MEASUREMENT OF THE THREE TOP QUARKS PROCESS WITH THE CMS DETECTOR AT THE LHC

by

Dylan O. Teague

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Physics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2025

Date of final oral examination: 12/04/2025

The dissertation is approved by the following members of the Final Oral Committee:

Kevin Black, Professor, Physics
Sridhara Dasu, Professor, Physics
Vernon Barger, Professor, Physics
Ellen Zweibel, Professor, Astronomy

Colophon

This document was typeset using LaTeX, adapting the thesis formatting create by William C. Benton. The bibliography was set to the CMS format which is in turn based on IEEE Transactions bibliography style. The typeface used in this document is Crimson Text, an open-source typeface created by Sebastian Kosch, inspired by Garamond. For math equations, the TeX Gyre Termes Math typeface was used, developed by the open-source project GUST and based on the URW Nimbus Roman No9 L typeface developed by URW Studio.

Acknowledgments

This PhD has been an effort of solitary work and study while simultaneously being supportive and collaborative, in no small part because of the Covid-19 pandemic. I havean acute awareness of the true utility of the both friends and coworkers who stood with me during these years because of this unique experience. First and foremost, I must thank my mom and dad, Rowena and Phil Teague, for their encouragement and help in furthering my physics education from the very start. Without them taking interest in my journey, it wouldn't exist. I must thank my advisor, Kevin Black, for his support and confidence in this analysis as well as the other professors, Sridhara Dasu, Tulika Bose, and Matt Herndon, and scientist, Sasha Savin. My collaborators, Deborah Pinna and Daniel Li, in particular, have worked to realize this analysis and have been critical of my work to make it into its best form. I must also thank my Wisconsin cohort, Ganesh Parida, Abby Warden, Susmita Mondal, Anagha Aravind, He He, Victor Shang, Wren Vetens, Abhi Mallampalli, as well as other, elder Badgers who helped me when I started, Kenneth Long in particular.

In my personal life, the friends I made at CERN have kept me sane and been resources for discussion on physics, philosophy, and other inane topics hashed over beer and french fries every Friday at the CERN cafeteria. Without them, I don't think I would have grown as much as I have as a scientist. For that, I thank Grace Cummings, Andres Delannoy, Bing Liu, Janina Nicolini, Clyde Laforge, Gwen Gardner, and Jonas Eschle. Special attention has to be given to Kevin Sedlaczek, who I have had the most discussions and who was one of the best men at my wedding. The other best man, Samuel Higginbotham, was the sole person I saw during the pandemic, and as our friendship flourished through the years that followed, he has been my greatest advocate, most helpfully resource, and my best friend through and through.

Lastly and most importantly, I have to thank my wife, who I met and married during my PhD while she is finishing hers as well. Our mutual bond in doctoral work has been the primary driver in keeping me going through this hard, long process. While a PhD represent only a tiny step in increasing human knowledge, I will always have the joy of remembering back on these times as the best and most consequential step for me by choosing to spend my life with her.

Contents

MEASUREMENT OF THE THREE TOP QUARKS PROCESS WITH THE CMS DETECTOR AT THE LHC

Dylan O. Teague

Under the supervision of Professor Kevin Black
At the University of Wisconsin-Madison

This thesis presents the first analysis measuring the three top quark decay process at the Large Hadron Collider. The measurement was performed on events with two same-sign leptons and three or more leptons on data taken by the Compact Muon Solenoid detector during the 2016–2018 data taking period corresponding to 137 fb^{-1} of data with $\sqrt{s} = 13$ TeV. The analysis found the three top process to have an overall significance of 0.182σ and a CLs 95 % upper limit of 28.1 fb. These results were also uses to reinterpret the data to place limits on possible beyond Standard Model physics models that predict enhancements to the three top quark cross-section.

Chapter 1

Introduction

Particle physics, or the study of subatomic particles, is a rich field of physics. While particle physics represents a fourth of all Nobel Prizes in the past 75 years, it only became recognizable in its current iteration in the past century which is quite new compared to the more mature disciplines such as astronomy. The branch owes its origins to the discovery of the atom and quantum mechanics, but the start of what is considered modern, fundamental particle physics is quite fuzzy. One could argue it began with the discovery of the first fundamental particle, the electron, or even later with events such as the discovery of the first subnuclear particles in the 1940s and '50s. This ambiguity demonstrates that particle physics has seen large, paradigm shifting discoveries over the past century that have changed not only our understanding of the subatomic, but of the fundamental nature of the universe.

The study of particles started obviously with the discovery of the first "particle": the atom. While the idea of the atom was seriously considered as a scientific theory in the 19th century by chemists owing to discoveries such as the periodic table, the atomic structure was not truly studied until 1897 with the discovery of the electron by J. J. Thomson.[1] Through his work, the first atomic model was created. This model was further refined with Rutherford's gold foil experiment in 1911 which suggested the existence of an atomic nucleus.[2] By today's categorization of research, these findings would be considered atomic and/or nuclear physics considering the scale they study, these distinctions did not exist and could only once researchers of the time investigated further and probe deeper into the atom.

Since then, the current roster of known particles has expanded to a staggering number, most only being observed at a particle accelerator. Similar to how the periodic table changed the understanding of the atom, the Standard Model tamed this zoo of particles. The Standard Model notes the elementary particles of nature and how they interact with each other, allowing the menagerie of particles to be organized as one of these elementary particles or combinations of them. It makes predictions that have been tried and tested for decades making it is now one of the most robust and resistant theories ever created.

Currently, the world of fundamental particle physics centers around CERN, or the European Center

for Nuclear Research, an international research facility located in Geneva, Switzerland. Built in 1954, CERN has been a leading institution for furthering particle physics, marking itself in the annals with the discovery of W and Z bosons in 1983 [3–6] and the Higgs boson in 2012 [7, 8]. CERN still distinguishes itself to this day by being the only current accelerator performing elementary high energy physics experiments.

The question *du jour* is what comes next. After the discovery of the Higgs boson, experimental tests at CERN have resulted in more confirmation of the Standard Model, but no new physics has been found despite theorists' speculation that the Standard Model is incomplete. After the confirmation of electroweak theory, or the unification of electromagnetism and the weak force, theorists have supposed that all forces can be unified, suggesting our current, non-unified model is incorrect, or at least incomplete. Even if there is no grand unification, the Standard Model breaks down at large energy scale such as at the Planck scale where gravity would have an appreciable effect. Similarly, dark matter has been experimentally confirmed but has never been directly observed. If dark matter is a particle, this would suggest the Standard Model necessarily is incomplete and at the very least requires an expansion to make room for the new particle. For the experimentalist, there are two main approaches to look for these new physics models: search for new physics proposed in a specific theorist's paper or perform precision measurements to strength test the Standard Model.

The first approach lacks certain motivations because there are currently no clues for the form of this new physics, let alone how to search for it. New theories tend to exist in unmeasured areas of Standard Model exactly so they don't disturb well tested ground while allowing a way to explain new phenomenon. Popular theories, such as supersymmetry, have large phase-spaces with many tuneable parameters, so, while they have been tested extensively and have had many of their iterations excluded, these "tent-pole" theories have not been outright ruled out. The theories themselves remain popular because they contain interesting features, but for many, the most obvious and elegant versions have been excepted leaving only stranger extensions and permutations. This has caused some theorists to prefer making models that are smaller and more generic in their extension of the Standard Model.

The second approach eschews looking for new physics to testing the Standard Model directly. This takes the form of either making more precise measurements of known values or testing yet unmeasured portions of the Standard Model. In investigating the frontiers of the Standard Model, similar regions are investigated as with searches for new physics, both testing portions of the particle physics landscape yet untouched, but the perspective is slightly different: one concentrates on model building while the other confirms our current model's validity. This slight paradigm shift may not lead to different implementations, but it does often lead to different focuses. In fact, some of the more exciting recent results have come from these Standard Model validation studies, such as tension found in the recent W boson mass measurement[9, 10].

For this thesis, a combination is used with an emphasis on measuring the Standard Model. Here,

we are looking into the yet unplumbed three top quark process, a type of decay firmly allowable in the Standard Model. While predicted by the Standard Model, this final state is exceptionally rare and thus no analysis has been attempted to measure this fleeting process, making this a first of its kind.

Chapter 2

Theory

For describing the three top quark state, we must couch it within the context of a measurement. But to make any measurement, a good theoretical understanding is needed to describe how it is produced and decays. Particle physics, thankfully is benefited by having a rich theoretical base.

2.1 The Standard Model

Particle physics is described by a theory called the Standard Model. In the most simple terms, the Standard Model enumerates all known elementary particle and their interaction, which are shown in Figure 2.1, with all particles being divided into two groups: bosons, or integer spin particles, and fermions, or spin-½ particles. In addition, the Standard Model describes how all the particles interact with each other where the interactions are the electromagnetic force, weak force, and strong force.

2.1.1 Gauge Bosons

The integer spin bosons contain the spin-0 Higgs boson, and the spin-1 photon, W/Z bosons, and gluons. These spin-1 bosons are all force carrying gauge bosons or particles whose underlying field is invariant under a gauge transformation. A gauge transformation is a special transformation that can be applied to our model without changing the underlying physics. This invariance, or symmetry, in our model means there is a redundancy in the field description. In classical physics, symmetries lead to conservation laws through Noether's Theorem [11],

and we observe something similar in quantum mechanics, but because our fields are quantized, the symmetry rules affect both the gauge and the underlying wavefunction itself. This means one can describe all the forces in the standard model by the symmetries in the wavefunction, which can all be described by Lie group theory. Because, in quantum field theory, particles are also quantizations of the field, the individual force carrying bosons can be read off from the dimension of the symmetry group

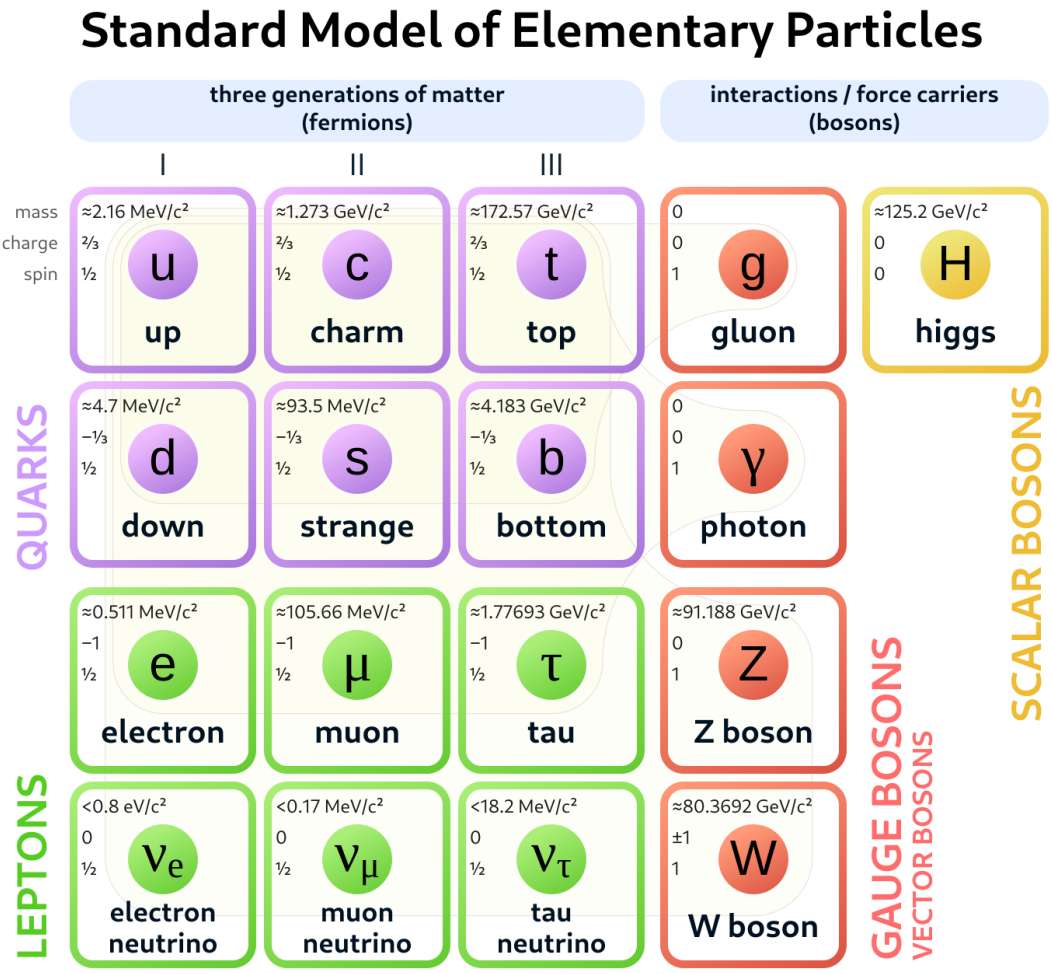

Figure 2.1: A pictorial representation of the Standard Model, showing the particles and their properties in a Euler diagram with each zone showing which forces the particle experiences.

of said gauge field. The details of how gauge symmetries appear in the Standard Model are discussed in Appendix A.

For the electromagnetic force, the underlying symmetry is U(1) group, which is 1-dimensional, so the field is quantized by one particle, the photon. The weak force symmetry is SU(2), which is a 3-dimensional group, thus the force is quantized by three particles, the Z and W^+/W^- bosons. The representation of the SU(2) symmetry in the Standard Model is 2×2 meaning the weak force acts upon a doublet of particles which happens to be a pair of fermions. The strong force has an SU(3) gauge symmetry which is 8-dimensional, so it has eight quantized particles and acts on a charge triplet which only acts on a subset of fermions called quarks. This charge triplet, or the charges of the strong force,

are named after the fundamental colors: red, green, and blue. When all three charges are combined, it produces a neutral charge, or what is often referred to as colorless. The eight force carrying particles of the strong force are all called gluons and each represent charge-anticharge combinations (such as $\bar{b}g$) that allow the transfer of color charge between particles. The model that describes the strong force is called Quantum Chromodynamics, or QCD, to allude to the color forces that govern it. While QCD is just the model that describes the strong force, QCD and strong force are often used interchangeably.

2.1.2 Higgs Boson

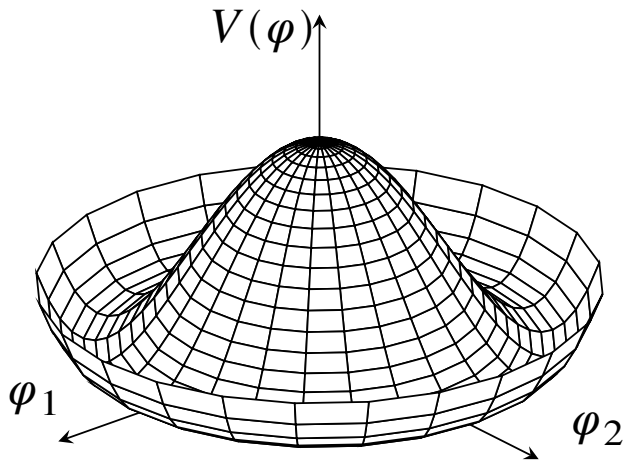

Figure 2.2: The Higgs mechanism's Mexican hat potential. Here, the trivial extremum at the origin is an unstable maximum, while the circle of points at the brim of the hat is the minimum, or where the theory ends up settling. Since there are infinite, identical minima, the theory has to arbitrarily choose one point, leading to spontaneous symmetry breaking.

Within the bosons, there is the slightly different Higgs boson. First, it is a spin-0, so it acts as a scalar with regard to the Lorentz symmetry of the Dirac equation. Second, the Higgs boson isn't associated with a force field, but rather arises from the field that gives mass to all fundamental particles, namely the Higgs field [12–14]. In the Standard Model, there is an overall potential field, φ called the Mexican hat potential (seen in Figure 2.2) which has a degenerate, non-trivial minimum. This means nature, in finding the minimum, has to pick a specific value along the "brim" of the hat. This breaks the overall symmetry of the model in a process called spontaneous symmetry breaking. Because the minimum is non-zero, it is convenient to recenter our potential around this new point, i.e. $\varphi = \nu + h$, where ν is the vacuum expectation value (VEV) or chosen minimum and h is the Higgs field or the new potential field centered at the VEV. For the Higgs field observed, it has the form of an $SU(2)$ group that couples with the electromagnetic and weak forces.

When our model is centered at this new minimum, it causes interesting results. The electromagnetic

and weak fields interact with both the Higgs field and the constant VEV term, so we introduce interaction of the form vA^2 as well as the typical hA^2 terms. Since the VEV is just a constant, this interaction acts like a mass term. Depending on where the Higgs field is centered (out of the infinite possibilities from the spontaneous symmetry breaking), one of the fields will remain massless while the other three gain the mass term[†]. In the Standard Model, the Higgs boson interacts with a merged electroweak force, breaking its underlying structure to create the form of the photon in the electromagnetic force and the massive W and Z bosons in the weak force. The Higgs potential also interacts with the fermions in the Standard Model giving them mass by a similar mechanism. In this case, the mass of each fermion is proportional to the VEV and its coupling strength to the Higgs boson, or what is called its Yukawa coupling.

2.1.3 Fermions

Opposite the bosons are the fermions which make up the normal matter found in nature such at the humble electron and up/down quarks (which combine to form protons and neutrons). While the gauge bosons describe the underlying force they facilitate, the fermions are what the gauge bosons or the forces act upon. Because of this, the fermions are divided up by what forces they experience. The strong force only interacts with the fermions called quarks while the other particles have no color charge and are called leptons. In contrast, all particles interact weakly[‡], where the force acts on a doublet of fermions. For quarks, the doublets are made up pairs of up-type quarks with down-type quarks while for the leptons, the doublets are made of a neutrino-charged lepton pair, where the doublets are represented each column of Figure 2.1. Quarks and leptons have three "generations" of these weak doublet pairs, but the primary difference between generations is the mass of the contained particles, later generations having larger masses. For quarks, the up-type quarks, going up in generation, are the up, charm, and top quarks; the down-type quarks are the down, strange, and bottom quarks; the charged leptons are the electron, muon and tau. Each neutrino has a corresponding lepton type, but are commonly not distinguished from one another, especially for experiments not sensitive to them. Unsurprisingly, because of the difference in mass for each generation, all normal matter is made of 1st generation particles, i.e. electrons and up and down quarks. As mentioned before, the fermions gain mass through spontaneous symmetry breaking of the Higgs mechanism. Each particle also has a quantum number associated with it called its flavor (e.g. the strange quark has a Strangeness quantum number) that is often used for creating approximate conservation laws that are useful for some calculation. For leptons interactions, the combined quantum number of the charge lepton plus neutrino pair is almost conserved, but neutrino oscillation breaks it. For quarks, their quantum numbers are conserved in all strong interactions, and

[†]When fields (in our case the weak and electromagnetic fields) interact with the potential field, one of them will be collinear with the VEV vector while the other will be normal. The field collinear will interact with the field quadratically as in the typical potential case (i.e. no mass), while the other fields interact along the flat, or brim, portion of the potential, gaining the mass term.

[‡]Up to handedness.

the combined weak doublet quantum number is almost conserved, but the weak force mixes doublets through the Cabibbo-Kobayashi–Maskawa, or CKM matrix, breaking the quark flavor and quark doublet conservation generally.

The interactions of the particles is fully described in the Standard Model Lagrangian. With that addition, the Standard Model is more-or-less described. More nuances in the theory (such as handedness, CP violation, etc.) exist, but this description gives a fairly holistic review.

2.2 The Top Quark

We are measuring the production of three top quarks, so it is most interesting to understand the top quark, its production, and its decay specifically. It is helpful to delve a little deeper into the top quarks properties and history.

The top quark ranks as one of the stranger pieces to the Standard Model. It is the heaviest elementary particle in the Standard Model by a wide margin. Coming in with a mass of roughly 175 GeV, the particle is 50 GeV heavier than the next most massive particle, the Higgs boson. Further, the mass of this quark is orders of magnitude larger than the other quarks. Looking at the masses of the quarks in each generation, we see the up and down quark masses are 2.2 MeV and 4.7 MeV, the charm and strange quarks masses are 1.28 GeV and 0.096 GeV, and the top and bottom masses are 173.1 GeV and 4.18 GeV. While there is a noticeable mass discrepancy in each generation, the top quark is nearly 50 times heavier than its partner, the bottom quark. This alone isn't a problem, but it is a queer fact that makes the top quark distinct among its peers.

The top quark's large mass has the added consequence that the top quark decays extremely quickly. Most quarks have a relatively low mass and will combine with each other to make hadrons or color neutral bound states held together by the strong force. Because the strong and electromagnetic forces conserve quark flavor, hadrons will typically decay by a weak interaction, specifically W boson decay.[†] The weak interaction is suppressed because most hadrons do not have the energy to create the W boson on-shell. Creating the off-shell W boson imparts a $1/m_{\{W}^2$ penalty to the decay amplitudes which leads to increases the decay time or, in other words, surprisingly long lifetimes. The top quark, oppositely, has enough mass to create an on-shell W boson, so the top quark can immediately decay resulting in a vanishingly small lifetime, estimated around 10^{-25} s. The consequence is there isn't enough time for the strong force to create a top quark bound state, so it does not hadronize. While this makes measurement more difficult, it does allow the top quark to act as a tool for studying properties of the Standard Model. Any measurement made with a top quark is a raw measurement and not filtered through the process of hadronization.

[†]If the hadron contains a quark-antiquark pair, pair annihilation via any of the forces is possible, but this is the exception rather than the rule.

Another quirk of the top quark is how it decays. As previously mentioned, the weak force allows for flavor changing of quarks between the particles in each weak doublet, but it can actually decay across generations. Because of differences in the weak force's quark eigenstates and the observed mass eigenstates, the weak force can violate conservation of weak double numbers. This means these eigenstates are mixed, and the amount of mixture between flavors, and thus the probability of flavor changes, is described in the CKM matrix[†] shown below:

$$V_{\text{CKM}} \equiv \begin{bmatrix} V_{\text{ud}} & V_{\text{us}} & V_{\text{ub}} \\ V_{\text{cd}} & V_{\text{cs}} & V_{\text{cb}} \\ V_{\text{td}} & V_{\text{ts}} & V_{\text{tb}} \end{bmatrix} \quad \rightarrow \quad |V_{\text{CKM}}| = \begin{bmatrix} 0.97435 & 0.22501 & 0.003732 \\ 0.22487 & 0.97349 & 0.04183 \\ 0.00858 & 0.04111 & 0.999118 \end{bmatrix}$$

Unsurprisingly, values on the diagonal of the CKM matrix (i.e. staying within the original weak doublet) are close to 1 meaning most weak decays happen between quarks of the same generation as is normally expected. For calculating the probability of a certain decay, the probability is proportional to the square of the CKM value, or $|V|^2$, so the $t \Rightarrow b$ transition happens for 99.8 % of the top decays, or top decays almost always decays to a bottom quark. This incidentally leads to bottom quarks having surprisingly long lifetime for their mass because available decays to lighter quarks are suppressed by the small CKM values, V_{cb} and V_{ub}.

2.2.1 History of the Top Quark

As an aside to the discussion of the top quark is the historical context of its initial theoretical proposal to its discovery. The top quark was theorized in the 1970s even though it was not discovered until 1995. In 1964, the first experimental result for CP[‡] violation was observed in kaon decays [16]. While the weak force breaks C and P symmetries individually, it was theorized that the combined CP symmetry was always preserved. In an experiment created by James Cronin and Val Fitch at Brookhaven, neutral kaons and their decays were studied. Neutral kaons come in two varieties, $K^0 = \bar{\text{s}}\text{d}$ and $\overline{K}^0 = \bar{\text{d}}\text{s}$, which also antiparticles of each other. From these two kaon states, one can create CP kaon eigenstates, namely

$$K_1 = \frac{1}{\sqrt{2}}(K^0 - \overline{K}^0)$$
$$K_2 = \frac{1}{\sqrt{2}}(K^0 + \overline{K}^0)$$

[†]These values are fit from multiple measurements to the expected Standard Model CKM matrix, with the most current fit being collated by the Particle Data Group [15], with individual sources for each analysis are found in Chapter 12 on their website.

[‡]CP here refers to a specific symmetry of a physical system. C represented the charge conjugation transformation which swaps particles for antiparticles and vice versa. P represents the parity transformation or inversion of all spacial coordinates. Before this experiment, it was theorized that all interactions were invariant under the combined CP transformation.

in which the K_1 decays primarily through $K_1 \rightarrow \pi^+\pi^-$ and $K_1 \rightarrow 2\pi^0$, while the K_2 decays by $K_2 \rightarrow \pi^+\pi^-\pi^0$, $K_2 \rightarrow 3\pi^0$, and $K_2 \rightarrow \pi^+\ell\nu$. The K_2 has a disadvantaged decay chains, where the leptonic decays are suppressed by the $\Delta I = 1/2$ rule and the three pion state, which requires the extra pion to conserve CP, leave only a slight energy excess for the decay which similarly suppresses its decay. Experimentally, this difference in neutral kaon decay time was observed, so the states were referred to as K-short (K_S) and K-long (K_L). If nature preserved CP symmetry absolutely, we would expect the observed K-short and K-long states to be identical to the CP eigenstates, or $K_S = K_1$ and $K_L = K_2$. In 1964, Fitch and Cronin found that, occasionally, K_L decayed through a two pion channel like K_S, meaning these kaons aren't pure CP eigenstates, but rather mixtures of K_1 and K_2. This proved that the weak force violates CP, even if just very slightly.

Theorists now had an interesting result that needed explanation. With an astounding amount of insight, Makoto Kobayashi and Toshidhide Maskawa wrote a paper in 1972 proposing that, with three generations of quarks, CP violation could naturally be explained [17]. Nearly a decade previous, Nicola Cabibbo proposed a mixing angle for weak interactions to explain strange quark decays (before the discovery of the charm quark) called the Cabibbo angle [18]. Kobayashi and Maskawa, in their paper, expanded the utility of the Cabibbo angle by proposing three generations of particles with a corresponding 3×3 quark mixing matrix. This matrix, now known as the CKM matrix after all three theorists, can be parameterized by three angles and a CP violated term. Kobayashi and Maskawa wrote their seminal paper before the discovery of the charm quark, so their idea required a doubling the number of quarks to explain this slight CP violation found in kaons. While the top quark wasn't named in the paper and was simply a construct, the charm quark discovery in 1974 [19, 20] and the bottom quark discovery in 1977 [21] made the top quark discovery all but inevitable.

To find this elusive quark, the Tevatron, or a proton-antiproton collider located at Fermilab, was conceived, constructed, and turned on by 1983. Around that time, the LEP accelerator at CERN found the W and Z boson, but also put loose limit on the top quark mass from indirect calculations and measurements [22]. Once the Tevatron started to take data, physicists began analyzing the data to find the elusive quark in $t\bar{t}$ (pronounced as "t-t-bar") events or events with a pair of top quarks because this is the most probable production mechanism that conserves charge and flavor. While $t\bar{t}$ is the most probable way to search for the top quark, the consequence is now the required production energy has doubling, and a smaller energy allowance generally decreases production rate making the search more difficult. The Tevatron, with its proton-antiproton beam, had the benefit of higher expected $t\bar{t}$ production, but it had to contend with the difficulties associated with producing and intensifying the antiproton beam. After years of acquired knowledge and technical improvements to the accelerator, the total energy and luminosity of Tevatron reached passed its designed quantities leading the DØ and CDF collaborations at Fermilab to announce the top quark discovery in 1995 [23, 24].

2.2.2 Production Mechanism

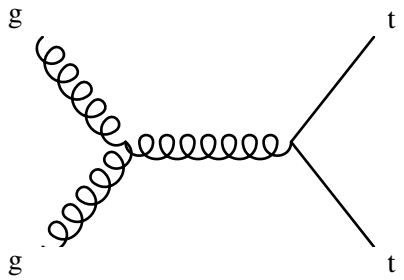

Figure 2.3: Example Feynman diagram for the production of $t\bar{t}$.

We are aiming to measure top quarks in a detector, so it is worth describing how top quarks are typically created. Because of the conservation of quark quantum number under strong interactions, the top quark is typically produced in pairs from a simple gluon interaction, as shown in Figure 2.3, and indeed, the Tevatron discovered the top quark from a $t\bar{t}$ pair. Expanding on this, we can investigate further by looking at processes that produce different numbers of top quarks.

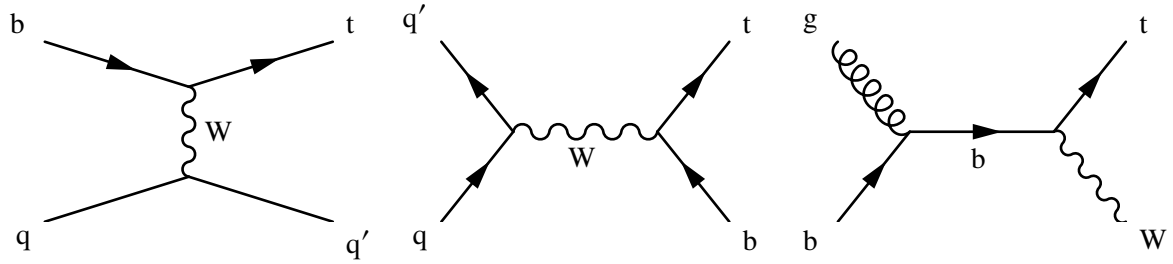

Figure 2.4: Example Feynman diagrams for the three production modes of single top quark: t-channel (left), s-channel (middle), and tW (right).

Single top production isn't excluded as might be suggested, but a weak interaction is necessary because the top quantum number has to be broken. Surprisingly the production rate for single top is about the same order of magnitude as $t\bar{t}$ production. While the weak decay penalizes the single top production, the smaller necessary mass leads to a smaller rate, but not significantly so. With single top, there are three main production mechanisms: s-channel, t-channel, and tW (with their Feynman diagrams shown in Figure 2.4). Here, tW refers to the final state of a top quark and W boson, s-channel to resonant-like decays, and t-channel to scattering-like decays (here the s- and t- come from Mandelstam variables which are ways contractions of the momentum variables and correspond to particular Feynman diagram [25]). For the s-channel diagram, it is obviously possible to produce in a proton-proton collider as necessary for this analysis, but the other production methods show a curious b quark in the initial state that is seemingly misplaced since the proton, presumably, only contains up and down quarks. Although protons are

indeed made up of two up quarks and one down quark, each hadron contains a sea of quarks and gluons that allow for more exotic interactions, all of which are described by the parton distribution function.

Parton Distribution Function

Quantum field theory calculations are in principle performed by calculating observables using the path-integral formulation, but in practice, the path integrals are impossible to evaluate for all but the simplest cases, so approximations must be made. The typical way is expanding the path-integral perturbatively in terms of the coupling constant, α, of the different forces where each term in the expansion corresponds to some Feynman diagram. While only an approximation, this perturbative method allows one to calculate to arbitrary precision. Even with very few orders of correction, one can reach higher precision than a currently measurable, most exemplified by the g-2 anomalous magnetic moment of the electron which has the 5^{th} order expansion experimentally validated to the billionth place [26]. The first terms in the expansion are often called tree-level or Leading Order (LO) to contrast higher order diagrams[†] that contain loops or divergences. These loops are of particular annoyance for calculation because they introduce free momenta that have to be integrated out.

These loop corrections are dependent on the energy scale of the interaction. Roughly, we expect interactions of higher energy scales to have larger contributions from these $\text{N}^{\text{n}}\text{LO}$ interactions because the larger energy mitigates the penalty from the loop interactions. But only the outgoing particles can be measured, so effects of the internal loop correction make the basic interaction itself appear stronger at higher energy levels[‡]. To express this idea, the interaction strength, which set by the coupling constant, α, is by convention allowed to vary by energy scale, or $\alpha \to \alpha(\mu)$, paradoxically making our coupling constant variable. This phenomenon of the coupling constant changing with energy scale is referred to as the running of the coupling. For an example, the coupling constant for electromagnetism is the fine-structure constant which classically has a value of $\alpha_{\text{FS}} \approx 1/137$, but at the electroweak energy scale, or $\mu_{\text{EWK}} = 90\,\text{GeV}$, we find $\alpha_{\text{FS}}(\mu_{\text{EWK}}) \approx 1/127$.

This property can lead to problems if the coupling constant becomes greater than 1 since our perturbative calculations will suddenly fail. For the electromagnetic and weak forces, our particle accelerators do not produce enough energy for this to be a problem, but for the strong force, things breaks down for different reasons. Opposite of the electroweak force, the strong force is asymptotically free, or the strong coupling constant decrease at energy scales, a fact was discovered by David Gross, Frank Wilczek, and David Politzer [27, 28]. For sufficient energies, the coupling constant falls below 1 and perturbation

[†]These next to leading order diagrams are abbreviated NLO, and subsequent orders of precision follow convention of NNLO, N^3LO, etc..

[‡]This is a simplification. In renormalization, there are bare quantities in the model that are divergent where renormalization removes these infinities by setting an energy scale and subtracting the divergent parts out. The renormalization group gives us a way to adjust the theory from our initial energy scale to another by way of the beta-function, which tells how much these renormalized quantities change with the energy scale.

theory works, but at the lower energy scales, our calculations would fail. In fact, because the coupling constant is large, the force increases at low energy scales meaning QCD bound states (i.e. hadrons) do not allow the isolation of a bare quark. In trying to remove a quark, the hadron fragments because the energy needed to separate the quark is greater than the energy required to form a new quark-antiquark pair thus creating new hadrons. This bound state property is called confinement.

Asymptotic freedom means hadrons, in their ground state, contain a mess of quarks and gluons binding it together. Using the language of Feynman diagrams, the N^nLO diagrams contribute more than the LO diagrams, so interactions containing multiple gluon and quark loops dominate. This also means interactions with hadrons can be quite complicated because a particle can interact with either the valence quarks, i.e. the primary quarks that make up the hadron, or the sea of quarks and gluons binding it together. To model any interaction with a hadron, one needs to calculate a probability for interaction with a valence quark or the gluons or virtual quarks hidden in the QCD mess. These probabilities are expected to depend on the energy scale of the interaction as well as the amount of momentum transferred to and from the hadron by the probing particle. The parton distribution function (PDF) is defined to quantify the relative probability of each interaction for different energy scales and momentum transfers, as seen in Figure 2.5.

The PDF has no theoretical derived equation that models it in part because of the non-perturbative nature of QCD, so models are analytically created to match experiment. These are called Deep Inelastic Scatter (DIS) experiments, and they involve hitting hadrons with a probe, such as an electron, to see the types of hadrons produced for different momentum transfers between the probe and hadron. From the resulting particles as well as their energy and momentum, we can construct our PDF best reproduce the relative abundance seen. This was first investigated in 1967 at SLAC, but further DIS experiments have been run to refine our models [30]. For single top production, the PDF is the mechanism that allows for the t-channel and tW production where the interaction is initiated by a b quark in the proton. Looking at the b quark PDF in Figure 2.5, these interactions are unlikely except for large energy scales and very inelastic collisions. The PDF shows there is a mechanism for producing single top quarks, but it is suppressed.

2.2.3 Three Top

With a clear understanding of the production mechanisms for $t\bar{t}$ and single top, processes with more top quarks are produced roughly through a combination of these two. For three top quarks, it is analogous to a $t\bar{t}$ and single top production, and for four top quarks, it is analogous to two $t\bar{t}$ productions. While single top production has a slightly smaller cross-section compared to $t\bar{t}$, three top and four top events have orders of magnitude smaller cross-section (where the concept of the cross-section is explained in full in Appendix B). This is a byproduct of the additional interaction vertices as well as the significant mass requirement. Even so, single top [31, 32], $t\bar{t}$ [23, 24], and recently, four top quark events have been

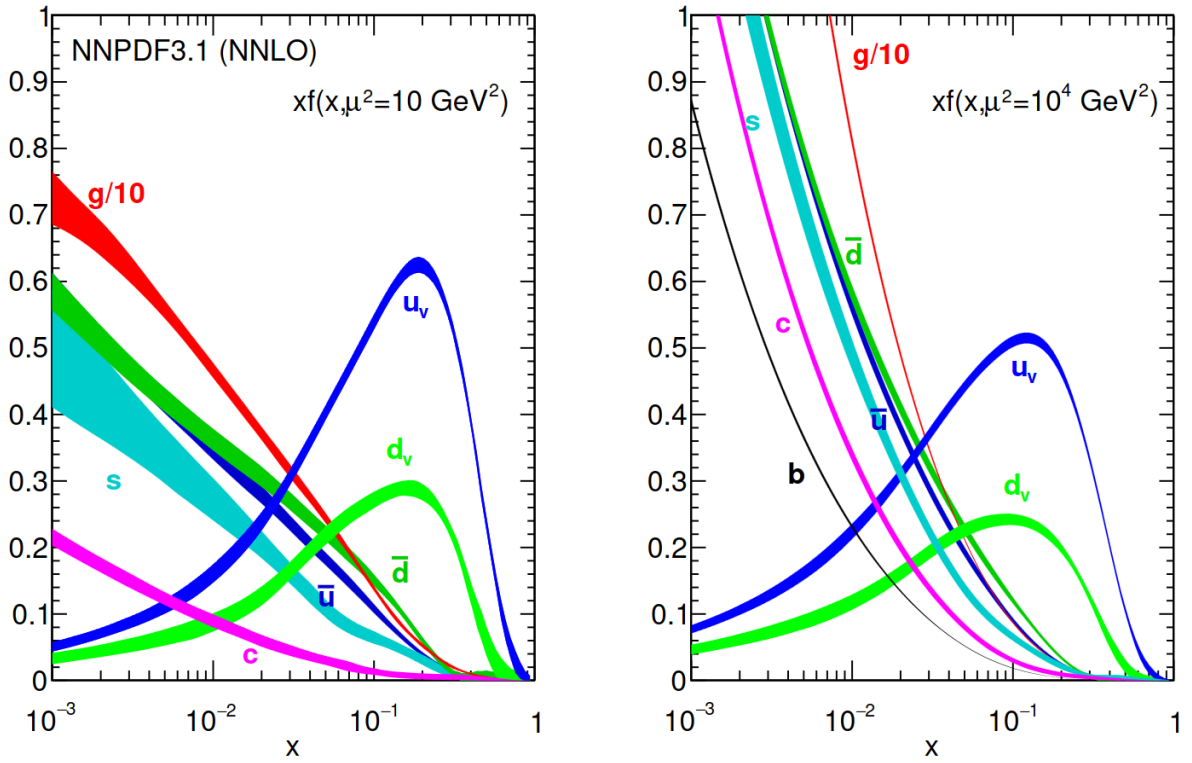

Figure 2.5: Example parton distribution functions (PDF) for a proton at different energy scales μ^2, left at $\mu = \sqrt{10}$ GeV and the right at $\mu = 100$ GeV, where 100 GeV is roughly the electroweak scale, or when electromagnetism and the weak force merge. Here, the x-axis is represented by the Bjorken variable x which is roughly the fraction of momentum transferred, higher value represented complete momentum transfer, i.e. elastic scattering, and lower value representing harder collision or inelastic collision where new hadrons are created. The y-axis, xf, is a scaled version of a PDF f. A different PDF exists for each quark flavor and gluon, as shown. [29]

observed [33, 34], but three top quarks have never been seen or even researched. The cross-sections for the four different top processes are shown in Table 2.1, and as can be seen, the three and four top processes are four to five orders of magnitude smaller than the $t\bar{t}$ and single top processes.

Because the three top quarks process is roughly a $t\bar{t}$ and single top process, its primarily decay modes corresponding to the three modes for single top decay: tttb, corresponding to s-channel, tttq corresponds to the t-channel decay, and tttW which corresponds to the tW decay. For simplicity, tttb and tttq are combined into a single tttj or three top plus quark final state. The Feynman diagrams of these different three top processes are shown in Figure 2.6 and are easily matched to its corresponding single top Feynman diagram from Figure 2.4. Looking at the overall cross-sections of these processes, we find that the tttW production mode has roughly twice the cross-section to that of tttj.

From these three *production* modes, we can investigate the different *decay* modes. The most obvious

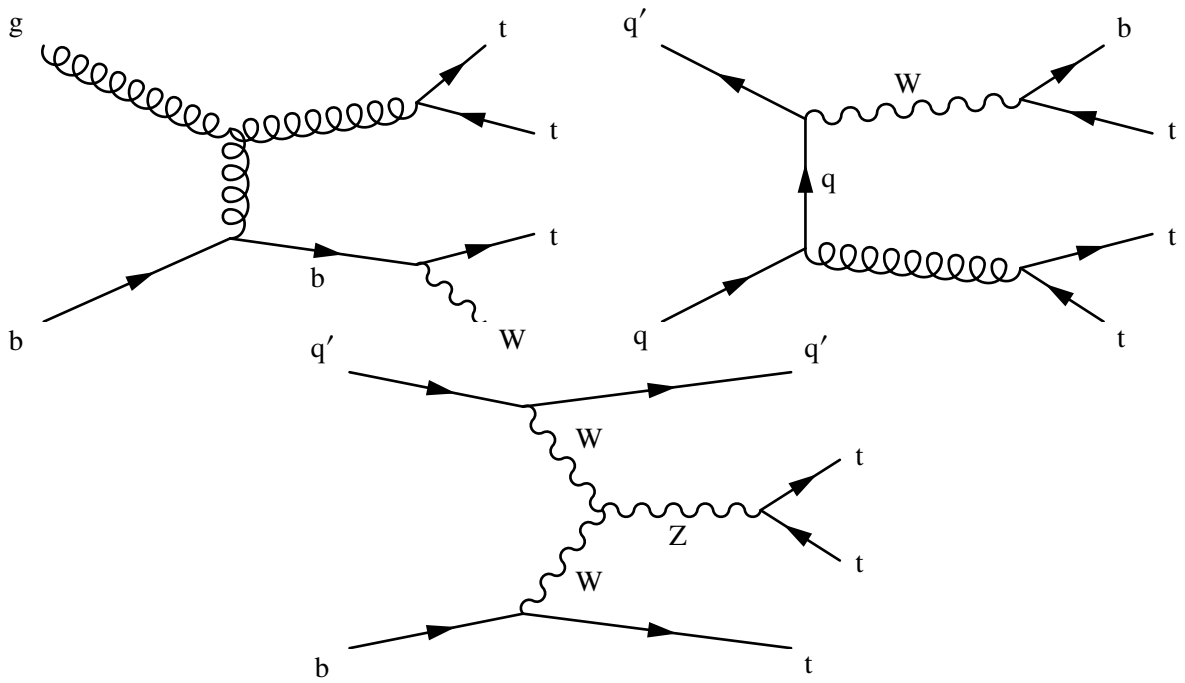

Figure 2.6: Example Feynman diagrams for the three production modes of three top quark: tttW (top left), tttb (top right), and tttq (bottom).

Table 2.1: Total cross-sections for the top quark production modes at 13 TeV for proton-proton interactions. These numbers are approximate, but give an idea of the abundance expected per process.

Process	Cross-section (fb)
$t\bar{t}$	830 000
Single Top	300 000
$t\bar{t}t\bar{t}$	13.0
tttW	1.3
tttj	0.7

way for classification is by number of leptons, especially since leptons are easily identified in a particle detector. Before, discussing further, it is useful to make a quick aside about what we mean by lepton in this classification.

What is Meant by Lepton?

There are three types of charged leptons (electrons, muons, and taus) and three types of neutrinos. Neutrinos, being chargeless, are difficult to detect, so they are considered lost energy for our investigation. Electrons and muons are easily detected because of their long lifetimes and particular signature in a detector. Taus, which have a mass of 1.7 GeV, decay quickly with a mean

lifetime of around 3×10^{-13} s. Even considering relativistic effects, this means most taus do not travel long enough distance to be directly measured in the detector before they decay, so taus are necessarily measured by their decay products. Taus decay into a W boson and a tau neutrino, and the W boson decays either into a quark pair, creating a hadron, or a lighter lepton-neutrino pair. The hadronic decay mode has particular properties that make it possible to identify as coming from a tau, but the leptonic decay channel has fewer distinguishing features so cannot be differentiated from other muons and electrons. Therefore, taus are only clearly identified through the hadronic decay channel. This analysis does not include hadronic taus, so, for a classification of different decay modes, leptons are in reality only a count of the number of muons and electrons in the event.

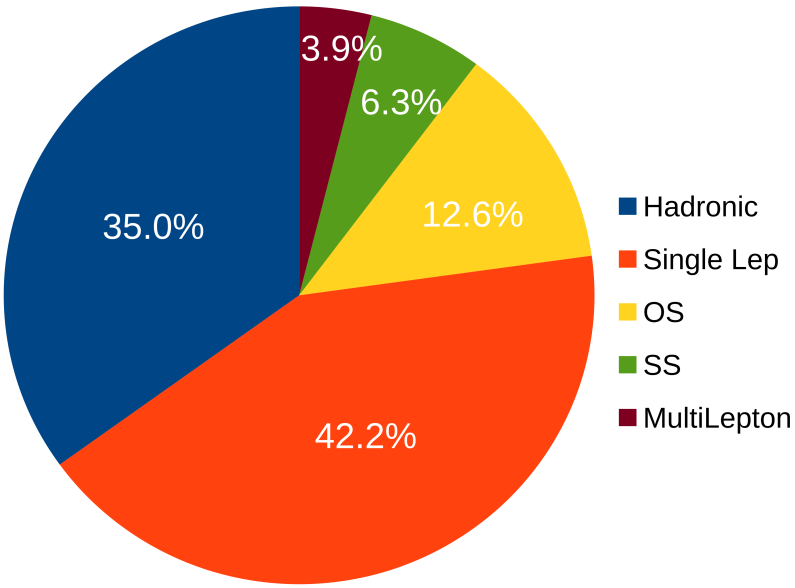

Figure 2.7: Branching fractions for different decay modes of three top quarks, classified based on number of leptons. Here, leptons means only light leptons, or electrons and muons, or taus that decay leptonically. Taus that decay hadronically are classified as hadrons for simplicity. As an additional note, while there are only three top quarks, the multilepton branches fraction includes decays with three and four leptons, the fourth lepton coming exclusively from the tttW production.

