CERN-THESIS-2015-414

A SEARCH FOR THE $T\bar{T}H\left(H \to B\bar{B}\right)$ CHANNEL AT THE LARGE HADRON COLLIDER WITH THE ATLAS DETECTOR USING A MATRIX ELEMENT METHOD

BY

AUSTIN T. BASYE

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Professor A. X. El-Khadra, Chair
Professor T. Liss, Director of Research
Professor J. N. Eckstein
Assistant Professor V. I. Martinez Outschoorn

Abstract

A matrix element method analysis of the Standard Model Higgs boson, produced in association with two top quarks decaying to the lepton-plus-jets channel is presented. Based on 20.3 fb^{-1} of $\sqrt{s}$=8 TeV data, produced at the Large Hadron Collider and collected by the ATLAS detector, this analysis utilizes multiple advanced techniques to search for $t\bar{t}H$ signatures with a 125 GeV Higgs boson decaying to two b-quarks. After categorizing selected events based on their jet and b-tag multiplicities, signal rich regions are analyzed using the matrix element method. Resulting variables are then propagated to two parallel multivariate analyses utilizing Neural Networks and Boosted Decision Trees respectively. As no significant excess is found, an observed (expected) limit of 3.4 (2.2) times the Standard Model cross-section is determined at 95% confidence, using the CLs method, for the Neural Network analysis. For the Boosted Decision Tree analysis, an observed (expected) limit of 5.2 (2.7) times the Standard Model cross-section is determined at 95% confidence, using the CLs method. Corresponding unconstrained fits of the Higgs boson signal strength to the observed data result in the measured signal cross-section to Standard Model cross-section prediction of $\mu = 1.2 \pm 1.3(\text{total}) \pm 0.7(\text{stat.})$ for the Neural Network analysis, and $\mu = 2.9 \pm 1.4(\text{total}) \pm 0.8(\text{stat.})$ for the Boosted Decision Tree analysis.

To my wonderful wife, Erin,
without whom none of this would have been possible

Acknowledgments

Standing on the precipice, but a few short paces from my goal, I can't help but think of all the individuals who have made this possible. I truly would not have made it this far without the timely and selfless investments from countless friends and family, mentors and colleagues, and a providential good Samaritan here and there. As such, I would like to take a moment to explicitly recognize but a fraction of these individuals and the contributions they made to this document. Understanding that any such list of finite length is subject to omission, this section is by no means complete.

Professional Acknowledgements

As this work was supported, in part, by the United States Department of Energy, under the grant numbers DE-SC0009932 and DEFG02-91ER40677, I would first like to thank the US DOE and, by obvious extension, the US Taxpayer for their support of fundamental scientific research. In a reality of finite resources, decisions must be made regarding the extent to which we collectively pursue speculative lines of inquiry regarding our universe and our place in it, irrespective of its economic application. As such I hope it is understood that I have taken this social investment very seriously, having sought to provide a substantive return in the form of this document.

I would like to recognize and thank the members of my Doctoral Committee who have contributed their valuable time and expertise on behalf of this endeavor. I thank Professor El-Khadra for her excellent classes and notes which greatly eased the writing of Chapter 2. I would like to thank Professor Eckstein for his enlightening pedagogical approach to Jackson's *Classical Electrodynamics* in addition to his refreshing curiosity during my preliminary exam. I thank Professor Martinez Outschoorn for our discussions at CERN regarding my work and for her genuine desire to see her students succeed. I would also like to thank her for her willingness to serve on my committee when Professor Grosse Perdekamp could not. Finally, I would like to extend my sincere thanks to Professor Grosse Perdekamp. His contribution to this document cannot be overstated. I arrived at UIUC principally due to the standard Dr. Grosse Perdekamp set as a member of the UIUC Physics Faculty. Furthermore, he frequently served as a much needed confidant during the inevitable trials of graduate school, providing wise

council for which I will be eternally grateful.

This dissertation would have never come to fruition were it not for the wise council, flexibility, and general adventurousness of my advisor, Professor Tony Liss. His confidence, or at least his ability to disguise his lack thereof, in my work allowed me to take the risks that ultimately payed off in the form of this document. His ostensibly 'hands off' approach allowed me to creatively explore and follow the trails I thought were interesting, while his occasional correction brought me back to the big picture when necessary. I simply could not have chosen a better mentor during this process.

I would like to give special recognition to the members of the ATLAS HSG8 $\left(H \to b\bar{b}\right)$ working group. Many plots in this document were generated using the tools collectively developed and maintained by its members. Furthermore, tremendous effort was expended to accommodate my work on the Matrix Element Method, ultimately resulting in its publication. Elements of this document which rely on the work of these individuals are denoted with the reference to the public paper [1] or its internal supporting document [2]. Finally, I would like to explicitly mention the huge contribution of Olaf Nackenhorst. Our mutual collaboration on the Matrix Element Method made an ostensibly impossible approach not only possible, but profitable. His exceptional physics and software acumen were instrumental in the MEM's development and ultimate success, following nearly 3 years and 3 million CPU-hours of intense analysis.

For their timely discussions, suggestions, and interest in my work, I would like to thank the members of the UIUC High Energy Physics group, Professors Neubauer, Errede, Hooberman, and Martinez Outschoorn. I have enjoyed seeing the evolution of the department over the last 5 years and based on its current trajectory, I expect UIUC to play a increasingly important role at ATLAS.

For their camaraderie in the Subatomic and QFT foxholes and their superb sparing during our preparation for the Qualifying Exam, I would like to thank my fellow *brothers-in-arms*: Daniel Jumper, Bryan Dannowitz, Noah Schroeder, Josiah Walton, and R. Evan McClellan. As iron sharpens iron, so one man sharpens another.

I would also like to thank those who played a role in my gravitation toward physics in general. Mrs. Elaine Gwinn was an exceptional physics teacher at Canyon High School, representative of the teaching staff at CHS in general. In anticipation of taking her class the following year, I surreptitiously obtained copies of the current year's project assignments, allowing me a year's head-start in my preparation. This advantage, along with the help of the rest of *Team Overkill* (Chris Roberts, Matt Farley, Jeremy Glover, and Tony Ithaca) allowed us to creatively design and engineer some exciting and borderline-dangerous contraptions, introducing me to the excitement of the unknown.

Finally, taking this excitement with me to Abilene Christian University, Professors Rusty Towell and Donald Isenhower directed it into the field of high energy physics. Following my first summer at Brookhaven National Laboratory,

I knew that physics was the best place for me to explore the unknown. The toys were bigger, the questions were harder, and the solutions were necessarily 'unconventional,' to put it mildly. There was truly no better place to encourage creativity and a passion for exploration, and to develop the technical skills and experience to creatively follow that passion.

Personal Acknowledgements

Without the social support from countless friends and family, I would have surely retreated to some other line of work. Physics is indeed a cruel mistress. As such, I would like to thank those groups and individuals that encouraged my work from the very beginning until now. In doing so, I wish for them to take some amount of ownership in this achievement, for whatever it is worth, as it required a truly collective effort.

I would first like to thank the membership at Philo Road Church of Christ in Urbana-Champaign, Illinois. Their encouragement during our first few years of graduate school was truly helpful during a very challenging time in our lives. Erin and I are indebted to the things we learned at Philo Road which have allowed us to grow together in our marriage.

Our relocation to Geneva, Switzerland was tremulous, to put it mildly, and would have ended before it started without the family at Crossroads Church, in Ferney-Voltaire, France. Erin and I are eternally grateful for their support and encouragement. The family at Crossroads, at various times, provided food, furniture, and lifelong friendships, not to mention the occasional ethnic adventures. I don't profess to know the explicit nature of what lies beyond this life, but I expect that the community at Crossroads has many aspects in common with it.

No list would be complete without the recognizing the substantial financial and emotional support provided by our parents, Tom and Cherise Basye and Glenn and Marcia Neimeyer. Few are blessed with two excellent parents, fewer still are blessed with four. Being one of the privileged few, I would like to thank them all for everything they have done for us, things which are to numerous to recount.

To my beloved wife Erin, I would like to extend my heartfelt thanks and acknowledgement. She, far more than anyone else, made the necessary sacrifices to see this work come to fruition. As such, I hope that she takes the requisite amount equity in this accomplishment, for it is not insubstantial.

Finally, to Him who is able to do immeasurably more than all we ask or imagine, according to His power that is at work within us, to Him be glory in the church and in Christ Jesus throughout all generations, for ever and ever! May this document stand as a testament to that which is immeasurably done.

Contents

Chapter 1

Introduction

The inexpressible drive to expand the boundaries of what is known is a very primitive feeling. Presumably this trait has been selected and enhanced over generations as it has undoubtably bestowed an advantage in life over those who are merely content with the status quo. What's more, as our cumulative understanding improves, so to does the rate at which we accumulate understanding. The 20th century was a whirlwind of advancement, innovation, and change, in large part thanks to the fundamental discoveries regarding our universe. Unfortunately, we need not look too far to find circumstances in which our command of the natural world has ended in indescribable suffering. Yet, despite these horrific turns, our curiosity compels us to press on into the unknown. This context of exploration and discovery for the principal good of humanity has provided the required motivation for the development of this dissertation. As such, this dissertation seeks to play some small part in the planting of the metaphorical trees under which our collective progeny will find shade, comfort, and hopefully peace, irrespective of this dissertation's immediate utility.

From the moment the earliest humans were capable of recognizing self, questions were asked regarding our ultimate place in this universe. Naturally, this question spawned other fundamental questions concerning, the nature of our universe, the way it evolves with time, and the hierarchy of complexity that appears even in our tiny corner of the cosmos. These questions have motivated a tremendous amount of intellectual effort on behalf of generations of our predecessors, brining us to where we are today. In their current form, the core theories consist of the *Standard Model* and the theory of *General Relativity*, two strikingly beautiful yet mutually dissonant theories, which collectively answer our most fundamental questions, to the best of their ability. Huge swaths of human experience are encoded in these theories, and yet, like a photo of Van Gogh's *Haystacks* and *Wheatfields* in an art appreciation textbook, they simply don't quite capture the essence of the masterpiece; but not for lack of trying.

As testaments to humanity's curiosity, tools have been developed to explore the reaches of our universe. One such tool exists 500 feet under the mountains of Geneva, Switzerland. The *Large Hadron Collider*, a 10 billion dollar microscope, serves as our setting. The professional environment of thousands of scientists and engineers from many disparate nations, creeds, and peoples provides the

many characters necessary for the success of this endeavor. Finally, the siren call of seemingly insurmountable challenges provides the subtext.

This dissertation seeks to contribute to the fund of human understanding by developing and presenting novel techniques for the measurement of the Standard Model parameter which describes the coupling between the top quark and the Higgs boson. Utilizing data accumulated in 2012 by the ATLAS detector at the Large Hadron Collider, this dissertation will present analysis methods which focus on the efficient identification of relevant experimental signatures which will aid in the accurate measurement of the lepton-plus-jets channel of the Standard Model $t\bar{t}H\left(H \to b\bar{b}\right)$ process. Finally, this dissertation seeks to provide sufficient detail to aid in the further evolution of its core methods for future applications at the LHC and other experimental venues. It is organized as follows:

- Chapter 2 provides the necessary theoretical context for this analytical endeavor, in addition to its motivating influence.
- Chapter 3 presents a detailed overview of the Large Hadron Collider and the ATLAS detector in addition to other indispensable tools required for this analysis.
- Chapter 4 presents a generic analysis method for identifying a signal process amidst a composite SM background before applying the method to the $t\bar{t}H\left(H \to b\bar{b}\right)$ signal.
- Chapter 5 introduces the general techniques involved with simulating the expected signal and background events in addition to presenting the explicit methods used to obtain this analysis' final result.
- Chapter 6 explicitly defines the objects, event selection criteria, and analysis region criteria utilized to obtain this analysis' final result.
- Chapter 7 introduces and presents the Matrix Element Method, the principal component of our analysis approach.
- Chapter 8 presents the development our implementation of the Matrix Element Method by presenting selected studies of the Matrix Element Method response.
- Chapter 9 presents two multivariate approaches for utilizing the output of the Matrix Element Method.
- Chapter 10 presents the sources of systematic uncertainty which were considered for this analysis.
- Chapter 11 presents the final results of this analysis, utilizing the techniques defined in Chapter 7 and Chapter 9, taking into account the sources of uncertainty presented in Chapter 10.

- Chapter 12 summarizes the results from Chapter 11 and presents a brief, speculative look to the future given the insights gained from this analysis.

Finally, the results of this analysis, specifically the search for the presence of a $t\bar{t}H\left(H \to b\bar{b}\right)$ signal in the 20.3 fb^{-1}, $\sqrt{s} = 8$ TeV dataset produced at the Large Hadron Collider and collected by the ATLAS detector, have been submitted for publication to the *European Physical Journal C*. [1]

Chapter 2

Theoretical Foundation and Motivations

Thanks to the epistemological framework of the *scientific method* and the intellectual effort of our predecessors, there exists a fund of knowledge that informs, directs, and motivates our current intellectual pursuits. This fund of knowledge spans countless academic disciplines allowing humanity to improve our collective condition as our understanding of our universe expands providing a foundation for deeper expeditions into the unknown. The theoretical foundation for this particular study is that of the Standard Model. This chapter seeks to provide the necessary theoretical context for this analytical endeavor, in addition to its motivating influence.

The following sources have been tremendously useful in writing this section: Peskin and Schroeder's *An introduction to quantum field theory* [3], Halzen and Martin's *Quarks and Leptons* [4], Griffiths' *Introduction to elementary particles* [5], the *Review of Particle Physics* [6], and Dr. El-Khadra's Subatomic Physics course notes, *Fall 2010, PHYS 570*.

2.1 The Standard Model

The Standard Model of high-energy physics (SM) codifies our present understanding of the most fundamental descriptions of matter and its interactions in the cosmos, excluding gravity, in its present incarnation. In many ways it resembles the periodic table of the elements in that it groups the constituent entities of particle physics in accordance with similarities in the ways in which these entities interact with each other. These similarities imply an underlying set of symmetries which govern a surprisingly small set of fundamental particles and interactions.

It is convenient to first classify the entities of the SM by their respective *spin* quantum numbers. Particles that have integer valued spin are called bosons, and particles with half-integer spin are known as fermions. In general terms, the bosons are localized excitations of the force fields contained in the SM, while the fermions are the localized excitations of matter fields whose interactions are mediated by the bosons.

The bosons are typically categorized according the role they play as mediators of a particular interaction. Table 2.1 shows the five observed bosons of the

Properties of the Standard Model Bosons

Mediator	**Symbol**	**Interaction**	**Mass** $\left[\mathrm{GeV}/c^2\right]$	**Q**	**Color Charge**	**Spin**
Gluon	g	Strong	0	0	Eight Possible Values	1
Photon	γ	EM	0	0	-	1
W boson	$W^{\pm}$	Weak	80.385 ± 0.015	± 1	-	1
Z boson	Z	Weak	91.1876 ± 0.0021	0	-	1
Higgs boson	H^0	EWSB	125.7 ± 0.4	0	-	0

Table 2.1: Table of the boson properties of the Standard Model. The boson masses are determined using direct measurements. [6] The electromagnetic interaction is abbreviated as EM, and the Yukawa interactions are abbreviated as EWSB after Electroweak Symmetry Breaking. Finally, **Q** denotes the electric charge of the particle in units of e.

SM along with a few of their properties including the interaction they mediate. Of particular note are their respective spin values. The first four bosons in the table all have spins of one. This implies that the forces they mediate are all described by vector fields originating from the requirement of local *gauge* invariance, as we shall see later. These bosons are thus referred to as the gauge bosons of the SM. The most recently observed boson, the Higgs boson, is unique in this regard as it has zero spin, implying its mediation of a scalar valued field, the only one of its kind observed in the SM.

In addition to the fundamental interactions of the SM, we find descriptions of fermions, particles that are similarly classified according to their respective quantum numbers. These quantum numbers point to the existence of symmetry groups which govern the boson and fermion interactions in the SM. Apart from the half-integer spin designation, fermions can be further subdivided into *quark* and *letpon* groups. The distinction between the two is largely driven by their principal interactions. Both groups interact weakly, but the leptons do not carry any color-charge, and therefore do not participate in the chromodynamic interaction, mediated by the gluon. Furthermore, the neutrinos do not interact electromagnetically, unlike the rest of the leptons and the quarks. Table 2.2 presents the three generations of fermions in the SM in addition to some of their properties.

2.2 The Theoretical Foundation of the Standard Model

In precise terms, the SM is a quantum field theory in which a Lagrangian density can be expressed that is manifestly gauge and Lorentz invariant pursuant to the constraints of unitarity and locality. These constraints, coupled

Properties of the Standard Model Fermions

Spin$=\frac{1}{2}$		**1st Generation**		**2nd Generation**		**3rd Generation**	
	Q		Mass $\left[\mathrm{MeV}/c^2\right]$		Mass $\left[\mathrm{MeV}/c^2\right]$		Mass $\left[\mathrm{GeV}/c^2\right]$
Quarks	$+\frac{2}{3}$	u	$2.3^{+0.7}_{-0.5}$	c	1275 ± 25	t	$173.21 \pm 0.51 \pm 0.71$
	$-\frac{1}{3}$	d	$4.8^{+0.5}_{-0.3}$	s	95 ± 5	b	4.18 ± 0.03
Leptons	-1	e^-	$0.51\cdots \pm 1.1 \cdot 10^{-8}$	μ^-	$105.6\cdots \pm 3.5 \cdot 10^{-6}$	τ^-	$1.7\cdots \pm 0.16 \cdot 10^{-3}$
	0	ν_e	$< 2 \cdot 10^{-6}$	ν_μ	< 0.19	ν_τ	< 0.0182

Table 2.2: Table of the fermion properties of the Standard Model. The quark masses are determined using the $\overline{MS}$ scheme, except for the top quark, t, which is determined from direct measurements (with associated statistical and systematic uncertainties). The neutrino mass exclusion limits are quoted at 95% confidence, except for the ν_μ which is quoted at 90% confidence. [6] Finally, **Q** denotes the electric charge of the particle in units of e.

with observational data, impose a SM based in the gague symmetry group $G_{SM} = SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. In this representation, the $SU(3)_C$ group describes the strong interaction, conserving the *color* quantum number, C. The $SU(2)_L \otimes U(1)_Y$ group describes the electroweak interaction which unifies the electromagnetic and weak interactions. For this interaction, the left-handed, L, isospin (T^3) and the hypercharge, Y, are conserved. The remainder of this section is devoted to describing the general framework of these groups, and their contributions to the SM Lagrangian.

2.2.1 Quantum Electrodynamics

Quantum Electrodynamics (QED) describes a quantized field theory characterizing the electromagnetic interaction. This interaction is mediated by the photon, γ, and involves all SM constituents with non-zero electric charge, a conserved quantity of the interaction. Interestingly, via the field theory approach, electrodynamics emerges from the requirement that a Lagrangian defining free fermions be locally gauge invariant, and not simply globally gauge invariant. This requirement is synonymous with the assertion that such a Lagrangian be a member of the unitary group, $U(1)$, a circle group, consisting of all complex numbers with an absolute value equal to unity, $\theta \mapsto e^{i\theta}$.

To show the emergence of QED from the local $U(1)$ symmetry requirement, we begin with the Dirac Lagrangian density, Eq. 2.1. This Lagrangian describes free fermion fields, ψ, parameterized by some mass m, which will allow quantized solutions of the form Eq. 2.2 (and it's Hermitian adjoint). Here, $a^s_{\mathbf{p}}$ denotes the fermion annihilation operator for the state $|\mathbf{p}, s\rangle$, for a particle with some spin, s, and some four-momentum $\mathbf{p}$. The antifermion creation operator is similarly denoted as $b^{s\dagger}_{\mathbf{p}}$. Finally, plane wave solutions for the fermion and antifermion

elements are denoted as $u^s(p)e^{-ip\cdot x}$ and $v^s(p)e^{ip\cdot x}$ respectively:

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi \tag{2.1}$$

$$\psi(x) = \int \frac{d^3p}{(2\pi)^3}\frac{1}{\sqrt{2E_{\mathbf{p}}}}\sum_s \left(a^s_{\mathbf{p}}u^s(p)e^{-ip\cdot x} + b^{s\dagger}_{\mathbf{p}}v^s(p)e^{ip\cdot x}\right). \tag{2.2}$$

It is manifestly evident that $\mathcal{L}_{\text{Dirac}}$ is invariant with respect to a global change of phase, $\psi \to e^{i\theta}\psi$. However, it can be shown that it isn't invariant with respect to a local phase change, defined by Eq. 2.3:

$$\psi \to e^{i\theta(x)}\psi \qquad\qquad \bar{\psi} \to e^{-i\theta(x)}\psi. \tag{2.3}$$

We can immediately see that under such a transformation, the mass term is locally invariant, but the kinetic term's derivative results in an additional term from the application of the product rule, shown in Eq. 2.4:

$$i\bar{\psi}\gamma^\mu\partial_\mu\psi \to i\bar{\psi}\gamma^\mu\partial_\mu\psi - \bar{\psi}\psi\gamma^\mu\partial_\mu\theta. \tag{2.4}$$

To make the kinetic term invariant, we can add a field, A_μ, to a modified derivative that transforms to cancel the offending term. This is done in the following steps, using g to represent the coupling strength of the field A_μ:

$$\partial_\mu \to D_\mu = \partial_\mu + igA_\mu(x) \tag{2.5}$$

$$A_\mu(x) \to A_\mu(x) - \frac{1}{g}\partial_\mu\theta(x). \tag{2.6}$$

The final result, is a new Lagrangian density which describes a field that is locally gauge invariant:

$$\mathcal{L}_{\text{Dirac}} \to \bar{\psi}(i\gamma^\mu D_\mu - m)\psi = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - g\bar{\psi}\gamma^\mu A_\mu\psi. \tag{2.7}$$

The gauge field introduced in Eq. 2.5 describes the photon, the quantized excitation of the electromagnetic field. To encode the possibility of observed free photon states, a kinetic term is added to the Lagrangian density, which is manifestly gauge invariant, which we designate $\mathcal{L}_{EM}$. It is interesting to note that a mass term of the form, $\frac{1}{2}MA^\mu A_\mu$, would not be gauge invariant, suggesting that the gauge boson associated with the A^μ field, the photon, is massless, which is consistent with observation:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{2.8}$$

$$\mathcal{L}_{\text{EM}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{2.9}$$

Pulling everything together, we see that the final Lagrangian density for QED, Eq. 2.12, can be broken into the following terms: a free fermion term, $\mathcal{L}_{\text{Dirac}}$, a free electromagnetic field term, $\mathcal{L}_{\text{EM}}$, and a term governing the interactions between the free fields, $\mathcal{L}_{\text{Interaction}}$:

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{2.10}$$

$$= \left[i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi\right] - \left[g\bar{\psi}\gamma^\mu A_\mu\psi\right] - \left[\frac{1}{4}F_{\mu\nu}F^{\mu\nu}\right] \tag{2.11}$$

$$= \mathcal{L}_{\text{Dirac}} + \mathcal{L}_{\text{Interaction}} + \mathcal{L}_{\text{EM}}. \tag{2.12}$$

This processes of promoting global gauge symmetries to local gauge symmetries, and subsequently utilizing the compensating fields to define the mediators of the resulting interaction terms serves as the template for further quantum field theories which underpin the SM.

2.2.2 Quantum Chromodynamics

Quantum Chromodynamics (QCD) governs the interactions between quarks and gluons. This quantized field is responsible for the bound quark states known as hadrons, such as protons and neutrons. In similar fashion to QED, QCD identifies a symmetry group and promotes a global gauge invariance of the group to a local gauge invariance requiring the addition of a mediating gauge field. However, unlike QED, QCD has a number of unique peculiarities which result in a much more complex quantum field theory.

The symmetry group QCD falls under is the special unitary Lie group of degree 3, or $SU(3)$, concerning the quantity known as *color charge*, due to its analogous nature to additive color mixing. Thus, each quark comes in one of three 'colors' (red, green, blue). This symmetry group is therefore $3^2 - 1$ dimensional and permits eight orthogonal unitary transformations which can be described by eight unique generators. Here, a generator is defined as an infinitesimal transformation with which successive applications can 'generate' a finite transformation. In the fundamental representation, the eight generators, T_α, are characterized by the Gell-Mann matrices, λ_α, where $\alpha = \{1, 2, 3, \cdots, 8\}$. These generators express a complex commutator structure, Eq. 2.14, making QCD, and by extension the SM, a non-Abelian gauge theory:

$$T_\alpha = \frac{\lambda_\alpha}{2} \tag{2.13}$$

$$[T_\alpha, T_\beta] = f_{\alpha\beta\delta}T_\delta. \tag{2.14}$$

Again, starting with the Dirac Lagrangian, Eq. 2.15, we specify a fermion field parametrizing the free quarks, $j = 1, 2, 3$ (red, green, blue), that participate

in the $SU(3)$ gauge interaction. Further, we specify the orthogonal transformations which characterize the $SU(3)$ gauge interaction, U, and require that local transformations leave the Lagrangian invariant:

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}_j \gamma^\mu \partial_\mu \psi_j - m\bar{\psi}_j \psi_j \tag{2.15}$$

$$\psi \to U\psi \tag{2.16}$$

$$\bar{\psi} \to U^\dagger \bar{\psi} \tag{2.17}$$

where:

$$U = e^{i\theta_\alpha(x) T_\alpha} \tag{2.18}$$

$$UU^\dagger = \mathbb{I}. \tag{2.19}$$

As before in the QED example, invariance is broken by the derivative term, Eq. 2.20, necessitating the inclusion of additional fields in an appropriate covariant derivative, Eq. 2.21, where g_s represents the strong force coupling constant:

$$\mathcal{L}_{\text{Dirac}} \to \mathcal{L}_{\text{Dirac}} + i\bar{\psi}_j \gamma^\mu U^\dagger (\partial_\mu U) \psi_j \tag{2.20}$$

$$\partial_\mu \to D_\mu = \partial_\mu - i g_s T^\alpha A^\alpha_\mu(x) \tag{2.21}$$

where:

$$T^\alpha A^\alpha_\mu \to U T^\alpha A^\alpha_\mu U^\dagger - \frac{i}{g_s} U^\dagger (\partial_\mu U). \tag{2.22}$$

Finally, we add a kinetic term for the mediators of the interaction in the same form as before. Again, we find symmetry requirements for a massless gauge boson, specifically, the gluon:

$$\mathcal{L}_{\text{QCD}} = \mathcal{L}_{\text{Dirac}} + \bar{\psi}_j \gamma^\mu T^\alpha A^\alpha_\mu \psi_j - \frac{1}{4} F^\alpha_{\mu\nu} F^{\mu\nu\alpha} \tag{2.23}$$

where:

$$F^\alpha_{\mu\nu} = \partial_\mu A^\alpha_\nu - \partial_\nu A^\alpha_\mu + g_s f^{\alpha\beta\delta} A^\beta_\mu A^\delta_\nu. \tag{2.24}$$

Expanding the gluon kinetic term in Eq. 2.23 provides an additional surprise. The non-Abelian commutation structure results in terms proportional to $g_s A^3$ and $g_s A^4$. These terms reveal the existence of three gluon and four gluon self-couplings, revealing the color-charged nature of the gluon. These terms have no analog in QED, and arise purely from the non-Abelian nature of the field theory.

In addition to the presence of gluon-gluon couplings, the non-Abelian nature of QCD gives rise to the qualitative condition of *asymptotic freedom*. This is of crucial importance as the strong coupling constant, g_s, is greater than one for momentum transfers, q^2, less than approximately 1 GeV. Known as *quark*

confinement, this behavior precludes the application of perturbation theory for calculating scattering amplitudes, effectively sidelining the Feynman calculus developed for QED. However, asymptotic freedom characterizes a phenomenon in which the strong coupling decreases with ever increasing momentum transfers, opposite to that of the QED momentum transfer-coupling strength relationship. This subsequently allows for the application perturbative methods in high energy regimes. Extensive phenomenology originating from the complex structure QCD plays an important role in this analysis, from the momentum structure of proton's constituents to the evolution of hard-scattered quarks, subjects which will be addressed in more detail later in this document.

2.2.3 Electroweak Interactions and the Spontaneous Symmetry Breaking

Quantum electrodynamics and quantum chromodynamics are based on the $U(1)$ and $SU(3)$ Lie group symmetries respectively. What remains is a description of the weak interactions that drive, for example, beta decay and the thermonuclear engines at the heart of every star.

In the same way that QED and QCD can be described under their Lie groups, the weak interactions can be described under the $SU(2)$ Lie group. Further, it can be shown that the electromagnetic interactions driving QED can be unified with the weak interactions to form a gauge invariant field theory described under the $SU(2)\otimes U(1)$ Lie group. This unified interaction is known as the electroweak interaction and marks a significant departure from the previous gauge invariant field theories of QED and QCD. First, the electroweak interaction is chiral, in that it couples differently to right and left-handed fermions, effectively violating parity symmetry. Second, the electroweak gauge symmetry is spontaneously broken, giving rise to additional Goldstone bosons that result in massive gauge bosons. Finally, the electroweak eigenstates of the quark fermion fields are not simultaneously mass eigenstates, leading to the mixing of up-type and down-type quarks.

Before we start to build a unified electroweak interaction, we shall first investigate its chiral property. From experiment we see a preferential coupling of weak interactions to left-handed leptons and right-handed anti-leptons. This preference was principally shown by the 1956 experiment led by C. S. Wu. In this experiment, cobalt-60 atoms undergoing weak-force mediated beta decay preferentially emitted electrons in a direction opposite to that of the nuclear spin vector. [7] This result shows that the mediator of beta decay couples preferentially to the resulting left-handed electrons and the right-handed anti-electron neutrinos, violating invariance under the parity operator, and thus violating parity symmetry. Further experiments show that this effect is *maximal*, in that charged current weak interactions do not couple at all to right-handed leptons and left-handed anti-leptons.

The experimental evidence of parity violation provides the basis for a *vector minus axial* theory, $V - A$. When we look at all the possible bilinear covariant combinations of γ-matrices, we see the following, $\bar{\psi}\{1, \gamma^\mu, \sigma^{\mu\nu}, \gamma^\mu\gamma^5, \gamma^5\}\psi$, where $\sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu, \gamma^\nu]$ and $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. Each of these possible covariants transform uniquely with the application of the parity operator. However, the chiral projection operators, $P_{\{R,L\}} = \frac{1}{2}\left(1 \pm \gamma^5\right)$, are required to achieve the necessary left-handed couplings seen in experiment. Adding the vector component, γ^μ, from QCD and QED, provides the final $V - A$ form of the weak interaction's charged current, $J^{\text{CC}}_{\text{weak}}$, shown by Eq. 2.25:

$$J^{\text{CC}}_{\text{weak}} \propto \bar{\psi}\gamma^\mu\left(1 - \gamma^5\right)\psi \tag{2.25}$$

$$\propto \bar{\psi}\gamma^\mu\psi - \bar{\psi}\gamma^\mu\gamma^5\psi \tag{2.26}$$

$$\propto [\text{Vector}] - [\text{Axial}]. \tag{2.27}$$

In the same manner as QED and QCD, we shall now build the electroweak interaction by requiring local gauge invariance. However, experimental evidence tells us to expect massive gauge bosons, something not permitted in either of the gauge theories already presented. Thus, identifying a natural way for gauge bosons to obtain mass becomes paramount. Our starting point will be the Glashow-Weinberg-Salam (GWS) model of electroweak interactions which satisfy the $SU(2)_L \otimes U(1)_Y$ symmetry containing three gauge bosons, A^α_μ where $\alpha = \{1, 2, 3\}$, from the generators of the $SU(2)_L$ group, and one gauge boson, B_μ, from the $U(1)_Y$ group. Using this notation, L denotes the coupling of the $SU(2)_L$ bosons to *left-handed* fields only, while the Y in $U(1)_Y$ stands for *hypercharge*, the conserved quantity of the $U(1)_Y$ symmetry. Starting with these fields, we can define a Lagrangian density describing their kinetic energy in a similar manner to QED and QCD. It can be seen in Eq. 2.29, that the $SU(2)_L$ symmetry is non-Abelian, resulting in a non-trivial commutator structure, $f^{\alpha\beta\delta}$:

$$\mathcal{L}_{\text{Kin.}} = -\frac{1}{4}A^\alpha_{\mu\nu}A^{\mu\nu\alpha} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \tag{2.28}$$

where:

$$A^\alpha_{\mu\nu} = \partial_\mu A^\alpha_\nu - \partial_\nu A^\alpha_\mu + g_e f^{\alpha\beta\delta}A^\beta_\mu A^\delta_\nu \tag{2.29}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \tag{2.30}$$

We can now introduce the matter fields in an appropriate chiral representation, using the chiral projection operators to describe the fermions through their left-hand and right-hand components:

$$\psi_{\{R,L\}} = \frac{1}{2}\left(1 \pm \gamma^5\right)\psi \tag{2.31}$$

$$\bar{\psi}_{\{R,L\}} = \frac{1}{2}\left(1 \mp \gamma^5\right)\bar{\psi}. \tag{2.32}$$

Taking these fermion fields, we can now define a covariant derivative, Eq. 2.33,

and an associated Lagrangian density for this effective electroweak interaction, Eq. 2.35. Neglecting the fermion mass terms of the form, $\bar{\psi}_L\psi_R$, leaves us with a Lagrangian density dependent on vector and axial interactions only:

$$D_\mu\psi_{\{L,R\}} = \left[\partial_\mu + ig_1\sum_{\alpha=1}^{3} T^\alpha_{L,R}A^\alpha_\mu + ig_2\frac{1}{2}Y_{L,R}B_\mu\right]\psi_{\{L,R\}} \tag{2.33}$$

where:

$$\left[T^\alpha_{L,R}, T^\beta_{L,R}\right] = if^{\alpha\beta\delta}T^\delta_{L,R} \tag{2.34}$$

$$\mathcal{L}_{\text{EW}} = -\frac{1}{4}\sum_{\alpha=1}^{3} A^\alpha_{\mu\nu}A^{\mu\nu\alpha} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + \bar{\psi}_L i\gamma^\mu D_\mu\psi_L + \bar{\psi}_R i\gamma^\mu D_\mu\psi_R. \tag{2.35}$$

Thus, using the following definitions, we can present a chiral $V-A$ theory with two massless charged currents, $W^\pm$, and two massless neural currents, W^0 and B^0. This theory explains the experimental data via its inclusion into Fermi's four-fermion effective theory. However, further experimental data shows that both charged current bosons, and one of the neutral current bosons, $W^\pm$ and Z^0, must be massive, implying a broken symmetry:

$$W^\pm_\mu = \frac{A^1_\mu \pm iA^2_\mu}{\sqrt{2}}; \qquad T^\pm = \frac{T^1_\mu \pm iT^2_\mu}{\sqrt{2}} \tag{2.36}$$

$$W^0 = A^3_\mu \tag{2.37}$$

$$B^0 = B_\mu. \tag{2.38}$$

Spontaneous Symmetry Breaking

It is clear that the unbroken $SU(2)_L \otimes U(1)_Y$ symmetry of the GWS model provides many of the necessary pieces for a unified electroweak theory consistent with data, except for the massive gauge bosons. Thus, this symmetry must be broken. Breaking this symmetry explicitly with mass terms input by hand would render the theory non-renormalizable and non-unitary. Thus, to maintain these requirements the symmetry must be broken *spontaneously*. This can occur when a ground state exists that is not invariant under some symmetry transformations. This means that, though the total Lagrangian is symmetric, the physical sates which exist near or at the ground state might not be.

A simple example of spontaneous symmetry breaking occurs when we start with a Lagrangian density for a $U(1)$ symmetric field and introduce a complex scalar field, ϕ, with an appropriately parameterized potential, $V(\phi)$:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + D_\mu\phi^* D^\mu\phi - V(\phi) \tag{2.39}$$

$$V(\phi) = -\mu^2(\phi^*\phi) + \lambda(\phi^*\phi)^2. \tag{2.40}$$

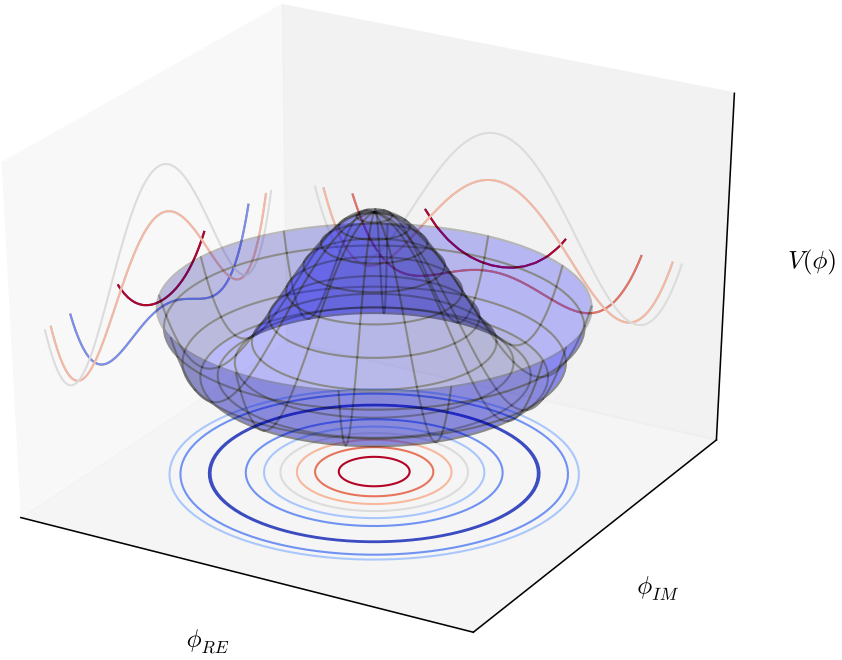


Figure 2.1: A mesh plot of a complex scalar field potential with a non-zero ground-state expectation value. Such a potential is described by Eq. 2.40.

When $\mu^2 > 0$ and $\lambda > 0$, the potential, while remaining rotationally symmetric with respect to $U(1)$ about the origin, develops a rotationally symmetric minimum displaced from the origin by ϕ_0. An example of such a configured potential is shown in Figure 2.1, resulting in a ground state at:

$$\phi_0 = \left(\frac{\mu^2}{\lambda}\right)^2 . \tag{2.41}$$

Expanding the scalar field around ϕ_0, and decomposing the complex scalar field into real and imaginary components gives Eq. 2.42:

$$\phi(x) = \phi_0 + \frac{1}{\sqrt{2}}\left(\phi_1(x) + i\phi_2(x)\right) . \tag{2.42}$$

We now rewrite the potential in terms of the expanded scalar field such that the real portion of the field, ϕ_1, acquires a mass term, squared in ϕ, where as the complex portion, ϕ_2, does not. Here, the mass of the ϕ_1 boson is $m = \sqrt{2}\mu$:

$$V(\phi) = -\frac{\mu^4}{2\lambda} + \frac{1}{2}\cdot 2\mu^2\phi_1^2 + \mathcal{O}\left(\phi_i^3\right) . \tag{2.43}$$

Looking closer at the kinetic term for the scalar field in the vicinity of ϕ_0, but neglecting terms cubic and quartic in A_μ, ϕ_1, and ϕ_2, we find a very important effect of the shifted ground state. We see that the gauge boson develops a squared term proportional to $e^2\phi^2$, which was not explicitly defined prior to the

spontaneous symmetry breaking:

$$|D_\mu \phi|^2 = \frac{1}{2}(\partial_\mu \phi_1)^2 + \frac{1}{2}(\partial_\mu \phi_2)^2 + \frac{e}{\sqrt{2}}\phi_0 \cdot A_\mu \partial^\mu \phi_2 + e^2 \phi_0^2 A_\mu A^\mu + \cdots . \quad (2.44)$$

Thus, the photon from the $U(1)$ gauge field, has picked up a mass term proportional to the vacuum expectation value, ϕ_0, of the spontaneously broken complex scalar field. Its mass is therefore equal to, $m_A = 2e^2\phi_0^2$ given the general form of mass terms, $\Delta\mathcal{L} = \frac{1}{2}m_A^2 A_\mu A^\mu$. The ϕ_2 field thus donates its degree of freedom to the massless photon, allowing it to have a non-zero mass. This scalar boson is a known as a Nambu–Goldstone boson, after its discoverers Yoichiro Nambu and Jeffrey Goldstone. Though its role is significant, the Goldstone boson does not appear as an independent, physical particle, unlike the the ϕ_1 boson. This can be readily shown by a $U(1)$ transformation of the complex scalar field such that for all points, x, $\phi(x)$ is real valued. This gauge transformation places $\phi(x)$ into the *unitarity gauge*, effectively removing ϕ_2 from the theory.

2.2.4 The Higgs Mechanism

The process of extending the Abelian example of spontaneous symmetry breaking to non-Abelian theories was explored by Higgs [8]; Guralnik, Hagen, and Kibble [9]; and, Englert and Brout [10]; eventually coming to be known as the *Higgs Mechanism*. By the summer of 2012, sufficient physical evidence of its existence was obtained so as to award Peter Higgs and François Englert with the 2013 Nobel Prize in Physics.

Revisiting the GWS theory, with the Higgs Mechanism machinery in hand, we can begin the process of utilizing a spontaneously broken scalar field to provide the necessary gauge boson masses. We similarly begin with the $SU(2)_L \otimes U(1)_Y$ symmetry group, and introduce an additional $U(1)$ gauge symmetry, described by a complex scalar field in the spinor representation of $SU(2)$, with an associated conserved hypercharge of $Y = +1/2$. This results in the transformations defined by Eq. 2.45:

$$\phi \rightarrow e^{i\alpha^a T^a} e^{i\beta/2}\phi \quad (2.45)$$

$$T^a = \frac{\sigma^a}{2}. \quad (2.46)$$

Further, if we presume that this field has the proper potential structure necessary to develop a non-zero vacuum expectation value in the manner described in the section on spontaneous symmetry breaking, the vacuum expectation value, ϕ_0, will acquire the form of Eq. 2.47. Further, the gauge transformations described by Eq. 2.48 will leave ϕ_0 invariant. We will continue to refer

to v as the Higgs field *vacuum expectation value*:

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{2.47}$$

$$\alpha^1 = \alpha^2 = 0, \qquad \alpha^3 = \beta. \tag{2.48}$$

We now apply the GWS covariant derivative, Eq. 2.33, to this scalar field, and evaluate the matrix product in the vicinity of the vacuum expectation value:

$$D_\mu \phi = \left[\partial_\mu + i g_1 \sum_{\alpha=1}^{3} T_L^a A_\mu^a + i g_2 \frac{1}{2} Y B_\mu \right] \phi. \tag{2.49}$$

Evaluating the covariant derivative product about the vacuum expectation value reveals the structure shown in Eq. 2.50. Cleaning up a little leaves the relevant terms shown in Eq. 2.51, were we denote the partial Lagrangian evaluated at v as $\Delta\mathcal{L}$:

$$\begin{aligned}(D_\mu \phi)^\dagger D^\mu \phi = \frac{1}{2} \phi_0^\dagger \left[\partial_\mu + i g_1 T_L^a A_\mu^a + \frac{i g_2}{2} Y B_\mu \right] \cdot \ldots \\ \left[\partial^\mu + i g_1 T_L^b A^{\mu b} + \frac{i g_2}{2} Y B^\mu \right] \phi_0\end{aligned} \tag{2.50}$$

$$\begin{aligned}\Delta\mathcal{L} = \frac{1}{2} \begin{pmatrix} 0 & v \end{pmatrix} \left[g_1 T_L^a A_\mu^a + \frac{g_2}{2} Y B_\mu \right] \cdot \ldots \\ \left[g_1 T_L^b A^{\mu b} + \frac{g_2}{2} Y B^\mu \right] \begin{pmatrix} 0 \\ v \end{pmatrix}.\end{aligned} \tag{2.51}$$

Carrying through the matrix product reveals the emergence of the gauge boson mass terms from the spontaneously broken scalar gauge field symmetry:

$$\Delta\mathcal{L} = \frac{v^2}{8} \left[g_1^2 \left(A_\mu^1 \right)^2 + g_1^2 \left(A_\mu^2 \right)^2 + \left(-g_1 A_\mu^3 + g_2 B_\mu \right)^2 \right]. \tag{2.52}$$

The bosons can be subsequently grouped in such a way as to reveal the structure of the couplings seen in data, similarly corresponding to the masses of the observed massive gauge bosons. The three massive bosons correspond to the two charged current $W^\pm$ bosons, and the neutral Z^0 boson. The fourth vector

field, orthogonal to Z^0_μ, describes the neutral photon, γ, and remains massless:

$$W^\pm_\mu = \frac{1}{\sqrt{2}}\left(A^1_\mu \mp iA^2_\mu\right) \qquad m_W = g_1\frac{v}{2} \tag{2.53}$$

$$Z^0_\mu = \frac{\left(g_1A^3_\mu - g_2B_\mu\right)}{\sqrt{g_1^2+g_2^2}} \qquad m_Z = \frac{v}{2}\sqrt{g_1^2+g_2^2} \tag{2.54}$$

$$A_\mu = \frac{\left(g_1A^3_\mu + g_2B_\mu\right)}{\sqrt{g_1^2+g_2^2}} \qquad m_\gamma = 0. \tag{2.55}$$

The two neutral currents can be reconfigured via the appropriate trigonometric identities to show the relative mixing of the A^3_μ and B_μ in terms of a 'weak mixing' angle parameter, θ_w. This angle is thus defined as, $\sin\theta_w = \frac{g_2}{\sqrt{g_1^2+g_2^2}}$:

$$Z^0_\mu = \frac{\left(g_1A^3_\mu - g_2B_\mu\right)}{\sqrt{g_1^2+g_2^2}} = A^3_\mu\sin\theta_w - B_\mu\cos\theta_w \tag{2.56}$$

$$A_\mu = \frac{\left(g_1A^3_\mu + g_2B_\mu\right)}{\sqrt{g_1^2+g_2^2}} = A^3_\mu\sin\theta_w + B_\mu\cos\theta_w. \tag{2.57}$$

Utilizing the photon coupling constant and relating it to it known coupling to the electric charge as, e, we can identify the electric charge quantum numbers, Q, from the conserved quantities of the GWS theory:

$$e = \frac{g_1g_2}{\sqrt{g_1^2+g_2^2}} \tag{2.58}$$

$$Q = T^3 + Y. \tag{2.59}$$

Finally, using the experimentally verified masses of the W and Z^0 bosons, we can identify the Higgs vacuum expectation value, v, which sets the scale for electroweak unification:

$$v = \frac{2m_W}{g_1} = 246 \text{ GeV}. \tag{2.60}$$

However, we have insufficient theoretical constraints to determine the mass of the physical boson associated with the Higgs Mechanism. This boson, known as the Higgs boson, has a mass which is therefore a free parameter of the theory. Recently, a new scalar boson signal, consistent with the SM Higgs boson signature, has been observed with a mass of 125.7 ± 0.4 GeV [6] adding significant experimental weight behind the Higgs Mechanism's place in the SM.

2.2.5 The Masses of the Fermions

Up until this point, the fermion fields of the GWS theory have been simply treated as a means to show the action of the GWS covariant derivative for the $SU(2)_L \otimes U(1)_Y$ gauge symmetry. Furthermore, as this development of

the GWS theory did not originate from the Dirac equation as was the case for QED and QCD, the mass terms of the fermion fields have been conspicuously absent. Fortunately, an additional function of the Higgs Mechanism provides the necessary terms by which the fermion masses can be generated.

Before we introduce the generation of the fermion masses by the Higgs Mechanism, it is important to introduce the fermion objects in accordance with the $SU(2)_L \otimes U(1)_Y$ symmetry and its conserved quantities charge, Q, weak-isospin T^a, and hypercharge, Y, which are each related to the previously introduced Gell-Mann–Nishijima relation, Eq. 2.59. This relationship allows for the categorization the fermions according to their respective quantum numbers. Furthermore, as the GWS theory is manifestly chiral, two tables exist for the left-handed (Table 2.3) and right-handed (Table 2.4) projections of the GWS fermion eigenstates.

Furthermore, according to the $SU(2)_L$ symmetry, we can form the left-handed fermions into doublet-eigenstates of the weak-isospin quantum number, $T = \frac{1}{2}$, and the right-handed fermions into singlet states, $T = 0$. Equations 2.61 through 2.64, show the left and right handed states for the leptons, $\mathscr{L}$, and the quarks, $\mathcal{Q}$.

$$\mathscr{L}_L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \tag{2.61}$$

$$\mathcal{Q}_L = \begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b \end{pmatrix} \tag{2.62}$$

$$\mathscr{L}_R = e, \mu, \tau \tag{2.63}$$

$$\mathcal{Q}_R = u, d, c, s, t, b \tag{2.64}$$

Beginning with the leptons, we may construct a gauge invariant coupling which links together the left-handed and right-handed components, via the complex scalar field in its spinor representation. Such a term, in the case of an electron, is shown in Eq. 2.65. Of important note here is how the Higgs field spinor and the $SU(2)$ doublet contract with one another, in addition to the cancelation of the $U(1)$ hyper-charges. Couplings of this nature are known as Yukawa couplings, due to their similarity with the Yukawa interaction describing inter-nucleon forces:

$$\Delta\mathcal{L}_e = -\lambda_e \mathscr{L}_L^{e\dagger} \phi \mathscr{L}_R^e + \text{h.c.} \tag{2.65}$$

$$= -\lambda_e \begin{pmatrix} \nu_e & e \end{pmatrix}_L \phi \, e_R + \text{h.c.} \tag{2.66}$$

Taking Eq. 2.66, and replacing ϕ with the expression for the Higgs field vacuum expectation value gives us the mass term Eq. 2.68, with an apparent mass of $m_e = \frac{\lambda_e v}{\sqrt{2}}$. Thus, the free fermion mass terms of the SM have been

Quantum Numbers of the Left-Handed Fermions

	Left-Handed	Q	T	T^3	Y	**Left-Handed**	Q	T	T^3	Y
Quarks	u_L, c_L, t_L	$+\frac{2}{3}$	$\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{3}$	$\bar{u}_L, \bar{c}_L, \bar{t}_L$	$-\frac{2}{3}$	0	0	$-\frac{4}{3}$
	d_L, s_L, b_L	$-\frac{1}{3}$	$\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{3}$	$\bar{d}_L, \bar{s}_L, \bar{b}_L$	$+\frac{1}{3}$	0	0	$+\frac{2}{3}$
Leptons	$\nu_L^e, \nu_L^\mu, \nu_L^\tau$	0	$\frac{1}{2}$	$+\frac{1}{2}$	-1	$\bar{\nu}_L^e, \bar{\nu}_L^\mu, \bar{\nu}_L^\tau$	Sterile (if they exist)			
	e_L^-, μ_L^-, τ_L^-	-1	$\frac{1}{2}$	$-\frac{1}{2}$	-1	$\bar{e}_L^+, \bar{\mu}_L^+, \bar{\tau}_L^+$	$+1$	0	0	$+1$

Table 2.3: Table of the left-handed fermions and their associated quantum numbers as defined by the GWS theory of electroweak interactions. Here, Q denotes the electromagnetic charge, T is the total weak-isospin, T^3 is the third component of the weak-isospin, and Y is the hypercharge quantum number associated with the $U(1)_Y$ symmetry. It is important to note that weak charged currents couple only to particles with non-zero weak-isospin. Thus, left-handed anti-fermions do not couple to the $W^\pm$ bosons.

Quantum Numbers of the Right-Handed Fermions

	Right-Handed	Q	T	T^3	Y	**Right-Handed**	Q	T	T^3	Y
Quarks	u_R, c_R, t_R	$+\frac{2}{3}$	0	0	$+\frac{4}{3}$	$\bar{u}_R, \bar{c}_R, \bar{t}_R$	$-\frac{2}{3}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{3}$
	d_R, s_R, b_R	$-\frac{1}{3}$	0	0	$-\frac{2}{3}$	$\bar{d}_R, \bar{s}_R, \bar{b}_R$	$+\frac{1}{3}$	$\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{1}{3}$
Leptons	$\nu_R^e, \nu_R^\mu, \nu_R^\tau$	Sterile (if they exist)				$\bar{\nu}_R^e, \bar{\nu}_R^\mu, \bar{\nu}_R^\tau$	0	$\frac{1}{2}$	$-\frac{1}{2}$	$+1$
	e_R^-, μ_R^-, τ_R^-	-1	0	0	-1	$\bar{e}_R^+, \bar{\mu}_R^+, \bar{\tau}_R^+$	$+1$	$\frac{1}{2}$	$+\frac{1}{2}$	$+1$

Table 2.4: Table of the right-handed fermions and their associated quantum numbers as defined by the GWS theory of electroweak interactions. Here, Q denotes the electromagnetic charge, T is the total weak-isospin, T^3 is the third component of the weak-isospin, and Y is the hypercharge quantum number associated with the $U(1)_Y$ symmetry. It is important to note that weak charged currents couple only to particles with non-zero weak-isospin. Thus, right-handed fermions do not couple to the $W^\pm$ bosons.

replaced with free fermion couplings, λ_i, to the scalar Higgs field:

$$\Delta\mathcal{L}_e = -\frac{1}{\sqrt{2}}\lambda_e \begin{pmatrix} \nu_e & e \end{pmatrix}_L \begin{pmatrix} 0 \\ v \end{pmatrix} e_R + \text{h.c.} \tag{2.67}$$

$$= -\frac{1}{\sqrt{2}}\lambda_e v e_L^\dagger e_R + \text{h.c.} + \text{interaction terms.} \tag{2.68}$$

If we reconfigure the Higgs field as oscillations of a physical particle about the vacuum expectation value, and subsequently transform into the unitarity gauge, we can reinterpret this particular coupling. Making the necessary transformations, the Higgs field, $\phi(x)$, becomes:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{2.69}$$

As a result of this transformation, the fermion coupling terms acquire the form of Eq. 2.70, making explicit the relationship between a fermion's mass and its coupling to the physical Higgs boson:

$$\Delta\mathcal{L}_f = -m_f f^* f \left(1 + \frac{h}{v}\right). \tag{2.70}$$

This framework can be further extended to generate the quark masses. Its application is straight forward in the case of the down-type quarks, however, the invariant anti-symmetric tensor, ϵ^{ab} must be employed to appropriately generate the mass terms for the up-type quarks:

$$\Delta\mathcal{L}_\mathcal{Q} = -\lambda_d \mathcal{Q}_L^{d\dagger} \phi \mathcal{Q}_R^d - \lambda_u e^{ab} \mathcal{Q}_L^{au\dagger} \phi^b \mathcal{Q}_R^u + \text{h.c.} \tag{2.71}$$

$$= -\frac{\lambda_d v}{\sqrt{2}} d_L^\dagger d_R - \frac{\lambda_u v}{\sqrt{2}} u_L^\dagger u_R + \text{h.c.} + \text{interaction terms} \tag{2.72}$$

$$= -m_d d^* d \left(1 + \frac{h}{v}\right) - m_u u^* u \left(1 + \frac{h}{v}\right). \tag{2.73}$$

However, when we extend the number of quark generations beyond simply u and d, additional terms allow for the generations to mix with one another. Thus the Higgs couplings and the gauge couplings cannot be simultaneously diagonalized. This mixing however can be described by a single angle parameter for the first two generations, but must be generalized to three angles and a CP violating phase in the three generation case. As such it is helpful to generalize transformations between the mass eigenstates and the electroweak eigenstates via a unitary transformation matrix known as the Cabibbo–Kobayashi–Maskawa (CKM) matrix. If we let

$$u_L^i = (u_L, c_L, t_L), \qquad d_L^i = (d_L, s_L, b_L), \tag{2.74}$$

we can describe the change of basis from the electroweak basis, u^i and d^i, to the

mass eigenstate basis, u'^i and d'^i, via the following unitary transformations:

$$u_L^i = U_u^{ij} u_L'^j, \qquad d_L^i = U_d^{ij} d_L'^j. \tag{2.75}$$

This allows us to then write the charged electroweak current in terms on the mass eigenstates:

$$J_{W^+}^\mu = \frac{1}{\sqrt{2}} u_L^{i*} \gamma^\mu d_L^i \tag{2.76}$$

$$= \frac{1}{\sqrt{2}} u_L'^{i*} \gamma^\mu \left(U_u^\dagger U_d\right) d_L'^j \tag{2.77}$$

$$= \frac{1}{\sqrt{2}} u_L'^{i*} \gamma^\mu \left(V_{ij}\right) d_L'^j. \tag{2.78}$$

The parameters of the CKM matrix are experimentally determined and result in the following matrix elements [6]:

$$V_{\text{CKM}} = \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \tag{2.79}$$

$$V_{\text{CKM}} = \begin{pmatrix} 0.97427 \pm 0.00014 & 0.22536 \pm 0.00061 & 0.00355 \pm 0.00015 \\ 0.22522 \pm 0.00061 & 0.97343 \pm 0.00015 & 0.0414 \pm 0.0012 \\ 0.00886 \pm 0.00033 & 0.0405 \pm 0.0012 & 0.99914 \pm 0.00005 \end{pmatrix}. \tag{2.80}$$

2.3 Motivation for the measurement of $t\bar{t}H\left(H \to b\bar{b}\right)$

With the final pieces in place from the previous sections, we are left with an amazingly accurate model of the visible universe. For example, Figure 2.2 compares a few SM predictions with observations made by the ATLAS detector at the LHC. The SM explains observed phenomena over a staggering breadth of energy scales, yet there are still questions left to be answered. As amazingly accurate as the SM is, there still remain discrepancies that point to the existence of a more general, more *ultimate* theory of nature. As a result, efforts to throw back the curtain obscuring such a theory is of utmost importance to theorists and experimentalists alike. Thus, the search for new physics (NP) continues.

