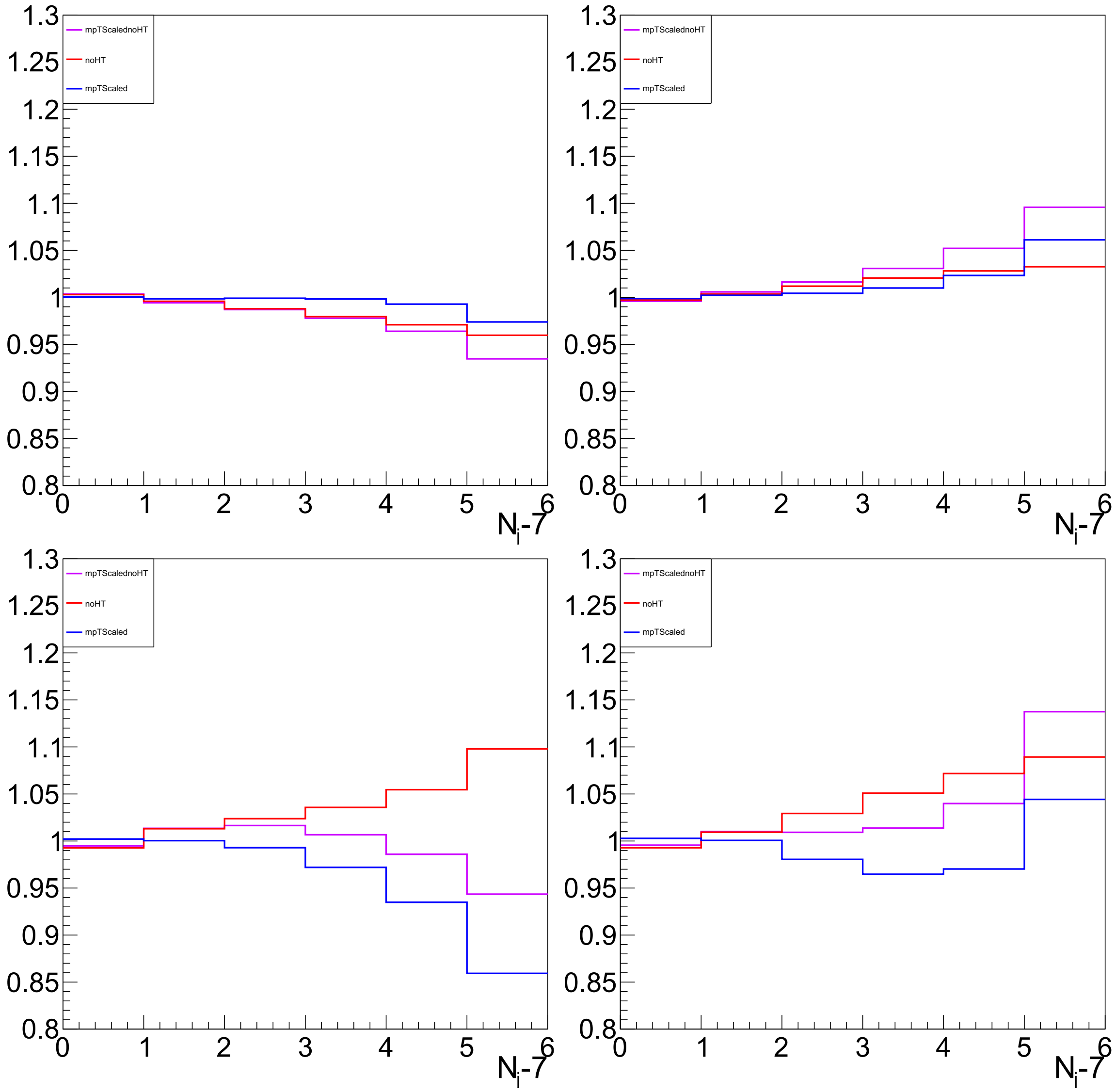


Figure A.21: The ratio $R = R_{\mathrm{syst}}/R_{\mathrm{nom}}$, which shows the N_{j} shape differences introduced by scaling all jet mass and p_{T} by 0.95 for the four NN bins: D_1 (top left), D_2 (top right), D_3 (bottom left), and D_4 (bottom right). The x axis shows N_{j} shifted down such that $N_{\mathrm{j}} = 7$ is the first bin, here noted by 0. Shown in purple is the mass-p_{T} scaled shape (no H_{T} reweighting), in red is the shape change resulting from only the removal of H_{T} reweighting, and blue is the ratio of the former to the latter.