Top quarks, as mentioned, decay almost exclusively into a bottom quark and W boson. For discriminating channels by the number of leptons, we must simply count the number of leptonic decays of the W bosons. Noting relevant decay percentage, or branching fractions, for W bosons and tau decays[†] as well

[†]For W bosons, $B(\mathrm{W} \to \ell\nu) \approx 0.33$. A nuance is, for $\mathrm{t} \Rightarrow \mathrm{W} \Rightarrow \ell$, the W boson can produce electrons, muons, and taus, each at a rate of $0.33/3 \approx 0.11$. For $\tau \Rightarrow \mathrm{W} \Rightarrow \ell$, the branching fraction is the same, but the W boson produced from the tau

as combinatorics, we find the following results in Figure 2.7.

While the hadronic and single lepton final states dominate, making up roughly 77 % of the decays, they are more difficult to probe because of difficulties in reducing other, similar looking processes that have the same final state. But, particles are typically created in neutral pairs to conserve charge and other quantum numbers, so an effective way of reducing these other processes is requiring two leptons which have the same charge. For our signal, this takes the form of same-sign dilepton events as well as multilepton events (which naturally contain a same-sign pair). While these two production modes only make up roughly 10 % of the three top decays, this choice gives significant sensitivity. Using this approach is how we will be investigating the three top quark production.

2.2.4 Beyond Standard Model

While the Standard Model has been tested and confirmed in numerous experiments [35], there are still several open questions in particle physics that the Standard Model does not answer. The largest and most cited include the existence of dark matter and the hierarchy problem.

Astronomical experiments have observed anomalies in multiple places including in galactic rotation curves [36], observed gravitation lensing effects [37], and mass measurements of the collision of the Bullet cluster [38] that has collectively been hypothesized to be caused by what is called dark matter. While dark matter seems to exist in astronomical data, it has never been observed in a particle accelerator. All these tests come from observations of dark matter's gravitational effect, not from light which indicates that it does not interact electromagnetically. In accelerators, gravity is too weak to observe, but astronomical data has not ruled out dark matter's ability to interact weakly or strongly. If it did interact through the strong or weak force (or another unknown mechanism), we would expect it to be detectable in an accelerator such as by being produced in a collision as lost energy. Currently, dark matter is not included in the Standard Model, so there is a clear rift between astronomical and particle observations if dark matter is indeed a particle.

Another issue is the so-called hierarchy problem, or the large observed difference in strength between gravity and all the other forces [39]. Compared to the three forces described in the Standard Model, gravity is quite weak. If one measures a force's strength based on its coupling constant, the effective gravitational coupling constant is many orders of magnitude weaker than all the other forces. Not only this, but for gravity to unify with the other forces, *à la* electroweak unification, it is expected to happen at the energy scale of gravity, or at the Planck mass (around 10^{19} GeV). While this isn't a problem per se, there are undesirable consequences. Most famously, when calculating loop corrections to the Higgs propagator, one finds some of these corrections are divergent. To preserve the measured Higgs boson mass at the Planck scale, either these divergent terms need to be canceled through fine-tuning of various parameters in the model, or the Standard Model would require additional interaction terms to remove

only has the energy to decay into an electron or muon, leading to a rate of roughly $0.33/2 \approx 0.17$.

these divergences. While fine-tuning does not invalidate the Standard Model, it seems unlikely and suggests an incompleteness in the high energy regime.

Because of these issues, much of the work of current particle theorists is to postulate new particles outside the Standard Model while experimentalists probe for them in the hopes of finding new physics that could answer these questions. For this sort of physics, we call it Beyond Standard Model (BSM) physics. While the production of three top quarks is predicted by the Standard Model, there are BSM mechanisms that could cause an increase to its cross-section.

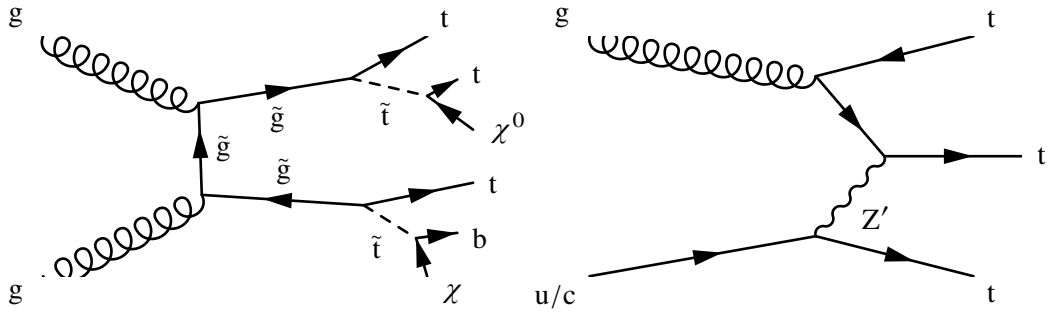

Figure 2.8: Example Feynman diagrams of possible BSM processes that could enhance the three top quark production cross-section. On the left is a SUSY model where a gluino ($\tilde{g}$, the gluon's supersymmetric partner) decays to a stop ($\tilde{t}$, the supersymmetric top) and a top, and the stop decays into either a chargino, $\chi^{\pm}$ and a bottom quark, or a neutralino, χ^0, and a top quark. The chargino and neutralino, here, are composite supersymmetric partners of the bosons. The diagram on the right is an example of a FVNC Z′ model where the Z′ changes the up-type quark into a top quark, therefore changing quark flavor. Both of these Feynman diagrams would cause enhancements to the overall three top cross-section.

Most BSM models that consider three top as an avenue for probing new physics consider the odd number of top quarks it produces. Producing three top quarks requires a weak decay and typically a bottom quark in the PDF which acts as a suppression for three top processes. Neutral interaction chains allow for flavor changing in the Standard Model, but they are heavily suppressed. For instance, a simple Z boson interaction does not allow for flavor changing, so flavor changing is only allowed through a loop diagram, but they are heavily inhibited by the Glashow–Iliopoulos–Maiani (GIM) mechanism which describes these flavor changing neutral interaction chains [40].

For this reason, three top BSM models suggest a new, flavor violating neutral current (FVNC) mechanism to boost the overall production rate. Since the three top quark process is quite small and has the possibility of a FVNC interaction, it is an attractive probe for testing this mechanism.

The most popular and widely studied BSM model is called Supersymmetry (SUSY), which postulates that there exists a symmetry between fermions and bosons and thus each particle has a supersymmetric partner thus doubling the current roaster of particles. [41]. The emergence of so many new particles in SUSY leads to innumerable variety in the theory that allows for changes in the three top production

cross-section through SUSY interactions. For our case, a FVNC process is possible through a cascade decay. If we allow stops (or the top quark's supersymmetric partner) to change quark flavor, we can create three top quark processes through decays of gluinos (or gluon supersymmetric partners) as shown in Figure 2.8.

While supersymmetry has decades of research, we can also achieve the same results with more generic and simple models. These generic models tend to be effective field theories, or simply, models that only work in specific, lower energy scales, as an approximation to a possibly more complete model. These effective field theories eschew theoretical completeness for producing a simplified model that is valid well up to some energy cutoff, Λ. Using effective field theories allows theorists to create interesting models that don't require the same theoretical rigor required for ultraviolet-safe theories that must be valid at energy scales past the cutoff, Λ.

Another, model that allows for FVNC is the Z' model (pronounced as "Z-prime models") which postulates a Z boson-like particle that can have different properties, such as flavor violation [42] where its Feynman diagram is shown in Figure 2.8. Z' models minimally extend the Standard Model with a Z boson like particle which becomes a common refrain in a variety of BSM theories such as with effective field theories.

These theories have been investigated in many papers such as those by Khanpour, Barger et al., and Cao et al. [43–45]. The models suggest extensions that increase the overall three top cross-section up to two orders of magnitude. While these new models would require new particles and interactions, the papers also suggest a limited change in kinematics meaning these models can be approximated by simple cross-section enhancements. This is in conjunction with the limited resolution of our measurement because of large uncertainties and low expected yields. This means, with our measurement, we can also reinterpret the data with each BSM model to either discover new physics or set limits on these models.

Chapter 3

Experimental Apparatus

All experiments are beholden to their experiment devices. With each generation of technology, the tools of the experimentalist expand, but this power is limited by the error associated with. This means experimentalists must understand both their device and the errors and uncertainties inherent in it to squeeze the most from every measurement. For fundamental particle physics research, the top of the line research is done at CERN, right outside of Geneva, Switzerland, using experimental devices that are not only the largest and most complicated of their class, but represent some of the most complicated engineering feats performed by man.

3.1 The Large Hadron Collider

At CERN, everything centers around the synchrotron particle collider called the Large Hadron Collider (LHC), measuring roughly 27 km in circumference, bursting across the Swiss boarder into France, which is used for colliding protons against protons for particle physics experiments. The accelerator itself is located roughly 100 m underground, using the 1980s tunnels dug for the Large Electron–Positron Collider or LEP. While the LEP collided electron-positron bunches, the LHC collides the heavier protons.

The particle beam itself is created from a supply of hydrogen which is ionized and then accelerated through a chain of steps shown in Figure 3.1 to progressively add energy. First, the protons are accelerated through one of the LINACs, or linear accelerators, then into the Proton Synchrotron Booster, then the Proton Synchrotron (PS), then the Super Proton Synchrotron, and finally the LHC itself. The LINAC raises the initial proton to 50 MeV and then raised to 1.4 GeV, 25 GeV, and 450 GeV by each step in the accelerator chain before reaching the final design energy of 7 TeV per beam of the LHC. This whole accelerator complex also gives a glimpse of CERN history, the PS being the first synchrotron to be built at CERN in 1959, the SPS being started in 1976, and the LEP tunnel, as mentioned before, laying the ground for the eventual LHC.

The LHC itself is primarily made of 1232 dipole magnets which apply the turning force on the beam, 392 quadrupole magnets which squeeze the beam to focus and reduce diffusive effects, and 16 RF cavities

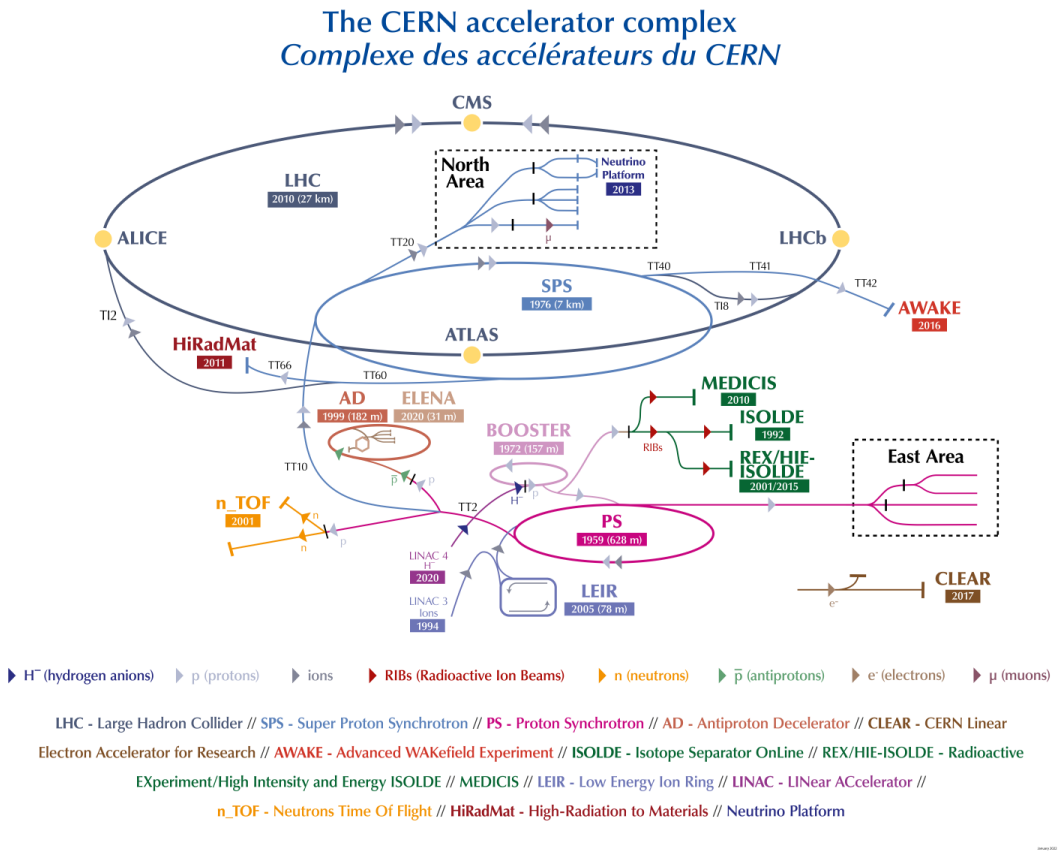

Figure 3.1: Representation of the full LHC accelerator complex. The date of the first usage of each accelerator is shown, giving a small history of CERN. Currently, the accelerator chain for LHC injection starts at the LINAC 4, proceeds to the Booster, PS, SPS, and finally the LHC. [46]

which accelerate the beam. Through the whole acceleration train, the protons are grouped into packets called bunches, each containing roughly 10^{11} protons. For filling the LHC, each bunch is injected into the LHC at specific times, with minimum spacing of 25 ns until the ring has 2835 bunches, or roughly 80 % filled, where the gaps come from general injection limitations and to leave space to dump the beam in cases of emergency. The 25 ns spacing means there is an expected particle collision rate of 40 MHz (ignoring the bunch gaps), which, including quantities such as beam geometry, leads to an instantaneous luminosity of roughly 10^{34} cm^{-2} s^{-1}.

The LHC takes data in three or more year campaigns called Runs where the collider operate roughly continuously from March to December. These Runs are interleaved with two years of downtime for upgrades and maintenance. The first run, or Run 1, stared in 2010 and concluded in 2012, Run 2 lasted from 2015 to 2018, and Run 3 started in 2022 and is planned to finish mid-2026. As the understanding of the accelerator has increased, it has been run more efficiently and has push towards its theoretical limits,

which is apparent from the overall increase of total luminosity year-on-year. For this analysis, we use Run 2 data, or data collected from 2016 through 2018[†].

There are several experiments on the LHC ring, the largest being: LHCb, a forward detector aimed at studying hadrons; ALICE, a detector specializing in heavy ion experiments; and CMS and ATLAS, the two "general purpose" detectors. For this analysis, data was taken on the CMS detector.

3.2 CMS Experiment

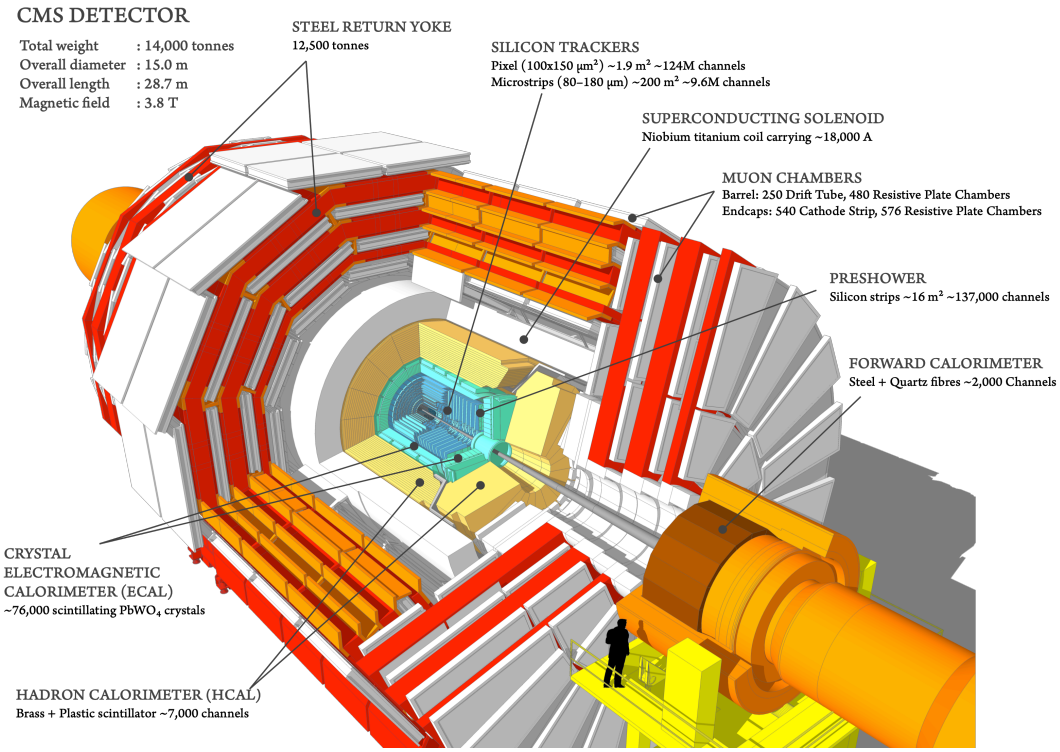

Figure 3.2: The CMS detector here is shown in full to give an idea of its size and basic construction. The detector is made of a barrel in the center and the two endcaps which fit snugly to fully contain all the particles coming from the main collision. [47]

CMS, or the Compact Muon Solenoid, which is schematically shown in Figure 3.2, gets its name from its large muon system (making up roughly half of its radius), its large superconducting magnet, and its size, being half the radius yet double the weight of the other general purpose detector, ATLAS. Most recorded collisions come from the proton-proton beam with analyses covering numerous topics like BSM searches, precision measurements, and even hadronic related experiments. The detector is

[†]2015 data is often omitted. This is partially because of the low luminosity collected as well as the issues that arose from collecting the first events at the higher 13 TeV energy after the addition of many upgrades.

also used during heavy ion collisions, which are performed typically at the end of each data-taking year. Because of the wide assortment of needs from the detector, its design was made to be flexible and robust, hence "general purpose".

As is typical with most modern particle detectors, the main components are a magnet, tracker, and calorimeters. The magnet bends charge particles in the detector. The tracker measures the now curved paths of the charged particles to measure their momentum. The calorimeter measures the energy of the particles, primarily by slowing them down so they interact and decay, so their total energy can be collected and measured. Because of the different decay properties of leptons and hadrons, the calorimeter is split into an electromagnetic and hadronic component, each specializing in measuring their respective particles. While these components can detect and measure most particles, muons act as an outlier. Their mass and interaction strength puts them into a "Goldilock" zone where they are light enough to have a long lifetime but heavy enough to punch through most of the detector without stopping. To remedy this, a dedicated muon detector is included.

An extra design issue is hermeticity. For a particle detector to be effective, it must capture all the particles created in the interaction, so everything can be reconstructed. This means the most optimal detector would be spherical to cover the full angular range, but this optimal, spherical detector would be difficult to fabricate. CMS was conceived first and foremost to discover the Higgs boson, so special design considerations were made with this purpose. When looking at simulated Higgs boson decays, it creates sprays of particles that chiefly fly perpendicular to the beam axis, meaning the detector should be optimized to measure particles in the transverse plane. Because of this, CMS is constructed as a cylinder, or barrel, around the beam with two end caps plugging the cylinder to keep the detector as hermetic as possible while prioritizing detection of the particles of most interest.

This describes the whole of CMS, but it is worth going into details of each component to explain how each part of the detector turns electronics and signals into particles.

3.2.1 Silicon Tracker

Trackers are used for reconstructing the path of the particle in the detector. With the path, which is bent by the magnetic field, the momentum of the particle can be inferred by the radius of curvature. Because the tracker is trying to reconstruct the path, the measurement process needs to be as non-invasive as possible to avoid interfering with the particle's momentum. To achieve this, trackers are often made of an easily ionizable material that doesn't sap much energy away from the incoming particles, while creating the measured current via the stripped electrons. This leads to two main classes of trackers: silicon based and gaseous based. The gaseous type benefits from being cheap and easily scalable while silicon based systems have high resolution and radiation hardness but are very expensive. For this reason, CMS is built with silicon based trackers at its core for the highest possible resolution while the outer tracking muon systems are made of gaseous tracking systems because lower resolution needs and the sheer space

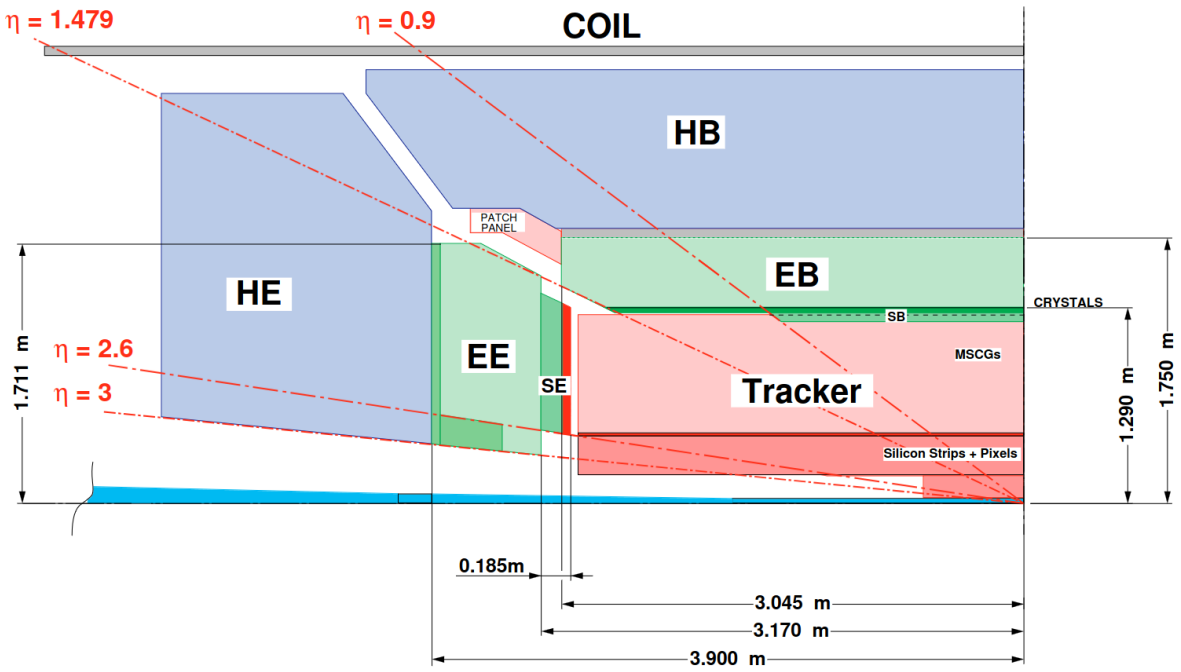

Figure 3.3: This shows a schematic of the inner part of the CMS detector, or the subdetectors located inside the magnet, in this case, the silicon tracker (in red), the electromagnetic calorimeter (in green), and the hadronic calorimeter (in blue). The naming convention here is a letter for the subdetector (e.g. H for hadronic calorimeter) and a B or E for barrel or endcap respectively. As an addendum, this diagram is an early configuration, so in the tracker, MSCG (a type of detector) was replaced with full silicon detectors, and the SB or barrel preshower was not installed. [48]

needing coverage.

Resolution demands decrease as the particle travels outwards because of geometric realities. Curvature measurements depend on the angular deviation from a straight path at a given point, but for the same angular displacement, the transverse distance increases with the radius. This means the error in the curvature and thus momentum is proportional to the inverse of the radius or $\sigma_p \propto 1/r$. To achieve a similar level of uncertainty in all parts of the detector, higher resolution is needed closer to the beam while one can get away with lower resolution farther away. The cost of a silicon detector tells a similar story: the more resolution, the more expensive. The physics and finances lead to the same conclusion: the highest resolution detectors should be placed in the core and coarser detectors farther out. While it might seem sensible to place the trackers as far from the beam axis as possible so less precision is needed, many of the particles created in the main interaction decay quickly, so to best understand the physics, tracking needs to start as close to the beam axis as possible. Therefore, the tracker is composed of two parts: the pixel inner tracker and silicon strip detectors surrounding it.

The pixel silicon detector, whose schematic is shown in Figure 3.3, is currently made of four layers in

the barrel and three layers in each of the endcaps. [49] The pixel detector, as the name suggests, acts like a camera with millions of tiny "pixels," each doped so it can easily be ionized by passing charge particles. The silicon is partitioned into 100 μm × 150 μm pixels that individually read out charge. The first layer of pixel detectors sits 3 cm from the beam axis while the fourth 16 cm from the beam axis.

Further out from the pixel detector are the silicon strip detectors, which are grouped into the inner barrel, the outer barrel, the inner endcap, and the outer endcap regions. Each subsection contains a variety of strip trackers and layerings (i.e. single vs double-sided) to accommodate the needs for the specific region. Combined, there are a total of ten barrel layers and nine endcap layers. The strip detectors contain doped silicon that charge particles can easily ionize, but unlike the pixel detector, the readout is done on long strips instead of individual pixels. While this might seem to lose a whole axis of resolution, the readout time recovers the long axis positioning. After the electron-hole pair are created, they drift towards to two ends of the strip, so the difference in readout timing can be used to infer their location using a simple estimation using the drift velocity. Further, the strips detectors are cross-hatched to reproduce a full 2D location, even without this drift time trick. While resolution for the pixel trackers is greater, the strip detectors compensate it at larger radii, so the combined silicon detector can resolve very high momentum objects with high efficiencies.

Calorimeter

A calorimeter is simply a device that measures or estimates the total energy of particles. In particle physics, this is achieved by combining two main technologies: a stopping material and a scintillating material. The first problem of calorimetry is collecting all the energy. To completely collect the energy of a particle, it must be slowed and eventually stopped within the detector else energy will be lost and thus unmeasured. For the detector to stay hermetic, this necessitates a stopping material to induce decays that will halt and fully contain the particles. For measuring the stopping potential of a material, we use a measure called the interaction length or the characteristic distance a particle travels on average before interacting with the medium. While the stopping material contains the decay, we need a way of actual measuring of the energy, so we use scintillators to achieve this. Scintillators are materials that absorb energy and release light in predictable frequency bands similar to spectral lines. By measuring the intensity of the light released by the scintillator, the energy deposited can be calculated. While other types of materials can achieve this goal, scintillators are the most common with modern calorimeters. The total energy of a particle is thus estimated from the scintillator measurement, compensating for any energy lost in the stopping material. From these two parts, we have created our calorimeter.

3.2.2 Electromagnetic Calorimeter

The electromagnetic calorimeter, or ECAL, is a detector primarily used for measuring the energy of electrons and photons. Prompt electrons coming from a high energy collision lose their energy primarily

by bremsstrahlung, or the radiation of photons, while photons lose most of their energy through pair-production. Each electron and photon decays into two daughter particles, so both will take, on average, half of the total energy. This means we can easily model the number of particles at each decay step by an exponential and, by corollary, the average energy per particle by a negative exponential. More explicitly, for each step n, the mean energy of each electron and photon can be modeled as:

$$E_n = \frac{E_0}{2^n}$$

After a sufficient number of steps, each particle will be in the energy regime where they interact primarily through ionization of the surrounding material until they totally stop. In fact, it is only after the particle is at a low enough energy that the stopping force increases causing the particle to abruptly lose its energy and stop, reproducing the characteristic Bragg's peak. This inflection point where the mode of decay changes happens in the MeV range. For an expected 100 GeV electron, this means we'd expect roughly 16 decay steps before the electron has $\sim$ 1 MeV of energy and transitions to an ionizing particle. To describe this distance, we use the radiation length, X_0, or mean length before an electromagnetic interaction. To ensure our detector is hermetic, our calculation gives a lower bound for the length of our ECAL at $16X_0$ (for 100 GeV electrons and photons at least), but this estimation gives the average distance for decay to transition from radiative to ionizing and does not include the distance traveled while ionizing. In practice, calorimeters are made much longer to guaranty capture of even the highest energy electrons and photons expected in any LHC collision.

For choosing the stopper, it helps to use a material with a short X_0 to save on space and potential cost. Bremsstrahlung and pair production are both processes which cannot happen in vacuum and need an observing particle or field to initiate meaning the interaction length is affected by the atomic number, Z, of the material. This means higher Z materials have shorter interaction lengths. CMS chose lead-tungstate ($PbWO_4$) crystal for its small interaction length ($X_0 = 0.89\,\text{cm}$), as well as its scintillating properties. Lead-tungstate is a transparent crystal, so all the scintillated light can be read out by a photomultiplier at the end of the crystal. This means the crystals act as both the stopping and scintillating material, so there is none of the usual energy loses in the stopping material itself making it a simple and compact device.

The ECAL is separated into a barrel and endcap portion, each using crystals of dimension 2.18 cm × 2.18 cm × 23 cm. The 23 cm corresponds to roughly $26X_0$ of length and the cross-sectional size means the angular resolution is a $\sigma_\phi = 0.0175$ [48]. The lead tungstate crystals have to be grown meticulously over a long time to ensure high optical qualities with little defects or the scintillation and transmission of light will be less effective. This equated to years of growth to produce the tens of thousands of crystals currently in CMS now.

3.2.3 Hadronic Calorimeter

Hadrons react differently in matter compared to electrons which explains the split between the hadronic calorimeter (HCAL) and the ECAL. Hadrons, being QCD bounded objects, can interact via the strong interaction with the nuclei of the material they are traveling through allowing for a different type of interaction. When high energy hadrons interact strongly, the decay is much more fragmented, typically creating a conical spray of new hadrons through spallation. This spray is aptly called jets[†]. Because hadrons fragment into multiple particles when they decay compared to electronic decays which produce only two byproducts, the stopping power is much greater, requiring far fewer interaction lengths for stopping where this stopping measure is called the nuclear interaction length, λ_0.

The difference between λ_0 and X_0 plays an interesting role in HCAL construction. The nuclear interaction length tends to be an order of magnitude larger than the electromagnetic interaction length and does not have an empirical formula like X_0, but, as a rule of thumb, denser materials equate to smaller λ_0. This means an HCAL needs to be thick and dense, so cheap, thick metal works just as well as the fancier lead-tungstate in the ECAL. Since metal doesn't have inbuilt scintillation, dedicated scintillators are sandwiched between the layers of metal making it a sampling calorimeter, or a detector that only takes measurements at specific spacial intervals.

HCALs must also contend with added complexity in their energy detection and estimation. In fact, when considering using a sample calorimeter, the incurred loses from energy deposits in the stopping material seem concerning, but after compensating through estimations, the error in the calculated energy is not much more significant than other inherent sources. The largest effect comes from energy imparted to the nuclei of the stopping material as hadrons interacts; this absorbed energy is non-trivial and cannot be measured. Similarly, neutrons, which have a lifetime of nearly 15 minutes, can be created in the decay chain, but they survive longer than the electronics in the HCAL process a given event so this energy to be lost. Other complications exist meaning our jet energy measurement has inherent energy loses which must be properly estimated to reconstruct the total energy. Any jet measurement tool will have losses that must be accounted for, so the sampling calorimeter does not hurt the efficiency while being a cheaper alternative to something like the ECAL's crystal construction.

For the CMS HCAL, it is divided into a barrel, endcap, forward, and outer region [50]. The barrel and endcap regions are similar in range to the ECAL and tracker, but the forward region adds detecting material past the endcap for measuring very forward jets while the outer region sits uniquely outside the magnet in the barrel region. The barrel, endcap, and outer regions use copper alloy and steel for the absorbers and plastic scintillators while the forward region, because of the larger flux and Cherenkov radiation (which can be measured), uses a copper-quartz matrix. The size of each scintillating readout corresponds to $\sigma_\phi = 0.085$.

[†]This nomenclature is extended to describe prompt hadrons and their byproducts found at the LHC.

3.2.4 The Magnet

The magnet, while not a detecting element of CMS, is at the heart of its design. The chief constraints to consider with any magnet is its strength and size. The strength is an obvious consideration since it determines to what extent it bends charged particles, allowing the momentum and charge to be measured from the curvature of its track. The stronger the magnet, the sharper particles will be bent in the detector. This is important because as a particle's momentum increases, the deflection in the magnetic field decreases, as can be seen from the elementary radius of curvature calculation for a constant magnetic field B:

$$r = \frac{p}{qB}$$

When the geometry of the path becomes flatter, i.e. the deflection smaller, the overall error of the momentum measurement will increase. To accurately measure high momentum particles, either one needs a stronger magnet or a denser tracker module.

The second large consideration is the magnet's size. The magnet is block of dense metal which absorbs particles so that energy is lost. Unlike the HCAL, which contains scintillators, the magnet does not allow for the similar interleaving of scintillating material for reading out energy, thus the magnet must be accommodated for any energy calculations. CMS solves this by putting the magnet outside the detecting material. For a cylindrical magnet, as used in CMS, the field is strongest and most uniform directly in middle of the cylinder, further incentivizing having the main detecting components completely inside the magnet. The difficulty is reaching large field strengths with a large radius.

To solve these issues, the tracker and calorimeters were made to be as small as possible while keeping hermeticity of detection for all particles (save muons); this is what leads to the compactness of CMS. More concretely, CMS uses a superconducting magnet with an operating field strength of 3.8 T, diameter of 5.9 m, and length of 13 m [51].

While the magnet field is strongest in the middle, the magnet is large enough that the residual field outside the cylinder has a substantial strength of 2 T. This outside field is used in conjunction with the muon system to perform additional tracking. The iron structural support of the muon system, called the return yoke, does not stop the field strength and thus adds a geometric complicity to the field lines, but it extends them to better concentrate the field in the actual muon detectors.

3.2.5 Muon Detector

The muon detector is the outermost portion of the detector and is located totally outside the magnet. The ECAL and HCAL are constructed to stop all particles, but muons are the exception. Muons are heavy enough to lose most of their energy through ionization rather than radiative decays as with electrons, but this stopping power is too weak to stop typically produced muons in the space of the CMS detector. The muon systems don't even attempt to stop the muon but rather just track them. By construction, only

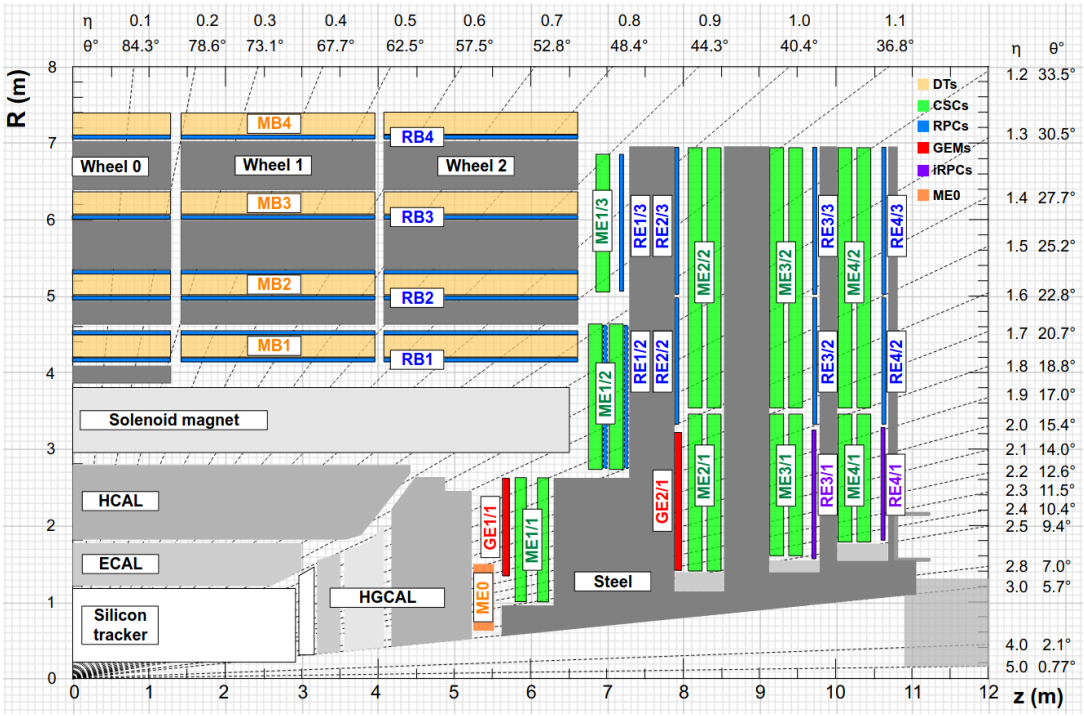

Figure 3.4: This figure shows a schematic layout of the muon detector, showing the placements of the currently installed subsystems. Here, the subsystems are labeled first by a letter denoting the detector type (e.g. G for GEMs), the second letter denoting barrel or endcap (B or E respectively), and then a number denoting the which layer the detector is on, the numbers increasing the further out. The endcap detectors have an additional number to represent the radial layer within each layer of the wheel. Here, the diagram shows both the installed GE1/1 (or innermost GEM detector) while the other GEM detectors, the GE2/1 and ME0, are slated to be installed during the next long shutdown the of the LHC. [52]

muons are expected to be found in the outer detector, so the mass (and thus the energy) can be inferred, removing the need for calorimetry. This high purity and additional detector information makes muon identification and measurement excellent.

The tracker is also useful for filtering out extraneous particles that fly through the muon detector. Most particles created in the detector are light enough to be stopped or heavy enough to decay quickly, but neutrons, have abnormally long decay times because neutrons have a mass only slightly above that of a proton (only about 1 MeV more), thus highly suppressing their decay. After only a few minutes of LHC collisions, neutrons populate the detector hall, creating a fairly homogeneous cloud. Once these neutrons decay, the proton and electron decay byproducts interact with the muon detector, but accurate reconstruction can remove these spurious tracks by discarding paths that do not originate from the beam axis.

The muon system is the most diverse subdetector in CMS, housing four different technologies of gaseous tracking detectors in the full suite: Drift Tubes (DT), Resistive Plate Chambers (RPC), Cathode

Strip Chambers (CSC), and Gas Electron Multipliers (GEM) [53]. For the central region, DT and RPC are use, while for the endcap, CSC, RPC, and GEM. The details of their placements are shown in Figure 3.4. All the gaseous detectors work on the same premise: a voltage is applied across a gaseous medium that ionizes when a charge particle passes through. They are distinguished by the different geometries used to induce the electromagnetic avalanche that amplified the signal enough to be read out.

The DT are made of multiple tubes with a high voltage anode wire running through the middle. The cylindrical structure of the electric field, with the cathode on the cylinders edge, induces the avalanche for signal amplification. The drift time can be used to get a more precise position for where the particle crossed, but no position information can be found along the DT's major axis. To combat this, individual tubes are place cross-hatched in multiple layers to provide full tracking.

The CSC, similar in design to the DT, are made of multiple anode wires arrange in a large chamber in rows with cathode plates segmented into strips, perpendicularly set to the wires. This structure is repeated in multiple layers. Given the wires and strips are perpendicular to each other, the detector achieves the 2D resolution without the need for multiple tubes. This, combined with the possible fan-like geometry for the construction, allows for better resolution in the endcap as well as fast response time even in high luminosity regions, making the CSC the main detector in the endcap region.

The RPC in contrast, does not use wires, but two sheets of Bakelite separated by a thin layer of gas. The two plastic layers are raised to a high voltage by metal cathodes and anodes outside the plates to induce the avalanche in the gas layer. The signal is then read out through the highly resistive plastic layer. The advantages of RPC are their fast response time compared to the wired gaseous detectors as well as the resistive plate isolating any charge saturation meaning more events can be readout for a given time period. This makes the RPC well suited for triggering (see Section 3.2.6) and are used in conjunction with the DT and CSC for a robust measuring.

The GEM detectors used a different geometry still. The GEM detector uses three foils set to increasingly larger voltages which are punched with microscopic holes in a hexagonal pattern [52]. These holes act as the location for the electron avalanche for signal amplification. The readout is done on the back-plate through different channels to gain x-y accuracy of the track placement. The main benefits of the GEM detector are their size and radiation hardness. For future, high luminosity runs, there is a worry that the wired detector will degrade or even stop working with the increased radiation. The foil of the GEM, in contrast, is much more robust against pitting or abrasion cause by radiation, so they become a perfect redundancy for future high luminosity runs while adding addition resolution and triggering possibilities. The detector is also very thin because of the foil construction. The iron return yoke has some additional unfilled space, but not much, so the svelte profile of the GEM allows it to be slotted into the available gaps in the endcaps' very forward regions. The GEM detector is the newest detector in the muon system, added in Run 3.

3.2.6 Trigger

While all the parts of the detector mentioned previously are integral for measuring particles, the LHC has a bunch crossing every 25 ns, equating to a rate of roughly 40 MHz. Even if the data saved per collision was made impractically small (say 1 kB), after one second, there would be 40 GB of data, and a whole days worth of data taking would amount to over a petabyte of data. Using this conservative estimate, the needed disk space would outstrip what even the most sophisticated data-centers could handle, not to mention most I/O electronics cannot reach the necessary throughput. To top this off, most of the data would be fairly useless containing events with little interesting physics or possibly no hard collision at all. This suggests some filtering is necessary to isolate only the intriguing events and to discard the rest, a process called triggering.

To define what events are worth saving, we need to explain what is "interesting" physics. The *raison d'être* of the CMS was to discover the Higgs boson, so it was designed with Higgs boson final states in mind, namely $H \to ZZ \to 4\ell$ and $H \to \gamma\gamma$. Obviously, this suggests a large focus should be on lepton and photon detection. Additionally, other colliders have probed the Standard Model before, but primarily at lower energy levels. For example, the Tevatron probed collision energies of 2 TeV, or one order of magnitude less than the LHC. Any event that creates high energy particles, especially at energies larger than what could be found at other experiments, is potentially worth investigating. From this, the main triggers, or thresholds, require events to contain: high energy jets, a lepton or photon; multiple leptons, jets, and/or photons; or large MET (a concept touched on in Section 4.2.4). For implementing the trigger, CMS uses a two-step system, the first is performed by hardware called the Level-1 (L1) trigger and the second by software called the High Level Trigger (HLT).

For the L1 trigger, the electronics are spread across the different subsystems, creating both quick reconstruction of particles and decision-making if the event passes any trigger threshold. While track reconstruction could be implemented in hardware today, during Run 2, the installed trigger was not capable of this in the timescales needed, so triggering is only utilizes information from the muon system and calorimeters. The calorimeter trigger performs a quick energy reconstruction by doing basic clustering of hits found to divide the energy deposits into different particles. Using locations of energy deposits (e.g. percentage of energy in ECAL vs HCAL), the calorimeter trigger can create a basic identification of the objects and sorts them based on quality. The muon trigger works first individually between each of the subsystems, creating muon candidates and then merges them to create a sorted list based on momentum as well as concordance between the subsystems. The calorimeter and muon trigger information is then shipped to the global trigger which combines it all to make a final decision on which event should be kept. This process takes place every bunch crossing with the caveat that the trigger decision isn't made before the next collision. While a collision happens every 25 ns, the particles generally take more time to propagate and decay in the detector, so, at any given time, there are multiple particles worth of collisions present in the detector, each radiating outwards like ripples in a pond. So, the information has

to be buffered to collect the full event. By the constraints of the hardware, the 40 MHz input is reduced to at most 100 kHz.

Next, the HLT performs a more robust reconstruction to make further trigger evaluations. The HLT uses only software, so it does not have the same tight time constraint as the L1 trigger, but it is still limited by the buffered data space as well as limits on how fast it can write out events. For the HLT, it aims to reduce the accepted rate to 100 Hz to 1000 Hz. To save computation resources, the reconstruction is done in stages, first filtering out events using similar information used by the L1 trigger, later incorporating finer granularity from the muon system and calorimeters, and lastly integrating the tracker information. This staged process ensures that bad events are removed before more costly calculations are performed. Since the HLT is a software based trigger, there is also flexibility with the triggers it can consider per run, and thus, year to year, the list of HLT triggers can change to meet the needs of analyses, to address efficiency problems, or to adjust data acceptance. Once the HLT has made its decision, the detector information is saved to disk where offline algorithms can be run to produce the data a physicist will use.

3.2.7 Luminosity

A last, necessary portion of the detector is the luminosity measurement tools. While the previously mentioned subdetectors collect event quantities, this information is useless without the global luminosity. Namely, we have no way of determining the expected number of events for our simulation, and thus cannot perform our statistical tests

From a basic probability standpoint, calculating the expected number of events of a given process is simple. Given a probability, P, the total expected yield for a process is $N_{\text{expected}} = N_{\text{total}} P$. The cross-section σ, as mentioned in Appendix B, is the effective area of collision and is used in place of the production probability. This means, in turn, the number of collisions is measured by the luminosity of our beam, $\mathscr{L}_{\text{tot}}$, leading to the equation:

$$N = \mathscr{L}_{\text{tot}} \sigma$$

Similarly, for finding the total rate of an event production, we can look at the instantaneous luminosity[†]. From this equation, we can calculate the luminosity from the event rate, which our luminosity detectors measure, and the cross-section. In this case, the cross-section isn't for any particular process, but the total rate through our luminosity detectors because we are trying to calculate the cross-section associated with the whole beam. Because of this, it helps to measure the luminosity in the very forward region where our primary, inelastic interaction do not bias the calculation which should come from common and frequent supervising collisions.