With the discovery of a new boson, consistent with the SM Higgs boson, with a mass of 125 GeV, effort has turned to measuring its various characteristics. For example, Figure 2.3 shows the measured signal strengths of the various SM Higgs boson couplings. Of particular interest is the coupling between the Higgs boson and the top quark which is conspicuously absent from Figure 2.3. This particular parameter of the SM has very strong cosmological implications, partially due to the effect it has on the effective Higgs field potential, providing insight regarding

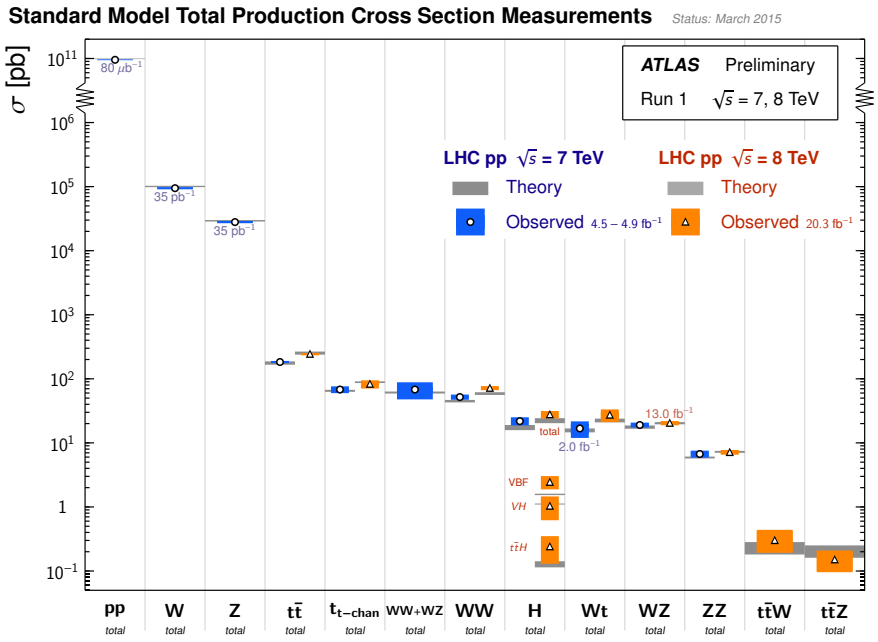


Figure 2.2: Standard Model observation summary using LHC data at $\sqrt{s}$=7 and 8 TeV. [11]

the scale of potential new physics. [12] Furthermore, there exists the possibility that as-yet unidentified massive fermions could contribute to the fermion loops that effectively couple the Higgs to the massless gluon and photon. [13–16] Finally, and most generally, it is important to open up as many perspectives on the Higgs as possible; exploring the unknown provides further opportunities for serendipitous discovery. Motivated by these aims, we set our sights on the top-Higgs Yukawa coupling, λ_t.

As mentioned previously, the Yukawa coupling is defined by the SM to be proportional to the mass of the top quark and consequently of order one. As a result, this parameter can be explored through multiple methods. First, we can assume that the Higgs-fermion couplings work as the SM predicts and use top quark mass measurements to pin down λ_t according to the relation derived earlier:

$$m_t = \lambda_t \frac{v}{\sqrt{2}} \approx \lambda_t \cdot 174 \text{ GeV} \tag{2.81}$$

$$\lambda_t = \frac{m_t}{174 \text{ GeV}} \approx 0.995. \tag{2.82}$$

Though accurate measurements of the top quark mass are essential to validating the SM, its presumption of the mass generating mechanism *a priori* renders it incapable of validating the mass generating feature of the Higgs Mechanism.

A second method utilizes cross section measurements of gluon-fusion based Higgs production and di-photon Higgs decay. As mentioned above, these effec-

tive couplings of the Higgs to massless bosons proceed through fermion loops, which in principal should be dominated by the top quark loop contribution. However, this indirect measurement of λ_t as a direct coupling of the Higgs boson and the top quark presumes no new physics contributions to the loop. Even in the case that rates are found to be consistent with SM prediction, the ability for loop contributions to interfere with one another leaves open the possibility that NP contributions could conspire with a non-SM λ_t to produce an effective coupling ostensibly consistent with the SM.

A final possibility is the direct measurement of the direct coupling between the top quark and the Higgs boson at tree level. This occurs in the $t\bar{t}H$ production mechanism where the Higgs is produced in association with a top-anti-top quark pair. As multiple tree-level diagrams contribute, this process includes top-fusion Higgs couplings as well as Higgs radiation from one of the top quarks. However, unlike the previous method, this coupling explicitly links a top quark and the Higgs boson, resulting in production cross sections dependent only on λ_t and other known SM couplings. It is this final method that we shall explore in this dissertation thanks to its few complicating assumptions.

Unfortunately, far from the low-hanging fruit of the production mechanisms used to currently observe the Higgs, the $t\bar{t}H$ production mechanism comes with a slew of complicating factors resistant to all but the most ambitious and technical analysis approaches. This mechanism is the rarest of the 'common' Higgs production mechanisms and is *nearly* indistinguishable from the far more prolific SM $t\bar{t}+$ jets backgrounds. Fortunately, this analysis is up to the challenge at hand.

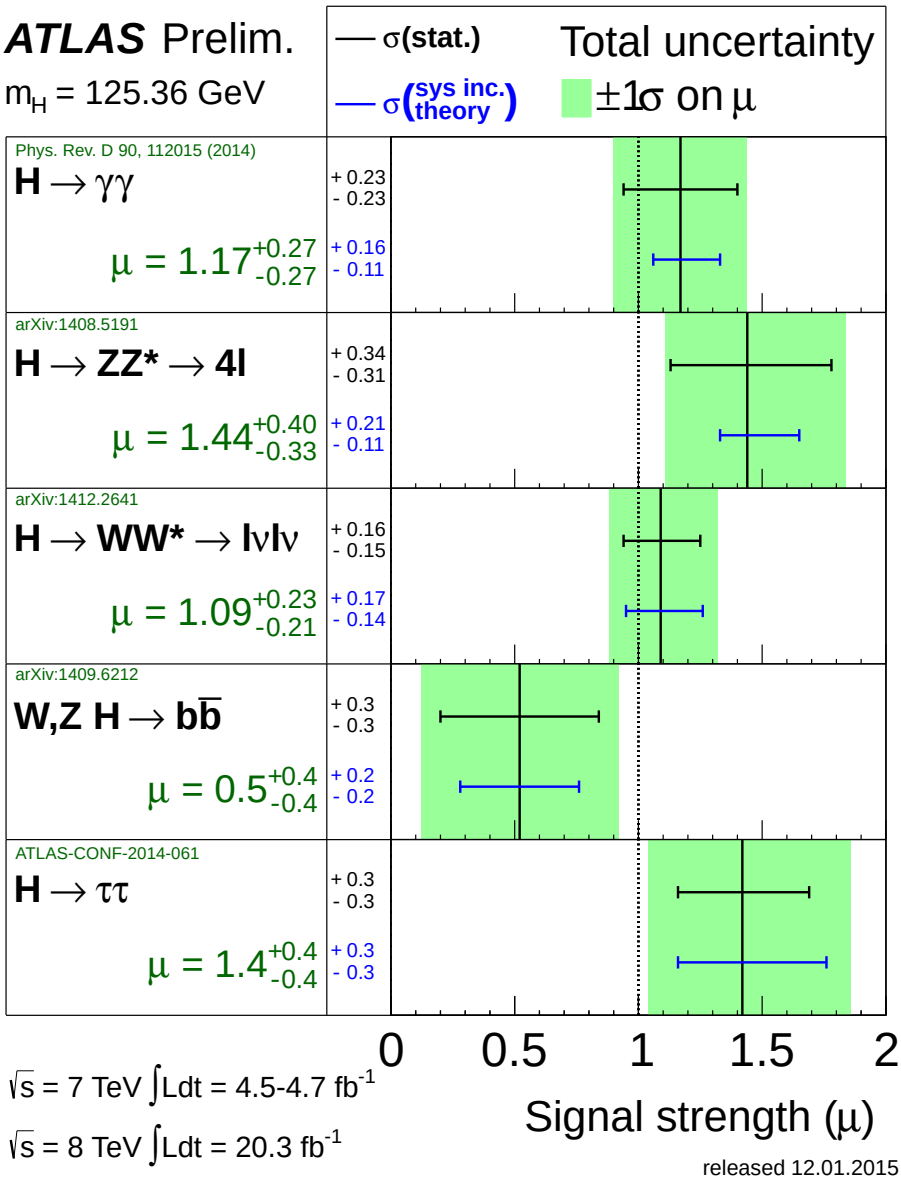


Figure 2.3: Higgs-like boson observation summary using LHC data at $\sqrt{s}$=7 and 8 TeV. [17]

Chapter 3

ATLAS and the Large Hadron Collider

In 1984, the year following the discovery of the W and Z bosons at CERN's Super Proton Synchrotron (SPS), the *Workshop on the Feasibility of Hadron Colliders in the LEP Tunnel* took place. It was there at the University of Lausanne in Lausanne, Switzerland that the intellectual and administrative foundations for the Large Hadron Collider (LHC) were laid.

The LHC would be constructed in the 27 km long Large Electron-Proton Collider (LEP) tunnel outside of Geneva, Switzerland and would be composed of superconducting magnets with an initial collision energy goal of 10 TeV in the center-of-mass frame. (A TeV, an abbreviation for tera-electron volt, represents the amount of kinetic energy gained by an electron after traversing an electric potential difference of one volt, one trillion times.) A two ring pp collider was favored over the alternative single ring $p\bar{p}$ collider in part to reach higher luminosities on the order of 10^{34} cm^{-2} sec^{-1} . And finally, its physics case was made, to ascertain "the deep origin of mass" and "the relation between masses and symmetry breaking processes." In other words, the search for the Higgs boson had begun in Europe. [19]

In 1992 a *Letter of Intent* [20] was published proposing the development of '**A T**oroidal **L**HC **A**pparatu**S**,' or ATLAS, experiment following the merger of the EAGLE and ASCOT collaborations. In the following years prototypes of various detector sub-systems were designed and tested in parallel with the construction of ATLAS's service infrastructure.These efforts culminated in 1999 with the release of the *Technical Design Report* [21].

Components of what would eventually become the ATLAS Detector were developed at member institutions across the globe over the course of the next few years. Starting in 2003 those finished components would start to be shipped to CERN for installation in the ATLAS experimental cavern adjacent to CERN's Meyrin facility.

Finally, on the 23rd of November, 2009, 25 years following the initial planning workshop, data was taken capturing the first pp collisions of the Large Hadron Collider. These first collisions registered a humble center-of-mass collision energy, or $\sqrt{s}$, of 900 GeV. However, by the start of Run 1, ATLAS was poised to collect an unprecedented sample of $\sqrt{s}$=7 TeV, pp collisions [22] marking the start of the LHC era in fundamental particle physics.

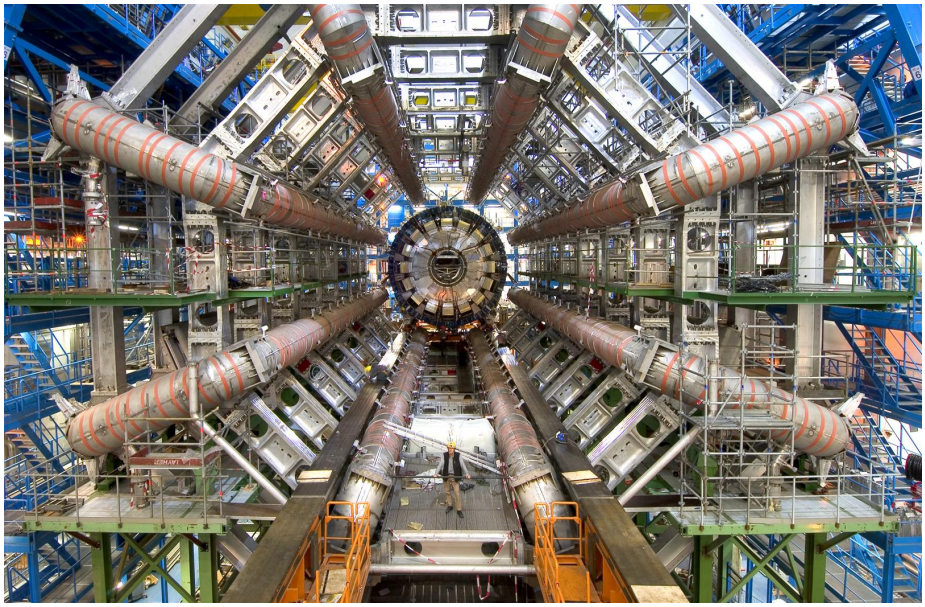


Figure 3.1: The iconic ATLAS toroid magnet just prior to the installation of the calorimeters. This particular magnet, made up of 8 superconducting electromagnetic coils, is the largest at ATLAS and stores 1.1 GJs of magnetic energy during normal operation. Its purpose is to provide the necessary magnetic field for precision muon momentum measurements in the muon spectrometer. [18]

The remainder of this chapter will cover the technical aspects of the Large Hadron Collider and the ATLAS detector, the underlying physical principles that govern their fundamental applications, as well as the general conventions and nomenclature common to collider-based experimentation.

3.1 The Large Hadron Collider

3.1.1 Design Information

The Large Hadron Collider's purpose is to accelerate hadronic matter, protons or atomic nuclei, to large fractions of the speed of light. Upon reaching the desired energy, the LHC collides that matter such that the various experiments at the LHC can analyze the debris from these collisions. To that end, the LHC is comprised of two, counter-propagating beam-lines. The beams of hadronic material are circulated in evacuated beam pipes, surrounded by high-field superconducting magnets. The beams intersect one another at the center of each experimental cavern. These crossing points are referred to as interaction points and constitute the focal point of each experimental apparatus. At the LHC, there are 4 experiments, as can be seen in Figure 3.2. ATLAS [24], at Point 1, and CMS [25], at Point 5, are general purpose detectors designed to study a wide range of physical phenomena using large general purpose detectors and detector technologies. The other two experiments, ALICE [26], at Point 2, and

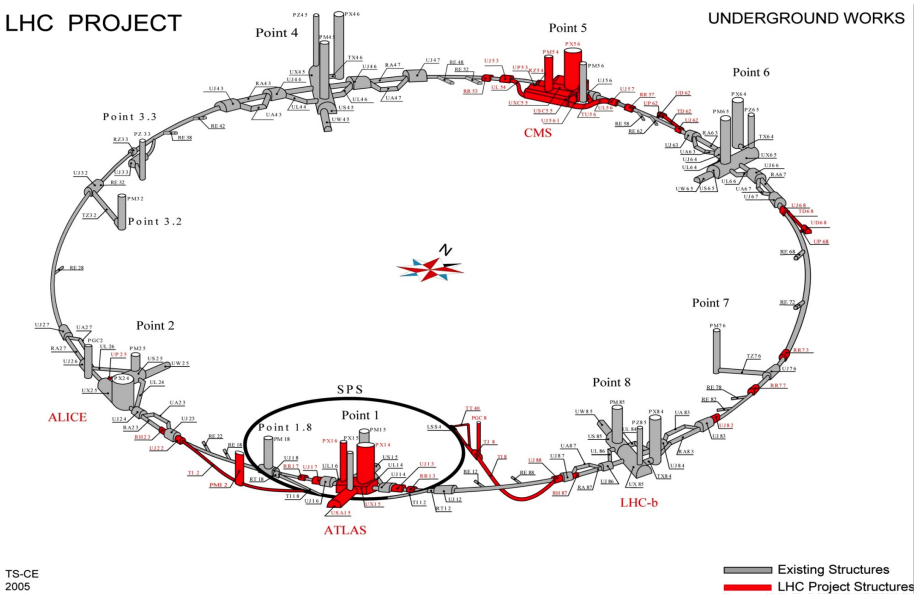


Figure 3.2: A rendered model of the LHC tunnel and cavern system. The grey structures represent the existing infrastructure inherited from the LEP collider, while the red structures represent the new works commissioned for the LHC. At 27 km in circumference and between 40 and 170 meters underground, the LHC tunnel represents a significant accomplishment in terms of civil works in its own right. Figure taken from "CERN Civil Engineering [23]."

LHCb [27], at Point 8, on the other hand are each specialized detectors built with the express purpose of uncovering insights regarding heavy ion collisions and B-meson physics respectively. Points 3 and 7 contain equipment used to collimate the beam, while Point 6 contains the Beam Dump System . Point 4 contains the accelerating cavities described in Section 3.1.1.

The design goals for the Large Hadron Collider advance the particle physics frontier along two primary axes. First, the LHC is designed to operate at an unprecedented collision energy of 14 TeV in the center-of-mass frame of reference. This is crucial as higher collision energies allow for the observation of heavier objects which require greater and greater momentum transfers.

Along the other axis, the LHC seeks to provide a high-luminosity source of these collisions on the order of 10^{34} cm^{-2} sec^{-1} or 100 nb^{-1}s^{-1}. The instantaneous luminosity, given in units of cm^{-2} sec^{-1} , is roughly proportional to the number of collisions occurring at a given time such that the time-integrated luminosity defines a measure of the total number of generated collisions. Such a high luminosity is necessary to generate enough data to adequately study the rare processes of interest.

To achieve this, the LHC was designed to collide protons in discrete packets referred to as bunches with a crossing frequency of 40 MHz. Thus, at full design luminosity, each bunch crossing occurs at 25 ns intervals. Further, each bunch contains up to 10^{11} protons. Coupled with an inelastic proton-proton

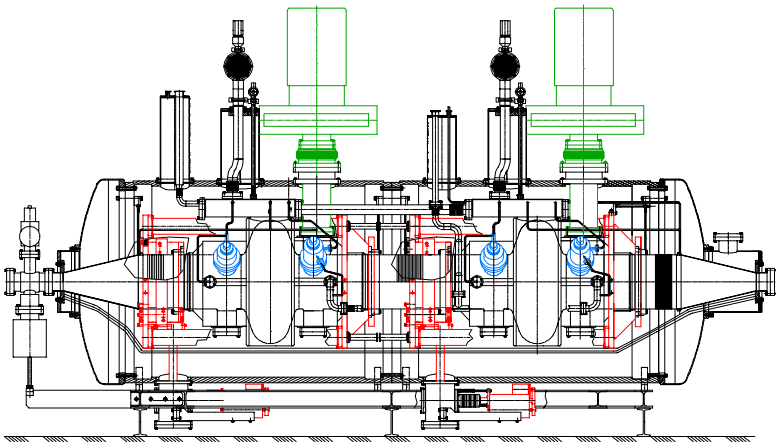


Figure 3.3: A technical layout of a 2-cell prototype superconducting radio frequency cavity. The actual production system located at Point 4 contains 8 single cell cavities, 4 per cryo-module, per beam. [28]

cross-section on the order of 70-80 mb, the LHC produces nearly 10^9 inelastic proton-proton events per second.

Using the 8 TeV run in 2012 as an example, the LHC was able to provide a peak instantaneous luminosity of 7.7×10^{33} cm^{-2} sec^{-1} . Integrated over the full 2012 proton-proton run, a total of 22.8 fb^{-1}of data was delivered of which 20.3 fb^{-1}was collected by the ATLAS detector and deemed to be of sufficient quality for physics analysis. During this run, there were, on average, 21 distinct proton-proton interactions per bunch crossing.

Acceleration

To accelerate the hadrons the LHC utilizes a superconducting radio frequency (RF) cavity system located at Point 4. Each beam includes two cryostats which house and cool four single-cell superconducting cavities for a total of 8 cells per beam. The cryostats operate at 4.5 K and each cavity supplies 1 MV at injection and up to 2 MV at maximum power. The RF energy in each cavity is supplied by a dedicated, 300 kW klystron located in the adjacent UX45 cavern. A technical layout of a prototype two-cell, superconducting RF cavity is show in Figure 3.3. [28–30]

During injection and ramp-up, these cavities supply the orbiting hadrons with kinetic energy by generating an electric potential gradient from which the charged hadrons continually 'fall.' One of the benefits of the LHC's circular architecture is that accelerated particles can interact with these components on numerous orbits around the ring allowing for steady and controlled ramp up to a specified energy and subsequent 'storage' when the desired energy is achieved. At full power, the integrated beam energy, for proton fills, is approximately 362

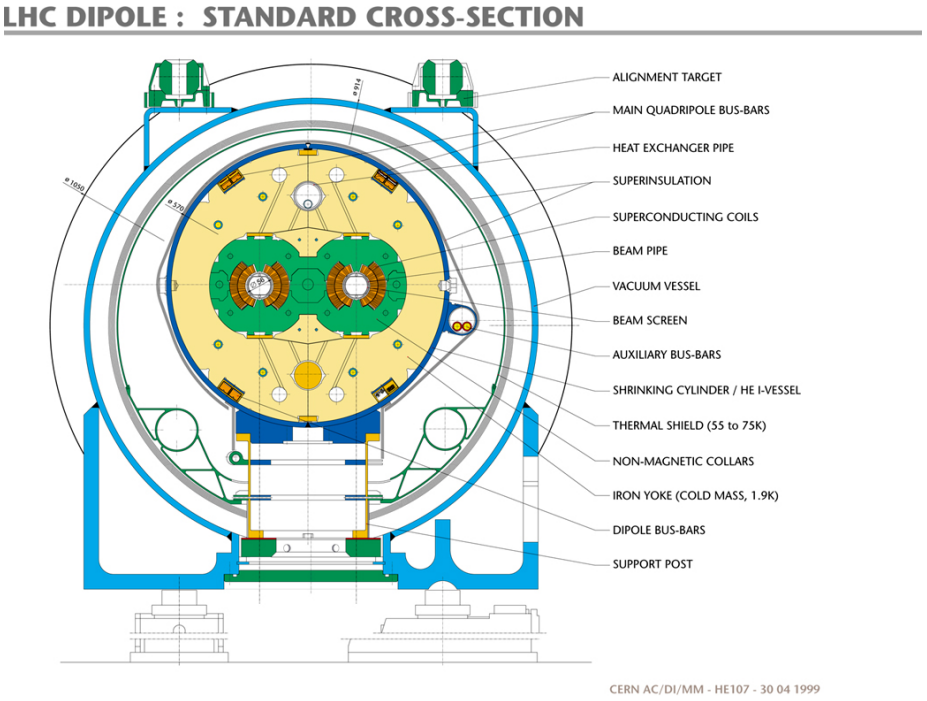


Figure 3.4: A technical cross-section of an LHC dipole magnet. As can be seen in this figure, the requisite cryostat and its associated services and safety mechanisms comprise the bulk of the observed magnet assembly volume. [31]

MJ, or alternatively 0.58 A. [28–30]

Circulation and Beam Alignment

The most crucial aspect of a particle collider's design is the magnet configuration. High quality magnets are important as extremely high and uniform magnetic fields are necessary to maintain the particles' circular trajectory. As the particle's energy and momentum increase, so too must the magnetic fields in order to maintain the appropriate orbit radius and to prevent the particles from colliding with the inside walls of the beam pipes. To accomplish this, the LHC utilizes 1232 dipole magnets arranged end to end along the 27 km circumference of the beam pipe. [29]

To generate the required field strengths, each 3.5×10^4 kg superconducting dipole magnet utilizes approximately 11 kA of electrical current at full design power, resulting in a 8.3 T field perpendicular to the beam's plane of curvature. To keep each magnet in its superconducting state at 1.9 K, liquid helium is utilized in each magnet-section's cryostat. These large cryostats, and their requisite services and safety mechanisms, comprise the bulk of the observed magnet assembly volume and mass, as can be seen in Figure 3.4. [29]

In addition to the dipole magnets used to maintain precise orbital alignment, quadrupole magnets are utilized to focus the beam immediately prior to collisions at the interaction points. This is crucial as the smaller the respective

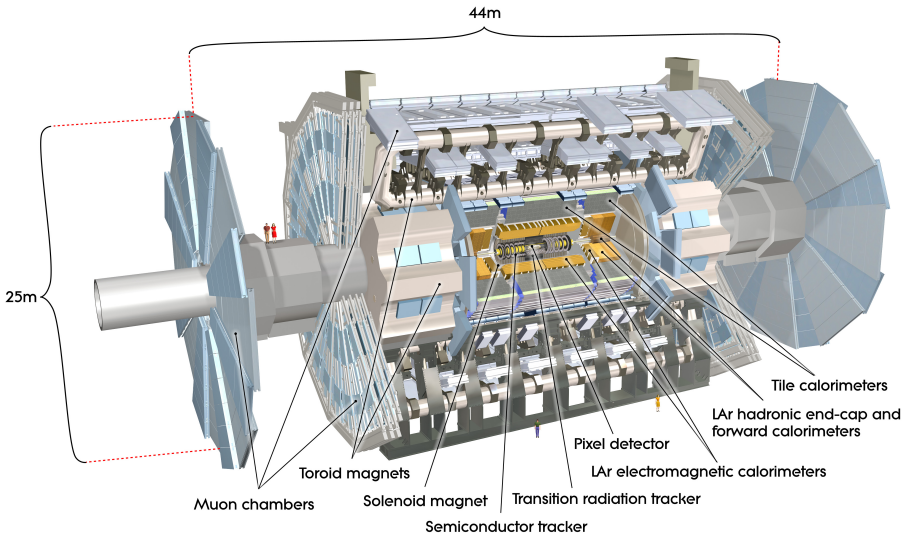


Figure 3.5: A rendered model of the ATLAS experimental apparatus with human facsimiles for scale and major subsystems identified. [24]

beam profiles, the larger the number of collisions per beam crossing, resulting in a higher instantaneous luminosity. In sum, the total amount of energy stored in the LHC's magnetic field is approximately 11 GJ. [29]

3.2 The ATLAS Experiment

3.2.1 Detector Architecture

Coordinate System

The ATLAS detector is roughly cylindrical and thus utilizes a natural coordinate system for this topology. Regarding common terminology when referencing the ATLAS architecture, the 'end-cap' refers to the circular end sections of the cylinder and the 'barrel' refers to the radially distributed components in-between the end-cap sections along the axis of symmetry.

The beam pipe runs along the cylindrical axis of symmetry with the interaction point midway between each end-cap. This interaction point defines the origin of the right-handed coordinate system used at ATLAS. In cartesian coordinates, the positive z axis runs along the beam pipe, the positive x axis points directly toward the center of the LHC ring, and the positive y axis points directly up. In spherical coordinates, the azimuthal angle ϕ and polar angle θ are defined as usual with the slight caveat that $\phi \in (-\pi, \pi]$, and follow the

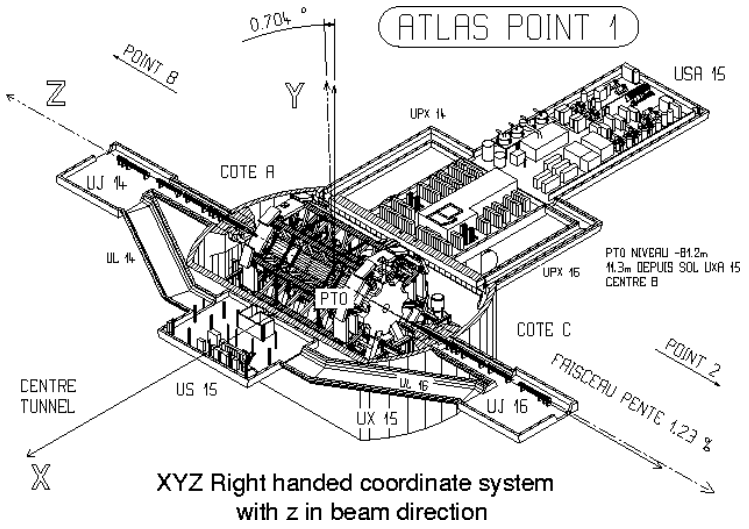


Figure 3.6: A technical schematic displaying the ATLAS right handed coordinate system. In this figure, the ATLAS detector is positioned in a cut-away of the LHC's Point 1 experimental cavern, approximately 100 meters underground. The structures perpendicular to the beam-line are caverns used to house miscellaneous service equipment from data-acquisition nodes to cryogenics. [24]

usual transformation rules shown in Eq. 3.1.

$$x = r\sin\theta\cos\phi, \qquad r = \sqrt{x^2 + y^2 + z^2} \tag{3.1a}$$

$$y = r\sin\theta\sin\phi, \qquad \theta = \arccos\left(\frac{z}{r}\right) \tag{3.1b}$$

$$z = r\cos\theta, \qquad \phi = 2\arctan\left(\frac{y}{\sqrt{x^2 + y^2} + x}\right) \tag{3.1c}$$

In the momentum the basis, the default representation resembles the cylindrical topology of the ATLAS detector. The radial component is called p_{T}, the transverse component of the momentum, ϕ defines the radial angle within the x-y plane just as before, and η defines the pseudo-rapidity, an asymptotic mapping to the azimuthal angle θ. This representation is compared to the Cartesian representation in Eq. 3.2.

$$p_x = p_{\mathrm{T}}\cos\phi, \qquad p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} \tag{3.2a}$$

$$p_y = p_{\mathrm{T}}\sin\phi, \qquad \eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \tag{3.2b}$$

$$p_z = p_{\mathrm{T}}\sinh\eta, \qquad \phi = 2\arctan\left(\frac{p_y}{\sqrt{p_x^2 + p_y^2} + p_x}\right) \tag{3.2c}$$

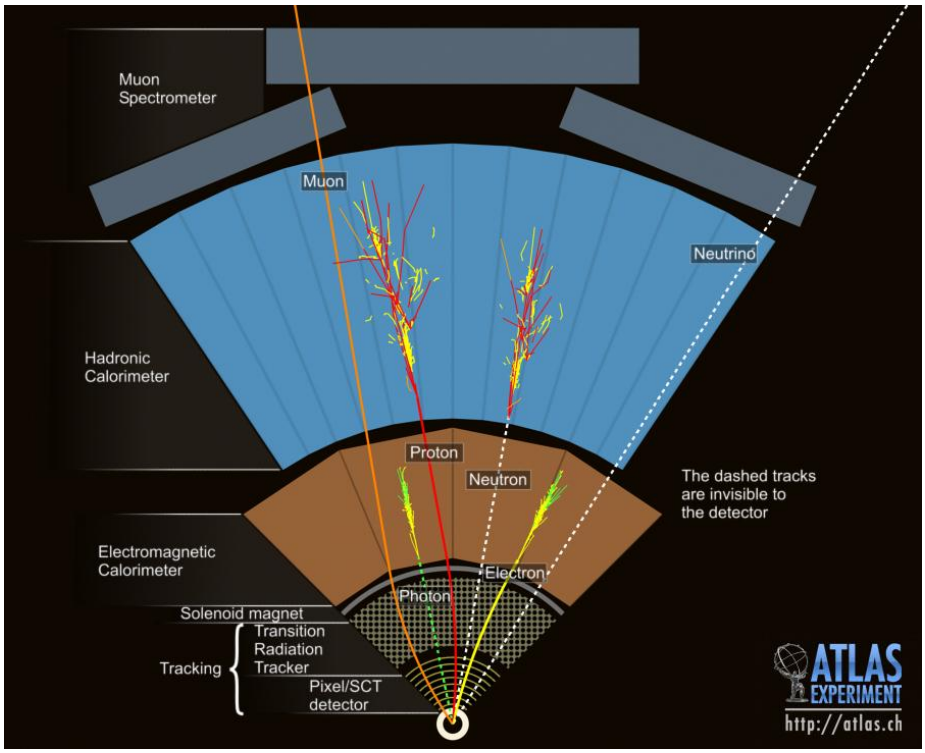


Figure 3.7: A cartoon depicting the various detection systems of the ATLAS detector in addition to the behavior of commonly observed signals from the hadron collisions. [32]

Put simply, the transverse components of ejected particle momenta, p_{T}, in collider experiments represent deep-inelastic effects of the parent collisions as the colliding hadrons have relatively little transverse momentum to contribute to final-state particles. Similarly, in the high energy limit, η very closely approximates the rapidity, y. Furthermore, the regions defined by large $|\eta|$ are called the *forward* regions where as the region defined by small $|\eta|$ is defined as the *central* region.

As shown in Figure 3.7, the cylindrical design of ATLAS allows for a nearly hermetic detection volume consisting of, in increasing distance from the interaction point: an inner station tasked with charged particle tracking and vertexing, an electromagnetic calorimeter for measurement and identification of electrons and photons, a hadronic calorimeter for measurement and identification of hadron-based particle jets, and finally a muon spectrometer for measurement and identification of high-p_{T} muons. The hermetic nature of the ATLAS detector, and of the EM and hadronic calorimeters in particular, further allow for the precise measurement of missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$. This signature is suggestive of the presence of the weakly interacting neutrinos which pass through ATLAS undetected while still carrying away energy. The definition and calculation of $E_{\mathrm{T}}^{\mathrm{miss}}$ is explored in more detail in Section 6.4. Finally, the inner detector and muon spectrometer elements exist in their respective magnetic fields to fa-

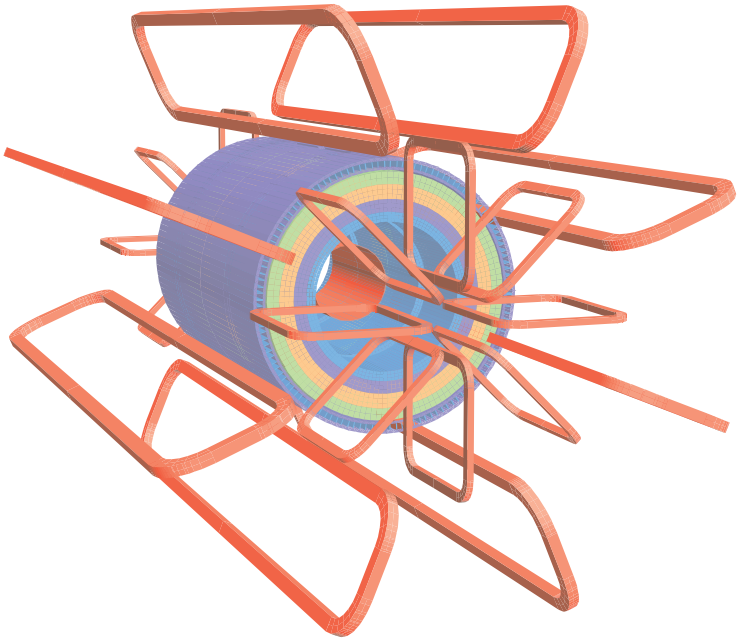


Figure 3.8: A rendered model of the ATLAS central solenoidal magnet surrounded by the 8-coil barrel and end-cap toroidal magnets. [24]

cilitate momentum measurements based on measured track curvatures due to the Lorentz force. Each of these elements will be discussed in further detail in the subsequent sections.

Magnet System

In the field of experimental particle physics, large magnets play a very fundamental role. In fact, much of the design and performance goals of an entire experiment revolve around the magnet design and topology. ATLAS is no exception. ATLAS utilizes two primary magnet systems. Each of these systems can be further sub-divided into their geometric components for a total of four superconducting magnets. The magnets were designed to cover a volume 22 m in diameter and 26 m long and store a total of 1.6 GJ of magnetic energy. A rendered model of the ATLAS magnet system can be seen in Figure 3.8.

The first system is the large ATLAS toroidal magnet, pictured in Figure 3.1, from which ATLAS gets its name. As it's name suggests, the system supports a large toroidal magnetic field in the region covered by the muon spectrometer. It is comprised of the barrel toroid, which gives ATLAS it's distinct appearance, and the two end-cap toroids at each end, which were the largest independent sub-assemblies of the ATLAS architecture. The toroid magnets provide between 1.5-5.5 Tm of bending power in the central region and approximately 1 to 7.5 Tm in the end caps.

The barrel toroid has an inner diameter of 9.4 m and an outer diameter of 20.1 m and is 25.3 m long and is the largest single magnet at ATLAS. It has

8 distinct coils and a total assembly mass of 830 tons. It takes five weeks to cool the cold mass of the barrel toroid down to an operating temperature of 4.6 K at which point it permits a nominal current of 20.5 kA, this corresponds to a 1.1 GJ magnetic field. In the event of a magnet quench, all of the magnetic field's 1.1 GJ of energy can be safely absorbed by the cold mass of the assembly bringing its temperature to, on average, 58 K. A restart can take place following the 50 hours required to bring the cold mass back to the nominal temperature of 4.6 K.

Compared to the barrel toroid, the end-cap toroids are much smaller both in size and in stored magnetic energy. Covering the end-cap sections of the ATLAS muon spectrometer, each end-cap torrid measures 5.0 m in length and 10.7 m in outer diameter and store 0.25 GJ of magnetic energy. Built in their dedicated cryostats on the surface and weighing in at 240 tons apiece, each end-cap toroid magnet had to be lowered into the ATLAS cavern via the primary access shaft in one of the most crucial operations during construction.

The second system is the central solenoid magnet which provides a 2 T axial magnetic field, parallel to the beam, in the volume containing the inner detector. In order to provide the calorimeters, situated outside the magnet, with the best possible resolution, the mass and material thickness of the magnet had to be kept to a minimum. With an inner diameter of 2.46 m, an outer diameter of 2.56 m, and a mass of only 5.7 tons, this was successfully achieved. In total, the central solenoid magnet contributes approximately 0.66 radiation lengths worth of material prior to the calorimeter at nominal incidence. To accomplish this, the magnet shares a vacuum vessel with the Liquid Argonne (LAr) calorimeter, thereby eliminating two vacuum walls. [24]

Inner Detector

The Inner Detector (ID) comprises the vast bulk of signal channels used at ATLAS. Situated very close to the beam pipe and thus the interaction point, the ID is tasked primarily with tracking and vertexing. In other words, the ID is supposed to identify the tracks of charged particles ejected from the collisions. These charged particle tracks are then used to identify the collisions points themselves, also known as the primary vertices. This poses a considerable challenge given the number of charged particles ejected per collision in addition to the ever increasing numbers of collisions per bunch crossing. In expectation of ever increasing number of primary vertices and their subsequent tracks given the high luminosity goals of the LHC, efforts were made to maximize the spatial granularity of the channels so as to resolve individual tracks and primary vertices in such an environment. To this end, the ID consists of three different detector subsystems: the pixel detector, the silicon microstrip layers (SCT) and the transition radiation tracker (TRT), each highlighted in Figure 3.10. Taken together, the components of the ID cover a cylindrical vol-

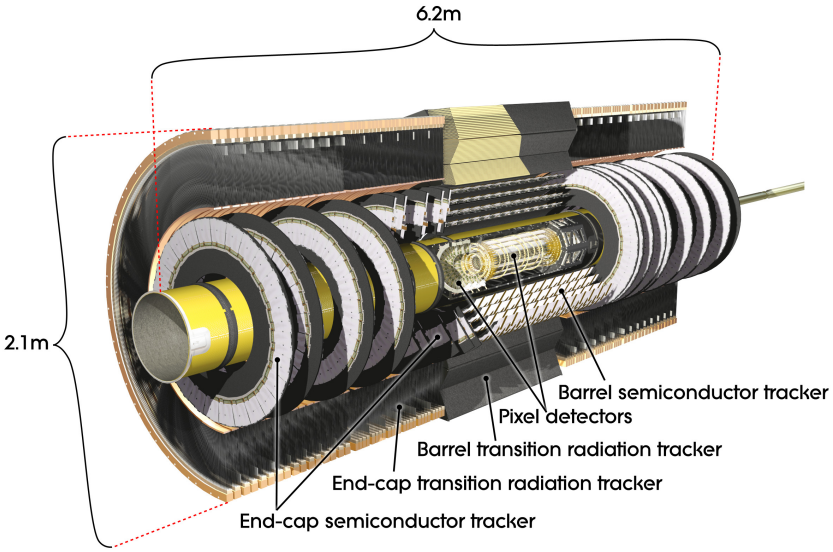


Figure 3.9: A rendered model of the ATLAS inner detector assembly.[24]

ume up to $|\eta| \leq 2.5$ and measure charge particles with a momentum resolution of $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} = 0.05\% p_{\mathrm{T}} \oplus 1\%$.

The innermost subsystem is the pixel detector which is composed of 3 layers which range from 5.1 to 12.3 cm away from the center of the beam pipe, or the nominal interaction point. It is a silicon based detection technology utilizing oxygenated n-type wafers with readout pixels on the n^+-doped side of the detector. All 1744 pixel sensors are identical and have a minimum granularity of 50×400 μm^2. In total, the pixel detector contributes 80.4 million readout channels.

Outside of the pixel system is the SCT detector. The SCT is composed of traditional silicon based detection layers with AC-coupled readout strips. The SCT contributes an additional 6.3 million channels of readout from four cylindrical barrel layers and two, 9 disk, end cap layers, with an intrinsic accuracy in the bending plane of 17 μm , as compared to 10 μm for the pixel detector or 130 μm for the TRT.

Finally, external to the SCT is the TRT. The TRT, unlike the pixel and SCT detectors, utilizes a gas volume as an ionization medium. The TRT is composed of polyimide straw tubes filled with a 70% Xe, 27% CO_2, 3% O_2 gas mixture. A 31 μm diameter, gold-plated, tungsten anode wire is strung inside the tube along the axis of symmetry. The conductive film lining the inner walls of the straw tubes is held at -1530 V with respect to ground and the anode wire. This allows for a gas amplification gain of 2.5×10^4. The TRT contributes 351 thousand channels, and covers a cylindrical volume from an inner radius of 554 mm from the beam interaction point to 1106 mm, the largest active volume of the ID.

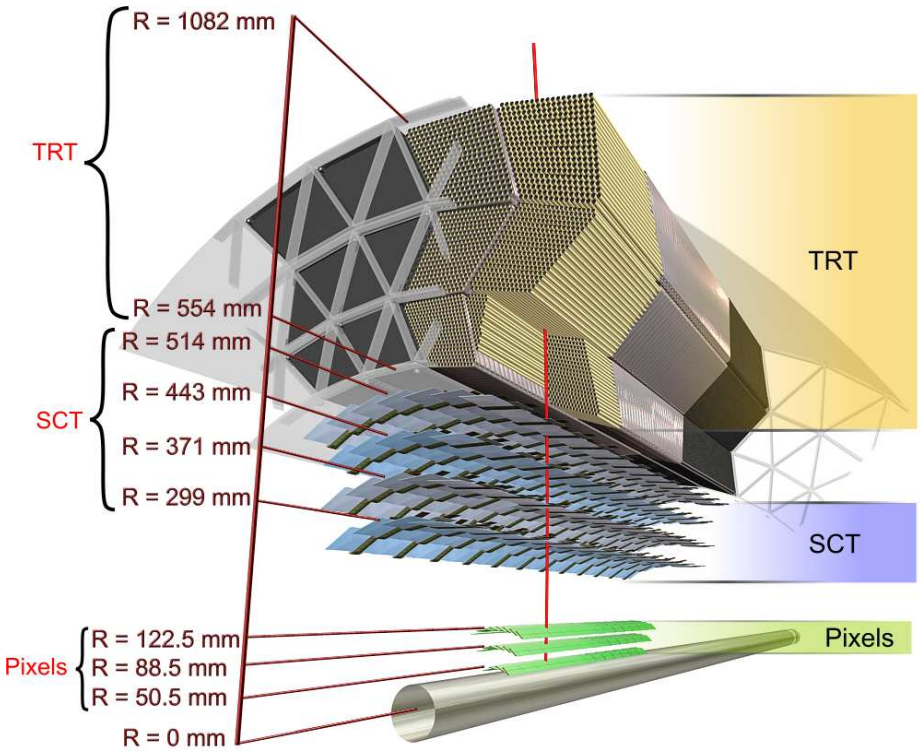


Figure 3.10: A rendered model of the inner detector component technologies.[24]

Calorimetry

Directly beyond the ID are the electromagnetic (EM) and hadronic calorimeters. These systems' primary responsibility is the detection, identification, and measurement of electrons, photons, jets, and missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$). To ensure good jet energy and $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution, as much of the energy produced in the collision must be measured by the calorimeters as possible. This motivates the hermetic design seen at ATLAS. A rendered model of the ATLAS calorimeters can be seen in Figure 3.11.

In general, calorimeters operate by the detection of energy deposits in some active medium, sandwiched between alternating layers of some passive medium. The active layers are typically comprised of some detection material such as a scintillation or ionization medium. The passive layers are typically comprised of some high Z material to encourage further interactions and thus energy capture by the calorimeter. The calorimeters at ATLAS operate under the same principle.

The EM calorimeter sits just beyond the solenoid magnet and is arranged into a barrel assembly ($|\eta| < 1.475$) and an end-cap assembly ($1.375 < |\eta| < 3.2$). It is at minimum 22 radiation-lengths in thickness in the barrel section and 24 radiation-lengths thick in the forward end-cap region. The EM calorimeter is comprised of a lead-liquid-argon (LAr) detector with accordion-shaped kapton electrodes and lead absorbers symmetric in ϕ. Having a granularity in η and

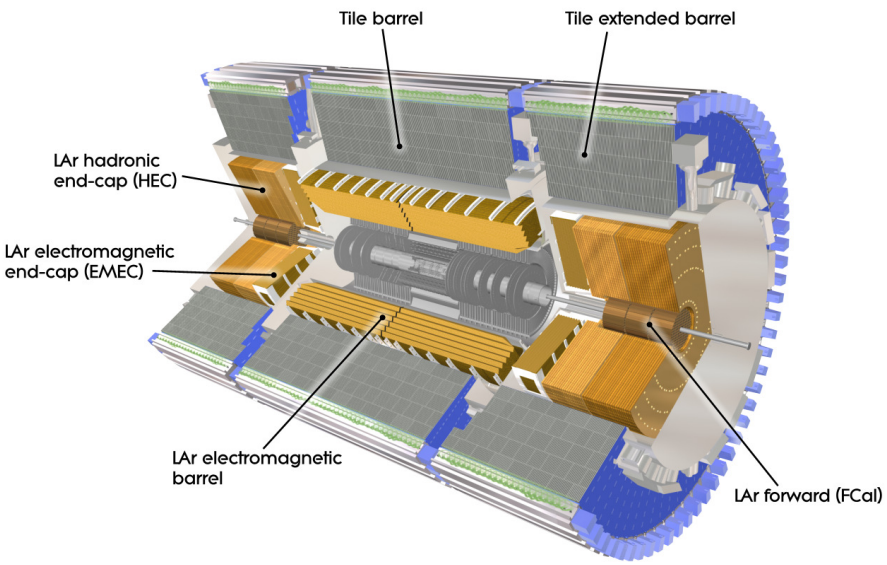


Figure 3.11: A rendered model of the composite ATLAS calorimetry subsystems.[24]

ϕ on the order of 0.025 x 0.025, the EM calorimeter contains approximately 64 thousand readout channels. The EM calorimeter is capable of measuring energy deposits to an accuracy of $\sigma_E/E = 10\%E/\sqrt{E} \oplus 0.7\%$.

The hadronic calorimeter is purposed with measuring the hadrons produced as prompt partons fragment following a hard interaction, and utilizes the same general detection mechanism of layered interaction and active media as the EM calorimeter. The hadronic calorimeter is composed of: the Tile Calorimeter, which utilizes alternating steel passive layers and scintillating tiles in the active layers; the LAr Hadronic End-Cap calorimeter, which uses the same lead-liquid-argon technology as the EM calorimeter; and the LAr Forward calorimeter which uses tungsten and copper instead of lead. In total, the hadronic calorimeter adds an additional 13 thousand channels of readout. The hadronic calorimeter achieves a nominal jet energy resolution of $\sigma_E/E = 50\%E/\sqrt{E} \oplus 3\%$ in the barrel and end-cap regions and $\sigma_E/E = 100\%E/\sqrt{E} \oplus 10\%$ in the forward regions.

Muon Spectrometer

Finally, the Muon Spectrometer (MS) extends from the calorimeters radially outwards 6 m. The MS is responsible for detecting and measuring high momentum muons utilizing a toroidal magnetic field and gaseous detector technologies. Subdivided into four sections, the two innermost in η form the barrel region, and the two outer most sections form the end-caps. The barrel region is comprised of 3 concentric stations of detector modules each consisting of an array of monitored drift tube (MDT) chambers. The second station of MDTs are sandwiched

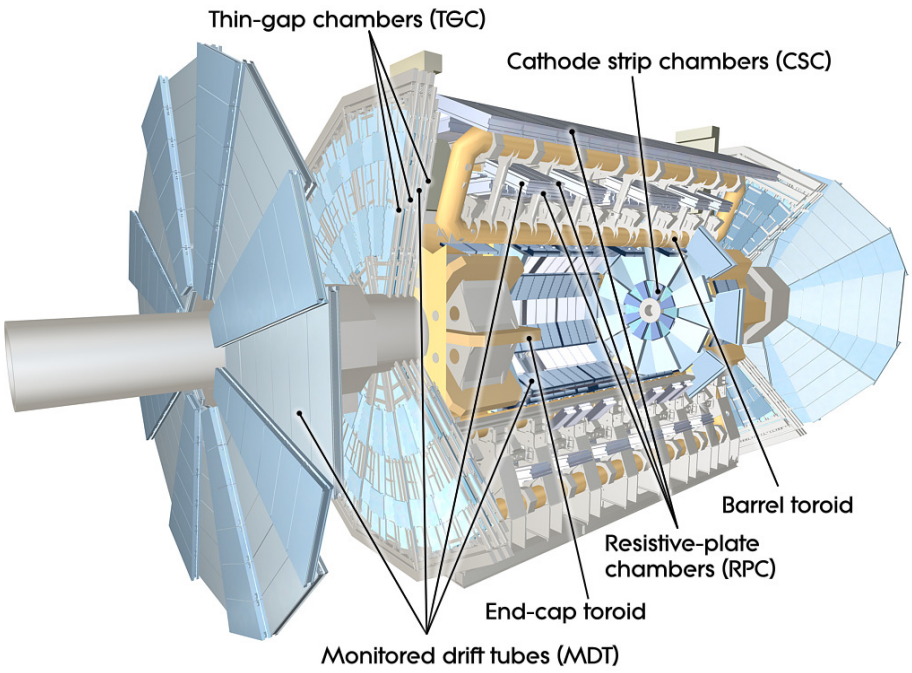


Figure 3.12: A rendered model of the ATLAS Muon Spectrometer.[24]

in-between resistive plate chambers (RPCs) used for the muon trigger. The end-caps also use MDTs, but cathode strip chambers (CSCs) are used in the more forward regions where thin gap chambers (TGCs) are used for triggering. Considering alignment effects, the MS approaches 10% momentum resolution at $p_{\mathrm{T}} = 1$ TeV. Everything combined, the MS provides more than 1 million channels of readout and a muon momentum resolution of $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} = 10\%$ at $p_{\mathrm{T}} = 1$ TeV [33, 34].

The MDTs are used for the precision muon momentum measurements and cover most of the MS acceptance. The implementation of the MDTs in their tracking and momentum measurement roles is demonstrated in Figure 3.13. The MDT utilizes tubes oriented perpendicular to the bending plane. These tubes, filled with $\mathrm{Ar/CO_2}$ (93%/7%) at 3 bar, are held at ground with respect to the 3080 V gold-plated tungsten-rhenium anode wire with a diameter of 50 μm . Incident ionizing particles liberate outer-shell elections in the gas, which themselves begin to ionize further gas molecules following their acceleration in the radial electric field, resulting in a measurable signal and an associated drift time between the trigger signal and the collection of the first elections at the anode. Each MDT layer consists of 2 multilayers which each consist of 3 layers of tubes. The average tube has a resolution of approximately 80 μm in the bending plane.

The CSCs perform the precision measurements in the far forward regions of the end-caps where $|\eta| > 2$. In this region, the hit rates exceed the safe operation thresholds of the MDTs ($150\mathrm{Hz/cm^2}$). The CSCs developed for this purpose

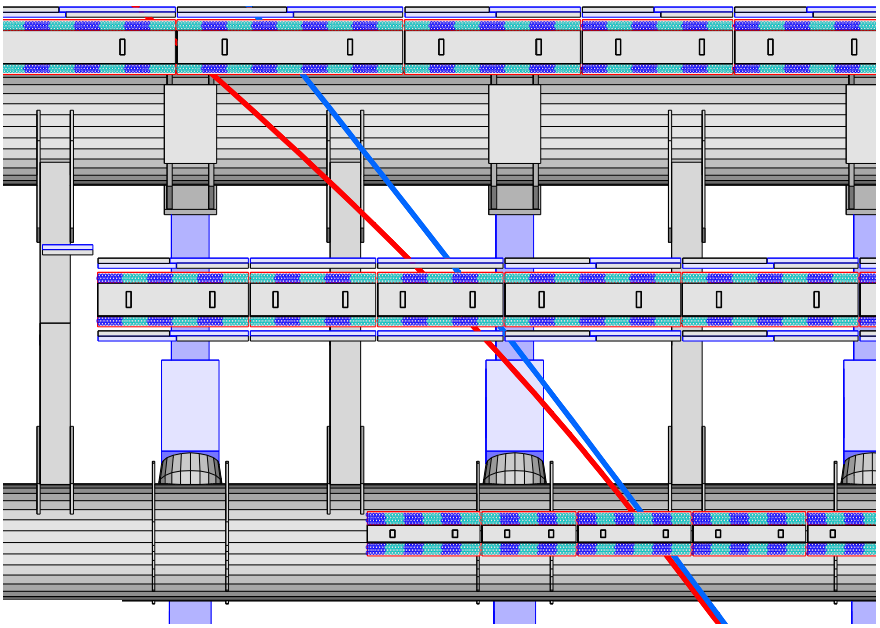


Figure 3.13: A rendered model of the ATLAS Muon Spectrometer (MS) with a 4 GeV(red) and 20 GeV(blue) track in the bending plane of the toroid magnet. Here the inner, middle, and outer layers of the MS. The light and dark alternating tube cross-sections show the granularity of each front-end electronic card known as a mezzanine card. [24]

are therefore expected to withstand rates up to 1000Hz/cm^2. Much like the MDTs, the CSCs similarly utilize a gaseous active medium, in this case, Ar/CO_2 (80%/20%), however, instead of independent gas volumes for each channel like the tubes of the MDT, the CSC chambers utilize a single volume per chamber layer. In each CSC chamber, there are 4 gas volumes each with 250 or 402 gold-plated tungsten-rhenium anode wires, depending on whether it is a large or small chamber. These wires are separated from each other and the readout plane by 2.5 mm. As their name suggests, the readout planes are comprised of cathode strips which run perpendicular to the anode wires and perpendicular to the bending plane. Due to the high multiplicity in the forward regions, the granularity of the CSC is much higher than the MDTs resulting in a resolution in the vicinity of 60 μm in the bending plane.