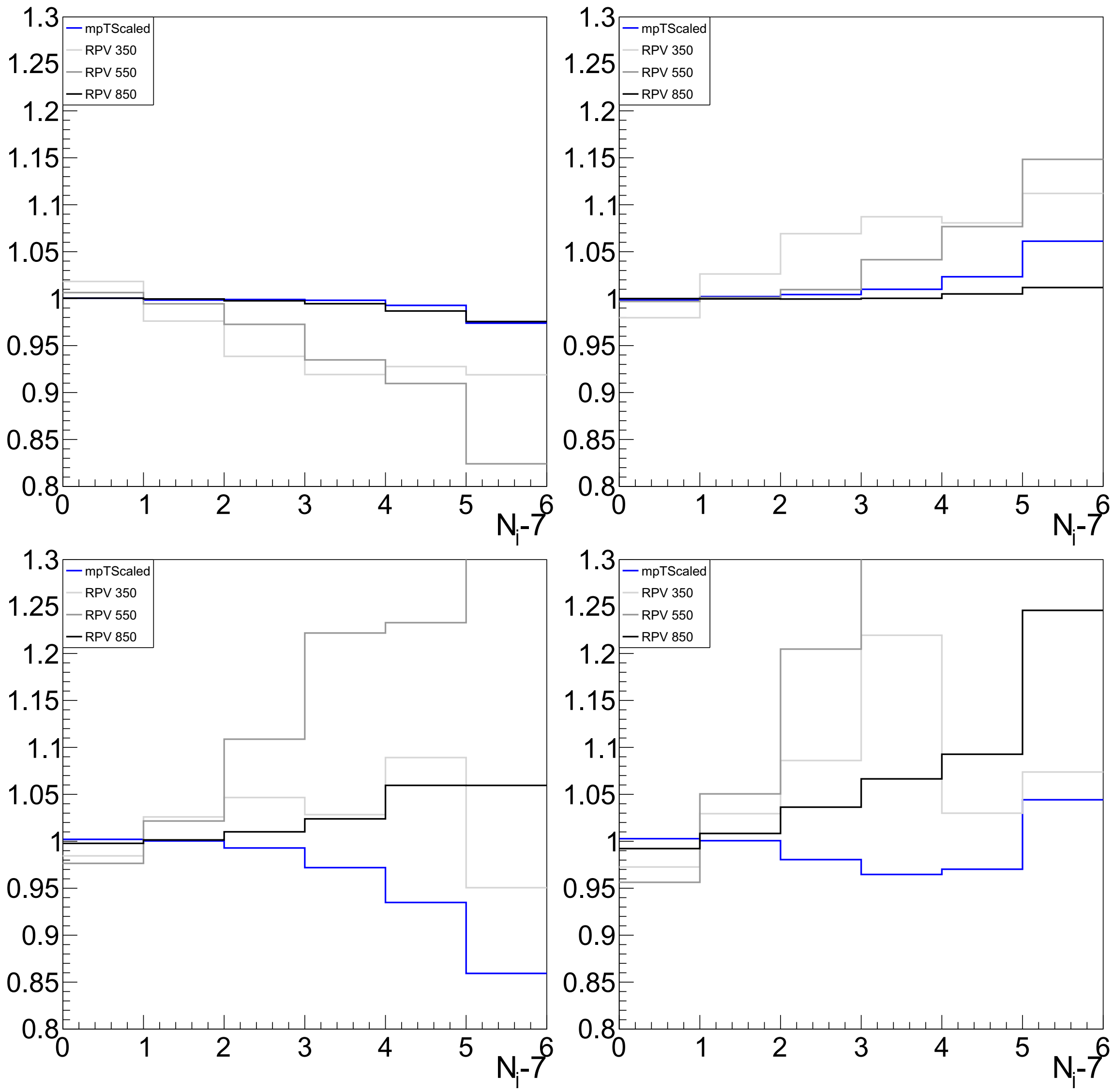


Figure A.22: The ratio $R = R_{\mathrm{syst}}/R_{\mathrm{nom}}$, which shows the N_{j} shape differences introduced by scaling all jet mass and p_{T} by 0.95 for the four NN bins: D_1 (top left), D_2 (top right), D_3 (bottom left), and D_4 (bottom right). The x axis shows N_{j} shifted down such that $N_{\mathrm{j}} = 7$ is the first bin, here noted by 0. As a direct comparison, the induced shape differences in the same ratio are shown when the signal N_{j} shape is added to the $\mathrm{t\bar{t}}$ shape (treating signal as a systematic uncertainty). This is done for three masses for the RPV SUSY model. Overall, the jet mass-p_{T} rescaling induces changes similar to how signal would induce changes.

BIBLIOGRAPHY

[1] CMS Collaboration, "Measurements of properties of the Higgs boson in the four-lepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV," CMS Physics Analysis Summary CMS-PAS-HIG-19-001, 2019. https://cds.cern.ch/record/2668684

[2] CMS Collaboration, "A measurement of the Higgs boson mass in the diphoton decay channel," *Phys. Lett. B*, vol. 805, p. 135425, 2020.

[3] Wikipedia contributors, "Standard model — Wikipedia, the free encyclopedia." https://en.wikipedia.org/wiki/Standard_Model

[4] T. Lancaster and S. J. Blundell, *Quantum field theory for the gifted amateur*. Oxford: Oxford University Press, 2014. https://cds.cern.ch/record/1629337

[5] M. E. Peskin and D. V. Schroeder, *An introduction to quantum field theory*. Boulder, CO: Westview, 1995, includes exercises. https://cds.cern.ch/record/257493

[6] C. P. Burgess and G. D. Moore, *The standard model: A primer*. Cambridge University Press, 2006.