[†]The instantaneous luminosity, which is the number of events per area per second relates to the total luminosity by a simple integral $\mathscr{L}_{\text{tot}} = \int \frac{d\mathscr{L}}{dt} dt$ where the instantaneous luminosity is the derivative of regular luminosity.

For the rate measurement tools, thankfully, many of the previously mentioned detectors can pull double duty, acting as a luminosity measurement tool in addition to its normal function. The pixel tracker and very forward HCAL act as duplicate instrument for measuring the flux and getting a basic luminosity measurement. Further, there are specialized luminosity detectors installed in CMS such as the pixel luminosity telescope (PLT). The PLT is located inside the main tracker module at an $|\eta| \approx 4.2$[†] and is made of three rings of silicon detectors that look for coincidence hits through each ring that are produced in the very forward region.

For finding this total cross-section visible through the luminosity detectors, it can be derived completely from beam parameters, so dedicated tests called Van der Meer scans are run periodically for measuring the necessary parameters, scanning the beams against each other and measuring the change in observed rates as the beams change position, thus creating a model of the beam's shape. From these Van der Meer scans, the luminosity for all three years was found to be $35.9\,\text{fb}^{-1}$ [54], $41.5\,\text{fb}^{-1}$ [55], and $59.6\,\text{fb}^{-1}$ [56].

[†]The coordinate η is explained in detail in Appendix C.

Chapter 4

Particle Reconstruction

Once all of the data is collected, said data still needs to be converted a useful form for performing an analysis. In its most raw form, the CMS output is a series of voltage spikes measured from different parts of the detector, and the next step, called reconstruction, transforms this into a collection of particles, identifying and quantifying kinematic properties of each with the highest possible accuracy.

Most of this is carried out "offline" on a farm of machines that run a set of algorithms for identification and reconstruction. Because it is just algorithms, the data can be reinterpreted with higher fidelity as methods become more refined. With the current iteration used for Run 2 data is called UltraLegacy. UltraLegacy is the successor to the individual, year-by-year reconstructions used when the data was first produced, leading to the advantage of a consistent reconstruction for all data in Run 2.

4.1 Vertices and Pileup

Before the reconstruction can start, we must identify the vertices in the event to find our hard collision and initiate the whole reconstruction chain with this information. For any recorded collision, there is the primary hard interaction from where the triggered particles originate[†] and the other, softer collisions from other protons in the beam interacting. These collisions that are not the main collision are called the Underlying Event (UE). UE interactions are made up of elastic proton collisions, soft inelastic collisions, and even initial state radiation from the primary interaction, but all of these UE particles are essentially not a part of the primary interaction and need to be filtered out to properly understand the event. To do this, the vertexing algorithm finds all the vertices created in the beam interaction, tags the primary interaction by selecting the vertex with the largest energy, and removes objects coming from the other, UE vertices.

The first issue to tackle is how to define a vertex in the first place. The vertices are found from the intersection of tracks, but because of slight error in positioning, reconstructed tracks, which are essentially

[†]While there is a possibility of two hard interactions for a single event, it is low enough that it is not considered in normal reconstruction.

a line representing the path of the particle, can originate from the same vertex but not intersect. If our tracks don't intersect, we have no obvious way to even define our vertices. To start this process, we filter out tracks we don't expect are the primary or UE interactions. These initial interactions create primary vertices (PV) which can be distinguished from secondary vertices (SV), or those created in the cascade decays. We expect the SV and their associated tracks to be located radially away from the beam axis since these cascade decays will occur after some travel time. This means, to filter out tracks originating from SVs, we can find the beam spot, and remove tracks sufficiently far away from the beam or interaction point (IP). In practice, tracks are assigned a point based on the closest ascent of the track to the beam axis From here, the points will be arranged in a Gaussian, and we remove tracker hits that are $\sigma_{IP} > 4$, or more than four standard deviations away from the IP where the rest are considered for PVs.

To further simplify matters, the remaining tracker point are flattened to the z-axis. Once we have created vertices, we still need to solve the problem of tracks not correctly being associated with each other. To remedy this, we use a clustering algorithms to recombine the tracks, namely the deterministic annealing algorithm [57]. It works similarly to actual annealing: a "temperature" is slowly lowered from a set max point, iteratively minimizing the "free-energy," thus relaxing the system into a clustering. In metal, this process relates to how the organization of its grain, but in this case, the minimization clusters vertices to create our final PVs. For the free-energy, the partition function is based on the χ^2 of the tracks to the vertex they're associated with, but weighted by the probability that a given track is associated to a vertex, or heuristically:

$$F = T \log Z \propto T \log \left(\sum_i \exp \left[-\frac{\chi_i^2}{T} \right] \right)$$

Using this metric, at high temperatures, all tracks are associated as a single vertex and when the temperature approaches zero, each track individualizes to its own, separate vertex. This means, as the temperature decreases, vertices split as it becomes "energetically" favorable for more vertices to exist. Thus, the algorithm works by starting at a sufficiently high temperature and iteratively lowering it to a set value as vertices split to the final PV collection. This final temperature is analytically determined, and currently used value is $T_{\text{min}} = 4$.

After the tracks are clustered into each PV, the vertex with the largest summed energy emanating from it is considered the primary interaction point. The other PVs have tracks associated with jets which are called pileup, so these tracks are removed from the event. For reference, during Run 2, the number of pileup vertices averaged around 35 per bunch crossings. While pileup can be filtered, out, the neutral portion of the pileup interactions is only seen in the HCAL. The HCAL does not have enough resolution to associate the neutral portion of pileup to their interaction points, so the neutral pileup contribution has to be estimated. For this, we assume that pileup jets are produced isotropically in the HCAL and with a predictable decay chain that results in 65 % charged jets, 10 % neutral jets, and 25 % photonic byproducts. From this assumption, the total neutral pileup energy can be estimated from the charge

pileup, so the neutral energy portion can be subtracted uniformly from the HCAL for any subsequent energy calculations.

4.2 Particle Flow

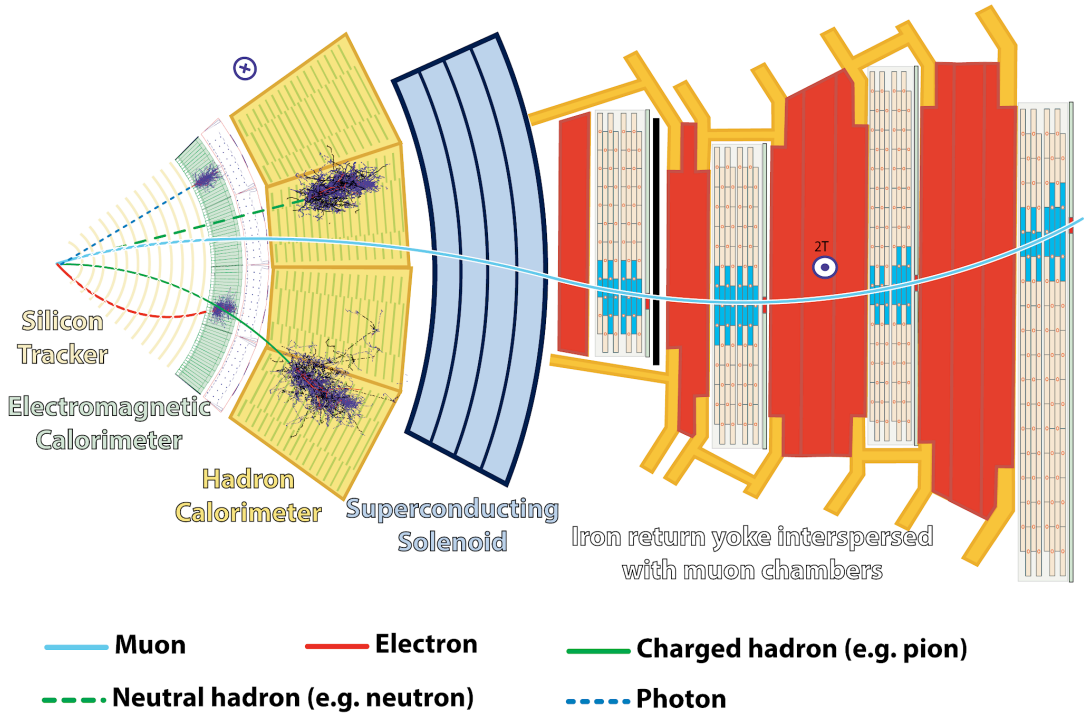

Figure 4.1: Overview of the particle flow algorithm. Each type of particle reconstructed at CMS is identified by the culmination of the signature found in each subdetector. This means reconstructed particles are come from a holistic look of the detector rather than isolating information to a single subdetector. [58]

The essence of the CMS particle reconstruction is an algorithm called Particle Flow (PF) [58]. The idea behind PF is simple: use all parts of the detector to identify particles and refine their kinematic values. In other words, PF creates global tracks of particles throughout the detector where the response of each subdetector are used as shown in Figure 4.1. While seemingly obvious, PF has only become commonplace with technological advances which allow for large amounts of cross-communication of subdetector information. Additionally, the PF algorithm only works effective where there is a high and similar enough resolution between subdetectors. If the resolution is too low, particle candidates merge together disallowing proper identification, and if one of the subdetectors has much worse resolution compared to the others, particles cannot be easily correlated between the different detectors. CMS was created with the intent of using the PF algorithm, so it excels at this, allowing for high reconstruction efficiencies for each particle.

Because this analysis does not use photons or taus, they will not be discussed.

4.2.1 Muon

Muons are the easiest particle to reconstruct because of the support of the large, dedicated muon system. Muons are primarily identified by the tracks in the tracker and the muon system, both of which the reconstruction algorithm tries to match together. When a muon is matched in both the tracker and muon systems, it is called a global muon since it uses the whole detector.

The efficiency of global muon reconstruction is very high, around 99 %, and the rate of particles faking muon is small as well. Charged hadrons can sometimes make it to the muon system, but most are filtered out by large matching energy deposits in the HCAL, especially in the outer HCAL. The larger problem with muon reconstruction is identifying the source. For analyses, we are concerned with prompt leptons, or those created in the primary interaction, but muons can be created as a byproduct of a W boson decay from any other particles in the event and be misidentified as prompt. These "nonprompt" muons are usually be filtered out by their low energy and tracks originating from a SV, but some can pass typical muon requirements, poisoning the overall real, prompt muon efficiency. To tackle these nonprompt muons, techniques are used for identifying and removing nonprompt muons and leptons in general which is discussed later in Chapter 6 as well as Chapter 7.

4.2.2 Electron

Electrons have a fairly clean signal with a track, large ECAL deposit, and little to no HCAL or muon system hits. Further, using the energy and momentum values coming from the tracker and ECAL, electrons can be validated based on a mass calculation, using the mass to veto against any hadronic resonance that might be misidentified as an electron. Electrons also may make it to the HCAL, but compared to any jet, the ratio of ECAL to HCAL deposits will be larger than for most jets, allowing this metric to sort electrons from hadronic candidates and vice versa.

The larger issue with electron identification comes from the inherent bremsstrahlung decay mechanism found in electrons. CMS tries to induce this decay process in the ECAL with the high-Z material, but bremsstrahlung decay is stochastic, or happens randomly, meaning many of these decays happen in the tracker. This decay disrupts the track, so these electrons will not have a clean arc, but will rather include a kink. To fit the tracks, a fitting method called Gaussian Sum Filter (GSF) is applied to approximate the electron's path and identify points of photon emission. A χ^2 of the fit is also used as a quality quantifier of the fitted track. Another issue is the emitted photon carries away part of electron's energy but does not leave a track. With the GSF fit, the photon's direction can be inferred and ECAL hits can be traced back to the electron, so the full energy of the initial electron can be properly reconstructed.

The combination of the tracker and ECAL means the efficiency for electrons can reach a level of roughly 85 %.

Electrons do have a bevy of particles that decay similarly to it. First, photons that convert to electrons early can be misidentified as electrons, but using the χ^2 of the fit as well as vetoing electrons that have multiple missing tracker hits can reduce the overall photon misidentification rate. Most charge hadrons can be filtered out using the ECAL to HCAL ratio, but pions, because of their small mass, are more likely to be stopped in the ECAL and be mistaken for an electron. To deal with this, machine learning algorithms which are trained on the different quantitative factors of the PF electron candidates are used to filter these false electrons out and this is discussed more in Section 6.1.4. Similar to muons, nonprompt electrons can be misidentified as prompt electrons, but the same techniques used for mitigating this in muons can be similarly applied to electrons.

A last problem which is discussed in more detail in Section 7.3 is charge identification. Unlike muons, which have generally clean tracks, electrons must contend with possibly bent GSF tracks in reconstruction. The electron's charge is inferred from the direction the particle bends in the magnetic field so, while rare, the GSF track can misidentify the charge because of the bremsstrahlung interfering with the ability to identify the correct bend of the electron. This rate is quite small and methods in this analysis are applied to reduce and model this charge misidentification.

4.2.3 Jets

Jets are the last major particle identified by PF, and they require the combination of tracks, ECAL, and HCAL hits to create a full picture. Jets are expected to have most of their energy deposit in the HCAL with some in the ECAL, so the ratio of energy deposits is used as a metric to avoid being misidentified as an electron. Within each jet, there are neutral and charge portions coming from various parts of the decay chain, so neutral hadrons will be found in charged jets and vice versa. The main reconstruction method works by first removing all tracks and calorimeter hits associated with electrons, muons, and photons so only the jet information is left over. The charged portions of jets appear in the tracker and HCAL, so the two are associated with each other. The neutral hadrons only appear in the HCAL and do not make tracks, so they are isolated by subtracting the neutral pileup energy from the HCAL.

Because jets are made of a cascade decay of hadrons originating from a single object, the individual neutral and charged hadrons found need be recombined to reconstruct the jets present in the prompt interaction. For achieving a clustering, we use an algorithm called anti-k_T [59]. The anti-k_T algorithm works by creating a novel distance metric for hadrons, h_i, h_j, using the inverse of the p_{T} of the hadrons in its definition:

$$d(h_i, h_j) = \min\left(\frac{1}{p_{\mathrm{T}}^2(h_i)}, \frac{1}{p_{\mathrm{T}}^2(h_j)}\right)\frac{\Delta R(h_i, h_j)^2}{R^2}$$

Where ΔR is the distance between the two hadrons (as defined in Appendix C) and R is radius of the

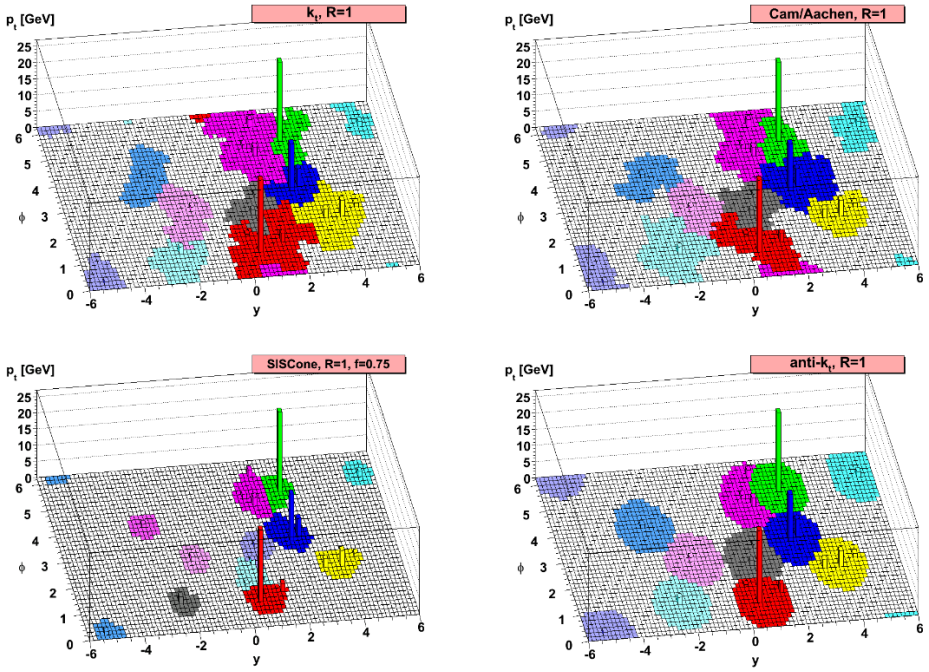

Figure 4.2: Example of different clustering algorithms with anti-k_T in the bottom right. The anti-k_T algorithm has the benefit of clustering the subjets into a consistent shape, unlike other algorithms which can create final jets with bizarre boundaries that can meld into other surrounding jets.

final jet. The distance metric allows lower p_{T} hadrons to be considered "closer" to high p_{T} hadrons compared to similarly spaced low p_{T} hadrons, thus the soft hadrons are grouped with hard hadrons. This is expected as hard hadrons typically slough off lower energy hadrons as they travel through the detector, or hadrons will naturally cluster such that we remove issues with infrared and collinear subjets. The algorithm works by iteratively clustering the closest hadron pairs together and substituting their sum in the pool of hadrons. The clustering continues until all jet candidates have a distance greater than the jet size limit, R. The result of this clustering can be seen in a toy case in Figure 4.2. At CMS, jets are defined to have a radius of $R = 0.4$, so they are typically called AK4 jets to reference this.

4.2.4 Missing Transverse Energy

While PF reconstruct most all particles, there is one that CMS cannot directly measure: neutrinos. Neutrinos have no charge and their cross-section with most matter is very small, so they are invisible to CMS and their momentum is lost. This lost energy is detrimental to reconstructing the event, but we can leverage conservation of momentum to construct a proxy for the lost neutrino energy. The colliding protons travel along the z-axis, so the cylindrical symmetry of the system means the total momentum in

the transverse plane should be zero. We should then expect any imbalance observed corresponds to the missing neutrinos. This imbalance is called the missing transverse energy or MET:

$$E_{\mathrm{T}}^{\mathrm{miss}} = -\sum \overrightarrow{p_{\mathrm{T}}}$$

The caveat is that MET inherently loses the neutrino information because it is the total p_{T} sum, so individual neutrinos cannot be identified from MET, only the total vector sum. Further, any mismeasurements in the energy of particles will contribute to the MET adding overall noise to the "true" MET. This mismeasurement energy, or "fake" MET, is usually small, so requiring events to contain a modest total MET should remove events with only fake MET and leave only events with actual neutrinos. The MET also acts as a useful tool in searches for new physics since possible BSM particles are expected to not interact with the detector (or it would have been discovered already), so new physics signals are characterized by their large MET. These two reason show why events with a large MET typically have interesting physics, and, as mentioned in Section 3.2.6, why MET is a useful quantity for triggering.

Chapter 5

Simulation

After reconstruction is performed, we still need a way to perform a statistical analysis on said data. To do this, CMS uses simulated data to model the expected data, so statistical tools can compare the two. This simulation need to model processes produced in a proton-proton collision and reproduce how these processes would appear in the detector itself. The first step requires an understanding of kinematics and production rates of different processes while the second requires a keen understanding of the detector, its conditions, and its response to the different stimuli. Combining these two disparate goals can seem difficult, but the simulation process can be cleverly factorized to create high quality results in a step-by-step method.

Splitting our modeling into stages is necessary because of the changing of modeling at different energy scales, μ, of the interaction. As stated in Section 2.2.2, quantities such as the coupling constant change at higher energies meaning the physics behaves differently. For the electromagnetic and weak forces, this is not problematic, but QCD acts as a large hurdle. QCD is asymptotically free, or its coupling constant decreases with energy scale, so it is nonperturbative at low energies but becomes perturbative at energies above the so-called QCD scale ($\Lambda_{QCD} \approx 1$ GeV). This means we have a prescription for simulating our data: step-wise simulate the data at the different energy regimes using the proper theoretical model at each point. The distinct stages in the QCD particle evolution are shown in Figure 5.1 and also described as:

- $\mu > \Lambda_{QCD}$: Hard interaction or Feynman diagrams model processes with simple perturbation calculations
- $\mu \approx \Lambda_{QCD}$: Hadronization or quarks are turned into hadrons, use the Lund string model
- $\mu < \Lambda_{QCD}$: Low energy decays or hadrons are confined, can use hadron physics to model

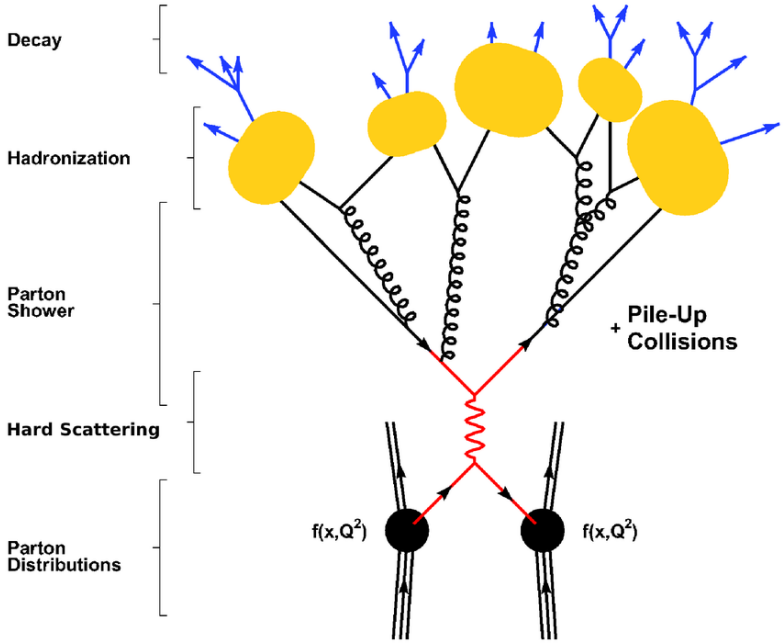

Figure 5.1: Diagram of the basic evolution of quarks and hadrons in a particle collision. This evolution matches the decrease in the energy scale of the interaction from the hard interaction where QCD is perturbative, to the final decay process where QCD is nonperturbative.

5.1 Hard Interaction

The LHC has a center of mass energy of 13 TeV, so the expected interaction energy scale is well above the QCD scale meaning the interaction can fully be modeled by perturbation theory and Feynman diagrams. In fact, hadrons aren't even considered in this first step because hadronization doesn't happen until the QCD energy scale or when the gluon binding energy increases enough for confinement to reappear. While QCD has a complex $SU(3)$ group structure involving three color charges, calculating quantities from these Feynman diagrams is a known and solvable problem. For a motivated individual, tree level calculations can even be done by hand. Thankfully, computer programs have been developed that can perform the calculations and produce simulations of the different Feynman diagrams to higher orders of precision (i.e. to NLO and higher). For doing this, the simulation program generates the necessary Feynman diagrams, including loop diagrams if needed for NLO+ accuracy, and uses simple combinatorial tricks of quark and lepton substitutions when possible to speed up calculation. These substitutions of quarks, gluon, leptons pose little problem since relative constants and the correct PDF values simply need to be swapped out. This means many processes only have a handful of diagrams that require actual amplitude calculations.

After the diagrams and their amplitudes are produced, the kinematics need to be modeled. While probability distributions can be theoretically created for all the different inputs and output momenta,

this becomes unwieldy, so it is instead modeled by a technique called Monte Carlo (MC) simulation. MC simulation takes a probability density function and models it by creating discrete events by randomly seeding values from it. The benefit of this method is it reduces the computation overhead of propagating the probability distributions through each step of the simulation and reconstruction along with converting the probability into discrete events which are analogous to the actual data. To model a specific physics process, it is sufficient to generate enough MC events to fully capture the different features of the particular process, and the actual expected yield is reconstructed by weighting each event such that the total sum of the weights equals the expect number of events: $N = \mathscr{L}\sigma$.

For the actual QFT MC simulation, the main parameters and outputs are simply the momentum of the incoming and outgoing particles[†] as well as spin values. There are further parameters that can be tweaked such as the value of coupling constants, PDF model used, mass values of different particles, and much more, but these are often set to agreed values with the possibility of varying them to simulate the overall uncertainty in these parameters. At this step of the simulation, the modeling is performed solely using the information from the Feynman diagram, but with spin information included, this takes the form of a matrix. For this reason, the output of this first step of hard interaction modeling is often called matrix level.

For actually reating the MC simulation, the programs MadGraph and Powheg are the most common. MadGraph [60, 61], which was originally made for tree-level Feynman diagram generation, now allows for NLO MC generation with a utility called aMC@NLO. One difficultly with NLO calculation is parton showering. While loops in NLO serve as a calculational burden, NLO diagrams also include infrared, or soft emissions. In the next step of the generation called showering, the algorithm simulates the radiation of particles from the quarks and gluons in the event which poses a problem because NLO matrix level diagrams with infrared emissions may be double counted with LO showered events. MadGraph includes multiple algorithms to help with matching and removing these double counted events, but it often requires some events to have a negative weight to cancel out these introduced, infrared divergences [62]. The other generator, Powheg [63], which stands for Positive Weight Hardest Emission Generator, orders its objects the energy which in its pipeline, allows for easier matching so all or nearly all weights can be positive, avoiding the negative weight issue. With the differences in generation methods, Powheg tends to be used for generating processes with top quarks, such as $t\bar{t}$, because of its better accuracy at modeling the gluon induced $t\bar{t}$ (as seen in Figure 2.3) and the soft showers coming from it, while MadGraph is used more generally for all other event generation.

[†]While the protons each have 7 TeV of energy, the interaction is between the partons, and the individual partons can have arbitrary momentum fraction of the total proton momentum.

5.2 Hadronization

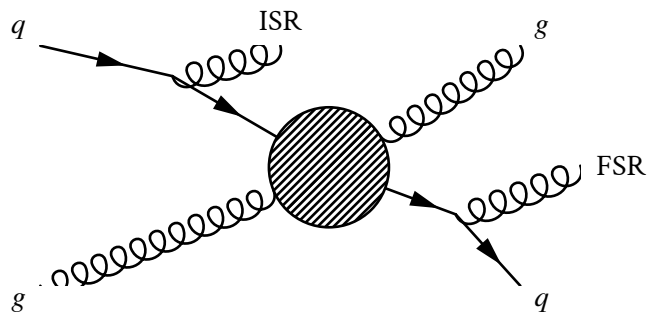

Figure 5.2: Representation of initial state radiation (ISR) and final state radiation (FSR). Here, the blob represents the hard, main interaction. ISR is produced before the hard interaction from a QCD interaction with a very energetic initial parton while the FSR are addition jets that are produced from the splitting of output partons.

After the main interaction, the overall energy scale of the interaction decreases meaning events transition from a perturbative scale to a nonperturbative scale for QCD in what is called hadronization. As alluded to in discussing matrix element generation with MadGraph and Powheg, this is not the full story as there is the addition of parton showering, but both the parton showering and hadronization are handled together by a software called Pythia8 [64, 65].

Parton showering happens after while the energy scale is still above the QCD scale where additional gluons and quarks shed off particles in the main Feynman diagrams. Either this happens before the primary interaction or after as diagrammatically shown in Figure 5.2. High energy partons act like water drops, the larger/more energetic they are, the more easily they will split and fragment. This means we expect large quantities of hadrons in each event, either from the primary interaction or particles surrounding it. The resulting jets coming before the main interaction are called initial state radiation or ISR and the jets coming after the main interaction are called final state radiation or FSR. Because ISR comes from the initial partons, it typically is made of higher energy jets while the FSR typically is made of lower energy jets.

After the parton showering, the event is propagated to a point the energy scale falls below the QCD scale, thus we cannot consider bare quarks singular elements. Now, we have to collect the quarks together to make bound hadrons. Because this isn't perturbative, only heuristic models are used to describe the hadronization process. The main method for describing the hadronization process is the Lund string model [66]. It operates on the assumption that the strength of the QCD field lines between particles is linear, so the field lines can be modeled by strings or tubulars. Based on the color charge assigned to each quark, connections are made between the shortest color neutral strings to create the basis for initial hadrons. From this basic model, the string will fragment based on the energy availability to create further

quark-antiquark pairs from the energy stored in the strings between the different quarks. This means the fragmentation can be fully simulated to great effect through a step-wise iteration of time, looking for any favorable splitting of the hadrons.

As mentioned, both of these steps are performed using PYTHIA8, which is run on the matrix level data to produce an output file that contains the full event level information, most notably the resultant hadrons calculated.

5.3 Decay

Once PYTHIA8 has created the hadrons and inserted any ISR or FSR associated with the decay, the base interaction is complete. Any reconstruction algorithm is essentially trying to reproduce the information at this step, and it is the "picture" of te event that CMS tries to take. But we need to incorporate our "camera" into the simulation. Our simulation pipeline must propagate our event through a simulation of the CMS detector, decaying the particles through the detector and reading out the products as they would be by the different subdetector. By replicate the event's measurement in MC, both it and data can be reconstructed and compared for future statistical tests. The knock-on effect is event information before and after the detector simulation becomes distinct: there is the "true" event structure produced by the generators, say, MADGRAPH and PYTHIA8, compared to the reconstructed event structure, i.e. particle information that has been run through the detector then reconstructed.

This simulation step is the most taxing since we must model the interactions of the particles with all the different materials in the detector considering how the responses change with energy, type, and decay mechanism for all the particles. Thankfully, this is all well understood and simulated using a tool called GEANT4 [67–69].

GEANT4 allows for the geometry of the CMS detector to be input, including the type of material and type of detector to simulate both the interaction of the particles with the medium and the readout of the respective detectors. These two steps, often called DigiSim or simulation plus digitization, are, as mentioned, the most computationally intensive process of the simulation pipeline, but this is in large part because of the emmense size and complexity of the detector and particles that travel through it. While several optimizations are made to increase process speed, this step is inherently complex and time-consuming if one wants to avoid sacrificing too much accuracy.

For this DigiSim step, performance tests are run regularly to ensure that the modeling of the detector response is correct, especially if certain simplifying assumptions in the simulation are made. Another issue that needs to be monitored and updated is the change of the different subdetectors' response and performance. Unsurprisingly, after near continuous exposure to high energy radiation, parts of the detector start to degrade meaning their efficiency will start to decrease. As an example, the lead-tungstate crystals become more opaque with radiation meaning the amount of light transmitted to the photomultipliers

will decrease, so reconstruction algorithms need updated to reflect these changes. These performance change happens continuously and smoothly, but the MC events are produced all at once, so any performance changes have to be emulated such that the performance profile matches, on average, the full, real dataset.

Up to this point, the hard collision has been fully described and simulated using MC methods, but this is only one pair of particles within the whole beam, and the UE particles have not been accounted for. We expect the other proton collisions to not effect the primary interaction, so adding this last piece of information ends up not being difficult. In what is called the premixing step, we forgo producing each event with the pileup information included, and, instead, simulating millions of the pileup jets and add them into each MC simulated event. We know that the average number pileup vertices for Run 2 was around 35, but this can be dynamically adjusted if pileup tagging algorithms change, allowing the MC's pileup content to be adjusted without reprocessing the whole MC event. Here, we've offset some computational overhead by splitting and remerging these parts without any substantial loss in accuracy.

The last step for the event is simulating the trigger response for the events, which is mainly applying the L1 trigger since it is implemented in hardware. The HLT has less problems because it is a software trigger, so it can run on the MC exactly as would be done for a data event, and thus requires no dedicated simulation software. After all these steps, we have a fully simulated event which can be put through the same reconstruction algorithms described in Chapter 4.

Chapter 6

Object Definitions

Before we start performing the brunt of the analysis, we need to define what the objects in the detector are. The PF reconstruction algorithm turns the basic voltages and timings from the detector into 4-vectors and a particle type, but these definitions are very loose to accommodate the spectrum of analyses at CERN. For our case, a balance must be struck in our lepton definition: reduce incorrectly identified particles by increasing the requirements or keep high particle efficiency by reducing them. For a particular analysis, the analyzer needs to choose any additional definitional requirements for each particle to optimize for their particular needs such as the final state under consideration.

With this in mind, this chapter goes through the different requirements placed on the different physics objects.

6.1 Leptons

As stated in Chapter 2, leptons only refers to the lighter leptons, or electrons and muons. Taus are not considered in this analysis, whether for event selection or rejection. While hadronic taus might be useful, even for vetoing events with taus, this analysis is limited in the number of events, so any tau requirement would reduce the overall signal efficiency for what might be negligible gains.

One of the most important features of this analysis is the data-driven background estimation using measurements of misidentification rates, detailed more in Chapter 7, so most of the lepton requirements are aimed to better understand and reduce false or misidentified leptons. The first variable used for taming the misidentification rate is called the isolation.

6.1.1 Lepton Isolation

The isolation variable is designed to filter leptons originating from jets, or nonprompt leptons. Non-prompt leptons are expected to be in proximity or even inside a jet's HCAL hits, so the isolation variable leverages this to, in essence, reject leptons that have too much hadronic energy near it. To do this, we

define the isolation variable as the ratio of the hadronic momentum to the lepton momentum:

$$I_{\text{rel}} = \frac{\sum_R p_{\text{T}}(h)}{p_{\text{T}}(\ell)}$$

where here, the sum of momentum is all hadrons, h, within a cone of radius $R = 0.4$ to match the typical radius used in the jet clustering algorithm. Using this definition, we see that a prompt, isolated lepton should have a small I_{rel}, close to 0, while a nonprompt lepton is expected to have a larger I_{rel}. This version of isolation, called the relative isolation, works well enough, but we can improve it with a couple of modification. First, it can be noted that the jet momentum is overestimated because any neutral pileup energy is counted towards the total jet energy. This would mean an event with a large amount of pileup could falsely cause all leptons to have a larger isolation meaning the leptons in these events would be more likely to be rejected. To combat this, we can subtract the neutral pileup from this sum using the average neutral pileup density, ρ. As an extra precaution, logic is included to guard against cases when the pileup density is greater than the amount of neutral energy detected within the cone. Second, as the lepton's energy increases, the energy deposits are expected to become more collimated, or focused, meaning its energy deposits would be concentrated into a smaller cone. To leverage this, we can allow the cone size to decrease with p_{T}; for our case, this takes the form $R = 10/p_{\text{T}}(\ell)$ (where the R is constrained to the range [0.05, 0.2]). This helps for reducing false positives in jets heavy events. A last nuance is the cone area, $\mathcal{A}_{\text{eff}}$, which is necessary for calculating the neutral pileup contamination. The effective area of the cone will change with η because of the nonlinear nature of this measure, but the p_{T} dependent radius also means the area needs to be adjusts when subtracting out the pileup. All together, we create the new p_{T}-dependent lepton isolation called the "mini-isolation" defined as [70]:

$$I_{\text{mini}} = \frac{\sum_R p_{\text{T}}(h^{\pm}) + \max\left\{0,\ \sum_R p_{\text{T}}(h^0) + p_{\text{T}}(\gamma) - \rho\mathcal{A}_{\text{eff}}\right\}}{p_{\text{T}}(\ell)}$$

where $h^{\pm}$, h^0, and γ represent the charge hadrons, neutral hadrons, and photons respectively. For all leptons in this analysis, they must have a $I_{\text{mini}} < 0.4$, which represents a fairly loose isolation requirement. A tighter isolation could be used for larger nonprompt rejection, but we opt for using a specific machine learning variable developed by the ttH analysis called the ttH−MVA [71].

6.1.2 ttH−MVA ID

While the isolation variable typically works well, nonprompt leptons from high cross-section processes, such as $t\bar{t}$ or Drell-Yan[†], are produce at a large enough rate that most three top signal events would simi-

[†]Drell-Yan is the name given to Z/γ^* boson production through the s-channel. It is notable for being one of the largest processes produced at the LHC along with W boson production (called W +jets often) and events with solely jets (often called QCD events).

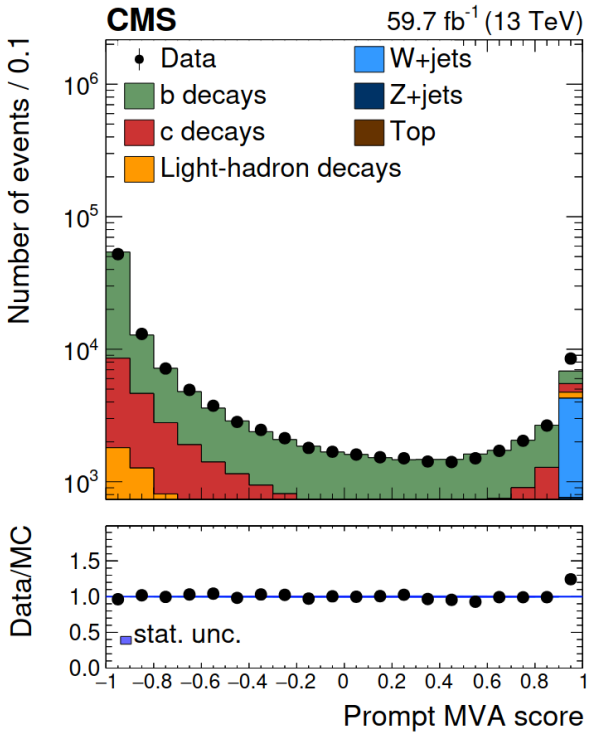

Figure 6.1: Example plot of the ttH−MVA from the tth analysis. This plot is taken from a multijet region from 2018 data. Since nonprompt leptons are lepton originating from jets, the MVA is trying to separate the real leptons in the W+Jets, Z+Jets, and top processes from the nonprompt leptons that have quark induced origins. b-jets, which are the heaviest quarks found in hadrons, contaminate the signal region some, but the prompt and nonprompt leptons are separated quite cleanly, much more than possible with simply the isolation variable.

larly be removed to by the isolation requirement necessary reduce these backgrounds. Machine learning ends up being the solution because it can reach higher efficiencies than the isolation alone. The ttH analysis developed a multivariate analysis (MVA) tool that calculates a discriminant score to distinguish prompt leptons from nonprompt using multiple input event variables [72]. A plot of the ttH−MVA from the ttH group's paper is shown in Figure 6.1, and one can see how it separates leptons coming from jets and real leptons. For our purposes, the MVA improves our filtering efficiency compared to using solely the mini-isolation allowing us to remove more nonprompt background while keeping more signal. The ttH−MVA is produced using a tool called a boosted decision trees or BDT [73] (of which, BDTs are discussed in more detail in Section 8.3). The inputs to the BDT include the lepton kinematics, isolation variable, vertex distance variables, other lepton ID scores, and the DeepJet score of the closest jet (which DeepJet is discussed in Section 6.2.2). While designed to be generic for distinguishing prompt leptons from nonprompt leptons, the ttH−MVA works well for the three top final state considering three top and ttH processes shared similar final objects and kinematics owing to the top-pair leg of both. The

details of how the ttH−MVA is used and the nonprompt method as a whole are described in Section 7.2.

6.1.3 Lepton Categorization

Once we have a way to remove nonprompt leptons from our signal region, we have the basis for the definition of our leptons, but we have another necessary component. As discussed in much more detail in Chapter 7, the nonprompt background, or events containing a nonprompt lepton, benefits greatly from being modeled directly from data as opposed to be simulated using MC. To achieve this data modeling, we must create different categories for our leptons, namely a "fake" and "tight" definition. It also becomes advantageous to have a "loose" defined lepton for discerning different regions of phase space. From this, we have three different categories of leptons: "loose", "fake", and "tight".

For these three categories, they are defined to be successively more selective, i.e. a fake lepton has all the same requirements as the loose leptons if not more stringent. This means all fake leptons are also loose leptons, and all tight leptons are also fake and loose leptons. As a summary of each category:

- Loose leptons: very basic requirements used for event vetoing, or ensuring no additional leptons are present in an event
- Fake leptons: leptons almost passing the tight category, but failing to pass the ttH−MVA requirement; these are used for estimating the nonprompt background
- Tight leptons: the basic, but most restrictive lepton, used for signal extraction

In defining each lepton categories, several IDs are used, where IDs in this context are collections of requirements on kinematic variables related to a lepton that reach some signal efficiency and/or background rejection. At CMS, dedicated groups study the leptons to devise and release these IDs, usually with the naming convention of loose, medium, and tight, where the loose ID has the highest signal efficiency, but lowest rejection while tight ID has the opposite. With the categories and IDs in mind, we can create the definitions for the electrons and muons used in this analysis.

6.1.4 Electrons

Electrons are primarily identified using an MVA produced by CMS. Electrons have numerous identification hurdles, as described in Section 4.2.2, so a sufficiently complex method is needed to separate true electrons from particles misidentified as electrons. This MVA combines multiple variables relating to the electron to create a discriminant that partitioned into so-called working points, or thresholds that lead to certain efficiencies in the electrons that can be used as an ID. For this analysis, the loose electrons used the "MVAloose" working point while the fake and tight leptons use the "MVA90" working point (where the 90 here corresponds to the expected signal efficiency).

For the kinematics of the electrons, loose electrons must have $p_{\mathrm{T}} > 5\,\mathrm{GeV}$ to meet typical triggering thresholds while fake and tight electron must have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$. To remove any spurious electrons, they

must have $|\eta| < 2.5$ to account for the limits of the ECAL (the η variable being explained in Appendix C).

To remove photons faking electrons which are not eliminated by the MVA ID, we limit the number of missing tracker hits. Electrons should pass through every layer of the tracker, creating a hit in each; these high energy photons do not ionize the tracker silicon, so they should have missing hits in the different layers of the tracker, at least until the photon pair-produces two electrons. Loose electrons can only have one missing hit while fake and tight electrons are not allowed to have any missing hits. An additional conversion veto developed by electron experts is applied to reduce selecting electrons coming from photons so we are sure our selected electrons have a high purity.

To remove electrons whose charge is difficult to measure accurately, we require that electrons pass a tight charge ID, also created by electron experts. While this tight charge requirement reduces the occurrence of incorrect charge identification, a non-negligible number of opposite-sign events can be categorized as passing the same-sign requirement. Since the charge misidentification modeling is known to be inaccurate in MC simulation, a data-driven background estimation method is used to model these events which is described in Section 7.3.

Additional requirements are made to improve the quality of the electrons, mainly related to the tracks distance to the PV, but also to make the MVA ID'd electrons look closer to the trigger level electrons.

Fake electrons are distinguished from the tight electrons using the ttH−MVA. For electrons that have a ttH−MVA score above a threshold (for this analysis, the threshold was found to best be at 0.65), they are categorized as tight, while if their ttH−MVA is below the threshold are categorized as fake. Fake electrons are treated slightly different to tight electrons, namely by using a different, corrected p_{T} and having an additional requirement on the p_{T}-ratio, where $p_{\mathrm{T}}^{\mathrm{ratio}} = p_{\mathrm{T}}(\mathrm{j})/p_{\mathrm{T}}(\ell)$ having $p_{\mathrm{T}}(\mathrm{j})$ be defined as the p_{T} of the geometrically closest jet to the lepton. This is discussed in more detail in Section 7.2.1.

6.1.5 Muons

Muons are identified according to a "cut-based" ID provided by the muon experts. A cut-based ID is created by applying a series of requirements to list of variables as opposed to the more complicated MVA used for electrons. In general, muons require less complicated tools than electrons because of their high efficiency afforded by the dedicated muon system. For this analysis, the loose cut-based ID is used for defining loose muons while the medium cut-based ID is used for the fake and tight muons.

The p_{T} requirements for muons are the same as electrons, namely loose muons must have $p_{\mathrm{T}} > 5$ GeV and fake/tight muons must have $p_{\mathrm{T}} > 20$ GeV. The η requirements is slightly different from electrons, namely $|\eta| < 2.4$ to account for the different size limits of the muon detector itself.

Similar to the electrons, muons have a tight charge requirement, but it is simply calculated by taking the ratio of the muon p_{T} error to the muon p_{T} and requiring it to be less than some experimentally found value. Charge misidentification is less of an issue with muon because of their suppressed bremsstrahlung and global tracks which at to cross-check the charge. With the tight charge requirement, the muon

charge misidentification is reduced to a rate less than 10^{-7}, as found in MC simulation, thus muon charge flips are not considered in this analysis.

For distinguishing between fake and tight muons, the process is the same as with electrons: muons with ttH−MVA > 0.65 are identified as tight and those that don't pass this requirement are labeled as fake. Similarly, there is an addition p_{T}-correction and $p_{\mathrm{T}}^{\mathrm{ratio}}$ threshold requirement applied to fake muons, which is discussed in full in Section 7.2.1

6.2 Jets

Jets are reconstructed using the anti-k_T algorithm using a radius of $R = 0.4$, but an additional process called charged hadron subtraction or CHS is used to reduce pileup contamination [74]. Charge hadron subtraction is the name for the process of removing all charged tracks associated with pileup vertices. This means clustering is only performed on tracks coming from the primary interaction. Further, jets are required to pass the "tight" jet ID requirement, which is a cut-based ID developed by jet experts at CMS, including factors such as HCAL to ECAL energy deposits. While the HCAL extends to $\eta = 5$, jets are only considered in the central region, or $|\eta| < 2.4$. The tracker and ECAL coverage only extends to the central region, so forward jets only have the HCAL for reconstruction. This means central jets usually have higher purity because of the additional detector information available for identification. Since this analysis is so jet intensive, it is important enough to require only central jets for the extra fidelity. Further, this η requirement helps to keep concurrence between jets and b-jets (described below in Section 6.2.2), which needs the central detector information.