For triggering RPCs are used to complement the MDTs in the barrel region, while TGCs are used exclusively in the forward end-cap regions due to their superior rate capability. These technologies sacrifice momentum resolution in exchange for excellent timing resolution and rate capability. The RPCs, unlike the other technologies use no wires and instead utilize resistive plates to hold a potential difference over an active gas volume. The gas volume and electric field acts as an amplifier as in the previous technologies and read-out is accomplished via a capacitive coupling between readout strips outside the active volume and a charge-depleted region on the resistive plates. These charge-depleted regions develop as liberated ions and electrons interact with the plates. This process achieves an acceptable timing resolution of 1.5 ns, and a dead time on the order of 3 ns. The TGCs on the other hand operate on the same principal as the CSCs, however, to optimize the timing and rate capability, the TGCs utilize a much smaller gas volume of 1.4 mm from anode wire to the readout planes and a smaller wire pitch of 1.8 mm. This geometry, coupled with the high electric field around the anode wires (2900 V), allow for acceptable rates and time resolution, 4 ns as compared to the CSC with 7 ns.

3.2.2 Triggering, Acquisition, and the ATLAS Data Model

Detection to Event Reconstruction

As described in Section 3.1.1 the high instantaneous luminosity provided by the LHC allows for the sooner observation of increasingly rare processes. However, this environment imposes formidable challenges for data acquisition efforts. As shown in Table 3.1, a nominal bunch crossing rate of 40 MHz results in an average of 30 proton-proton collisions every 25 ns. The ATLAS trigger and data acquisition system must then identify those events which are deemed to be useful for physics analysis at such a rate so as to saturate the available 200 event/s IO bandwidth. If at any stage a selection requirement is not fulfilled, the event is vetoed and all the event's data is purged from any and all buffers and

Data Acquisition Rates

Operation Stages	**Rate** [Hz]	**Processing Time** [seconds]
Bunch Crossing	4.0×10^{7}	2.5×10^{-8} †
L1 Trigger	7.5×10^{4}	2.5×10^{-6}
L2 Trigger	3.5×10^{3}	4.0×10^{-2}
Event Filter	2.0×10^{2}	4.0×10^{0}

Table 3.1: Elements of the ATLAS data acquisition system and their respective rates and processing times. † Proton-proton collisions occur up to once every 25 ns, giving context to the further processing windows. For instance, in the maximum time allowed to acquire a L1 trigger decision, 100 subsequent bunch crossings have taken place. [35]

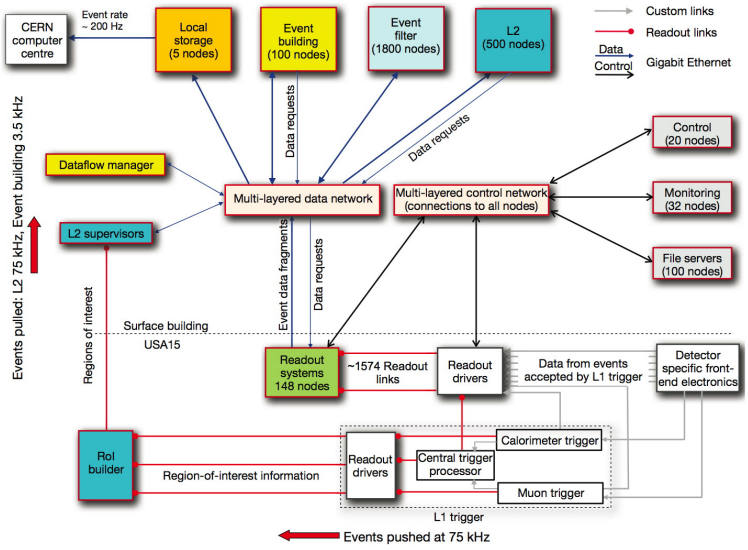


Figure 3.14: A flow control diagram of the ATLAS data from the front-end electronics, through the trigger processors and data acquisition system, to local storage at the CERN computing center. [24]

memory, forever lost. It is thus imperative to save the CPU and IO-expensive operations for those events that fulfill the tightest selection criteria, eliminating the events failing the loosest cuts first. This process is accomplished in thee stages as described below.

First, the level 1 (L1) trigger searches for the primary signatures of interesting events. These signatures are typically based on the detection of high-p_{T} objects such as muons, elections, photons, jets and τ-leptons, or a predefined multiplicity of those objects. The L1 trigger is hardware-based and uses coarse grained information from the calorimeter and muon trigger chambers to define local regions of interest (ROIs) which define the geometric locations of objects which caused a given trigger to fire. As shown in Figure 3.14, data generated by the L1 trigger is passed to the Central Trigger Processor (CTP) which makes the overall L1 accept decision for a given event. Following an accept decision, the accumulated ROIs are sent downstream to the level 2 trigger. The process from L1 to L2 is quite challenging given the 25 ns windows between bunch crossings. In fact, the time of flight for a relativistic muon from collision point to the outer layers of the muon spectrometer is frequently longer than 25 ns. As a result, trigger information is stored in pipeline memories in custom electronics on or near the trigger detectors. Concerns regarding the cost and complexity of the electronics, in addition to the reliability of the stored data in a hard radiation environment motivate the use of the shortest pipeline possible. This, in turn, imposes a constraint on the L1 trigger decision latency of 2.5 μs. Both the calorimeter L1 and muon L1 triggers comfortably meet this requirement with 2.1 μs response times.

The CTP receives data from the calorimeter and muon trigger chambers upon which it bases its trigger accept decision. This decision is determined by the fulfillment of a pre-defined trigger condition. These conditions can be combined for up to 256 distinct trigger items which make up the trigger menu for a given operating period. Finally, for each trigger item, information regarding the item's priority, a mask, and a pre-scale factor are stored. These parameters allow the CTP to prioritize certain signatures over others during a given run. However, as these parameters manipulate the trigger acceptance, they are logged every few minutes corresponding to discrete LuminosityBlocks (LBs). These parameters are then pulled from a database so as to accurately determine the recorded luminosity.

The level 2 trigger (L2) accepts ROIs from the L1 trigger following the associated event's acceptance by the L1 CTP. The ROIs, containing full detector granularity, are transmitted along dedicated networking links to the L2 Supervisor which coordinates the computation of the L2 trigger decision. The full event information is concurrently aggregated and transferred by the read-out system (ROS) in preparation for the L2 trigger decision. The ROIs are transferred from the L1 trigger at approximately 75 kHz while full events are transferred at roughly 3 kHz. The L2 trigger is software based, and runs above ground at

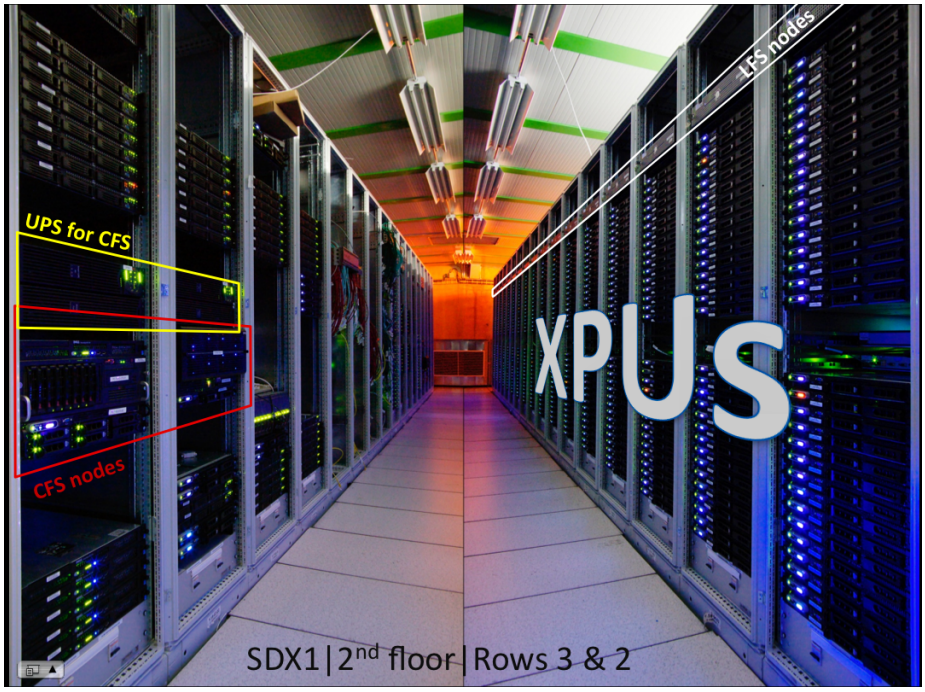


Figure 3.15: Rows 3 and 2 on the second floor of SDX1, the on-site computing facility for ATLAS. On the left are elements of the Central File Servers (CFS) and on the right are processing units of the High Level Trigger (HLT) farm. [36]

on-site compute facility, SDX1, on approximately 500 compute nodes and is the first component of the high-level trigger (HLT) shown in Figure 3.15.

Finally, the Event Filter (EF), the final component of the HLT, applies further object selection utilizing the full ATLAS event reconstruction and analysis application. It runs, as shown in Figure 3.14, on roughly 1600 compute nodes at SDX1. Just prior to the EF, the Event Builder takes the byte-code stream from events passing the L2 trigger and assembles the RAW event structure and passes it off to the EF. If an event passes a given selection encoded in the EF trigger menu, the event is tagged and appended to the appropriate physics stream, and stored as RAW data files at the local storage sub-farm outputs (SFOs) to await periodic transfer to the CERN computing center. Events are finally stored at a rate of 200 Hz as shown in Table 3.1. This results in an event-rate reduction of over 7 orders of magnitude.

ATLAS Data Preparation

With the data safely stored at CERN's Computing Facility, the work starts to prepare the data for analyzers. This effort revolves around two principle efforts. First, to identify the precise conditions of the collision source when the data was taken. This takes into account the number of collisions per bunch crossing, the beam spot location, the beam profiles, and ultimately, the exact integrated luminosity per LuminocityBlock (LB). The second effort seeks to determine the

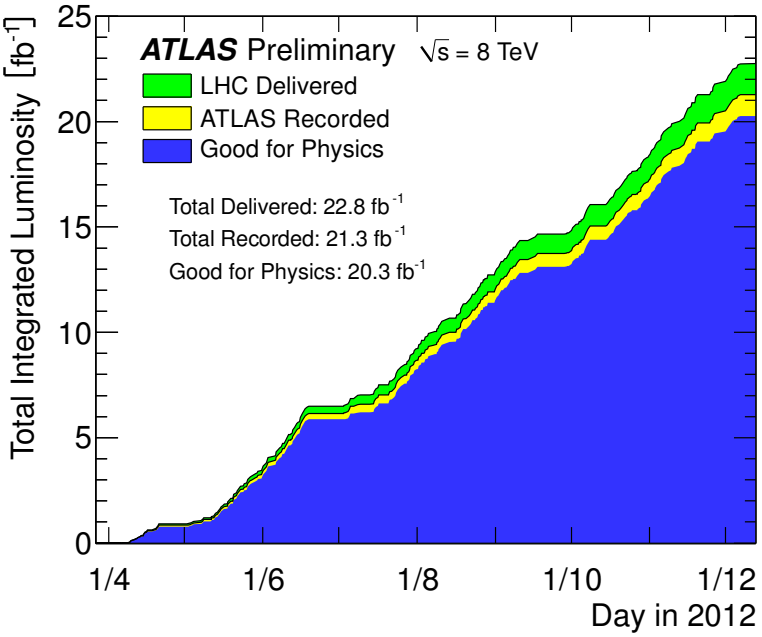


Figure 3.16: "Cumulative luminosity versus time delivered to (green) [and] recorded by ATLAS (yellow), and certified to be good quality data (blue) during stable beams and for [proton-proton] collisions at 8 TeV centre-of-mass energy in 2012. The delivered luminosity accounts for the luminosity delivered from the start of stable beams until the LHC requests ATLAS to put the detector in a safe standby mode to allow a beam dump or beam studies. The recorded luminosity reflects the DAQ inefficiency, as well as the inefficiency of the so-called 'warm start': when the stable beam flag is raised, the tracking detectors undergo a ramp of the high-voltage and, for the pixel system, turning on the preamplifiers. The data quality assessment shown corresponds to the All Good efficiency shown in the 2012 DQ table. The luminosity shown represents the preliminary 8 TeV luminosity calibration. Data quality has been assessed after reprocessing." [37]

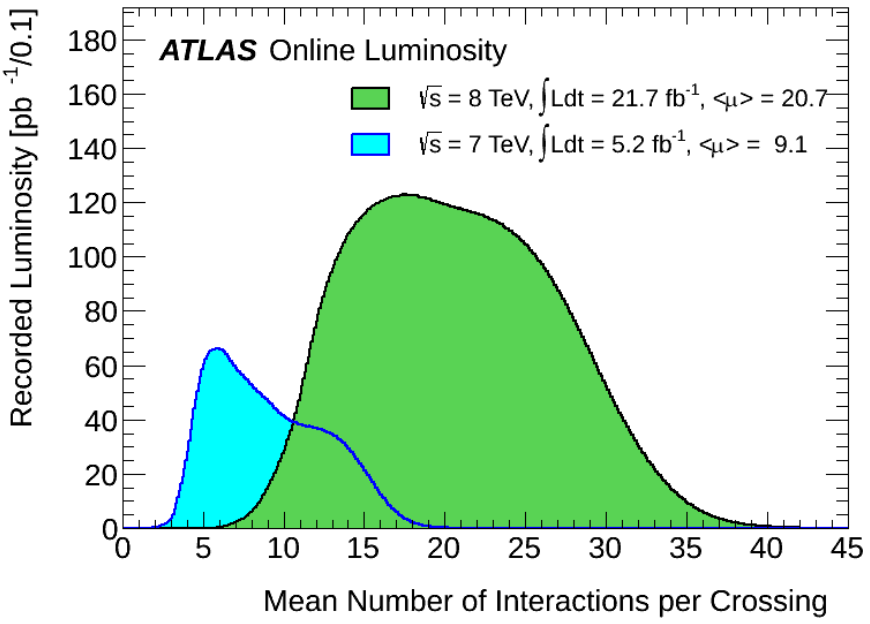


Figure 3.17: "The luminosity-weighted distribution of the mean number of interactions per crossing for the 2011 and 2012 data. This shows the full 2011 and 2012 *pp* runs. The integrated luminosities and the mean mu values are given in the figure. The mean number of interactions per crossing corresponds the mean of the poisson distribution on the number of interactions per crossing calculated for each bunch."[37]

precise condition of the ATLAS detector when the data was taken. This takes into account any losses in acceptance or detector coverage due to defects or malfunctions in detector operations, in addition to the settings of the trigger menu and pre-scales. On longer scales, the ATLAS geometry and alignment can be updated, periodically reducing the measured resolution of the affected objects. Estimates for all these contributing factors are continually updated during the data-taking period and distributed to analyzers in the form of Good-Run-Lists, integrated luminosity calculations, and per-LB pile-up (the number of additional collisions per event) conditions. Examples of this information is shown in Figure 3.16, where the cumulative integrated luminosity is shown as a function of data-taking days, and in Figure 3.17, where a luminosity-weighted distribution for the number of interactions-per-crossing is shown.

Finally, the data is distributed to various computing facilities all over the globe for distributed analysis. To accomplish this efficiently, derived data formats are used by which the RAW data is transformed into more useful formats. Some data, unneeded by some analyses is removed, some derived variables are added, and at its final stage, the data is takes the form of ROOT n-tuples, a common format for data-analysis in high-energy physics. Examples of some of these formats are shown in Table 3.2 along with their per-event data-size in megabytes per event (MB/event). The distributed nature of the data takes advantage of a number of global computing resources for storage and analysis, and brings us to the implementation of the Worldwide LHC Computing Grid (WLCG).

3.2.3 Worldwide LHC Computing Grid

The Worldwide LHC Computing Grid (WLCG) provides the necessary computing resources to store, distribute and analyze the physics data generated by the various LHC experiments. It is organized into four levels of abstraction or "Tiers" 0 through 3. Each tier defines a class of computing facility and its primary role in the data processing and analysis infrastructure.

For ATLAS, the Tier 0 facility is the CERN Data Center to which the raw ATLAS data is transmitted and stored. This particular tier is common to all the LHC experiments but it provides less than 20% of the total computing capacity of the WLCG. Apart from storing the raw LHC data, CERN's Tier-0 facilities also provide sufficient computing resources to fully reconstruct the RAW data from detector signals into physics objects, such as muons, elections and jets. Finally, the Tier-0 facility transmits copies of all RAW and reconstructed data to the Tier-1 sites for further processing and safe keeping.

ATLAS has approximately 10 Tier-1 sites located in member states across the globe, typically at national laboratories. Each Tier-1 facility is linked to CERN's Tier-0 and other Tier-1 facilities via the LHC Optical Private Network, a dedicated 10 Gb/s optical network. These sites collectively store a copy of all RAW ATLAS data and provide a considerable amount of comput-

IO Event Sizes

Data Format	**Size** [MB/event]
Raw Data	1.6
Simulated Raw Data	2.0
ESD	0.5
Simulated ESD	0.5
AOD	0.1
DESD-NTUP	$\lesssim 0.1$

Table 3.2: Commonly used data formats and their respective sizes per event. Here, the 'Simulated' qualifier denotes events which have been generated by a typical Monte-Carlo generator, the output of which has been sent through the entire ATLAS simulation and reconstruction chain. RAW data is unmanipulated byte-stream data from the ATLAS data acquisition system. Event Summary Data (ESD) refer to events in an object-oriented representation which has been reconstructed and subsequently striped of some unnecessary RAW data. Analysis Object Data (AOD) refer to formats which further remove unnecessary data leaving only the commonly utilized analysis objects. Finally, DESD-NTUP refers to event stored as a series of branches in a ROOT n-tuple tree suitable for analysis using common high-energy physics analysis software. [35]

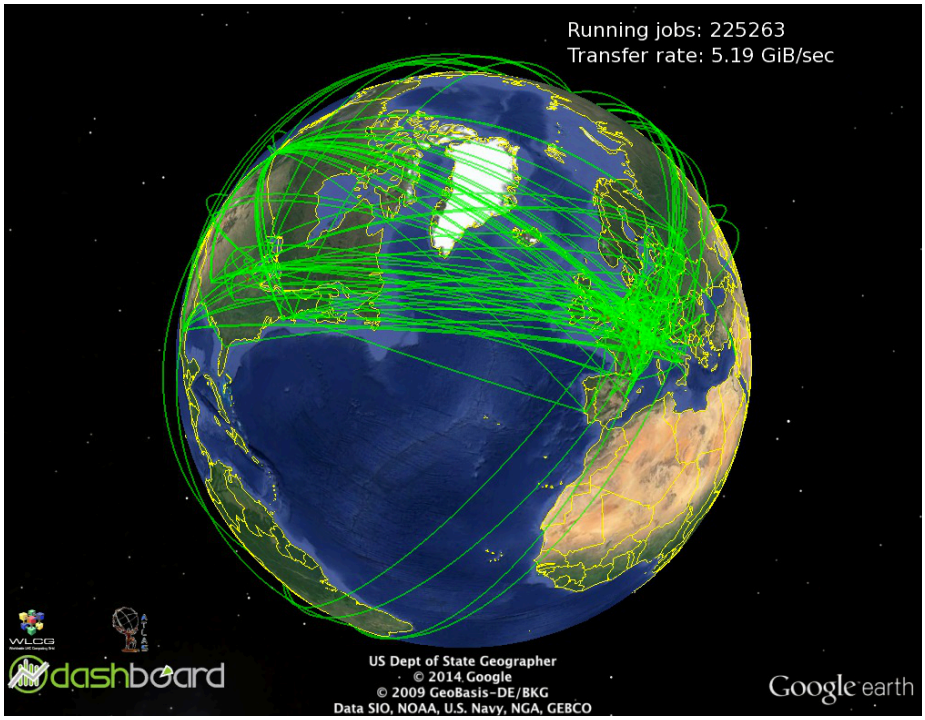


Figure 3.18: A rendered model of Earth with the current data site-to-site transfers represented as green arcs from source to destination as of 1200 GMT, Tuesday, 14th of October 2014. In addition, the current number of running jobs is shown in the top right corner as well as the total data transfer throughput.

ing resources dedicated to calibrating and reprocessing this data. They also store the simulated data used for analysis at downstream tiers. Further, each Teir-1 provides the necessary support to maintain round-the-clock operations of the WLCG. Finally, the Tier-1 facilities act as hubs to their respective Teir-2 facilities, thus forming a primary artery for the WLCG computing cloud.

Tier-2 facilities are typically associated with member universities or computing facilities that don't participate as a Tier-1 which can provide sufficient computing, storage, and support recourses for given analyses tasks. Each Tier-2 is connected to one Tier-1 facility via a dedicated networking channel for data-transfer and support needs. These facilities provide the necessary compute nodes to run the numerous analyses underway for the LHC experiments. The bulk of the analysis computation occurs at the Tier-2 facilities, though a portion of working nodes are appropriated for analysis tasks at the Tier-1 and Tier-0 facilities.

Finally, scientists are free to connect to the Grid via their local machines or department clusters. These are known as Tier-3 resources. Machines under this designation typically provide storage and processing in service to the specific analysis goals of the given institution.

Chapter 4

The Search for $t\bar{t}H$ at $m_H = 125$ GeV

To identify and measure a possible $t\bar{t}H$ signal predicted by the SM, the $t\bar{t}H$ rate is measured assuming a Higgs boson mass of 125 GeV, consistent with direct Higgs boson observations in other channels at the ATLAS [38] and CMS [39] experiments. This chapter will first present a generic analysis method for identifying some signal process, α, amidst a composite SM background. This generic analysis will then be applied to the search for the $t\bar{t}H$ process in particular, providing a roadmap for subsequent chapters.

4.1 Null and Signal Hypothesis Definition

An experiment designed to search for the hypothesized existence of some process ($pp \to X_\alpha$), which we will name α, at a hadron collider must first define an observable and an associated set of expectations consistent with the presence or absence of the process in question.

The observable common to nearly all process *searches* in collider experiments is the observed cross-section, $\sigma_\alpha(pp \to X_\alpha)$. This observable is proportional to the number of α-*events* produced per pp collision, also referred to as an event rate. Thus, each pp collision that results in a primary vertex and a collection of observed objects, X_α, associated with that vertex is considered to be an α-event.

For this process, σ_α is measured by summing the number of observed X_α events produced from a known number of pp collisions. To this end, the number of observed events determined to have originated from the α process, N_α, is divided by the integrated luminosity, $\mathcal{L} = \int L \, \mathrm{d}t$, of the dataset in question, provided by the relationship shown in Eq. 4.1 and Eq. 4.2, where $\mathcal{L}$ is the time-integrated luminosity:

$$N_\alpha = \sigma_\alpha \int L \, \mathrm{d}t \tag{4.1}$$

$$\sigma_\alpha = \frac{\alpha}{\mathcal{L}}. \tag{4.2}$$

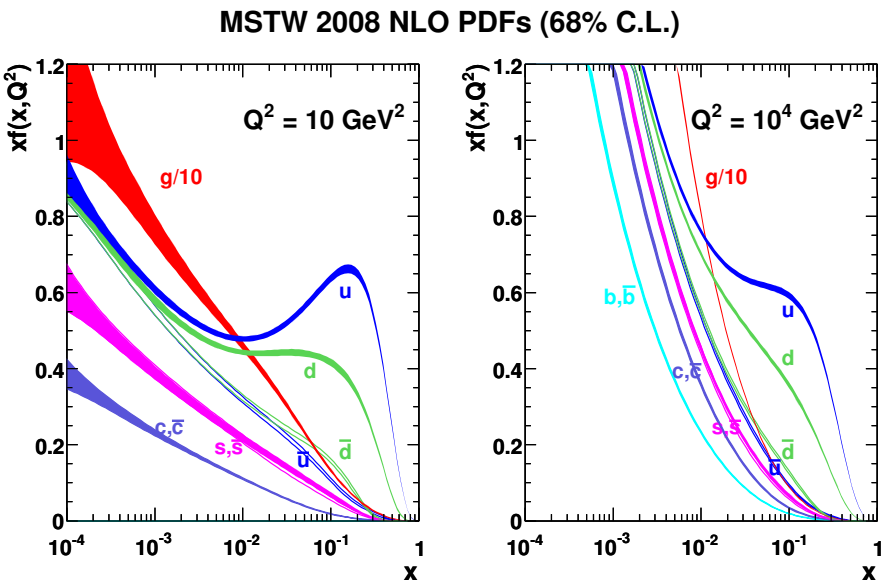


Figure 4.1: The MSTW 2009 NLO PDF sets using two Q^2 configurations of 10 GeV^2 and 10^4 GeV^2. [40, 41]

4.1.1 Parton Distribution Functions

As our hypothetical collider collides protons, the theoretical production cross-section must take into account the composite nature of these protons. An additional factor is therefore applied which effectively describes the interacting partons of the initial-state as constituents of their parent proton with some fraction of that proton's momentum, x_f. Further, the flavors of these initial-state partons are summed over, thus capturing all available configurations contributing to α production.

The method to include parton information is described by Eq. 4.3. It utilizes a set of parton distribution functions (PDFs), $f_a\left(x_a\right)$ and $f_b\left(x_b\right)$, for each initial-state parton, a and b. The PDFs, shown here, are parameterized at a momentum scale, Q^2, which is characteristic of the hard process $\sigma_\alpha\left(ab \to X_\alpha\right)$. For example, the MSTW 2009 NLO PDF sets [40, 41] using two Q^2 configurations of 10 GeV^2 and 10^4 GeV^2 are shown in Figure 4.1.

$$\sigma_\alpha\left(pp \to X_\alpha\right) = \sum_{ab} \iint \mathrm{d}x_a \mathrm{d}x_b \; f_a\left(x_a, Q^2\right) f_b\left(x_b, Q^2\right) \sigma_\alpha\left(ab \to X_\alpha\right) \quad (4.3)$$

4.1.2 Standard Model Backgrounds

An additional factor in determining σ_α is the presence of background processes that mimic the signature of the α process by contributing events which appear similar to the X_α events. These events are referred to as *background* events, in contrast to *signal* events which do originate from the α process.

To allow for the presence of background events, Eq. 4.2 is transformed into Eq. 4.4, given Eq. 4.5. Here N^{obs} is the number of observed events of a given final-state and N^{α} and N^{bkg} represent the expected events from *signal* and *background* processes respectively:

$$\sigma^{\text{obs}} = \frac{N^{\text{obs}}}{\mathcal{L}} \tag{4.4}$$

$$N^{\text{obs}} = N^{\alpha} + N^{\text{bkg}}. \tag{4.5}$$

Recomputing Eq. 4.4 and Eq. 4.5 in terms of our original observable then gives us Eq. 4.6 and Eq. 4.7:

$$N^{\alpha} = N^{\text{obs}} - N^{\text{bkg}} \tag{4.6}$$

$$\sigma_{\alpha} = \frac{N^{\text{obs}} - N^{\text{bkg}}}{\mathcal{L}}. \tag{4.7}$$

Depending on the nature of the background contamination for a given final-state, N^{bkg} can either be directly measured from data in designated control regions or it can be modeled using theory and Monte-Carlo simulation techniques. It is important to note that these same simulation techniques can be used to estimate N^{α} as well.

4.1.3 Selection and Acceptance Effects

Looking at Eq. 4.7, we see that uncertainty on σ_{α} is dependent on the uncertainties inherent to the measurement of the inclusive number of events, N^{obs}, as well as on our understanding of the background contribution, σ^{bkg} and the integrated luminosity $\mathcal{L}$.

These uncertainties are subsequently divided into two categories, statistical and systematic uncertainties. Statistical uncertainties originate from the stochastic nature of quantum interactions and the finite amount of data under consideration. Systematic uncertainties typically originate from uncertainties in the theoretical models used and uncertainties in the detector response and its simulation.

As our goal is to identify and characterize the presence of the α process, we are incentivized to configure our experiment such that the uncertainties are minimized and the statistical difference between the *signal plus background* hypothesis and the null, *background only*, hypothesis is maximized. As the statistical uncertainties of the contributions are *Poisson* distributed, this corresponds to maximizing Eq. 4.8:

$$f\left(N^{\alpha}, N^{\text{bkg}}\right) = \frac{N^{\alpha}}{\sqrt{N^{\text{obs}}}}. \tag{4.8}$$

It is important to keep in mind, however, that this treatment does not include uncertainties related to systematic affects present in the analysis which

could easily shift the optimal signal and background contributions away from the maximum value of Eq. 4.8. To constrain the remaining systematic uncertainties, control regions are defined which attempt to isolate and characterize these uncertainties. Minimizing both statistical and systematic uncertainties ultimately increase the sensitivity of an analysis.

This process corresponds to identifying and *selecting* final-state topologies for which the measurement of N^{α} has the least amount of total uncertainty. This might correspond to selecting final-states with the highest relative $N^{\alpha}/N^{\text{bkg}}$ or it might correspond to selecting the best measured final-states with the least systematic uncertainty. In any case, a *selection* criteria is defined that an event must satisfy to be considered a part of a measurement region. As such, this event selection imposes some acceptance, A_S, based on the fiducial regions of the detector geometry, and some efficiency, ϵ_S, unique to a given source of events. For example, suppose we have two sources of backgrounds, the first, A, is effectively identical to the signal, and the second, B, only passes the selection when it is mis-reconstructed, described by Eq. 4.9. Eq. 4.11 shows the application of the event selection, S, on the signal process by transforming the inclusive cross-section, $\sigma_{\alpha}^{\text{inc}}$, into a measured cross-section, $\sigma_{\alpha}^{\text{meas}}$. Thus, the measured result is then defined by transforming Eq. 4.7 into Eq. 4.11:

$$\epsilon_S^{\alpha} = \epsilon_S^{\text{A}} > \epsilon_S^{\text{B}} \tag{4.9}$$

$$\sigma_{\alpha}^{\text{meas}} = A_S \, \epsilon_S^{\alpha} \, \sigma_{\alpha}^{\text{inc}} \tag{4.10}$$

$$\sigma_{\alpha}^{\text{meas}} = \frac{N^{\text{obs}}}{\mathcal{L}} - A_S \cdot \left(\epsilon_S^{\text{A}} \sigma^{\text{A}} + \epsilon_S^{\text{B}} \sigma^{\text{B}} \right). \tag{4.11}$$

Background sources like, A, in the above example, which have identical final-states to the desired signal, are referred to as *irreducible backgrounds* as simple selection criteria are unable to improve Eq. 4.8. In contrast, backgrounds which can be reduced by defining an appropriate selection, S, are known as *reducible backgrounds*.

These background contributions can be estimated by numerous methods. Frequently, backgrounds are estimated by using theoretical models of the background processes in question to generate events and then simulating the resulting detector response to the generated events. These backgrounds, generally referred to as Monte-Carlo backgrounds, carry with them the inherent uncertainties attributed the models used. In other cases, backgrounds are estimated by measuring control regions which are known to provide reasonable facsimiles of the backgrounds present in the measurement region. A common example of this is the side-band-subtraction technique, where the regions adjacent to a signal peak, the sidebands, are used to estimate the number background events in the peak itself.

4.1.4 Observable-Based Hypotheses

Thus, we reach the final form of our measurement taking into account the following elements. First, the expected signal contribution, $\sigma_\alpha\left(pp \to X_\alpha\right)$, and its selection efficiency, ϵ_S^α, are estimated, typically using a particular theoretical model. The number of expected events from both reducible and irreducible backgrounds, N^{bkg}, are estimated using Monte-Carlo techniques or through control region measurements. The geometrical acceptance, A, of the detection apparatus is defined. Finally, the integrated luminosity of the dataset, $\mathcal{L}$, is estimated. Taking each of these pieces, we can then define the measurement and the competing hypotheses. The measurement of $\sigma_\alpha\left(pp \to X_\alpha\right)$ is shown in Eq. 4.12:

$$\sigma_\alpha\left(pp \to X_\alpha\right) = \frac{N^{\text{obs}} - N^{\text{bkg}}}{A_S^\alpha \epsilon_S^\alpha \cdot \mathcal{L}} \tag{4.12}$$

4.1.5 Application to the $t\bar{t}H$ Search

Starting from the previous section, it is fairly straightforward to apply the methods introduced to the search for the $t\bar{t}H$ process by exchanging α and $t\bar{t}H$. However, as the $t\bar{t}H$ process is predicted by the standard model we can go one step further and define a signal strength parameter, μ, which, as the name suggests, characterizes the results in terms of the SM prediction. This transforms Eq. 4.12 into Eq. 4.13, and results in a convenient definition of the signal hypothesis, $\mu = 1$, the null hypothesis, $\mu = 0$, and a parameter estimation hypothesis that finds the most probable value for the signal strength, $\mu = \hat{\mu}$:

$$\mu = \frac{N^{\text{obs}} - N^{\text{bkgs}}}{A_S^{t\bar{t}H}\, \epsilon_S^{t\bar{t}H} \cdot \mathcal{L}} \cdot \frac{1}{\sigma_{t\bar{t}H}^{SM}} \tag{4.13}$$

$$\mu = \frac{\sigma_{t\bar{t}H}^{meas}}{\sigma_{t\bar{t}H}^{SM}} \tag{4.14}$$

With these competing hypotheses defined, we can then define a figure of merit by which we can quantitatively assess the state of the analysis without appealing to the data and incurring potential bias. This figure is the expected limit on μ. This expected limit is the magnitude of μ that we can exclude, with 95% confidence, assuming the observed data looks exactly like the simulated backgrounds. This limit is generally driven by statistical or systematic uncertainties, depending on the analysis, and thus, choices that reduce either of these typically improve the limit, and by extension, the sensitivity of the measurement.

With the measurement and a figure of merit defined, we can now proceed to select a target final-state with optimal sensitivity to the presence of the $t\bar{t}H$ process.

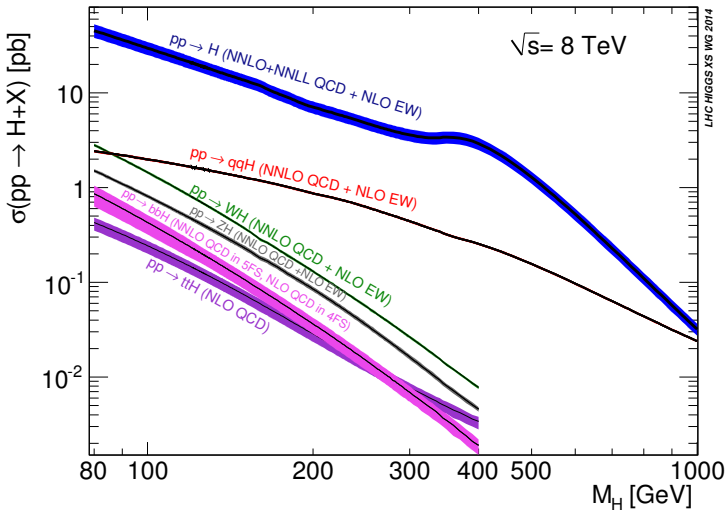


Figure 4.2: Higgs boson production mechanisms and their respective cross-sections as a function of Higgs mass at 8 TeV. [42]

4.2 Search Channel Definition and Motivation

In the $t\bar{t}H$ case, a signal hypothesis can be defined for many possible final-states, each with pros and cons with regard to their experimental sensitivity. These final-states are dependent on the different decays of the top and Higgs particles and occur at different rates. Special attention must be paid to these branching fractions in particular as $t\bar{t}H$ production is relatively rare and thus statistics limited for many channels utilizing the 20.3 fb^{-1} dataset. The various Higgs boson production mechanisms and their respective cross-sections are shown in Figure 4.2.

With regard to the topologies of the Higgs decays, four final-states targeted at ATLAS are depicted in Figure 4.4. These four were chosen here in particular to highlight the differing strategies for $t\bar{t}H$ measurement. In the case of the electroweak decays described by Figures 4.4(a) and 4.4(c), large branching fractions provide needed statistics for a measurement, given the small $t\bar{t}H$ production cross-section. The $H \to b\bar{b}$ and $H \to WW$ channels contribute 57.7% and 21.5% of the total decay width respectively, with a 125 GeV Higgs boson. However, the presence of similar decay products from the top pairs make identifying the correct Higgs-daughter objects very difficult. On the other hand, the decay described by Figure 4.4(b) is statistics limited due to its small branching fraction of 0.228%; however, it provides a clean signal in the form of two high-p_{T} photons which can be used to cleanly reconstruct the Higgs. Last but not least, the decay shown by Figure 4.4(d) provides an order of magnitude more events than the di-photon decay, but also suffers from a similar object multiplicity issue as do the first two Higgs decay channels. Due to the relatively few anticipated $t\bar{t}H$ events present in the 8 TeV dataset on which this analysis

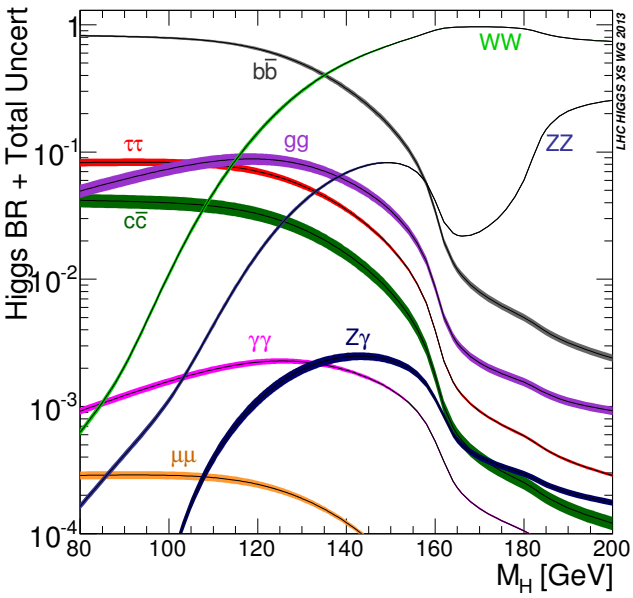
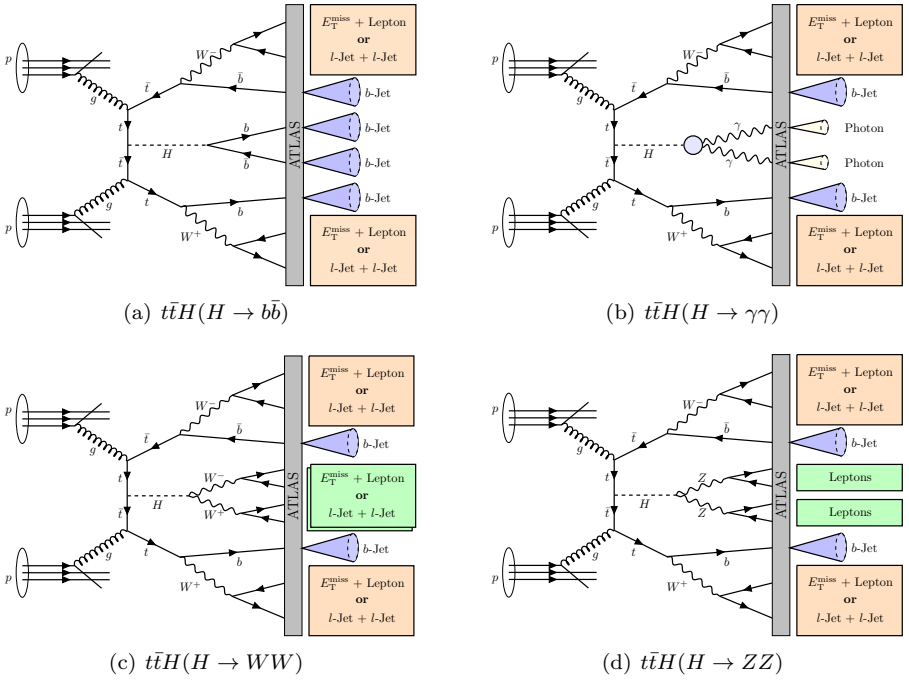


Figure 4.3: Higgs boson decay channels and their respective branching fractions as a function of Higgs mass. [42]

is based, we choose to focus our measurement on the $t\bar{t}H$ ($H \to b\bar{b}$) channel due to its dominant decay fraction. Unfortunately, as this channel contains four b-jets, reconstruction of the Higgs boson invariant mass becomes extremely difficult, as it is not clear *a priori* which two are the b-jets from the Higgs boson. As a result, side-band-subtraction techniques cannot be used to estimate the irreducible background. These issues with the b-jet multiplicity must then be solved with various multivariate analyses techniques, such as the one described in Chapter 7. The relative branching fractions of these Higgs decay channels and others are shown as a function of Higgs boson mass in Figure 4.3.

After choosing the decay channel for the Higgs, the top pair final-state has to be chosen. Each top decays, almost exclusively, to a W boson and a bottom quark. The W decays leptonically to a lepton and a neutrino, or hadronically to a pair of quarks. Tau-leptons can appear as a particle shower, similar to the jets from quarks in the hadronic case. Thus, they are categorized separately from the muon and electron channels. The leptonic branching fractions for the W are approximately 11% for each lepton flavor, and 67% for a hadronic decay. As our signal has two top quarks, two W decays result in three categories of final-states, as we ignore the W decays to tau leptons. These categories are defined as: the di-leptonic channel with two high-p_{T} leptons and four b-jets; the lepton-plus-jets channel with a single high-p_{T} lepton, two light-flavor jets (l-jet), and 4 b-jets; and finally the all-hadronic channel, with four light-flavor jets, four b-jets, and no leptons. These are depicted in Figure 4.5, along with the respective branching fractions for each category.

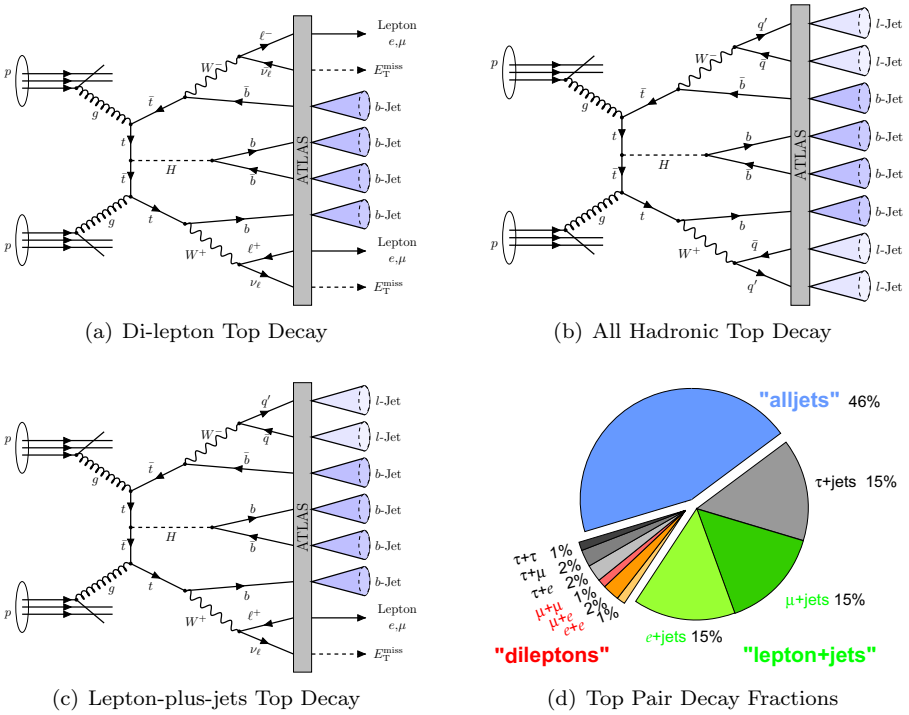
For this analysis, the lepton-plus-jets channel, Figure 4.6, was chosen for two



(a) $t\bar{t}H(H \to b\bar{b})$ (b) $t\bar{t}H(H \to \gamma\gamma)$

(c) $t\bar{t}H(H \to WW)$ (d) $t\bar{t}H(H \to ZZ)$

Figure 4.4: Cartoons of selected $t\bar{t}H$ Higgs decay channels and their respective final-state topologies. l-jet denotes a light-flavor jet.



(a) Di-lepton Top Decay (b) All Hadronic Top Decay

(c) Lepton-plus-jets Top Decay (d) Top Pair Decay Fractions

Figure 4.5: Cartoons of selected $t\bar{t}H$ ($H \to b\bar{b}$) Top decay channels and their respective final-state topologies. The pie chart shows each channel's respective branching fraction. l-jet denotes a light-flavor jet.

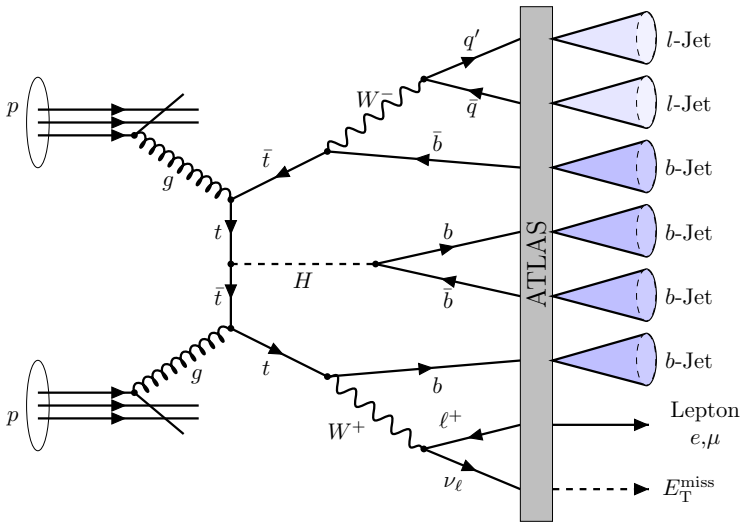


Figure 4.6: The lepton-plus-jets $t\bar{t}H$ ($H \to b\bar{b}$) channel, the target for this analyses, chosen for its large branching fraction and clean single-lepton trigger signature. l-jet denotes a light-flavor jet.

primary reasons. First, the branching fraction of this final-state is relatively large at 30% for the summed electron and muon sub-channels, as opposed to the di-lepton channel's 4%, though not as good as the fully hadronic channel at 46%. Secondly, the single high-p_{T} lepton allows for efficient triggering thus significantly reducing the contamination by multi-jet QCD backgrounds in which high-p_{T} leptons are generally absent. However, this channel does come with its own challenges, chief of which is the potential mis-classification of the light and heavy flavor jets, in addition to large $t\bar{t}$ backgrounds described in Section 4.3.

4.3 Expected Standard Model Backgrounds

The lepton-plust-jets $t\bar{t}H\left(H \to b\bar{b}\right)$ channel benefits primarily from the relatively large branching fractions of the combined ($H \to b\bar{b}$) and lepton-plus-jets decay channels, preserving the few events originating from the rare $t\bar{t}H$ production process. The price to pay for this, however, is the comparatively huge SM backgrounds and a heavy dependence on quark flavor identification.

By far, the main source of backgrounds originate from SM $t\bar{t}$ events produced in association with at least two additional jets. In terms of number of events at 8 TeV, for every $t\bar{t}H$ event produced at ATLAS, approximately 1900 $t\bar{t}$ pairs will be produced. Fortunately, by utilizing the object definitions and selection criteria defined later in Chapter 6, this background can be reduced to 26 background events for every expected signal event.

The inclusive $t\bar{t}$ background can be subdivided into three orthogonal components based on the flavor content of the additional jets. Events with extra b-flavor jets are classified as an irreducible $t\bar{t} + b\bar{b}$ background class, while

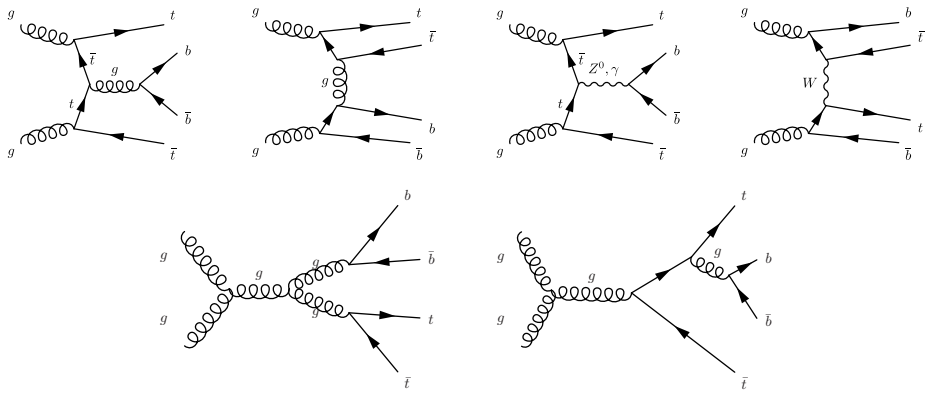


Figure 4.7: Example Feynman diagrams of the irreducible $t\bar{t}$+$b\bar{b}$ Standard Model background. For additional diagrams, including $q\bar{q}$ production modes, see Figure 7.3 and Figure 7.4.

the remainder which pass the event selection constitute a reducible $t\bar{t} + c\bar{c}$ or $t\bar{t}$ + Light-Flavor component with misclassified extra-jets, allowing them to pass the selection in the signal region. Feynman diagrams of the irreducible component are shown in Figure 4.7.

Ultimately, the sensitivity of this analysis is driven by the theoretical uncertainty on the $t\bar{t} + b\bar{b}$ cross-section and on the kinematic modeling of the $t\bar{t}$ background events in general. With this in mind, a novel multivariate analysis is developed with the principle goal of kinematically differentiating $t\bar{t}H$ events from $t\bar{t} + b\bar{b}$ events. To the extent that this is possible, statistical constraints can be put on the $t\bar{t} + b\bar{b}$ uncertainty based on $t\bar{t} + b\bar{b}$ enriched regions of the measurement. The better understanding of both the signal and the irreducible background provided by this method subsequently increases this analysis' sensitivity to the SM $t\bar{t}H$ production cross-section. This multivariate analysis is defined in Chapter 7 and the improvement attributed to this method is shown in Chapter 11.

Chapter 5

Signal and Background Simulation

The generation and simulation of the various samples used at ATLAS is a monumental computing challenge. As such, ATLAS and CERN devote a tremendous amount of manpower to develop the required software and computing resources needed to keep up with continuing demand for simulated data. The express purpose behind this effort is to provide analysts with the most accurate simulation of the physical data as possible, incorporating everything from theoretical and phenomenological effects to detector resolution, efficiency and acceptance effects.

5.1 Event Generation and Simulation Overview

In the most general terms, the simulation of a given event proceeds through three primary phases. In the first phase, a process is generated utilizing the matrix element information for a given process by which the scattering amplitudes can be calculated. Also, parton distribution functions (PDFs) are used to encode the status of the colliding initial sate objects. The second phase handles non-perturbative elements of of the final-state evolution such as the parton shower and hadronization. In addition, this phase introduces additional effects independent of the process under consideration such as initial and final-state radiation (ISR,FSR). Finally, the third and final phase takes the generated events from the previous phases and passes them through a detector simulation and reconstruction framework which effectively creates simulated raw data based on the ATLAS detector model.

5.2 $t\bar{t}H$ $(H \to b\bar{b})$ Modeling and Simulation

To model $t\bar{t}H$ events, matrix elements are first defined with the HELAC-Oneloop [43] package corresponding to a next-to-leading-order (NLO) QCD approximation. To model the parton shower, POWHEG BOX [44–46] is used as an interface for PYTHIA 8.1 [47] to shower the HELAC-Oneloop output. Samples utilizing this method are defined as the POWHEL samples and represent the baseline simulation for the signal process. These samples include all possible SM Higgs decay modes inclusively and utilize the CT10NLO PDF

set [48, 49] parameterized with factorization (μ_F) and renormalization (μ_R) scales set to $\mu_0 = \mu_F = \mu_R = m_T + m_H/2$. The top quark mass is set to 172.5 GeV and samples are produced for a range of Higgs boson masses, $m_H = \{115, 120, 125, 130, 135\}$ GeV. PHOTOS 2.15 [50] and TAUOLA 1.20 [51] are used to simulate photon radiation and tau decays respectively.

All $t\bar{t}H$ samples are simulated using the full ATLAS detector geometry and response using GEANT4 [52] , except for the samples used for training which utilizes a faster, approximate simulation of the calorimeter response. This includes the modeling of multiple simultaneous pp interactions to simulate the anticipated pile-up during the 8 TeV run. Finally, simulated events are passed through the entire ATLAS reconstruction software. In some circumstances, the results are corrected with scale factors such that object identification and reconstruction efficiencies as well as energy scales and resolutions are consistent between data and simulation in predefined control samples. The theoretical cross-section and branching fractions for these samples are calculated at NLO and are available in Reference [53].

5.3 $t\bar{t}$+Jets Backgrounds Modeling and Simulation

Following the trigger-based preselection, the largest contributing background originates from $t\bar{t}$+jets production. This is a particularly crucial background as it is composed of $t\bar{t}$ pairs produced in association with additional jets, some of which originate from additional b-quarks. Top quark pairs that are produced in association with two b-quarks constitute an irreducible background with an identical final-state topology as the signal process for this analysis. Furthermore, the ratio between $(t\bar{t} + b\bar{b})$, $(t\bar{t} + c\bar{c})$, and $(t\bar{t} + \text{Light-Flavor})$ events produced in the inclusive $t\bar{t}$+jets sample represents an extremely challenging theoretical hurdle. As a result, the largest systematic uncertainty for this analysis is the uncertainty on the theoretical cross-section of the $t\bar{t} + b\bar{b}$ background.

The simulated samples for this background are produced inclusively by interfacing POWHEG [44] and PYTHIA 6.425 [54] for NLO matrix element generation and showering respectively. In accordance with many top quark analyses, all POWHEG +PYTHIA events, except for those events classified as $t\bar{t} + b\bar{b}$ events, are corrected to the top-quark-p_{T} and $t\bar{t}$-p_{T} differential cross-sections measured from 7 TeV data by using a sequential reweighing procedure by which events are re-weighted first with the $t\bar{t}$-p_{T} factor, and then with the top-quark-p_{T} factor. The improvement due to this reweighing procedure can be seen in Figure 5.1, based on the improved data to simulation agreement in the H_T and jet multiplicity distributions. Here, H_T is defined as the scalar sum of all jet E_T in the event. The region pictured is comprised of all events which pass the analysis event selection and have at least 6 jets and exactly 2 b-tagged jets, consistent

with the $t\bar{t}$ + Light-Flavor final-state.