[7] D. Goldberg, *The standard model in a nutshell*. Princeton, NJ: Princeton University Press, 2017. http://cds.cern.ch/record/2244785

[8] M. Robinson, *Symmetry and the Standard Model: mathematics and particle physics*. New York: Springer, 2011. https://cds.cern.ch/record/1399152

[9] A. Djouadi, "The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model," *Phys. Rept.*, vol. 459, pp. 1–241, 2008.

[10] R. N. Mohapatra, *Unification and supersymmetry: the frontiers of quark-lepton physics; 3rd ed.*, ser. Graduate texts in contemporary physics. Berlin: Springer, 2003. http://cds.cern.ch/record/737303

[11] P. Langacker, *The standard model and beyond; 2nd ed.*, ser. High energy physics, cosmology and gravitation. CRC Press, 2017. http://cds.cern.ch/record/2256595

[12] M. Kuroda, "Complete Lagrangian of MSSM," 1999.

[13] S. P. Martin, "A Supersymmetry primer," pp. 1–98, 1997, [Adv. Ser. Direct. High Energy Phys. 18, 1 (1998)].

[14] A. Breskin and R. Voss, *The CERN Large Hadron Collider: Accelerator and Experiments*, 2009. https://cds.cern.ch/record/1244506

[15] CMS Collaboration, "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software," CERN-LHCC-2006-001 ; CMS-TDR-8-1, 2006. http://cds.cern.ch/record/922757

[16] M. Hoch, "CMS pixel detector," 2014, CMS Collection. http://cds.cern.ch/record/1977415

[17] CMS Collaboration, "CMS Technical Design Report for the Pixel Detector Upgrade," CERN-LHCC-2012-016. CMS-TDR-11, 2012. https://cds.cern.ch/record/1481838

[18] M. Brice, T. Virdee, and T. Camporesi, "Images of CMS HCAL Forward Calorimeter (HF)," 2008, CMS Collection. https://cds.cern.ch/record/1431489