In addition to the tight jet ID, an additional pileup jet ID is applied. This ID is a centrally produced MVA made to distinguish prompt jets from pileup jets. The typical vertexing methods can filter out most pileup, but the vertex clustering algorithm can still lead to extraneous jets being considered prompt. This pileup jet ID helps to reduce the pileup contamination significantly. This ID was trained and is only applied to jets with $p_{\mathrm{T}} < 50\,\mathrm{GeV}$ since most pileup will be soft and/or will not scatter enough to have a large proportion of its momentum in the transverse plane.

Lastly, while the isolation requirement is applied to leptons to remove nonprompt leptons, jets may still be located near leptons, especially in events with numerous jets, as is expected for our signal region. To avoid including these soft jets in proximity to the leptons, all jets within $\Delta R < 0.4$ of a lepton are removed. All of these requirements give a high purity collection of jets, but the jets still need to be calibrated, most notably with the jet energy corrections.

6.2.1 Jet Energy Corrections

The jet energy corrections, or JEC, are the corrections to jet energy and momentum to fix for differences between data and MC as well to correct the measured energy to match more closely the true prompt jet

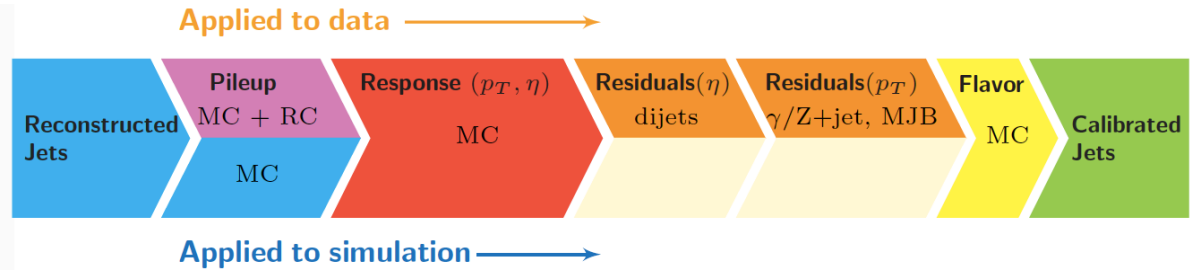

Figure 6.2: Schematic of the steps for applying the jet energy corrections (JEC). Notably, both MC and data have corrections applied, where the MC is corrected to better match observed jets while data is corrected to better calibrate the energy to match the true, matrix level particle energy. This correction is split into four primary steps: pileup removal; detector response corrections; residual shape corrections matching data distributions to known processes; and quark flavor corrections. Typically, this flavor correct is excluded because of the marginal benefits.

value. To achieve this, CMS opted for using a factorized approach in which the JEC is split into multiple factors corresponding to different effects and are each applied sequentially as shown in Figure 6.2. The first correction, which is applied to data and MC, corrects for any energy contamination coming from pileup. Even with the CHS, which removes the charged pileup, the neutral portion must be removed. This is partially estimated using MC, but additional corrections are found comparing data to MC using minbias events, or events collected with the loosest trigger, to best estimate the pileup interaction. After that, the MC jets are corrected against their truth values so the response of p_{T} and η is similar between the reconstructed values and the truth values. Lastly, the corrections are applied to fix difference found between data and MC which are found by jet experts at CMS and are determined by checking differences agreement in different regions, such as a Z +jet region. These last corrections, since they fix overall data/MC differences, are what are used for systematic uncertainties in the final fit.

It has also been observed that jets in data have a worse overall momentum resolution compared to MC, so we perform a last correction, so our jets have the same overall resolution. Measuring the resolution in MC is simple: take the ratio of reconstructed p_{T} to the generator level p_{T}, and from this, the standard deviation, σ_{JEC}, represents the spread of momentum. For data, we do not have the true p_{T}, so it is estimated in dijet events, looking at the difference in energy between the two jets. Dijet events are required to have two equal p_{T} jets by momentum conservation, so the difference can be used to extrapolate a resolution from the standard deviation of dijet p_{T} subtraction. After making slight adjustments for the measurement techniques, we've created a measure of the resolution which confirms the differences between data and MC. To correct this discrepancy, MC samples are smeared to match the observed resolution.

The jet energy resolution is corrected using a combination of two different methods depending on if the object is a true jet, i.e. there is matching generator-level jet, or not. If the jet is real and has associated

generator-level jet, the p_T is rescaled as:

$$p_T \to p_T + (s_{JER} - 1)\left(p_T - p_T^{gen}\right)$$

Where the s_{JER} term represents the scale factor determined by jet experts to scale the MC jet energy resolution (JER) to match data. For jets that do not have a corresponding generated jet, the above procedure cannot be used. To get around this, we smear the jet's p_T stochastically using a Gaussian distribution, $\mathcal{N}(0, \sigma_{JEC})$, with mean of 0, and standard deviation of σ_{JEC} (measured in MC):

$$p_T \to p_T \left[1 + \mathcal{N}(0, \sigma_{JER})\sqrt{\max(0, s_{JER}^2 - 1)}\right]$$

After applying the JER corrections, the jets are properly calibrated and should give both excellent data-to-MC agreement and an estimation of the true jet's p_T and energy.

6.2.2 Bottom Jets

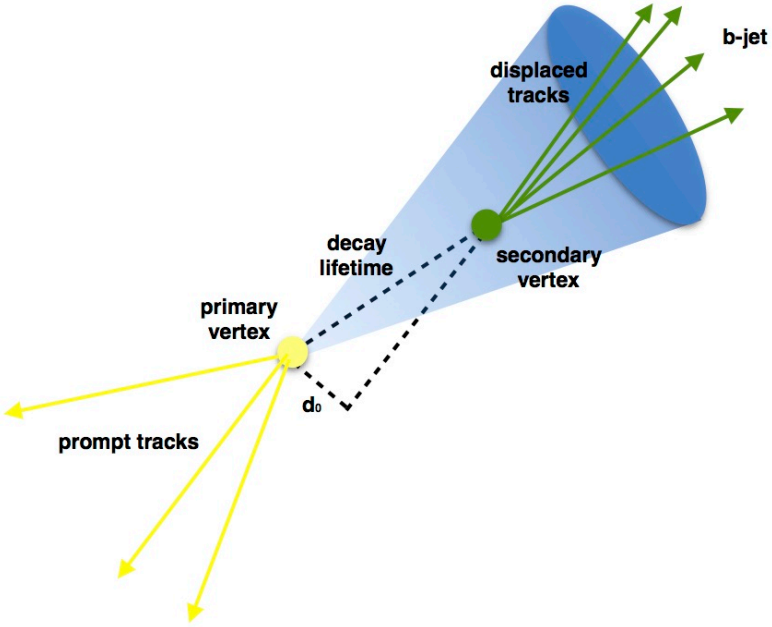

Figure 6.3: Idealized picture of a b-jet decay showing the displaced secondary vertex coming from the long lifetime of the b quark. While the DeepJet algorithm does not just use the vertex displacement, this variable is the most important for distinguishing b-jets from other, lighter flavor jets.

Once jets are calibrated, we can perform an additional selection on the jets to decide which are b-jets, or jets containing a b quark. This is particularly important because each top quark decays into a bottom

quark and W boson[†], so we expect at least three b-jets in our signal events. To make the determination if a jet contains a b quark, we use an MVA created by CMS called DeepJet [75]. The principle behind b-tagging, even before the advent of MVAs, comes from the long lifetime of the bottom quark. Because the CKM matrix value V_{tb} is nearly 1, bottom quark decays into lighter quarks, i.e. a charm or up quark, are greatly suppressed, leading to b-jets having a very long lifetime relative to their mass. This means b-jets will have a characteristic decay length, so b-jets are identifiable by a secondary vertex displaced by typical length as shown in Figure 6.3. While this works well in isolation, charm jets, have roughly the same decay times as b-jets, so additional variables must be incorporated to distinguish bottom and charm jets which is what the DeepJet MVA does. For every jet, a DeepJet discriminant value is determined, and bottom quark experts defined working points corresponding to different signal efficiencies labeled as loose, medium, and tight, similar to lepton IDs. For this analysis, a b-jet is defined as passing the DeepJet medium working point.

To account for any discrepancies in the MC and data distributions related to these b-jets, weights are applied to the event. These weights use a prescription called iterativeFit where the weights are designed to weight the MC events so the MVA discriminant shape matches with data. Normally, scale factors fix data/MC differences for a particular working point, deriving the scale factors to ensure the b-jet kinematics and yields are the same. For the iterativeFit method, scale factors attempt the same kinematic and yield matching while reshaping the MC's MVA discriminant. This means we can use the raw discriminant as a variable in the analysis framework, and having this variable at our disposal is powerful for distinguish jets as well as background from signal. It does come with the caveat: the scale factor derivation introduces multiple sources of systematic error, as is discussed in Section 9.2.3. Additionally, because the DeepJet MVA depends on the jet's p_T, for each JEC systematic, the scale factors for DeepJet must be recalculated to account for the difference in tagging efficiencies.

6.3 Missing Transverse Momentum

The missing transverse momentum, or MET is defined as the negative sum of the transverse momentum from all reconstructed objects in the event, or:

$$E_T^{miss} = -\sum \vec{p_T}$$

While this equation is correct, this "raw" MET is quite simple since it does not contain any of the typical correction (e.g. JEC and pileup) that are handles in reconstruction. This leads to the different levels of MET corrections:

- Raw MET: direct calculation from HLT objects

[†]While the branch fraction for $t \to bW$ is not 100 %, we assume so for simplicity.

- Type-0 MET: using CHS to remove charge pileup contamination
- Type-1 MET: propagate JECs of jets to the MET calculation
- Smeared MET: propagate JER smearing of the jets to the MET calculation

Each step of corrections applied to the MET correspond roughly the order that corrections applied to the jets themselves, especially since a majority of the MET, or at least variation, comes from the jets themselves. One might be concerned about the reliability of the MET since the MET is derived from a vector sum of many corrected quantities. This means the MET could vary wildly between each step in the correction ladder. For analyses with a large sensitivity to MET, extra precautions and studies may need to be performed for validation, however, this analysis only uses the MET as a preselection criterion, so we are not too concerned. As a check, we looked at the variations observed in the MET from the systematic variations of the JEC, and minimal changes were found leading to confidence in our MET calculation.

Because MET is only in the transfer plane, it is characterized solely by its momentum (or energy as the naming suggests) and its transverse angle ϕ. It is also created from all objects in the event, so it is truly a global variable, which means it is sensitive to an event's quality because it needs the full detector information to be accurately estimated. To combat this, events are given quality flags of certain detector or event features at any given time called MET filters. All recommended MET filters are applied to assure that only good events are chosen, and the MET is well behaving in all events. Additional corrections to the x- and y-components of the MET are included to account for an observed modulation of the MET ϕ distribution causing it to deviate from an expected isotropic distribution. The cause of this modulation is not well understood, but suspected to be caused by factors such as differences in detector response, general misalignments of subdetectors, and positioning of the beam spot in collisions.

6.4 Scale Factors

With all the object fully defined for the analysis, there is a final issue of validating and correcting these particles. Inherent in any simulation are differences between it and actual data, and as the variables become more complex, these differences can amplify and/or change in non-intuitive ways. This requires studies to test the overall agreement of objects between MC and data and help in creating correction factors to be applied to the simulation to better match data. Any discrepancies between data and MC could come from multiple sources that change from analysis to analysis, so disentangling these effects can prove challenging. To simplify this problem, the most common procedure is call tag-and-probe.

Tag-and-probe works using two concepts: first, there exists some "standard candle" process that is well-behaved and distinct, and second, we can accurately "tag" some particle (i.e. feel confident we have a low particle misidentification rate), and use our confidence in the tagged particle to "probe" the signal efficiency of another particle. A visual of the basic idea is summed up by Figure 6.4. While this may sound

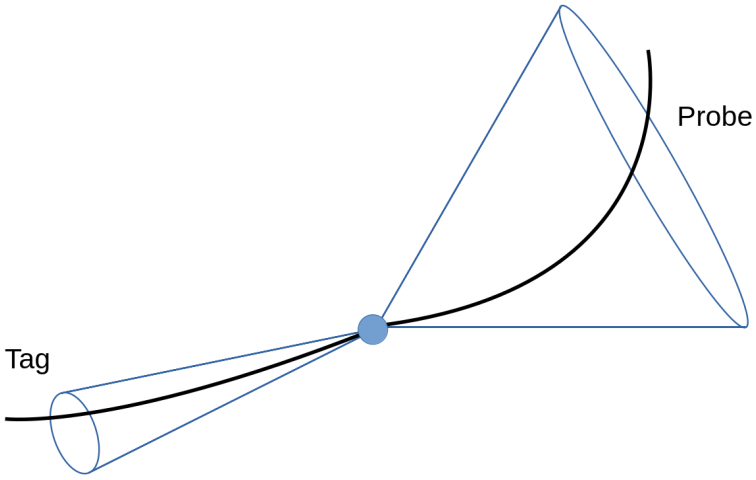

Figure 6.4: Rough picture of the tag-and-probe method. Namely, a well identified, good quality particle (represented by the small energy cone) is tagged, while a second particle that could be much looser, is labeled as the probe and used for efficiency measurements. This set up works based on the principle that there is some kinematic relationship between the tag and probe particles, either by momentum conservation (such as will dijet events) or by a mass relationship (such as a dimuon pair originating from a Z boson).

quite theoretical, the method is obvious because, at the energy scales of the LHC, we have the Z boson which acts as a perfect standard candle. Z boson production, or Drell-Yan, is produced readily at the LHC, has a strong resonant peak mass of roughly 90 GeV, and has a small decay width of 2.5 GeV. Further, the Z boson decays into a pair of opposite charged particles, so one of these particles can act as a tag while the other can be probed for efficiency measurements. While this mass relation is unnecessary for performing tag-and-probe, it adds extra requirements to the particle selection which reduces the overall contamination from misidentified particles that normally plagues efficiency measurements. Other particle decays can be used as well, such as other clean resonances such as J/ψ, or certain, predictable decays such as dijet events.

The steps for performing a tag-and-probe study for leptons are:

- Look for events with two same flavor, opposite charge leptons that have a dilepton mass close to the Z boson mass.
- Define some very strict tag definition, and identify and label the leptons that pass this tag requirement (making the other lepton the probe).
- Using the probe, perform the efficiency measurement needed.

For most cases, our efficiency measurement is used to test a lepton ID, so it would take the form of measuring the efficiency of a loosely defined lepton passing the lepton ID.

We expect the efficiencies to be different between data and MC, so we need to correct for this dis-

crepancy. While the most accurate method would be to correct the simulation parameters directly, so our simulation matches data precisely, this is not feasible considering the numerous sources for efficiency differences as well as the pure logistics required to reproduce the simulation after every change in the parameter space. For most efficiency calculations, the difference in data and MC is less than 5 %, so the preferred method is using event weights. Event weights are simply scale factors applied to the overall weight for each event to "push" the MC kinematic shapes to better match what is observed in data. This basic procedure can be extended by having the event weight vary based on the particle's kinematics, such as its p_{T} and η. For the tag-and-probe method, extracting these event weights turns out to be simple:

$$w(p_{\mathrm{T}}, \eta) = \frac{\epsilon_{\mathrm{data}}(p_{\mathrm{T}}, \eta)}{\epsilon_{\mathrm{MC}}(p_{\mathrm{T}}, \eta)}$$

Where ϵ represents a given efficiency. More complex features can be added to this method, but the core remains the same. For every particle mentioned in this chapter, there are a suite of scale factors applied to every MC event to account for the difference in data-MC agreement for all the IDs used. Thankfully, since most of these IDs are commonly used in CMS by analyzers, there are dedicated teams, namely the ones that produce the IDs, that additionally produce the necessary scale factors. For IDs that are specific to a particular analysis, the analyzers themselves have to produce their own scale factors. For this analysis, we produced bespoke trigger and ttH−MVA scale factors.

It should be noted that these scale factors also carry uncertainty in them. First, the scale factors affect the overall weight of each event, so the weights need to be properly propagated to the statistical uncertainty of the MC samples. Second, the scale factors have systematic uncertainties associated with them. The groups that produce each scale factor studies these sources of uncertainty that enter into the measurement and propagate them to the scale factors to produce uncertainty bands. For utilizing these uncertainties, the analyzer can run their analysis with the up and down[†] shifted versions of the scale factors to create new event weights corresponding to the uncertainty bands, which, in turn, creates a "shifted" version of every kinematic variable representing the uncertainty for the particular scale factor. This is the heart of how we handle uncertainty estimation, which is discussed more thoroughly in Chapter 9

[†] up and down here means the $\pm\sigma$ variations for a particular systematic uncertainty.

Chapter 7

Background Estimation

The largest background in our signal region, as well as one of the largest concerns, is the so-called "fake" background. While the different algorithms used in reconstruction aim to remove particles identified incorrectly as meticulously as possible, it is inevitable that some background events will look enough like the signal to pass our requirements. For most analyses, these fake backgrounds are not large enough to be of much concern, but this analysis must be treated differently because of the same-sign requirement. While normally fake backgrounds can be ignored because the relative yields are much smaller than the total expected yield, the same-sign requirement reduces our background's yield enough that the fake backgrounds becomes substantive. Thus, the same-sign requirement simultaneously removes most extraneous backgrounds as well as necessitating specific modeling for fake processes.

To properly model these faking processes, we use a data-driven method.

7.1 Method

MC samples exist to model all possible decays (even those involving misidentified particles), but their quality can be limited. First, the number of events produced is typically lacking. The source of these fake backgrounds are those with large cross-sections such as $t\bar{t}$, W+Jets, and Drell-Yan. The signal region requirements combined with the small expected misidentification rate means the selection efficiency is low enough that typically, only a handful of raw events remain to model yields and kinematics. Second, these misidentification rates are not always modeled correctly. By definition, these incorrectly identified particles are abnormal since they arise from the reconstruction algorithms identifying particles incorrectly. This means the fake processes are only modeled as well our simulations can predict how particles will be misidentified. While the simulation does a fair job, it is currently limited. For these reasons, we turn to modeling these processes using the data itself. Modeling from data reverses both of these issues by producing the sample in a sideband region with more statistics and handling modeling better because

we are using data directly.[†] Not only this, but we gain the added benefit of reduced systematic errors for free. Because the data-driven background only has systematic errors coming from the background estimation method itself which, in the end, tend to be smaller than traditional MC uncertainties.

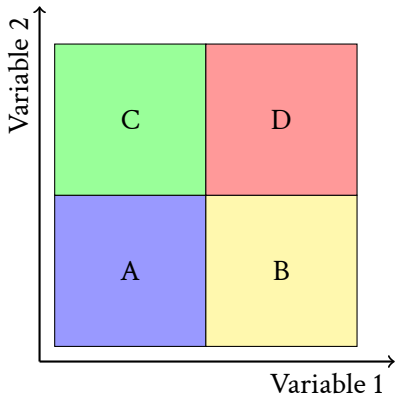

Figure 7.1: Example setup of the ABCD method. After a preselection is applied to the data, the phase space is split into four regions, (ABCD as shown), where these four regions are defined by some threshold on two variables (here, Variable 1 and Variable 2). In practice, the signal region, let us say D, can have its yield estimated by using ratios of the other three regions under the assumption that the kinematics of the region in A and C are similar to that in B and D, (i.e. across the requirement applied on Variable 1, and the same for Variable 2).

While switching to a data-driven method seems like the panacea, it comes with complications. The obvious problem is that data is not labeled for each process, so there is no way deciding which events came from a fake process and which didn't. To solve this, problem we used the technique called the ABCD method.

The ABCD method is simple: create a 2D phase space from two kinematic variables or criteria, and divide it up into four distinct regions using a threshold on the two variables, as shown in Figure 7.1 (labeled ABCD, where the method gets its name). From this, you can use ratios of the events in three of the quadrants to estimate the total in the fourth. The principle behind this is based on the assumption that basic kinematics should be the same across our phase space, especially the smaller, and thus more homogeneous, our considered space is. If our two variables are independent (or sufficiently so), we should expect the efficiency of passing or failing certain requirements should be the same across the space as well.

Put more concretely, after making an initial selection and defining a Variable 1 and Variable 2 that are independent of each other (using Figure 7.1 as our template), we should expect the follow relationships to hold:

$$\frac{N_C}{N_A} = \frac{N_D}{N_B} \qquad \text{and} \qquad \frac{N_B}{N_A} = \frac{N_D}{N_C}$$

[†]Because we are using data, there is no modeling used in the reconstruction at all. All potential forms of error thus come from the estimation method itself, as outlined further in the chapter.

Where N_X represents the number of events in X region. These relationships are expected to hold when the two variables are independent, so Variable 1 should look the same in Regions A+B as in Regions C+D, and similarly for Variable 2. So, if we identify region D as our final signal region, we can estimate the number of events by:

$$N_D = N_C \frac{N_B}{N_A}$$

This estimation method allows us to predict the background yield of some process using data without biasing ourselves by looking at the data in our signal region. Heuristically, we interpret region A and B as a sideband region where a "misidentification rate", or the efficiency of events that pass the Variable 1 requirement is measured. With this rate, we can "transfer" events from region C to region D using the ratio $w = \frac{N_B}{N_A}$. To summarize, Variable 1 acts as a definition for what are considered misidentified and Variable 2 acts to define our sideband region.

A clarification of nomenclature: in performing the background estimation, we use a transfer factor of $w = \frac{N_B}{N_A}$, but we measure the so-called misidentification rate in the sideband region where the misidentification rate is the efficiency for events with these misidentified particles to pass our signal requirements, defined as $f = \frac{N_B}{N_A+N_B}$. The misidentification rate is an efficiency, so it is bounded from 0 to 1 unlike the transfer factor, so it is much easier to interpret. When the transfer factor is needed for the actual estimation process, it can be reproduced using the relation:

$$w = \frac{N_B}{N_A} = \frac{f}{1-f}$$

Overview

For this analysis, there are two backgrounds that are estimated by this method: the nonprompt and charge-misId backgrounds. As the namings suggests, the nonprompt background models leptons that simulate the isolation/promptness of signal leptons, and the charge-misId background models leptons whose charge is mislabeled. For the generic pattern of the ABCD method, the requirement on the Variable 1 in our diagram acts to remove all "bad" leptons, but can't totally, so the misidentification rate aims to measure of the percentage of misidentified leptons that pass this requirement.

This does mean that measuring the misidentification rate in the sideband region, the misidentification rate should only be measured on nonprompt leptons, so some precaution needs to be taken to ensure the purity of leptons. The following sections describe of the extraction and validation of these misidentification rates:

7.2 Nonprompt Lepton

Nonprompt leptons are defined as leptons not originating from the main process but rather in a following cascade decay. Usually these leptons come from jet decays where the lepton would be embedded in the jet cone and typically soft since they take only a small portion of the jet's total energy. As was discussed in Section 6.1, we remove the nonprompt leptons mainly through requirements on the mini-isolation and ttH−MVA. While this remove a majority of them, the overall nonprompt contamination is quite appreciable. These nonprompt leptons, coming primarily from large cross-section backgrounds such as $t\bar{t}$, W+Jets, and Drell-Yan, allow these processes to pass the same-sign requirement since a nonprompt lepton can be produced from any jet and thus has an even chance of matching the charge of the other lepton in the event. Even for a tight baseline selection, the rate is appreciable and requires dedicated modeling.

For applying the ABCD method, there are two primary steps: choosing the defining variable (i.e. Variable 1) and then measuring the misidentification rate in a dedicated sideband region (i.e. region created by Variable 2). The ttH−MVA worked to distinguish prompt from nonprompt leptons, so it is use for defining the fake and tight categories for the lepton. For defining our sideband region where the misidentification rate is measured, we choose a jet enriched region, or what is typically called a QCD multijet region[†] because it will have plenty of events which, if they contain leptons, should come solely from jets.

While we have the skeleton for a misidentification rate measurement, we include extra steps to improve the accuracy of our misidentification rate measurement. First, there is a bias induced in our nonprompt leptons based on what sort of jet a lepton is produced from. To account for this, the fake categorized leptons have their p_{T} corrected. Further, in our QCD multijet region, there is prompt contamination. We can subtract the expect number of prompt events when calculating the misidentification rate, but our measurement will rely fully on our prompt estimation. To mitigate this, we correct the expected prompt yield by fitting it to data in a dedicated region. Lastly, to confirm our overall misidentification rate, we compare our data-driven results to expected results in a closure region. With these tweaks, we have the outline for a full misidentification rate measurement:

1. Derive a fake lepton p_{T} correction
2. Fix the prompt lepton yield by fitting it to data in an addition "fit" region
3. Measure the misidentification rate by subtracting the fixed prompt lepton yield in our measurement region
4. Test the overall data-driven method by comparing it to expected yields in a closure region

[†]For processes that contain no tree level electroweak interactions, only strong interactions, they are called QCD events. These events are expected to have many jets and no real leptons.

7.2.1 Momentum Correction

To start, we first have to address a source of bias. Nonprompt leptons mainly come from jets and are expected to have a different p_T spectrum compared to prompt lepton, but the type of jet the lepton originates from also has an effect. If the lepton is produced from a light flavor jet versus a b-jet, the fraction of p_T the lepton takes will be different. This is not a problem in principle, but in applying our misidentification rate, we may produce the wrong kinematics because the jet flavor in our measurement region will be different from in our signal region. Without addressing this issue, our background will be skewed by the flavor composition as well as the difference in p_T fraction between the different jet flavors of the mother jet to the lepton.

To remove the bias, we could use the mother jet's p_T (or the p_T of the closest jet) as opposed to the lepton's p_T to remove the flavor dependency, but this replaces one problem with a new one, namely the mother jet having a much larger momentum than real measured lepton's p_T. So, we define the new fake lepton p_T as a proportion of the closest jet's p_T, rescaling the jet p_T by a constant to match the observed tight lepton's p_T. This gives us our corrected fake p_T as follows:

$$p_T(\ell) \rightarrow f \cdot p_T(\mathrm{j}) = f \cdot \frac{p_T(\ell)}{p_T^{\mathrm{ratio}}}$$

Where $p_T^{\mathrm{ratio}} = p_T(\ell)/p_T(\mathrm{j})$ or the ratio of p_T the lepton takes and f is the factor used to rescale our fake p_T to match our tight p_T.

For calculating this scaling factor, f, it is done by fitting the jet p_T to the lepton p_T across the ttH−MVA spectrum. The ttH−MVA includes the lepton's p_T as one of its inputs, so we expect the average p_T of lepton to change with different MVA scores, therefore fitting the running average p_T of the two types of leptons would cause mismodeling. Instead, we look at the average p_T for different ttH−MVA values as in the top plots in Figure 7.2. By looking at the average of the p_T, the discrepancy between the jet p_T and the tight lepton p_T becomes evident, where the discontinuity represents the difference in p_T definition at the threshold of fake and tight leptons. Choosing the scale factor, f, is a matter of finding a value that allows the average p_T's to match at the fake-tight defining ttH−MVA value. To avoid any noise skewing the results, a linear fit is done for values around the threshold for both the distribution, and the scale factor is fixed such that the fit lines match. This fitting procedure removes some noise but is sensitive to the manner the fit is done (range in which the fit is done, order of fitting function, fitting only one-sided vs two, etc.). So, we limit the final p_T correction factor to steps of 0.025 to remove any spurious fine-tuning. The result of the fit are seen in bottom plots of Figure 7.2 where the average p_T now aligns for the fake and tight leptons.

It should be noted that in our fake lepton definition, there is an additional requirement on the p_T^{ratio} to reduce the number of soft leptons. A lepton could come from a high energy jet but only take a small percentage of its energy which would lead to the soft lepton having a corrected p_T much larger than its

original p_T. The isolation and ttH–MVA requirements remove most of these very soft leptons, but by requiring $p_T^{ratio} > 0.4$, the rest are removed and none can skew the calculation of the p_T-correction factor. After the p_T-correction factor is calculated, the new p_T is used in place of the nonprompt lepton's origin p_T for all kinematic requirements and calculation, with the small exception of the trigger requirement. Because the trigger requirement is based on the originally calculated p_T, the fake leptons must pass the usual trigger requirement as well as a loose baseline on the "raw", or true, p_T to ensure our leptons would pass our triggers, even without the p_T correction.

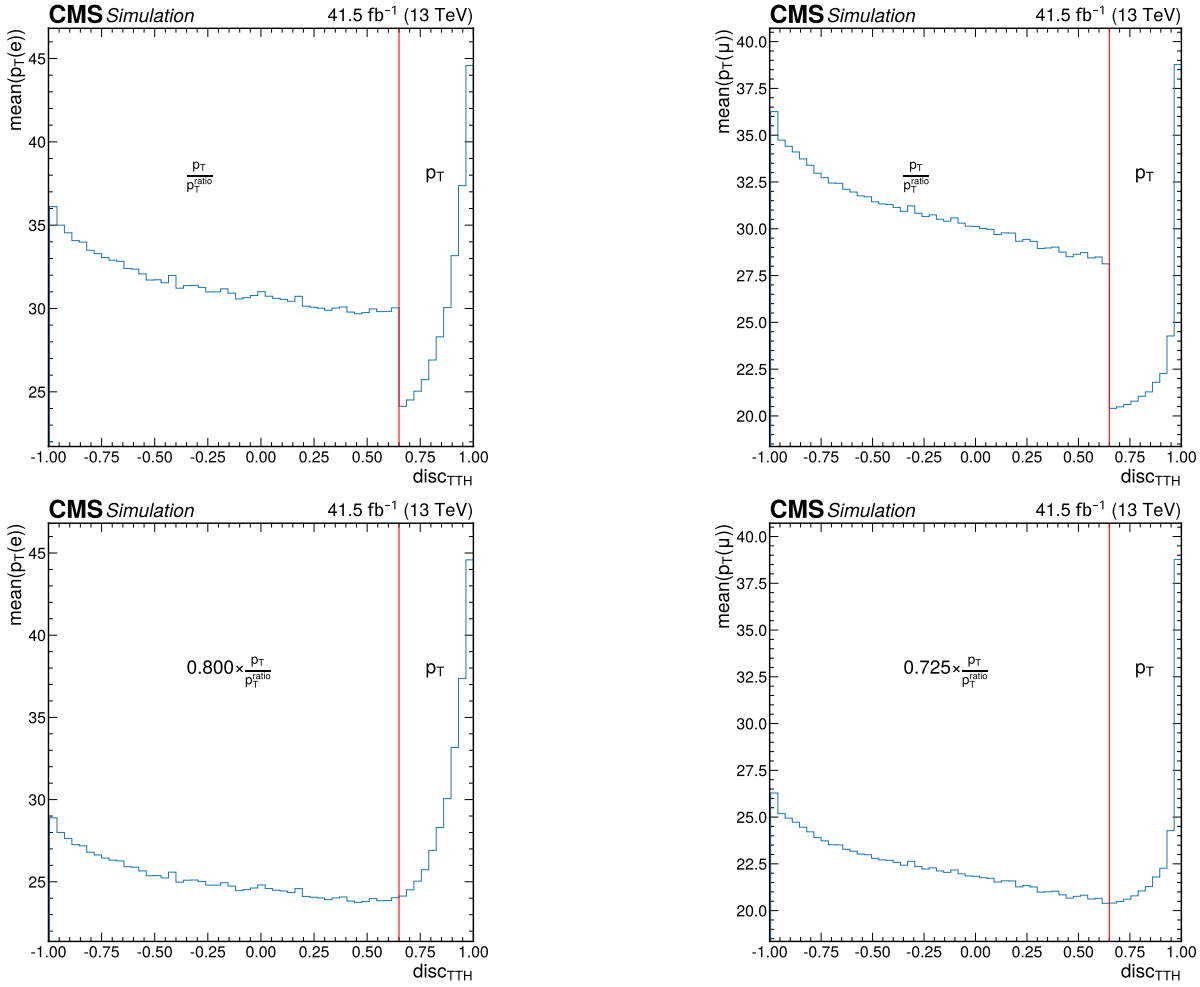

Figure 7.2: Example of the p_T correction factor for 2017 samples. Above is the average p_T for the ttH–MVA spectrum, noting that p_T used for ttH–MVA < 0.65 is the mother jet p_T while for ttH–MVA > 0.65, the lepton's p_T. Below shows the fake lepton's p_T correction to match the tight lepton's p_T as needed. The plots are for both lepton types: electrons on the left, muons on the right.

7.2.2 Template Fit

After we fix the fake lepton's p_{T}, we can start to create our measurement region. Our measurement region requires: one lepton, at least one recoiling jet (to balance the momentum), and QCD specific requirements, more specifically, low MET. We need to ensure we are measuring only using nonprompt leptons, so keep high purity, we opt to subtract the prompt MC samples from data to best approximate only the nonprompt contribution in data. We might wish to simply tighten our requirements until our measurement region contains no prompt leptons, but this has the same problem in reverse to our signal region. To remove all the prompt leptons, in the end, there would be little nonprompt leptons for our misidentification rate measurement. Subtracting the prompt contribution adequately addresses this problem as long as our MC can model the actual prompt lepton contribution, so we settle on fitting the prompt contribution to data to ensure proper modeling of the MC samples. The backgrounds that correspond to the prompt contribution in our QCD multijet region are W+Jets, $\mathrm{t\bar{t}}$, and Drell-Yan, where W+Jets dominates. We collectively call these samples EWK while the jet dominated events that contain the nonprompt leptons are called QCD. The EWK is simulated with NLO samples and is generally well modeled, but our fit allows for minor fixes to the overall normalization in the final prompt subtraction. For fitting the EWK, both MC samples for EWK and QCD are fit to data, but we only allow their overall yields to change in what is called a template fit (since we are fitting the overall template shapes of our samples). This fit is also performed in a region orthogonal to our measurement region defined by inverting our MET requirement, so the region is enriched in prompt leptons. EWK samples are expected to have little to no fake leptons, so this fit is also performed only on tight leptons. In summary, the requirements for our template fit region become:

- Pass single lepton trigger
- One tight lepton
- $N_{\mathrm{j}} = 1$ with $\Delta R(\mathrm{j}, \ell) > 1$
- $E_{\mathrm{T}}^{\mathrm{miss}} > 30\,\mathrm{GeV}$

The two MC groups, QCD and EWK, are fit to data using with the m_{T} distribution which is defined as the transverse mass of the lepton with the MET, or:

$$m_{\mathrm{T}} = \sqrt{2 p_{\mathrm{T}} E_{\mathrm{T}}^{\mathrm{miss}} (1 - \cos \Delta\phi)}$$

This variable emulates the W boson's mass since a leptonic W boson decay will produce a lepton and neutrino (which we model the MET). Further, we expect our largest contribution in the EWK samples to be W+Jets, so there should be a broad, but noticeable peak at the W boson mass or around 80 GeV. This is in contrast to QCD which has no real, prompt MET, so the m_{T} should look similar the lepton p_{T} spectrum, i.e. an exponential. The two groups have distinct shapes, so the fitting should have no problems related to correlation. For the fitting, we analytically find the two normalization factors that minimize

the chi-squared with data. The fit is performed for different p_T and η ranges to account for differences in the scaling in those variables, so we use the same $p_T - \eta$ binning as the actual misidentification rate. As expected, the QCD sample has an exponential shape while the EWK peaks around the W boson mass, becoming more prominent at higher p_T as the lepton takes a larger proportion of the W boson decay energy. The results of this fitting process are shown in Figure 7.3. The fit has EWK template fit values close to 1, as expected with only minor deviations in the phase space. The QCD has much poorer template fit value, but we expect this considering QCD essentially is our nonprompt background, showing the necessity for dedicated modeling. The QCD background is only used for its shape information, so this means the larger deviation in the QCD template fit values is of no large consequence for our EWK fit.

7.2.3 Measured Misidentification Rate

Now that the pieces are set, the actual misidentification rate measurement can be performed in the measurement region. For the measurement region, we require one lepton, at least one recoiling jet, and low MET, and add an addition requirement of low m_T to help further reduce the prompt contamination. Our signal region should contain lots of b-jets as well to select the three top signal, so the nonprompt should originate primarily from $t\bar{t}$ as opposed to our measurement region which is dominated by nonprompt from generic QCD induced decays. These QCD events chiefly contain lighter jets, so we should expect our nonprompt decays to be primarily initiated by lighter flavor jets as well. To further reduce this bias, we can force our QCD events to include a larger proportion of b-jets by requiring at least one of the recoiling jets in the event (i.e. a jet with $\Delta R > 1$ from the lepton) to also be a b-tagged jet. This naturally forces a larger proportion of the QCD events to come from $b\bar{b}$ decays (where the second b quark becomes the nonprompt lepton), more closely matching the expected environment of the signal region. The specific working point used for the b-tagged jet was decided by looking at the overall agreement in the closure region (discussed in the next section), minimizing the chi-squared between the expected nonprompt yield using the ABCD method and MC. In summary, our measurement region is defined as:

- Pass single lepton trigger
- One fake or tight lepton
- At least on recoiling jet ($\Delta R(j, \ell) > 1$), passing loose b-tagging WP
- $E_T^{miss} < 30\,\text{GeV}$
- $m_T(\ell) < 55\,\text{GeV}$

To measure of the misidentification rate, events are split up into different p_T and η bins where the misidentification rate is measured in each individually. Doing so better captures the kinematics associated with the misidentification rates. We would expect considering reconstruction algorithm will have different efficiencies for different p_T and η values more generally because of the geometry of the detec-

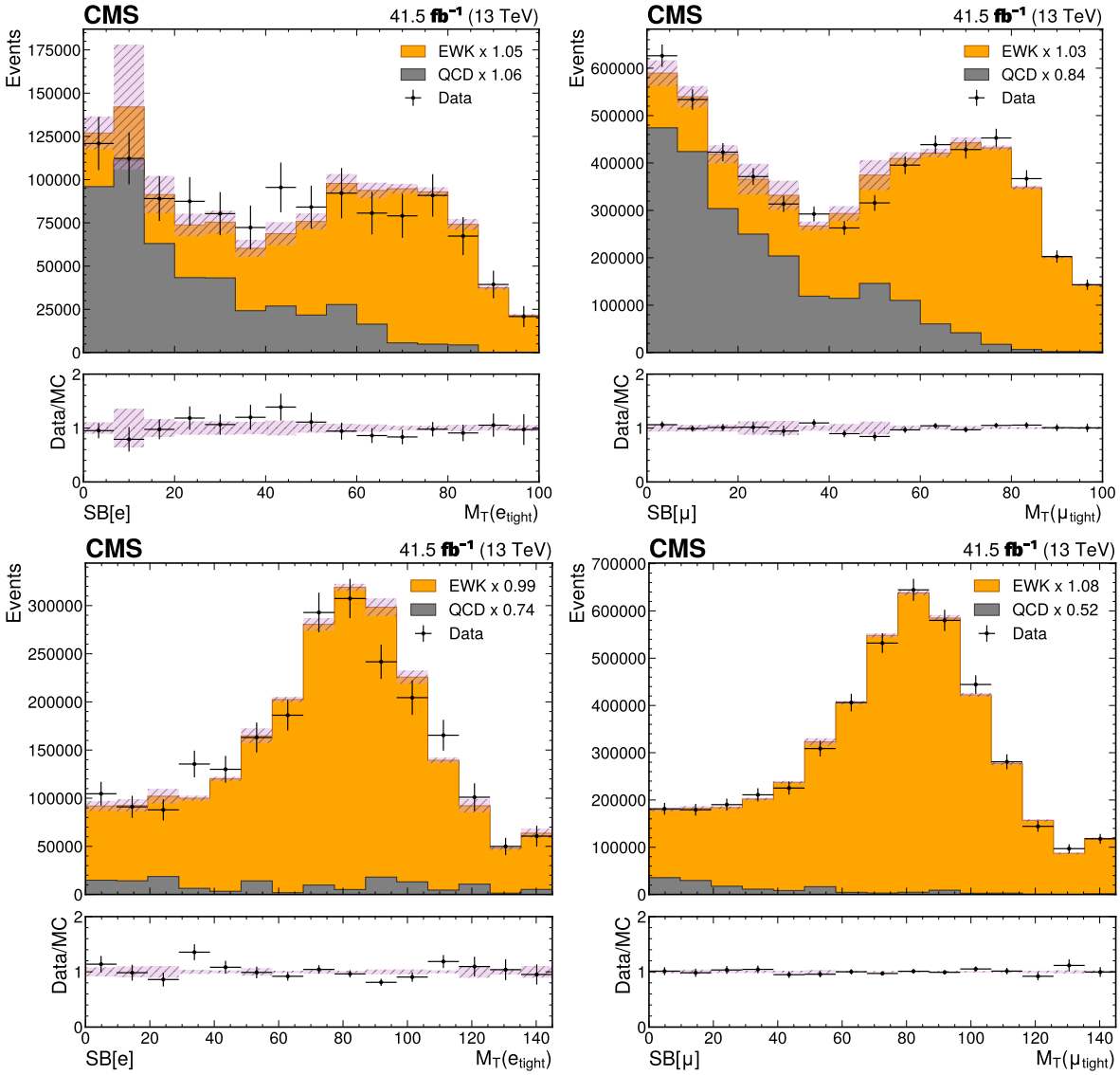

Figure 7.3: Representative m_{T} plots from 2017 data with events with electrons on left and muons on right. The top plots show low p_{T} fits, containing leptons with 20 GeV $< p_{\mathrm{T}} <$ 25(30) GeV for electrons (muons) while the bottom shows leptons with $p_{\mathrm{T}} > 50$ GeV. Both rows show leptons with $|\eta| < 0.8(1.2)$ for electrons (muons). The other $p_{\mathrm{T}} - \eta$ plots are omitted for space, but show similar fitting.

tor, algorithms depending on these variables, as well as many other causes. In each individual p_{T}, η bin, the misidentification rate is measured by taking the ratio of tight passing nonprompt leptons to the total nonprompt leptons where nonprompt leptons are modeled by the total data with the rescaled EWK MC

subtracted out. This can be summarized as:

$$f(p_{\mathrm{T}}, \eta) = \frac{N_T^{\mathrm{data}} - f \cdot N_T^{\mathrm{EWK}}}{(N_T^{\mathrm{data}} + N_F^{\mathrm{data}}) - f \cdot (N_T^{\mathrm{EWK}} + N_F^{\mathrm{EWK}})}$$

As a cross-check, the misidentification rate can also be measured using exclusively the QCD MC, but the low statistics and poor observed modeling in the template fit means we expect low precision, but it can give a rough idea of total agreement. With this, we have derived the misidentification rate, as shown in Figures 7.4.

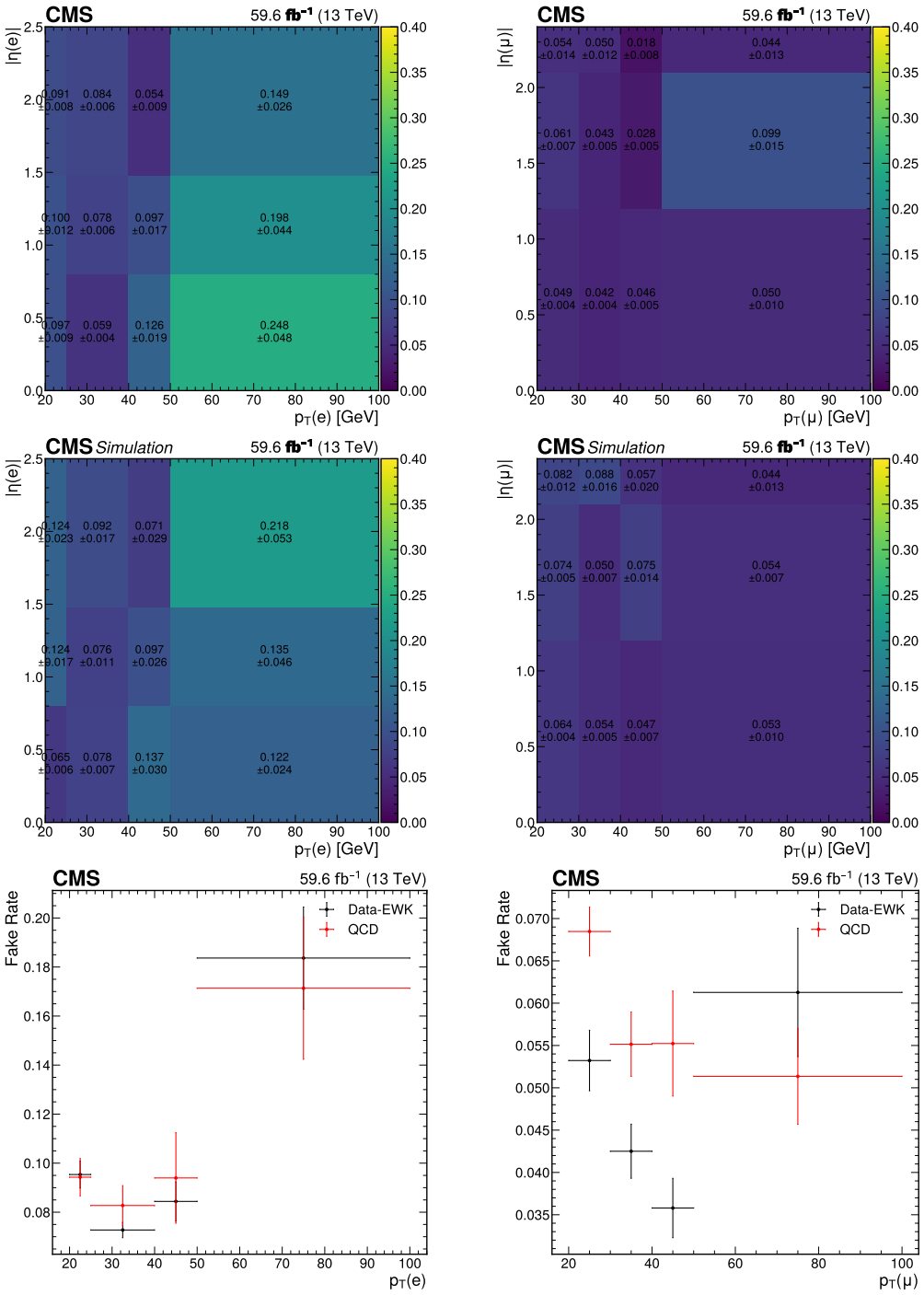

Figure 7.4: This shows the misidentification rate for 2018. Top row shows the rate as calculated using Data – EWK. The second row shows the misidentification rate as calculated using just QCD MC. The last row shows a comparison of the two methods in the p_{T} spectrum. On the left is the misidentification rates for electrons while the right shows the misidentification rates for muon.