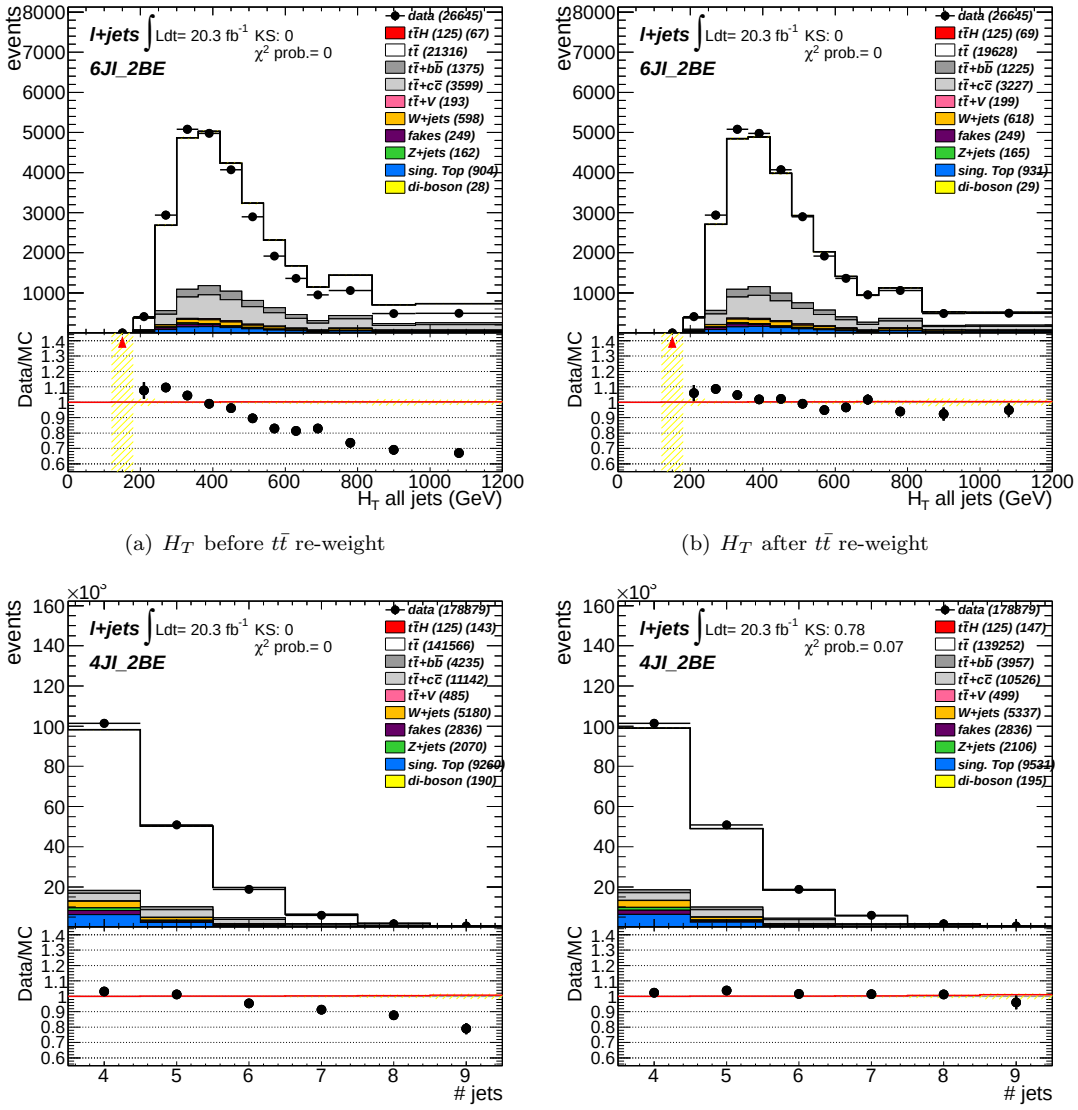
The $t\bar{t}+b\bar{b}$ events are corrected separately to a description given by SHERPA + OPENLOOPS [55, 56] NLO. In addition to POWHEG +PYTHIA , other event generation schemes were tested, including MADGRAPH +PYTHIA , and were shown to either be inferior or equivalent to the POWHEG +PYTHIA model. The classification of events into the $t\bar{t}+b\bar{b}$, $t\bar{t}+c\bar{c}$, and $t\bar{t}+$Light-Flavor categories is described in Section 5.3.1. Finally, the inclusive $t\bar{t}$+jets theoretical cross-section is computed using next-to-next-to-leading-order (NNLO) in QCD calculations which include resummation of the next-to-next-to-leading-logarithmic (NNLL) terms by TOP++ 2.0.

5.3.1 $t\bar{t}$+Jets Flavor Classification

The $t\bar{t}$+jets sample is further subdivided into three mutually exclusive samples containing $t\bar{t}+b\bar{b}$, $t\bar{t}+c\bar{c}$, and $t\bar{t}+$ Light-Flavor events. This categorization is based on a hadron-to-*particle-jet* matching algorithm. Particle jets are the progenitors of the reconstructed jets simulated by the ATLAS simulation framework. To be identified as a b-flavor *particle jet*, the candidate must have a p_{T} greater than 15 GeV and be matched to a b-hadron with a p_{T} greater than 5 GeV. A *particle jet* and a hadron are considered matched if the angular separation, or $\Delta R_{\eta\phi}$, is less-than or equal-to 0.4, where $\Delta R_{\eta\phi}(a,b) = \sqrt{(\eta_a-\eta_b)^2+(\phi_a-\phi_b)^2}$. Jets that directly originate from top quark decays are not considered for matching as the goal of this algorithm is to classify the 'extra' jets that do not come from top quark decays. The number of events classified into each category are shown in Figure 5.2.

The same procedure is repeated with remaining particle jets with respect to present charm hadrons. Finally, events with at least one b-flavor *particle jet* are placed in the $t\bar{t}+b\bar{b}$ category. Remaining events with at least one c-flavor *particle jet* are placed in the $t\bar{t}+c\bar{c}$ category, and all remaining events with no b- or c-flavored particle jets are placed in the $t\bar{t}+$ Light-Flavor category. Each of these categories are ultimately treated as distinct samples with unique normalizations despite the inclusive nature of their production.

To obtain a more precise description of the $t\bar{t}+b\bar{b}$ sample, the SHERPA + OPENLOOPS [55, 56] NLO reweighing scheme mentioned in Section 5.3 is utilized. This method provides a NLO description of the $t\bar{t}+b\bar{b}$ sample in addition to providing systematic uncertainties to the $t\bar{t}+b\bar{b}$ modeling itself, providing it with the same footing had it been produced on its own. To accomplish this however, the $t\bar{t}+b\bar{b}$ sample must be further sub-divided based on the multiplicity of the b flavor *particle jets* in each event. Events are classified into categories with one to three b-jets, or with particle jets matched to more than one b-hadron, designated 'B-flavor' *particle jets*, denoting an unresolved gluon to $b\bar{b}$ splitting. Ultimately, this reweighing procedure results in excellent kinematic agreement in each category between POWHEG [44] and the NLO $t\bar{t}+b\bar{b}$ prediction, improv-



(a) H_T before $t\bar{t}$ re-weight

(b) H_T after $t\bar{t}$ re-weight

(c) Jet multiplicity before $t\bar{t}$ re-weight

(d) Jet multiplicity after $t\bar{t}$ re-weight

Figure 5.1: Data to simulation agreement before and after the application of the sequential $t\bar{t}$-p_{T} re-weighting method. The scalar sum of the transverse energy of all jets in the event, H_T, is shown before, (a), and after, (b), the re-weighting. In addition, the jet multiplicity for the event is shown before, (c), and after, (d), the re-weighting. This region includes all events which pass the analysis selection and have at least 6 jets, and exactly 2 b-tagged jets, consistent with a $t\bar{t}$ final-state. [2]

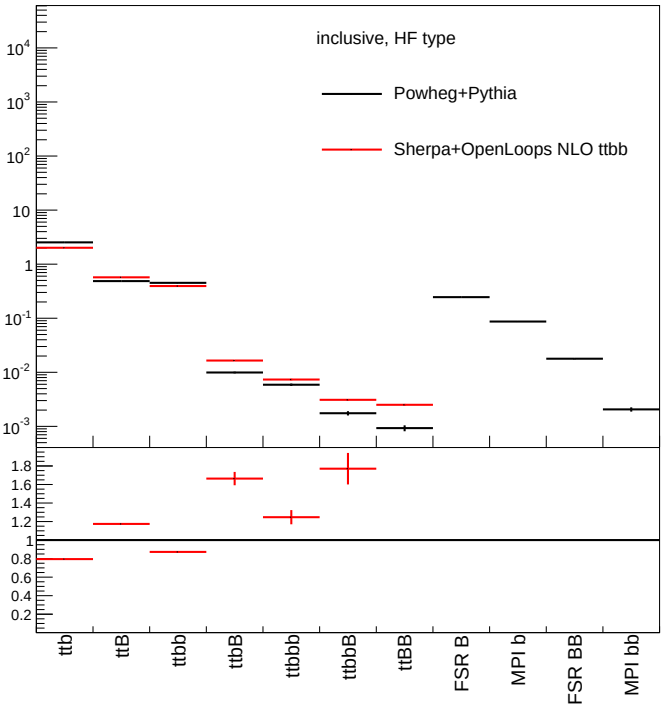


Figure 5.2: Plot showing a comparison between the different flavor components of $t\bar{t} + b\bar{b}$ under the SHERPA+OPENLOOPS NLO prediction with respect to the default POWHEG +PYTHIA prediction. The y-axis describes the number of events classified in each of the b-flavor categories along the x-axis. [2]

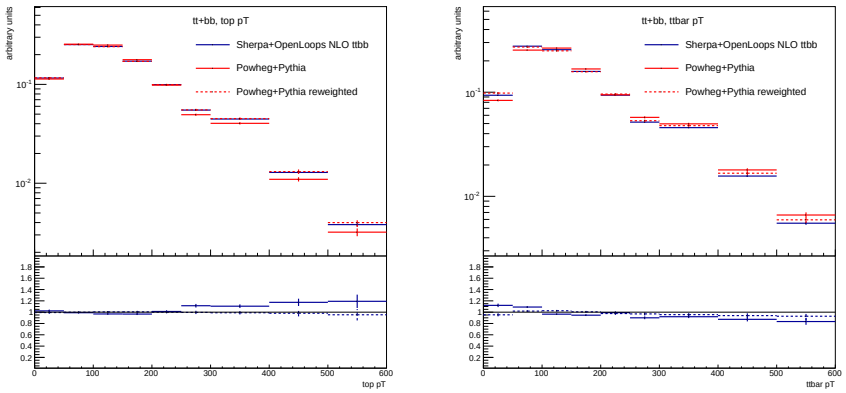


Figure 5.3: Plots showing the top-p_{T} (left) and $t\bar{t}$-p_{T} (right) before and after the POWHEG +PYTHIA to SHERPA+OPENLOOPS NLO re-weight showing good agreement beforehand. The areas have been normalized to unity to emphasize the shape differences. [2]

ing confidence in the default POWHEG +PYTHIA model, while simultaneously providing systematic handles specifically for the $t\bar{t}+b\bar{b}$ background. This comparison is shown in Figure 5.3.

5.4 Sub-leading, Reducible Backgrounds

The remaining backgrounds for the lepton-plus-jets $t\bar{t}H$ ($H \to b\bar{b}$) channel arise primarily due to W/Z+jets, diboson production, single-top production and $t\bar{t}$ production in association with vector bosons. These samples are evaluated from simulation and subsequently normalized to their respective theoretical cross-sections.

The W/Z+jets samples are produced by ALPGEN 2.14 [57] at leading order (LO) utilizing the CTEQ6L1 PDF set [58] . Parton showers and fragmentation are modeled by PYTHIA 6.425 [54] . Diboson production is showered with HERWIG 6.520 [59] . The 'MLM matching' algorithm [60] is used to avoid double-counting of the states produced by both the matrix element calculation and parton shower. As ALPGEN +HERWIG production does not include the hadronic decays of Zs, SHERPA + OPENLOOPS [55, 56] samples are generated with massive b and c quarks to provide these events. The $W$and Z samples are produced with up to five additional jets and are normalized to the NNLO [61] theoretical prediction. The Diboson sample is generated with up to three additional jets and is normalized to NLO [62] theoretical predictions.

Single-top production is modeled with POWHEG using the CT10NLO PDF set [48, 49] . This sample includes contributions from the s-channel, t-channel and associated Wt production modes. These samples are each showered with PYTHIA 6.425 [54] which utilize the CTEQ6L1 PDF set [58] . The MSTW2008 68% NNLO PDF set [40, 41] is then used to normalize the single-top samples to their NNLO theoretical predictions.

The $t\bar{t}+V$ samples, where $V = \{W, Z\}$, are generated with the MADGRAPH 5 LO [63, 64] generator interfaced to PYTHIA 6.425 [54] for showering and the CTEQ6L1 PDF set [58] . Each sample is then normalized to its respective NLO cross-section prediction.

The multijet background, consists of events with a lepton that originates either from converted photons or jets faking an election, or from soft, semi-letponic decays of b and c hadrons faking prompt muons. These contributions, being extremely hard to properly model, are derived from control regions using the 'Matrix Method' [65]. Using this method, kinematic distributions and normalizations are derived by measuring the relative efficiencies of leptons from known real and fake sources under graduated selection criteria. After parametrizing these efficiencies in terms of relevant kinematic parameters, they are applied to the measured data to estimate their expected contributions.

Chapter 6

Object Description, Event Selection, and Region Definitions

Having defined the signal topology and the primary backgrounds expected to share that topology in Chapter 4, the task now falls to explicitly define what constitutes a muon, an electron, a jet, etc. These definitions are known as 'object definitions.' Once these objects have been defined, a set of selection criteria can be developed to eliminate as much of the reducible background as possible while retaining as much signal as possible. Further, control regions can be defined from which information can be gathered to constrain various systematic uncertainties. All object definitions and selection criteria mirror the collaboration-wide recommended standard definitions. This chapter proceeds to lay out those object, event, and region definitions and their respective motivations.

6.1 Event Trigger and Preselection

From April 2012 to January 2012, the LHC collided protons with a center-of-mass energy of $\sqrt{s}$=8 TeV. During this run, the data successfully recored by ATLAS amounted to a total integrated luminosity of 20.3 fb^{-1} out of the 22.8 fb^{-1} delivered by the LHC, see Figure 3.16. All events included in this data set were recorded during stable beam conditions coinciding with nominal operation of all detector subsystems required for the detection, identification, and reconstruction of jets, leptons, photons, and $E_{\mathrm{T}}^{\mathrm{miss}}$.

As the lepton-plus-jets final-state for $t\bar{t}H$ ($H \to b\bar{b}$) is characterized by a single, high-p_{T} lepton produced with multiple jets, a single-lepton trigger was utilized. As the trigger definitions change over course of the data-taking periods, multiple triggers are used. The p_{T} thresholds used are either 24 or 36 GeV for muons, and either 24 or 60 GeV for electrons. The lower-p_{T} triggers include isolation-cuts which are not utilized for the higher-p_{T} triggers. In all cases, the trigger selection requirements are looser than the final object definition requirements. As a result of this trigger configuration, events are stored with a range of jet multiplicities. Events with varying jet multiplies will help define the various control regions introduced in Section 6.5.1.

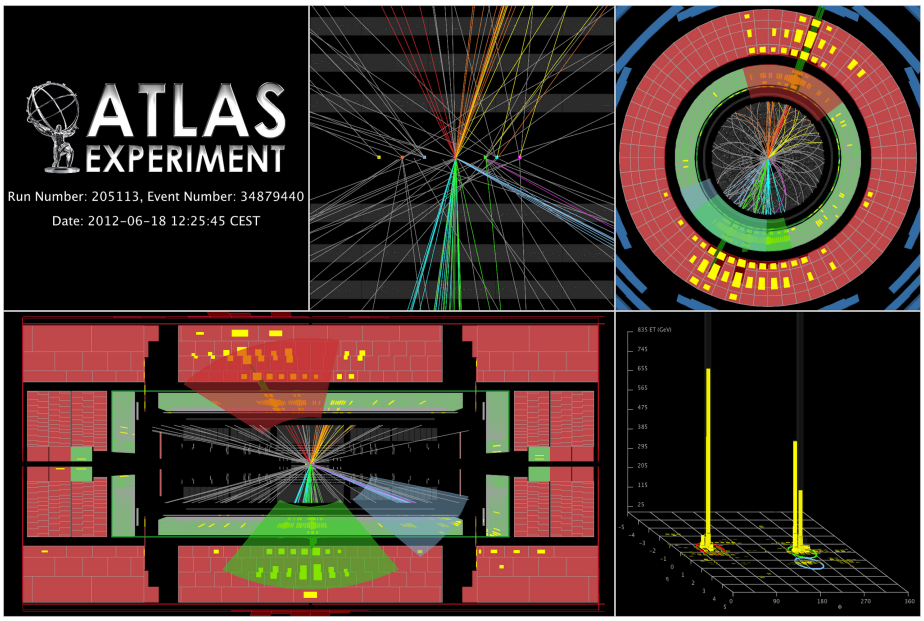


Figure 6.1: The highest-mass central dijet event with the highest-p_{T} jet collected by the end of June, 2012 (Event 34879440, Run 205113): the two central high-p_{T} jets have an invariant mass of 4.1 TeV, and the highest-p_{T} jet has a p_{T} of 2.1 TeV, and the sub-leading jet has a p_{T} of 1.9 TeV. The missing E_{T} and SumET for this event are respectively 63 GeV and 4.2 TeV. Only tracks with $p_{\mathrm{T}} > 400$ MeV are displayed. The event was collected on June 18th, 2012.

6.2 Jet Reconstruction and Selection

Jets are the detected remnants of high energy quarks and gluons. These objects are typically recognized by a high-multiplicity cone of charged particles produced, nearly collinearly with the parent particle, in the tracking regions of a detector. Jets also leave large energy deposits in the calorimetry. Figure 6.1 shows an event display of the highest-reconstructed-mass pair of jets recorded at ATLAS by the end of June in 2012. The top right pane and lower left pane show the η, ϕ and η, z projections of the charged particle tracks respectively, revealing the cone shaped showers of charged particles characteristic of jet objects. For more information regarding the subsystems involved, the ATLAS Inner Detector and calorimetry are described in Section 3.2.1.

Jet reconstruction algorithms then build the jet candidates from the energy deposits in the calorimeter. As jets are the result of a evolving interaction of highly energetic particles, it is often difficult to a assign a given energy deposit to a potential jet candidate. To handle this, jets are reconstructed using a jet reconstruction algorithm which respects certain phenomenological aspects of parton behavior, and thus provide a consistent definition of a jet between theorists and experimentalists. For this analysis, the anti-k_t algorithm [66–68] is used with a 'radius parameter' of 0.4. This algorithm uses calibrated topological neighbors and is collinear and infrared safe. Prior to the jet reconstruction, the

utilized topological neighbors are calibrated using the LCW local calibration scheme. [69, 70] Reconstructed jets are then calibrated to the measured jet energy scale (JES). [71]

Following all corrections and calibrations, this analysis requires all jets to have a p_{T} greater than 25 GeV and a $|\eta|$ less than 2.5. Jets whose p_{T} is less than 50 GeV are further required to satisfy a "jet vertex fraction" (JVF) cut. Jets with a JFV of 50% or greater are retained, while the rest are removed. A jet's JVF quantifies the fraction of tracks with $p_{\mathrm{T}} > 1$ GeV that are associated with the jet and originate from a vertex compatible with the primary collision vertex. This variable provides an efficient way to remove jets from secondary pp interactions.

Finally, to prevent double counting, jets in close proximity to electrons are removed, as no distinction between the two is made at the time of reconstruction. Jets within a $\Delta R_{\eta\phi}$ cone of 0.2 of a defined and selected electron are removed, then, any electron within a $\Delta R_{\eta\phi}$ cone of 0.4 of a remaining jet are removed. Here, $\Delta R_{\eta\phi}$ is defined as $\Delta R_{\eta\phi}(a,b) = \sqrt{(\eta_a - \eta_b)^2 + (\phi_a - \phi_b)^2}$.

6.2.1 Jet Flavor Tagging

Selected jets are subsequently classified into one of two categories based on the probable provenance of the jet. Jets originating from a bottom quark in the final-state, are classified as b-jets. Whereas, jets originating from the lighter flavor quarks or gluons produced in the final-state are labeled as *light flavor* jets. This classification process is performed by a b-jet identification algorithm referred to as a b-tagging algorithm.

To distinguish between the two, certain kinematic differences are exploited. The chief difference comes from the disparity in lifetimes between b-hadrons and their lighter counterparts. Due to the CKM suppression of flavor-changing charged currents in the SM, b-hadrons have significantly longer lifetimes. As a result, b-jets, and their constituent charged particles, often contain a displaced vertex, secondary to the primary collision vertex. In addition, there are numerous shower shape and composition differences that imply differing flavor content. To capture all this information into a single observable, a multivariate analysis [72] is preformed to classify the observed jet flavor for different efficiency working points.

For this analysis, the b-tagged jets and light flavor jets are treated as distinct but similar objects. Jets are tagged at a 70% working point which corresponds to a 70% tagging efficiency for jets originating from b-quarks. This working point similarly corresponds to a charm quark rejection factor of $\sim$5 and a light quark rejection factor of $\sim$130. This working point is characterized using simulated $t\bar{t}$ events with a nearly identical jet selection criteria [73]. Plots characterizing the response of the utilized tagging algorithm are shown in Figure 6.2.

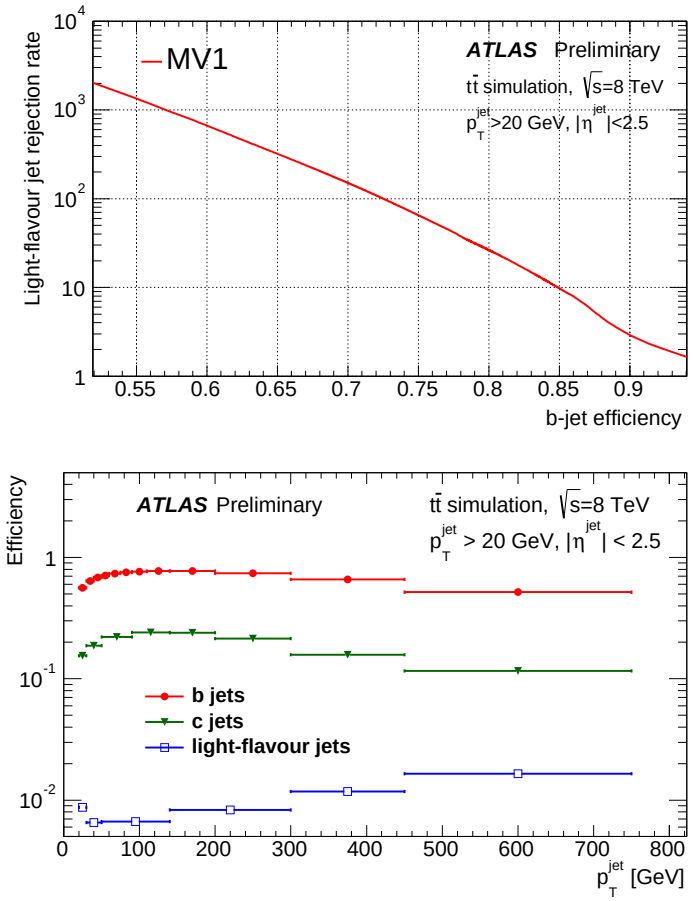


Figure 6.2: Plots showing the b-tagging response of the MV1 tagging algorithm. The plot on the top shows the light flavor jet rejection factor as a function of b-tagging efficiency. The plot on the bottom shows the efficiency response for the 70% working point as a function of jet p_{T} for light-flavor, b, and c-jets. For both plots, evaluated jets have a $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$. [73]

6.3 Lepton Reconstruction and Selection

6.3.1 Electron Objects

Electrons produced in the high energy pp collisions at ATLAS typically present as charged particle tracks in the Inner Detector (ID) accompanied by energy deposits, or clusters, in the electromagnetic calorimeter. For this analysis, selected electron candidates are identified by a collection of selection criteria over a collection of selected variables, optimized by the ATLAS e/γ combined performance group [74]. Of the available working points, the TIGHT classification was chosen. The TIGHT classification is a cut-based identification method parameterized in $|\eta|$ and E_T. A subset of the MEDIUM and LOOSE selections, the TIGHT selection includes tightened cuts on more variables. The efficiencies and scale factors for the electron identification as a function of E_T and the mean number of interactions $\langle\mu\rangle$ are shown in Figure 6.3.

For this analysis, the reconstructed electron candidates are required to have a corrected p_{T} greater than 25 GeV, a cluster-based $|\eta| < 2.47$, and a longitudinal impact parameter within 2 mm of the primary vertex. Candidates that are reconstructed within the transition region of the calorimeter, $1.37 < |\eta| \leq 1.52$, are excluded.

Electrons are required to satisfy isolation requirements which set maximum limits on the amount of energy allowed in adjacent cells. Energy deposits in adjacent cells within some $\Delta R_{\eta\phi}$ cone of the original electron cluster are summed, excluding energy deposits directly associated with the electron. This sum is then required to be below a given threshold which depends on the electron $|\eta|$ and E_T and a given efficiency working point. Further, this isolation is defined both for the scalar sum of E_T from adjacent cells as well as scalar summed p_T of nearby tracks.

For this analysis, requirements are placed on both E_T and p_T isolation. In the former case, E_T within a $\Delta R_{\eta\phi}$ cone of 0.2 must be below the 90% working point threshold. In the latter case, p_T from nearby tracks within a $\Delta R_{\eta\phi}$ cone of 0.3 must be below the 90% working point threshold. These thresholds are shown in Tables 6.1 and 6.2 as piecewise functions, binned in the $|\eta|$ and E_T of the electron.

6.3.2 Muon Objects

A muon is a charged lepton similar in all respects to the electron except for its mass, which is ∼200 times greater. High-p_{T} muons from the hard process are able to penetrate the calorimeters as they do not interact strongly and are relatively massive. This leaves a relatively pure collection of muons which then leave charged tracks in the Muon Spectrometer (MS). At ATLAS, muons are detected and characterized by both the inner tracking chambers of the Inner Detector (ID) and the MS, described in Section 3.2.1. Following track reconstruction in

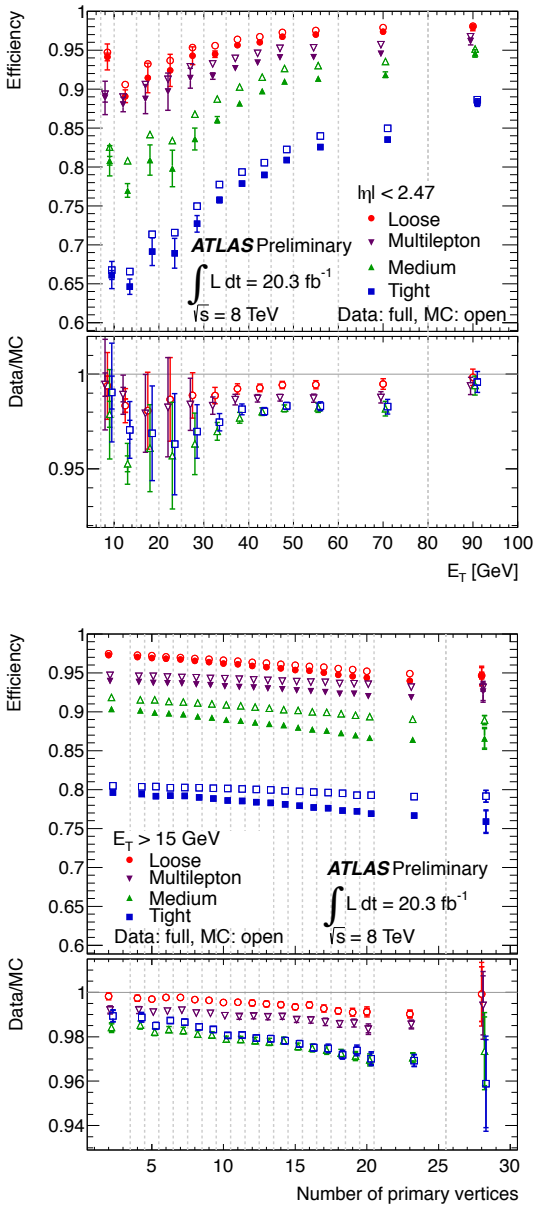


Figure 6.3: Plots showing the electron identification efficiency as a function of E_T and the mean number of interactions $\langle\mu\rangle$ for identified electron objects reconstructed using the LOOSE, MULTILEPTON, MEDIUM, and TIGHT selection criteria. These plots were obtained from the same $\sqrt{s}$=8TeV data as that used in this analysis. [74]

Electron E_T Isolation Requirements [GeV]

E_T Bins GeV	$\|\eta\|$ Bins ≤ 0.6	≤ 0.8	≤ 1.15	≤ 1.37	≤ 1.52	≤ 1.81	≤ 2.01	≤ 2.37	≤ 2.47
10 - 15	3.25	3.05	2.95	3.1	–	2.65	3.25	2.8	2.55
15 - 20	3.35	3.0	3.3	3.25	–	2.6	2.85	2.45	2.35
20 - 30	2.75	3.0	3.05	3.35	–	2.6	2.75	2.45	2.3
30 - 40	2.75	2.8	3.0	3.35	–	2.7	2.85	2.5	2.4
40 - 50	2.5	2.55	2.75	3.05	–	2.3	2.6	2.4	2.5
50 - 60	2.6	2.8	2.8	3.25	–	2.45	2.75	2.45	2.4
60 - 70	2.7	3.05	3.15	3.95	–	2.55	2.85	2.4	2.65
70 - 80	2.85	3.1	3.15	4.15	–	2.65	2.95	2.85	3.45
80 - 120	2.9	3.45	3.75	4.85	–	3.15	3.45	3.0	3.5

Table 6.1: The 90% working point electron E_T isolation requirements using a $\Delta R_{\eta\phi}$ cone of 0.2, shown here as a function of electron E_T and $|\eta|$. Bins with a requirement of '–' denote the excluded calorimeter transition region.

Electron p_T Isolation Requirements [GeV]

E_T Bins GeV	$\|\eta\|$ Bins ≤ 0.6	≤ 0.8	≤ 1.15	≤ 1.37	≤ 1.52	≤ 1.81	≤ 2.01	≤ 2.37	≤ 2.47
10 - 15	1.3	1.35	1.2	1.1	–	1.2	1.15	1.0	1.0
15 - 20	1.15	1.15	1.25	1.2	–	1.0	1.0	1.0	1.0
20 - 30	1.0	1.05	1.05	1.0	–	1.0	1.0	1.0	1.0
30 - 40	1.0	1.0	1.0	1.0	–	1.0	1.0	1.0	1.0
40 - 50	1.0	1.0	1.0	1.0	–	1.0	1.0	1.0	1.0
50 - 60	1.0	1.0	1.0	1.0	–	1.0	1.0	1.0	1.0
60 - 70	1.0	1.0	1.0	1.0	–	1.0	1.0	1.0	1.0
70 - 80	1.0	1.0	1.0	1.0	–	1.0	1.0	1.0	1.0
80 - 120	1.0	1.0	1.0	1.0	–	1.0	1.05	1.0	1.0

Table 6.2: The 90% working point electron p_T isolation requirements using a $\Delta R_{\eta\phi}$ cone of 0.3, shown here as a function of electron E_T and $|\eta|$. Bins with a requirement of '–' denote the excluded calorimeter transition region.

each of these regions separately, ID tracks are matched and combined with MS tracks to form a *combined* muon object. Of the reconstruction algorithms available at ATLAS, this analyses uses the MuID, or Chain 2, algorithm which is unique in that it re-fits the *combined* muon object using the matched MS and ID track information while taking into account the potential energy loss in the calorimeter. The detector response for Chain 2 combined muons is shown in Figure 6.4 as a function of both p_{T} and the mean number of interactions $\langle\mu\rangle$. [75]

Following their reconstruction, muons must satisfy selection criteria on their track quality as well as their kinematic and isolation properties. Muons must have a track quality consistent with a well-reconstructed track, interacting with a sufficient number of active elements in the ID, and its longitudinal impact parameter must be within 2 mm of the primary vertex. A muon candidate must have a $p_T > 25$ GeV and an $|\eta| < 2.5$. Further, the scalar sum of the track p_{T} in a $\Delta R_{\eta\phi}$ cone of 10 GeV/p_T^{μ}, centered on the muon candidate, must be less than 5% of the muon's p_{T}. Finally, muon candidates reconstructed within a cone of $\Delta R_{\eta\phi} = 4.0$ of the nearest defined jet object are removed from consideration to avoid selecting non-prompt muons from jet showers.

6.4 Missing Transverse Energy Definition

Missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$, originates from the production of prompt neutrinos in the hard process and subsequent decays. As neutrinos do not interact with the detectors at ATLAS, the energy carried by these particles is not directly measured. As a result, the vector sum of the transverse energy, E_T, is not zero as it should be given the conservation of momentum. Here, $E_T = E\sin\phi$, and in the high energy limit, $E_T = p_{\mathrm{T}}\sin\phi$. The vector equal in magnitude but opposite in direction to sum's residual vector is defined as the $E_{\mathrm{T}}^{\mathrm{miss}}$. The fact that there are several sources of transverse energy and several methods by which this energy can be lost, aside from prompt neutrinos, make this object particularly hard to define.

This analysis utilizes the MET_RefFinal calibration of the $E_{\mathrm{T}}^{\mathrm{miss}}$, as defined by the members of the ATLAS *jet*+$E_{\mathrm{T}}^{\mathrm{miss}}$ combined performance group [76]. In this calibration, topological clusters of energy deposits in the calorimeters are assigned to nearby reconstructed objects like jets and leptons. These clusters are then calibrated with prescriptions unique to their associated objects. The event's $E_{\mathrm{T}}^{\mathrm{miss}}$ is then defined as, $E_{\mathrm{T}}^{\mathrm{miss}} = -\sum E^{\mathrm{cluster}}\sin\phi^{\mathrm{cluster}}$. Though this analysis does not characterize or select events based on $E_{\mathrm{T}}^{\mathrm{miss}}$ information, the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution criteria is used to model the ATLAS detector response for the Matrix Element Method (MEM), the core technique presented in this analysis. This detector modeling, as a part of the MEM technique, can be reviewed in Section 7.2.3.

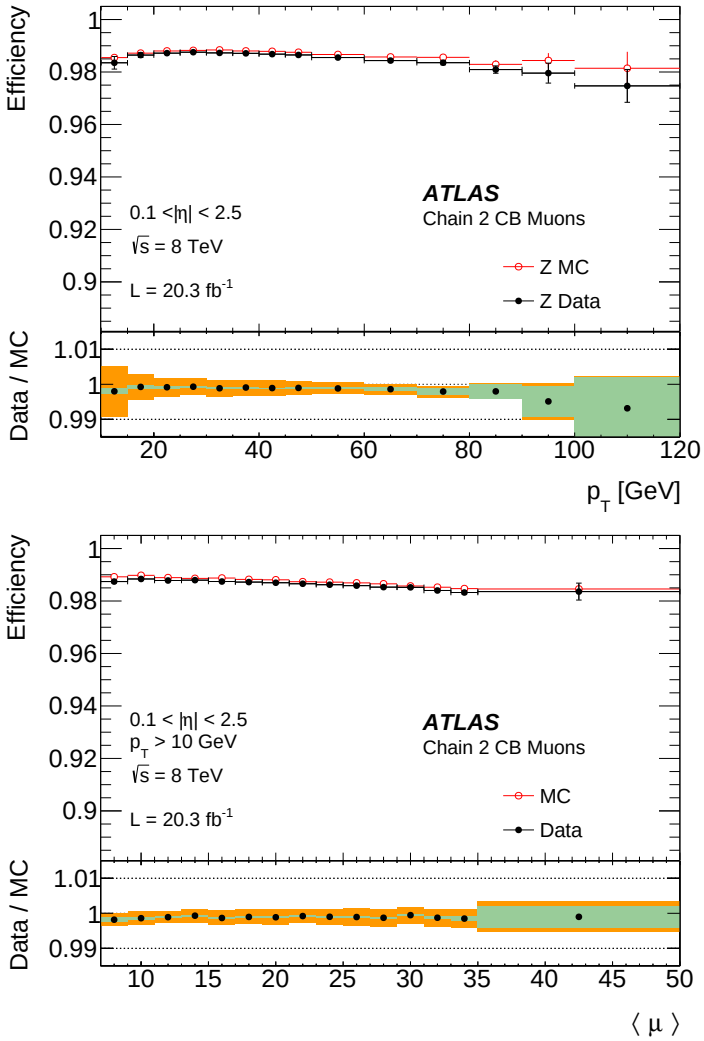


Figure 6.4: Plots showing the muon reconstruction efficiency as a function of p_{T} and the mean number of interactions $\langle\mu\rangle$ for the *combined* muon objects reconstructed using the 2nd reconstruction chain. Green regions in the ratio plots indicate the statistical uncertainty whereas the orange areas include also the systematic uncertainty. These plots were obtained from the same $\sqrt{s}$=8TeV data as that used in this analysis. [75]

6.5 Event Selection and Classification

Events which satisfy the necessary trigger requirements defined above are required to have at least one reconstructed vertex with at least five associated tracks. Each such vertex is then sorted by the sum of the squared momenta of its associated tracks. The vertex with the highest sum is subsequently taken as the primary vertex. Events with jet objects of $p_{\mathrm{T}} > 20$ GeV, which have been determined to have originated from the previous bunch crossing, or from calorimeter noise, are rejected. Events containing multiple leptons, or single leptons not matched to the the fired trigger are discarded to prevent overlap with the $t\bar{t}H$ ($H \rightarrow$ Multi-leptons) channels. Finally, events containing less than four defined jets are rejected.

6.5.1 Jet and b-tag Multiplicity-Based Analysis Model

Using the trigger and object definitions described in the previous sections, events are selected and classified based on their relevance to the lepton-plus-jets $t\bar{t}H$ ($H \rightarrow b\bar{b}$) final state. Events remaining after this selection are defined by a single lepton, either a muon or electron, and at least four defined jets.

To increase the sensitivity of this analysis, the remaining events are classified into 9 exclusive regions in jet and b-tag multiplicity. The purpose of this approach is to define distinct analysis regions, each with different background and signal compositions. The statistical combination of these regions will subsequently constrain the background normalization uncertainties in the signal rich regions by using available information in the background dominated regions.

The 9 regions are first characterized by the number, n, of jets in the event (4, 5, ≥ 6). Each n-jet category is subsequently categorized by the number of jets which have been b-tagged (2, 3, ≥ 4). Each region is subsequently named according to its n-jet and b-tag multiplicity, shown by Table 6.3. Each region's $S/\sqrt{B}$ ratio provides a means to identify 'signal rich' and 'signal depleted' regions. The yields for each region, taking into account statistical and systematic uncertainties prior to the likelihood fit are shown in Table 6.5. Each region will contribute a characteristic distribution to the multidimensional profiled likelihood fit, described in Section 11.1, for signal extraction. The characteristic distributions are chosen specifically to isolate certain systematic uncertainties in the case of the signal depleted regions, or to separate signal events from background events in the signal rich regions.

Table 6.4 describes the layout of the analysis regions and the characteristic distributions chosen for each. For many of the background dominated regions, the scalar sum of the jet transverse-energy, H_T^{had}, was chosen primarily due to its sensitivity to the jet energy scale (JES) and $t\bar{t}$ modeling systematics, allowing constraints to be placed on these uncertainties in the signal regions. H_T^{had} is also well behaved for all the contributing background processes across multiple jet and b-tag multiplicities, allowing for consistency across the primary back-

Analysis Region Names

Region	2 b-tags	3 b-tags	$\geq$4 b-tags
4 jets	(4j, 2b)	(4j, 3b)	(4j, $\geq$ 4b)
5 jets	(5j, 2b)	(5j, 3b)	(5j, $\geq$ 4b)
$\geq$6 jets	($\geq$ 6j, 2b)	($\geq$ 6j, 3b)	($\geq$ 6j, $\geq$ 4b)

Table 6.3: Names for each analysis region.

Analysis Region Fit Distributions

Region	2 b-tags	3 b-tags	$\geq$4 b-tags
4 jets	H_T^{had}	H_T^{had}	H_T^{had}
5 jets	H_T^{had}	HF-NN	NN
$\geq$6 jets	H_T^{had}	MVA	MVA

Table 6.4: Layout of the characteristic distributions chosen for each analysis region.

ground regions. The (5j, 3b) region is given a specialized neural network trained specifically to separate $t\bar{t} + b\bar{b}$ and $t\bar{t} + c\bar{c}$ contributions from $t\bar{t}$ + Light-Flavor contributions, HF-NN, with the goal of better constraining the heavy flavor normalizations. The (5j, $\geq$ 4b) region is given a specialized neural network trained specifically to separate $t\bar{t}H\left(H \to b\bar{b}\right)$ from $t\bar{t} + b\bar{b}$ contributions, NN, with the goal of better constraining the heavy flavor normalizations. Finally, the signal-rich ($\geq$ 6j, 3b) and ($\geq$ 6j, $\geq$ 4b) regions are provided with specialized multi-variate analyses (MVA) developed to discriminate between $t\bar{t}H\left(H \to b\bar{b}\right)$ and $t\bar{t} + b\bar{b}$, resulting in a cleaner measurement of the $t\bar{t}H\left(H \to b\bar{b}\right)$ signal strength. The goal of this dissertation is to introduce a specialized MVA designed specifically for this purpose and to present the improvement resulting from this effort. The presented method is introduced in Chapter 7.

Figure 6.5 shows the predicted relative $S/\sqrt{B}$ for each region. The blue bins signify the 'signal depleted' regions used as control samples. The red bins signify the 'signal rich' regions used for the extraction of the $t\bar{t}H$ signal. Figure 6.6 shows the relative background contributions for the dominant SM backgrounds in each region, and Figure 6.7 shows the relative signal contribution for the dominant Higgs boson decay modes.

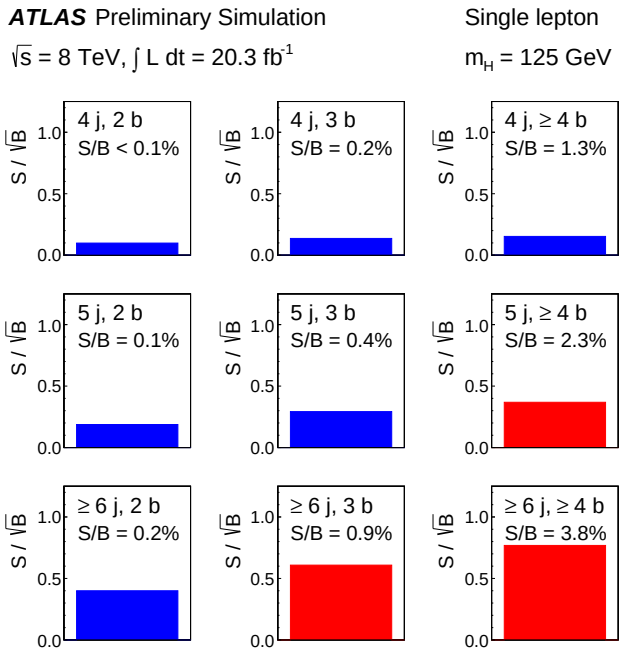


Figure 6.5: A plot showing the S/B and $S/\sqrt{B}$ of the various analysis regions based on jet and b-tag multiplicity. Here a Higgs boson mass of 125 GeV and SM cross-sections and branching ratios are assumed. [2]

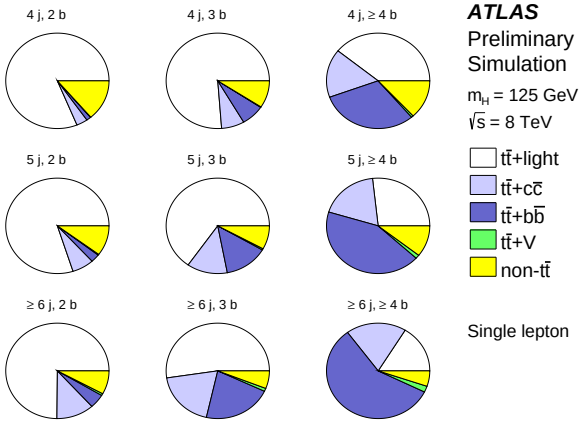


Figure 6.6: A plot showing the relative background contributions to each analysis region based on jet and b-tag multiplicity. Here a Higgs boson mass of 125 GeV and SM cross-sections and branching ratios are assumed. [2]

Pre-fit Event Yields per Analysis Region

	4 j, 2 b	4 j, 3 b	4 j, $\geq$ 4 b
$t\bar{t}H$ (125)	30.7 ± 2.8	12.9 ± 1.4	1.95 ± 0.29
$t\bar{t}$+ light	76 700 ± 7500	6170 ± 750	53.2 ± 12
$t\bar{t} + c\bar{c}$	4870 ± 3000	682 ± 390	21.2 ± 12
$t\bar{t} + b\bar{b}$	1840 ± 1100	680 ± 380	44.2 ± 25
W+jets	5120 ± 3000	225 ± 130	5.52 ± 3.3
Z+jets	1130 ± 600	50.2 ± 27	0.900 ± 0.56
Single top	4930 ± 640	337 ± 60	6.78 ± 1.6
Diboson	217 ± 71	11.5 ± 4.1	0.240 ± 0.12
$t\bar{t} + V$	122 ± 40	15.5 ± 5.1	0.890 ± 0.30
Multijet	1560 ± 620	102 ± 37	3.52 ± 1.3
Total	96 500 ± 9500	8280 ± 1100	138 ± 34
Data	98 049	8752	161

	5 j, 2 b	5 j, 3 b	5 j, $\geq$ 4 b
$t\bar{t}H$ (125)	40.9 ± 2.1	22.7 ± 1.8	6.22 ± 0.80
$t\bar{t}$+ light	37 600 ± 5500	3480 ± 520	60.8 ± 15
$t\bar{t} + c\bar{c}$	4300 ± 2400	810 ± 460	42.8 ± 25
$t\bar{t} + b\bar{b}$	1670 ± 880	886 ± 480	115 ± 63
W+jets	1940 ± 1200	135 ± 87	5.89 ± 3.9
Z+jets	405 ± 240	28.9 ± 17	1.47 ± 0.90
Single top	1880 ± 360	195 ± 41	8.32 ± 1.3
Diboson	96.5 ± 39	8.02 ± 3.4	0.400 ± 0.20
$t\bar{t} + V$	145 ± 48	26.5 ± 8.6	3.10 ± 1.0
Multijet	461 ± 170	69.9 ± 28	8.31 ± 3.7
Total	48 500 ± 7000	5670 ± 980	252 ± 75
Data	49 699	6199	286

	$\geq$ 6 j, 2 b	$\geq$ 6 j, 3 b	$\geq$ 6 j, $\geq$ 4 b
$t\bar{t}H$ (125)	63.7 ± 5.0	40.2 ± 3.5	16.5 ± 2.0
$t\bar{t}$+ light	18 800 ± 4400	2010 ± 460	52.4 ± 17
$t\bar{t} + c\bar{c}$	3730 ± 2100	846 ± 480	79.1 ± 46
$t\bar{t} + b\bar{b}$	1420 ± 770	974 ± 530	245 ± 130
W+jets	912 ± 620	96.7 ± 66	8.60 ± 6.2
Z+jets	183 ± 120	19.0 ± 12	1.54 ± 1.0
Single top	836 ± 220	122 ± 35	11.9 ± 3.7
Diboson	50.5 ± 24	5.98 ± 3.0	0.540 ± 0.27
$t\bar{t} + V$	182 ± 59	44.6 ± 14	8.45 ± 2.8
Multijet	181 ± 66	21.3 ± 7.6	1.09 ± 0.52
Total	26 400 ± 5800	4180 ± 1000	426 ± 150
Data	26 185	4701	516

Table 6.5: Pre-fit event yields for contributing processes and data in each of the analysis regions. The quoted uncertainties are the sum in quadrature of the statistical and total systematic uncertainties on the yields. [2]

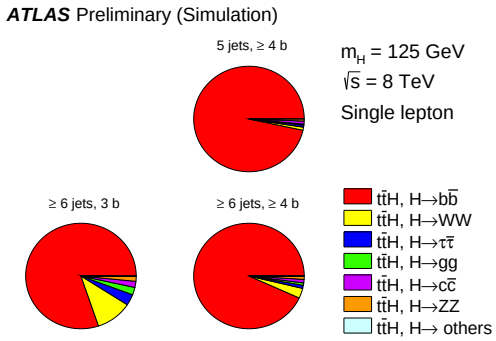


Figure 6.7: A plot showing the relative signal contributions from $t\bar{t}H$ production events based on dominant Higgs boson decay modes. Here a Higgs boson mass of 125 GeV and SM cross-sections and branching ratios are assumed. [2]

Chapter 7

The Matrix Element Method

The Matrix Element Method (MEM) was originally introduced as a means to accurately measure the top quark mass [77]. Following implementations at CDF and DØ [78–81], it became clear that, despite substantial technical hurdles, the MEM can serve as an excellent tool for analyzing collider based high-energy physics data on a number of topics from the Tevatron to the LHC.

7.1 Matrix Element Method Description

7.1.1 Introduction

The goal of the MEM is to utilize the most theoretical information possible [82] to infer aspects of a latent interaction given a measured final-state. To accomplish this goal the MEM is built upon the foundation of the Lorentz invariant formulation of Fermi's Golden Rule shown by Eq. 7.1. This rule defines the transition probability, Γ_{if}, from some initial-state i, to some final-state f. In so doing, Fermi's Golden Rule elegantly separates terms associated with quantum dynamics from those associated with initial and final-state kinematics:

$$\Gamma_{if} = (2\pi)^4 \cdot \prod_{i=1}^{n_i} \frac{1}{2E_i} \cdot \int d\boldsymbol{\Phi}_{\text{lips}} \cdot \left|\mathcal{M}_{if}\right|^2 . \tag{7.1}$$

In the above equation, $\left|\mathcal{M}_{if}\right|^2$, encodes the quantum mechanical effects during the interaction, whereas, $d\boldsymbol{\Phi}_{\text{lips}}$, encodes the density of states in the *Lorentz invariant phase space* measure and is defined by Eq. 7.2:

$$d\boldsymbol{\Phi}_{dlips} = \prod_{j=1}^{n_f} \frac{d^3\boldsymbol{p}_j}{(2\pi)^3\ 2E_j}\ \delta\left[\sum_i^{n_i} \boldsymbol{p}_i - \sum_j^{n_f} \boldsymbol{p}_j\right]\ \delta\left[\sum_i^{n_i} E_i - \sum_j^{n_f} E_j\right]. \tag{7.2}$$

Given the transition probability defined by Eq. 7.1, cross-sections for hypothesized interactions can be calculated. For example, the differential cross-section for some process, α, can defined as follows (assuming a center-of-mass frame where the incoming particle momenta, p_1 and p_2, obey $|p_1| = |p_2| = |p_i|$):

$$\boldsymbol{d}\sigma_\alpha\left(ab \to \boldsymbol{x}_f\right) = \frac{(2\pi)^4}{4|p_i|\sqrt{s}} \cdot \left|\mathcal{M}_\alpha\left(ab \to \boldsymbol{x}_f\right)\right|^2 \cdot d\boldsymbol{\Phi}_{dlips}. \tag{7.3}$$

Before making inferences about the process, α, we must take into account effects unique to the experimental apparatus. First, as described in Section 3.1, the Large Hadron Collider collided protons at a center of mass energy, $\sqrt{s}$, of 8 TeV. However, Eq. 7.3 is defined at the parton level. In other words, the initial-state participants, a and b, are fundamental particles unlike the protons being collided by the LHC. Thus, we must transform Eq. 7.3 from a parton level differential cross-section ($ab \to \boldsymbol{x}_f$) into a proton-proton level differential cross-section ($pp \to \boldsymbol{x}_f$). To do this we convolute Eq. 7.3 with a parameterized set of parton-density functions (PDFs). These PDFs contain the necessary information for all possible flavor and kinematic combinations of our initial-state, ab. This convolution is shown below, in Eq. 7.4:

$$\boldsymbol{d}\sigma_\alpha\left(pp \to \boldsymbol{x}_f\right) = \sum_{flavors} \int dx_a dx_b f_a(x_a, Q^2) f_b(x_b, Q^2)\, d\sigma_\alpha\left(ab \to \boldsymbol{x}_f\right). \quad (7.4)$$

In the above equation, the PDF, f_a, is a probability distribution function of the initial-state parton momentum fraction, x_a, the proton energy, $\sqrt{s}/2$, the parton flavor, a, and a momentum transfer scale, Q^2. A set of PDFs from the MSTW 2009 NLO [40, 41] collection are shown, in Figure 4.1, for two Q^2 configurations as an example.

In addition to transforming Eq. 7.3 into a pp differential cross-section, the final-state parton momenta must be convoluted with the anticipated detector response. To do this, a set of transfer functions are developed specifically modeling the ATLAS detector. These transfer functions provide a probability distribution of 'true' values, given the measured detector response, and can be applied to object momenta or directions. The transfer functions used in this analysis are parameterized in terms of transverse energy, E_T, and pseudo-rapidity, η, and are a part of the KLFitter software package [83].

A cartoon of the evolution of the $t\bar{t}H$ process and the terms describing each phase is illustrated in Figure 7.1. Taking each of these pieces into consideration, the matrix element technique can now be defined. Using the final pp differential cross-section from Eq. 7.4, convoluting it with the parameterized detector response for the various final-state objects, and choosing a suitable normalization condition, we can define a probability density function describing the likelihood that a reconstructed event, $\mathbf{x}_f$, originated from some process, α.

$$\mathcal{P}(\mathbf{x}_f|\alpha) = \frac{1}{\sigma_\alpha \epsilon A} \int d\mathbf{y}_f\; \boldsymbol{d}\sigma_\alpha\left(pp \to \boldsymbol{x}_f\right) W\left(\mathbf{y}_f|\mathbf{x}_f\right)\; d\boldsymbol{\Phi}_{dlips} \quad (7.5)$$

In Eq. 7.5, $W\left(\mathbf{y}_f|\mathbf{x}_f\right)$, represents a product of probability densities which map the reconstructed event four-vectors, $\mathbf{x}_f$, to potential 'true' values, $\mathbf{y}_f$, according to some modeled detector response, W. Also, $\sigma_\alpha\left(pp \to \boldsymbol{x}_f\right)$, is defined using PDFs as in Eq. 7.4. The probability is normalized by the expected total observed cross-section, σ_α, where the anticipated efficiencies, ϵ, and acceptances,

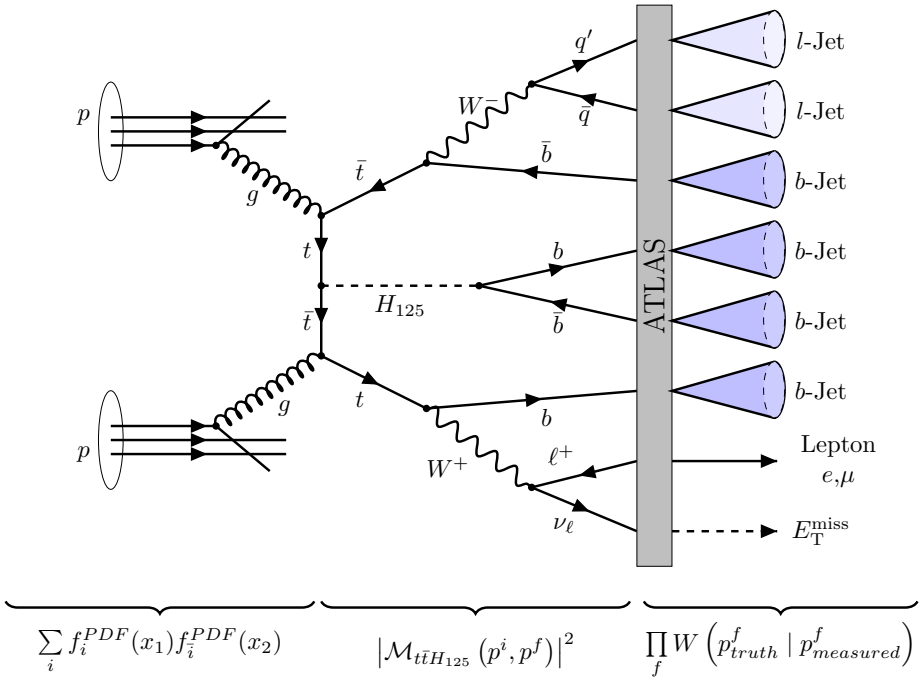


Figure 7.1: A diagram of the time evolution, from left to right, of a lepton-plus-jets $t\bar{t}H$ event as detected by the ATLAS detector. Below the figure, the corresponding term in the matrix element probability calculation is shown for the three primary phases. The first term contains the PDF information required to map from proton momenta to initial-state parton momenta. The second term, the Matrix Element, contains all the leading order information regarding the dynamics of the $t\bar{t}H$ scattering process. The final term maps the momenta of the final-state objects to expected measured momenta taking into account the ATLAS detector response.