[19] CMS Collaboration, "CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter," CERN-LHCC-2012-015. CMS-TDR-10, 2012. https://cds.cern.ch/record/1481837

[20] CMS Collaboration, "Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays," *JINST*, vol. 5, no. arXiv:0910.5530. CMS-CFT-09-015, p. T03021. 35 p, 2009. http://cds.cern.ch/record/1215500

[21] CMS Collaboration, "Particle-flow event reconstruction in CMS and performance for jets, taus, and $E_{\mathrm{T}}^{\mathrm{miss}}$," CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009. http://cdsweb.cern.ch/record/1194487

[22] CMS Collaboration, "Commissioning of the particle-flow event reconstruction with the first LHC collisions recorded in the CMS detector," CMS Physics Analysis Summary CMS-PAS-PFT-10-001, 2010. http://cdsweb.cern.ch/record/1247373

[23] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector," *JINST*, vol. 12, no. 10, p. P10003, 2017.

[24] S. Höche, "Introduction to parton-shower event generators," in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014): Boulder, Colorado, June 2-27, 2014*, 2015, pp. 235–295.

[25] C. Madrid *et al.*, "HCAL simulation tuning of the time slew and pulse shape for sipm readout channels," CMS Detector Note DN-17-036, 2017. http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20DN-2017/036

[26] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_{T} jet clustering algorithm," *JHEP*, vol. 04, p. 063, 2008.

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

[28] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet user manual,” *Eur. Phys. J. C*, vol. 72, p. 1896, 2012.

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

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

[31] CMS Collaboration, “Study of pileup removal algorithms for jets,” CMS Physics Analysis Summary CMS-PAS-JME-14-001, 2014. http://cds.cern.ch/record/1751454

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

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

[34] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV,” *JINST*, vol. 13, no. CMS-BTV-16-002. CMS-BTV-16-002-004. 05, p. P05011. 114 p, 2017. https://cds.cern.ch/record/2298594

[35] J. A. Evans and Y. Kats, “LHC Coverage of RPV MSSM with Light Stops,” *JHEP*, vol. 04, p. 028, 2013.

[36] J. Fan, M. Reece, and J. T. Ruderman, “Stealth Supersymmetry,” *JHEP*, vol. 11, p. 012, 2011.

[37] J. Fan, R. Krall, D. Pinner, M. Reece, and J. T. Ruderman, “Stealth Supersymmetry Simplified,” *JHEP*, vol. 07, p. 016, 2016.

[38] J. Fan, M. Reece, and J. T. Ruderman, “A Stealth Supersymmetry Sampler,” *JHEP*, vol. 07, p. 196, 2012.

[39] F. Chollet *et al.*, “Keras,” https://keras.io, 2015.

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

[41] G. C. Fox and S. Wolfram, “Observables for the analysis of event shapes in e^+e^- annihilation and other processes,” *Phys. Rev. Lett.*, vol. 41, pp. 1581–1585, 1978.

[42] J. D. Bjorken and S. J. Brodsky, “Statistical model for electron-positron annihilation into hadrons,” *Phys. Rev. D*, vol. 1, pp. 1416–1420, 1970.

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

[44] S. Chinnamgari, *R Machine Learning Projects.* Packt Publishing, 2019.

[45] E. Gerwick, T. Plehn, S. Schumann, and P. Schichtel, "Scaling Patterns for QCD Jets," *JHEP*, vol. 10, p. 162, 2012.

[46] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *Eur. Phys. J.*, vol. C71, p. 1554, 2011, [Erratum: Eur. Phys. J.C73,2501(2013)].

[47] ATLAS and CMS Collaborations, "Procedure for the LHC Higgs boson search combination in Summer 2011," CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, 2011. https://cds.cern.ch/record/1379837

[48] Wikipedia contributors, "689599.7 rule — Wikipedia, the free encyclopedia." https://en.wikipedia.org/wiki/68-95-99.7_rule

[49] B. A. Dobrescu and F. Yu, "Exotic Signals of Vectorlike Quarks," *J. Phys. G*, vol. 45, no. 8, p. 08LT01, 2018.