7.2.4 Closure Test

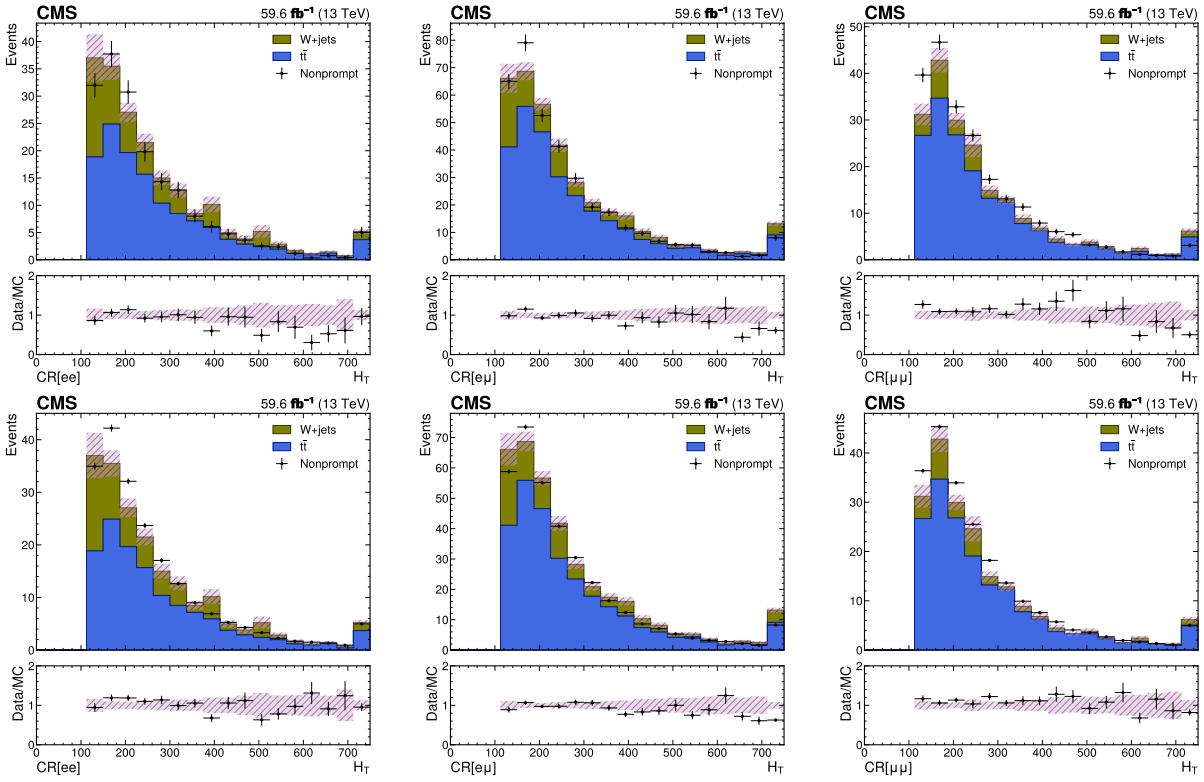

Figure 7.5: Closure plots for 2018 considering all three channels, ee (left), eμ (middle), and $\mu\mu$ (right). For the top plot, the nonprompt estimation is derived from data while for the bottom plot, the nonprompt estimation is derived from MC. For both methods, the estimation comes from applying the misidentification rates in the FF and TF regions to transfer to our closure region.

While the misidentification rate has been derived in a rigorous fashion, a cross-check should be performed to ensure it works well in our signal region and does not have any induced bias that must be addressed (such as the issue of $\mathrm{t\bar{t}}$ mentioned). This is done in the form of a closure test. We define our closure region to be similar to our signal region to ensure good performance for our signal extraction. The exact requirements are:

- Pass dilepton trigger
- Two same-sign leptons (fake or tight)
- $M_{\ell\ell} < 12\,\mathrm{GeV}$ and $|M_Z - M_{\ell\ell}| < 15\,\mathrm{GeV}$
- $N_j \geq 1$
- $H_T \geq 125\,\mathrm{GeV}^\dagger$
- $E_T^{miss} > 50\,\mathrm{GeV}$

†The H_T variable is a measure of hadronic activity in an event and is defined in Section 8.2.

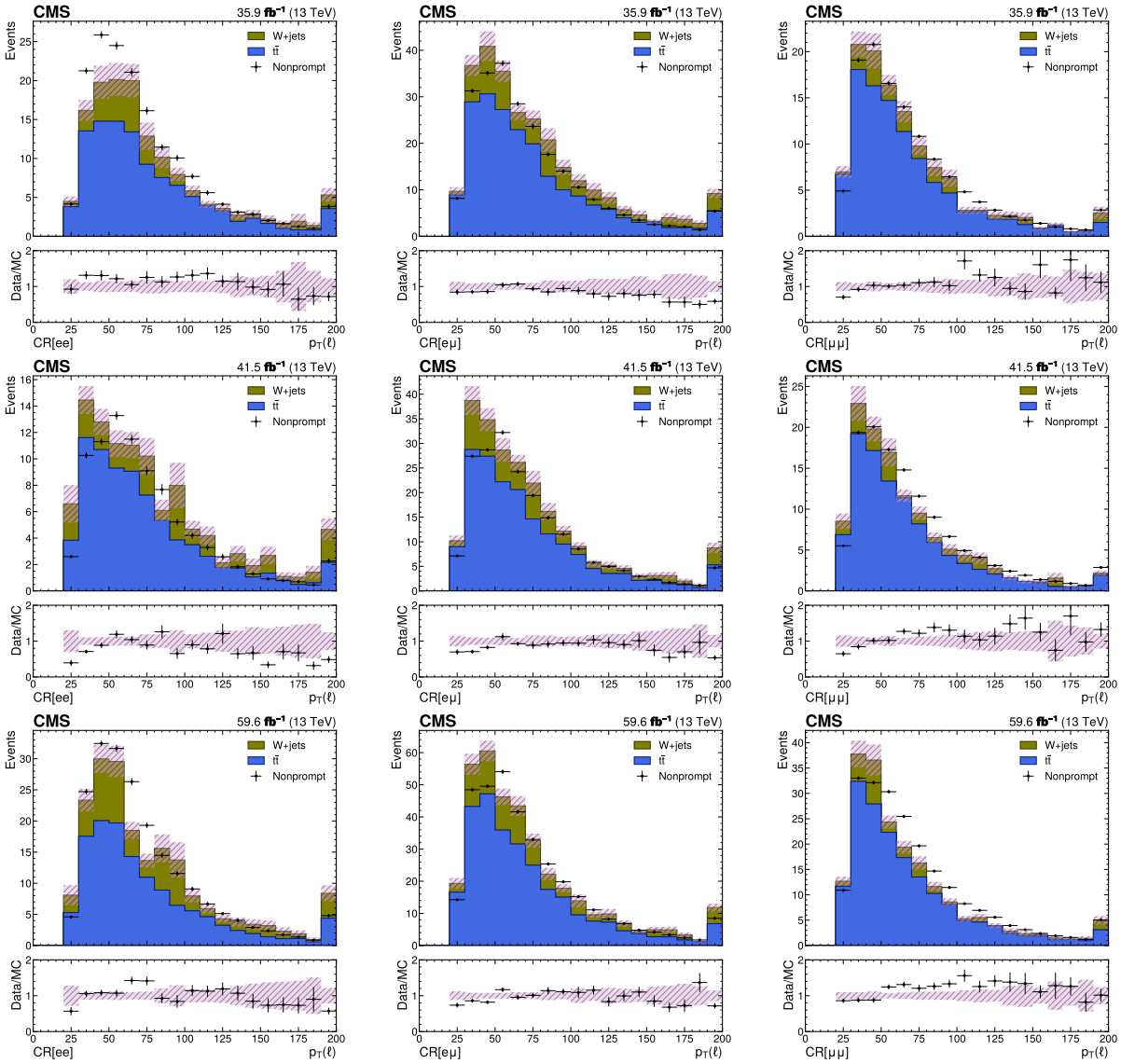

Figure 7.6: Closure plots of the p_{T} distribution for 2016 (top), 2017 (middle), and 2018 (bottom) considering all three channels, ee (left), eμ (middle), and $\mu\mu$ (right). Nonprompt here is derived from MC, applying the misidentification rates in the FF and TF regions to transfer to our closure region, so the comparison is MC vs MC.

Ideally, a closure test is performed using data from our transfer regions, fed through the ABCD method, and compared against the actual data. This proves to be difficult because we have no way to isolate just nonprompt event in data, as demonstrated by the complications in our the prompt subtraction. Instead, we opt to use two methods: compare data processed with the ABCD method against the MC and the MC processed with the ABCD method against MC estimation. The first method has the problem of

being an apples to oranges comparison, because the nonprompt modeling is not expected to be correct in MC, but we can get a basic comparison in our relaxed phase space. The second method comparing MC against MC, while carrying the same inherent flaws originating from using MC for modeling, is at least self-consistent.

To create our data-driven estimation, we must take events from regions containing fake leptons and "transfer" them to the signal region using the weights. For our two lepton closure region, this means our transfer regions are those with two fake leptons (FF), or one fake lepton (TF), while our signal region is that with two tight leptons (TT). There is some slight nuance with applying these transfer factors, especially when generalizing to the multilepton cases (i.e. TTT and TTTT regions) that are discussed in Appendix E.1.

The nonprompt background comes primarily from $t\bar{t}$ with W+Jets is subleading[†]. Our final closure comparisons are shown in Figures 7.5 and 7.6. Looking at the H_T distribution, which is unrelated to our expected nonprompt measurement, shows good agreement across the two methods giving us confidence. For the additional p_T distribution from the MC-MC test in Figure 7.6, we notice larger tensions in lower p_T bins, especially for the electron only channels. This is a by product of lower yields in the measurement region for electron only events as well as larger difficulties with electron reconstruction. Overall, the agreement is fair and can be managed by our uncertainties in the final fit, especially since we expect our signal to be dominated by higher p_T leptons as a byproduct of the large top quark mass.

While the data-driven method has no MC modeling systematic uncertainties, uncertainty comes from the ABCD method itself. There are multiple, probable sources of induce error, but we opt for a conservative method using a flat envelope uncertainty for the whole method. This uncertainty can be estimated with the closure test itself from the observed difference between the expected rate and the data-driven estimate. Looking at the overall agreement, a conservative 30 % flat uncertainty was chosen. The statistical errors of the misidentification rates themselves, while they can be enveloped into the flat uncertainty, are included to allow the misidentification rate in our final fit to vary based on lepton flavors.

7.3 Charge Misidentification

The major requirement for the signal region is two same sign leptons which removes the large $t\bar{t}$ and Drell-Yan backgrounds, but this is dependent on the two leptons having their charge identified correctly. A lepton's charge is primarily determined by the direction of the curvature of the lepton's track. Muons are reconstructed as global tracks, and they generally have suppressed bremsstrahlung, so their charge identification is quite good. Electrons only have the tracks from the silicon detectors for charge reconstruction and their smaller mass means they are more susceptible to bremsstrahlung decay. The tracks

[†]Drell-Yan, while one of these large cross-section processes, is primarily removed by the Z boson mass veto and should have little impact unlike the other samples.

are reconstructed using a Gaussian-sum filter algorithm which account for decays that divert the electron's path and momentum, but it is not precise enough to reduce the overall charge misidentification rate to zero. While the rate is small (around 10^{-4}), large cross-sections processes can offset it to generate a nontrivial number of events incorrectly misidentified as same-sign.

The MC modeling of the charge misidentification is typically poor because of the lack of events to properly model any kinematic shapes, so a data-driven method is used. The overall charge misidentification rate is smaller than the nonprompt misidentification rate by almost two orders of magnitude, so the expected charge-misId background is important but will not be the major background like the nonprompt. The nature of the faking mechanism does mean that the charge-misId background is only necessary for the same-sign dilepton channel and not the multilepton channel. For the dilepton channel, opposite sign processes such as $t\bar{t}$ can fake our signal requirements if an electron's charge is misidentified. For the multilepton region, our signal process only allows for mixed charge configurations, namely $(+,+,-)$ and $(+,-,-)$[†], i.e. all leptons must not have the same charge. For an event to fake the multilepton requirement, an electron's charge must be mismeasured from an event with three same-charged leptons, but this process has a lower cross-section than our signal and other expected Standard Model processes. Therefore, the charge-misId background is only estimated for the dilepton region.

The charge-misId background is estimated using the ABCD method. For defining the four regions, the misidentified defining variable (or Variable 1 as in Figure 7.1) represents same-sign or opposite-sign lepton events while the Variable 2 which sets the sideband regions is defined later. The difficulty in performing this measurement come in isolating and identifying leptons whose charge has been misidentified, especially in data. To avoid this complication, our charge-misId rate is performed in MC, and a closure test is performed to correct for any observed problems in the agreement. Putting this together, the steps for performing this study are:

1. Measure the charge-misId rate in MC
2. Create a closure region and apply the transfer factors to the data to create the estimation
3. Correct overall normalization issues stemming from MC only misidentification rate measurement

7.3.1 Measurement

The first hurdle for measuring the charge-misId rate is how to actually measure it because there is no easy way of distinguishing true charge misidentification in data. In MC, it is trivial because we have the generator or true details of the event, so the charge measurement's validity can be gathered by comparing the generator and reconstructed charges. If one wanted to measure this in data, one can isolate only truly misidentified events, but contamination and lack of statistics prove to be a similar issue to our other misidentification rate measurement. For the nonprompt, the misidentification rates ranged from 5 %

[†]With the tttW process, we can have events with four leptons since the additional W boson allows for another prompt lepton. The four lepton case adds the charge configuration of $(+,+,-,-)$, but the same point generalizes to this case.

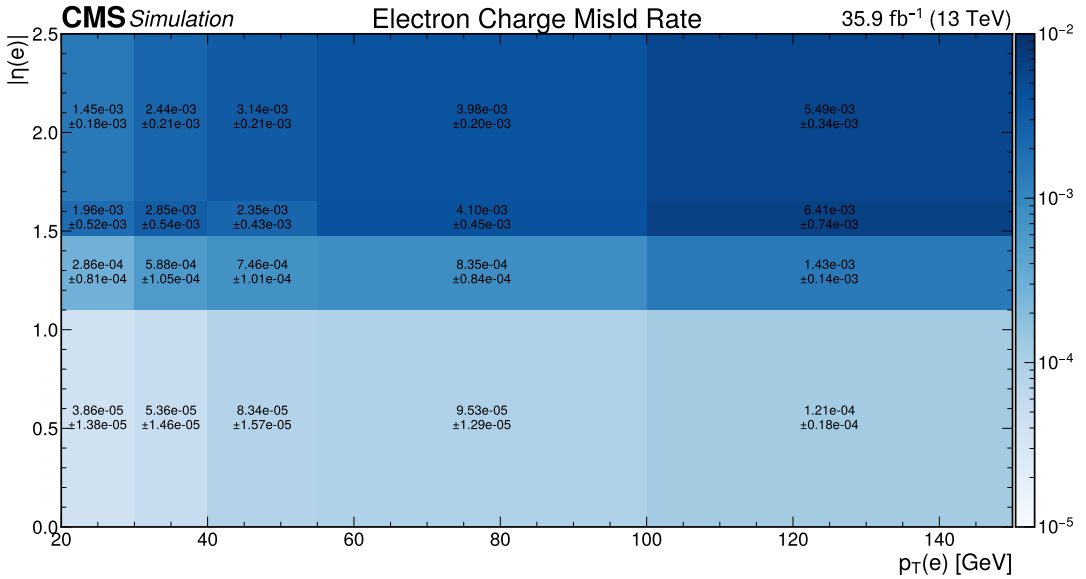

Figure 7.7: Example plot of the charge misidentification rates for 2016.

to 20 % depending on the kinematic region, but the charge-misId rate is orders of magnitude smaller, so uncertainties in isolating and removing contamination can be quite large. Because of these issues, it becomes preferable to measure the charge-misId rate using MC and then adjust the misidentification rates in a closure test to better match the data. This is justified since we expect MC should model the overall kinematics well, but might not produce the correct rate. Unlike the generation of nonprompt leptons, the mechanics for a lepton charge misidentification is much simpler, so the MC can model well enough, but the observed versus expected rate may differ. If the MC can produce the correct kinematic shapes for the charge-misId background, then the rate measured in MC only needs to be adjusted for its overall normalization.

Because we will being using the MC to measure the actual charge-misId rate, we have flexibility in choosing our measurement region because there is no purity issue. We know a majority of the misidentified events are coming from the large cross-section backgrounds (because they are the only backgrounds that can survive after the same-sign requirement), so the misidentification rate can be measured directly from those corresponding MC samples. To keep the misidentification rates as representative of our final phase space, we can just use a loosened version of our signal region. In the end our requirements for the measurement region are:

- Pass dilepton trigger
- Two tight leptons
- Veto events with more than two loose leptons
- $M_{\ell\ell} > 50\,\text{GeV}$

- $N_j \geq 2$
- $H_T > 100\,\text{GeV}$
- $E_T^{miss} > 25\,\text{GeV}$

Here, the H_T requirement is quite loose to increase the statistics of our main backgrounds. We additionally require the dilepton mass to be above 50 GeV to match the requirements used in the matrix level simulation for some backgrounds, namely the Drell-Yan. The actual measurement of the misidentification rate is performed using a binning dependent on p_T and η, similar to the nonprompt misidentification rate. The charge is measured by the curvature of the particle in the tracker, so the larger the p_T, the more difficult it should be for the charge to be reconstructed. Meanwhile, the detector is made as a cylinder, so measurements of momentum in the transverse plane are good while less accurate towards the beam-line, and thus making our misidentification rate depend on η. This, combined with difficulties in measurement at the interface between the barrel and the endcap (at $\eta \approx 1.45$ as seen in Figure 3.3), leads to larger misidentification in the "crack" as well as the endcap regions. The results of the misidentification rate measurements are shown for 2016, binned in p_T and η, are shown in Figure 7.7.

7.3.2 Closure Test

While the charge-misId rate is measured in a signal like region, for validation of the misidentification rates, we perform it in a $Z \rightarrow ee$ region. This Z boson region has lots of events due to its large cross-section and the mass peak isolates high purity electrons. To reduce other background contamination, we require two electrons that have a dilepton mass close to Z boson mass, a small MET, and low H_T. Even with these requirements, it is expected that there is a small amount of contamination from backgrounds such as diboson, but a large enough majority are Drell-Yan for our comparison. In summary, this gives the following requirement for our charge-misId closure region:

- Pass dilepton trigger
- Two same-sign tight electrons
- Veto events with more than two loose leptons
- $70\,\text{GeV} < M_{ee} < 115\,\text{GeV}$
- $H_T < 250\,\text{GeV}$
- $E_T^{miss} < 50\,\text{GeV}$

While the closure test is performed in a region with two same-sign leptons, the charge-misId background is transferred from the region with opposite sign leptons. The specifics of how the transfer factor is calculated are described in Appendix E.2. Before looking at the closure agreement, the previous point on the misidentification rate being measured from MC must be addressed. To fix the MC derived misidentification rates, we scale these rates based on the observed yield difference between the data-driven background and data itself in our closure region. We normally would have to consider the

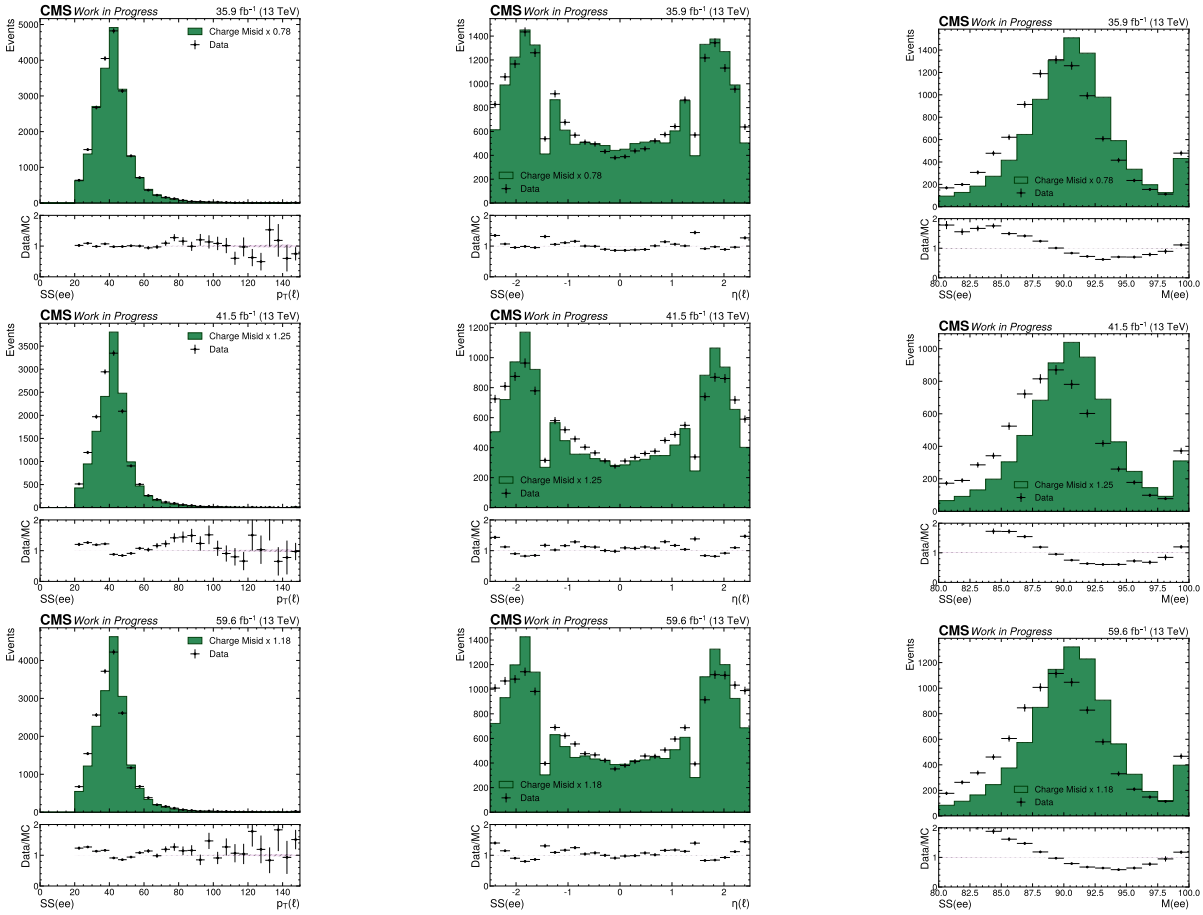

Figure 7.8: Closure of the charge-misId for 2016 (top), 2017 (middle), and 2018 (bottom) for different parameters of interest. The plots show the combined electron p_{T}, combined electron η, and dielectron mass (from left to right).

difference in the misidentification rates and transfer factor for this rescaling, but because of the very small charge-misId rate, we lose no accuracy by rescaling our misidentification rate using the naive method:

$$f \to f \cdot \frac{N_{\mathrm{data}}}{N_{\mathrm{MC}}}$$

The scale factors are done per year and are shown in Table 7.1.

Looking at the closure between the data-driven background and the actual data yields in Figure 7.8, agreement, while overall fair, has some notable discrepancies. For example, for the dilepton mass, the peak is slightly different between the data-driven method and data. This is expected because we are using normal electrons to model the momentum of electrons that underwent a bremsstrahlung decay strong enough to disrupt the charge reconstruction. The misidentification rate being binned in p_{T} helps mitigate this, but the disrupted track affects the p_{T} and angular reconstruction which causes a slight

Table 7.1: Data/MC scale factors for the charge-misId rates for all three years. Because the misidentification rate is measured in MC, the agreement is expected to be different from in data. This is compensated by scaling the misidentification rates by the expected charge-misId yield (using the ABCD method) and the observed value in our closure region.

year	obs/pred
2016	0.78
2017	1.25
2018	1.18

shift certain variables such as the dilepton mass peak. This can be seen by the observed mass which is slightly less than the expected Z boson mass at ~ 91.2 GeV. While this modeling issue exists, the overall discrepancy is small for this subleading background, so we expect a limited effect on the final result. The η distribution also shows slight differences, especially around the barrel endcap interface, suggesting the performance might improve with finer binning in the η distribution or better modeling of the detector at this interface, but considering the number of events used for the background estimation and the expected yield, this discrepancy is within an acceptable margin.

Since the background is modeled using data, we should expect systematic uncertainties to only come from the ABCD method itself. With the nonprompt background, the systematics were modeled to an overall flat, enveloping systematics and a pair of statistical errors to account for misidentification rate differences between the two lepton flavors. Only the electrons are modeled for the charge-misId background, so it is sufficient to use a single, conservative, flat uncertainty that would contain all sub-uncertainties. Again, we estimate it by looking at the overall agreement in the closure region, and we chose a value of 20 % for the total systematic error.

Chapter 8

Analysis Overview

With the requisite pieces in place, namely object definition and data-driven backgrounds, we may now start apply our selection to data to distinguish signal from background. Inevitably, there will be background processes that look too similar to our signal, so removing all the background is simply not feasible. Typically, for smaller and less complicated analyses, one can just apply multiple requirements on the events and use the final, total yields for the statistical testing in what is called the cut-and-count method. This is too limiting for our case because our signal is contaminated by large background processes which overwhelm our relatively small signal yield. Another method uses MC to construct a probability density functions (pdf)[†] of a kinematic variable that is compared to data. This process, called an unbinned fit, is quite robust but complicated, in particular in creating the fit model and error evaluation. A compromise of these two methods is using template histograms.

After creating our signal region, i.e. a phase space with a large signal proportion, we create histograms of certain kinematic distributions where the background and signal are distinct for our statistical tests. By using histograms, we essentially expand the single "bin" of the cut-and-count method, but avoiding the complexity induced by constructing a pdf from scratch. Instead, the histograms are used as a template for a pdf, merging the methodologies of the two competing methods. By using histograms, it additionally allows us to leverage the general shape information to separate signal from background as well as add constraints to background processes. Now to build the analysis, the primary steps are creating the signal region and then the template histograms used for the statistical tests

8.1 Processes

Before we can create our signal region, we first need to know what MC samples should be included so we can most efficiently remove them. Ideally, every Standard Model process would be simulated and fed into

[†]Annoyingly, parton distribution functions (PDF) have the same acronym as probability density functions (pdf), but for most cases, it is very apparent when one versus the other is being used. Regardless, to help distinguish the two, the capitalization rule shown here are used to clear any possible confusion.

the analysis, but in reality, only a subset are actually included for ease. MC samples at CMS are centrally produced and validated by experts, so they are made based on expected needs for any given analysis. Large cross-section backgrounds are made with hundreds of millions of events while more rare processes can have less than a million events. Therefore, any backgrounds will have an uncertainty larger than their expected yield and can generally be ignored. For example, a large cross-section background such as QCD multijet typically dominates because of its large cross-section, but the same-sign lepton requirement only leaves a couple to even zero raw MC events. Any effect of this sample is indistinguishable from general error in our modeling, thus it justifies removing them since they have no real bearing on our results.

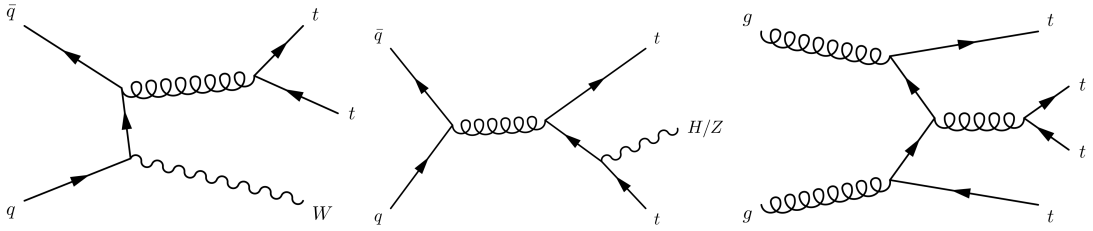

Figure 8.1: Feynman diagrams for the main background processes, which, from left to right, are: ttW, ttH or ttZ, and $t\bar{t}t\bar{t}$. ttH/ttZ are grouped together because their production mechanisms are very similar, and for this Feynman diagram, are interchangeable. All these processes can pass our same-sign requirement because of the top quarks and bosons in the event making them more difficult to remove from our signal region. The $t\bar{t}$+XY processes are excluded here, but have similar production mechanism, but with an extra boson compared to the $t\bar{t}$+X processes.

The first class of processes to consider are the irreducible backgrounds or processes that look similar to the signal and are thus not easily removed. We expect these to be backgrounds which contain top quarks. Those that contain top quarks are: single top, $t\bar{t}$, $t\bar{t}$+X, $t\bar{t}$+XY, and four top[†]. Single top and $t\bar{t}$ are generally excluded because of the same-sign lepton requirement, but $t\bar{t}$+X, which includes ttW, ttZ, and ttH, all can produce the same-sign leptons between a leptonic decay of the top and a leptonic decay from the boson. In fact, the $t\bar{t}$+X processes are the largest irreducible backgrounds in the analysis. $t\bar{t}$+XY similarly can pass the same-sign requirement, but the extra boson and vertices decreases the overall cross-section of these processes, so the overall yield is fairly small. Four top similarly has a close kinematic profile to three top, but has a comparatively smaller cross-section to the $t\bar{t}$+X processes. All of these processes Feynman diagrams are shown in Figure 8.1

The second category is the reducible backgrounds which have significantly different kinematic shapes but none-the-less contribute significantly to the signal region. A perfect example of a reducible background is the nonprompt background. This background makes up a majority of the expected yield, but can easily be discerned from our signal. While it might be advantageous to tighten our requirements to remove it completely would require hurting our overall signal efficiency. With the template histogram

[†]Here, X/Y here stands for a boson, namely a W boson, Z boson, or Higgs boson.

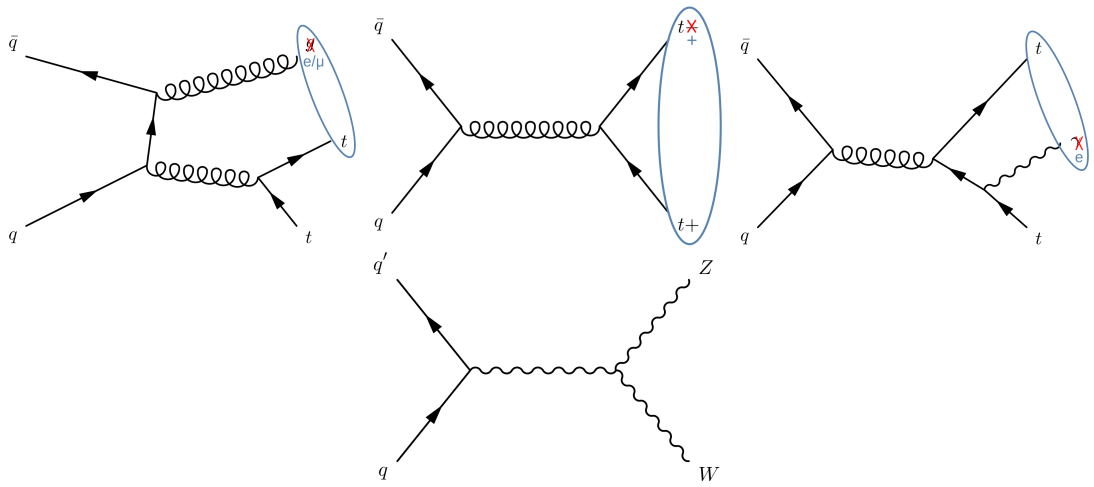

Figure 8.2: Examples of the common reducible backgrounds in our analysis, on top from left to right are nonprompt, charge-misId, X + γ, while on bottom is the WZ background. The lepton pair that passes the same-sign requirement is circled. The nonprompt background fakes our signal region by having a jet misidentified as a lepton. Charge-misId passes by having charge of an electron misidentified so an opposite-sign event becomes a same-sign event. The X + γ fakes our region by having an emitted photon be identified as an electron. The last group, rare, is a catch-all for remaining backgrounds with the WZ sample specifically is shown.

method, our statistical model can separate the two processes without loss of signal, so totally removing it is unnecessary.

In Figure 8.2, the Feynman diagrams for the major reducible backgrounds are shown, displaying the mechanism for passing the same-sign requirement. Excluding the data-driven backgrounds, we can see two main groups: what we call X + γ and rare. The X + γ refers to any process with a high energy photon in the event, where these events pass our selection because these high energy photons are misidentified as electrons. While our electrons have requirements to reduce these sorts of backgrounds, such as the missing hit and conversion vetoes (discussed in Chapter 6), their cross-sections means there will always be events that pass our requirements. The X + γ group is made up of Wγ, Zγ, t$\bar{\text{t}}\gamma$, but t$\bar{\text{t}}\gamma$ has the largest overall impact in the signal region owing to the top quarks in the event. For the rare group, this is a catch-all for any process not previously described that can be found in signal region but will individually contribution very little. This includes multiboson events (so diboson and triboson processes) as well as some of the previously mentioned, low yield top processes such as single top and t$\bar{\text{t}}$+XY, but the rare group is dominated mainly by the WZ process. Each individually may generally pass some part of the preselection, but will be greatly reduced by the overall preselection. In the end, the rare processes make up a small, but measurable difference to the overall yield.

Any other MC process is deemed too small, and if there is good agreement between data and MC, we can have confidence that this is indeed true. For testing, we look at a preselection region, or a part of

the phase space that is less stringent than the signal region where the signal is insignificant.

8.2 Preselection

A preselection region, as previously stated, is a phase space that has looser requirements to the signal region. The purpose of the preselection is to create a region where the signal contribution is much smaller as an overall percentage of the yield so MC and data can be compared. This is done to avoid any bias in the analysis procedure, else an analysis could be constructed to enhance the statistical significance of a signal artificially. In practice, all the analysis creation is done blinded, or without looking at data, and data is only considered after the methods are set or when considering regions that are signal insignificant to thus avoid bias. The preselection region avoids this concern by loosening the signal region so the overall agreement of the MC modeling can be tested while also not preemptively detecting the signal.

Te first requirement is a combination of dilepton and single lepton triggers. The dilepton triggers used have a first lepton p_{T} requirement around 25 GeV and a second lepton p_{T} requirement of 8 GeV to 12 GeV, depending on the lepton type. While sufficient, the single lepton trigger increases our overall efficiency. For combining the single and dilepton triggers, we simply OR them together, or accept any event where one of the triggers is activated. We do have to be sure the requirements applied to the event are always harsher than what is expected by the triggers to avoid any mismatch between data and MC. If the analysis selection is less stringent than the trigger requirement, data events which are simply not saved could be found in our simulation leading to possible disagreement. Also, the reconstructed event content may not match data when the trigger makes its decision, so requiring our events have more aggressive requirements further decreases any disagreement. This means our leptons must be defined with the trigger in mind. For our objects, the baseline p_{T} requirement is such that is above the lower lepton threshold, and an additional $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ is applied to the leading lepton. The overall disagreement is handled by scale factors which are calculated specifically for this analysis using the tag-and-probe method discussed in Section 6.4.

Besides the same-sign requirement, we expect our three top signal to contain three b quarks and three W bosons, so signal like events should have lots of jets and real MET coming from the neutrinos from the leptonic W boson decays. For this reason, lots of backgrounds can be removed or reduced by requiring multiple jets, large jet activity, and large MET. For modeling this overall "jet activity", a variable called the H_{T} or transverse hadronic energy is constructed. This quantity is formed by taking the scale sum of the p_{T} of all jets, or:

$$H_{\mathrm{T}} = \sum_{\mathrm{jets}} |p_{\mathrm{T}}|$$

By summing all jet p_{T}, we have a simple heuristic for the amount of jet energy in the event. For three top events, we might expect our H_{T} to be quite large. Considering a simple case, we might expect the

H_T to be $3 \times m_t \approx 525\,\text{GeV}$, or, removing the momentum lost to the leptonic W boson decays (which accounts for roughly half the top quark decay), we find an approximate value of $H_T \approx 350\,\text{GeV}$. This value for the average expected H_T is a ballpark figure, but gives an initial baseline.

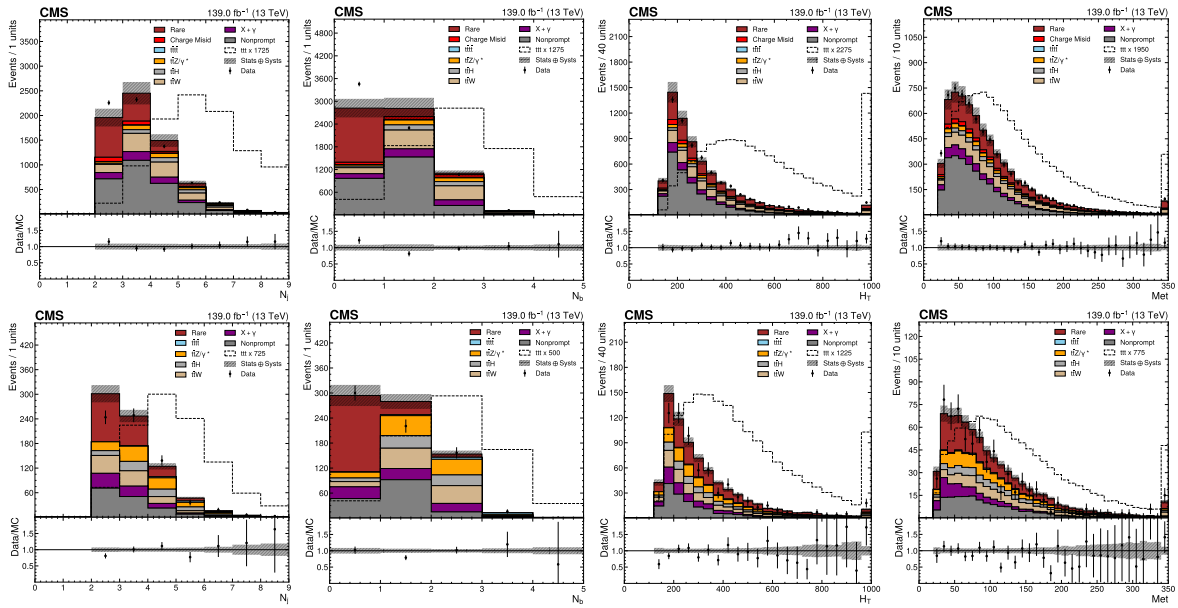

Figure 8.3: Preselection plots for the full Run 2 for the dilepton channel (above) and multilepton channel (below) showing the plots, from left to right: number of jets, number of b-jets, H_T, and MET. As can be seen, the overall agreement is good, and for these distributions, the signal is distinct from the other backgrounds. Here, the oddly named rare group dominates because of WZ events which generally do not have a b-jet, but this suggests additional requirements on the b-jets can remove large amounts of background without affecting the signal acceptance much at all.

Events used for the signal extraction must have two tight ID, same-sign leptons. For events with three or more leptons, the event is vetoed if three of the leptons have the same charge since such a charge configuration is forbidden in the three top production. Further, the mass of any pair of opposite charge, same flavor loose leptons in the event must be greater than 12 GeV and must not be within 15 GeV of the Z boson mass. These mass requirements remove any mass resonances, such as J/ψ at low masses and the Z boson. This is attractive because QCD vents will produce low mass resonances and ttZ will obviously produce a Z boson, so we can veto these processes with the mass requirement. The final preselection requirements are:

- $N_\ell^{\text{tight}} \geq 2$
- Leading lepton $p_T > 25\,\text{GeV}$
- $N_j \geq 1$
- $E_T^{\text{miss}} > 25\,\text{GeV}$
- $H_T > 150\,\text{GeV}$

- $m_{\ell\ell} > 12\,\text{GeV}$ and $|m_{\ell\ell} - m_Z| \geq 15\,\text{GeV}$ for any ee or $\mu\mu$ pair with opposite charge

From this preselection region, we can now look at the data-MC agreement as shown in Figure 8.3. We note that the agreement is good showing our modeling is fair. These distributions show in particular the most discerning variables, namely N_j, N_b, and H_T where we can see the signal separation in both the dilepton and multilepton channels. This is expected since our preselection region is quite loose, thus many backgrounds can be removed using simple additional requirements. From these distributions, we decide to define our signal region by tighten our requirements to: $E_T^{miss} > 50\,\text{GeV}$, $H_T > 250\,\text{GeV}$, and $N_b \geq 1$. These remove most all the WZ sample in the rare group as well as significant amounts of the nonprompt. After this step, we use another tool to extract more from the data.

8.3 Machine Learning Techniques

For the signal extraction, there are a few considerations that need to be made. First, the signal is very small meaning without a necessary amount of signal separation, it will be complete hidden by the background and its uncertainties. Second, and more subtle, is related to the four top background. While four top is not a significant background source, it produces four top quarks which matches our signal profile.

This has the troublesome side effect of limiting our ability to disentangle the four top background from our signal. For example, we might use a distinguishing variable such as the number of b-tagged jets to remove most backgrounds and leave a purer signal region, but the acceptance four top would be higher because of the addition top quark. This means simultaneously removing four top and other backgrounds while not reducing the signal largely is quite difficult. To make matters worse, the four top process has a cross-section roughly ×10 that of three top meaning for any statistical test, our signal will be heavily contaminated and possibly totally covered up.

To counteract these issues, the machine learning methods are used for signal separation. Machine learning is most generally defined as a self-guided algorithm that learns to categorize or make some decisions base on learning performed over a subset of data. This categorization can be broad, but the results created from machine learning tools turn out to be powerful, as can be seen with current large language models used for services like ChatGPT. [76] For this analysis, the most complicated versions of machine learning are not needed because our low expected event yield. For complex models to extract useful differences between two sources, there must be sufficient data for the model to train on, but our smaller training data necessarily limits the possible complexity of our final classifier. The type of machine learning algorithm chosen for this analysis is a Boosted Decision Tree (BDT) [73].

8.3.1 Description

A boosted decision tree is an advanced version of a decision tree where a decision tree is essentially a flow chart; each node is a requirement on an event variable that successively splits the data into the

final child nodes. This basic structure allows a data set to be split into arbitrary sized groups (depending on the tree size), but for our purposes, it splits our data into signal and background event. A decision tree reproduces, roughly, the procedure for a cut-and-count analysis, but with higher granularity. A simple extension of the decision tree is adding weights to each branch, or edge, which can be added up along each path to produce a final weight. This weight, after simple normalization, can act as a signal *vs.* background discriminant for fine-tuned separation rather than producing a binary result. This allows for weighting to give prominence to requirements that help separate signal-like events more effectively, so our decision tree can fully leverage all the available shape information.

The problem in producing a good decision tree lies in the computation difficulty. To make an effective decision tree, one needs to find the best node requirements and edge weights across the tree to create results that can outpace a simple cut-and-count analysis. As the number of considered variables increases, the phase space to optimize over becomes too huge to create high quality results. Even with optimization tricks, there is no way to avoid potential local minima that perform on par, or worse, than simpler methodologies. To remedy this, we use boosting. Instead of creating a single dense and complex tree, boosting algorithms create multiple smaller, simpler trees and combine them. Creating simple trees with moderate discriminating power is easy, and using the weights generated in each tree, we can "add" these trees together to get a combined result from the suite. Each of these trees in the ensemble can have their overall weights adjusted based on its individual discriminating power allowing for optimization. This boosting method also has the benefit of reducing the overall bias of the algorithm since building a single, large decision tree often induces biases coming from optimization into a local minimum.

For this analysis, we use the library called XGBoost, or the Extreme Gradient Boosting [77]. Gradient boosting is an extension where the least discriminating trees are modified at each step to make the overall result improve. This is performed using gradient decent or where the weights in the trees are adjusted based on the gradient of the error of result.

Because the BDT is just a collection of trees that are combined by adding the total weight of each tree's output, the BDT discriminant will generally be unbounded. In fact, with longer training and more subtrees, the bounds of the discriminant are expected to increase. The typical way of dealing with this is to then apply a logistic function on this "raw" discriminant to create a more manageable output that is constrained from 0 to 1.