A, are taken into account. This normalization scheme is chosen such that the condition defined in Eq. 7.6 holds true:

$$\int d\mathbf{x}_f \mathcal{P}(\mathbf{x}_f|\alpha) = 1. \tag{7.6}$$

With this probability developed, a model based likelihood can be defined for a given sample of events. In many analyses these probabilities are determined under multiple model hypotheses defined by some parameter of interest. Of recent note is the work done by the CDF and DØ collaborations to determine the precise mass of the top quark [78, 79]. Sample likelihoods are parameterized by hypothesized top quark masses, described in Eq. 7.7. The maximum likelihood parameterization is then chosen as the most likely top quark mass:

$$\mathcal{L}^{\text{Sample}}(M_T|\bar{\mathbf{x}}_f) = \prod_{i=1}^{\text{Events}} \mathcal{P}(\mathbf{x}_f^i|M_T). \tag{7.7}$$

Due to the success of this method in $t\bar{t}$ channels at the Tevatron and at the LHC, it is interesting to consider that the MEM could be adapted to Higgs searches in associated production modes and specifically for the lepton-plus-jets $t\bar{t}H$ ($H \to b\bar{b}$) channel. The richness of this final-state, in addition to the relatively high number of intermediate propagators in its associated matrix element motivates the basis of this analysis, that the matrix element method can be adapted into a powerful multivariate classifier. The following section explores this approach in more detail.

7.1.2 The Matrix Element Method as a Classifier

For this analysis, the MEM is utilized as an event classifier, as opposed to a parameter estimation technique. The differences in the scattering dynamics between the signal and irreducible background events arising from the Higgs propagator provide essentially the only handle for discrimination between $t\bar{t}+b\bar{b}$ and $t\bar{t}H$ events. Given the method's sensitivity to the leading order scattering dynamics, it is deemed worthwhile to investigate its potential as an event discriminant.

The MEM utilizes all model dependent information, *a priori*, to classify candidate events. In other words, the power of the MEM comes from leveraging the underlying model-dependent physics to discriminate signal events from background events. It does this while utilizing all the event kinematic information contained in the object four-vectors to generate a single probability value. This is an especially powerful feature in the context of analyses with complex final-states like $t\bar{t}H$. This is because the multiplicity of final-state objects factorially increases the number of possible kinematic variables on which to train a learning-based MVA. This requires the analyst to identify and motivate an optimal subset of those variables. Furthermore, unlike most other learning-based

MVAs, there is no explicit requirement for a simulated training sample, making the MEM simulation independent. In the present case however, we do incorporate functions to model the anticipated detector response which are based on simulated data. These functions are described in Section 7.2.3.

As the name would suggest, the goal of any classifier is utilize the available information to classify the input into one or more categories. As is often the case, instead of explicitly classifying an input as a member of a discrete category, most MVAs map elements to the real line such that an optimal cut can be determined which minimizes the amount of contamination of mis-classified inputs. In our analysis, the leading priority of a MEM based discriminant is to separate $t\bar{t}H$ events from $t\bar{t} + b\bar{b}$ events in our primary signal regions. This helps on two fronts. First, it improves the signal to background ratio in the 'signal rich' analysis regions. Secondly, it allows for a more precise estimation of the $t\bar{t} + b\bar{b}$ background. To this end, probabilities are calculated for the same event under different process hypotheses, $t\bar{t}$+$b\bar{b}$ and $t\bar{t}H$. These probabilities are then used to classify each event based on the Neyman-Pearson *lemma* [84] ratio of the $t\bar{t} + b\bar{b}$ and $t\bar{t}H$ probabilities on a per-event basis. This value is commonly referred to as an *event probability discriminant* (EPD). In general, two representations of this information are used by this analysis, $\mathrm{R_{ME}}$, and $\mathrm{EPD_{ME}}$. Both of these representations are defined below, using Eq. 7.5, and have features which lend themselves to different uses within the analysis:

$$\mathrm{R_{ME}}\left(\mathbf{x}_f \mid t\bar{t}H, t\bar{t} + b\bar{b}\right) = \frac{\mathcal{P}(\mathbf{x}_f|t\bar{t}H)}{\mathcal{P}(\mathbf{x}_f|t\bar{t} + b\bar{b})} \tag{7.8}$$

$$\mathrm{EPD_{ME}}\left(\mathbf{x}_f \mid t\bar{t}H, t\bar{t} + b\bar{b}\right) = \frac{\mathcal{P}(\mathbf{x}_f|t\bar{t}H)}{\mathcal{P}(\mathbf{x}_f|t\bar{t}H) + \mathcal{P}(\mathbf{x}_f|t\bar{t} + b\bar{b})}. \tag{7.9}$$

Finally, as the goal of this is analysis is improve sensitivity to the presence of the $t\bar{t}H$ signal, the ultimate figure of merit is the value of and uncertainty on the signal strength parameter, μ, defined by Eq. 4.14 in Section 4.1.5. Apart from the technical difficulties associated with determining this value, another impediment is that this value is defined only for the full data sample. Thus, for testing incremental improvements, a proxy is developed as a means of measuring classification performance and quantifying the discriminating power. As the sensitivity of the $t\bar{t}H$ measurement is roughly dependent on the relative purities in each bin, a measure which quantifies the degree of purity across the bins of a distribution would serve as a reasonable proxy. This relationship is captured in relative terms by a variable's *separation* power, shown by Eq. 7.10. Here we take normalized and binned signal, S_V, and background, B_V, distributions for some discriminating variable, V, and compute the normalized fraction of

non-overlapping area, $\mathcal{S}$:

$$\mathcal{S}_V = \frac{1}{2} \sum_{i}^{\text{bins}} \left| S_V^i - B_V^i \right|. \tag{7.10}$$

where the signal, S, and background B, distributions are normalized to unity:

$$\sum_{i}^{\text{bins}} B_V^i = 1, \qquad \sum_{i}^{\text{bins}} S_V^i = 1. \tag{7.11}$$

7.1.3 Technical Hurdles

The primary drawback to using the MEM to study complex final-states is that the method is exceedingly challenging from a technical standpoint. To perform the calculation defined in Section 7.1.1, many millions of multi-dimensional integrations must be performed over an extremely complex function which itself takes many CPU cycles to complete. In fact, the first implementation that was attempted took more than 24 hours to complete for a single event. As this calculation is to run over approximately 67 million events, it is clear that a tremendous amount of effort would have to be directed at optimizing the computational aspects of the MEM to make it feasible. Further development later brought the 24h/event run time down to 30 minutes per event, and eventually to 1.7 minutes per event, nearly three orders of magnitude improvement. Once the run times dropped below 30 minutes per event, effort was further placed into optimization of the classification power.

The initial phase of development focuses on decreasing the absolute run-time to a goal of 5 million CPU hours for the entire analysis, corresponding to a little less than 5 minutes per event. Given the nature of the MEM, two avenues are available. The first avenue is to simply decrease the number of times needed to call the matrix element calculation itself. This is the most broad avenue to take as improvements run the gamut from tightening the event selection so that fewer events need to be processed, to improving the convergence rate of the adaptive Monte Carlo (MC) integration technique. The second avenue is to improve the run-time performance of the matrix element calculation itself. This avenue ultimately provides the most substantial gains to our calculation run times.

The second phase of development is focused on improving the classification power of the technique without increasing the per-event run-times beyond what is technically feasible for the entire dataset. The avenues available here are to determine the separation power as a function of the jet-assignment pre-selection and as a function of the various approximations used to speed up the calculation. A more detailed explanation of how these improvements are achieved can be found in Section 7.3.

7.2 Implementation

It was clear from initial testing that existing MEM software wasn't up to the task of handling complex final-states without extensive modification. To push ahead independently, we started with an existing C++ based framework named MEMTool . This framework was initially developed for $t\bar{t}$ mass measurements, but was sufficiently extensible to adapt to the $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ processes. As an aside, it is important to note that, in the time since this our efforts first began, considerable progress has been made in this area by various other groups, parallel to our work.

In broad terms, MEMTool is designed to interface with multiple external shared libraries for various elements of the calculation. The GSL 1.16 [85] implementation of the VEGAS adaptive Monte-Carlo integration technique [86] is utilized as the numerical integration library for a number of reasons. First, it's optimization algorithm is very well defined and can be exploited via carful variable transformations that are similarly well defined in the context of the matrix element integrand. Second, it is widely used, validated, and documented. Finally, it performs very well while simultaneously being trivially easy to implement in the MEMTool framework. The PDF calculations are performed by the LHAPDF 5.9.1 [87] library which utilizes the CTEQ6 [58] parameterization. Though a variety of input and output formats are possible with the framework, this implementation exclusively utilizes ROOT 5.34/14 [88] for handling the file IO. Finally, methods generated by MadGraph 5 [89] are included to handle the explicit calculation of the process' squared matrix elements. Descriptions of these processes can be found in the following sections.

7.2.1 Signal Hypothesis Description

This analysis uses MadGraph 5 to generate the software which calculates the squared matrix elements for a given set of four vectors. The specific process used for the signal hypothesis is defined as a 125 GeV Standard Model (SM) Higgs boson, consistent with current observations from ATLAS [90] and CMS [91], produced in association with two top quarks.The top quarks are required to decay into the lepton-plus-jets channel, and the Higgs boson is required to decay to a pair of b-quarks. This results in a final-state with 8 total objects, four b-quarks, 2 light quarks, a lepton and a neutrino. Exploiting the charge conjugation and lepton universality symmetries of the SM allows us to limit the contributing diagrams to only those with an electron in the final-state. In a similar manner, diagrams where the Higgs boson radiates from one of the Ws are neglected, this particular constraint allows for more straight forward kinematic transformations without significantly impacting the calculation.

This process, once explicitly defined, allows MadGraph to generate a C based library containing the necessary functions for the calculation. Software written for MEMTool then imports these functions into the framework for

dynamic use at runtime. The contributing diagrams to the signal process defined above can be viewed in Figure 7.2.

7.2.2 Background Hypothesis Description

In a similar manner to the signal process defined in the previous section, the background hypothesis process is incorporated into the MEMTool framework. This background hypothesis is defined as a pair of top quarks produced in association with two additional b-quarks. This constitutes the final-state topology of our irreducible background $t\bar{t} + b\bar{b}$. The top quarks are similarly required to decay into the lepton-plus-jets channel as in the case of the signal process. However, the b-quarks are allowed to be produced via gluon splitting or via radiative processes, significantly increasing the number of required diagrams. These diagrams can be viewed in Figures 7.3 and 7.4.

7.2.3 Detector Response

A crucial aspect of the MEM is the fact that the squared matrix element, $|\mathcal{M}(ij \to X)|^2$ it is defined at the parton level. This is distinct from what is ultimately measured by the ATLAS detector. Objects which participate in the scattering dynamics defined by the squared matrix element, quarks in particular, are frequently altered by subsequent interactions during hadronization and by final-state radiation (FSR). Further, the act of measuring introduces further distortions due the finite resolutions inherent to any detection apparatus. To *transfer* the 'measured' four-momenta to the 'parton-level' four-momenta, a *transfer function* is defined which models the anticipated smearing effects unique to the detection apparatus. As described in Section 7.1.1, the expected detector response is approximated by a set of functions defined for each object class over different ranges of the ATLAS detector in η, mirroring the detector geometry.

Due to the extensible nature of the MEMTool framework, it is a simple task to import existing descriptions of the ATLAS detector response for use. These functions were initially designed for use with a kinematic-likelihood method named KLFitter [83]. The transfer functions themselves are parameterized separately for each object in η and are modeled with double Gaussian functions in δE, defined by Eq. 7.13. An example jet-energy transfer function is defined in Eq. 7.12. Here, the parameters, p_i, are fit given a clean sample of simulated events with well defined jet energies which are then unambiguously matched to the quarks present in the generated particle record. The functional form of an example transfer function is shown in Figure 7.5.

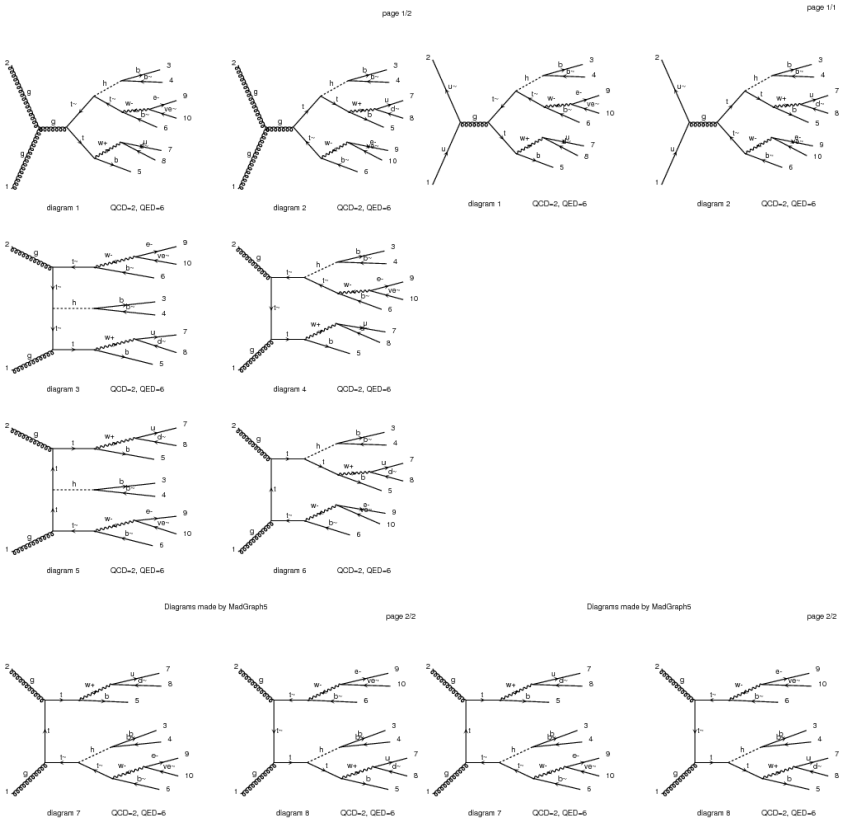


Figure 7.2: An illustration of the 12 contributing Feynman diagrams to the $t\bar{t}H$ process which defines the signal model. Here, the 10 gluon fusion diagrams and 2 quark annihilation diagrams describe the leading order picture of the $t\bar{t}H \to b\bar{b}+b\bar{b}+\ell\nu+qq'$ final-state topology. These diagrams, and their associated matrix element calculation functions, are created with MadGraph 5.

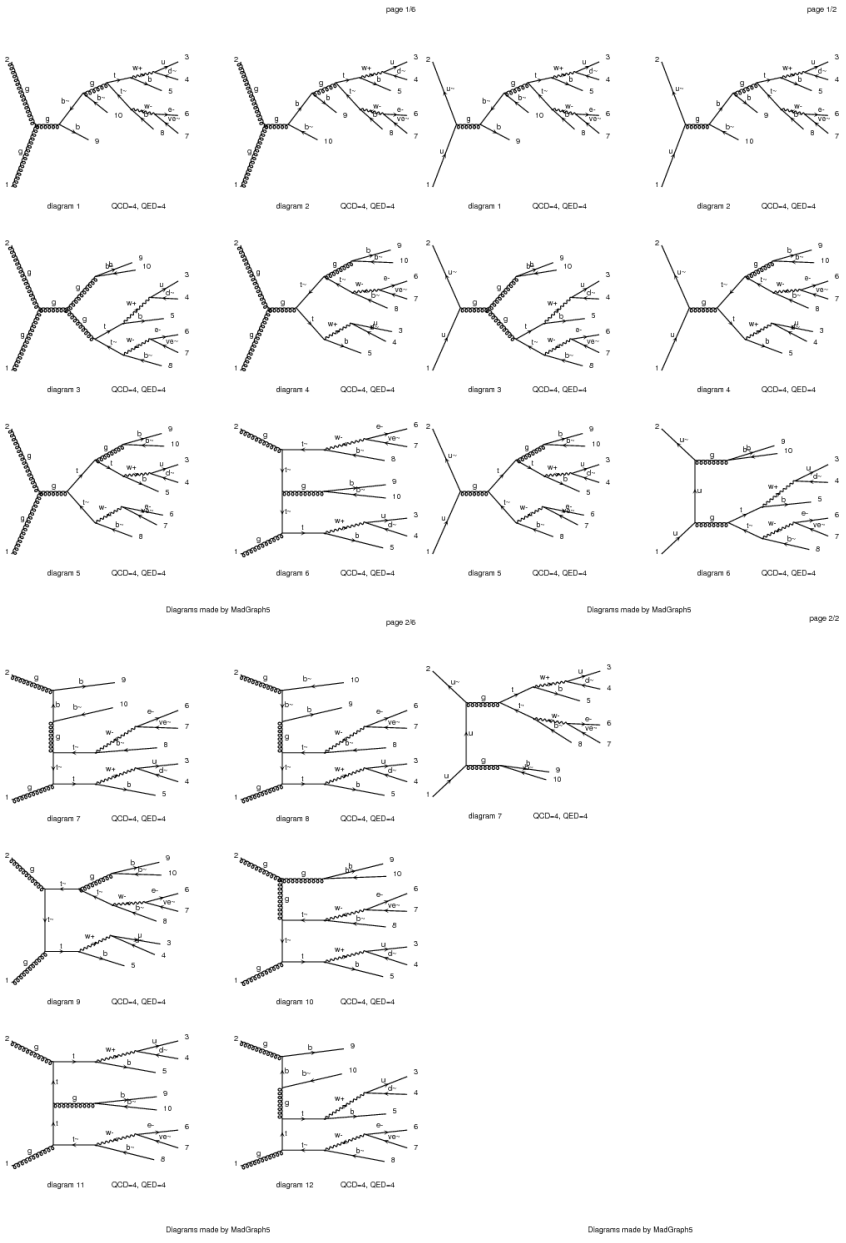


Figure 7.3: An illustration of a subset of Feynman diagrams which contribute to the $t\bar{t}b\bar{b}$ process which defines the background model. Here, 19 of the 40 total diagrams are shown. These diagrams describe the leading order picture of the $t\bar{t} + b\bar{b} \to b\bar{b} + b\bar{b} + \ell\nu + qq'$ final-state topology. These diagrams, and their associated matrix element calculation functions, are created with MadGraph 5.

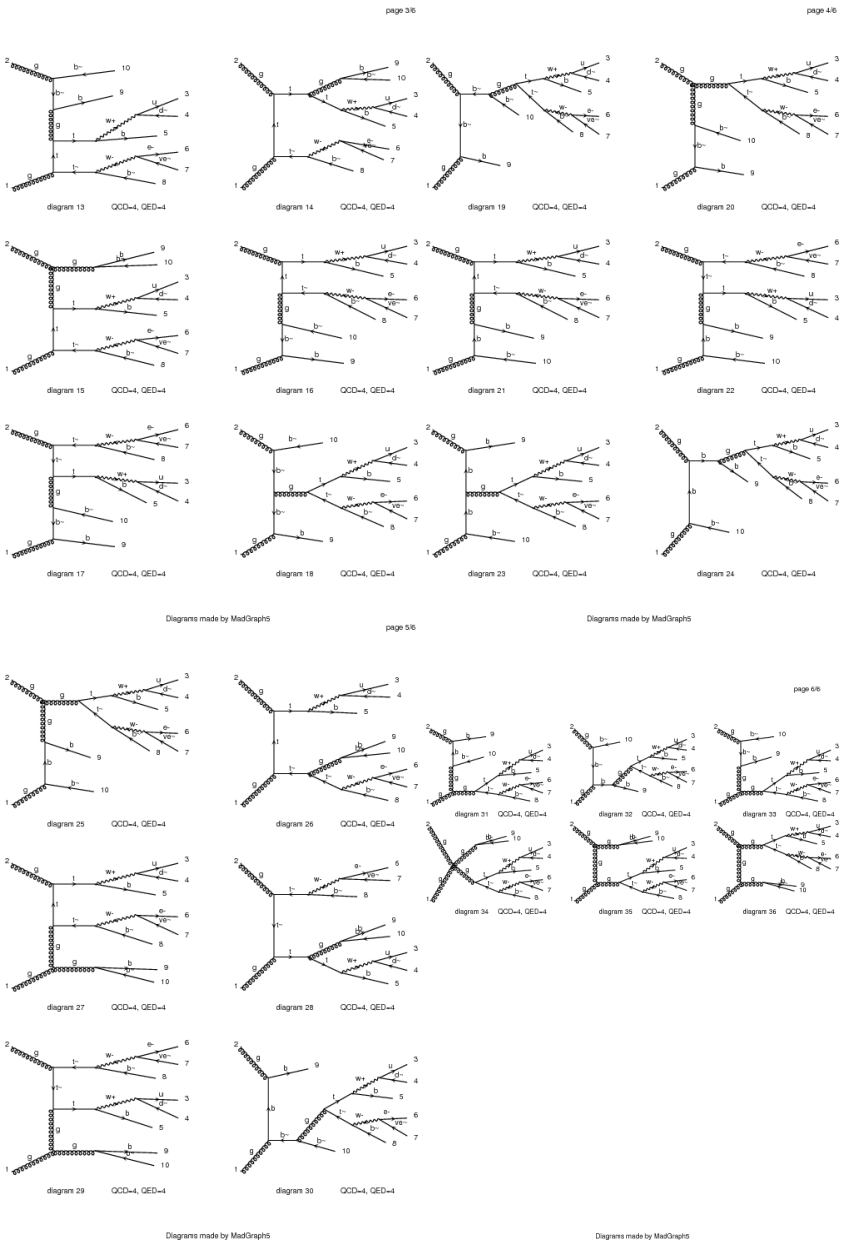


Figure 7.4: An illustration of a subset of Feynman diagrams which contribute to the $t\bar{t}b\bar{b}$ process which defines the background model. Here, the remaining 21 of the 40 total diagrams are shown. These diagrams describe the leading order picture of the $t\bar{t} + b\bar{b} \rightarrow b\bar{b} + b\bar{b} + \ell\nu + qq'$ final-state topology. These diagrams, and their associated matrix element calculation functions, are created with MadGraph 5.

Detector Response Configuration

Description		**Parameters**			
Parent	**Object**	E	η	ϕ	M
t_{Had}	b-Jet$_1$	W_b	δ	δ	δ
t_{Lep}	b-Jet$_2$	W_b	δ	δ	δ
H, g	b-Jet$_3$	W_b	δ	δ	δ
H, g	b-Jet$_4$	W_b	δ	δ	δ
$W_{q\bar{q}'}$	l-Jet$_1$	W_l	δ	δ	δ
$W_{q\bar{q}'}$	l-Jet$_2$	W_l	δ	δ	δ
$W_{\ell\nu}$	ℓ	δ	δ	δ	δ
$W_{\ell\nu}$	ν	W_ν	δ	δ	δ

Table 7.1: Configuration of the applied transfer functions at the final iteration of this MEM analysis. Here, δ's denote delta functions symbolizing the expectation that this quantity is well measured or known *a priori*, as with an analytical solution. The W_x's denote transfer functions parameterized for a given general object x.

$$W(E_p|E_m) = \frac{1}{\sqrt{2\pi}\,(p_2 + p_3 p_5)} \left(e^{\frac{-(\delta E - p_1)^2}{2p_2^2}} + p_3 e^{\frac{-(\delta E - p_4)^2}{2p_5^2}} \right) \tag{7.12}$$

$$\delta E = \frac{E_{\text{[parton-level]}} - E_{\text{[measured]}}}{E_{\text{[parton-level]}}} \tag{7.13}$$

It should be noted that each transfer function not explicitly defined as a delta function effectively increases the dimensionality of the integrand by one. The computation penalty for adding additional integration dimensions is significant. Thus, we adopt a familiar starting point among MEM analyses by transferring object directions with delta functions, effectively removing them from the integral phase space, as they are comparatively well measured. Furthermore, the lepton momentum is similarly determined to be well measured and is therefore transferred with a delta function. This effectively constrains our calculation to a 6 dimensional integrand with one transfer function defined for each jet energy. The chosen transfer functions for the MEM analysis are shown in Table 7.1.

Finally, in addition to providing the detector response description, the Gaussian response in jet-E provides a reasonable method by which to restrict the integration phase space. Initially, windows of 5σ, centered at the peak, were used. However, this was found to be sub-optimal in cases where contributions from the matrix-element propagators and contributions from the transfer-functions are sufficiently separated so as to result in a multi-modal integrand. To handle this situation, the integration windows are always expanded to incorporate the associated peaks coming from the matrix-element, when they happen to fall out

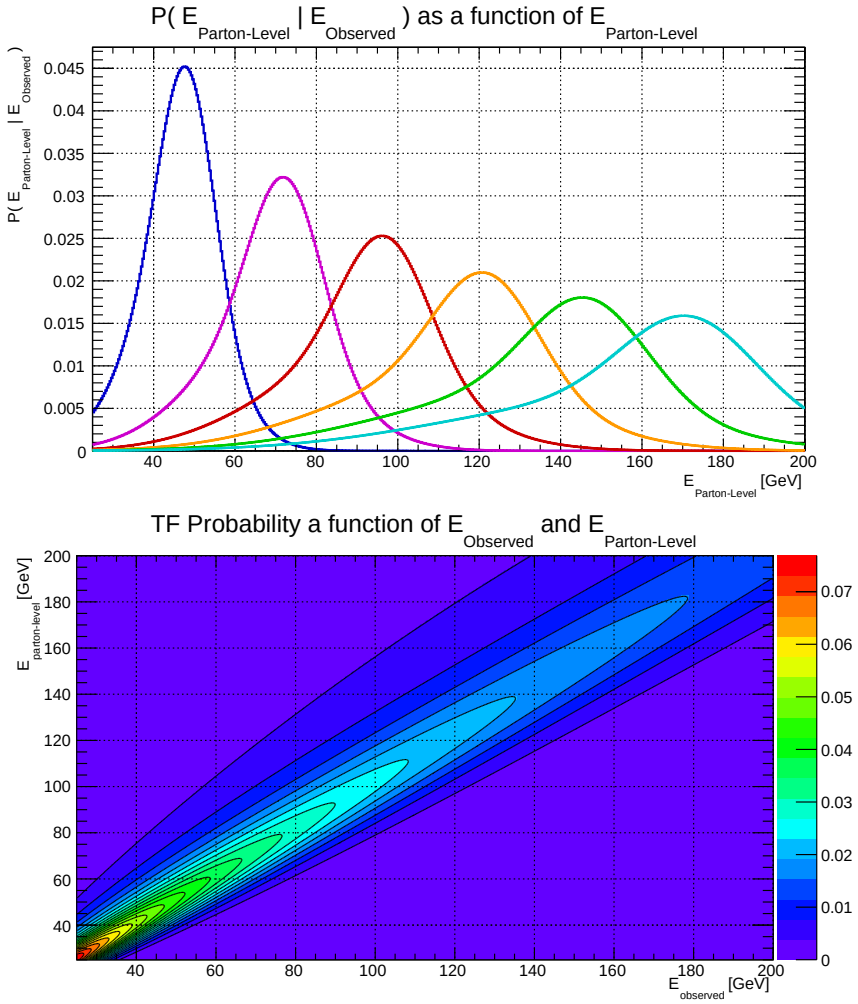


Figure 7.5: The modeled detector response as a function of the parton-level and observed jet energies for a b-jet with a pseudo-rapidity of $|\eta| \leq 0.8$. The top plot shows the probability (y-axis) that a jet of an observed energy, E_{observed}, could have originated from a parton with a different energy, $E_{\text{parton-level}}$ (x-axis). This is shown for 6 different observed jets of energies: 50 GeV (blue), 75 GeV (magenta), 100 GeV (red), 125 GeV (orange), 150 GeV (green), and 170 GeV (cyan). The second plot shows the 2 dimensional contours of the probability distribution function as a function of both the observed and parton-level jet energies.

of the nominal 5σ window.

7.2.4 Jet-Assignment Handeling

For every event, we calculate the event probability under a given hypothesized process for 12 unique jet-assignments. The algorithm designed to choose these 12 is detailed in Section 7.3.3. As our goal is to define an *event* level discriminant, we must identify a way with which to combine the jet-assignment probabilities. Our tests show that the arithmetic mean of the probabilities provide the best separation. However, we also compute all discriminating values with the maximum jet-assignment probability as well, for comparison sake. In all circumstances, within statistical uncertainties, the arithmetic mean performes better than the maximum jet-assignment probability, yet performance of the two formulations remains very consistent throughout the analysis. Definitions used in the formulation of our results are shown below.

For Eq. 7.14 and Eq. 7.15, $\mathrm{R}_{\mathrm{ME}}^{\mathrm{Avg}}$ and $\mathrm{EPD}_{\mathrm{ME}}^{\mathrm{Avg}}$, are the two discriminating variables introduced earlier, now defined using the arithmetic mean of the jet-assignment probabilities of the signal and background hypotheses. Using this notation, the jet-assignments are defined where γ_{p} is the p-th permutation, generating a set of jet-assignments P, and $\mathbf{x}_f^0$ represents the non-permuted, measured final-state four-momenta. Thus, $\mathbf{x}_f^{\mathrm{p}} = \gamma_{\mathrm{p}}\mathbf{x}_f^0$ where $\gamma_{\mathrm{p}}\mathbf{x}_f^0 \in \mathrm{P}$ and $\gamma_0 = \mathbb{I}$:

$$\mathrm{R}_{\mathrm{ME}}^{\mathrm{Avg}}\left(\mathbf{x}_f^0 \mid t\bar{t}H, t\bar{t}b\bar{b}\right) = \frac{\frac{1}{P}\sum_{\mathrm{p}=1}^{\mathrm{P}} \mathcal{P}(t\bar{t}H|\mathbf{x}_f^{\mathrm{p}})}{\frac{1}{P}\sum_{\mathrm{p}=1}^{\mathrm{P}} \mathcal{P}(t\bar{t}b\bar{b}|\mathbf{x}_f^{\mathrm{p}})} \tag{7.14}$$

$$\mathrm{EPD}_{\mathrm{ME}}^{\mathrm{Avg}}\left(\mathbf{x}_f^0 \mid t\bar{t}H, t\bar{t}b\bar{b}\right) = \frac{\frac{1}{P}\sum_{\mathrm{p}=1}^{\mathrm{P}} \mathcal{P}(t\bar{t}H|\mathbf{x}_f^{\mathrm{p}})}{\frac{1}{P}\sum_{\mathrm{p}=1}^{\mathrm{P}} \mathcal{P}(t\bar{t}H|\mathbf{x}_f^{\mathrm{p}}) + \frac{1}{P}\sum_{\mathrm{p}=1}^{\mathrm{P}} \mathcal{P}(t\bar{t}b\bar{b}|\mathbf{x}_f^{\mathrm{p}})}. \tag{7.15}$$

For Eq. 7.16 and Eq. 7.17, the discriminating variables $\mathrm{R}_{\mathrm{ME}}^{\mathrm{Max}}$ and $\mathrm{EPD}_{\mathrm{ME}}^{\mathrm{Max}}$ are computed using the maximum probability jet-assignment from the set of jet-assignment probabilities, $\mathcal{P}_{\mathrm{p}}$; $\mathrm{p} \in \mathrm{P}$, under both signal and background hypotheses. Here, the same permutation notation is used as above, where $\mathbf{x}_f^{\mathrm{p}} = \gamma_{\mathrm{p}}\mathbf{x}_f^0$ and $\gamma_{\mathrm{p}}\mathbf{x}_f^0 \in \mathrm{P}$:

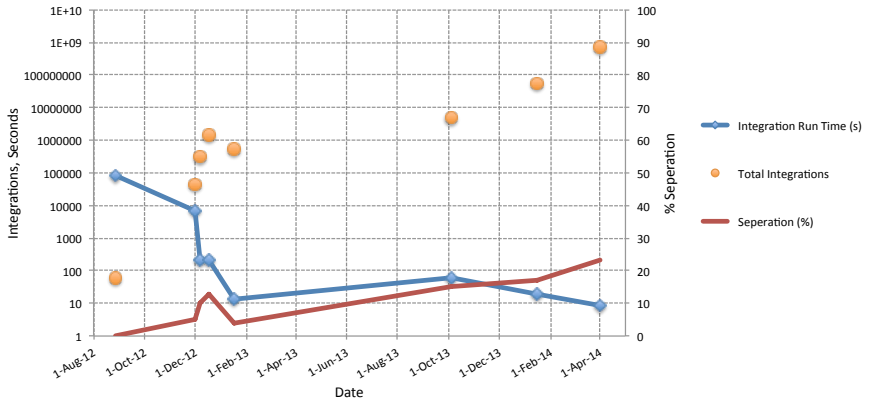


Figure 7.6: Improvement in run time and separation power of the MEM discriminating variables as a function of development time. The blue curve shows the improvement in per-event run times using the logarithmic scale on the left in *seconds*. The red curve shows the improvement in separation power using the linear scale on the right in sepration *percentage*, were separation is defined by Eq. 7.10. Finally, the orange points show the total number of integrations performed at each stage of development using the logarithmic scale on the left in *integrations*.

$$\mathrm{R}^{\mathrm{Max}}_{\mathrm{ME}}\left(\mathbf{x}^0_f \mid t\bar{t}H, t\bar{t}b\bar{b}\right) = \frac{\max\limits_{\mathcal{P}_{\mathrm{p}};\ \mathrm{p}\in\mathrm{P}} \mathcal{P}(t\bar{t}H|\mathbf{x}^{\mathrm{p}}_f)}{\max\limits_{\mathcal{P}_{\mathrm{p}};\ \mathrm{p}\in\mathrm{P}} \mathcal{P}(t\bar{t}b\bar{b}|\mathbf{x}^{\mathrm{p}}_f)} \tag{7.16}$$

$$\mathrm{EPD}^{\mathrm{Max}}_{\mathrm{ME}}\left(\mathbf{x}^0_f \mid t\bar{t}H, t\bar{t}b\bar{b}\right) = \frac{\max\limits_{\mathcal{P}_{\mathrm{p}};\ \mathrm{p}\in\mathrm{P}} \mathcal{P}(t\bar{t}H|\mathbf{x}^{\mathrm{p}}_f)}{\max\limits_{\mathcal{P}_{\mathrm{p}};\ \mathrm{p}\in\mathrm{P}} \mathcal{P}(t\bar{t}H|\mathbf{x}^{\mathrm{p}}_f) + \max\limits_{\mathcal{P}_{\mathrm{p}};\ \mathrm{p}\in\mathrm{P}} \mathcal{P}(t\bar{t}b\bar{b}|\mathbf{x}^{\mathrm{p}}_f)}. \tag{7.17}$$

7.3 Major Areas of Optimization

From the outset, the significant technical huddles introduced in Section 7.1.3 put the focus of our technical efforts squarely on improving the MEM CPU run times. In the context of our chosen adaptive numerical integration technique, VEGAS [86], the following sections present the results of our optimization and approximation studies. In addition, improvements of both run times and separation, defined in Eq. 7.10, are shown as a function of development time in Figure 7.6.

7.3.1 Dimension Reduction

Matrix element method runtimes are most sensitive to the dimensionality and relative size of the definite integral's phase space. This is not surprising as the 'Curse of Dimensionality' has hung over numerical integration methods since

their inception. Though Monte-Carlo integration methods significantly improve convergence rates beyond naïve cubature methods, this improvement is relative with respect to integrals with many more than 4 dimensions. In other words, performance improvements will be compounded with each dimension one can eliminate. Thus, the first step in any future MEM analysis is to identify integration dimensions that provide little discriminaton and to remove them.

In the case of this analysis, as shown in Table 7.1, parameters that are measured well are not convoluted with a modeled detector response. This is crucial as the $t\bar{t}H$ and $t\bar{t} + b\bar{b}$ matrix elements are functions of 40 different input parameters. These input parameters are made up of the four-momenta of the two initial-state objects, in addition to the four-momenta of the 8 final-state objects. Further, integration dimensions of the initial-state can be eliminated based on the assumption that the initial-state p_T^i is equal to zero, in addition to the requirements of conservation of energy and momentum, $\sum \mathbf{P_i} - \sum \mathbf{P_f} = 0$. Finally, we set the masses to zero for all objects except for the b-jets. As shown in Table 7.2, this leaves only 7 remaining dimensions to integrate. These dimensions remain as they either experience non-trivial detector responses, as in the case of the jet-energies, or because they are not measured at all, as is the case for the neutrino p_z.

7.3.2 Kinematic Variable Transformation

Once the optimal set of dimensions has been determined, the time comes to optimize the integrand. This is important due to the behavior of the VEGAS Monte-Carlo (MC) integration technique, which can be exploited. In general, MC techniques are based on the fact that an integral of some function, $f(\mathbf{x})$, over a definite interval, Φ, is equal to the volume of that interval, V_Φ, times the mean value of $f(\mathbf{x})$ in Φ, shown in Eq. 7.20.

$$I = \int_\Phi d\mathbf{x}\ f(\mathbf{x}) \tag{7.18}$$

$$I = \langle f(\mathbf{x}) \rangle_\Phi \cdot \int_\Phi d\mathbf{x} \tag{7.19}$$

$$I = \langle f(\mathbf{x}) \rangle_\Phi \cdot V_\Phi \tag{7.20}$$

Exploiting the *law of large numbers*, one can then determine the mean value of $f(\mathbf{x})$ by repeatedly sampling $f(\mathbf{x})$ at random points, $\mathbf{x}_i$, in Φ, shown in Eq. 7.22.

Generic Squared Matrix Element Phase Space

Description		**Parameters**			
Parent	**Object**	p_x	p_y	p_z	M GeV
p_1	i_1	$p_T^i = 0$	$p_T^i = 0$	$\sum E_i = \sum E_f$	0.0
p_2	i_2	$p_T^i = 0$	$p_T^i = 0$	$\sum p_z = 0$	0.0
—	—	E	η	ϕ	M GeV
t_{Had}	b-Jet$_1$	$\mathbf{d_1}$	δ	δ	4.7
t_{Lep}	b-Jet$_2$	$\mathbf{d_2}$	δ	δ	4.7
H, g	b-Jet$_3$	$\mathbf{d_3}$	δ	δ	4.7
H, g	b-Jet$_4$	$\mathbf{d_4}$	δ	δ	4.7
$W_{q\bar{q}'}$	l-Jet$_1$	$\mathbf{d_5}$	δ	δ	0.0
$W_{q\bar{q}'}$	l-Jet$_2$	$\mathbf{d_6}$	δ	δ	0.0
$W_{\ell\nu}$	ℓ	δ	δ	δ	0.0
—	—	p_x	p_y	p_z	M GeV
$W_{\ell\nu}$	ν	$\sum p_x = 0$	$\sum p_y = 0$	$\mathbf{d_7}$	0.0

Table 7.2: Pre-transformation configuration of the integration phase space. This table displays each of the 40 possible integration dimensions, and the assumption used to eliminate them from consideration. Dimensions with a δ signify those parameters which have been well measured and need no detector response convolution. The assumption that the initial-state transverse momenta is zero, $p_T^i = 0$, is used. In addition, constraints from momentum conservation, $\sum \mathbf{p} = 0$, further limit the available phase space. All these constraints and approximations leave $\mathbf{d_{1-7}}$ as the remaining integration phase space dimensions.

$$I = \langle f(\mathbf{x}) \rangle_{\Phi} \cdot V_{\Phi} \tag{7.21}$$

$$I \approx I_N^{MC} = \frac{V_{\Phi}}{N} \sum_{i=1}^{N} f(\mathbf{x}_i) \tag{7.22}$$

$$\lim_{N \to \infty} I_N^{MC} = I \tag{7.23}$$

where $\mathbf{x_i}$ are randomly drawn from Φ, and are therefore statistically independent:

$$\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \cdots, \mathbf{x}_N\} \in \Phi. \tag{7.24}$$

The expected uncertainty and subsequent convergence rate of the approximate MC solution, I_N^{MC}, with respect to the exact value of the integral is therefore determined by the variance of the sampled points using an unbiased estimate of the variance, σ_N. This yields an uncertainty which decreases by $\frac{1}{\sqrt{N}}$, where N, is the number of samples, shown in Eq. 7.25:

$$\delta I_N^{MC} = V_{\Phi} \frac{\sigma_N}{\sqrt{N}}. \tag{7.25}$$

VEGAS improves on the $\frac{1}{\sqrt{N}}$ convergence rate by utilizing an adaptive sampling method based on *importance sampling*. The principal behind *importance sampling* is shown in Eq. 7.26. The concept is that instead of sampling points from a uniform distribution over Φ, samples are pulled from a probability distribution, $p(\mathbf{x}_i)$, defined in Eq. 7.27. When samples are pulled from, $p(\mathbf{x}_i)$, the variance, and hence the uncertainty will vanish. This sounds silly at first as $p(\mathbf{x}_i)$ is defined using f and its definite integral over Φ, the very thing we are attempting to calculate. However, most *importance sampling* routines seek to simply use previous samples to approximate $p(\mathbf{x}_i)$ from which to pull future samples. This effectively results in more samples being pulled from regions of phase space with larger contributions, hence the etymology of *importance sampling* where more important regions are sampled comparatively more.

$$I \approx I_N^{MC} = \frac{V_{\Phi}}{N} \sum_{i=1}^{N} \frac{f(\mathbf{x}_i)}{p(\mathbf{x}_i)} \tag{7.26}$$

$$p(\mathbf{x}) = \frac{f(\mathbf{x})}{\int_{\Phi} f(\mathbf{x})} \tag{7.27}$$

The VEGAS algorithm seeks to approximate $p(\mathbf{x}_i)$ as the calculation proceeds, by storing projections of the integrand along each integration dimension for a total of kd regions, where k is the number of bins in the projection histogram and d is the number of dimensions. This is distinct from a cubature

approach which would need to store k^d regions, however, this means that VEGAS is optimal for functions which are nearly separable along each integration dimension and sub-optimal when the function is inseparable. An example of this sub-optimal behavior is shown in Figure 7.7 and in Figure 7.8 for an integration over the $t\bar{t}H\left(H \to b\bar{b}\right)$ squared matrix element. Here, the projections of $p(\mathbf{x}_i)$, are displayed at five different points during the integration, from top to bottom, in Figure 7.7. The left column displays the projections of the integrand as a function of the leading Higgs daughter b-jet energy. The middle column displays the projections of the integrand as a function of the sub-leading Higgs daughter b-jet energy. And the rightmost column displays the 2 dimensional projection of the integrand as a function of both Higgs daughter b-jet energies. As can be clearly seen, the region of importance is not aligned to either of these integration dimensions. As a result, VEGAS frequently samples the integrand in regions which do not contribute, effectively wasting CPU time. This is revealed by comparing Figure 7.8, which shows where VEGAS chose to sample, with the last row of Figure 7.7, which shows the important regions of the integrand. It is precisely this sub-optimal behavior which motivates the need for kinematic variable transformations.

The approach, described by Table 7.2, shows the parameters of interest as integration dimensions, in this case the jet energies. We know, however, that the functional form of our squared matrix element will be sharply peaked in regions of phase space in the vicinity of intermediate particle resonances. In other words, for the $t\bar{t}H_{125}$ process, when the two b-jets associated with the Higgs have an invariant mass in the vicinity of 125 GeV, we expect a much larger squared matrix element value than when this is not the case. Further, we know that the functional form of this peak, with respect to the two jet energies, is not separable, and is sub-optimal, as can be seen in Figure 7.7. However, with a variable transformation on the integration phase space, we can integrate over the Higgs mass, and the leading jet energy, a far more optimal configuration.

After considerable development, the kinematic transformation used for the remainder of this analysis was chosen due to its combination of excellent CPU performance and discriminating power. The seven-dimensional phase space of six jet-energies and the neutrino p_z, described by Table 7.2, is transformed into a six dimensional phase space described by Table 7.3. Here, the neutrino p_z integration is removed via the use of a *narrow width approximation* (NWA) which constrains the p_z to a value consistent with a parent W mass of 80.4 GeV. The remaining six dimensions are then broken up into pairs based on their parent objects. The two b-jets, originating from the two top quarks, are effectively left alone providing the $\mathbf{E_1}$ and $\mathbf{E_2}$ dimensions in Table 7.3. The pair of b-jets originating from the Higgs or gluon, g, depending on the hypothesized process, initially provides two dimensions over their respective energies, $\mathbf{E_3}$ and $\mathbf{E_4}$. Following the transformation, the sub-leading (lower p_T) b-jet energy dimension, $\mathbf{E_4}$, is converted into an integration over $\mathbf{M_X^2}$, the mass of the Higgs or gluon,

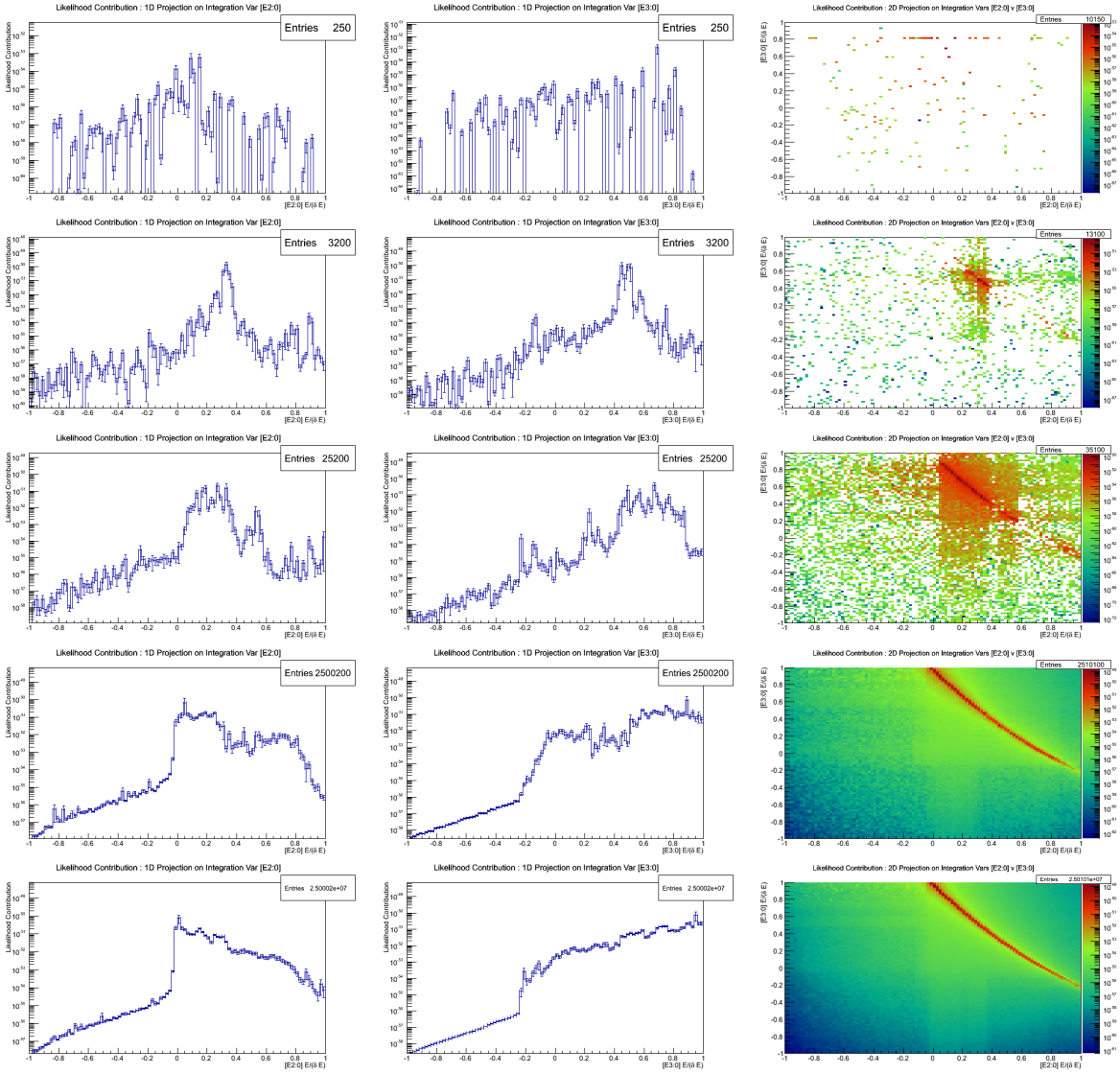


Figure 7.7: A example integration by the VEGAS algorithm using a pathological integrand. Here we see five sequential snapshots of the projected integrand, projected into the 2-D plane of Higgs-associated b-jet energies (right) and two 1-D projections into each of the two Higgs-associated b-jet energy dimensions (center and left) which represent orthogonal dimensions during this integration. This is a pathological case as the integrand is clearly inseparable along this pair of dimensions. The ridge in the upper-right quadrant of the 2-D integrand projection plots reveal the squared matrix element manifold consistent with a 125 GeV Higgs boson. Samples, which cost valuable CPU time, are, therefore, being wasted on non-contributing regions of phase space away from the Higgs ridge. One can also see slight artifacts from the VEGAS algorithm in the regions far from the peak (right column), though they are extremely small (note the logarithmic z-axis scale). The total sample distribution can be seen in Figure 7.8.

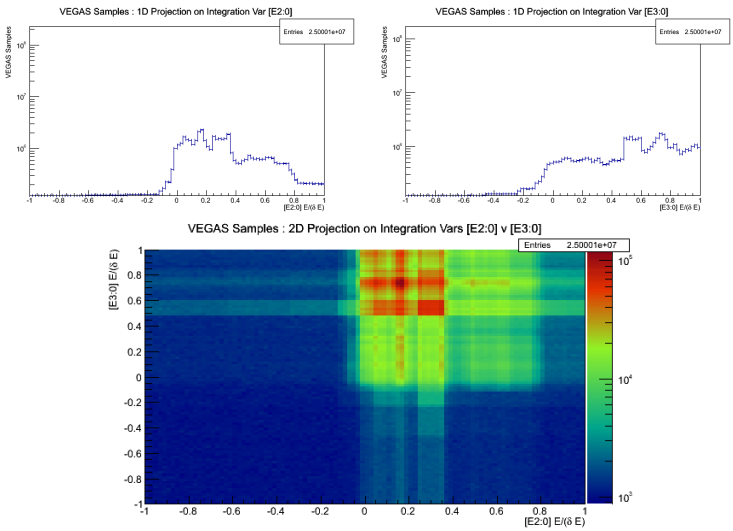


Figure 7.8: Total sample distribution from the integration in Figure 7.7. The location of each sample chosen by VEGAS is shown revealing the sub-optimal sampling regions far from the contributing ridge visible in Figure 7.7.

effectively aligning the mass ridge to the integration dimension. This approach is also taken for the two light-jet energies, $\mathbf{E_5}$ and $\mathbf{E_6}$, which are converted into integrations over $\mathbf{E_5}$ and $\mathbf{M}_W$. These transformations require a recalculation of the Jacobian determinant in addition to an additional scale factor for the leptonic W NWA. It is important to note that these transformations simply define the integral phase space from the perspective of VEGAS . In other words, for the calculation of the squared matrix element, the integration phase space point is transformed back into cartesian four-momenta and calculated with the full squared matrix elements.

7.3.3 Jet-Assignment Selection

In addition to decreasing the number of matrix element evaluations through more efficient numerical integration, we chose to reduce the number of needed integrations all together by limiting the number of jet-assignments for which we would calculate a probability. To perform the calculation, each measured final-state object must be assigned to one of the matrix element objects. This is non-trivial due to the fact that the measured event topology provides only generalized object descriptions. Thus, for an event containing 4 b-jets and 2 light-jets, the calculation must be performed on every unique jet-assignment consistent with the matrix element topology. In our case, our matrix element calculation leaves room for 12 unique assignments, given a 6 jet, 4 b-jet final-state. Here, we assume that assignments which only exchange the two b-jets assigned to the Higgs or the gluon, to be invariant. This picture changes, however, for events with greater than 4 b-jet candidates, for which the number of assignments increases factorially. To maintain a nominal 12 assignments for

Transformed Squared Matrix Element Phase Space

Description		**Parameters**			
Parent	**Object**	p_x	p_y	p_z	M GeV
p_1	i_1	$p_T^i = 0$	$p_T^i = 0$	$\sum E_i = \sum E_f$	0.0
p_2	i_2	$p_T^i = 0$	$p_T^i = 0$	$\sum p_z = 0$	0.0
—	—	E	η	ϕ	M GeV
t_{Had}	b-Jet$_1$	$\mathbf{d_1}$	δ	δ	4.7
t_{Lep}	b-Jet$_2$	$\mathbf{d_2}$	δ	δ	4.7
H, g	b-Jet$_3$	$\mathbf{d_3}$	δ	δ	4.7
H, g	b-Jet$_4$	$\delta\left(\mathbf{M_X^2} - (p_3^b + p_4^b)^2\right)$	δ	δ	4.7
$W_{q\bar{q}'}$	l-Jet$_1$	$\mathbf{d_5}$	δ	δ	0.0
$W_{q\bar{q}'}$	l-Jet$_2$	$\delta\left(\mathbf{M_W^2} - (p_1^l + p_2^l)^2\right)$	δ	δ	0.0
$W_{\ell\nu}$	ℓ	δ	δ	δ	0.0
—	—	p_x	p_y	p_z	M GeV
$W_{\ell\nu}$	ν	$\sum p_x = 0$	$\sum p_y = 0$	$\delta\left(M_W^2 - (p^\ell + p^\nu)^2\right)$	0.0

Table 7.3: The configuration of the integration phase space following the kinematic transformations. Here, the b-jet energies from their respective top quarks, $\mathbf{E_1}$ and $\mathbf{E_2}$, constitute the first two dimensions. The leading b-jet energy from the extra pair of b-jets, $\mathbf{E_3}$, along with the pair's invariant mass $\mathbf{M_X^2}$, where X is the Higgs boson, H, or the gluon, g, depending on the signal or background process respectively. The remaining two integration dimensions come from the leading light flavor jet, $\mathbf{E_5}$, and the hadronic W mass, $\mathbf{M_W^2}$.

every event, an algorithm is presented to veto inferior assignments prior to the matrix element calculation. In addition, this algorithm had to be sufficiently general to accommodate the $(\geq 6\text{j}, 3\text{b})$ region as well.

The algorithm starts by ranking the selected collection of b-jets by their MV1 *weight*. This, weight is defined by the MV1 b-jet tagging algorithm, a multivariate classification algorithm built to identify b-jets. The weight ranges from 0 to 1, with values near one signifying a more b-like jet. Following this sort, the top 4 most b-like jets are chosen and are subsequently permuted through the possible assignments to b-partons in the matrix element. If two or more b-jets are found to have the same MV1 weight, the jet with the largest p_T is preferred. The remaining b-tagged jets, if any, are then added to the set of non-tagged jets. Then, each pair of non-tagged jets is evaluated based on the pair's invariant mass in comparison to the W invariant mass. The pair with the mass most consistent with a W is chosen and subsequently assigned to the hadronically decaying W's light flavor partons.