8.3.2 Methodology

Our goal is to create a discriminant that separates signal from background. In addition to this goal, we also want to take care of the four top background in particular. Running the BDT training, naively, on the signal against all the backgrounds leads to a result where the background peaks around zero, as expected, but curiously, the four top background has a mean BDT score greater than our three top signal. Four top, in some sense, has all the hallmarks of a three top event but to an even larger degree. For example,

for a discriminating variable like N_{b}, the three top signal will have a larger value than background, but the four top background will have an even larger b-jet multiplicity, so it contaminates our most signal sensitive region. This also causes correlation problems for our final statistical fit. If both our signal and one of the backgrounds have the same shape in our final histograms, we've nullified the utility of the shape information and reproduced the cut-and-count method. It is very unlikely a background's shape will exactly match our signal shape, but large signal region contamination can roughly reproduce this problem meaning our four top background will be expected to strongly limit our final results without controlling for it.

The method to conquer this is to create a dedicated four top control region using a BDT thus limiting and constraining the amount of four top. The procedure is: train a BDT for the four top background (called the 4top BDT), create a four top control region using said BDT, and lastly, train our final signal BDT on the events not in the four top control region (called the 3top BDT). The four top control region is created by simply applying a selection on the 4top discriminant in a way that optimize to keep the overall signal significance of three top as large as possible (i.e. not removing too much signal from the signal region) while reducing the correlation between three top and four top as much as possible. There are methods for creating simultaneous discriminants with BDTs, but in testing, they had worse results than separating the training into two steps.

Both of these BDTs are produced in similar ways, using the same variables and optimization methods, but run on different sets of data and with different groupings of said data. For describing the basics of the method, each step will be described for the general generation and then specifically for the two BDTs.

8.3.3 BDT Creation

The first step is creating a list of important variables. For our case, we need an assortment of variables that both distinguish three top from the reducible and irreducible backgrounds with a particular focus on separating three top from four top. The most obvious variables for inclusion are ones related to jets because three top is most notable for the byproducts of the top decays, namely numerous jets and b-jets. For this reason, the most discerning variables are generally the number of jets, N_{j}, number of b-jets, N_{b}, and H_{T}. Further, similar variables can be added, such as $H_{\mathrm{T}}(\mathrm{b})$, or the scalar sum of p_{T} of all b-jets and other kinematic quantities that combine leptons and jets. For helping separate three top from four top, we incorporate topological variables, or variables that relate to where particles are located in relationship to each other. Typical variables of this sort are something like $\Delta R(p_1, p_2)$, or angular distance between two particles.

The variables are honed to a list of 40 by looking at their importance in a BDT training. Since the BDT is a summation of weights of multiple trees, the importance of a variable can be estimated by looking at the total weights associated with it by running test trainings. We come to the full list of used variables below:

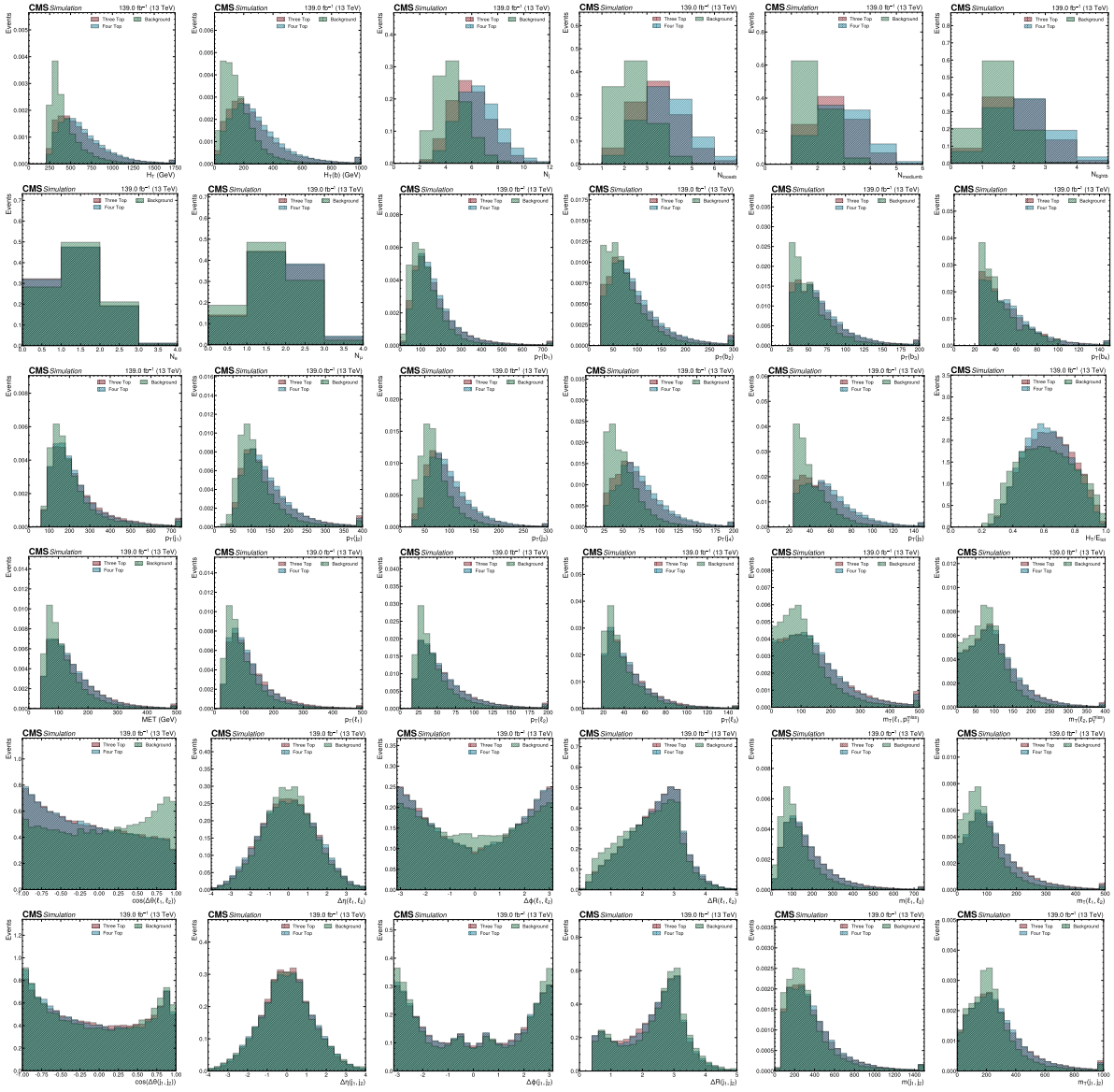

Figure 8.4: Here are a subset of the variables used in the BDT training with each plot showing a normalized shape of the signal, four top background, and the rest of the backgrounds. Some plots are composed of variables that might not have a valid value (such as $p_{\mathrm{T}}(\mathrm{b}_2)$ for events with only one b-jet), so a dummy value is used to indicate a missing value, but in these plots, these missing values are excluded to best compare the different kinematic shapes.

- N_{j}
- $N_{\text{loose b}}$
- $N_{\text{medium b}}$
- $N_{\text{tight b}}$
- $N_{\text{loose }\mu}$
- $N_{\text{loose e}}$
- H_{T}
- $H_{\mathrm{T}}(\mathrm{b})$
- $E_{\mathrm{T}}^{\text{miss}}$
- Centrality $(H_{\mathrm{T}}/E_{\text{tot}})$
- $p_{\mathrm{T}}(\mathrm{j}_{1-5})$
- $\text{Disc}_{\text{b-tag}}(\mathrm{j}_{1-5})$

- $p_{\mathrm{T}}(\ell_{1-3})$
- $m(\ell_1, \ell_2)$
- $\Delta R(\ell_1, \ell_2)$
- $m_{\mathrm{T}}(\ell_1, \ell_2)$
- $\cos\left[\Delta\theta(\ell_1, \ell_2)\right]$
- $m(\mathrm{j}_1, \mathrm{j}_2)$
- $\Delta R(\mathrm{j}_1, \mathrm{j}_2)$
- $m_{\mathrm{T}}(\mathrm{j}_1, \mathrm{j}_2)$
- $\cos\left[\Delta\theta(\mathrm{j}_1, \mathrm{j}_2)\right]$
- $\cos\left[\Delta\theta(\ell_{1-2}, \mathrm{j}_1)\right]$
- $m(\ell_{1-2}, \mathrm{b}_1)$
- $m_{\mathrm{T}}(\ell_{1-2}, E_{\mathrm{T}}^{\mathrm{miss}})$
- m_{T2}

Here, there are multiple variables used with complex, but useful definitions. m_{T} is the transverse mass (analogous to the p_{T}), $\mathrm{Disc}_{\mathrm{b-tag}}$ represents the b-tag discriminant. m_{T2} is a specialty variable created by calculating th m_{T} but estimating the real invisible momentum from the MET [78]. The centrality is a ratio of the H_{T} to the total jet energy, so it acts as a proxy for the proportion of energy in the transverse plane. A subset of these variables are plotted for the signal, background, and four top groups in Figure 8.4. While many of these variables will be correlated to each other, (e.g. H_{T} and N_{j} are strongly correlated since H_{T} will increase with more jets), BDTs are robust enough to avoid problems associated with correlating input variable, unlike many other forms of machine learning. This is in some part because of the simple nature of the decision trees themselves in addition to the boosting process. Indeed, when testing the performance of the BDT with and without certain variables, no significant increase in performance was noted by removing variables in a way that would suggest correlations having maliferous effects on the results

The next step is setting up the data for training. The typical procedure splits the data into a test set, training set, and validation set. The training set is used to actually train the BDT and should be a large as possible. The validation set is data not trained but used during the training process to cross-check the model's progress to avoid overfitting, or what is generally called overtraining. Overtraining is when the BDT finds patterns particular only to the training dataset and not an overall feature expected of the data at large. When a BDT is overtrained, this false pattern will result in good differentiation in the training set, but quite poor results when applied more generally. The validation set allows an unbiased test of the performance to avoid this. The test set is similar to the validation set, being the data that is not trained over, but it is used to see the final performance. Most trainings forgo the test or validation set in lieu of other, but that is predicated on the model being used to run over random input data. In our case, we only have the MC events we have produced, and for the final statistical fit, we cannot use any of the training data because these events would be inherently biased from the knowledge the BDT had during the training process, i.e. which events are signal or background. For this reason, the test set will act as the final verification of the BDT model as well as the data used for creating the final histograms while the training set and validation set are only used for the training itself.

To perform the split of the MC samples, we try to keep a split of 70 %, 27 %, 3 % for the test, training, and validation sets, or 30 % of the MC is removed for training, and of that 30 %, 10 % is used for validation during the training itself. While this breakdown is the ideal, the input data to the BDT needs to be proportioned such that the more important samples are prioritized in the training. To achieve

this, the number of raw MC events contributed to the training set is apportioned based on the expected yield in our signal region thus ensuring more important backgrounds have a larger presence in the training. This could be achieved using an event weight, but events weights that are too small or large can lead to strange effects in the training. The BDT iteratively trains over different subsets of data to avoid overfitting, but this means very small or large weights can produce poor performance depending on the subsample considered. This is especially problematic for backgrounds with many MC events, but low overall yield because they saturate the training steps and skew the result.

To split our data, we must calculate the expected proportion for the yield of each sample and set aside enough raw events for each background. For the total number of raw events used in the training, this was estimated using the $t\bar{t}$+X samples taking up the entire required 30 % of their events for training since they are the most important irreducible backgrounds. If a sample did not have enough raw events, 30 % of the events were taken, but each event was weighted to match the required amount. If a sample had less than 200 events, the sample was not included in the training. Last, for the data-driven backgrounds, the charge-misId was excluded because of its small expected yield while nonprompt is modeled with MC. In lieu of wasting the data events for training, the MC samples for $t\bar{t}$, Drell-Yan, and W+Jets were used with the nonprompt background estimation method applied. While the MC doesn't emulate nonprompt background perfectly, it models well enough for the basic signal-background separation.

With the basic pieces of the BDT ready, the training can be performed. To achieve the best results out of the BDT, there are multiple variables in XGBoost called the hyperparameters that affect the model creation. For perfecting our BDT results, we optimize these hyperparameters using Bayesian random search. Machine learning works, in some sense, as a minimizing algorithm that tries to reduce the loss, or some function that describes how "wrong" the results are. The random search tests each hyperparameter and changes them to find, roughly, what values minimize the loss, or performance, of the BDT. The variables that were optimized for were:

- `max_depth` — max length of the tree
- `gamma` — minimum change in loss per step
- `eta` — learning speed
- `eval_metric` — metric for minimization: logloss, rms-error, AUC, or mean absolute error
- `colsample_bytree` — subset of variables used for each training
- `min_child_weight` — minimum weight allowed for a branch of a tree
- `subsample` — subset of sample used for each training

Most of these variables regulate how each step of the decision tree creation is performed and how large each of the changes are allowed to be at each step of the fit. These are important for reducing overtraining and increasing the overall complexity of the BDT itself. This optimization is performed for both BDTs separately to ensure the best performance, especially since the different setup conditions lead to different responses of the training.

For the training of the two different BDTs, first, the 4top BDT was trained using the four top sample as the signal and all other samples as background. While the backgrounds had the number of events used based on the method mentioned above, we include as many three and four top events as possible because they are the processes we are concern with the most for demarcating our control region. We define this four top control region as all events with a BDT score greater than some threshold. We must define this control region to balance reducing the correlation between four top and three top while also keeping as many signal events as possible for the final signal extraction. To do so, tests were run to maximize the overall significance across the four top control region and the different signal regions while minimizing the correlation between three top and four top. Before including a four top control region, the overall correlation neared 99 %, but the best results came from a requirement that reduced the overall significance by less than 5 % while reducing the correlation to roughly 80 %. After defining our four top control region, the leftover training events are again used for creating the 3top BDT where our signal was trained against all other backgrounds, excluding the four top background.

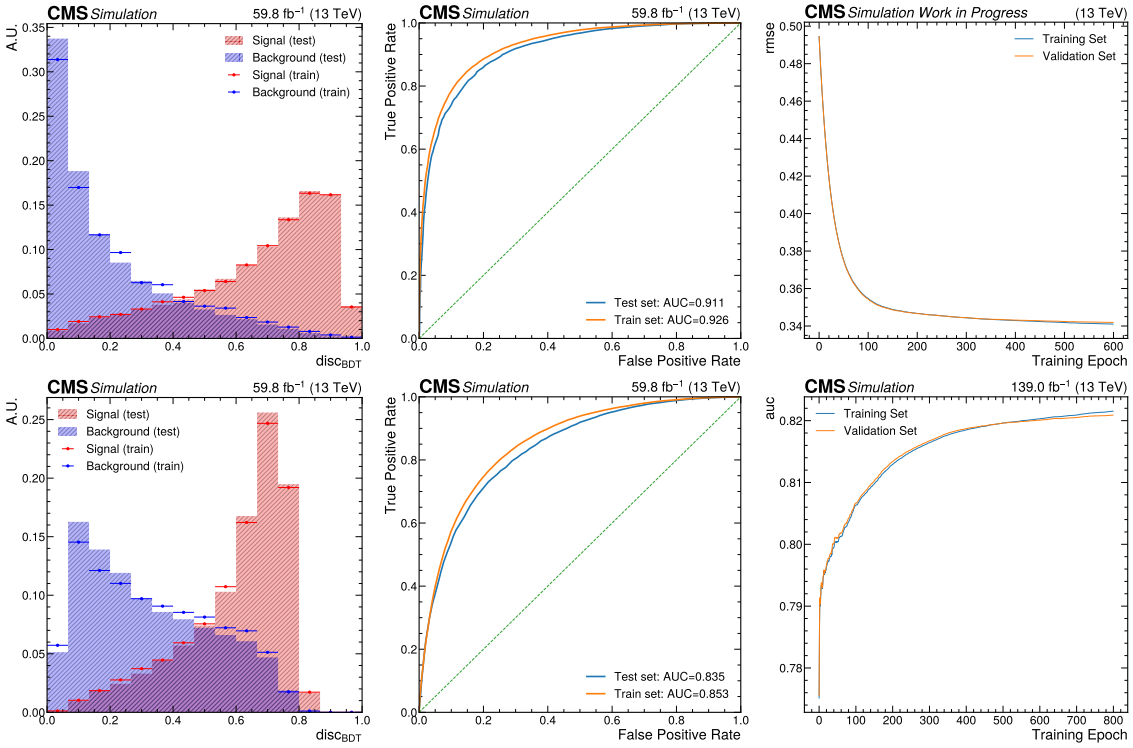

Figure 8.5: Plots showing the basic training output for the 4top discriminant (above) and the 3top discriminant (below). On the left is an example comparison of the different output shapes for training and test set to check for overtraining of the model. In the center shows the basic ROC curve of model performance. On the left shows the performance of the model over each training step, where the model stops after performance slows. For the plots shown, they are produced for 2018 as a representative model of the other years.

Each BDTs' performance can be checked by comparing the training set against the test set to confirm there is no overtraining. We'd expect markedly better performance of the training set if there was overtraining, but as shown in left most plots of Figure 8.5, the BDT responds similarly showing no bias. The right most plots show the overall performance of the model at every training steps, or epochs, where the efficacy of training levels out, letting XGBoost know when to stop. The last diagnostic tool used is the ROC curve, or the receiver operating characteristic curve [79] in the middle of Figure 8.5. The ROC curve is a plot of the true positive rate against the false positive rate. An almost perfect BDT will create a curve hugging the top left corner corresponding to nearly zero false positives or nearly all elements are correctly labeled, and a poor discriminator which guesses randomly would have a straight line ROC curve meaning an equal proportion of correct and incorrect choices. To quantify how well the discrimination works, one can integrate the ROC curve to find the area under the curve, or AUC. For our purely random discriminant, the AUC will be 0.5 meaning there is a 50 % chance of choosing correctly vs incorrectly, as opposed to our ideal case of AUC = 1 where our BDT chooses correctly 100 % of the time. For our 4top and 3top discriminants, we find an AUC of ~ 85 % and ~ 90 % respectively, showing an overall good separation.

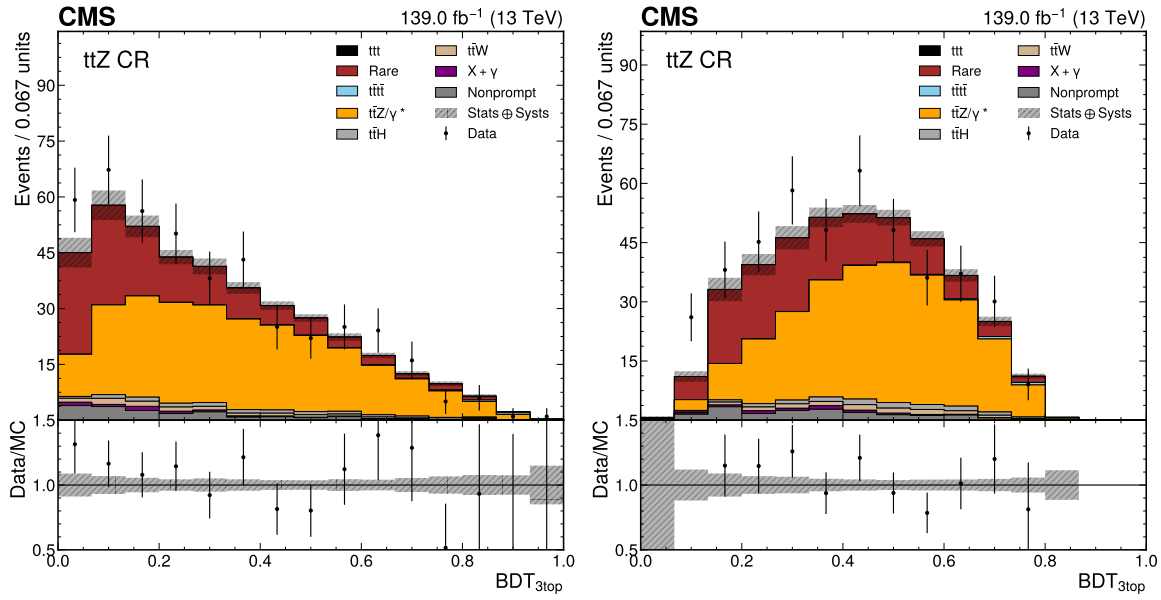

Figure 8.6: Plots showing the validation of the two used BDTs (four top left and three top right) in the ttZ control region. Because the ttZ control region inverts the Z boson mass veto and creates phase space that has a low signal yield that allows us to unblind and check the BDT's response in data. The agreement looks fair, especially in the four top and three top enriched regions i.e. closer to 1, where we want the best signal response.

For a last quality control, we can test the performance of the BDT on real data. While the previous checks tested for overtraining as well as general efficacy of the BDT, we haven't shown that the response

in data will match our expectations. With a device created from multiple variables, each slight tensions in agreement can compound in the final discriminant leading to a BDT that reacts differently in data. To test this, we must look at data in a separate region where signal should not exist in large proportion to avoid biasing ourself. We have the benefit already having a perfect choice for the phase space in the form of the ttZ control region. The details of this region are discussed in full in Section 8.4.2, but this is a control region created by inverting the Z boson mass veto, or it is a region created from events with a same flavor, opposite lepton pair with dilepton mass of $|m_{\ell\ell} - m_Z| < 15\,\text{GeV}$. Plotting the BDT discriminants in this region as in Figure 8.6, we notice fairly good agreement.

8.4 Fit Regions

Now we are confident in our BDT, we now have the ability to create our signal regions and any control regions and the kinematic variables that will be the inputs to our fit models. For each variable, we must bin it into histograms for input to the statistical software. The binning is optimized to increase signal significance while keeping at least five events in each bin to ensure the usual statistical approximations are valid. For this analysis, four regions were created to derive our final results: two control regions (ttZ and $t\bar{t}t\bar{t}$) and two signal regions (dilepton and multilepton) where the histograms for signal extraction are shown in Figure 8.7

8.4.1 Dilepton and Multilepton Regions

Our primary signal region is split into two parts to help with the fitting. The two regions are the dilepton and multilepton region, which are, unsurprisingly, defined by the number of leptons in each event. The dilepton contains the two lepton, same-sign events, while the multilepton region has all the events with addition leptons, so three or four. The dilepton region is expected to have roughly twice the number of events (going by the pie chart in Figure 2.7), but in practice, after selection, the dilepton region has roughly $\times 10$ the amount of data. Even with the smaller yield, the multilepton region, generally, is more signal significant in part because of the smaller background contamination. For this reason, the regions are split to leverage the stronger discriminating power of the multilepton region while still having numerous events for improving confidence in the dilepton region.

For the histograms, both use the 3top BDT, though each is binned differently based on the available statistics. As is discussed in Chapter 9, our statistical fit model works best when we have at least five events per bin, so the approximations used for calculating our statistical tests works. For this reason, the overall binning is fixed to maximize the total signal significance while keeping five scaled events in each bin at a minimum.

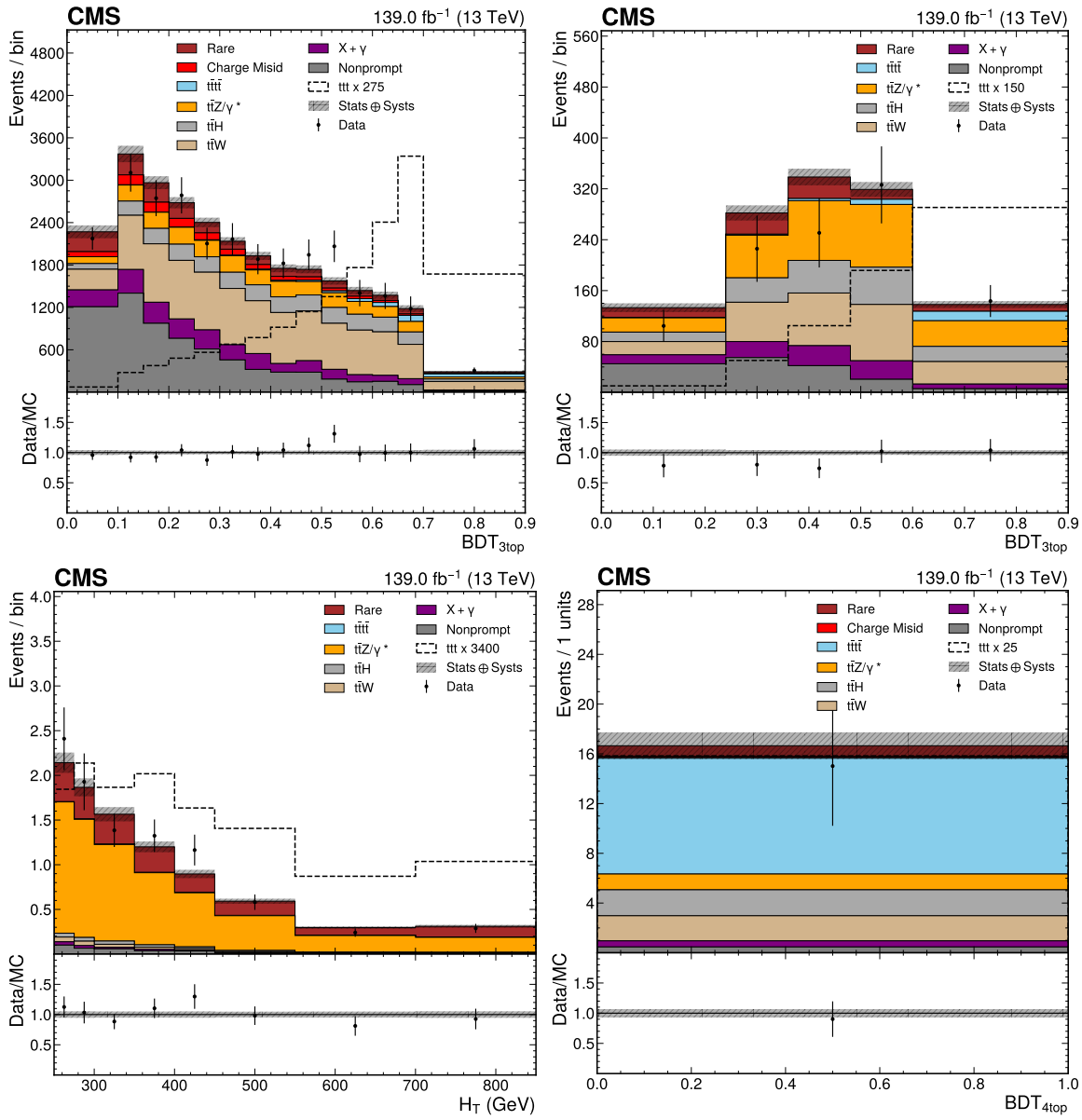

Figure 8.7: Final template histograms for the full Run 2. Starting from the top left going clockwise, the plots show the 3top BDT for the dilepton region, multilepton region, the 4top BDT in the four top control region, and the H_T distribution in the ttZ control region.

8.4.2 ttZ Control Region

The ttZ control region shares a similar selection as the multilepton region except the Z boson mass veto requirement is flipped, or events must contain three leptons with at least one pair of same flavor, opposite sign leptons which have $|m_{\ell\ell} - m_Z| < 15\,\text{GeV}$. Because of the same-sign requirement, this

control region must have at least three leptons, and thus we expect this region to be dominated by ttZ. We used this region previously for testing the data's response to our BDT, but we can extract more from it by using it to constrain the ttZ sample yield in our fit model. We can link the ttZ cross-section across histograms in different regions, but if we have a region saturated in a sample, the overall background can be constrained in the other regions.

Since the ttZ sample is being constrained using this region, we choose a fit variable where the ttZ is significantly different from other samples is needed. While a variable such as the Z boson mass peak might be work, the residual WZ background is the largest secondary background, so the H_{T} distribution is chosen to distinguish the two.

8.4.3 Four Top Control Region

The four top control region is created to help reduce correlation between the three top signal and four top background. This was also mentioned previously in the creation of the two BDTs, but the four top control region is defined as the phase space where events have a 4top BDT score greater than some threshold. Events that failed this requirement are what are considered the signal region. While other techniques exist to separate the four top background from our signal region, none can achieve true separation. Even when training the three top signal against the four top background directly, the greatest AUC was found to be less than 0.70, so with current techniques, there will always be a problem in separation. Of the different techniques tried, the currently used method of two tiers of BDTs was the simplest and highest performing, so other ways of handling this issue showed little to no improvement in our tests.

Chapter 9

Results

Now the final step in our analysis has been reach. Once we have our events region defined as well as our final histograms ready for fitting, we can now perform the statistical hypothesis testing. For this, CMS and CERN as a whole have developed statistical tools for making inferences about the data which, while complex, aim to extract the most from the data it collects.

9.1 Statistical Tests

For a typical analysis as CMS, the statistical tests are performed with the hypotheses:

- Null hypothesis: Only the background processes exist (B)
- Alternative hypothesis: signal and background exist (S+B)

For devising the fit, we use the maximum likelihood method. For building this method, we define the likelihood function which is the product of all the probability functions associated with our data. First, our data is binned into histograms where each bin is characterized by a Poisson distribution, $\mathcal{P}_{\text{Pois}}(n, \lambda) = \frac{\lambda^n}{n!} e^{-\lambda}$, where the n is the observed number of events and λ is the expected number of events. For our case, our expected number of events is the event yield modeled in simulation, but we actually have more flexibility in defining the expected number of events. Our statistical test is performed by comparing data to our two hypotheses conditions, but the signal may not be produced exactly as defined in the Standard Model; for example, the cross-section could follow some BSM model. To generalize the fitting to an arbitrary signal yield, we define a quantity called the signal strength which is a multiplicative factor on our signal cross-section to allow the signal yield to float:

$$\sigma_{\text{fit}} = \mu \sigma_{\text{theory}}$$

Using this signal strength, we can fit our histograms to data to find what cross-section for the signal matches best to data and restate our hypotheses as $\mu = 0$ for the null and $\mu \neq 0$ for the alternative. With this, we can represent our expected mean of our Poisson distribution as $\lambda = \mu S + B$. Using solely

Poisson distributions to describe the data is incomplete, so the second step for building the likelihood function is adding in the systematic uncertainties. Systematic errors are the collection of uncertainties inherent in the modeling, measurement, or theory as opposed to the statistical errors which represents the error in our statistical modeling and testing. The systematic errors are normally modeled by a Gaussian distribution, but Gaussian distributions allow for negative values since they are defined over $\mathbb{R}$, but this is not allowed for our strictly positive event yield. To avoid this, the log-normal distribution for our systematic errors, where the log-normal distribution is the Gaussian distribution of the log of the values. Its domain is the positive reals, so we avoid any strange issues in our fitting of the likelihood function. For a given systematic error, θ, the log-normal distribution is expressed as:

$$\mathcal{P}_{\text{LogN}}(m, \sigma; \theta) = \frac{1}{\theta\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln(\theta) - m)^2}{2\sigma^2}\right]^{\dagger}$$

To represent the systematic error generally, we collate them all into the vector $\boldsymbol{\theta}$, which their own log-normal distribution in the likelihood function. These systematic errors represent overall uncertainty in our background and signal yields or kinematic shapes, so our yields will depend on these errors. For simplicity of notation, we represent all our probability distribution functions as functions of just our free parameters, i.e. signal strength and errors, so our Poisson distribution for the i^{th} bin can be re-represented as:

$$\mathcal{P}_{\text{Pois}}(n_{\text{data};i}, \lambda = \mu S_i(\boldsymbol{\theta}) + B_i(\boldsymbol{\theta})) \to \mathcal{P}_{\text{Pois}}(n_{\text{data};i}, \mu, \boldsymbol{\theta};\ i)$$

Putting these parts together, we find our basic likelihood function is:

$$\mathcal{L}(\mathbf{n}_{\text{data}}, \mu, \boldsymbol{\theta}) = \prod_{\theta\in\boldsymbol{\theta}} \mathcal{P}_{\text{LogN}}(\theta) \prod_{i\in\text{bins}} \mathcal{P}_{\text{Pois}}(n_{\text{data};i}, \mu, \boldsymbol{\theta}; i)$$

Since these probability functions are multiplied together and involve exponential functions, we can reframe the likelihood as the negative log-likelihood, which simplifies the math significantly:

$$\begin{aligned}\ell(\mathbf{n}_{\text{data}}, \mu, \boldsymbol{\theta}) &= -2\ln\mathcal{L}(\mathbf{n}_{\text{data}}, \mu, \boldsymbol{\theta}) \\ &= -2\sum_{i\in\text{bins}} \ln\mathcal{P}_{\text{Pois}}(n_{\text{data};i}, \mu, \boldsymbol{\theta}; i) - 2\sum_{\theta\in\boldsymbol{\theta}} \ln\mathcal{P}_{\text{LogN}}(\theta)\end{aligned}$$

The original likelihood function, because it is simply a combination of all probability functions, has its floating parameters maximized when all the different probability functions are maximized. Put simply, the likelihood function is maximized for the most probable case. This is why the function is so powerful. The values for the free parameters, i.e. the signal strength, μ, and errors, $\boldsymbol{\theta}$, give us the estimators

[†]Similar to the typical Gaussian distribution, the log-normal distribution is categorized by a mean, m and variance σ^2. Usually, the notation uses μ for the mean, but, to avoid confusion with the signal strength, the mean is expressed using the variable m.

for these values, or the most probable value of said parameters given our data. Our measurement is concerned with the total three top cross-section, so we only care about the signal strength. While the errors are parameters that must be fit by necessity, they are not the parameter of interest so are often called nuisance parameters. For testing the different hypotheses, the likelihood functions value gives a proxy of how likely a certain data configuration is, so, after fitting our nuisance parameters in the two hypotheses, i.e. background only ($\mu = 0$) and signal plus backgrounds ($\mu \neq 0$), we can compare them to see which hypothesis is preferred. To calculate the estimators as well as the total likelihood value for the given data, we must use numeric minimizers because of the number of errors and bins in our fit is too large.

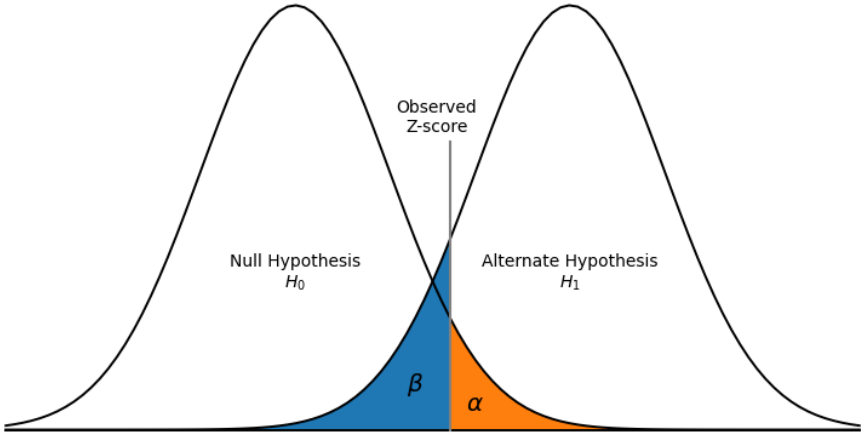

Figure 9.1: This picture gives the basics of how a statistical inference test is performed. Both hypotheses are idealized as Gaussians with means at their expected value for each hypothesis. For any observation, our confidence that the data is in agreement with the null hypothesis is represented by α and we reject the null hypothesis when α is below some threshold (conventionally 5 %). There is also the quantity β which represents the type II error or false negative rate. For a fixed α, the false positive rate shows how distinct the alternative hypothesis is from the null as well as how confident we are in the alternative hypothesis, if accepted.

While we now have a trick for finding the best fit for the signal cross-section, we still have not done any statistics, namely assigning confidence for accepting or rejecting our null hypothesis. To develop this, it's useful to introduce the statistical power. The canonical statistic inference method is performed by creating a probability density function (pdf) that describes data producing specific likelihood values. We reject the null hypothesis when the observed data has a probability of occurrence in our pdf below a certain, predetermined statistical significance, α. The statistical significance also represents the probability of type I error (false positives). This is the error associated with accepting the alternative hypothesis, but the null was actually true.

The other type of error is type II error (false negative) or accepting the null hypothesis when the

alternative hypothesis is true. It correspondingly is represented by the variable β. Both type I and type II error are demonstrated in Figure 9.1. The larger our value β is for a constant α, the more each hypotheses' pdfs overlap meaning both hypotheses are more similar to each other, so to make the strongest statistical statements, we want to minimize β, or otherwise, maximize $1-\beta$. Here, we call this last quantity, $1-\beta$, the power. By maximizing the power, we ensure our two hypotheses are statistically as disparate as possible, so any given data isn't "wasted" for cases where both the null and alternative hypotheses both seem probable. The Neyman-Pearson lemma states that the most powerful statistical tests are performed with the likelihood ratio [80], or:

$$\gamma(\mu, \mathbf{n}) = -2 \ln \frac{\mathcal{L}(\mathbf{n}, \mu, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}(\mathbf{n}, \hat{\mu}, \hat{\boldsymbol{\theta}})}$$

Where $\hat{x}$ represents the expectation value of variable x, and $\hat{\hat{\boldsymbol{\theta}}}$ in the context of $\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}})$ is the values of $\boldsymbol{\theta}$ that maximize the likelihood function for a given value of μ. This equation which uses the double hat is called the profile likelihood ratio and is the test statistic we use for this analysis.

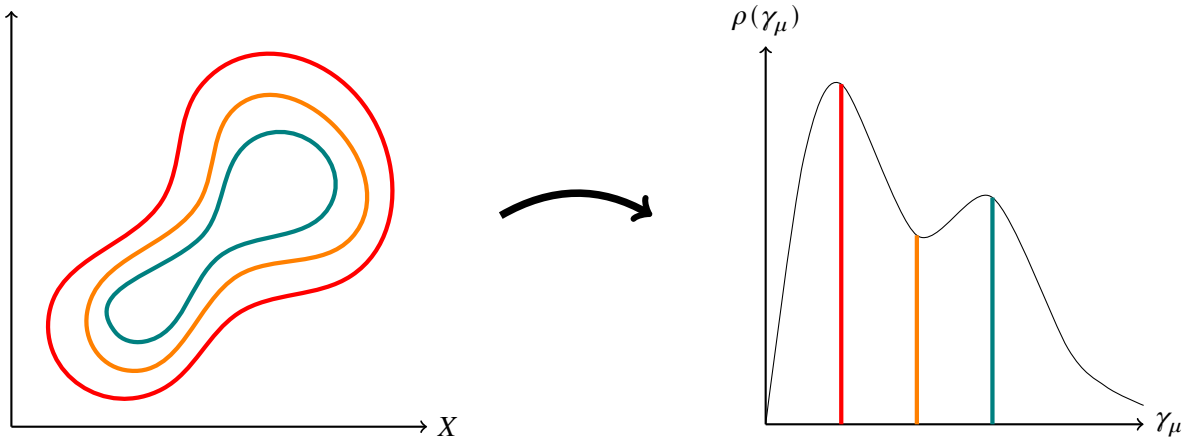

Figure 9.2: Visual example of the creation of the pdf for our test statistics. Given some data space X shown on the left, there will exist a curve set of data such that all produce the same value of γ_μ given a particular μ under consideration. From the relative frequency associated with every data point for every value of γ_μ, the pdf, $\rho(\gamma_\mu)$, can be produced. In practice, the points in the space X are simulated from the expected data.

A test statistic is simply a function that evaluates our data to a value and works to create the pdf used for evaluating our statistical tests. For analytically deriving a functional form of said pdf, $\rho(\gamma)$, it is typically impossible because γ is derived from a minimization for every set of data and μ under consideration, so, except for the simplest of cases, it must be numerically calculated. We estimate it using numerous "toy" datasets, or data that is randomly seeded from the expected distribution. As a nuance, our likelihood ratio is a function of the specific value of μ under consideration (which may be different

from the best fit value, $\hat{\mu}$) and the particular data from the total dataspace X, so it has the general form, $\gamma(\mathbf{n}_{\text{data}}, \mu)$. We are concerned with creating a pdf, $\rho(\gamma)$, which shows how our data is distributed for our test statistic, we must freeze μ to create the pdf. The process for creating it is shown in Figure 9.2. While this figure gives a generalize idea of producing γ_μ, our pdf is maximized when $\mu = \hat{\mu}$ considering our estimator is by definition the most probably value for our signal strength. Looking at our definition of the likelihood ratio, we find $\gamma(\mathbf{n}_{\text{data}}, \hat{\mu}) = 0$, so we expect our pdf to be largest near 0 and small for large γ_μ, i.e. an exponential shape.

Our pdf can be produced for every value of μ, lending us class of functions, but we are generally only concerned with the $\mu = 0$ case because our statistical tests are strength testing the validity of the null hypothesis. For performing the usual hypothesis testing, we look at the p-value of our observed data so we can see if it is greater than our set statistical significance α. The p-value can be calculated simply using the pdf:

$$p = \int_{\gamma_0^{\text{obs}}}^{\infty} \rho(\gamma_0) d\gamma_0$$

We can convert this p-value similarly into a z-score, or the equivalent number of standard deviations our data is from the mean. While most science fields opt for a 95 % confidence level, or $\alpha = 0.05$, CERN has discovery level agreed upon at a significance of 5σ or greater or a p-value of 3×10^{-7} for a one-tailed test. This is absurdly unlikely, but the sheer amount of data collected and analyzed means more stringent standards are necessary. The p-value represent the probability that an "extreme" set of data would naturally occur, so, with an HLT rate of 1 kHz, there is enough data collected and analyses performed that a 5 % threshold would lead to too many false positives. Many times the history of particle physics, a 3σ ($p \approx 0.3$ %) result has been found but subsequently disappeared with more data. Having a 5σ threshold means we can be extremely confident about with our result, so if an analysis can reach this lofty significance, the possibility of it being later falsified are vanishingly small.

If our results do not reach the 5σ level, we can still extract information, namely an upper limit. Here, the upper limit is the maximum allowable signal that is compatible with the data to some confidence level. This is basically performing the inverse of the significance test; we scan the different values of μ and look at the p-value for each given γ_μ, and the μ value that causes the p-value to fall below a certain confidence level gives us the maximum permitted signal that is deemed still compatible with the data. For defining a limit, we typically take the modest 95 % confidence level that is typical in science, but this translates to a more conservative evaluation since this is our certainty in rejecting our signal. Now we have an upper limit for the possible cross-section of our signal to 95 % confidence level, where future measurements could lower this limit until enough data was gathered for discovery. For actual testing at CMS, a modification called the CLs method is used which helps mitigate rejection of searches with smaller signal sensitivity[†] [81].

[†]For completeness, the CLs method uses a modified "p-value" of the form $\frac{p_{S+B}}{1-p_B}$, where p_{S+B} is the p-value derived from γ_μ while p_B is the p-value derived from γ_0. The $1 - p_B$ term increases the probability over the regular p-value to compensate

To create our pdf, we need to model our profile likelihood by populated it with hundreds to thousands of toy data to create a smooth enough distribution for making our statistical claims. This generally takes a lot of time and computation. Thankfully, assuming the values in each bin are large enough (around five or more events, similar to the typical Poisson-Gaussian approximation point), simplifying assumptions can be made that allows us to calculate most statistical values analytically [82]. For this analysis, while our signal is limited, the overall data yield is expected to be moderate, so our histograms can always be binned in such a way that we can use this asymptotic approximation without sacrificing any real significance in our tests.

To actually perform all these tests, the path is clear: we need to identify all sources of systematic and statistical uncertainty and create our test statistic. For the most simple uncertainties, we can model them as flat, log-normal uncertainties, or sources of error that do not affect the kinematics but will affect our overall normalization. A good example of this type of error is the systematic uncertainty on the luminosity. Most errors do change the kinematics and thus the overall shape of their histograms giving rise to the so-called shape uncertainties where the bin-by-bin uncertainty can vary. Pictorially, we imagine events flow between kinematic values from the nominal case to the up and down variations as opposed to the flat normalization shift up and down. The JEC uncertainties are a classic example that changes the kinematic shape, allowing the jet p_{T} distribution to have its mean shift up and down for the two variations. To model these shape uncertainties, each bin has a log-normal distribution corresponding to the bin's expected error with all bins correlated together to keep the total shape "connected" and roughly the same [83]. While complicated, this shape information can localize the effect of an uncertainty in our model so a portion of phase space that has high confidence doesn't erroneously have a large error.

Many uncertainties, especially less important ones, will have shape effects, but don't necessarily require the complexity of the shape uncertainties. Instead, we revert them to being modeled as a flat uncertainty with their error enveloped, i.e. the worst case error is used for each bin. This allows the modeling of these uncertainties to be simplified at the cost of a larger uncertainty. This can be justified for smaller systematics whose uncertainty will not significantly affect the results, especially if performing the studies to estimate the uncertainty would be complicated and time-consuming. The perfect examples in this analysis are the data-driven background uncertainties which have multiple sources of uncertainty that can enter leading to difficult and bespoke final errors that wouldn't likely improve results since the data-driven backgrounds are mainly relegated to signal insensitive portions of our phase space.

For actually performing our statistical tests, we use a software called Combine [84]. With this, we can enumerate all of our sources of uncertainty.

---

for low signal cases so overly harsh limits aren't placed these less sensitive models, making it generally more conservative.

9.2 Uncertainty

Sources of uncertainty are categorized into three types in COMBINE: statistical, normalization, and shape. The statistical uncertainty is taken from each bin in the histograms. This statistical error is relatively small for most bins because our MC samples have a large enough number of raw events to make the effective uncertainty small. While typically there would be a statistical error per sample for each bin in our histogram, we use the Beeston-Barlow lite method which creates a single error per bin for all MC samples [85]. The normalization uncertainties are simply nomenclature for the flat, log-normal uncertainties described before, so the error is just a normalization factor representing a one standard deviation displacement across the whole histogram. The shape uncertainties are exactly what was described in the previous section. These uncertainties need to be monitored and studied because they potentially can change the best-fit values in the fit by "bin migration" or moving the shape around as events change the bin they occupy in the signal distributions. This can cause strange effects in the fit that need to be well understood.

The fit is performed over the three years of Run 2, but over four eras, which are a subset of the years, as described in more detail in Appendix D. For each of the systematics, the uncertainties can be linked, or correlation, between the different eras in our model. For certain systematics, we would suppose the overall uncertainty would change between years, such as with differences in the detector year-to-year, but most uncertainties arise from our simulation and reconstruction and are thus derived identically between eras and should be correlated. When the data was first generated, the reconstruction changed year-on-year requiring most uncertainties to be handled individually for each year, but this analysis uses the UltraLegacy campaign data. UltraLegacy had the simulation and reconstruction performed exactly the same for each era, thus these systematics should be correlated. Correlation is helpful in our fit model because each systematics is constrained between eras, so they are disallowed from ballooning too large for a single era.

For the statistical error, as mentioned, these are automatically determined by the content of MC samples in each bin, so the software simply generates it. For the normalization and shape based uncertainties, they are summarized in Tables 9.1 and 9.2. To explain all of these uncertainties, the following sections give a quick overview of each, grouped into theory, object reconstruction, event description, and analysis method uncertainties.

9.2.1 Theory Uncertainty

Theoretical uncertainties are the first source in the Monte Carlo simulation chain. The sources of these theoretical uncertainties are: expected cross-sections, parton showering, renormalization and factorization scales, and parton distribution functions. Most all of them are inherently the same, being related to uncertainty in theoretical calculations based on unmeasured quantities such as the energy scale, but each

Table 9.1: Summary of all normalization uncertainty sources used for the analysis. Normalization rates listed represent a single standard deviation and are treated according to the log-normal prescription. While the table only lists the year 2016, in the fit model, 2016 is split into the two eras mentioned in Appendix D, but the overall uncertainty is the same between the 2016 eras, and thus the column is compressed for ease.

Source	Corr.	2016	2017	2018
Luminosity	—	0.7 %	2 %	1.5 %
Luminosity	Era	0.4 %	0.9 %	2.0 %
Luminosity	2017, 2018	–	0.6 %	0.2 %
Electron MVA ID	Era	1.5 %	1.5 %	1.5 %
Muon ID	Era	1 %	1 %	1 %
ttH–MVA	—	2 %/lep	2 %/lep	2 %/lep
Dilepton Triggers	—	3 %	3 %	3 %
$\sigma(t\bar{t}t\bar{t})$	Era	10 %	10 %	10 %
$\sigma(ttH)$	Era	20 %	20 %	20 %
$\sigma(ttZ)$	Era	20 %	20 %	20 %
$\sigma(X + \gamma)$	Era	20 %	20 %	20 %
σ(rare)	Era	20 %	20 %	20 %
Nonprompt Closure	—	30 %	30 %	30 %
Charge-MisId Closure	—	20 %	20 %	20 %

Table 9.2: Summary of all shape uncertainty sources used for the analysis. Here, the overall correlations are shown for each, where there are three types: era meaning the systematic is correlated across eras, process meaning it is correlated internal within the QCD and EWK induced processes, and partial meaning there are correlated and uncorrelated portions of the systematic. To help illustrate the systematics, the RMS average is also shown.

Source	Correlation	Approx. Unc.
L1 Prefiring (2016)	—	1 %
Pileup	Era	0.1 %
Jet Pileup ID	Era	0.1 %
Jet Energy Scaling	Partial	2.5 %
Jet Energy Resolution	Era	0.1 %
DeepJet	Partial	3 %
PDF	Era	0.2 %
PDF α_S	Era	5×10^{-2} %
ISR	Era, Process	0.5 %
FSR	Era, Process	1 %
Renormalization Scale (μ_R)	Era, Process	1 %
Factorization Scale (μ_F)	Era, Process	0.2 %
Muon ID	—	0.5 %
Electron ID	—	1 %
Nonprompt MisID Statistical Error	—	2 %

come individually from the different steps of the modeling of the simulation as laid out in Chapter 5.

Parton Shower

PYTHIA inserts ISR and FSR jets automatically during the parton showering step. These jets will heavily dependent on the strong coupling constant, α_S, for a given squared-momentum transfer, Q^2 used in the generation step, so there is an uncertainty for the mismodeling of these inserted jets. The uncertainty is expressed using a simple event weight that changes the relative abundance of events with certain ISR or FSR jets. To generate these weighs, the renormalization scale for QCD emissions in the initial and final states is varied by a conservative ratio of 2 (½) for the up (down) variation.

The FSR is correlated for all samples, but the ISR sources are correlated for all eras, but not across all samples. This is to model the observed difference between ISR produced in electroweak and QCD induced productions of top quarks. This is unsurprising since a process like $t\bar{t}$, which is induced by gluon fusion, should allow for quarks to shed off in initial part of the process, while a single-top process, which is initiated by electroweak decay, has less ability to generate jets before the main diagram. Most of our background processes are QCD initiated except the rare backgrounds since they include diboson, triboson, and single top events, all of which are electroweak processes.

Energy Scale

At generator level, a choice is made for the ultraviolet renormalization scale (μ_R) and the infrared factorization scale (μ_F) for performing calculations at matrix level. Changing these alters the interactions in ways mentioned before such as changing the coupling constant value for the different forces. The uncertainty is estimated using an up and down variation of 2 and ½, as with the parton shower uncertainty. An additional constraint on μ_F is,

$$\frac{1}{2}\mu_R < \mu_F < 2\mu_R.$$

This avoids underestimating the uncertainty due to cancellations between the μ_F and μ_R dependence. These energy scales are distinct from the parton shower uncertainties because they relate to the initial, matrix level interaction, but similarly to the ISR uncertainty in particular, the μ_R and μ_F sources are correlated by the background process groupings, namely QCD and EWK.

Parton Distribution Function

The parton distribution function, as discussed in Section 2.2.2, is set of functions that give the probability of an interaction with a parton in the proton for a given momentum transfer. It is inserted into our amplitude calculation to allow for the different types of decays allowed by the Feynman diagrams. Our three top signal requires a b quark for the initial parton in many decay modes, so it is sensitive to the PDF modeling. For CMS production, each PDF used comes from the LHAPDF library using primarily the

NNPDF 3.1 PDF set which is a five-flavor QCD scheme [29, 86]. These PDF uncertainties are difficult to calculate analytically, first because of numerous input features required for PDF modeling, and second, because the PDF is not analytically determined to begin with, so uncertainties have to be estimated based on modeling of previously taken data. The PDF is a set of functions, but it depends only on the parton flavor, energy scale, and scattering momentum. These values are set by the particular Feynman diagram and momentum of the event, so the PDF is simply a number in our calculation. This means the uncertainty modeling can be greatly simplified to MC method that varies the PDF model to create a distribution of possible PDF values. In our particular case, 100 of these different PDF variations are produced for each event and the error is derived from the standard deviation of these different "replica" PDF conditions, as they are called.

Separate from varying the PDF directly, we can also vary the strong coupling constant, α_S, which alters the coupling with the PDF. From this, there are two main PDF uncertainties derived and used in the analysis.

Cross-Section

Cross-sections, because of their fundamental nature in QFT calculations, are easily produced by the simulation to almost arbitrary precision. While he cross-section are typically calculated with the sample itself, more precise values, usually produced with higher order methods, are used in their place. While utilizing a more accurate cross-section should decrease the uncertainty in the process's yield, the lower order simulation will not have the same kinematic shape as its higher order version. Even so, generating MC samples at higher orders requires significantly more computational, so the expected difference is simply folded into the cross-section uncertainty. This means the cross-section uncertainty is a combination of this higher order uncertainty, any observed uncertainty from previous Standard Model measurements, and, most importantly, any yield uncertainty coming from the MC generation.

For this last point, we use the cross-section uncertainty to help in stabilizing the fitting to data, The other theoretical uncertainty sources discussed are quite conservative (using the ×½ and ×2 paradigm for uncertainty bands) and are expected to change the kinematic shapes and event yield. Instead, we opt for a procedure in which the other theoretical uncertainties have their up and down variations normalized to the nominal yield so they merely change the shape, removing any yield variations. This allows the cross-section uncertainty to takes care of any uncertainty in the absolute cross-section. This paradigm helps decorrelate these shape and normalization effects which can sometimes cause issues in the fitting. To compensate, all theoretical uncertainties use a fairly conservative error estimate, but after minimization, the best fit uncertainty is typically smaller, reflecting the "true" uncertainty. The cautious errors allow leeway for the minimization to not become over-constrained and be forced into a poorer fit.

9.2.2 Event Level Uncertainty

Event level uncertainties refer to effects that influence all events similarly, typically from global sources instead of event specific uncertainties. There are three sources categorized as such: the luminosity, Level-1 (L1) trigger prefiring, and pileup uncertainties. The luminosity uncertainty is applied as a flat normalization rate whereas the prefiring and pileup uncertainties have a shape dependence.

Luminosity

Luminosity, as discussed in Section 3.2.7, takes hits found in several subdetectors and converts these into an estimation of the instantaneous luminosity. From the assumptions made and the measurement tools themselves, there may be slight deviations from the actual luminosity, and this effect may differ from year to year as different calibrations (namely Van der Meer scans) were used. The actual luminosity uncertainty was decided by experts to be split into three separate components: one for the full Run 2, one specific to each year (or uncorrelated), and one solely for 2017 and 2018. The 2017–2018 uncertainty is added to decorrelate these years slightly from 2016 because it reflects the changes in the luminosity measurement techniques after 2016. The correlation percentages are listed in Table 9.1.

L1 Trigger Prefire

The L1 trigger prefire issue was a problem found throughout Run 2 which caused a slightly smaller data acceptance than usual. The basic problem was caused by both a timing mismatch in the ECAL and a limited timing resolution in the muon detectors. Both individually caused timing information for these subsystems to be out of sync from the other detectors. Collisions take place every 25 ns, but we don't expect most of these events to contain interesting physics. This is what justifies reducing the event rate from 40 MHz to 100 kHz in the L1 trigger, and implicit in this assumption is back-to-back collisions that pass our trigger are expected to be rare.

The L1 trigger rejects these sequential triggered events which helps to solve the issue of particle travel time. Most particles produced in each collision will not have fully decayed before the next collision starts, so particles from multiple collisions will coexist in the detector at any given time. To resolve which particles go with which events requires precise timing information and the removal of sequentially triggered events reduces the error rate of associating particles to the wrong event. The prefire issue, in which the time information was slightly off, led to some particles being mislabeled with the wrong collision causing accidental triggering of the neighboring events thus rejecting real, triggered data. While the drift in the ECAL timing was fixed in 2016, the muon resolution issue was present throughout Run 2. The overall effect is small, but MC has to be adjusted to account for this reduced data efficiency. This is achieved by using a scale factor corresponding to the expected prefire probability for a given event. To account for

any further discrepancy between data and MC coming from the prefire issue, systematic uncertainties on this scale factor were also calculated and applied in our fit model.

Pileup

Pileup, which is the collection of extra jets in the event is produced in different quantities from event to event. While generally unrelated to the total event kinematics, reconstruction algorithms as well as certain IDs are sensitive to pileup since results will differ depending on the ability of these algorithms to identify and reject said pileup. To fix the difference observed in the number of pileup, scale factors were derived by luminosity experts to scale MC events and remove the data/MC disagreement. These scale factors had uncertainties derived as well to account for variation of pileup in different phase spaces.

9.2.3 Object Reconstruction Uncertainties

Object reconstruction uncertainties correspond to methods used in reconstructing various physics objects as described in Section 6. Generically, if there is some scale factor that is applied to correct for data/MC agreement for a certain particle object ID, there is an accompanying uncertainty to account for the error linked with these scale factors.

Identification

Leptons and jets have multiple IDs that are applied to account for efficiency differences in data and MC, so uncertainty bands are also derived and applied. Most are centrally produced as shape uncertainties with the leptons having uncertainties from the muon ID and electron MVA ID while jets have the pileup jet ID, all being applied and used as recommended by the experts.

The ttH−MVA ID, on the other hand, is a bespoke scale factor produced specifically for this analysis. In addition, the L1 trigger and HLT acceptance for leptons will generally be different from the offline selection, so trigger scale factors were similarly calculated to account for the differences between trigger acceptances. Both uncertainties, we opted for an enveloping flat uncertainty. The other scale factors were derived by experts, so they had thorough studies performed to explore the nuances of all sources of uncertainty on the different kinematic variables. This analysis, which is expected to be dominated by other factors such as jet uncertainties, found it unnecessary to derive similar shape uncertainties, thus using simple flat uncertainties derived from basic uncertainty calculations is more than justified.

DeepJet Uncertainty

Bottom jets are tagged using an algorithm called DeepJet using the iterativeFit paradigm. This means the overall b-tagging MVA has its whole shape corrected for, not just its yield, to match data. While this is very useful for allowing the raw DeepJet MVA score to be used in our BDT training, it does come at

the cost of a larger systematic uncertainty. The DeepJet uncertainty is split into eight sources categorized by the true jet-type, light flavor (LF), heavy flavor (HF), and charm (CF), to account for differences in response to the MVA. Light flavor jets are significantly different from b-jets so are usually very easy to distinguish. Heavy flavor corresponds to the response of true b-jets which we would expect to have a high DeepJet score. Charm jets lie in an interesting spot where their masses and lifetimes are similar enough to b-jets to be labeled as one easily. Much of the effort of b-tagging actually goes into distinguishing bottoms from charms since lighter flavor jets can effectively be distinguished by typical secondary vertex and impact parameter requirements.

For the uncertainty implementation, the HF and LF uncertainties are further split into three different uncertainties: one to account for general systematic uncertainties in the scale factor extraction and two to account for the statistical error in the measurement and scale factor calculation. Normally, statistical uncertainties are handled by a single uncertainty, but the scale factors are calculated using more complex, polynomial form to best fix the overall shape of the MVA, so the statistical error is split into two sources for first and second order effects. For the charm uncertainties, there are only two sources calculated differently to account for the different way charm jets are treated in the DeepJet algorithm.

The scale factors only exist to fix the shape of the b-jet multiplicity and MVA distributions, but in their application, it often causes the normalization to change. To fix this, the total yield for each sample is fixed to match the yield before the b-tagging scale factors are applied. This means the scale factors only affect the b-tagging MVA shape though the scale factors will impact the total yields after some b-jet multiplicity requirement is applied. This normalization process is similarly applied to all b-tagging systematics.

JEC Uncertainty

The JECs introduce one of the more complex uncertainty sources, being split into 11 sources plus an addition uncertainty for the jet energy resolution. Most of the sources deal with different responses in the different parts of the detector (barrel, endcap, versus the front portions of the HCAL), while the others correspond to the general correction factors used to reconstruct the true, prompt level energy (as shown in Figure 6.2). Because of the different natures of the uncertainties, some are correlated amongst all eras while others are not. The uncertainties slightly shift each jet's energy leading to the interesting consequence of causing certain events to pass or fail our preselection and signal requirements because of changes to the H_{T} and jet multiplicity. This is opposed to all other systematics which are implemented using event weights, thus the different variations don't change the raw event count, only the weights used to calculate the final yield. This curiosity means these systematics require extra care in application.

Further, the effects of changing jet p_{T} have consequences on other quantities, most notably MET. MET must be recalculated for every JEC uncertainty, but, because it is a composite variable, the variations can have seemingly erratic changes. Looking at overall shape uncertainties, the MET distributions are

reasonable and the systematics do not highly affect the event acceptance. The DeepJet MVA is also sensitive enough to jet p_{T} variations that additional scale factors were derived for all JEC variations, and they are used in the analysis.

9.2.4 Data-Driven Background

There are two data-driven background methods used, nonprompt and charge-misId. Because these are data-driven methods, their uncertainties only arise from the ABCD method. While there are multiple potential sources to the error in the background estimation, we use a simple flat rate error to be conservative, estimating this uncertainty from the observed differences seen in the closure region: 30 % uncertainty applied to the nonprompt background and a 20 % uncertainty applied to the charge-misId background. The error rates are larger than the final expected, but it should be constrained to a more reasonable value after fitting.

For the nonprompt misidentification rate, there is the addition consideration of lepton flavor since the rate is measured individually for nonprompt muons and electrons. To decouple the different lepton uncertainties, the associated statistical errors in the two different misidentification rates for the leptons are also included. These uncertainties give the fit more room to adjust the rates between flavors as well as make fine adjustments in each $p_{\mathrm{T}} - \eta$ bin, constrained by the statistical errors in the rates observed.

9.2.5 Template Smoothing

A last step to ensure proper handling by the fitting software, we perform a smoothing on some systematic uncertainties. We generally expect our uncertainty to have some function form that smoothly deforms our kinematic variable from nominal shape to the shapes of the up and down variations[†]. Since these variations should also represent first order effects, the up and down variations should be roughly symmetric; this is not generally true, but for most uncertainties, this is the case. For judging the overall smoothness, it is often useful to look at the ratio of the up/down variation to the nominal distribution where it is more straightforward to see the relative smoothness and symmetrical between the two uncertainty bands.

These assumptions are not always found in our systematic. Because our distributions are created from discrete events that are binned into histograms from kinematic variables that have uncertainties, we have to contend with bin-by-bin errors. Bin-by-bin errors here correspond to changes to variable shape based on the binning used for the histogram. This can take the form of both a histogram not changing even when the input variables change because the changes are smaller than the resolution of the binning or histograms changing significantly between the same data because the binning could split data differently. Put differently, most uncertainties we have discussed are related to changing the total

[†]In more mathematical language, we might expect our uncertainty to parameterize our variable of interest, $f(x)$, such that for $t \in \mathbb{R}$ we can create a function $g(x,t)$ where $g(x,0) = f(x)$ and $g(x,\pm 1)$ are our up and down variations.

yield for a given bin, therefore being an error in the y-axis, while the bin-by-bin errors relate to the errors induced by the binning and relate to the error in the x-axis for our plots.

Most systematics don't have sizable bin-by-bin error, but our JEC uncertainties change the overall raw event acceptance as well as actually changing the kinematics of each event (in opposition with other systematics which only have an event weight), so our uncertainty bands have observed bin migration and overall jitter. This additional noise leads to the ratio plots varying wildly between a positive and negative changes that are not smooth or symmetric. This poses a problem for our fitting algorithm which has more difficulty seeing how events flow and change in the uncertainty bands meaning the fit demonstrates odd behavior or event fails. To remedy this, we use smoothing on these problematic histograms.

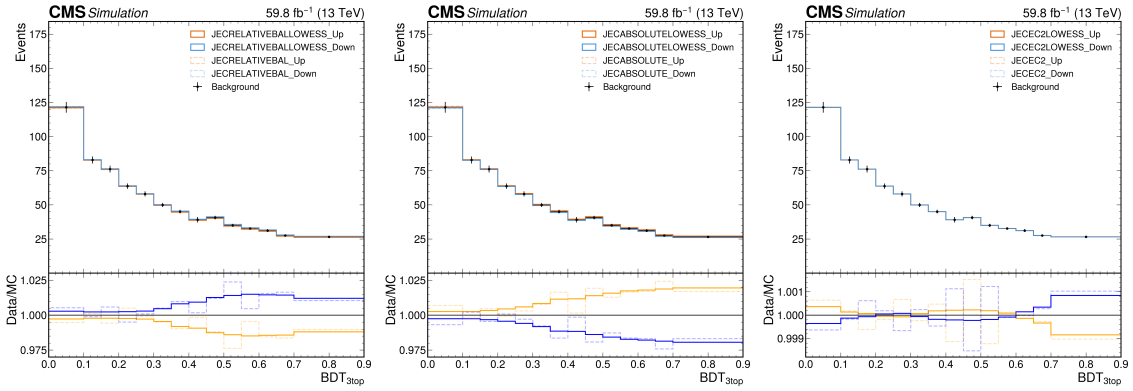

Figure 9.3: Example band plots for three JEC uncertainties in the dilepton region for 2018. The dotted line shows the original up and down variations while the solid line is the LOWESS smoothed variation.

For the smoothing, we use an algorithm called Locally Weighted Scatterplot Smoothing (LOWESS). In essence, it replaces all the values with a moving weighted average of the data within a certain window [87]. The wider this window, the smoother the result, the larger window averaging out the larger deviations. For implementation, the smoothing is performed on the ratio plot, so the effect size of the systematic is smoothed and not the shape in such a way that could cause the variation for a particular bin to be artificially large. In addition, the variations are symmetrized by taking the average the up and down variations to avoid any strange behaviors and ensure consistent results. As mentioned previously, the smoothing is applied to the JEC systematics and any odd systematic that causes trouble in the minimization step. Looking at output of the smoothing on a subset of systematics in Figure 9.3, the new uncertainty band captures the essence of the calculated variations while removing large, erratic swings in the uncertainty ratio, thus allowing for better fitting behavior. Especially in the EC2 systematic, shown on the far right in Figure 9.3, the variation flips from a positive to a negative variation, and the smoothing removes this wild behavior in favor of more sensible values.

9.3 Statistical Tests

Once we have our systematics uncertainties understood and modeled, everything is prepared for creating our likelihood ratio, so we can use it to perform our typical tests

9.3.1 Expected Results

Before looking at the data, we can create approximations of what we expect to see. This might be interesting because if the results from our data are very different from our expected results, we may suspect our overall modeling may be incorrect.

To generate these expected results, we need fake data to run over. As mentioned previously, we can generate fake data through MC techniques to make toy data, but often we are interested in the most likely result, not a specific possibility. For creating the most probable event, but this is just the one that maximizes our likelihood function, we use the Asimov dataset, or the values given by our MC simulation. Our model is build using the MC values as the template, so it should be no surprise that the expected most probable case is exactly the MC's results. Using Asimov dataset, we can calculate a significance and limits. In addition to the expected limits, we can also use toy data to calculate the CLs 95 % limit on μ for hundreds of thousands of cases. The distribution of the limit values allows us to find a $\pm\sigma$ and $\pm 2\sigma$ band for our expected value. From this, we can more accurately assess how well our data matches our expectation.

9.3.2 Fitting

Once we unblind ourselves and look at real data, the first step is to fit our model to it. We previously plotted our template histograms with data in Figure 8.7, but this was the raw output of our MC simulations and our uncertainty estimations and are thus called the pre-fit distributions. After fitting the free parameters of our profile likelihood function, the yields of each sample and total uncertainty will change to reflect the results of the fit. The output of the fit (i.e. the post-fit distributions) are shown in Figure 9.4. The agreement in the post-fit distributions shows marked improvement as the fit moves the different background groups around to best match the observed data and reducing the overall error bands, matching the expectation that the fit would constrain certain uncertainties that were overestimated.

Similarly, looking at the breakdown of the post-fit event yields in Table 9.3, we see overall good agreement between data and MC after the fit. Numerically, its obvious that the two control regions have cordoned off large proportions of four top and ttZ events into their respective regions as expected, where roughly 60 % of ttZ is contained in the ttZ control region and 25 % of the four top in the four top control region. Nonprompt and ttW dominate the background yield, but the nonprompt background is safely separated from the signal significant areas of the phase space while ttW, which has more contamination

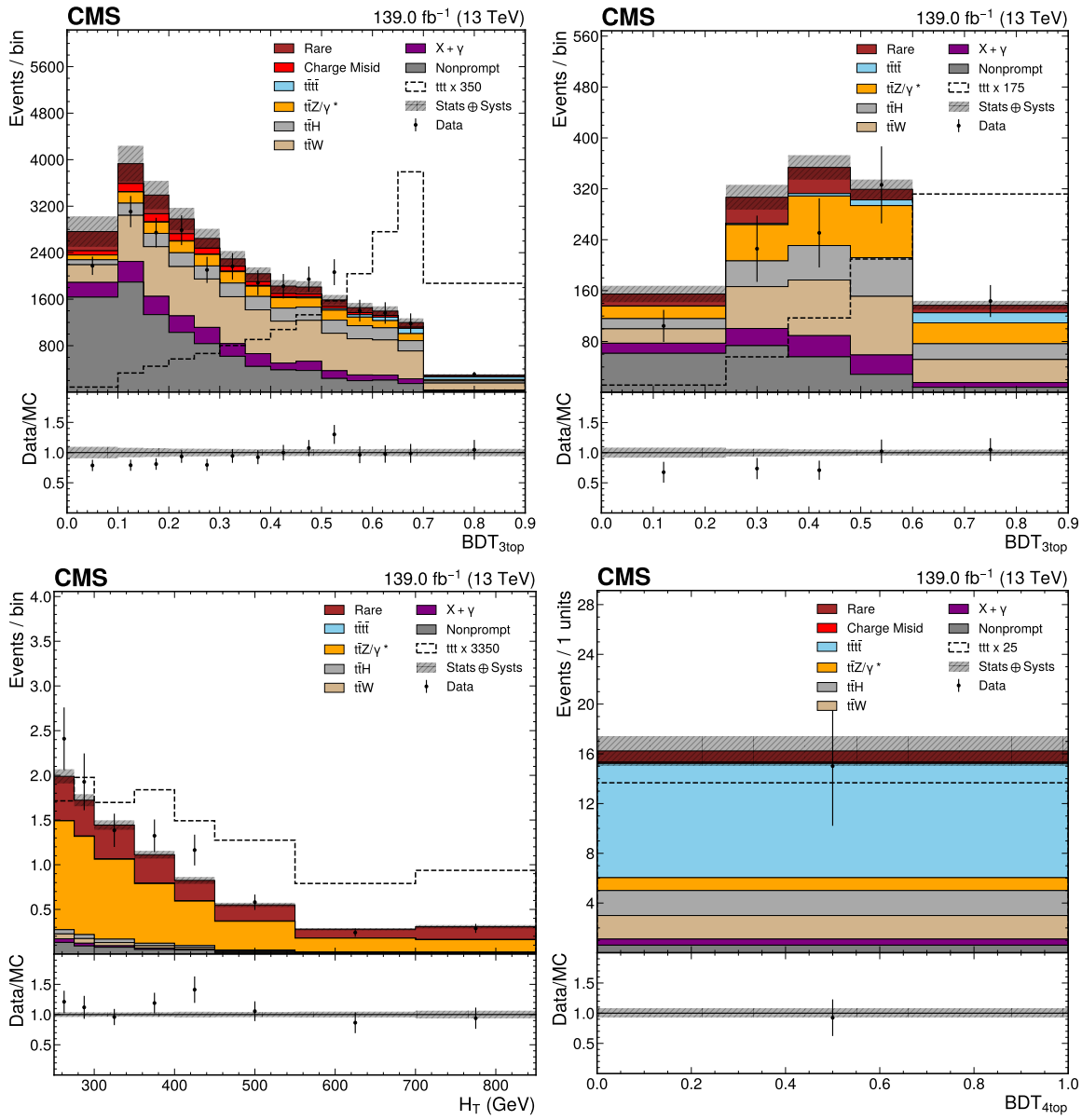

Figure 9.4: Final template histograms for the full Run 2 after the fit. Starting from the top left going clockwise, the plots show the 3top BDT for the dilepton region, multilepton region, the 4top BDT in the four top control region, and the ttZ control region.

in the high BDT score regions, has a much flatter shape, allowing the fitting algorithm to distinguish the two much more easily.

Table 9.3: Breakdown of total events and errors for each background groups in each region for full Run 2 after fitting.

Process	Dilepton	Multilepton	ttZ CR	$t\bar{t}t\bar{t}$ CR
Nonprompt	408.30 ± 13.70	26.54 ± 1.64	13.85 ± 0.72	0.47 ± 0.09
ttW	483.23 ± 7.11	43.61 ± 1.28	9.59 ± 0.26	2.04 ± 0.24
ttZ	138.24 ± 2.94	48.36 ± 1.74	284.62 ± 7.82	1.28 ± 0.16
ttH	145.14 ± 2.24	28.52 ± 0.83	9.72 ± 0.25	2.12 ± 0.20
X + γ	141.20 ± 4.52	15.87 ± 0.73	4.49 ± 0.25	0.49 ± 0.08
Rare	121.20 ± 3.66	17.24 ± 0.95	109.67 ± 3.64	0.87 ± 0.19
Charge-MisId	51.31 ± 1.77	–	–	0.09 ± 0.01
$t\bar{t}t\bar{t}$	23.57 ± 0.78	6.12 ± 0.34	2.04 ± 0.06	9.33 ± 0.67
Total Bkg.	1512.19 ± 15.22	186.26 ± 3.44	433.99 ± 7.99	16.69 ± 1.06
ttt	3.37 ± 4.84	0.77 ± 1.95	0.21 ± 0.27	0.56 ± 1.74
Data	1505	164	438	14

9.4 Final Results

After the fitting process, we can run the final statistical test to get our results. For the significance, we find a value of 0.182σ for full Run 2 results. Because this has not reached the necessary significance level, we cannot reject the null hypothesis and confirm the existence of the Standard Model three top process. While this result has not reached the 5σ point yet, more data will lead to discovery if the three top process is not spurious.

For the next step, we can calculate the observed upper limit. After running our test, we find a 95 % CLs upper limit on the cross-section of 28.1 fb. Comparing to the Standard Model cross-section of 2.05 fb, that equates to an upper limit of roughly $\mu = 15$. All of this is shown graphically in Figure 9.5. The limit plot shows good agreement of the data with the expected result, being totally within the 1σ band, so everything matches our prediction. While these results are generally small and not reaching the discovery threshold, it reflects the current limitations of the three top measurement, namely our small signal yield coupled with the high contamination of backgrounds and uncertainty in our signal sensitive regions.

9.5 BSM Interpretation

Included in this analysis is an evaluation of different BSM extensions discussed in Section 2.2.4. For most of the BSM models that induce enhancements to the three top cross-section, they are primarily caused by a flavor violating neutral current, or FVNC. This takes the form of a neutral particle that allows for flavor changing in its coupling, similar to the CKM matrix allowing for interactions outside the normal

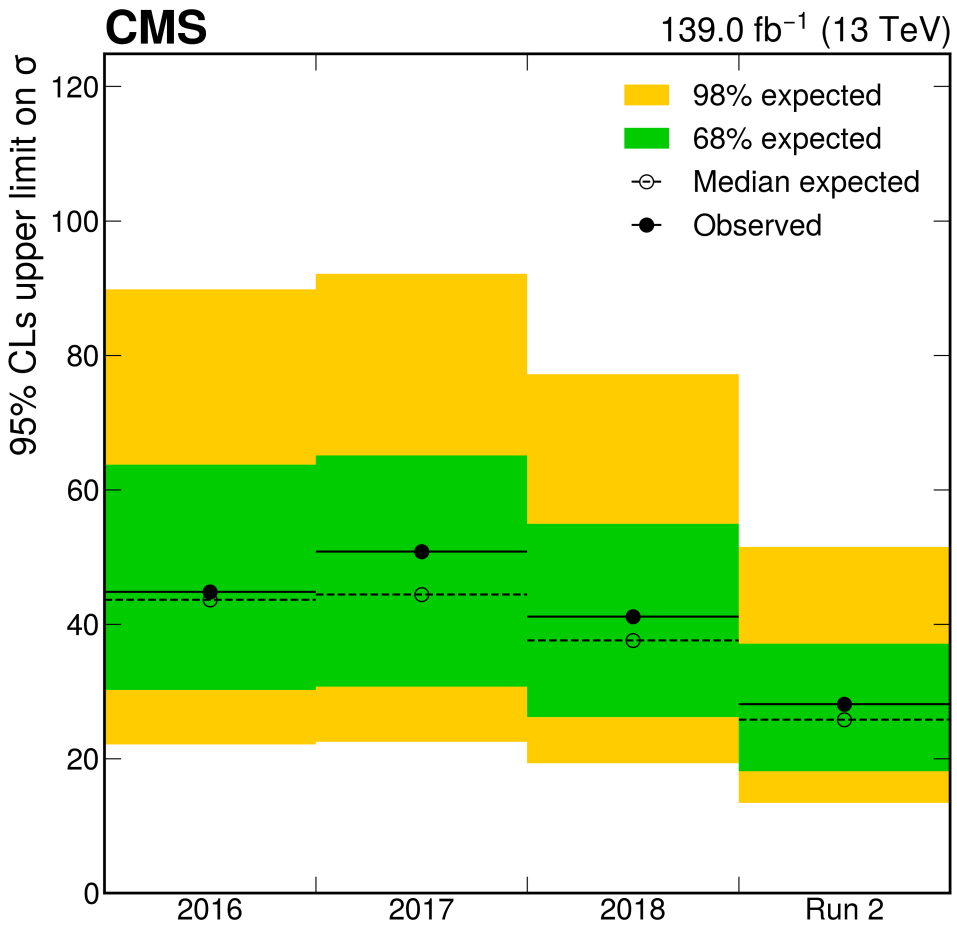

Figure 9.5: Plot of the expected and observed CLs 95 % upper limits on the three top cross-section. It is binned per year as well as for full Run 2 to show the change from year to year. All observations fall within the expected band showing good agreement.

weak doublets. This is in opposition to the standard models conception of the neutral currently, i.e. the Z boson, which conserves quark flavor in all interactions.

For studying these BSM models, a preferred way is to perform the hypothesis testing for each, leveraging each model's particular features and specific kinematics. For three top, the expected yield for the Standard Model process is small, and the uncertainties are large. This means our signal is fairly insensitive to the kinematic variations we'd expect from a BSM model, so these models can be effectively treated as simple augmentations to the typical three top cross-section. With this knowledge, the results found in our Standard Model measurement can be reinterpreted for a BSM model, allowing us to place limits of the cross-section for a BSM model with three top cross-section enhancement.

The papers by Barger et al. and Khanpour focus on FVNC models at 14 TeV as well as for future collider energies, making a direct limit setting impossible [43, 45]. The paper by Khanpour particularly

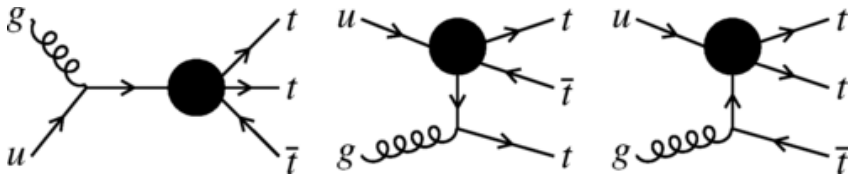

Figure 9.6: Basic Feynman diagrams that can produce the three top process using the uttt EFT coupling, shown as the large black circle. These diagrams are taken directly from the paper by Cao et al..

looks at possible flavor violation in the Standard Model itself look at top-gluon-quark (tqg) and top-quark-Higgs (tqH) couplings where similar increases could come by some FVNC mechanism. Current CERN studies have placed limits on these branching fractions, but three top is an interesting probe because three top can be produced uniquely by all these couplings, causing an enhancement to the overall cross-section. Most of these branches fractions have limits set to sub-percent levels, so even if the paper considered LHC energy scales, our limits would not be competitive. While not currently sensitive, in higher energy and luminosity accelerators, the three top may become a useful probe for these processes.

In Barger et al., they consider two theories, a Z′ theory and a minimal supersymmetric, or SUSY, theory. The Z′ theory proposes a similar FVNC extension where the Z′ is leptophobic, or doesn't couple to leptons, and biases its interactions towards coupling with later generation quarks. These constraints act as necessary additions because without them, we would have expected to see Z′ decay modes in single top processes. For the SUSY model, the large potential phase space supersymmetry affords us to consider gluino decays that have FVNC properties meaning a gluino-antigluino process can produce the necessary three top quarks in the final state.

Both the Z′ and SUSY models are only calculated at 14 TeV, leading to incompatibility for comparison, but they calculated a new three top cross-sections of 28 fb for the Z′ theory and 41 fb for the SUSY model. We are unable to make exclusion limits for these theories, but we can note that cross-sections will typically increase with center-of-mass energy, so our current limit of 28 fb would most likely exclude these models to the 95 % level if recalculated at $\sqrt{s} = 13$ TeV.

From the paper by Cao et al., they propose a FVNC effective field theory in the form of a three top plus up quark coupling (uttt) [44]. This four-point interaction, shown in Figure 9.6, naturally allows for a flavor violating (f_{FV}) and flavor conserving (f_{FC}) term (for the ut and tt couplings respectively). Because this is an effective field theory (EFT), it is not fully descriptive of all energy scales, so there is an associated energy scale, Λ, of its validity. An EFT essentially works by integrating out higher energy interactions to leave a simpler theory that can be tested. This means our scale, Λ, acts as the natural energy scale necessary for new physics, i.e. where the EFT breaks down, and a new theory must take over. The value of Λ not only can hint at what scales new physics is expected, but it hints at the mass of a potential new particle that would mediate the theorized interaction. These newly proposed ut and tt couplings can be mediated by a scalar or vector like interaction, both leading to slightly different physics.

Looking at 13 TeV CERN interactions, they calculated the modified three top cross-section as:

$$\sigma_{ttt}^{\mathcal{V}} = 0.9582 \times \left(f_{FV}^{\mathcal{V}} f_{FC}^{\mathcal{V}}\right)^2 \left(\frac{\text{TeV}}{\Lambda}\right)^4 \text{pb}$$
$$\sigma_{ttt}^{\mathcal{S}} = 0.3131 \times \left(f_{FV}^{\mathcal{S}} f_{FC}^{\mathcal{S}}\right)^2 \left(\frac{\text{TeV}}{\Lambda}\right)^4 \text{pb}$$

Using our observed final limit of 28.1 fb (any noticing the unit difference), we find the following inequality:

$$f_{FV}^{\mathcal{V}} f_{FC}^{\mathcal{V}} \leq 0.171 \times \left(\frac{\Lambda}{\text{TeV}}\right)^2$$
$$f_{FV}^{\mathcal{S}} f_{FC}^{\mathcal{S}} \leq 0.300 \times \left(\frac{\Lambda}{\text{TeV}}\right)^2$$

If we assumed $f_{FV} f_{FC} = 1$, or the basic scaling for our coupling constants, we can find a lower bound for the energy scale of this proposed interaction:

$$\Lambda^{\mathcal{V}} \geq 2.4\,\text{TeV} \qquad \Lambda^{\mathcal{S}} \geq 1.8\,\text{TeV}$$

Compared to the paper in question, which gives a 95% upper bound on the energy scale as $\Lambda^{\mathcal{V}(\mathcal{S})} \geq 0.910(773)$ GeV, our limits are much more stringent. For theory papers such as this, their limits are simple estimations which require many simplifying assumptions (particularly because the theoretical model is usually of more interest, not the estimated limits), but this does show the power of the statistical methods used by CMS and this analysis for making quite stringent limits. From these results, we can see that this generic EFT uttt coupling excludes all but a TeV scale BSM particle to exist which would require either significantly more data or energy to evaluate.

In summary, even with small overall final significance, our model is able to set fairly strong limits that can use to evaluate current theories proposed about this signal process.

Chapter 10

Conclusion

This analysis represents the first of it kind, probing the Standard Model looking for three top quark production at the LHC. We looked at both two same-sign leptons and multilepton events consistent with the three top quark process using machine learning techniques for signal extraction. The analysis was performed using Run 2 data, corresponding to 137 fb^{-1} of data, collected from the CMS detector at $\sqrt{s} = 13\,\text{TeV}$. From this data, we found an overall significance of 0.182σ and a 95 % CLs upper limit of 28 fb corresponding to no observation of the three top quark process.

While this result has not reached the 5σ point yet, the addition of more data will bear out to discovery, of course if three top process exists as specified in the Standard Model. A crude approximation of significance is the figure of merit or $S/\sqrt{B}$. Here, the simple derivation is: ignoring all systematic errors and considered cut-and-count experiment, the statistics would be modeled solely by a Poisson distribution. Because the mean and variance for a Poisson distribution are equal, the statistical error is the square root of the number of events observed. The expected significance for rejecting the null hypothesis, which is just the z-score, should simply be the amount of events in excess of the background only hypothesis (which is the number of signal events) divided by the uncertainty of our expectation. This nicely leads to the relation:

$$z = \frac{x - \mu}{\sigma} = \frac{(B + S) - B}{\sqrt{B}} = \frac{S}{\sqrt{B}}$$

as needed. We can see that as the luminosity grow, the yields of both the background and signal increase leading to the figure of merit increasing at a rate of $\sqrt{\mathscr{L}}$. This naturally means more data will increase the significance, so discovery is a matter of having enough data for the measurement's sensitivity. For our case, we expect our actual significance to be larger than the figure of merit in spite of the added uncertainties because we use shape information and the more complex likelihood ratio.

Therefore, the outlook for the future discovery looks quite bright. Currently, the LHC is approaching the end of Run 3 collisions which has collected over 300 fb^{-1} of data at a center of mass energy of 13.6 TeV. There is still the rest of 2025 and 2026 for data collection, so Run 3 looks to collect more data than Run 1 and 2 combined. While increased luminosity leads to more events, higher energy gen-

erally increases the rate for processes meaning the overall significance for the three top signal in a future analysis will only improve.

After the long shutdown starting at the conclusion of Run 3, the LHC will be undergoing upgrades for the High Luminosity LHC, or HL-LHC. The HL-LHC is the project name for the project of upgrading the LHC to achieve luminosity levels that are 5-10 times current levels, meaning over the next 10 years (or what will be Run 4 and Run 5), the LHC is expected to produce 3000 fb^{-1} of data. Using the crude figure of merit formula from before to calculate an expected significance, Run 3 data is expected to find at least a significance 35 % larger than our current results and, in the lifetime of the HL-LHC, it will at minimum increase the significance by ×5.

Current work is being performed to expand this analysis to Run 3 results to both increase current results and more thoroughly investigate possible BSM models that were only touched upon with this analysis. Because this analysis is a first of its kind, there are multiple avenues for improvement and optimization that are in the works for more competitive results. With Run 3 algorithms, identification of particles continues to improve leading to higher efficiencies. On that note, machine learning tools have also allowed for high performant and reliable taggers for composite particles such as W boson and top quark tagging which a future analysis could use for better background separation. Further, with the knowledge gained from this first analysis, better, more complex techniques can be employed for signal extraction and modeling of uncertainties. With this simple list of improvements for the Run 3 analysis, expected results can only improve.

In the more distant future, the HL-LHC harkens to a new era of precision measurement for the Standard Model and possibly even the discovery of new physics to answer the current questions of particle physics today. While this analysis did not reach the threshold for discovery of CERN of 5σ, the HL-LHC, with its boasted tremendous data output, would expect to advance current results by almost an order of magnitude. This, coupled with improved identification, analysis methods, and modeling, leads to promising outlooks for a three top quark discovery in the HL-LHC's lifetime.

Appendix A

Gauge Theory

Gauge theory is much deeper and complex than explained in the main body of this thesis, so this appendix adds more details of gauge theory and in the Standard Model.

The essence of the Standard Model interactions is gauge theory because all interaction in it are mediated by gauge fields, where a gauge field is simply a field that is invariant under a gauge transformation. Gauge transformations, which were originally defined in classical electromagnetism, refer to a type of transformations which leave the physics of the system unchanged. Since these transformations do not change our system, we can view them as a symmetry. Noether's theorem states that symmetries give rise to conversation laws [11], it becomes very interesting to understand our quantum mechanical system through these various gauge symmetries.

As a first step, we can look at the typical example. In classical electromagnetism, one can reframe Maxwell's equations by expressing the electric and magnetic fields in terms of a new 4-vector potential field $A_\mu = (\phi, \vec{A})$ such that:

$$E_i = -\partial_0 A_i - \partial_i \phi$$
$$B_i = \epsilon_{ijk} \partial^j A^k$$

While the A field in some sense derives our observables, the electric and magnetic fields E and B, it is not directly measurable. The 0^{th} entry, ϕ, which is the voltage, has no absolute value because only the *difference* in voltage can be measured, and the magnetic vector potential, $\vec{A}$, has a similar leniency in definition. This freedom comes from, in some ways, a lack of degrees of freedom in our theory. While the E and B fields have six distinct components, Maxwell's equations constrain them to only two independent components. Since our vector potential field has four components, this means it necessarily has redundancy. One can impose so-called "gauge" conditions that constrain our vector potential, though, we can express them more generally as classes of transformation that keep the underlying physics the same. This gauge condition can be written as $A_\mu \to A_\mu + \partial_\mu \Lambda(x)$ where we have freedom in our choice of $\Lambda(x)$, allowing us to change the particular "gauge" we consider, usually for mathematical necessity for

a particular calculation. Therefore, these gauge transformations act as a sort of symmetry in our vector field.