For events with 4 or more b-jets and 2 or more light jets, only 12 jet-assignments are possible, as the above algorithms will always output precisely 4 b-candidates and 2 light-candidates. However, for events with less than 4 b-jets in the final-state, such as is the case for the $(\geq 6\text{j}, 3\text{b})$ region, we must add a caveat. In this case the procedure is followed resulting in again, 4 b-candidates and 2 light candidates. However, this time, only 3 of the b-candidates are actually b-tagged by the MV1 tagging algorithm. Thus, we allow this untagged jet to be permuted with the 2 light jets, and *vice versa*, resulting in 36 unique assignments. In order to prune these down to the nominal 12, we employ a simple 400 sample integration using the GSL 1.6 [85] implementation of the MISER [92] MC algorithm on each jet-assignment. The assignments are then ranked by the result of this integration and the 12 assignments with the largest integrals are retained. This pruning step occurs in all circumstances where the number of remaining unique jet-assignments is greater than the desired number, which we designate as 12 for the entire analysis.

This jet-assignment selection is extensible to events with large jet multiplicities and has been shown to rarely misidentify the true jet-assignment while significantly reducing the total number of assignments required to process. As will be seen in Section 8.4 and in Section 8.3, this quite general selection algorithm is very close to optimal given the available computing resources.

7.3.4 *In Situ* Jet-Assignment Optimization

As the VEGAS MC technique is principally based on importance sampling, it proceeds through an iterative process of alternating integration and adaptation phases. It will integrate the function, and store intermediate results in the histograms described in Section 7.3.2. Then it will use these histograms to approximate $p(\mathbf{x}_i)$, defined before by Eq. 7.27, and adapt the phase-space seg-

mentation accordingly to concentrate the most samples in the most important regions of the phase space. This can be seen at different sequential iterations, from top to bottom, in Figure 7.7.

In practice, it does this with, in our case, three independent sets of samples therefore running three simultaneous integrations. Following the integration phase, VEGAS compares the variance of the independent integrations and provides a weighted estimate of the definite integral, its estimated uncertainty, and the χ^2/d.o.f. of that estimate, based on the three simultaneous runs. Common practice is to continue iterating the integration and adaptation phases until the desired relative uncertainty on the integral estimate is reached. In our case, a more robust set of termination requirements was required.

As was mentioned in Section 7.2.4, the event probability discriminants to be used are defined by the maximum or arithmetic mean of the jet-assignment probabilities. Thus, assignments which asymptotically approach a value far smaller than the current sum of assignment probabilities contribute comparatively little information to the event discriminant while still utilizing substantial computational resources. Thus, instead of simply terminating an integration based on an assignment's uncertainty, we also elect to terminate if and when an assignment's contribution to the total event probability is below a defined threshold.

Finally, we must identify a reasonable sampling density and a way to modify the number of samples per iteration to optimize the convergence rate. Our empirical studies determined that a good initial 'warm up' of 2048 samples is sufficient, along with further 2048 sample iterations as needed. Further, taking into account that the convergence rate for MC integrations typically ranges from the naïve MC rate of $\sqrt{n}^{-1}$ to a best case scenario of n^{-1}, we can conservatively estimate that increasing the samples by a factor of four for an iteration should halve the uncertainty.

Putting the above together, the following sample elevation and termination algorithm was developed. First, each jet-assignment is required to undergo a 2048 sample 'warm up' integration and at least one 2048 sample iteration. Following these first two integration passes, the integral uncertainty and χ^2/d.o.f. is polled to determine if it has reached an acceptable level. These requirements become less stringent the more iterations, and thus more samples, are performed. This corresponds to Tests 1-3 in Table 7.4. If the integral estimate is ever less than 10^{-200} or if the estimate is less than 0.1% of the cumulative probability, the calculation is terminated for the given jet-assignment. This corresponds to Tests 4 and 5. Finally, Test's 7 and 8 work to prevent a calculation from proceeding *ad infinitum*.

Finally, in the event that the calculation is not terminated but, the uncertainty is less than 10%, the number of samples for the next iteration is doubled. In the event that the uncertainty is acceptable, but the χ^2/d.o.f. is out of bounds, the number of samples for the next iteration is increased by 10%. These requirements further encourage convergence at a rate faster than otherwise, based on

Termination Conditions

Test	**Iteration**	**Integral Estimate**	**Rel. Uncert. %**	$\lvert\chi^2/\text{d.o.f.} - 1\rvert$	**Contribution Fraction %**
1	1	-	< 2.5	< 0.5	-
2	2	-	< 5.0	< 0.5	-
3	≥ 3	-	< 10.0	< 0.5	-
4	-	$< 1e-200$	-	-	-
5	-	-	-	-	< 0.1
6	≥ 2	-	-	-	< 1
7	≥ 6	-	< 15.0	< 0.5	< 5
8	18	-	-	-	-

Table 7.4: Sample elevation and termination conditions for a given jet-assignment integration. The test column identifies each condition and the order in which they are applied. The remaining columns display the different parameters defined during a given integration and the criteria that must be satisfied to terminate the calculation. As an example, for Test 7, if there have been at least 6 iterations, and the relative uncertainty is below 15%, the χ^2 is reasonable, and this jet-assignment contributes less than 5% of the cumulative event's probability sum, then it will be terminated.

empirical studies.

7.3.5 Squared Matrix Element Optimization

The final optimization procedure targets the squared matrix element itself by slightly modifying the functions generated by MadGraph 5. As is the case in the MadGraph 5 implementation, the squared matrix element calculation is decomposed into a sum of terms orthogonal in the helicity basis to then be summed and squared. Each term represents a unique helicity state in a set that contains 2^p configurations, where p is the number of participants or the number of external lines in diagrams to be calculated. In our case there are 2^{10} or 1024 total helicity states. Of these, the vast majority are non-physical returning zero when evaluated. Of the remainder, several will be non-unique due to symmetries present in the diagrams. Thus, of the total 1024 terms, very few are physical and unique. Finally, as these terms are all positive definite and thus do not interfere with one another during the summing and squaring procedure, it is possible to identify and remove states which do not contribute appreciably to the ultimate squared matrix element without significantly affecting its information content.

To take advantage of this understanding of the information content of the various helicity terms in the calculation, the following algorithm was developed. First, each analytically possible helicity state is stored in a list. Each helicity state structure stores its relative contribution and an integer value designating its degeneracy. After the first two calls to the calculation method, helicity states with zero relative contribution are pruned from the list. Then, states found to

be invariant are combined, one state is pruned from the list, while the other increments its degeneracy weight. For all subsequent calls, calculations are only performed on helicity states remaining in the list. The individual contributions are weighted by their respective degeneracies, summed, then squared.

Finally, a separate pruning method was developed which would remove states with sub-leading relative contributions to the final result. This method is called after first 100 calls to the matrix element calculation method, using those calls to generate a cumulative picture of each term's contribution. Once called, it proceeds to prune away all helicity states contributing less than 10% of the maximally contributing state. These terms are permanently removed from the calculation. From then on, every 2000 calls the method is called to further prune states which contribute less than 1% of the maximally contributing state. Our empirical studies have shown that this approximation procedure does not perceptibly change the discriminating power of the final discriminants. Yet, these approximations, taken together, reduce the per-sample evaluation time by nearly two orders of magnitude. It is important to note that these approximations are reinitialized with every integration. In other words, these approximations are unique and adaptive to every jet-assignment probability calculation.

Chapter 8

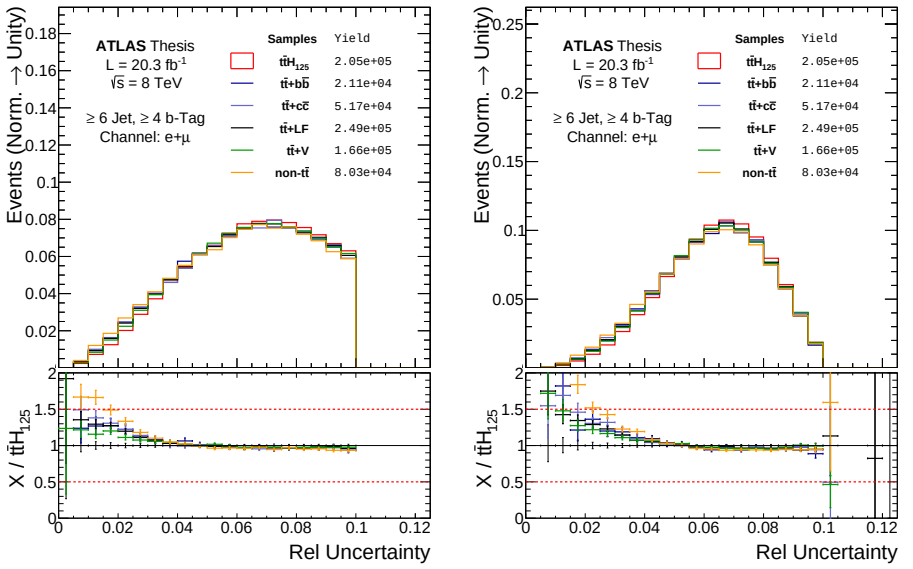
Matrix Element Method Development

This chapter presents the prominent validation and optimization studies performed throughout the development of this MEM implementation. Particular focus will be placed on the implementation which was used to obtain the final result, which includes implementing $t\bar{t}H$ matrix elements with a 125 GeV Higgs boson, $t\bar{t}H_{125}$. These studies will be grouped by their primary focus into the following sections.

8.1 Integration and Integrand Studies

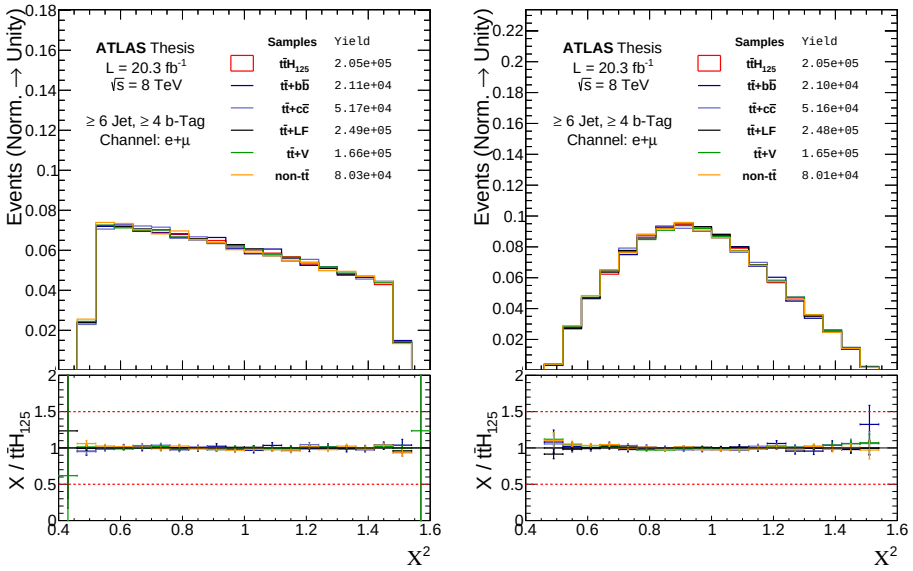
Initial testing revolved around characterizing the integrand and identifying methods to improve convergence. One such method was the application of the integration termination algorithm described in Section 7.3.4. To validate this algorithm, the relative uncertainty and VEGAS's χ^2 value are stored for each individual integration. In addition, the number of times the integrand is sampled for each jet-assignment is stored. The plots in Figures 8.1-8.5 show the results of these studies which portray the stable operation of the matrix element integration framework.

In addition to identifying the correct termination conditions, it is crucial to identify kinematic variables that significantly influence the value of the jet-assignment probability. As the goal is to reduce run times by reducing the dimensionality of the integrand, it is important to identify dimensions that can be removed without significantly affecting the final result of the integration. To accomplish this, the probability result from a given jet-assignment is plotted as a function of a kinematic variable derived from that jet-assignment. This is then repeated for all the jet-assignments for all events in the various samples. The result is a projection of the probability onto a kinematic parameter space defined by the integrated jet-assignments. Large variance in the probability as a function of the kinematic parameter denotes a strong dependence between the two. For instance, these studies motivate the decision to integrate over the Higgs boson mass, as opposed to using a Narrow Width Approximation (NWA) as the NWA neglects valuable information near the Higgs resonance peak. Selected kinematic variables from these studies are shown in Figures 8.6-8.9.



(a) Max L($t\bar{t}H$) Assignment

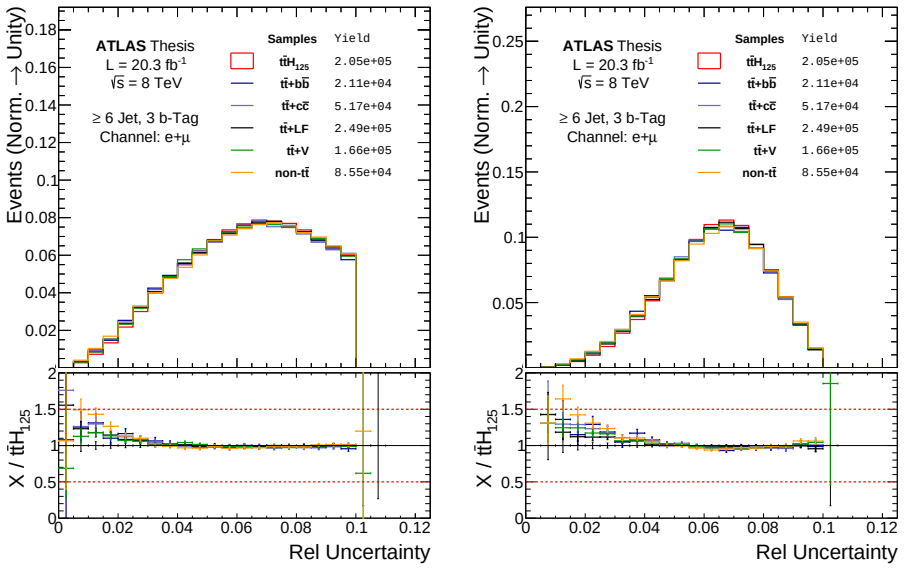
(b) Mean L($t\bar{t}H$)



(c) Max L($t\bar{t}H_{125}$) Assignment

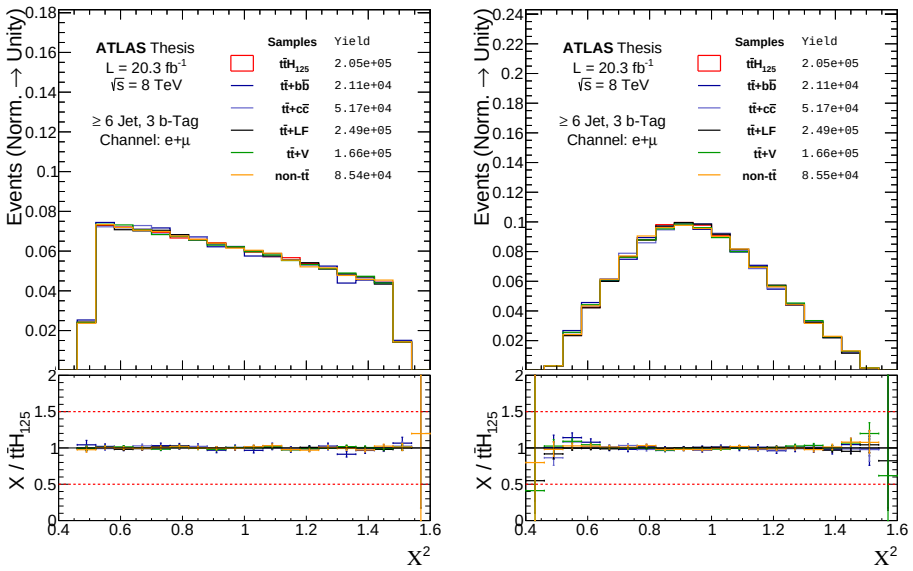
(d) Mean L($t\bar{t}H_{125}$)

Figure 8.1: The $t\bar{t}H\left(H \to b\bar{b}\right)$ matrix element integration quality parameterized by the GSL 1.6 implementation of the VEGAS algorithm, in the $(\geq 6\text{j}, \geq 4\text{b})$ region. Figures (a) and (b) show the relative error of the maximum-probability jet-assignment, (a), and the weighted relative error on the arithmetic mean of the jet-assignment probabilities, (b). Figures (c) and (d) show the the χ^2 for the maximum-probability jet-assignment, (c), and the weighted χ^2 for the arithmetic mean of the jet-assignment probabilities, (d). In all these plots, the relative shapes of all the different background samples are compared by normalizing the area of all the samples to unity. The ratio plots below compare the background samples to the $t\bar{t}H_{125}$ sample. The quoted yields represent the total number of unweighted events which were integrated.



(a) Max L($t\bar{t}H$) Assignment

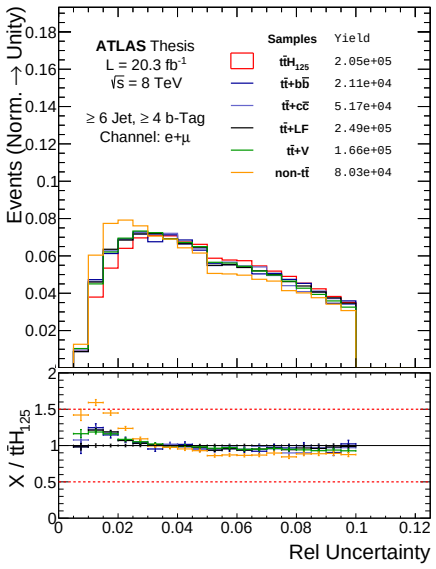
(b) Mean L($t\bar{t}H$)



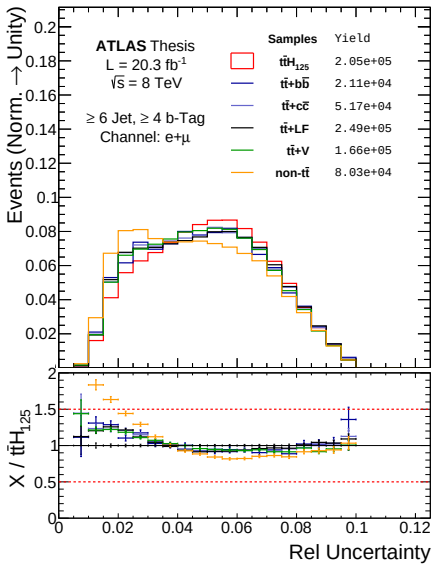
(c) Max L($t\bar{t}H_{125}$) Assignment

(d) Mean L($t\bar{t}H_{125}$)

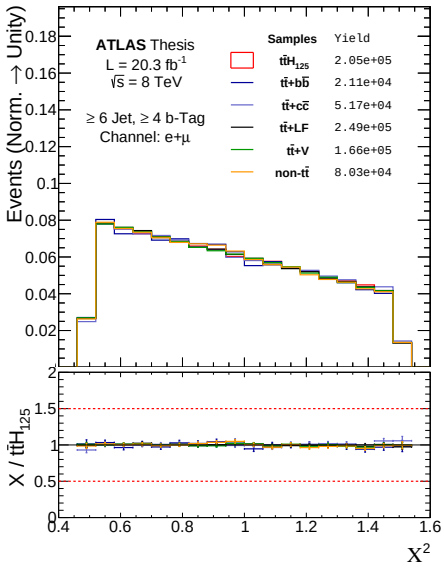
Figure 8.2: The $t\bar{t}H\left(H \to b\bar{b}\right)$ matrix element integration quality parameterized by the GSL 1.6 implementation of the VEGAS algorithm, in the $(\geq 6\text{j}, 3\text{b})$ region. Figures (a) and (b) show the relative error of the maximum-probability jet-assignment, (a), and the weighted relative error on the arithmetic mean of the jet-assignment probabilities, (b). Figures (c) and (d) show the the χ^2 for the maximum-probability jet-assignment, (c), and the weighted χ^2 for the arithmetic mean of the jet-assignment probabilities, (d). In all these plots, the relative shapes of all the different background samples are compared by normalizing the area of all the samples to unity. The ratio plots below compare the background samples to the $t\bar{t}H_{125}$ sample. The quoted yields represent the total number of unweighted events which were integrated.



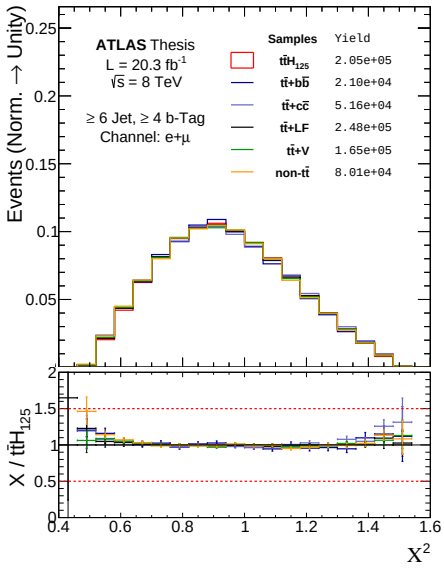
(a) Max L($t\bar{t} + b\bar{b}$) Assignment



(b) Mean L($t\bar{t} + b\bar{b}$)

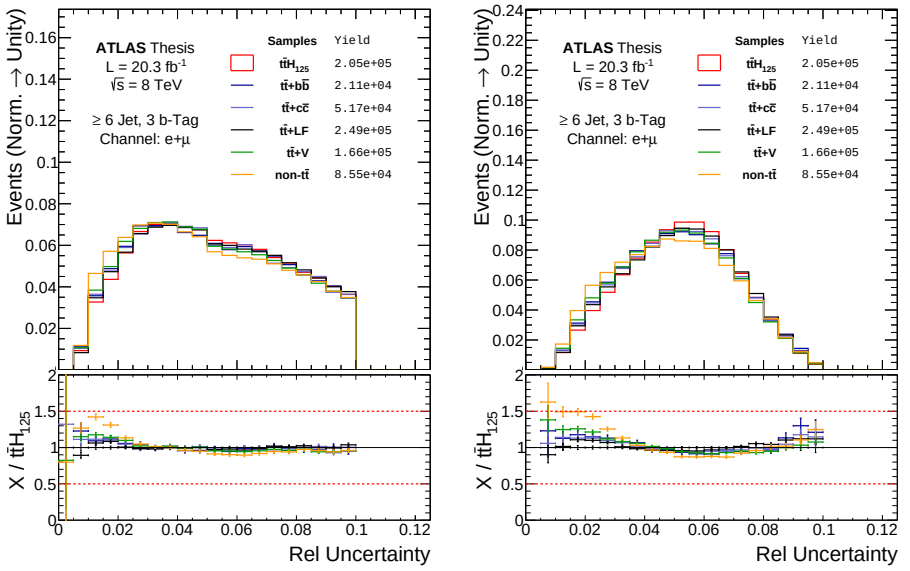


(c) Max L($t\bar{t} + b\bar{b}$) Assignment



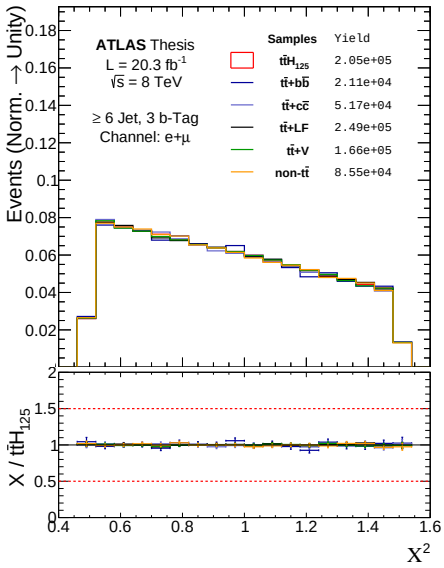
(d) Mean L($t\bar{t} + b\bar{b}$)

Figure 8.3: The $t\bar{t} + b\bar{b}$ matrix element integration quality parameterized by the GSL 1.6 implementation of the VEGAS algorithm, in the ($\geq$ 6j, $\geq$ 4b) region. Figures (a) and (b) show the relative error of the maximum-probability jet-assignment, (a), and the weighted relative error on the arithmetic mean of the jet-assignment probabilities, (b). Figures (c) and (d) show the the χ^2 for the maximum-probability jet-assignment, (c), and the weighted χ^2 for the arithmetic mean of the jet-assignment probabilities, (d). In all these plots, the relative shapes of all the different background samples are compared by normalizing the area of all the samples to unity. The ratio plots below compare the background samples to the $t\bar{t}H_{125}$ sample. The quoted yields represent the total number of unweighted events which were integrated.

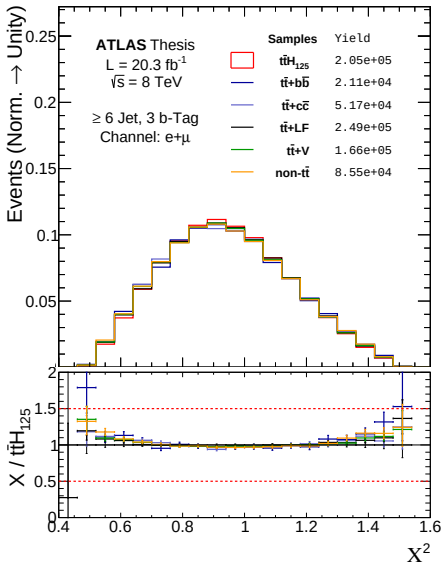


(a) Max L($t\bar{t} + b\bar{b}$) Assignment

(b) Mean L($t\bar{t} + b\bar{b}$)

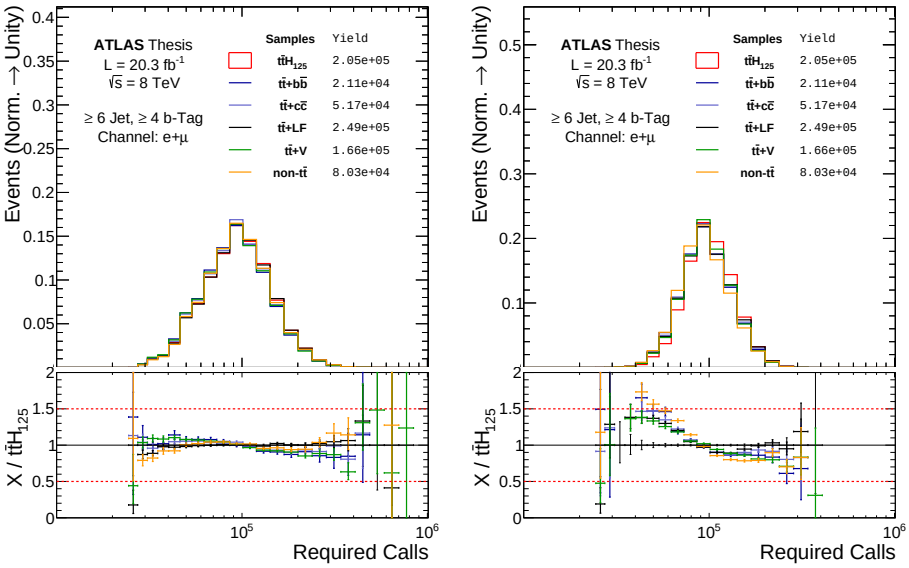


(c) Max L($t\bar{t} + b\bar{b}$) Assignment



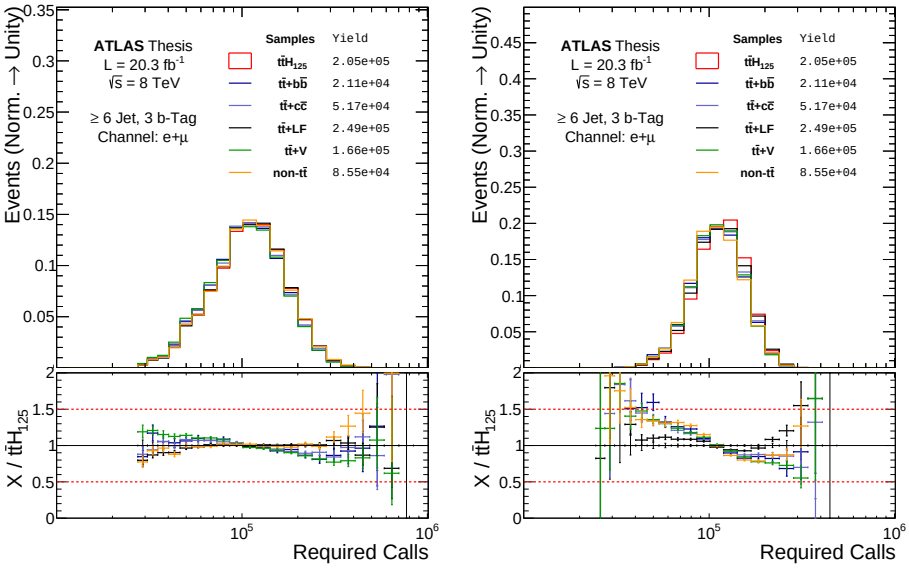
(d) Mean L($t\bar{t} + b\bar{b}$)

Figure 8.4: The $t\bar{t} + b\bar{b}$ matrix element integration quality parameterized by the GSL 1.6 implementation of the VEGAS algorithm, in the ($\geq$ 6j, 3b) region. Figures (a) and (b) show the relative error of the maximum-probability jet-assignment, (a), and the weighted relative error on the arithmetic mean of the jet-assignment probabilities, (b). Figures (c) and (d) show the the χ^2 for the maximum-probability jet-assignment, (c), and the weighted χ^2 for the arithmetic mean of the jet-assignment probabilities, (d). In all these plots, the relative shapes of all the different background samples are compared by normalizing the area of all the samples to unity. The ratio plots below compare the background samples to the $t\bar{t}H_{125}$ sample. The quoted yields represent the total number of unweighted events which were integrated.



(a) L($t\bar{t}H_{125}$) Integration

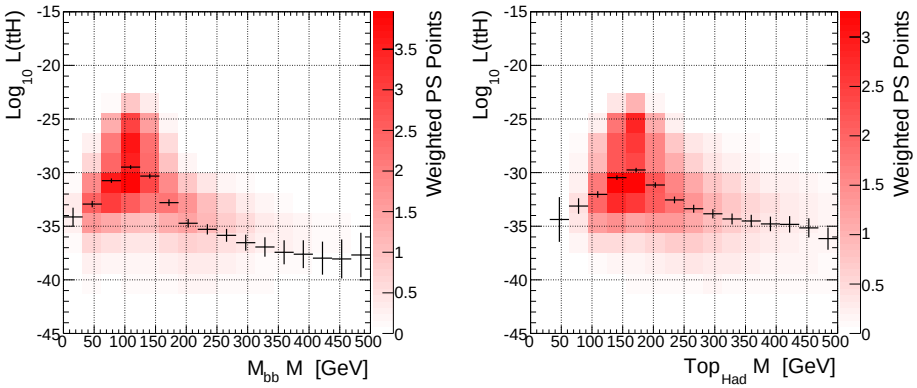
(b) L($t\bar{t} + b\bar{b}$) Integration



(c) L($t\bar{t}H_{125}$) Integration

(d) L($t\bar{t} + b\bar{b}$) Integration

Figure 8.5: Total integration samples per event for the various samples used in this analysis. The required calls per event are shown for the $t\bar{t}H\ (H \to b\bar{b})$ matrix elements for the $(\geq 6\text{j}, \geq 4\text{b})$ region, (a), and the $(\geq 6\text{j}, 3\text{b})$ region, (c). In addition, the required calls per event are shown for the $t\bar{t} + b\bar{b}$ matrix elements for the $(\geq 6\text{j}, \geq 4\text{b})$ region, (b), and the $(\geq 6\text{j}, 3\text{b})$ region, (d). In all of these plots, the relative shapes of all the different background samples are compared by normalizing the area of all the samples to unity. The ratio plots below compare the background samples to the $t\bar{t}H_{125}$ sample. The quoted yields represent the total number of unweighted events which were integrated.



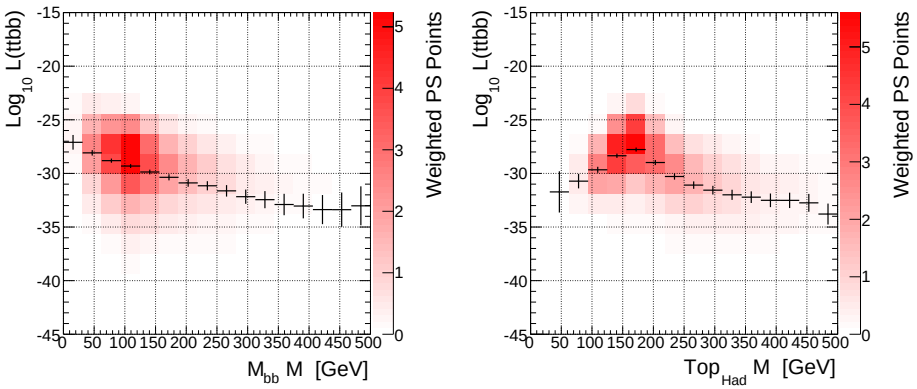
(a) $m_H \mapsto L(t\bar{t}H_{125})$

(b) $m_T(\text{Had}) \mapsto L(t\bar{t}H_{125})$

(c) $m_W(\text{Had}) \mapsto L(t\bar{t}H_{125})$

(d) $m_T(\text{Lep}) \mapsto L(t\bar{t}H_{125})$

Figure 8.6: Jet-assignment probabilities for the $t\bar{t}H\left(H \to b\bar{b}\right)$ *process*, projected onto kinematic variables for the $t\bar{t}H_{125}$ *sample* in the $(\geq 6\text{j}, \geq 4\text{b})$ region. The abscissa axis denotes the range of a given kinematic variable determined by a given jet-assignment. The ordinate axis shows the range of probabilities calculated for the given jet-assignment. The color axis shows the total number of jet-assignments. The black curve represents the mean probability values for each abscissa axis bin. Figure (a) is of particular importance in differentiating $t\bar{t}H\left(H \to b\bar{b}\right)$ events from $t\bar{t} + b\bar{b}$ events due to the expected Higgs mass resonance in the former. This is in contrast to (a), which utilizes the $t\bar{t} + b\bar{b}$ matrix elements with no Higgs resonance.



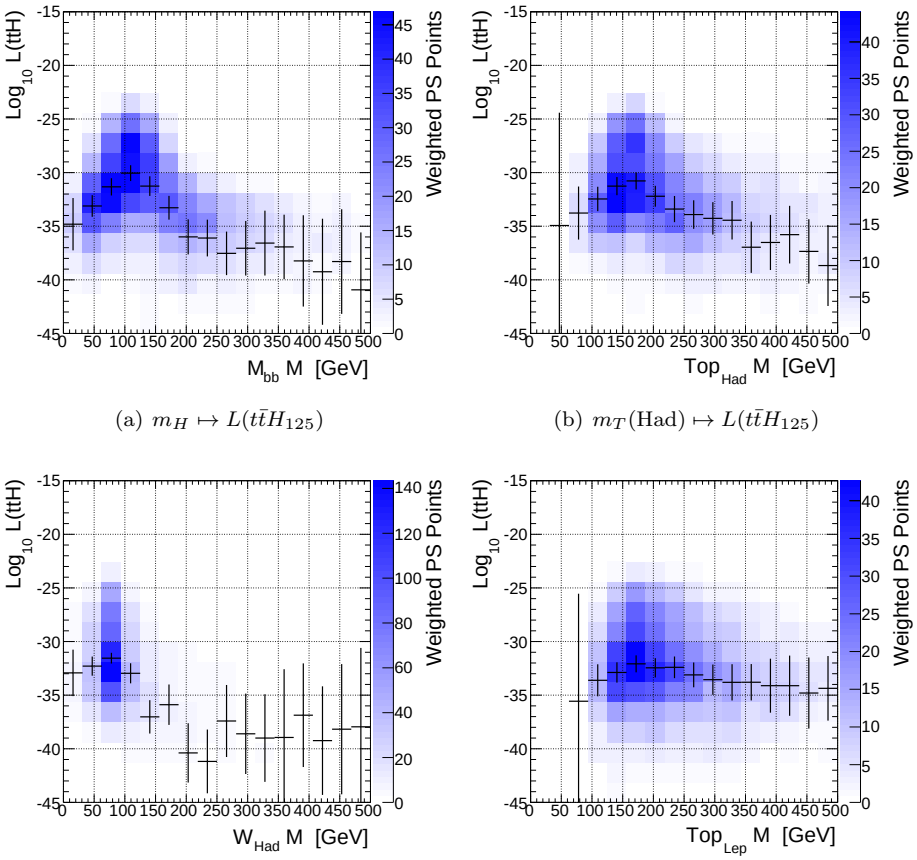
(a) $m_H \mapsto L(t\bar{t} + b\bar{b})$

(b) $m_T(\text{Had}) \mapsto L(t\bar{t} + b\bar{b})$

(c) $m_W(\text{Had}) \mapsto L(t\bar{t} + b\bar{b})$

(d) $m_T(\text{Lep}) \mapsto L(t\bar{t} + b\bar{b})$

Figure 8.7: Jet-assignment probabilities for the $t\bar{t} + b\bar{b}$ *process*, projected onto kinematic variables for the $t\bar{t}H_{125}$ *sample* in the $(\geq 6\text{j}, \geq 4\text{b})$ region. The abscissa axis denotes the range of a given kinematic variable determined by a given jet-assignment. The ordinate axis shows the range of probabilities calculated for the given jet-assignment. The color axis shows the total number of jet-assignments. The black curve represents the mean probability values for each abscissa axis bin. Figure (a) is of particular importance in differentiating $t\bar{t}H\left(H \to b\bar{b}\right)$ events from $t\bar{t} + b\bar{b}$ events due to the expected Higgs mass resonance in the former. This is in contrast to (a), which utilizes the $t\bar{t}H\left(H \to b\bar{b}\right)$ matrix elements which include the Higgs resonance.



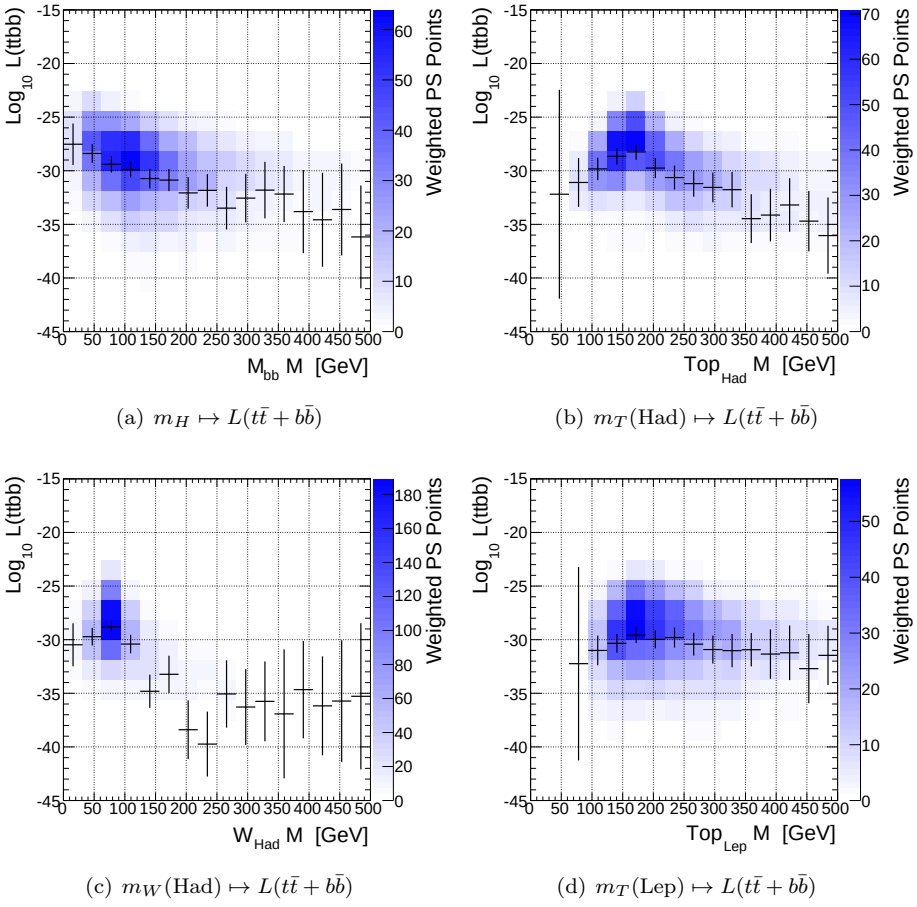
(a) $m_H \mapsto L(t\bar{t}H_{125})$

(b) $m_T(\text{Had}) \mapsto L(t\bar{t}H_{125})$

(c) $m_W(\text{Had}) \mapsto L(t\bar{t}H_{125})$

(d) $m_T(\text{Lep}) \mapsto L(t\bar{t}H_{125})$

Figure 8.8: Jet-assignment probabilities for the $t\bar{t}H\left(H \to b\bar{b}\right)$ *process*, projected onto kinematic variables for the $t\bar{t} + b\bar{b}$ *sample* in the $(\geq 6\text{j}, \geq 4\text{b})$ region. The abscissa axis denotes the range of a given kinematic variable determined by a given jet-assignment. The ordinate axis shows the range of probabilities calculated for the given jet-assignment. The color axis shows the total number of jet-assignments. The black curve represents the mean probability values for each abscissa axis bin. Figure (a) is of particular importance in differentiating $t\bar{t}H\left(H \to b\bar{b}\right)$ events from $t\bar{t} + b\bar{b}$ events due to the expected Higgs mass resonance in the former. This is in contrast to (a), which utilizes the $t\bar{t} + b\bar{b}$ matrix elements with no Higgs resonance.



(a) $m_H \mapsto L(t\bar{t}+b\bar{b})$

(b) $m_T(\text{Had}) \mapsto L(t\bar{t}+b\bar{b})$

(c) $m_W(\text{Had}) \mapsto L(t\bar{t}+b\bar{b})$

(d) $m_T(\text{Lep}) \mapsto L(t\bar{t}+b\bar{b})$

Figure 8.9: Jet-assignment probabilities for the $t\bar{t}+b\bar{b}$ *process*, projected onto kinematic variables for the $t\bar{t}+b\bar{b}$ *sample* in the ($\geq$ 6j, $\geq$ 4b) region. The abscissa axis denotes the range of a given kinematic variable determined by a given jet-assignment. The ordinate axis shows the range of probabilities calculated for the given jet-assignment. The color axis shows the total number of jet-assignments. The black curve represents the mean probability values for each abscissa axis bin. Figure (a) is of particular importance in differentiating $t\bar{t}H\left(H \to b\bar{b}\right)$ events from $t\bar{t}+b\bar{b}$ events due to the expected Higgs mass resonance in the former. This is in contrast to (a), which utilizes the $t\bar{t}H\left(H \to b\bar{b}\right)$ matrix elements which include the Higgs resonance.

8.2 Discrimination Studies

Once an intuition was developed regarding the number of needed integration samples and the general form of the integrand in the jet four-momentum phase space, the task turned to optimizing the separation power of the matrix element variables. The primary discriminants, defined in Section 7.1.2, had to then be developed and assessed. These studies primarily focus on identifying the optimal use of the resulting jet-assignment probabilities in defining a total event probability discriminant.

The first study seeks to characterize the relative effects of the normalization parameters which define the respective $t\bar{t}H\left(H \to b\bar{b}\right)$ and $t\bar{t} + b\bar{b}$ probabilities. In Eq. 7.5 we see that the probability, $\mathcal{P}(\mathbf{x}_f|\alpha)$, is properly normalized by, $(\sigma_\alpha \epsilon A)^{-1}$, where σ_α is the total cross section for the process α, ϵ, the efficiency associated with measuring events from the α process; and, A, the geometrical acceptance of the detector. Therefore, we can redefine the the jet-assignment probability to be the product of the normalization factor, N_α, and the un-normalized integral, $L(\mathbf{x}_f|\alpha)$, which we refer to as the jet-assignment's *likelihood*. As the symbol suggests, N_α is dependent on the process α, but it is the same normalization for all jet configurations, $\mathbf{x}_f$. Taking this one step further, because our discriminants are defined utilizing only two probabilities associated with the $t\bar{t}H\left(H \to b\bar{b}\right)$ and $t\bar{t}+b\bar{b}$ processes, we can factor $N_{t\bar{t}H}$ and $N_{t\bar{t}+b\bar{b}}$ into a combined normalization factor which we named $\boldsymbol{\lambda}$. This process is defined in Eq. 8.2 and Eq. 8.3:

$$\mathcal{P}(\mathbf{x}_f|\alpha) = \frac{1}{\sigma_\alpha \epsilon A} \int d\mathbf{y}_f \; \boldsymbol{d}\sigma_\alpha \left(pp \to \boldsymbol{x}_f\right) W\left(\mathbf{y}_f|\mathbf{x}_f\right) \; d\boldsymbol{\Phi}_{dlips} \tag{8.1}$$

$$\mathcal{P}(\mathbf{x}_f|\alpha) = N_\alpha L(\mathbf{x}_f|\alpha) \tag{8.2}$$

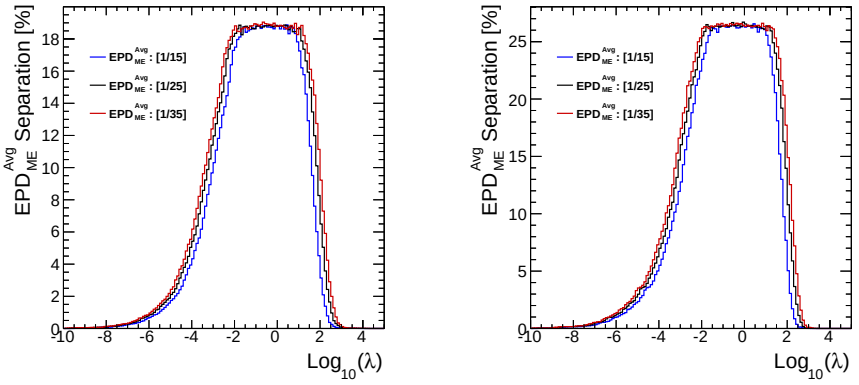
$$\boldsymbol{\lambda} = \frac{N_{t\bar{t}H}}{N_{t\bar{t}+b\bar{b}}}. \tag{8.3}$$

Thus, our primary discriminating variables for this analysis can be re-defined in terms of the un-normalized *likelihood* and $\boldsymbol{\lambda}$. Eq. 8.4 shows this variable transformation applied to the Neyman-Pearson ratio, R_{ME}, and Eq. 8.5 shows this transformation applied to the event probability discriminant, EPD_{ME}. In both cases, the ratio of normalization parameters is replaced by the single parameter $\boldsymbol{\lambda}$:

$$\mathrm{R_{ME}}\left(\mathbf{x}_f \mid t\bar{t}H, t\bar{t}+b\bar{b}\right) = \boldsymbol{\lambda}\frac{L(\mathbf{x}_f|t\bar{t}H)}{L(\mathbf{x}_f|t\bar{t}+b\bar{b})} \tag{8.4}$$

$$\mathrm{EPD_{ME}}\left(\mathbf{x}_f \mid t\bar{t}H, t\bar{t}+b\bar{b}\right) = \frac{L(\mathbf{x}_f|t\bar{t}H)}{L(\mathbf{x}_f|t\bar{t}H) + \boldsymbol{\lambda}L(\mathbf{x}_f|t\bar{t}+b\bar{b})}. \tag{8.5}$$

To evaluate the effect of $\boldsymbol{\lambda}$ on the final discriminants, $\boldsymbol{\lambda}$ is scaled by 15



(a) $(\geq 6\text{j}, 3\text{b})$ (b) $(\geq 6\text{j}, \geq 4\text{b})$

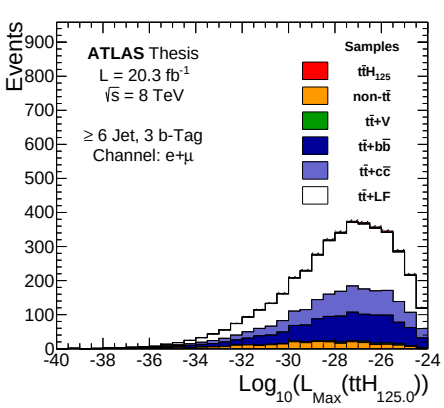
Figure 8.10: Separation dependence on $\boldsymbol{\lambda}$ in $(\geq 6\text{j}, 3\text{b})$ and $(\geq 6\text{j}, \geq 4\text{b})$ regions. Here, the EPD_{ME}^{Avg} distribution is defined using the $\boldsymbol{\lambda}$-parametrization shown in Eq. 8.5. The separation is then calculated between the $t\bar{t}H_{125}$ and $t\bar{t} + b\bar{b}$ samples using this discriminant. That separation is then measured for multiple values of $\boldsymbol{\lambda}$. As the plots show, $\boldsymbol{\lambda}$ does not effect the separation over the fiducial region, roughly $-2 < \text{Log}_{10}(\boldsymbol{\lambda}) < 2$. This holds for histograms with, 15 (*blue*), 25 (*black*), and 35 (*red*) bins.

orders of magnitude. In principal, as $\boldsymbol{\lambda}$ represents a scaling of the ratio, we do not expect any dependence on $\boldsymbol{\lambda}$, in terms of actual discrimination power. However, due to the finite number of bins in the distribution, extreme values of $\boldsymbol{\lambda}$ will degrade the apparent separation as more events are pushed into fewer total bins. The limiting case occurs when a value of $\boldsymbol{\lambda}$ results in all events appearing in the same bin, resulting in a separation of zero, where separation is defined by Eq. 7.10. This behavior is shown explicitly in Figure 8.10. The value for $\boldsymbol{\lambda}$ chosen for this analysis is $\boldsymbol{\lambda} = 0.23$, the central value for the fiducial region defined roughly by $-2 < \text{Log}_{10}(\boldsymbol{\lambda}) < 2$.

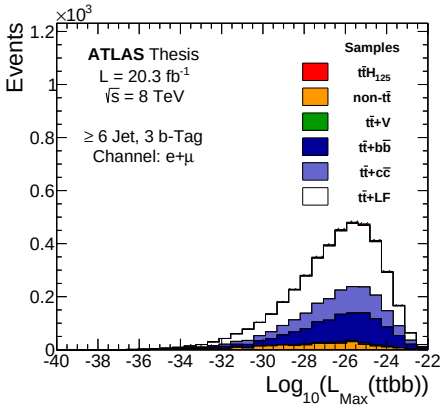
Figures 8.11 and 8.12 display the expected yields with respect to the unnormalized signal and background likelihoods defined previously. These likelihoods are shown both for the jet-assignment with the maximum likelihood, as well as the average likelihood for all considered jet-assignments. These likelihoods are then used as inputs to the final discriminating variables defined by Eqs. 8.4 and 8.5.

Figures 8.13 and 8.14 display the expected yields with respect to the final discriminating variables defined by Eqs. 8.4 and 8.5. These variables are shown both for the jet-assignment with the maximum likelihood, as well as the average likelihood for all considered jet-assignments.

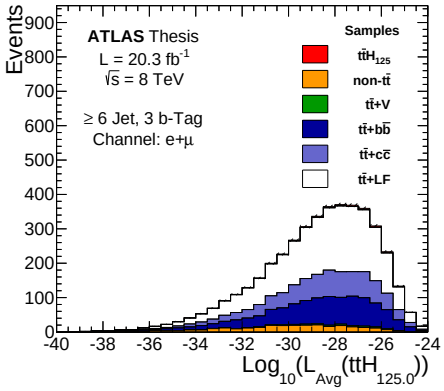
Figures 8.15 and 8.16 display the expected discrimination power with respect to the final discriminating variables. These variables are shown both for the jet-assignment with the maximum likelihood, as well as the average likelihood for all considered jet-assignments. For these plots, separation is defined by



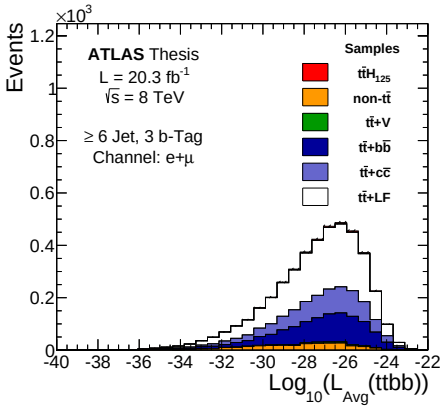
(a) Max L($t\bar{t}H_{125}$) Assignment



(b) Max L($t\bar{t} + b\bar{b}$) Assignment

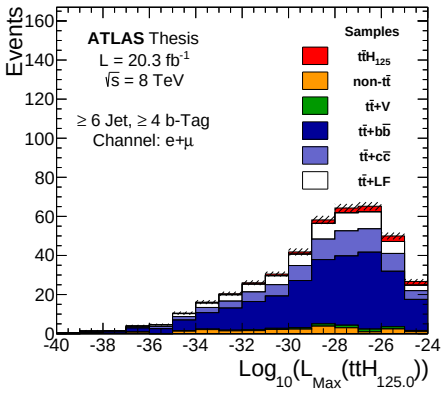


(c) Average L($t\bar{t}H_{125}$)

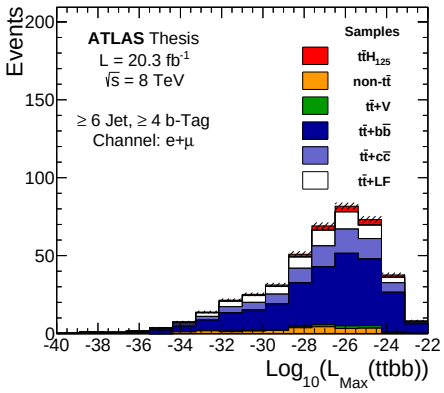


(d) Average L($t\bar{t} + b\bar{b}$)

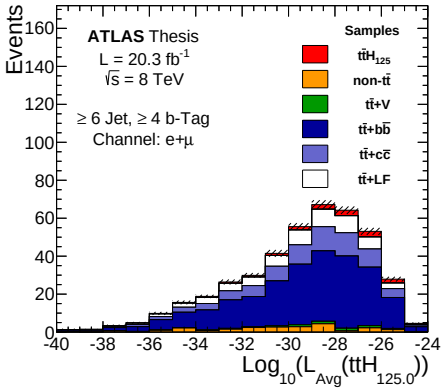
Figure 8.11: Expected un-normalized likelihood response using the "max likelihood jet-assignment" and "average jet-assignment likelihood" representations in the $(\geq 6\text{j}, 3\text{b})$ region. Here the $t\bar{t}H_{125}$ sample is normalized to the expected SM cross section, consistent with a SM signal strength, $\mu = 1$.



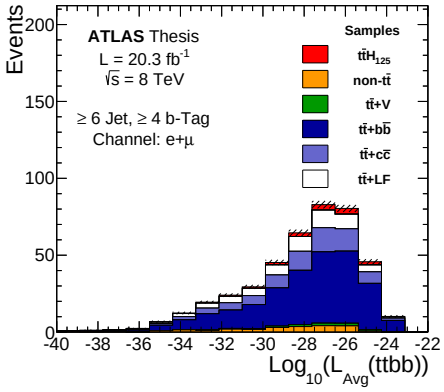
(a) Max L($t\bar{t}H_{125}$) Assignment



(b) Max L($t\bar{t} + b\bar{b}$) Assignment

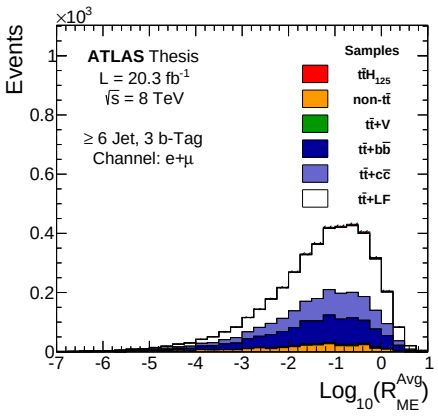


(c) Average L($t\bar{t}H_{125}$)

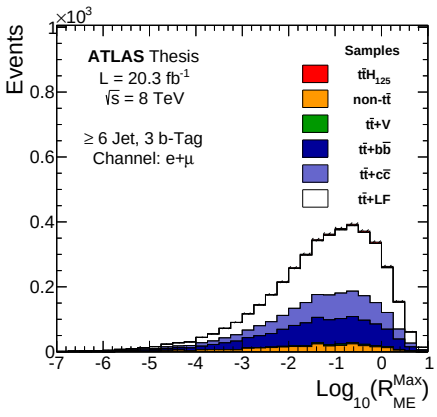


(d) Average L($t\bar{t} + b\bar{b}$)

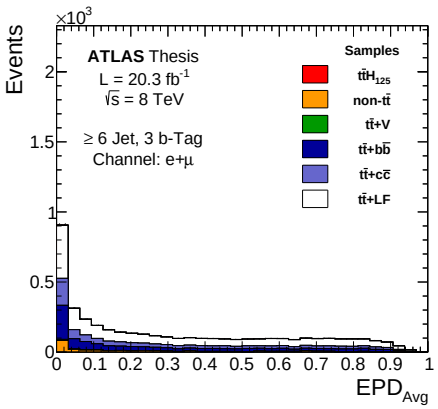
Figure 8.12: Expected un-normalized likelihood response using the "max likelihood jet-assignment" and "average jet-assignment likelihood" representations in the ($\geq$ 6j, $\geq$ 4b) region. Here the $t\bar{t}H_{125}$ sample is normalized to the expected SM cross section, consistent with a SM signal strength, $\mu = 1$.