For quantum field theory, similar gauge symmetries can be found in the three different measured forces: strong, weak, and electromagnetism, but they are more complicated. To build up similar gauge transformations in quantum field theory, we can start again with electromagnetism. First, we need the Lagrangian for particle physics with electromagnetism included:

$$\mathscr{L} = \overline{\psi}(i\partial\!\!\!/ - m)\psi + e\overline{\psi}A\!\!\!/\psi - \frac{1}{4}|F|^2$$

The first term is the Dirac Lagrangian or the free particle kinematics, the second term is the particle interaction with the electromagnetic field, and the last is the electromagnetic self interaction term. F, here, is the classical electromagnetic tensor and its indices are simply the E and B fields, so this term is totally invariant under the previous gauge transformation. Unfortunately, the same gauge transformation does not leave the interaction term invariant:

$$\begin{aligned}
\mathscr{L} \to\ & \overline{\psi}(i\partial\!\!\!/ - m)\psi + e\overline{\psi}A\!\!\!/'\psi - \frac{1}{4}|F'|^2 \\
&= \overline{\psi}(i\partial\!\!\!/ - m)\psi + e\overline{\psi}\left(A\!\!\!/ + \gamma^\mu \partial_\mu \Lambda(x)\right)\psi - \frac{1}{4}|F|^2 \\
&= \mathscr{L} + e\overline{\psi}\gamma^\mu \partial_\mu \Lambda(x)\psi
\end{aligned}$$

The only way to compensate this is if this gauge symmetry is related to some symmetry in the wavefunction itself. We expect the wavefunction to be symmetric under some transformations (an obvious example being Lorentz invariance), so we might suspect it might be similarly symmetric under gauge transformations to reproduce the classical result. To build up to this, we might consider a generic phase transformation, $e^{i\phi}$, which keeps our Lagrangian the same. This phase transformation is called a global phase transformation. We can expand this to local transformation by allowing our phase to depend on the position, or:

$$\psi \to e^{i\alpha(x)}\psi \approx (1 + i\alpha(x))\,\psi$$

Where, for a small phase, we have the approximation shown. This construction is the basics of a Lie group transformation, or the group that describe continuously differentiable transformations that satisfy some symmetry conditions. In our case, since quantum mechanics requires unitarity and for our wave function to be normalized, our group transformations are from the special unitary Lie group, $\mathrm{SU}(N)$, where N is the dimension of the group. Applying these Lie transformations to our wavefunction, most

terms allow the Lie transformation to pass through except the derivative:

$$\begin{aligned}
\mathscr{L} \rightarrow\ & \overline{\psi}'(i\partial\!\!\!/ - m)\psi' + e\overline{\psi}'A\!\!\!/\psi' - \frac{1}{4}|F|^2 \\
&= (e^{-i\alpha(x)}\overline{\psi})(i\partial\!\!\!/ - m)(e^{-i\alpha(x)}\psi) + e(e^{-i\alpha(x)}\overline{\psi})A\!\!\!/(e^{i\alpha(x)}\psi) - \frac{1}{4}|F|^2 \\
&= e^{-i\alpha(x)}\overline{\psi}\left[(i\partial\!\!\!/ e^{i\alpha(x)})\psi + e^{i\alpha(x)}(i\partial\!\!\!/\psi)\right] - m\overline{\psi}\psi + e\overline{\psi}A\!\!\!/\psi - \frac{1}{4}|F|^2 \\
&= \mathscr{L} - \overline{\psi}\left(\partial\!\!\!/\alpha(x)\right)\psi
\end{aligned}$$

So again, our Lagrangian isn't invariant. One way to remedy this is to connect this Lie group transformation acting on the wavefunction to the gauge transformation acting on the vector field. In fact, if we require $\alpha(x) = eA(x)$, the terms cancel, and our Lagrangian is invariant meaning we can reproduce gauge theory through the marriage of local Lie group transformation with the classical gauge transformation. This allows us to associate each force with a local symmetry group of the wavefunction that dictates the structure of the gauge force. The summation of this theory of gauge forces in particle physics is called Yang-Mills theory [88].

To extend this result to other forces, we must more generally consider the Lie structure. A Lie group is characterized by its Lie algebra which itself is a group made of a set of operators, T_a, which obey the multiplication rule of the Lie bracket, $[T_a, T_b] = if_{abc}T_c$, where the multiplication is a defined the structure constants, f_{abc}. The electromagnetic example was much simpler in structure because the underlying group structure was Abelian, i.e. commutative, but the $\mathrm{SU}(N)$ groups are generally non-Abelian. This Lie algebra is a local description of the Lie group where the full structure can be reproduced through the exponential mapping: $G \equiv \exp\left(ig\frac{\theta_a}{2}T_a\right)$ for $G \in \mathrm{SU}(N)$, where θ_a is the rotation in the a$^{\text{th}}$ direction and the g is a constant placed for convenience[†]. This exponential mapping works because the Lie group is continuous and differentiable, so any transformation can be created by repeated application of the local transformations from the algebra.

Once this correspondence is created, our gauge field can be represented as an element of the Lie algebra, or $A^\mu = A^\mu_a T_a$. When our theory is quantized, we find that each basic operator, T_a, of the Lie algebra act as the axes for the force, so each element of the algebra corresponds to a particle associated with the force. The non-commutativity of our generalized case does mean we have to be more careful in defining our gauge transformation. Starting from our wavefunction transformation, $\psi \rightarrow G\psi$, it can be shown that our non-Abelian gauge transformation must have the following form to ensure our Lagrangian is invariant:

$$A^\mu \rightarrow GA^\mu G^{-1} - \frac{i}{g}(\partial^\mu G)G^{-1}$$

[†]In general, this constant can be used to normalize the group terms for specific uses. For our usage, as we saw in the electromagnetic example, this constant g will be the charge to allow for the relation between the Lie group and gauge transformation.

In general relativity, we expect our theory to be invariant under coordinate transformations, but the normal derivative breaks this invariance. To fix this, a term called the connection[†] is added to act to correct for the curvature of the manifold between points that the normal derivative does not account for. This is important for creating covariant, or invariant, derivative, i.e. a derivative that acts the same as a normal derivative in flat space. In our Lagrangian, after transforming our wavefunction, we find extra terms coming from the Lie group transformation. In the same way the curvature causes our coordinate system to change from point to point, our local gauge transformation induces error, so under the same prescription, we can create a gauge covariant derivative. For Yang-Mills theory, this is simply

$$\mathcal{D}_\mu = \partial_\mu - igA_\mu = \partial_\mu - igA^a_\mu T^a$$

The gauge transformation changes the local phase of the wavefunction, so our derivative similarly needs to compensate for these phase differences between points, necessitating a covariant derivative as in general relativity. This correction term is exactly the gauge force's interaction term which is why the different symmetries of the wavefunction give rise to the gauge forces. The forces themselves are the connections used to keep our theory invariant as expected. This means we can rewrite our Lagrangian in a form similar to the free particle solution because our covariant derivative accounts for the "curvature" caused by the interaction:

$$\mathcal{L} = \overline{\psi}(i\not{\mathcal{D}} - m)\psi - \frac{1}{4}|F|^2$$

Yang-Mills theory gives a generalized format for describing gauge invariant forces using an arbitrary Lie group structure. With it, we can easily formulate the general form of the force field as being spanned by the Lie algebra where each gauge bosons is associated with the axis of one of the basis vectors T_a of the algebra. From this, we can formulate the Lagrangian with the self-interaction term and interaction term which becomes part of the covariant derivative. From the Lagrangian, the Feynman rules can be read off for doing calculations. The Standard Model describes the electromagnetic, strong, weak forces and the underlying structure for the three forces are the $\mathrm{U}(1)$, $\mathrm{SU}(2)$, and $\mathrm{SU}(3)$ Lie groups respectively. This creates the full Standard Model structure which is described as $\mathrm{U}(1) \times \mathrm{SU}(2) \times \mathrm{SU}(3)$.

As a last note, because $\mathrm{U}(1)$ is Abelian, we have $f^{abc} = 0$ for all entries, so the field tensor, F, is trivially made of the derivative terms, $\partial^\mu A^\nu$. For the weak and strong force, the field tensor contains addition $gf^{abc}A^b_\mu A^c_\nu$ terms meaning in the Lagrangian allows for three and four vertex interactions. For the strong force, which is asymptotically free, it is exactly these three and four vertex terms that cause problems in loop calculations, meaning the strong force cannot be evaluated using perturbation theory when $\alpha_S(\mu) > 1$.

[†] In general relativity, this is formally the Christoffel symbols.

Appendix B

Cross-Section

In particle physics, and particularly in this thesis, the concept of the cross-section is used extensively to describe the production rate for different processes. While this interpretation of the cross-section is correct, the more common image is a 2D slice of an object. It is worth exploring deeper why we use the cross-section and how we come to the particle physic understanding of it.

The first, important note for understanding the cross-section is remembering that particle physics is a quantum theory and therefore statistical in nature. Any general quantum process is described by the wave function which is interpreted as a probability density for the particle once squared, or $|\psi(x)|^2 = \rho(x)$. In quantum field theory, through second quantization, our equations are modified, but the overall probabilistic nature of the theory remains. This means, to find a process's production rate, it is equivalent to find its production probability. From the standpoint of an experimentalist, it is more useful to couch this production rate within the context of a detector, so we consider the probability that a resulting particle will pass through a certain portion of our detector, not just the production probability. In classical physics, this is expressed with the cross-section, σ.

For a classical scattering experiment, there would a target and a beam, and any particles from the beam scattered by the target are collected in detectors located at specific angular positions. The cross-section represents the function of how the target will scatter incoming particles. From a physical standpoint, this description of the cross-section is identical to the geometric cross-section, or is the 2D collideable area seen by beam. For instance, for a billiard ball, the cross-section is just the profile area, πr^2, representing the possible area a particle could hit and scatter off of[†]. Because of the importance of knowing where the beam will be scattered, the total cross-section is distinguished from the differential cross-section, $\frac{d\sigma}{d\Omega}$, which represents the portion of the total cross-section that scatters the beam into a specific solid angle. For a force field, this definition becomes much more abstract because there is no physical scattering area, but the same computational methods can be applied to produce what ends up being an effective area for interacting with the field. This is shown more concretely with the toy example in Fig-

[†]Technically, the cross-section will depend on the beam particle's area as well as the target particle's area. This point isn't important for the discussion, but worth clarifying.

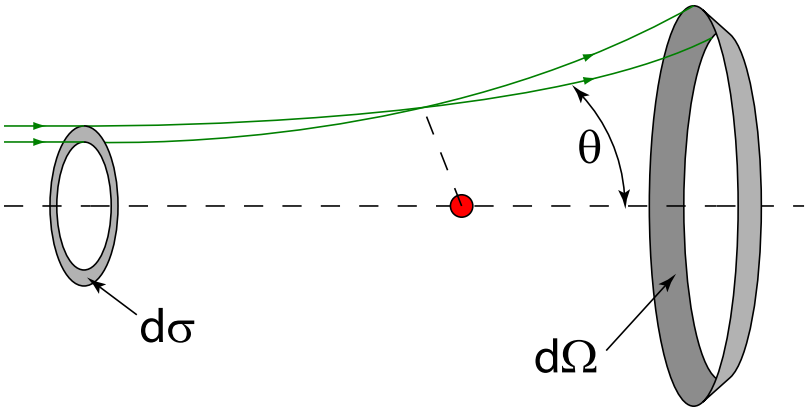

Figure B.1: Representation of the cross-section for a simple scattering of a beam against some force centered at the origin. The differential cross-section, $\frac{d\sigma}{d\Omega}$ is pictorially represented as the ratio of the initial differential area the beam travels through, $d\sigma$ divided by the differential angular area the beam is scatter into, $d\Omega$.

ure B.1, where the differential cross-section is represented by the total differential area that scatters the beam to the necessary solid angle.

From the differential cross-section, one can obviously reproduce the total cross-section by integrating over all angles. If we wanted to understand this more geometrically, we note the total cross-section corresponds to the area of the "target" that results in the particular scattering of interest. Applying this for the most basic case where we assume our force field is an electric field leads an infinite total cross-section. While at first concerning, this result is expected because the field's range is also infinite, so all particles will be scattered by the field, even if only by a minuscule amount making the area that scatters particles similarly infinite. This might prove problematic for our experiment, but in particle physics, we are mainly concerned with hard, inelastic processes that are short ranged and produce specific output particles, all of which having a finite total cross-section.

While the cross-section allows us to quantify how our target will scatter our beam, this is only half of the story. To model the observed rate, the scattering depends on the intensity of the beam as well. Intensity, or luminosity, is a measure of the total number of particles in the beam and the 2D density of particles in the beam (i.e. how tightly the beam is packed). To reproduce a scattering rate, it is simply a combination of the luminosity and cross-section, or the production rate, r, is :

$$r = \sigma \frac{d\mathscr{L}}{dt}$$

Given an instantaneous luminosity $\frac{d\mathscr{L}}{dt}$. Ignoring the luminosity, the cross-section serves as a proxy for the total production rate for a particular process in our detector.

It is convention to measure this cross-section in the strange unit of barns (b) that follows the relation

$1\,\mathrm{b} \equiv 10^{-24}\,\mathrm{cm}^{-2}$. Barns were first invented as a unit for measuring the cross-section of atomic nuclei to scattering from a particle beam, (the typical story being its first use was with the Manhattan Project). Barns use typical SI prefixes, and, for our experiment, are of the scale of pico and femtobarns, or pb and fb. For ease of calculation, our luminosity uses the even more unusual unit of inverse femtobarns, or fb^{-1}.

Appendix C

Detector Coordinates

CMS is a cylindrical detector, optimized for measuring particles in the transverse plane, and the particles of interest are relativistic and originating from a single vertex. Because of this, CMS uses a coordinate system unique to particle physics: (p_T, φ, η) or a modified cylindrical coordinate system.

In typical cylindrical coordinates, (ρ, φ, z), the coordinates are the radius from the z-axis, ρ, the azimuthal angle, φ, and the typical z-axis distance. While optimal for describing positions, the main concern of a particle physicist is describing 4-momenta, $(E, \vec{p})$. Most equations, such as for the cross-section, are parameterized in terms the incoming and outgoing momenta of particles or require calculations using 4-momentum vectors, so absolute positioning is unnecessary for most cases. This means our coordinate system needs to be modified to describe the vector of momentum originating from the primary interaction point.

First, the radius is substituted with the p_T or transverse momentum. The CMS's design is optimized to detect in the transverse plane, so the main variable measured is the transverse momentum. Because the p_T is so well measured and most particles will be around the transverse plane, it is also customary to refer to the p_T in lieu of the full momentum when describing a particle's kinematics. The p_T is not equal to the full momentum (considering it does not contain the p_z component), but the average p_T is close to full momentum because of the higher occupancy in the center of the detector.

For the azimuthal component, the angle stays the same as in the cylindrical system, but the z component is not described with p_z. Instead, another angular component, η, or the pseudorapidity is used. The reason stems from Lorentz invariance. For collisions in the LHC, the initial proton's z-momentum cancels, but for the resulting particles of interest, this is generally not the case. Because the initial interaction happens with the individual quarks and gluons in the protons, the inelastically produced particles can have imbalanced momenta as different proportions of the energy are imparted to the created particles *vs.* staying in the proton itself. To compare similar collision events effectively, one would apply a boost transformation along the z-axis to compensate for these differences. This leads to the utility of creating z-boost invariant quantities or a coordinate system more compatible with these quantities. With that,

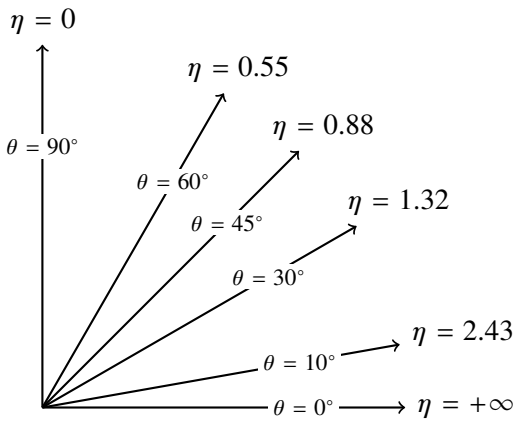

Figure C.1: Plot of values for the pseudorapidity, and the relationship between the polar angle θ and the pseudorapidity.

one comes to a definition of a coordinate called the rapidity, y:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right)$$

The rapidity allows for a simple z-boost transformations, or, given a boost velocity β:

$$y' = y - \tanh^{-1}(\beta)$$

And more importantly, the difference of rapidities is Lorentz invariant meaning two particle decays have a convenient Lorentz invariant quantity.

The pseudorapidity, η, is a slight modification of the rapidity, replacing the energy with the total momentum:

$$\eta = \frac{1}{2} \ln \left(\frac{p + p_z}{p - p_z} \right)$$

Highly relativistic particles have momenta much larger than their mass meaning the pseudorapidity approaches the rapidity as $m/E \to 0$. Since the pseudorapidity is derived solely from the momentum vector, it has the added benefit of being a purely geometric measurement and can be defined using a "polar" angle θ, as shown in Figure C.1, or mathematically:

$$\eta = -\ln \left(\tan \left(\frac{\theta}{2} \right) \right)$$

And this quantity is very easy to measure. The definition means that $\eta = 0$ corresponds to the transverse plane and $\eta = \pm\infty$ correspond to either ends of the beam axis. Inverting our equation, we can look at

the Taylor expansion of θ in terms of η around 0:

$$\theta = 2\tan^{-1}(e^{-\eta}) = \frac{\pi}{2} - \eta + \mathcal{O}(\eta^3)$$

Or, around 0, θ and η change at the same rate meaning η looks like an angle. This is used for angular definitions such as the distance parameter ΔR:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\varphi^2}$$

This distance parameter roughly represents the angular distance between two momentum vectors in the detector. For a given particle, this is illustrated by imagining a cone around the momentum vector with a certain radius r; vectors within the cone have $\Delta R < r$ and vectors outside of it have $\Delta R > r$. Further, because this variable is derived from η and φ, which are both Lorentz invariant under z-boosts, ΔR is also Lorentz invariant which makes it very useful for helping define different physics objects.

Appendix D

Run 2 Eras

Run 2 is composed of the data taken during 2016, 2017, and 2018, corresponding to roughly 137 fb^{-1} of data[†]. As understanding of the LHC grew, it was able to deliver more luminosity while similarly, CMS had less downtime leading to the noticeable difference in data taken between 2016 (36 fb^{-1}) versus 2018 (60 fb^{-1}). MC samples were produced each year with detector responses and simulation quantities set to match the overall data observed that year. As mentioned in Chapter 5, these presets are averages since detector response will change within each year over time to effects such as radiation damage or changes to input voltages for different subdetectors. Overall, the data/MC agreement has been observed to be quite good.

After the competition of Run 2, a campaign called UltraLegacy was made to converge the different calibrations and algorithms used between years into a newly reprocessed set of data and MC that would be compatible to be combined for a full Run 2 analysis. In looking at the UltraLegacy samples, one will notice that the data isn't split into the three years, but four distinct parts or eras caused by UltraLegacy bisecting the 2016 data into two, roughly equal parts. The primary reason for this comes from the issues stemming from change of the LHC colliding every 50 ns to 25 ns for Run 2.

At the start of Run 2, fewer tracks were observed than expected in the silicon detector at these higher rates, and the cause was found to come from highly ionizing particles. Most models for a charged particle's energy loss comes from the Bethe-Bloch formula, shown in Figure D.1, where notably, there is a point of minimum energy loss corresponding to the particle primarily ionizing. This minimum is broad, so particles in the valley of this distribution are expected to spend most of its lifetime in this energy range. This is what causes the typical Bragg's peak where energy dissipation is low until the sudden total energy deposit when the particle is stopped. For muons, the minimum valley is in the expected momentum range for particles at the LHC which empirically shows why muons so easily punch through the detector.

In contrast, you can have highly ionizing particles which instead collide inelastically causing a cascade of particles that further ionize the material they are in. This is analogous to the typical hadronic decays

[†]Run 2 technically started in 2015, but these events are typically not included because its low luminosity and problems associated with calibration after the long shutdown.

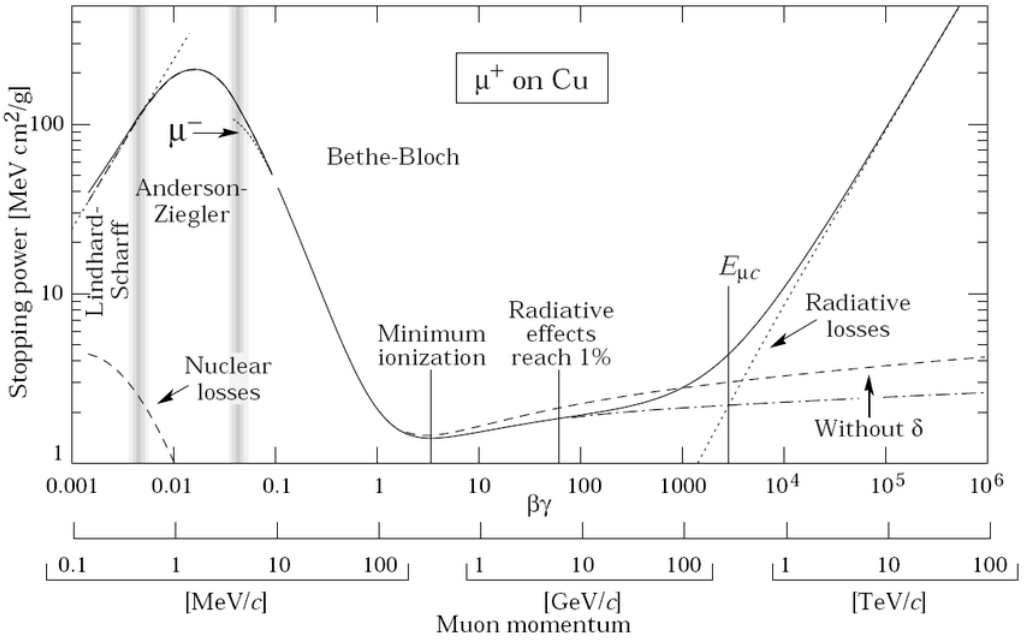

Figure D.1: Bethe-Bloch formula, specifically for a muon traveling through copper. While the minimum ionization point is around 1 GeV, the broad valley from 0.1 GeV to 100 GeV has only a roughly 10 MeV cm^{-1} stopping power when scaled by the density, or muons at the LHC generally are not stopped by the detector. While this graph is specifically for copper, the distribution is similar for other stopping materials such as steel or brass.

in the HCAL mentioned in Section 3.2.3. This sort of process is not expected to happen frequently in the tracker, but with higher luminosity, the rate of inelastic interactions increases. These highly interacting particles are not inherently a problem, but they create numerous ionizing particles localized in a portion of the tracker that can saturate the tracker readout and leads to general inefficiencies. Because of the time it took to understand this issue, the first half of 2016's data taking was plagued with these inefficiencies. Once found, the readout chip had its pre-amplifier feedback voltage changed, so the charge on the module could be reabsorbed quicker thus reducing the saturation that caused the tracker issues [89].

For modeling the data in simulation, this poses a problem because the particle reconstruction would be fundamentally different before and after this fix. Certain parts of the detector may have downtime, which is handled by removing these events, similar to the MET filters described in Section 6.3, or with correction factors as with the prefire issue described in Section 9.2.2. Because this issue was larger and splits the data in half, the solution was to simply simulate the data for these two conditions and treat them as separate "years." This is what lead to the creation of the four eras of data with 2016 split into 2016pre and 2016post. For some cases, 2016pre and 2016post are combined to make up for smaller data in each split year (only 20 fb^{-1} and 17 fb^{-1} respectively), but for all the statistical tests, the years are split into the four eras since the response in the model is expected to be different between the two 2016 data

taking eras.

Appendix E

Transfer Factor

For applying the transfer factors for the nonprompt and charge-misId data-driven backgrounds, there are several nuances to consider, primarily in the way the transfer factors are calculated. While the method shown in Section 7.1 is generally valid, complications come from expanding the method for multiple leptons in an event. The transfer factor is calculated slightly differently for each event for both data-driven methods in this analysis, so the details for both are listed in the following sections.

E.1 Nonprompt Transfer Factors

Our signal region involves multiple leptons, so we need to consider double counting in our application. This is a concern because the naive extensions to the transfer factor calculation leads to a combinatorial excesses or deficits.

The tight (T) and fake (F) categories for the leptons[†], while optimized to classify the leptons based on promptness, do not reflect the truth value of the leptons, i.e. prompt (p) or nonprompt (n). The tight definition is defined specifically to reduce the number of nonprompt leptons that are categorized as tight while also keeping most prompt leptons as tight, or $P(T|p) \to 1$ and thus $P(F|p) \to 0$. The basis of the data-driven method is that our identification isn't perfect, so there will be a nontrivial misidentification rate $P(T|n)$. Using these probabilities, we write our transfer factor as:

$$w = \frac{P(T|n)}{P(F|n)}$$

This holds because we measure our misidentification rate (and thus transfer factor) in a pure nonprompt region. This is why the prompt subtraction is so important. In our signal region, there will be prompt contamination, so we must reexpress N_T in terms of N_F using the true prompt and nonprompt yields, N_p and N_n, probabilities, and transfer factor to find how to properly estimate our signal region. With

[†]Here, fake (F) represents leptons that *only* pass the fake requirement. While tight leptons are a subset of fake leptons, for calculating transfer factors, the T and F regions are considered orthogonal, thus fakes in this context means exclusively fake leptons.

this in mind, we start with the following relation:

$$N_T = N_n P(T|n)$$
$$N_F = N_n P(F|n) + N_p P(F|p)$$

Here, we must remember that N_T is only concerned with the number of nonprompt events that pass our tight requirement, so N_p or number of prompt/signal events enters into the relation only when a real prompt lepton is misidentified. In fact, the $N_p P(T|p)$ term is effectively our signal. This means the $N_p P(F|p)$ term represents the prompt contamination that is incorrectly categorized. While not necessary for the proceeding math, we can ignore this prompt term because the amount of nonprompt events is much greater than the number of prompt events, or $N_p/N_n \sim \epsilon$ and prompt leptons are defined so $P(T|p) \to 1$, and by proxy, $P(F|p) \sim \epsilon$. Putting these together, we find:

$$\begin{aligned} N_F &= N_n P(F|n) + N_n \frac{N_p}{N_n} P(F|p) \\ &= N_n \left(P(F|n) + \mathcal{O}(\epsilon^2) \right) \end{aligned}$$

Or the prompt contamination to our misidentification rates is assumed to be at most a second order correct, so these prompt terms can safely be ignored for this, and further calculation[†]. Connecting the nonprompt/fake relation with our transfer factor, we find:

$$\begin{aligned} w \times N_F &= \frac{P(T|n)}{P(F|n)} \times N_n P(F|n) \\ &= N_n P(T|n) = N_T \end{aligned}$$

As expected.

For a two lepton case, this same process become more complicated. There are three possible configurations of our nonprompt leptons based on their reconstruction category: tight-tight (TT), tight-fake (TF), and fake-fake (FF), where the TT region is our signal region, and N_{TT} is the nonprompt yield we are trying to estimate. Normally, we would use the same method as with the single lepton case, applying our transfer factor to events in the TF region. While this transfers one fake lepton, both our TT and TF regions contain events that have two nonprompt leptons. For the TF region, this is a problem because the two nonprompt leptons are in this region because one passed the tight requirement while the other only passed the fake requirement. That means there are two choices for which lepton could be tight, so these events gain a factor of two. For the TT region, the double nonprompt events are in this region because both leptons were identified as tight, so there is no ambiguity in which nonprompt

[†]Again, it should be noted that this assumption is not necessary, the result is exactly the same with the addition prompt terms, but it does greatly reduce the mathematical complexity and avoids unnecessary complications.

passes the tight requirement as in the TF region. Thus, our naive method double counts the events with two nonprompt leptons.

To fix this, we need to subtract out the extra events with two nonprompt leptons from our weighted TF events, and, thankfully, this can be estimated by the number of events in the FF region. Put more mathematically, we can reproduce relationships between the number of tight/fake lepton events and the number of prompt/nonprompt lepton events, as done with the single lepton case. We can also use the same simplifying assumption of $N_p/N_n P(F|p) \sim \epsilon^2 \to 0$ for ease, and thus we can redefine $P(T) \equiv P(T|n)$, $P(F) \equiv P(F|n)$, and $P(T|p) = 1$:

$$\begin{aligned} N_{TT} &= N_{np} P(T) + N_{nn} P(T)^2 \\ N_{TF} &= N_{np} P(F) + 2N_{nn} P(T)P(F) \\ N_{FF} &= N_{nn} P(F)^2 \end{aligned}$$

Simply applying w to N_{TF} as done before, we notice a discrepancy of $N_{nn} P(T)^2$, or an excess of double nonprompt events. To reproduce a relation similar to the single lepton case, we need to subtract the double counted N_{nn} events which can be done using the FF region. This leads to this equality:

$$N_{TT} = w \times N_{TF} - w^2 \times N_{FF}$$

This logic can be extended to the three lepton case, and an alternating pattern for the weights can be observed for the different transfer regions. Similar in the above example, to simulate the TTT region, using only TTF region will lead to overcounting of event with two nonprompt leptons, but subtracting events from the TFF region leads to an undercounting of three nonprompt events, so we would need to add back the transferred FFF region to remedy this.

The repeated application of combinatorial corrections leads to alternating signs for each subsequent correction term, or, generally, an extra negative sign is needed for every even number of fake leptons in an individual event:

$$w = (-1)^{N_{\text{fake}}+1} \prod_{i \in \text{fakes}} \frac{f(p_{\text{T}}^i, \eta^i)}{1 - f(p_{\text{T}}^i, \eta^i)}$$

This becomes the final transfer factor used for the nonprompt background estimation

E.2 Charge-MisId Transfer Factor

As mentioned in the overview in Section 7.1, for transferring event, we use a transfer factor of the form $w = f/(1-f)$. For the charge-misId background, this approach needs a slight modification due to combinatorial differences. The transfer factor, in some sense, scales the weight of the event to account for one of the electrons flipping its charge where our signal region transfers electrons from the opposite sign

region into the same sign region. The problem comes from the charges of electrons being interchangeable, i.e. we do not discriminate between $(+,+)$ and $(-,-)$ events. For an electron-muon event, there is no problem, we can construct and use the typical transfer factor from the misidentification rate of the electron which is binned in p_{T} and η. For the electron-electron case, there are two ways the opposite sign events could be measured as a same sign event: either the first or the second electron's charge is flipped. If we were to flip the charge of both electrons, we would have an opposite sign event again, so our charge-misId background come strictly from a single electron's charge being flipped. This necessitates that our electron's transfer factor must be added instead of multiplied together to account for the indistinguishability. Our formula for our transfer factor in the electron-electron region is:

$$w = \frac{f_1}{1 - f_1} + \frac{f_2}{1 - f_2}$$

Bibliography

[1] J. J. Thomson, "Xl. cathode rays", *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **44** (October, 1897) 293–316, doi:10.1080/14786449708621070.

[2] E. Rutherford, "Lxxix. the scattering of alpha and beta particles by matter and the structure of the atom", *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **21** (May, 1911) 669–688, doi:10.1080/14786440508637080.

[3] UA1, "Experimental observation of isolated large transverse energy electrons with associated missing energy at", *Physics Letters B* **122** (February, 1983) 103–116, doi:10.1016/0370-2693(83)91177-2.

[4] UA2, "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the cern p collider", *Physics Letters B* **122** (March, 1983) 476–485, doi:10.1016/0370-2693(83)91605-2.

[5] UA1, "Experimental observation of lepton pairs of invariant mass around 95 gev/c2 at the cern sps collider", *Physics Letters B* **126** (July, 1983) 398–410, doi:10.1016/0370-2693(83)90188-0.

[6] UA2, "Evidence for z0->e+e- at the cern p collider", *Physics Letters B* **129** (September, 1983) 130–140, doi:10.1016/0370-2693(83)90744-x.

[7] CMS, "Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc", *Physics Letters B* **716** (September, 2012) 30–61, doi:10.1016/j.physletb.2012.08.021.

[8] ATLAS, "Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc", *Physics Letters B* **716** (September, 2012) 1–29, doi:10.1016/j.physletb.2012.08.020.

[9] T. Aaltonen et al., "High-precision measurement of the w boson mass with the cdf ii detector", *Science* **376** (April, 2022) 170–176, doi:10.1126/science.abk1781.

[10] CMS Collaboration, "High-precision measurement of the W boson mass with the CMS experiment at the LHC", technical report, CERN, Geneva, 2024. arXiv:2412.13872,

Submitted to Nature. All figures and tables can be found at http://cms-results.web.cern.ch/cms-results/public-results/publications/SMP-23-002 (CMS Public Pages).

[11] E. Noether, "Invariante variationsprobleme", *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* **1918** (1918) 235–257.

[12] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", *Physical Review Letters* **13** (October, 1964) 508–509, doi:10.1103/physrevlett.13.508.

[13] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323, doi:10.1103/PhysRevLett.13.321.

[14] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global conservation laws and massless particles", *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587, doi:10.1103/PhysRevLett.13.585.

[15] Particle Data Group Collaboration, "Review of particle physics", *Phys. Rev. D* **110** (2024), no. 3, 030001, doi:10.1103/PhysRevD.110.030001.

[16] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, "Evidence for the 2π decay of the k_2^0 meson", *Phys. Rev. Lett.* **13** (Jul, 1964) 138–140, doi:10.1103/PhysRevLett.13.138.

[17] M. Kobayashi and T. Maskawa, "Cp-violation in the renormalizable theory of weak interaction", *Progress of Theoretical Physics* **49** (February, 1973) 652–657, doi:10.1143/ptp.49.652.

[18] N. Cabibbo, "Unitary symmetry and leptonic decays", *Physical Review Letters* **10** (June, 1963) 531–533, doi:10.1103/physrevlett.10.531.

[19] J. J. Aubert et al., "Experimental observation of a heavy particle *j*", *Phys. Rev. Lett.* **33** (Dec, 1974) 1404–1406, doi:10.1103/PhysRevLett.33.1404.

[20] J. E. Augustin et al., "Discovery of a narrow resonance in e^+e^- annihilation", *Phys. Rev. Lett.* **33** (Dec, 1974) 1406–1408, doi:10.1103/PhysRevLett.33.1406.

[21] S. W. Herb et al., "Observation of a dimuon resonance at 9.5 gev in 400-gev proton-nucleus collisions", *Physical Review Letters* **39** (August, 1977) 252–255, doi:10.1103/physrevlett.39.252.

[22] G. Arnison et al., "Associated production of an isolated, large-transverse-momentum lepton (electron or muon), and two jets at the cern p collider", *Physics Letters B* **147** (November, 1984) 493–508, doi:10.1016/0370-2693(84)91410-2.

[23] CDF Collaboration Collaboration, "Observation of top quark production in $\overline{p}p$ collisions with the collider detector at fermilab", *Phys. Rev. Lett.* **74** (Apr, 1995) 2626–2631, doi:10.1103/PhysRevLett.74.2626.

[24] D0 Collaboration Collaboration, "Observation of the top quark", *Phys. Rev. Lett.* **74** (Apr, 1995) 2632–2637, doi:10.1103/PhysRevLett.74.2632.

[25] S. Mandelstam, "Determination of the pion-nucleon scattering amplitude from dispersion relations and unitarity. general theory", *Phys. Rev.* **112** (Nov, 1958) 1344–1360, doi:10.1103/PhysRev.112.1344.

[26] X. Fan, T. G. Myers, B. A. D. Sukra, and G. Gabrielse, "Measurement of the electron magnetic moment", *Phys. Rev. Lett.* **130** (Feb, 2023) 071801, doi:10.1103/PhysRevLett.130.071801.

[27] D. J. Gross and F. Wilczek, "Ultraviolet behavior of non-abelian gauge theories", *Phys. Rev. Lett.* **30** (Jun, 1973) 1343–1346, doi:10.1103/PhysRevLett.30.1343.

[28] H. D. Politzer, "Reliable perturbative results for strong interactions?", *Phys. Rev. Lett.* **30** (Jun, 1973) 1346–1349, doi:10.1103/PhysRevLett.30.1346.

[29] R. D. Ball et al., "Parton distributions from high-precision collider data: Nnpdf collaboration", *The European Physical Journal C* 77 (October, 2017) doi:10.1140/epjc/s10052-017-5199-5.

[30] E. D. Bloom et al., "High-energy inelastic $e-p$ scattering at 6° and 10°", *Phys. Rev. Lett.* **23** (Oct, 1969) 930–934, doi:10.1103/PhysRevLett.23.930.

[31] CDF Collaboration, "Observation of electroweak single top-quark production", *Phys. Rev. Lett.* **103** (Aug, 2009) 092002, doi:10.1103/PhysRevLett.103.092002.

[32] D0 Collaboration, "Observation of single top-quark production", *Phys. Rev. Lett.* **103** (Aug, 2009) 092001, doi:10.1103/PhysRevLett.103.092001.

[33] ATLAS, "Observation of four-top-quark production in the multilepton final state with the atlas detector", *The European Physical Journal C* **83** (June, 2023) doi:10.1140/epjc/s10052-023-11573-0.

[34] CMS, "Observation of four top quark production in proton-proton collisions at s=13tev", *Physics Letters B* **847** (2023) 138290, doi:https://doi.org/10.1016/j.physletb.2023.138290.

[35] R. L. Workman et al., "Review of particle physics", *Progress of Theoretical and Experimental Physics* **2022** (August, 2022) doi:10.1093/ptep/ptac097.

[36] F. Zwicky, "On the masses of nebulae and of clusters of nebulae", *The Astrophysical Journal* **86** (October, 1937) 217, doi:10.1086/143864.

[37] J. A. Tyson, R. A. Wenk, and F. Valdes, "Detection of systematic gravitational lens galaxy image alignments - mapping dark matter in galaxy clusters", *The Astrophysical Journal* **349** (January, 1990) L1, doi:10.1086/185636.

[38] M. Markevitch et al., "Direct constraints on the dark matter self-interaction cross section from the merging galaxy cluster 1e 0657-56", *The Astrophysical Journal* **606** (May, 2004) 819–824, doi:10.1086/383178.

[39] G. F. Giudice, "Naturally Speaking: The Naturalness Criterion and Physics at the LHC", p. 155–178. WORLD SCIENTIFIC, January, 2008. doi:10.1142/9789812779762_0010.

[40] S. L. Glashow, J. Iliopoulos, and L. Maiani, "Weak interactions with lepton-hadron symmetry", *Phys. Rev. D* **2** (Oct, 1970) 1285–1292, doi:10.1103/PhysRevD.2.1285.

[41] H. Miyazawa, "Baryon number changing currents", *Progress of Theoretical Physics* **36** (December, 1966) 1266–1276, doi:10.1143/ptp.36.1266.

[42] P. Langacker, "The physics of heavy Z' gauge bosons", *Rev. Mod. Phys.* **81** (Aug, 2009) 1199–1228, doi:10.1103/RevModPhys.81.1199.

[43] V. Barger, W.-Y. Keung, and B. Yencho, "Triple-top signal of new physics at the LHC", *Physics Letters B* **687** (apr, 2010) 70–74, doi:10.1016/j.physletb.2010.03.001.

[44] Q.-H. Cao, S.-L. Chen, Y. Liu, and X.-P. Wang, "What can we learn from triple top-quark production?", *Physical Review D* **100** (sep, 2019) doi:10.1103/physrevd.100.055035.

[45] H. Khanpour, "Probing top quark FCNC couplings in the triple-top signal at the high energy LHC and future circular collider", *Nuclear Physics B* **958** (sep, 2020) 115141, doi:10.1016/j.nuclphysb.2020.115141.

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

[47] A. Tumasyan et al., "A portrait of the higgs boson by the cms experiment ten years after the discovery", *Nature* **607** (July, 2022) 60–68, doi:10.1038/s41586-022-04892-x.

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

[49] CMS CollaborationV. Karimäki, et al., "The CMS tracker system project: Technical Design Report". Technical design report. CMS. CERN, Geneva, 1997.

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

[51] L. Rossi, "The LHC Superconducting Magnets",.

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

[53] CMS Collaboration, J. G. Layter, "The CMS muon project: Technical Design Report". Technical design report. CMS. CERN, Geneva, 1997.

[54] CMS Collaboration, "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at cms", *Eur. Phys. J. C* **81** (2021), no. 800, doi:10.1140/epjc/s10052-021-09538-2.

[55] CMS Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV", technical report, CERN, Geneva, 2018.

[56] CMS Collaboration, "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV", technical report, CERN, Geneva, 2019.

[57] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014) P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569. Comments: Replaced with published version. Added journal reference and DOI.

[58] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector. Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017), no. 10, P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965. Replaced with the published version. Added the journal reference and DOI. All the figures and tables can be found at http://cms-results.web.cern.ch/cms-results/public-results/publications/PRF-14-001 (CMS Public Pages).

[59] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-ktjet clustering algorithm", *Journal of High Energy Physics* **2008** (April, 2008) 063–063, doi:10.1088/1126-6708/2008/04/063.

[60] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *Journal of High Energy Physics* **2014** (July, 2014) doi:10.1007/jhep07(2014)079.

[61] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *Journal of High Energy Physics* **2014** (jul, 2014) doi:10.1007/jhep07(2014)079.

[62] S. Frixione and B. R. Webber, "Matching nlo qcd computations and parton shower simulations", *Journal of High Energy Physics* **2002** (June, 2002) 029–029, doi:10.1088/1126-6708/2002/06/029.

[63] C. Oleari, "The powheg box", *Nuclear Physics B - Proceedings Supplements* **205–206** (August, 2010) 36–41, doi:10.1016/j.nuclphysbps.2010.08.016.

[64] C. Bierlich et al., "A comprehensive guide to the physics and usage of pythia 8.3", 2022. https://arxiv.org/abs/2203.11601.

[65] CMS Collaboration, "Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements", *The European Physical Journal C* **80** (jan, 2020) doi:10.1140/epjc/s10052-019-7499-4.

[66] B. Andersson, G. Gustafson, G. Ingelman, and T. SjÃ¶strand, "Parton fragmentation and string dynamics", *Physics Reports* **97** (1983), no. 2, 31–145, doi:https://doi.org/10.1016/0370-1573(83)90080-7.

[67] S. Agostinelli et al., "Geant4—a simulation toolkit", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (July, 2003) 250–303, doi:10.1016/s0168-9002(03)01368-8.

[68] J. Allison et al., "Geant4 developments and applications", *IEEE Transactions on Nuclear Science* **53** (2006), no. 1, 270–278, doi:10.1109/TNS.2006.869826.

[69] J. Allison et al., "Recent developments in geant4", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **835** (November, 2016) 186–225, doi:10.1016/j.nima.2016.06.125.

[70] K. Rehermann and B. Tweedie, "Efficient identification of boosted semileptonic top quarks at the lhc", *Journal of High Energy Physics* **2011** (March, 2011) doi:10.1007/jhep03(2011)059.

[71] CMS Collaboration, "Search for ttH production in multilepton final states at sqrt(s) = 13 TeV",.

[72] CMS Collaboration, "Muon identification using multivariate techniques in the cms experiment in proton-proton collisions at sqrt(s)=13tev", *Journal of Instrumentation* **19** (feb, 2024) P02031, doi:10.1088/1748-0221/19/02/P02031.

[73] Y. Coadou, "Boosted Decision Trees", p. 9–58. WORLD SCIENTIFIC, February, 2022. doi:10.1142/9789811234033_0002.

[74] CMS Collaboration, "Pileup mitigation at cms in 13 tev data", *Journal of Instrumentation* **15** (sep, 2020) P09018, doi:10.1088/1748-0221/15/09/P09018.

[75] E. Bols et al., "Jet flavour classification using deepjet", *Journal of Instrumentation* **15** (dec, 2020) P12012, doi:10.1088/1748-0221/15/12/P12012.

[76] P. P. Ray, "Chatgpt: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope", *Internet of Things and Cyber-Physical Systems* **3** (2023) 121–154, doi:10.1016/j.iotcps.2023.04.003.

[77] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system", in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, p. 785–794. ACM, August, 2016. doi:10.1145/2939672.2939785.

[78] C. G. Lester and A. J. Barr, "mtgen: mass scale measurements in pair-production at colliders", *Journal of High Energy Physics* **2007** (December, 2007) 102–102, doi:10.1088/1126-6708/2007/12/102.

[79] M. Tec, "Roc analysis for classification and prediction in practice", *Journal of the American Statistical Association* **120** (January, 2025) 585–586, doi:10.1080/01621459.2024.2423434.

[80] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses", *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **231** (February, 1933) 289–337, doi:10.1098/rsta.1933.0009.

[81] A. L. Read, "Presentation of search results: theclstechnique", *Journal of Physics G: Nuclear and Particle Physics* **28** (September, 2002) 2693–2704, doi:10.1088/0954-3899/28/10/313.

[82] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *The European Physical Journal C* **71** (February, 2011) doi:10.1140/epjc/s10052-011-1554-0.

[83] J. S. Conway, "Incorporating nuisance parameters in likelihoods for multisource spectra", 2011. https://arxiv.org/abs/1103.0354.

[84] A. Hayrapetyan et al., "The cms statistical analysis and combination tool: Combine", *Computing and Software for Big Science* **8** (November, 2024) doi:10.1007/s41781-024-00121-4.

[85] R. Barlow, "Event classification using weighting methods", *Journal of Computational Physics* **72** (1987), no. 1, 202–219, doi:https://doi.org/10.1016/0021-9991(87)90078-7.

[86] A. Buckley et al., "Lhapdf6: parton density access in the lhc precision era", *The European Physical Journal C* **75** (March, 2015) doi:10.1140/epjc/s10052-015-3318-8.

[87] W. S. Cleveland, "Lowess: A program for smoothing scatterplots by robust locally weighted regression", *The American Statistician* **35** (February, 1981) 54, doi:10.2307/2683591.

[88] C. N. Yang and R. L. Mills, "Conservation of isotopic spin and isotopic gauge invariance", *Phys. Rev.* **96** (Oct, 1954) 191–195, doi:10.1103/PhysRev.96.191.

[89] M. Jansova, "Search for the supersymmetric partner of the top quark and measurements of cluster properties in the silicon strip tracker of the CMS experiment at Run 2". PhD thesis, University of Strasbourg, 2018. Presented 27 Sep 2018.