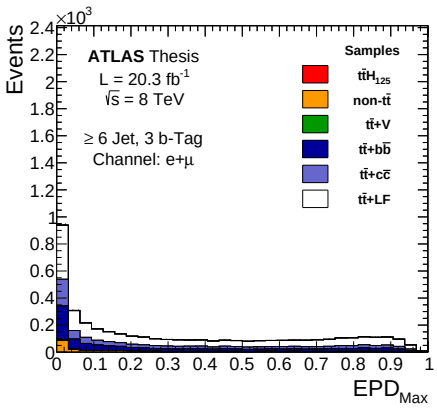
(a) Average L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$)



(b) Max L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$) Assignments

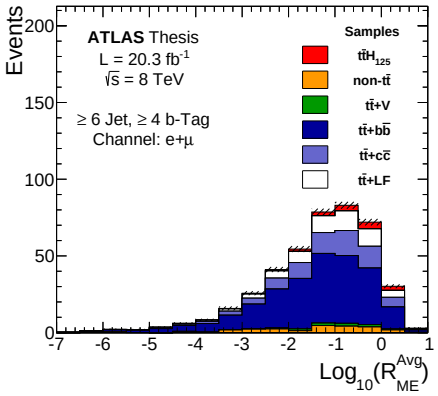


(c) Average L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$)

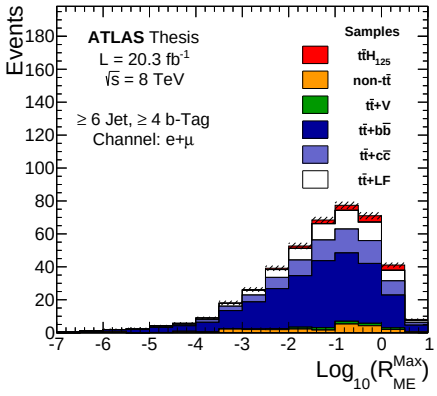


(d) Max L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$) Assignments

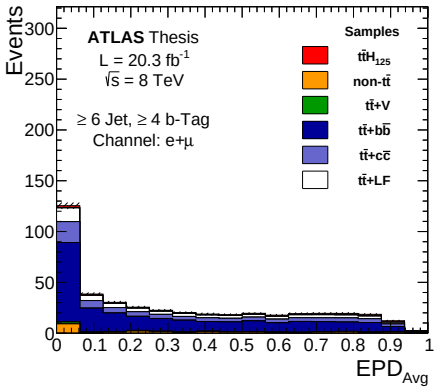
Figure 8.13: Expected discriminating variable response using the "max likelihood jet-assignment" and "average jet-assignment likelihood" representations in conjunction with the $\boldsymbol{\lambda}$ parametrization in the $(\geq 6\text{j}, 3\text{b})$ region. Here the $t\bar{t}H_{125}$ sample is normalized to the expected SM cross section, consistent with a SM signal strength, $\mu = 1$.



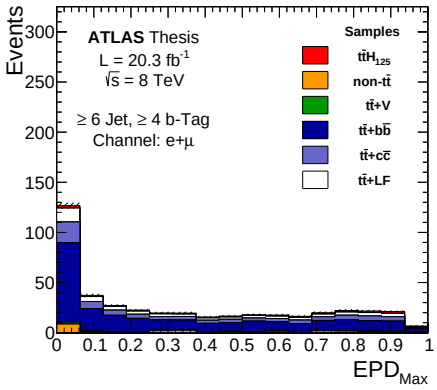
(a) Average L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$)



(b) Max L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$) Assignments



(c) Average L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$)



(d) Max L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$) Assignments

Figure 8.14: Expected discriminating variable response using the "max likelihood jet-assignment" and "average jet-assignment likelihood" representations in conjunction with the $\boldsymbol{\lambda}$ parametrization in the $(\geq 6\text{j}, \geq 4\text{b})$ region. Here the $t\bar{t}H_{125}$ sample is normalized to the expected SM cross section, consistent with a SM signal strength, $\mu = 1$.

Eq. 7.10, and is quoted below the yields in the legends. We see here that the "average jet-assignment likelihood" construction consistently out-performs the "max likelihood jet-assignment" construction.

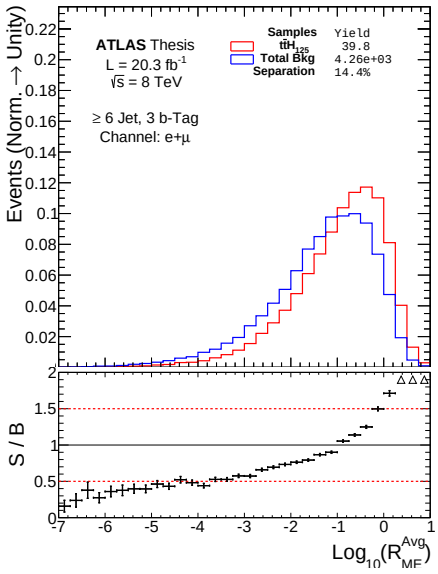
Figure 8.17 and Figure 8.18 display the expected shape response with respect to the final discriminating variables. These variables are shown both for the jet-assignment with the maximum likelihood, as well as the average likelihood for all considered jet-assignments. For these plots, all samples are normalized to unity to show the varying shapes of the different samples. In the lower ratio plots, we see the bin-by-bin 'X/$t\bar{t}H_{125}$' ratio, where X denotes the present background contributions corresponding to the included legends. Here we see that the $t\bar{t}+V$ and $t\bar{t}$ + Light-Flavor backgrounds are more similar to the $t\bar{t}H_{125}$ signal than the $t\bar{t}+c\bar{c}$ and $t\bar{t}+b\bar{b}$ backgrounds. This is helpful as the $t\bar{t}$ + Light-Flavor background in particular is comparatively well estimated from other, non-signal regions, while $t\bar{t}+c\bar{c}$ and $t\bar{t}+b\bar{b}$ backgrounds must be estimated in regions where signal events are more concentrated.

8.3 Truth Studies

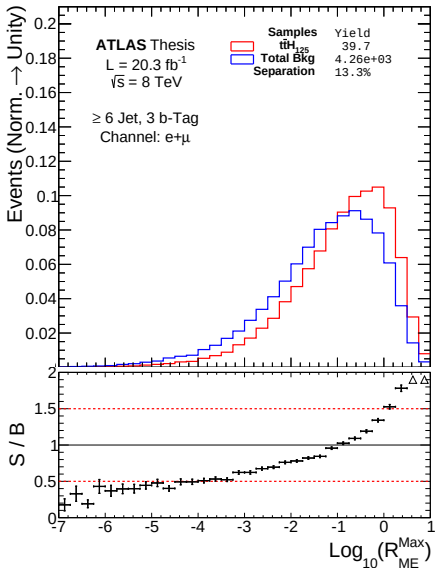
Once the integration was stable and important features of the integrand over the four-momentum phase space were identified, the ability of the MEM to correctly identify and assign the various final-state objects present was quantified. First, we provided our MEM with a generated $t\bar{t}H\left(H \to b\bar{b}\right)$ event. As this event was generated using Monte-Carlo (MC) techniques, we know, *a priori*, from which parents the final-state quarks originated. We then used the MEM to calculate the likelihood for all possible, unique jet-assignments for the present, final-state jets. If the MEM performed well, we would expect the jet-assignment with the maximum likelihood to contain jets close to the initially generated quarks, nearly all of the time. This is not universally true, given low-probability pathological cases that contain far-off-shell parent objects, but is true on balance.

The four-momentum of particles which participate in the hard-process are stored in the simulated event's data structure. These are commonly referred to as 'truth' objects, as they represent the 'true' final-state particles of the hard-process. These are distinct from 'reconstructed' objects which are modeled with the expected detector response of ATLAS and reconstructed with the ATLAS common reconstruction software. As a result, reconstructed objects include resolution, scale, reconstruction, and pile-up effects from ATLAS that don't affect the truth objects.

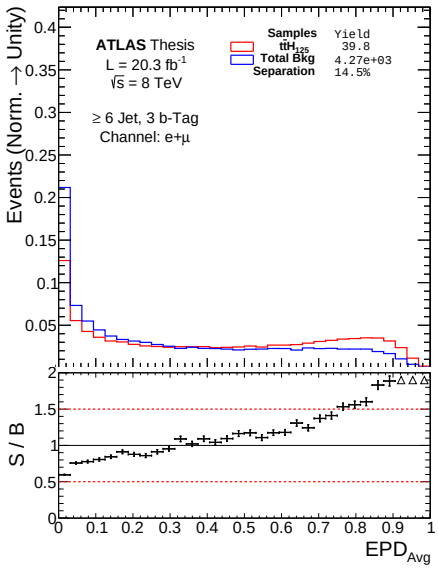
The fraction of times in which the correct object assignment is chosen is known as its assignment efficiency. First, we used the MEM to calculate a likelihood for each jet-assignment. The jet-assignment with the largest likelihood is then deemed the the most likely assignment. Each measured object in that assignment is then 'matched' with the appropriate truth object using a $\Delta R_{\eta\phi}$ cone of 0.4 centered on the truth object. Reconstructed objects assigned to



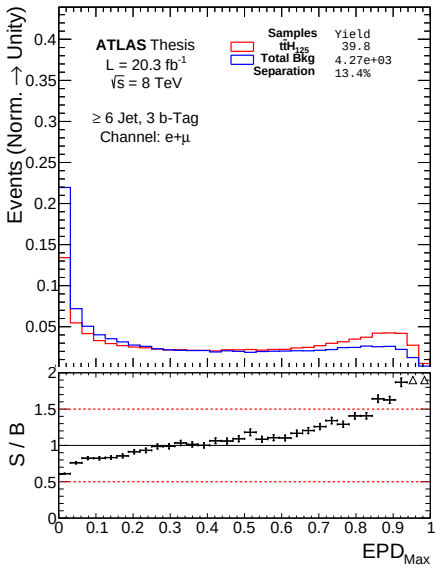
(a) Average L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$)



(b) Max L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$) Assignments

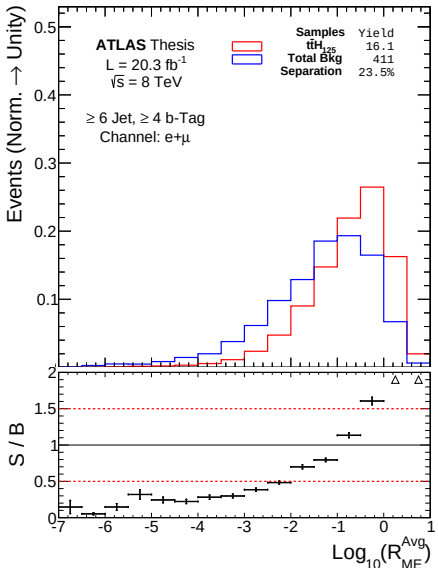


(c) Average L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$)

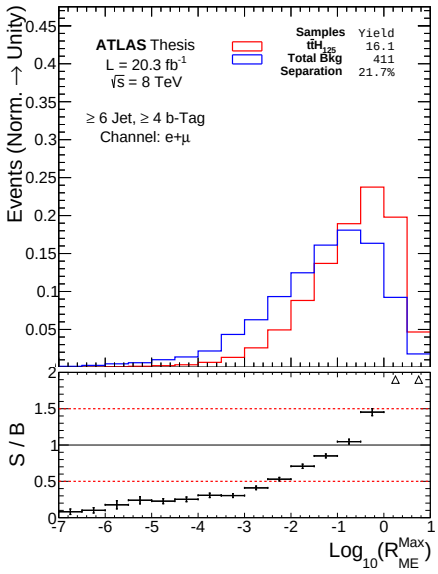


(d) Max L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$) Assignments

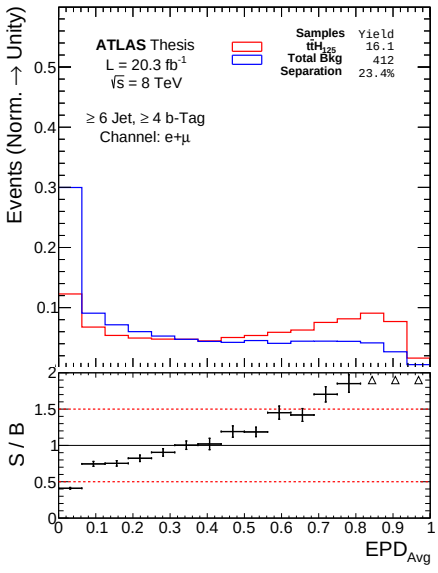
Figure 8.15: Expected discrimination using the "max likelihood jet-assignment" and "average jet-assignment likelihood" representations in conjunction with the $\boldsymbol{\lambda}$ parametrization in the $(\geq 6\text{j}, 3\text{b})$ region. Here the $t\bar{t}H_{125}$ sample and the remaining backgrounds are each normalized to unity so as to show the respective shapes introduced by the discriminating variable. Separation is defined by Eq. 7.10, and is quoted below the yields in the legends. In the ratio plots, signal over background is shown by the black curve, including statistical error.



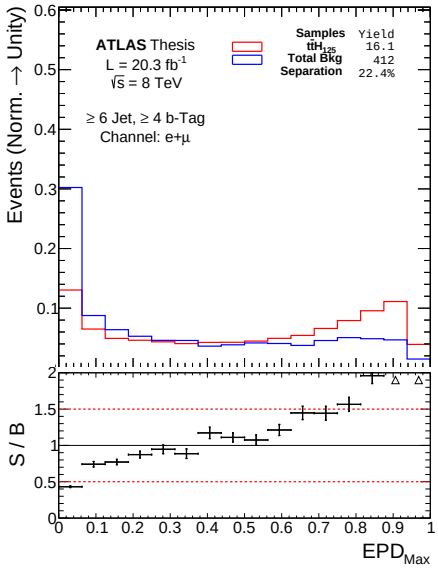
(a) Average L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$)



(b) Max L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$) Assignments

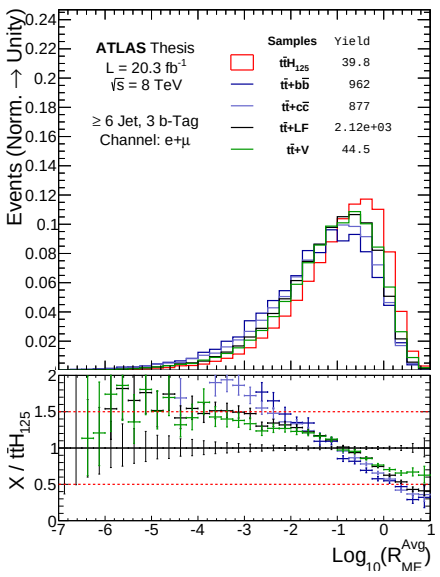


(c) Average L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$)

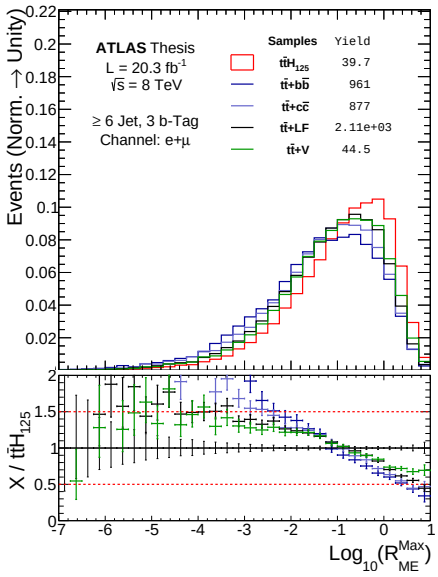


(d) Max L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$) Assignments

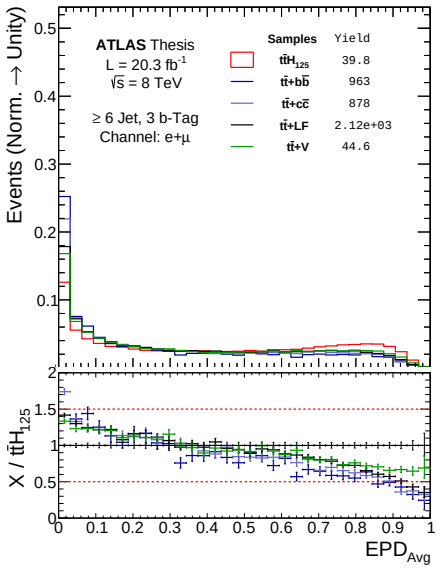
Figure 8.16: Expected discrimination using the "max likelihood jet-assignment" and "average jet-assignment likelihood" representations in conjunction with the $\boldsymbol{\lambda}$ parametrization in the $(\geq 6\text{j}, \geq 4\text{b})$ region. Here the $t\bar{t}H_{125}$ sample and the remaining backgrounds are each normalized to unity so as to show the respective shapes introduced by the discriminating variable. Separation is defined by Eq. 7.10, and is quoted below the yields in the legends. In the ratio plots, signal over background is shown by the black curve, including statistical error.



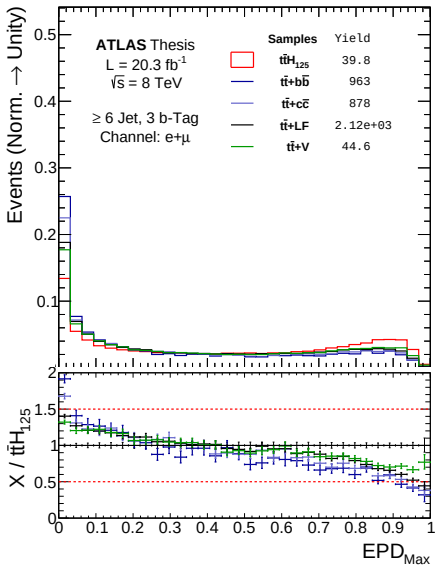
(a) Average L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$)



(b) Max L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$) Assignments

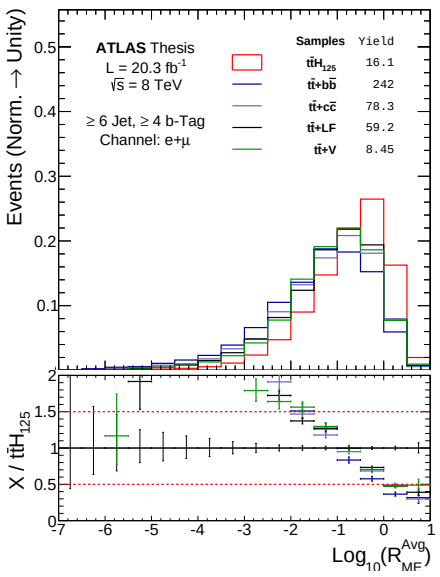


(c) Average L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$)

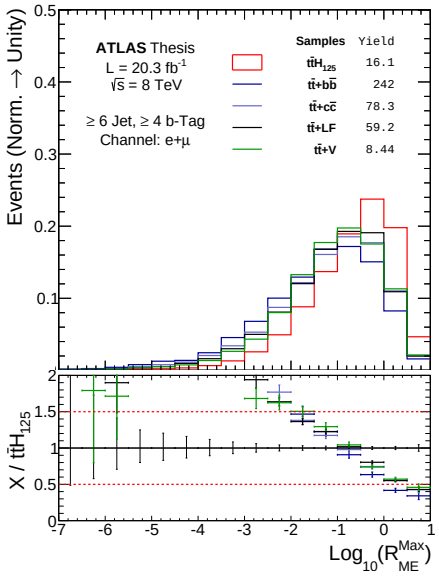


(d) Max L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$) Assignments

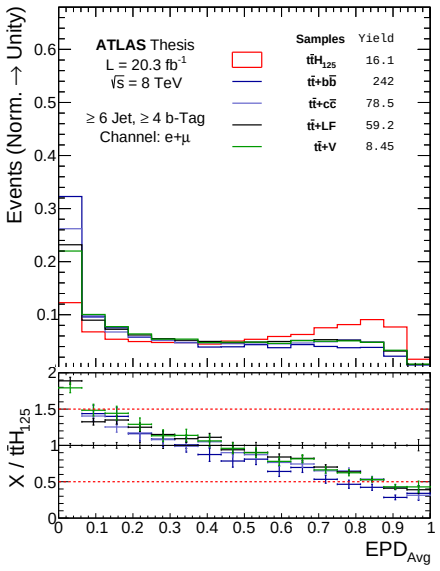
Figure 8.17: Expected discriminant response of the contributing backgrounds using the "max likelihood jet-assignment" and "average jet-assignment likelihood" representations in conjunction with the $\boldsymbol{\lambda}$ parametrization in the ($\geq$ 6j, 3b) region. Here the $t\bar{t}H_{125}$ sample and the remaining backgrounds are each normalized to unity so as to show the respective shapes introduced by the discriminating variable. In the lower ratio plots, each curve represents the X/ $t\bar{t}H_{125}$ ratio corresponding to the included legend.



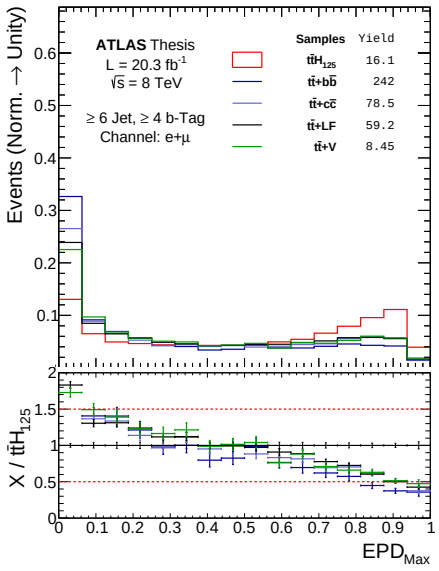
(a) Average L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$)



(b) Max L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$) Assignments



(c) Average L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$)



(d) Max L($t\bar{t}H_{125}$) and L($t\bar{t}+b\bar{b}$) Assignments

Figure 8.18: Expected discriminant response of the contributing backgrounds using the "max likelihood jet-assignment" and "average jet-assignment likelihood" representations in conjunction with the $\boldsymbol{\lambda}$ parametrization in the ($\geq$ 6j, $\geq$ 4b) region. Here the $t\bar{t}H_{125}$ sample and the remaining backgrounds are each normalized to unity so as to show the respective shapes introduced by the discriminating variable. In the lower ratio plots, each curve represents the X/ $t\bar{t}H_{125}$ ratio corresponding to the included legend.

truth particles that lie in this cone are considered to be 'matched,' while those that lie outside the cone are considered 'unmatched.' The denominator is chosen to be the number of events where all the truth objects have been successfully reconstructed and pass the relevant object selection. This denominator represents 29.4% of all $t\bar{t}H\left(H \to b\bar{b}\right)$ events in the $(\geq 6\text{j}, 3\text{b})$ region and 43.3% of all $t\bar{t}H\left(H \to b\bar{b}\right)$ events in the $(\geq 6\text{j}, \geq 4\text{b})$ region. More information on the preparation of this sample is located in Appendix B.

Figures 8.1-8.3, present these assignment efficiencies of the MEM given a sample of events containing the generated truth objects specifically using the $t\bar{t}H\left(H \to b\bar{b}\right)$ matrix elements. The total system assignment efficiency (defined by the fraction of events where every object is correctly assigned, neglecting the neutrino) is shown in Table 8.3. Here we see that, including both lepton channels, the MEM chooses the correct assignment 26.5% and 33.2% of the time, for the $(\geq 6\text{j}, 3\text{b})$ and $(\geq 6\text{j}, \geq 4\text{b})$ regions respectively. Furthermore, we see in Table 8.1 that we correctly identify the Higgs boson 42.9% and 47.6% of the time, for the $(\geq 6\text{j}, 3\text{b})$ and $(\geq 6\text{j}, \geq 4\text{b})$ regions respectively. These key assignment efficiencies modestly outperform competing kinematic assignment and χ^2 methods, though at the expense of considerable additional computation time.

In addition, these studies allow us to identify any performance trends of this technique across different b-tag multiplicity regions and lepton channels. Ultimately, we see that the expected trends are present. We see little difference in the assignment efficiencies between lepton channels and we see diminished performance in the $(\geq 6\text{j}, 3\text{b})$ region as compared with the $(\geq 6\text{j}, \geq 4\text{b})$ region due, in part, to the increased number of unique jet-assignments. Additional truth studies are detailed in Section B.2 and in Section B.3. Of particular note are the assignment efficiencies when the $t\bar{t} + b\bar{b}$ matrix elements are used to choose the best jet-assignment.

Finally, we can use the above assignment efficiencies and truth information to define a confusion matrix. A confusion matrix compares the true jet-assignment with the jet-assignment chosen by the MEM. This allows us to identify circumstances in which the MEM swaps a b-jet from the Higgs with a b-jet from one of the top quarks. This gives us further information regarding the performance of the general technique, identifying potential issues with the assignment pre-selection and detector response. No major pathologies are present.

8.4 Jet-Assignment Pre-Selection Studies

Further studies were performed to identify potential improvements based on optimizing the jet-assignment pre-selection. The premise was that, given known properties of the parent particles, like the Top and W masses, heuristics could be developed to veto obviously incorrect assignments. This would result in fewer total jet-assignments to calculate and a larger relative contribution com-

Truth Object Match [$L^{\max}(t\bar{t}H_{125})$]

Intermediate State Match Efficiency[%]

Channel	W_{H}	W_{L}	$\mathbf{Top_H}$	$\mathbf{Top_L}$	H_{125}
≥ 6 J, $= 3$ b, e	63.4	49.5	55.6	50.0	43.2
≥ 6 J, $= 3$ b, μ	63.8	47.6	55.9	49.7	42.6
≥ 6 J, $= 3$ b, $e + \mu$	63.6	48.6	55.8	49.8	42.9
≥ 6 J, ≥ 4 b, e	69.7	49.9	58.6	51.7	48.0
≥ 6 J, ≥ 4 b, μ	69.7	48.0	59.3	51.3	47.1
≥ 6 J, ≥ 4 b, $e + \mu$	69.7	49.0	58.9	51.5	47.6

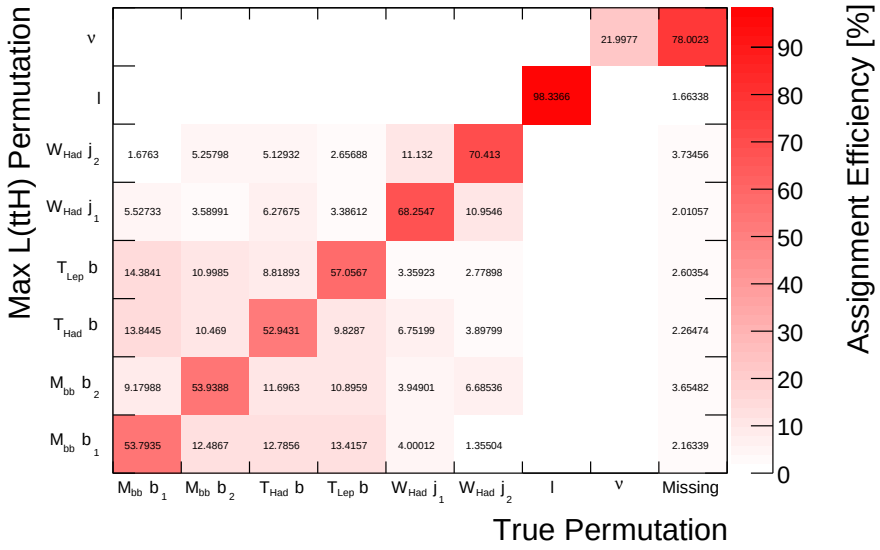
Table 8.1: Intermediate state matching efficiency, in percent, to truth objects, using the maximum signal likelihood jet-assignment. Truth objects are defined as the original particles created by the event generator prior to the detector simulation. The events used to determine these efficiencies are required to have all truth objects be matched to reconstructed jets which successfully passed the initial object selection.

Truth Object Match [$L^{\max}(t\bar{t}H_{125})$]

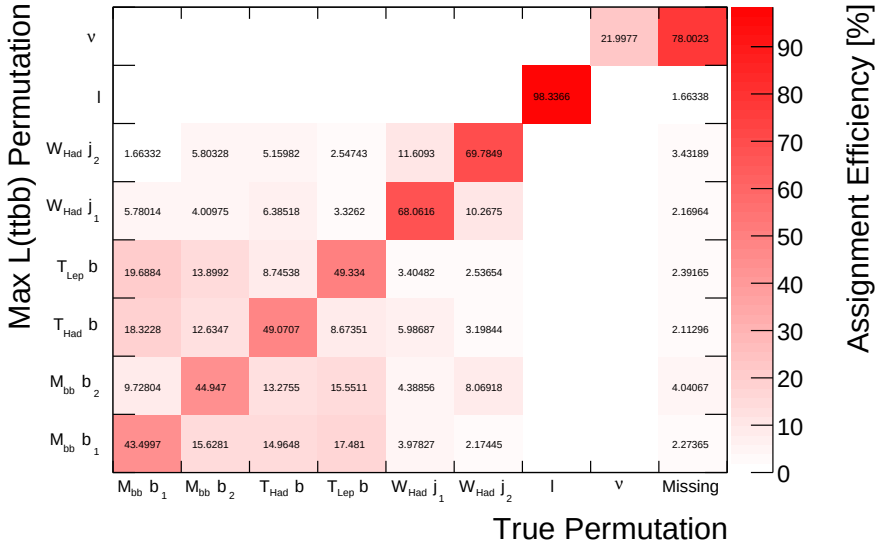
Full Final State Match Efficiency[%]

Channel	**A** H_{125} b	**B** H_{125} b	$\mathbf{Top_H}$ b	$\mathbf{Top_L}$ b	**A** $W_{\mathbf{H}}$ **Jet**	**B** $W_{\mathbf{H}}$ **Jet**	l	ν
≥ 6 J, $= 3$ b, e	53.9	53.7	52.9	57.2	69.0	70.7	98.4	21.7
≥ 6 J, $= 3$ b, μ	53.7	54.2	53.0	56.9	67.5	70.1	98.3	22.3
≥ 6 J, $= 3$ b, $e + \mu$	53.8	53.9	52.9	57.1	68.3	70.4	98.3	22.0
≥ 6 J, ≥ 4 b, e	57.2	58.0	59.7	60.5	74.5	75.0	98.6	21.9
≥ 6 J, ≥ 4 b, μ	56.6	57.5	58.1	60.2	73.5	74.8	98.4	23.0
≥ 6 J, ≥ 4 b, $e + \mu$	56.9	57.7	58.9	60.4	74.0	74.9	98.5	22.4

Table 8.2: Full final state matching efficiency, in percent, to truth objects, using the maximum signal likelihood jet-assignment. Truth objects are defined as the original particles created by the event generator prior to the detector simulation. The events used to determine these efficiencies are required to have all truth objects be matched to reconstructed jets which successfully passed the initial object selection. Jets labeled with an A are the leading jets in p_T, and jets labeled B are sub-leading.

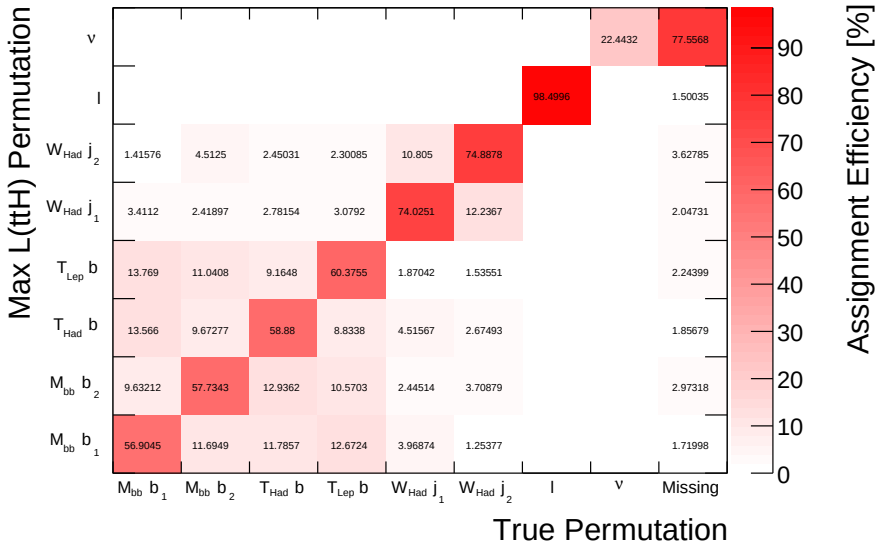


(a) Max L($t\bar{t}H_{125}$) Assignment; ($\geq$ 6j, 3b)

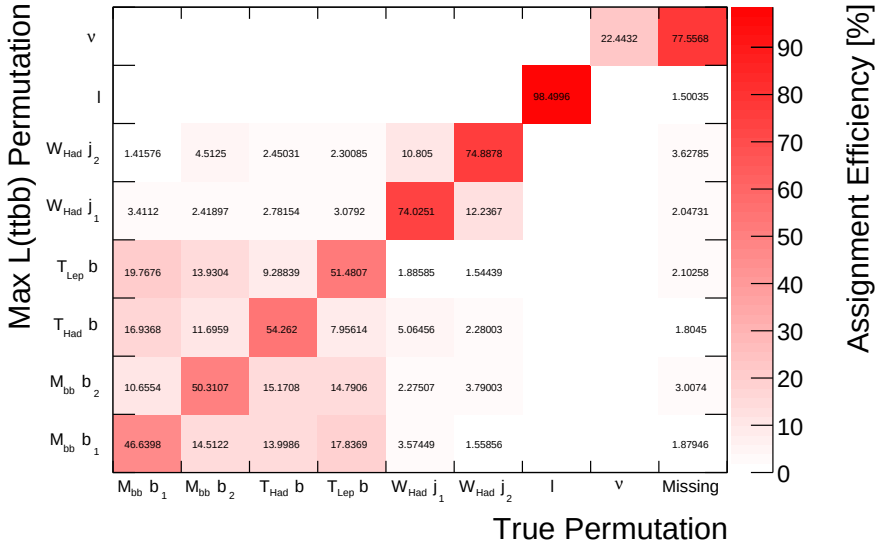


(b) Max L($t\bar{t} + b\bar{b}$) Assignment; ($\geq$ 6j, 3b)

Figure 8.19: Confusion matrices for the maximum $t\bar{t}H_{125}$ likelihood, L($t\bar{t}H_{125}$), and maximum $t\bar{t} + b\bar{b}$ likelihood, L($t\bar{t} + b\bar{b}$), jet-assignments in the ($\geq$ 6j, 3b) region. The abscissa axis categorizes the true identities of the jets, while the ordinate axis gives the assignment chosen by the MEM. Diagonal elements show the level of agreement between the true assignments and those chosen by the MEM. The off-diagonal elements show the prevalence of different assignment pathologies. These matrices allow us to characterize these pathologies in the context of the kinematic transformations. The rightmost column, labeled 'Missing,' shows the percentage of chosen jet-assignments with an object which was not matched to any 'truth' object. For the jets, this could mean that the jet assigned by the MEM may have originated from a radiated parton or from a jet not associated with the initial hard-process.



(a) Max L($t\bar{t}H_{125}$) Assignment; ($\geq$ 6j, $\geq$ 4b)



(b) Max L($t\bar{t} + b\bar{b}$) Assignment; ($\geq$ 6j, $\geq$ 4b)

Figure 8.20: Confusion matrices for the maximum $t\bar{t}H_{125}$ likelihood, L($t\bar{t}H_{125}$), and maximum $t\bar{t} + b\bar{b}$ likelihood, L($t\bar{t} + b\bar{b}$), jet-assignments in the ($\geq$ 6j, $\geq$ 4b) region. The abscissa axis categorizes the true identities of the jets, while the ordinate axis gives the assignment chosen by the MEM. Diagonal elements show the level of agreement between the true assignments and those chosen by the MEM. The off-diagonal elements show the prevalence of different assignment pathologies. These matrices allow us to characterize these pathologies in the context of the kinematic transformations. The rightmost column, labeled 'Missing,' shows the percentage of chosen jet-assignments with an object which was not matched to any 'truth' object. For the jets, this could mean that the jet assigned by the MEM may have originated from a radiated parton or from a jet not associated with the initial hard-process.

Truth Object Match [$L^{\max}(t\bar{t}H_{125})$]

Composite Match Efficiency[%]

Channel	**1** W	**2** W	**1** H_{125}	**2** H_{125}	**2 Top**	**Full Top Had**	$t\bar{t}$	**Total**
≥ 6 J, $= 3$ b, e	80.5	59.3	67.3	40.3	36.2	42.1	30.4	26.2
≥ 6 J, $= 3$ b, μ	78.9	58.7	68.1	39.8	36.3	41.6	30.5	26.8
≥ 6 J, $= 3$ b, $e + \mu$	79.7	59.0	67.7	40.0	36.2	41.8	30.4	26.5
≥ 6 J, ≥ 4 b, e	82.2	67.4	70.7	44.5	43.8	49.3	38.0	33.4
≥ 6 J, ≥ 4 b, μ	82.6	65.7	69.5	44.6	43.7	48.0	37.2	33.0
≥ 6 J, ≥ 4 b, $e + \mu$	82.4	66.5	70.1	44.6	43.7	48.7	37.6	33.2

Table 8.3: Composite state matching efficiency, in percent, to truth objects, using the maximum signal likelihood jet-assignment. Truth objects are defined as the original particles created by the event generator prior to the detector simulation. The events used to determine these efficiencies are required to have all truth objects be matched to reconstructed jets which successfully passed the initial object selection. 1 X, denotes the matching of *at least* one of X's daughters, and 2 X, denotes the matching of both of X's daughters.

ing from the correct jet-assignment. In practice however, it was seen that each formulation of the suggested heuristics performed similarly to or worse than the nominal implementation presented earlier in this chapter, in terms of discriminating power.

Table 8.4 describes the various veto heuristics studied. Table 8.5 shows the results of implementing the veto heuristics. Finally, Table 8.6 shows us the information content of sub-leading jet-assignments. This explains the benefit to using averaged, as opposed to maximum, likelihoods in the construction of the discriminating variables.

Taken all together, it is seen that these veto rules do not significantly improve the measurement, but do have the potential to negatively impact the analysis by reducing the number of selected events by vetoing all the possible jet-assignments. This precise behavior is also found in the $(\geq 6\text{j}, 3\text{b})$ region. As a result, none of these heuristics are implemented in the final analysis. However, this study does provide valuable insight regarding the non-negligible contribution of sub-leading jet-assignments to the final discriminant.

8.5 Maximum Likelihood Jet-Assignment Studies

In addition to studying the separating power of the final discriminants, which are based on the integrated likelihoods, we decided to inspect the kinematic variables derived from the maximum likelihood jet-assignments. If the likelihoods were totally accurate, we would expect any discrimination from these kinematic

Assignment Veto Descriptions

Veto Name	**Requirement**
M_{lb}	$(\boldsymbol{p}_{\text{Lep}} + \boldsymbol{p}_{\text{Top-Lep}b})^2 > m_T^2$
M_{q1b}	$(\boldsymbol{p}_{W\text{-Had}q_1} + \boldsymbol{p}_{\text{Top-Had}b})^2 > m_T^2$
M_{q2b}	$(\boldsymbol{p}_{W\text{-Had}q_2} + \boldsymbol{p}_{\text{Top-Had}b})^2 > m_T^2$
$M_{\text{Top-Lep}}$	! (150 GeV $\leq m_{\text{Top-Lep}} <$ 200 GeV)
$M_{\text{Top-Had}}$	! (150 GeV $\leq m_{\text{Top-Had}} <$ 200 GeV)

Table 8.4: Jet-assignment veto descriptions based on reconstructed Top and W masses. Here $q1$ and $q2$ are the leading and sub-leading light-jets.

Jet-Assignment Pre-Selection Studies – Assignment Vetos

Heuristic	$\mathcal{S}^\dagger$ [%]	$\Delta\mathcal{S}^\dagger$ [%]	**S**	**B**	$\frac{\text{S}}{\sqrt{\text{B}}}$
Nominal	**23.4**	**0.0**	**16.1**	**412**	**0.793**
M_{lb}	23.4	0.0	16.0	409	0.791
M_{q2b}	23.6	0.2	15.9	408	0.787
M_{q2b} & M_{lb}	23.3	-0.1	15.8	404	0.786
M_{q1b}	23.0	-0.4	15.4	391	0.779
M_{q1b} & M_{lb}	23.0	-0.4	15.2	387	0.773
M_{q1b} & M_{q2b}	23.1	-0.3	15.1	382	0.773
M_{q1b} & M_{q2b} & M_{lb}	23.0	-0.4	15.0	377	0.773
$M_{\text{Top-Had}}$	22.0	-1.4	11.3	261	0.699
$M_{\text{Top-Lep}}$	19.3	-4.1	10.3	273	0.623
$M_{\text{Top-Lep}}$ & $M_{\text{Top-Had}}$	18.4	-5.0	6.89	160	0.545

Table 8.5: Jet-assignment pre-selection study results for the ($\geq$ 6j, $\geq$ 4b) region. The separation, $\mathcal{S}^\dagger$, is defined as before, in Eq. 7.10. The signal yields, S, and background yields, B, are also shown as well as the total region significance defined as $\text{S}/\sqrt{\text{B}}$. The results are sorted in descending order based on the veto's signal selection efficiency.

Jet-Assignment Pre-Selection Studies – Number of Assignments

Heuristic	$\mathcal{S}^\dagger$ [%]	$\Delta\mathcal{S}^\dagger$ [%]	**S**	**B**	$\frac{\text{S}}{\sqrt{\text{B}}}$
Top 1	22.3	-1.1	16.1	412	0.793
Top 3	23.1	-0.3	16.1	412	0.793
Top 6	23.4	0.0	16.1	412	0.793
Top 9	23.4	0.0	16.1	412	0.793
All 12	23.4	0.0	16.1	412	0.793

Table 8.6: Separation as a function of number of averaged ranked jet-assignments for the ($\geq$ 6j, $\geq$ 4b) region. This shows us the information content of sub-leading jet-assignments which explains the benefit to using averaged, as opposed to maximum, likelihoods in the construction of the discriminating variables.

variables to be very strongly correlated with the discrimination in the final discriminants. However, we know that many approximations were utilized in the calculation of the likelihoods. Thus it was hypothesized that some information lost to the approximations may be recoverable through the kinematics of the reconstructed maximum likelihood jet-assignments. This hypothesis was studied and found to be a promising avenue to augment the MEM.

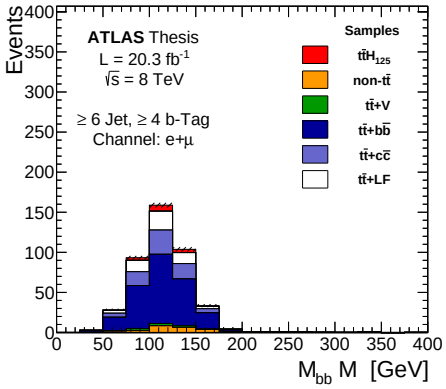
The following plots will show that many kinematic variables provide discrimination power that is largely orthogonal to the information contained in the likelihoods. Later chapters will explore ways to include these kinematic variables into the final multi-variate analysis. For all ($\geq$ 6j, $\geq$ 4b) region figures explicitly mentioned below, identical figures are included for the ($\geq$ 6j, 3b) region.

Figure 8.21 shows the expected yields for the reconstructed parent-particle masses derived from the maximum $t\bar{t}H_{125}$ likelihood jet-assignment. Figure 8.23 shows the same under the maximum $t\bar{t} + b\bar{b}$ likelihood jet-assignment. These plots show good agreement between the expected particle masses and those reconstructed from the MEM assignments.

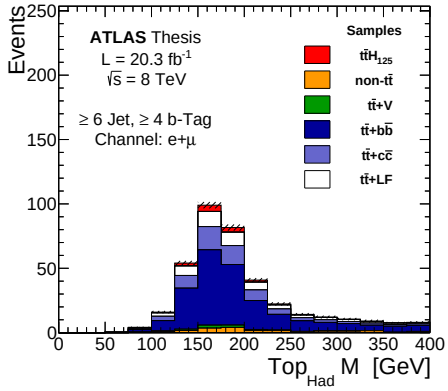
Figure 8.25 and Figure 8.28 show the discrimination resulting from the differing kinematics of the maximum likelihood jet-assignments for the $t\bar{t}H_{125}$ and $t\bar{t} + b\bar{b}$ matrix elements respectively. Of particular note is the invariant mass of the pair of b-tagged jets not associated with the top quark. This mass, M_{bb}, behaves very differently between the maximum likelihood $t\bar{t}H_{125}$ jet-assignment, 8.26(a), and the maximum likelihood $t\bar{t}+b\bar{b}$ jet-assignment, 8.28(a). This is not surprising as the signal likelihood expects a 125 GeV Higgs, where the other does not. What is surprising however, is the amount of separation found in the maximum likelihood background jet-assignment, 8.28(a). We believe this structure results from the $t\bar{t}+b\bar{b}$ matrix elements preferentially selecting the b-tagged jets to satisfy the top quark mass constraints. This then leaves the correct two Higgs b-tagged jets to be reconstructed into the M_{bb} excess seen at 125 GeV. Similar differences are found in other kinematic variables as well, including the difference in η between the reconstruvted M_{bb} particle, and the reconstructed top quark farthest, in angular distance, from the M_{bb}, referred to as $\text{top}_{\text{Max}\Delta R}$, as well as the $\cos\theta$ of the leading b-tagged jet assigned to the M_{bb} pair.

8.6 Data and Simulation Agreement

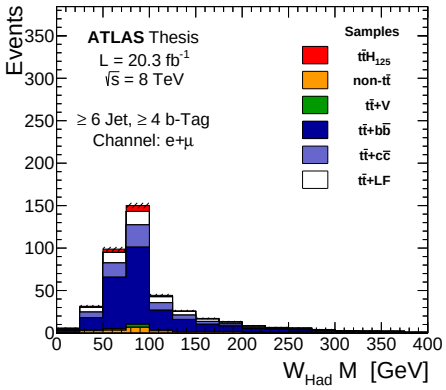
Ultimately, the MEM had to be validated with actual ATLAS data. The purpose of this validation was to determine the fidelity of the simulated backgrounds with respect to the data in the absence of potential signal events. In addition, it was important to identify any unexpected issues that might cause the MEM to behave differently for data as compared to the modeled backgrounds. This would be evident in the bin-by-bin ratio of the background yields and the data yields, typically represented by an additional ratio plot. This 'data to Monte-



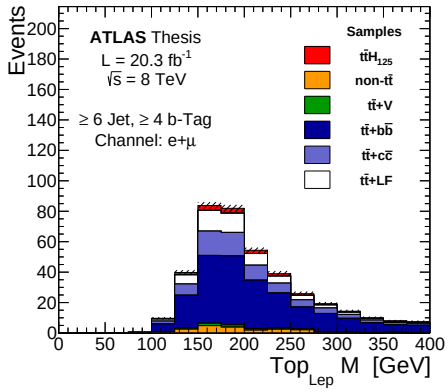
(a) Max L($t\bar{t}H_{125}$) Assignment



(b) Max L($t\bar{t}H_{125}$) Assignment

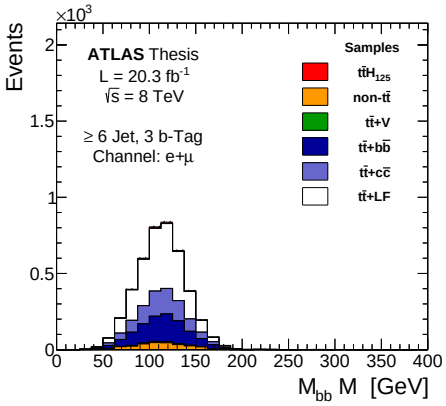


(c) Max L($t\bar{t}H_{125}$) Assignment

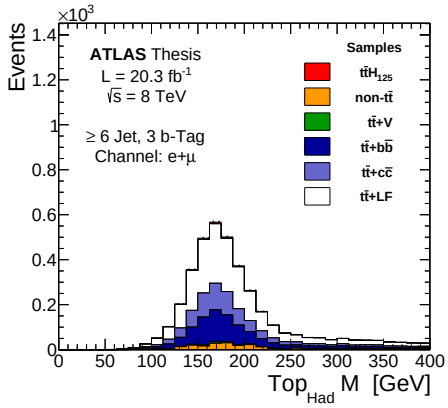


(d) Max L($t\bar{t}H_{125}$) Assignment

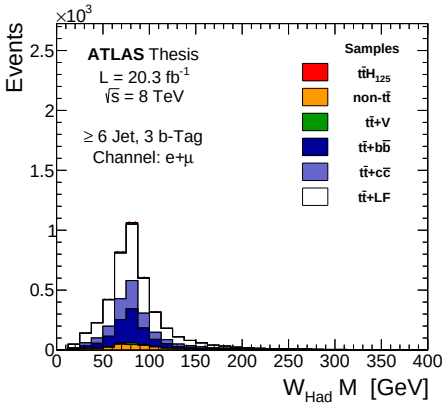
Figure 8.21: Expected kinematic distributions from maximum $t\bar{t}H_{125}$ likelihood jet-assignments in the ($\geq$ 6j, $\geq$ 4b) region. These plots show reconstructed parent masses consistent with those present in the $t\bar{t}H_{125}$ matrix element.



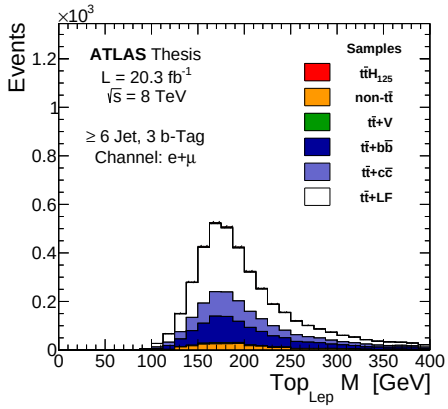
(a) Max L($t\bar{t}H_{125}$) Assignment



(b) Max L($t\bar{t}H_{125}$) Assignment

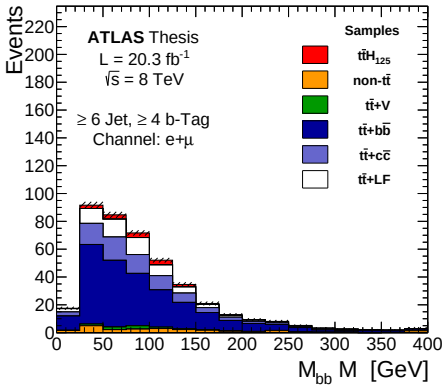


(c) Max L($t\bar{t}H_{125}$) Assignment

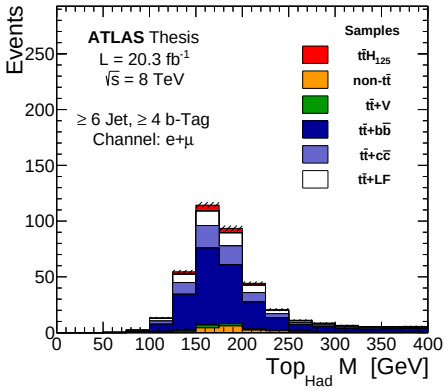


(d) Max L($t\bar{t}H_{125}$) Assignment

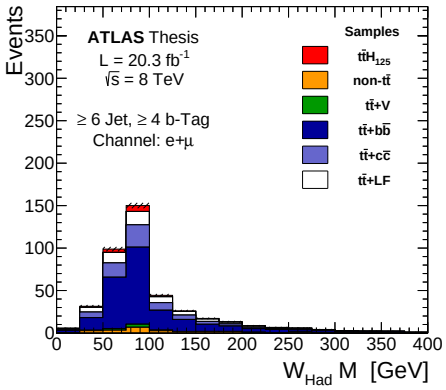
Figure 8.22: Expected kinematic distributions from maximum $t\bar{t}H_{125}$ likelihood jet-assignments in the $(\geq 6\text{j}, 3\text{b})$ region. These plots show reconstructed parent masses consistent with those present in the $t\bar{t}H_{125}$ matrix element.



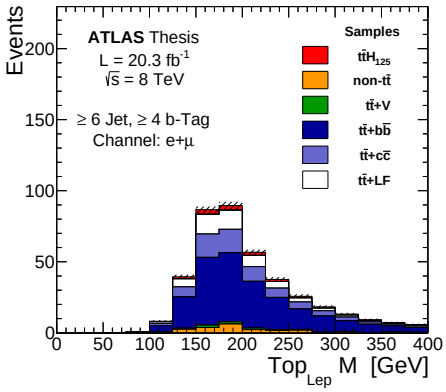
(a) Max L($t\bar{t}+b\bar{b}$) Assignment



(b) Max L($t\bar{t}+b\bar{b}$) Assignment

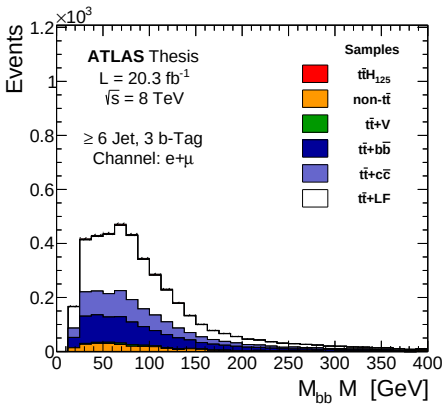


(c) Max L($t\bar{t}+b\bar{b}$) Assignment

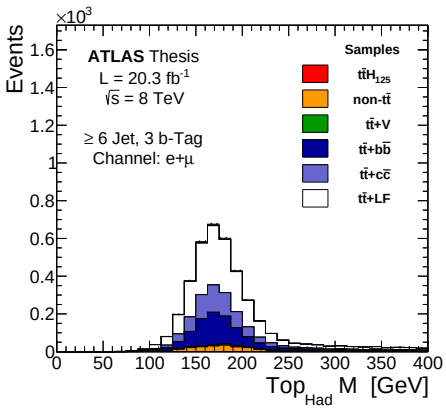


(d) Max L($t\bar{t}+b\bar{b}$) Assignment

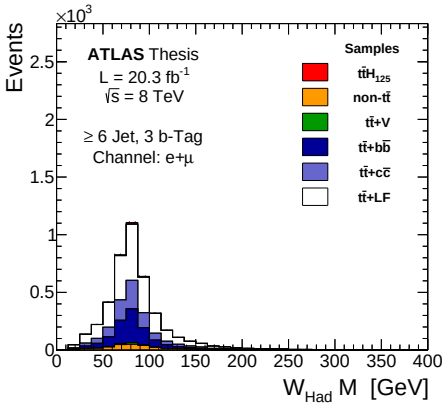
Figure 8.23: Expected kinematic distributions from maximum $t\bar{t}+b\bar{b}$ likelihood jet-assignments in the ($\geq$ 6j, $\geq$ 4b) region. These plots show reconstructed parent masses consistent with those present in the $t\bar{t}+b\bar{b}$ matrix element.



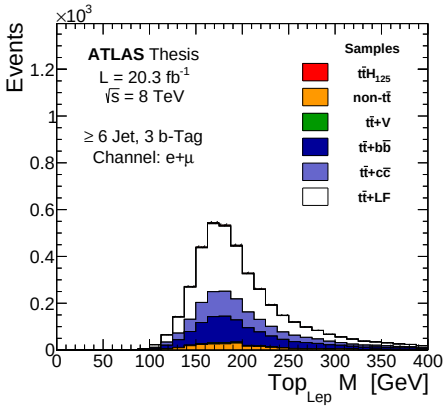
(a) Max L($t\bar{t}+b\bar{b}$) Assignment



(b) Max L($t\bar{t}+b\bar{b}$) Assignment

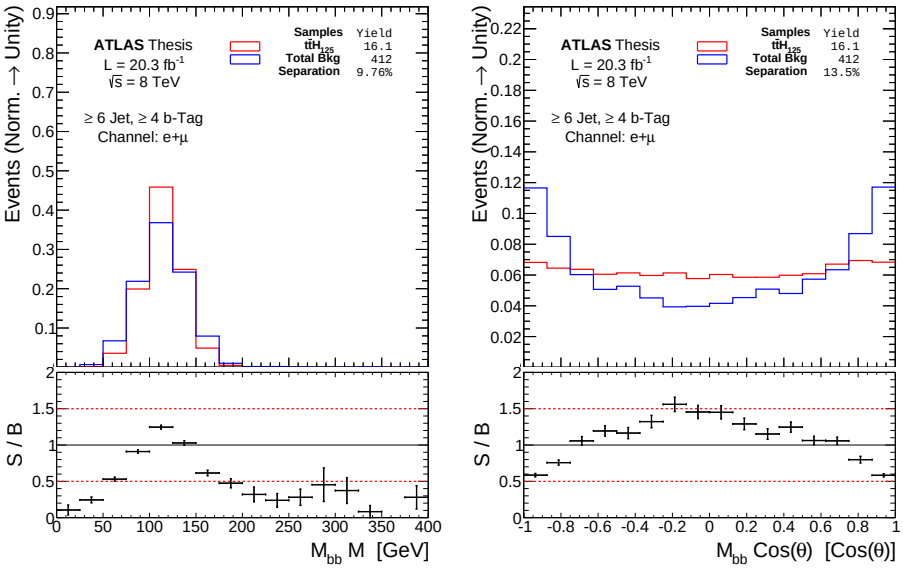


(c) Max L($t\bar{t}+b\bar{b}$) Assignment



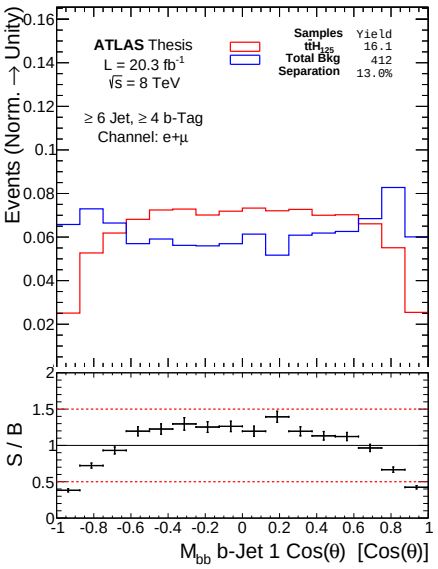
(d) Max L($t\bar{t}+b\bar{b}$) Assignment

Figure 8.24: Expected kinematic distributions from maximum $t\bar{t}+b\bar{b}$ likelihood jet-assignments in the $(\geq 6\text{j}, 3\text{b})$ region. These plots show reconstructed parent masses consistent with those present in the $t\bar{t}+b\bar{b}$ matrix element.

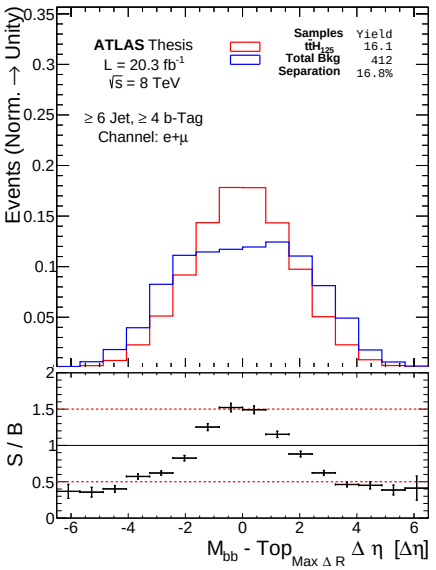


(a) Max L($t\bar{t}H_{125}$) Assignment

(b) Max L($t\bar{t}H_{125}$) Assignment

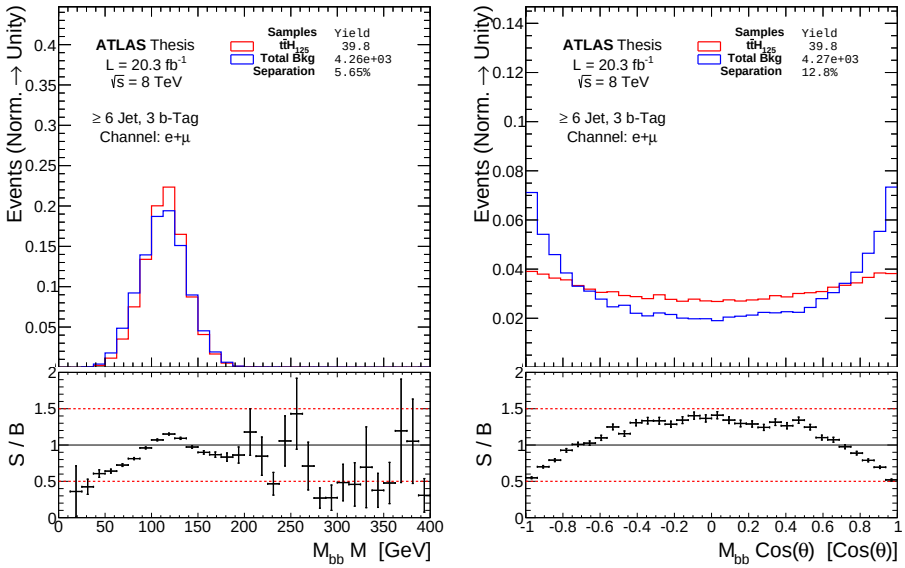


(c) Max L($t\bar{t}H_{125}$) Assignment



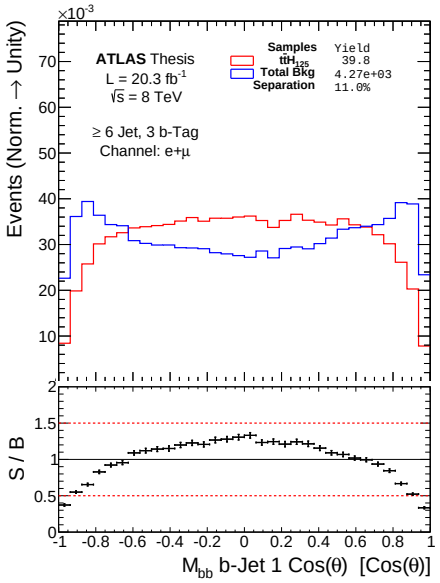
(d) Max L($t\bar{t}H_{125}$) Assignment

Figure 8.25: Discriminating kinematic responses from maximum $t\bar{t}H_{125}$ likelihood jet-assignments in the ($\geq$ 6j, $\geq$ 4b) region. These plots show selected kinematic variables which exhibit significant discrimination. Here the $t\bar{t}H_{125}$ sample and the remaining backgrounds are each normalized to unity so as to show the respective shapes of the different samples. Separation is defined by Eq. 7.10, and is quoted below the yields in the legends. In the ratio plots, signal over background is shown by the black curve, including statistical error.

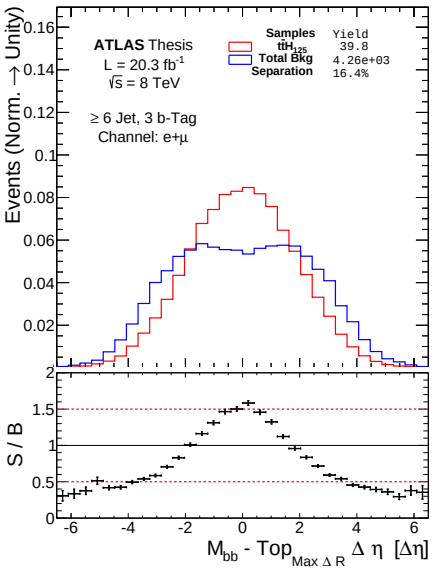


(a) Max L($t\bar{t}H_{125}$) Assignment

(b) Max L($t\bar{t}H_{125}$) Assignment

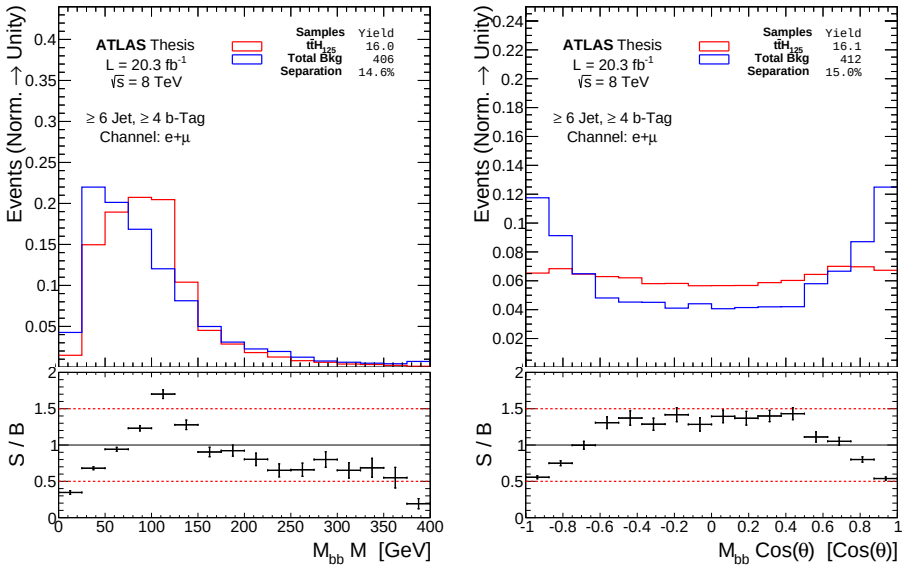


(c) Max L($t\bar{t}H_{125}$) Assignment



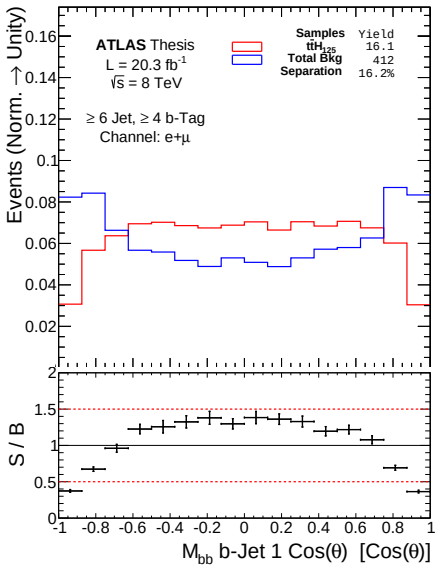
(d) Max L($t\bar{t}H_{125}$) Assignment

Figure 8.26: Discriminating kinematic responses from maximum $t\bar{t}H_{125}$ likelihood jet-assignments in the ($\geq$ 6j, 3b) region. These plots show selected kinematic variables which exhibit significant discrimination. Here the $t\bar{t}H_{125}$ sample and the remaining backgrounds are each normalized to unity so as to show the respective shapes of the different samples. Separation is defined by Eq. 7.10, and is quoted below the yields in the legends. In the ratio plots, signal over background is shown by the black curve, including statistical error.

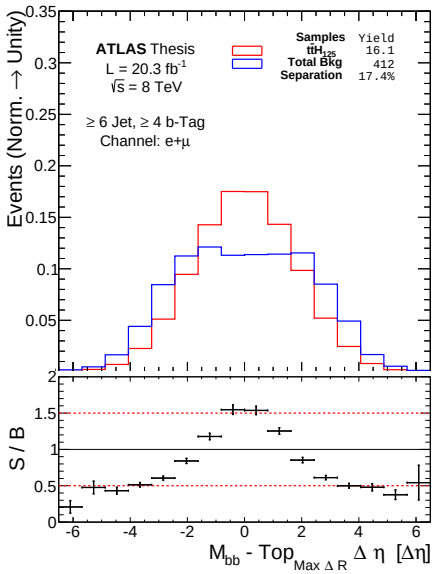


(a) Max L($t\bar{t} + b\bar{b}$) Assignment

(b) Max L($t\bar{t} + b\bar{b}$) Assignment

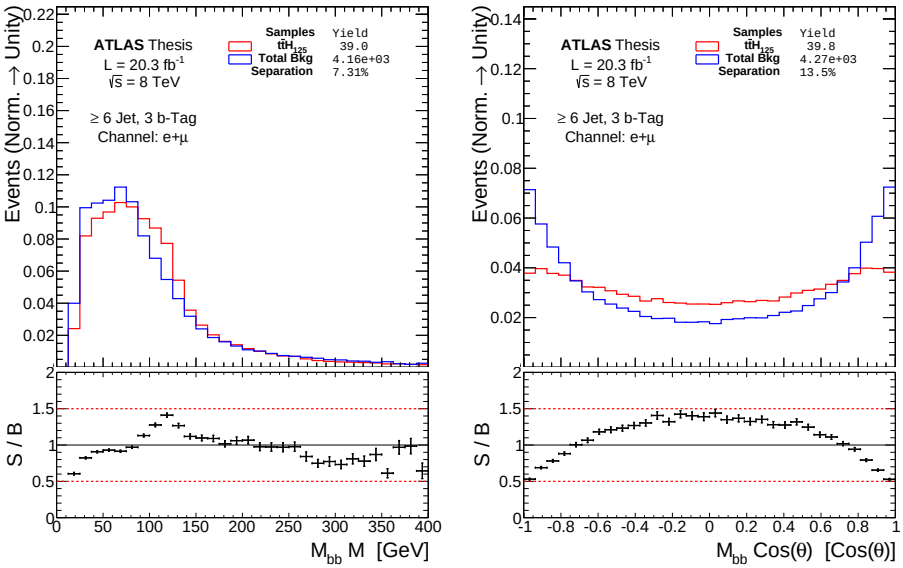


(c) Max L($t\bar{t} + b\bar{b}$) Assignment



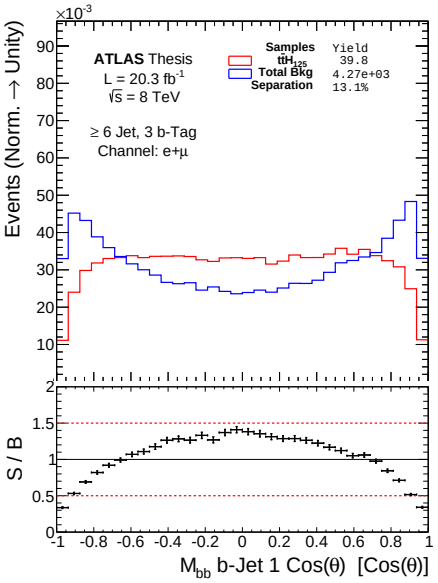
(d) Max L($t\bar{t} + b\bar{b}$) Assignment

Figure 8.27: Discriminating kinematic responses from maximum $t\bar{t} + b\bar{b}$ likelihood jet-assignments in the ($\geq$ 6j, $\geq$ 4b) region. These plots show selected kinematic variables which exhibit significant discrimination. Here the $t\bar{t} + b\bar{b}$ sample and the remaining backgrounds are each normalized to unity so as to show the respective shapes of the different samples. Separation is defined by Eq. 7.10, and is quoted below the yields in the legends. In the ratio plots, signal over background is shown by the black curve, including statistical error.

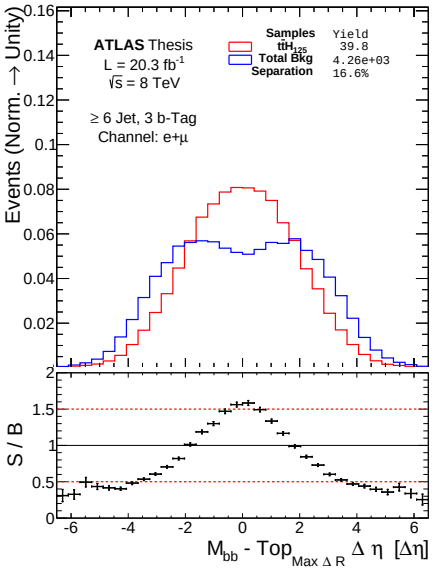


(a) Max L($t\bar{t} + b\bar{b}$) Assignment

(b) Max L($t\bar{t} + b\bar{b}$) Assignment

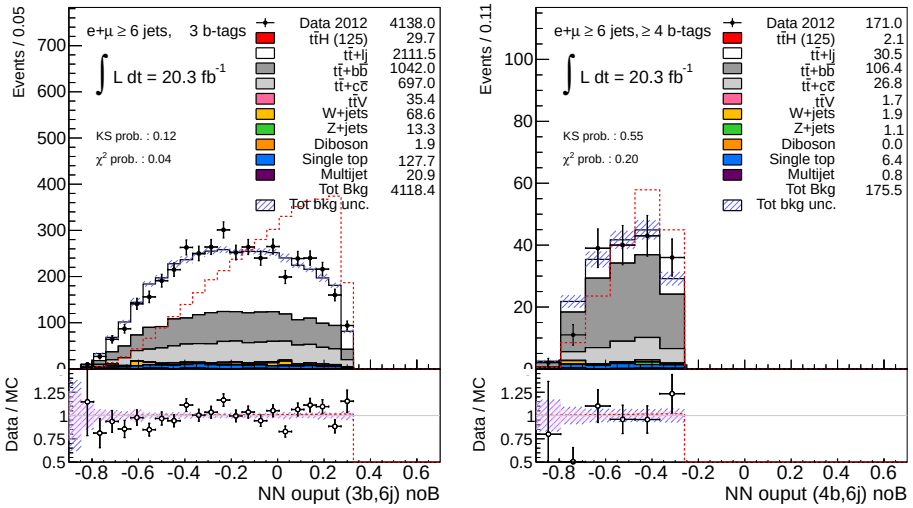


(c) Max L($t\bar{t} + b\bar{b}$) Assignment



(d) Max L($t\bar{t} + b\bar{b}$) Assignment

Figure 8.28: Discriminating kinematic responses from maximum $t\bar{t} + b\bar{b}$ likelihood jet-assignments in the ($\geq$ 6j, 3b) region. These plots show selected kinematic variables which exhibit significant discrimination. Here the $t\bar{t} + b\bar{b}$ sample and the remaining backgrounds are each normalized to unity so as to show the respective shapes of the different samples. Separation is defined by Eq. 7.10, and is quoted below the yields in the legends. In the ratio plots, signal over background is shown by the black curve, including statistical error.



(a) $(\geq 6\text{j}, 3\text{b})$ Region

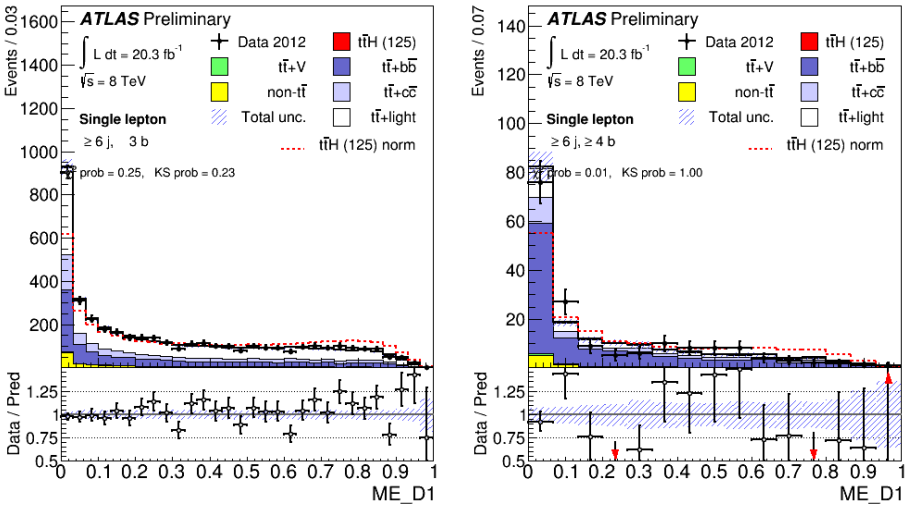
(b) $(\geq 6\text{j}, \geq 4\text{b})$ Region

Figure 8.29: Blinded kinematic neural network distributions showing the signal-depleted regions used for simulation validation in the $(\geq 6\text{j}, 3\text{b})$ region and the $(\geq 6\text{j}, \geq 4\text{b})$ region.

Carlo' agreement thus becomes a crucial validation step prior to attempting to identify the presence of a signal.

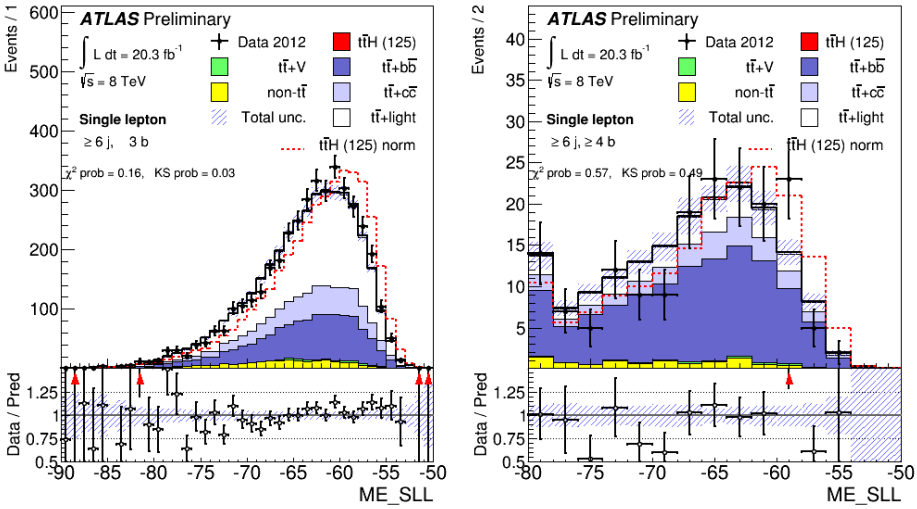
The validation of data to Monte-Carlo agreement is performed on pre-defined control regions so as not to bias the final result. The data is categorized into two distinct sub-regions in the $(\geq 6\text{j}, \geq 4\text{b})$ and $(\geq 6\text{j}, 3\text{b})$ regions. These sub-regions are defined using a cut on the artificial neural network output designated for those regions resulting in an unblinded, signal-depleted 'control region' and a blinded, signal-enriched 'measurement region.' This cut ensures that the two unblinded control regions contain no more than 2% signal contamination. Figure 8.29 shows the unblinded, signal-depleted control regions used for validation. The artificial neural network used to make these cuts is defined in Reference [93].

Using the blinded data and simulation events selected by the artificial neural network cut, two variables of particular importance are validated. The first is the final discriminant, EPD_{ME}^{Avg}, defined above. The blinded control plots for this variable are shown in Figure 8.30. The second is the common logarithm of the sum of all jet-assignment likelihoods, named SSLL. The blinded control plots for this variable are shown in Figure 8.31. These two variables, in particular, are included in the original artificial neural network based on their mutual information content. This neural network will be described in more detail in Section 9.1.



(a) $(\geq 6\text{j}, 3\text{b})$ Region (b) $(\geq 6\text{j}, \geq 4\text{b})$ Region

Figure 8.30: Blinded control plots for the EPD_{ME}^{Avg} variable in the $(\geq 6\text{j}, 3\text{b})$ region and the $(\geq 6\text{j}, \geq 4\text{b})$ region. Generally good agreement is seen. Slight excesses in the $(\geq 6\text{j}, 3\text{b})$ region were traced to an anomalous excess in the muon channel common to all analyses studying this particular final-state topology.



(a) $(\geq 6\text{j}, 3\text{b})$ Region (b) $(\geq 6\text{j}, \geq 4\text{b})$ Region

Figure 8.31: Blinded control plots for the SSLL variable in the $(\geq 6\text{j}, 3\text{b})$ region and the $(\geq 6\text{j}, \geq 4\text{b})$ region. Generally good agreement is seen. Slight excesses in the $(\geq 6\text{j}, 3\text{b})$ region were traced to an anomalous excess in the muon channel common to all analyses studying this particular final-state topology.

Chapter 9

Multivariate Analysis Application

From the outset of this analysis, it was clear that significant effort would have to be expended to identify powerful discriminating variables to separate signal from background. Fortunately, the richness of the $t\bar{t}H\left(H \to b\bar{b}\right)$ final-states prove to be nearly ideal cases for leveraging multivariate techniques. These techniques combine many variables with marginal separating power into a single output distribution with far superior separating power.

Of the nine analysis regions, defined in Table 6.3, four utilize an advanced multivariate analysis technique with which to discriminate between signal and background events, shown in Table 6.4. Two of these four regions are analyzed with the Matrix Element Method (MEM): the 6 jet inclusive, 3 b-tag, ($\geq$ 6j, 3b) , region and the 6 jet inclusive, 4 b-tag inclusive, ($\geq$ 6j, $\geq$ 4b) region. In both regions, the MEM information is incorporated into an MVA. The resulting output distributions are then used for the final fit and subsequent measurement. This chapter introduces two distinct methods for utilizing this MEM information.

9.1 Neural Networks with Matrix Element Information

The first method utilizes two distinctly trained Artificial Neural Networks (NNs), one for the ($\geq$ 6j, 3b) region and one for the ($\geq$ 6j, $\geq$ 4b) region. An Artificial Neural Network, sometimes referred to as a Multilayer Perceptron, is a multivariate technique which uses a layered topology of artificial neurons for function approximation.

The NN illustrated in Figure 9.1 is a cartoon of a three layer, feed forward neural network, with a $N = \{2, 4, 1\}$ node topology. This network utilizes nodes which mimic the biologic function of neurons, the core components of biological nervous systems, called perceptrons. Each perceptron is comprised of a characteristic univariate *activation function*, $A(x)$, and a vector of weights $\mathbf{w} = \{w_1, w_2, \cdots, w_n\}$ associated with each input. This *activation function* takes as input the dot product of the vector of weights and the vector of inputs from the preceding node, and returns a real valued scalar to pass to subsequent nodes.

The neural network used in this analysis is implemented by NeuroBayes [94],

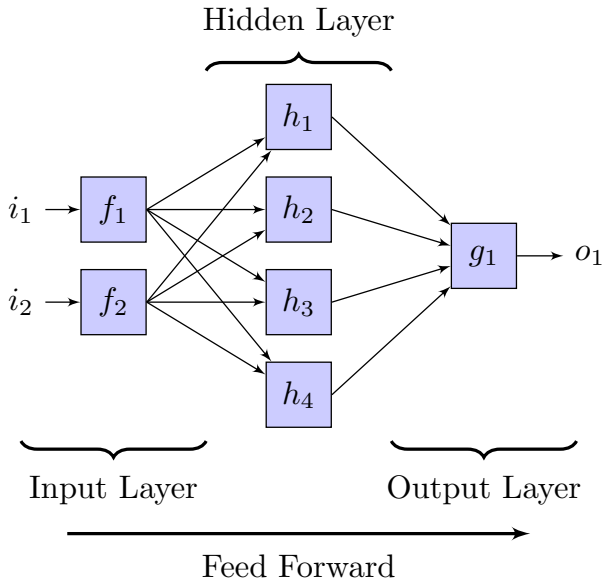


Figure 9.1: A diagram of a three layer artificial neural network with two input nodes, four hidden layer nodes, and a single output node operating in a feed-forward configuration.

a propriety, general purpose multivariate analysis framework. It is configured as a 3-layer, feed-forward network constructed from N_i input nodes, $N_i + 2$ hidden layer nodes, and a single output node, where N_i is equal to the number of defined inputs. Figure 9.1 is a representation of this network when $N_i = 2$. The neural network is trained using a backwards-propagated, supervised approach. The activation functions used are sigmoid functions defined by the hyperbolic tangent of the argument. Further, the optimization algorithm that drives the learning process utilizes an approximated Newton's method referred to as the BFGS algorithm [95]. Other hyper-parameters for the NeruoBayes networks are shown in Table 9.1.

For training, a specialized Monte-Carlo (MC) training sample is generated for each of the $t\bar{t}H$, $t\bar{t}+b\bar{b}$, $t\bar{t}+c\bar{c}$, and $t\bar{t}+$Light-Flavor processes. The training MC samples have the same number of events as the MC samples used in the measurement though they were generated independently. However, these samples are simulated using the fast-simulation configuration of the ATLAS detector model which approximates the calorimeter response resulting in a marginally inferior simulation, though of sufficient quality for training purposes. This NN is explicitly trained utilizing a signal sample of $t\bar{t}H_{125}$ events and a background sample of $t\bar{t}+b\bar{b}$ events.

During the training phase, signal and background events in the training samples are *labeled* with a numerical mapping, $f(\mathbf{i}) = \{-1, 1\}$, such that for signal events, $f(\mathbf{i}_{\text{Sig}}) = 1$, and for background events, $f(\mathbf{i}_{\text{Bkg}}) = -1$. This function replicates a classification scheme that the NN attempts to approximate as it learns. The final output following training, $F(\mathbf{i})$, maps a given input,

NeuroBayes Hyper-Parameters

Hyper-parameter	**Value**
# of variables	N_I
# nodes in hidden layer	$N_I + 2$
Update weights interval	50 events
Learning speed factor	1
Maximum learning speed	0.01
# of iterations	100

Table 9.1: Full details of the NeuroBayes settings for the four Neural Networks.[2]

i, of unknown origin (background or signal) to some real number between -1 and 1 with signal events populating the region near 1 and background events populating the region near -1, under ideal circumstances.

9.1.1 NN Variable Selection

To select the input variables used for each of the NNs, an interactive process is used to identify the most powerful ensemble of variables from an exhaustive list of possible inputs. This is accomplished using NeuroBayes' input variable ranking feature. During this process, the input variables are first ranked by their separation power defined similarly to Eq. 7.10. Then they are transformed such that the variance of each variable's distribution is equal to unity and mean is equal to zero. Once the inputs are prepared, a succession of runs are conducted for which one variable is excluded at a time. The results of these runs allow for the ranking of the variables in terms of final separation dependence, assuming the input variables are sufficiently uncorrelated. The least important variable is dropped and the process starts over with the smaller ensemble of inputs. This procedure is followed until significant losses in separation power are identified.

In addition to 2 MEM variables, 10 additional kinematic variables are identified as having considerable impact on the separation values in both the $(\geq 6\text{j}, 3\text{b})$ region and the $(\geq 6\text{j}, \geq 4\text{b})$ region. This method is also utilized to construct the two other NNs in the $(5\text{j}, 3\text{b})$ and $(5\text{j}, \geq 4\text{b})$ regions, though these are not trained with any MEM inputs. Table 9.2 defines all the variables used for all four NNs, with the MEM inputs highlighted explicitly. Table 9.3 shows the respective rankings of these variable in the context of each regions NN. Finally, Figures 9.2 and 9.3 respectively show the pre-fit and post-fit data to Monte-Carlo agreement of the input variables, whereas Figure 9.4 shows the separations provided by all the variables used by the $(\geq 6\text{j}, \geq 4\text{b})$ region's NN, in order of importance as determined by NeuroBayes.

Input Variables used for the Four Neural Networks

Variable	**Definition**
$^{\dagger}\mathrm{EPD}_{ME}$	Newman-Pearson MEM discriminant
$^{\dagger}SSLL$	Summed MEM signal log-likelihoood
Centrality	Sum of the p_{T} divided by sum of the E for all jets and the lepton
H1	Second Fox-Wolfram moment computed using all jets and the lepton
$m_{\mathrm{bb}}^{\mathrm{min\ \Delta R}}$	Mass of the combination of two b-tagged jets with the smallest ΔR
N_{40}^{jet}	Number of jets with $p_{\mathrm{T}} \geq 40$ GeV
$\Delta R_{\mathrm{bb}}^{\mathrm{avg}}$	Average ΔR for all b-tagged jet pairs
$m_{\mathrm{jj}}^{\mathrm{max\ p_T}}$	Mass of the combination of any two jets with the largest vector sum p_{T}
$\mathrm{Aplanarity}_{\mathrm{b-jet}}$	$1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor built with only b-tagged jets
$H_{\mathrm{T}}^{\mathrm{had}}$	Scalar sum of jet p_{T}
$m_{\mathrm{jj}}^{\mathrm{min\ \Delta R}}$	Mass of the combination of any two jets with the smallest ΔR
$\Delta R_{\mathrm{lep-bb}}^{\mathrm{min\ \Delta R}}$	ΔR between the lepton and the combination of two b-tagged jets with the smallest ΔR
$m_{\mathrm{bj}}^{\mathrm{min\ \Delta R}}$	Mass of the combination of a b-tagged jet and any jet with the smallest ΔR
$m_{\mathrm{bj}}^{\mathrm{max\ p_T}}$	Mass of the combination of a b-tagged jet and any jet with the largest vector sum p_{T}
$m_{\mathrm{uu}}^{\mathrm{min\ \Delta R}}$	Mass of the combination of two untagged jets with the smallest ΔR
$p_{\mathrm{T}}^{\mathrm{jet5}}$	Fifth leading jet p_{T}
$\Delta R_{\mathrm{bb}}^{\mathrm{max\ p_T}}$	ΔR between two b-tagged jets with the largest vector sum p_{T}
$m_{\mathrm{bb}}^{\mathrm{max\ m}}$	Mass of the combination of two b-tagged jets with the largest invariant mass
$p_{T,\mathrm{uu}}^{\mathrm{min\ \Delta R}}$	Scalar sum of the p_{T}'s of the pair of untagged jets with the smallest ΔR
m_{jjj}	Mass of the jet triplet with the largest vector sum p_{T}
$\Delta R_{\mathrm{uu}}^{\mathrm{min\ \Delta R}}$	Minimum ΔR between two untagged jets
$m_{\mathrm{bb}}^{\mathrm{max\ p_T}}$	Mass of the combination of two b-tagged jets with the largest vector sum p_{T}

Table 9.2: The variables considered for use within the Neural Network analysis. Variables designated with a † utilize information from the Matrix Element Method. [2]

Input Variable Assignments and Rankings for the Four Neural Networks

	Analysis Regions			
Variable	$(\geq 6\text{j}, \geq 4\text{b})$	$(\geq 6\text{j}, 3\text{b})$	$(5\text{j}, \geq 4\text{b})$	$(5\text{j}, 3\text{b})$
$^{\dagger}\text{EPD}_{ME}$	1	10	-	-
Centrality	2	2	1	-
$p_{\text{T}}^{\text{jet5}}$	3	7	-	-
$H1$	4	3	2	-
$\Delta R_{\text{bb}}^{\text{avg}}$	5	6	5	-
$^{\dagger}SSLL$	6	4	-	-
$m_{\text{bb}}^{\text{min }\Delta\text{R}}$	7	12	4	4
$m_{\text{bj}}^{\text{max p}_{\text{T}}}$	8	8	-	-
$\Delta R_{\text{bb}}^{\text{max p}_{\text{T}}}$	9	-	-	-
$\Delta R_{\text{lep}-\text{bb}}^{\text{min }\Delta\text{R}}$	10	11	10	-
$m_{\text{uu}}^{\text{min }\Delta\text{R}}$	11	9	-	2
$\text{Aplanarity}_{\text{b}-\text{jet}}$	12	-	8	-
N_{40}^{jet}	-	1	3	-
$m_{\text{bj}}^{\text{min }\Delta\text{R}}$	-	5	-	-
$m_{\text{jj}}^{\text{max p}_{\text{T}}}$	-	-	6	-
$H_{\text{T}}^{\text{had}}$	-	-	7	-
$m_{\text{jj}}^{\text{min }\Delta\text{R}}$	-	-	9	-
$m_{\text{bb}}^{\text{max p}_{\text{T}}}$	-	-	-	1
$p_{T,\text{uu}}^{\text{min }\Delta\text{R}}$	-	-	-	3
$m_{\text{bb}}^{\text{max m}}$	-	-	-	5
$\Delta R_{\text{uu}}^{\text{min }\Delta\text{R}}$	-	-	-	6
m_{jjj}	-	-	-	7

Table 9.3: The lists and rankings of the variables in each of the analysis regions. Variables designated with a † utilize information from the Matrix Element Method. [2]

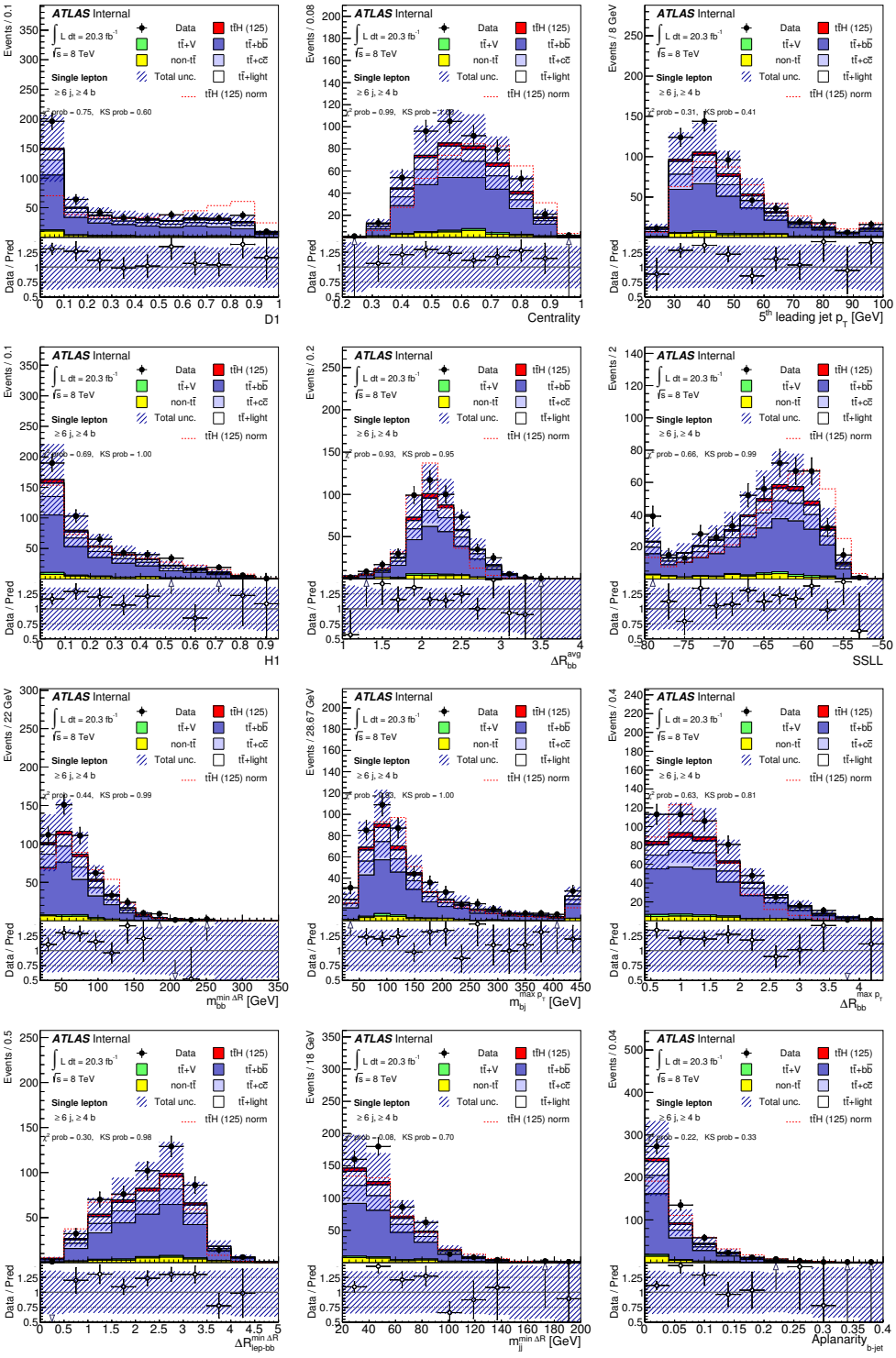


Figure 9.2: Pre-fit, data to Monte-Carlo agreement of the 12 input variables for the ($\geq$ 6j, $\geq$ 4b) region. The variables are ordered in terms of importance from left to right, then top to bottom. Here, D1 is equivalent to EPD_{ME}. [2]

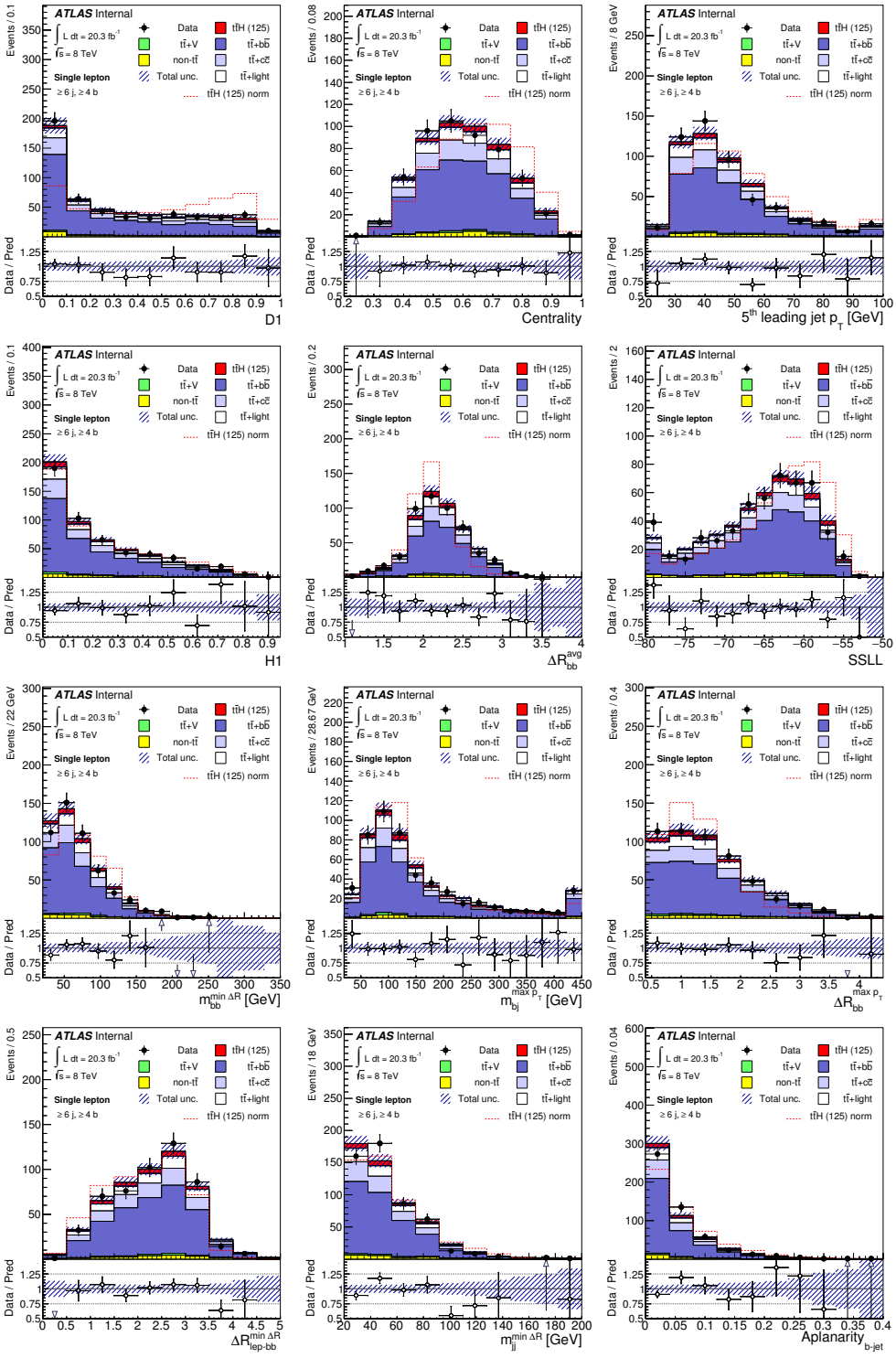


Figure 9.3: Post-fit, data to Monte-Carlo agreement of the 12 input variables for the ($\geq$ 6j, $\geq$ 4b) region. The variables are ordered in terms of importance from left to right, then top to bottom. Here, D1 is equivalent to EPD_{ME}. [2]

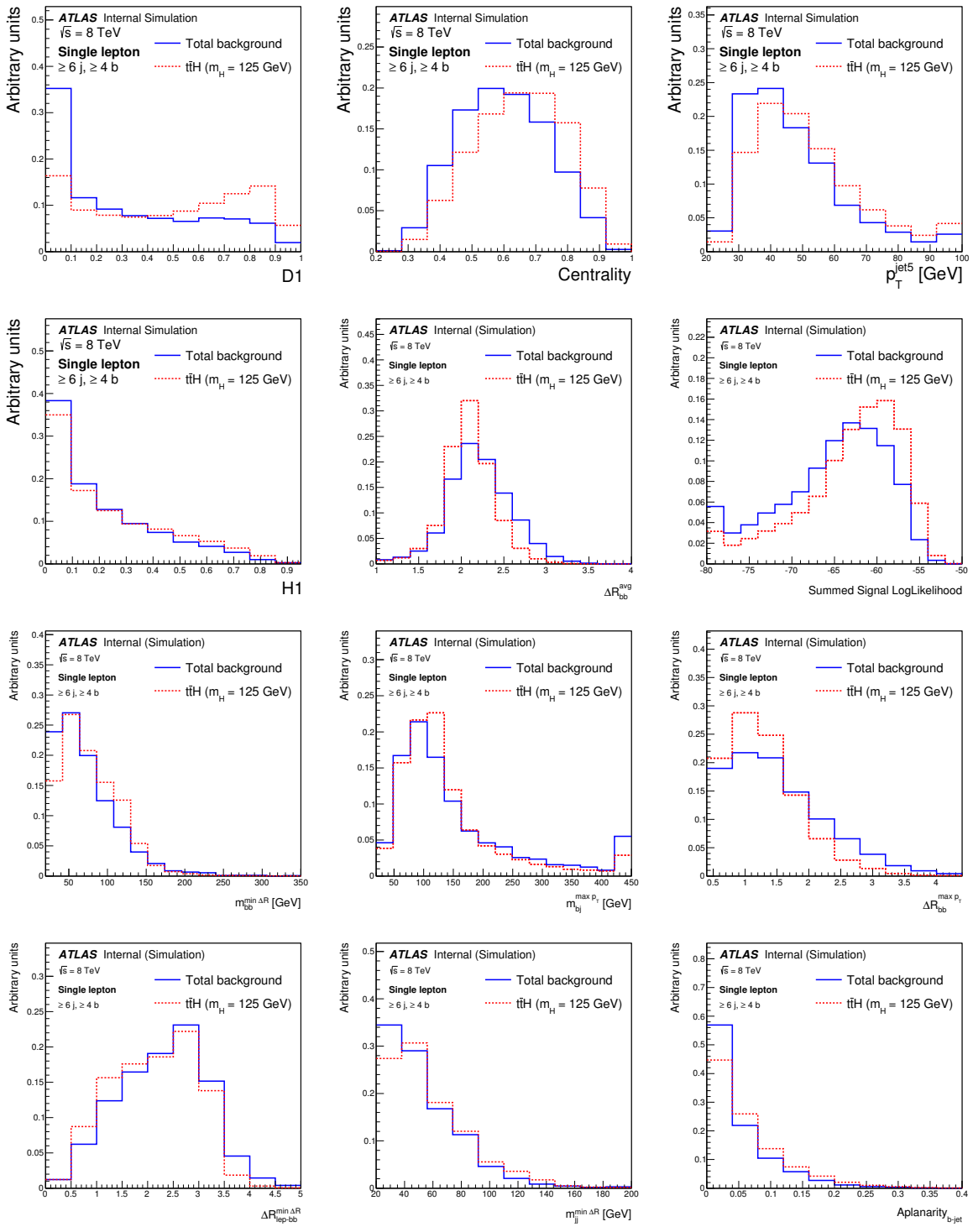


Figure 9.4: Separation power of the 12 input variables for the ($\geq$ 6j, $\geq$ 4b) region. The variables are ordered in terms of importance from left to right, then top to bottom. Here, D1 is equivalent to EPD_{ME}. [2]

Separation Power With and Without MEM Variables

MVA Configuration	$(\geq 6\text{j}, 3\text{b})$	$(\geq 6\text{j}, \geq 4\text{b})$
Kinematic NN - Without MEM	26.3%	29.7%
Kinematic NN - With MEM	27.8%	32.7%
Relative Improvement	5.7%	10.1%

Table 9.4: Neural Network separation power improvement due to the inclusion of two additional MEM variables, EPD_{ME} and $SSLL$.

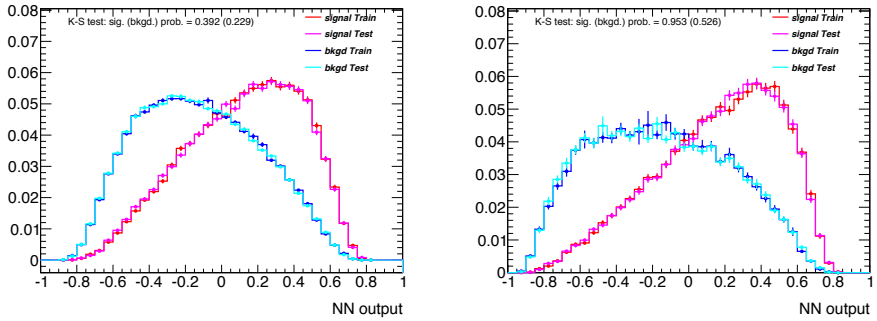


Figure 9.5: Overtraining tests for the $(\geq 6\text{j}, 3\text{b})$ (left) and $(\geq 6\text{j}, \geq 4\text{b})$ (right) regions. Here red and blue curves show the NN response for the signal and background samples used for training, where the violet and cyan curves show the signal and background samples used for testing. [2]

9.1.2 NN Performance

The resulting performance in the two signal regions, $(\geq 6\text{j}, 3\text{b})$ and $(\geq 6\text{j}, \geq 4\text{b})$, represents a large improvement over any one individual input variable, which is expected. As shown in Table 9.4, the addition of the MEM variables further improves these regions' separation power by approximately 6% and 10% respectively. The propensity of machine learning algorithms to over-fit the training data and therefore overestimate their performance is known as *overtraining*. Figure 9.5 shows the overtraining tests for these two regions where the NN output for both the training samples and test samples are overlaid showing a consistent response between the two. Here we see a consistent response between inputs which were seen during training and unseen test inputs. Finally, Figure 9.6 shows the separation power of the final NN output with and without the MEM information.

9.2 Matrix Element Based Boosted Decision Tree

During the development of the MEM it became evident that an additional MVA could be used to improve its separating power. This new MVA would addition-

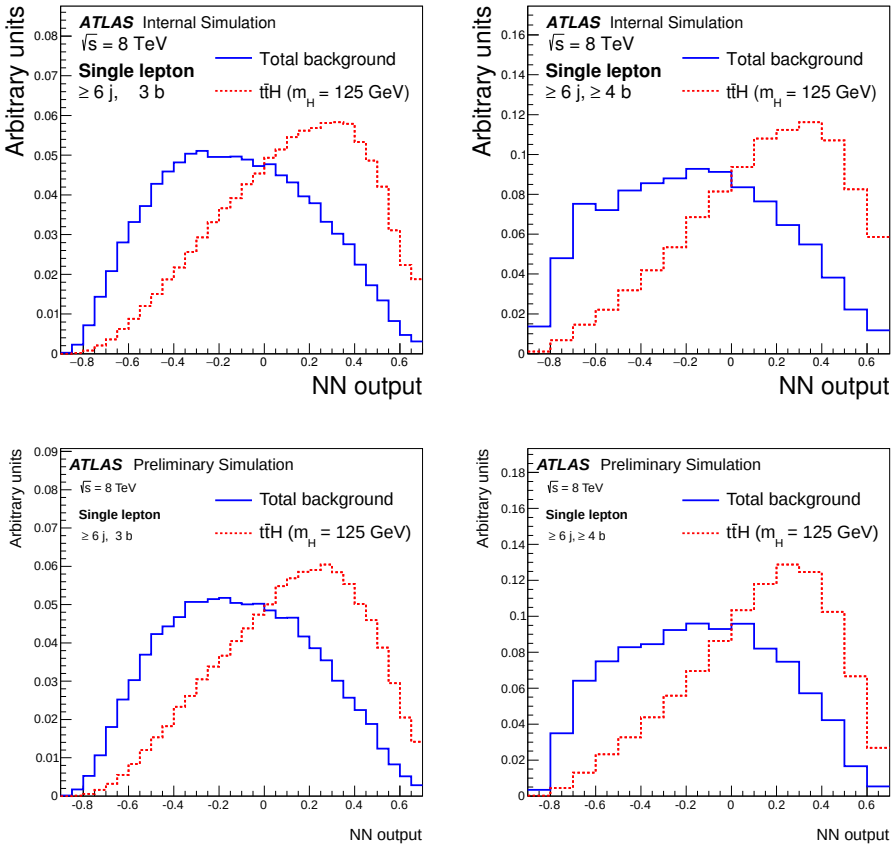


Figure 9.6: Separation plots for the $(\geq 6\text{j}, 3\text{b})$ (left) and $(\geq 6\text{j}, \geq 4\text{b})$ (right) regions with, and without, MEM information. The plots on the top row show the NN response without the inclusion of the 2 MEM variables, while the bottom row shows the NN response with the MEM variables. [2] Separation values can be seen in Table 9.4.

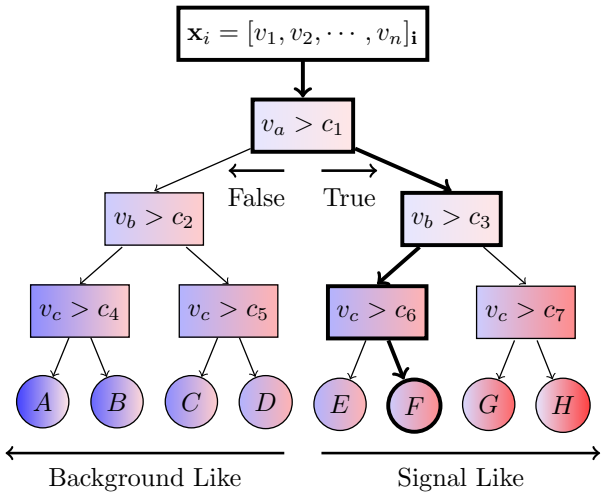


Figure 9.7: A diagram of a seven node (rectangular nodes) decision-tree with three decision-layers and eight output leaves (circular nodes). Each decision layer routes events, $\mathbf{x_i}$ into some leaf node, Leaf $= \{A, B, \cdots, H\}$, based on some event input variable, v_j, and an associated cut value, c_k, at each binary-decision node. A hypothetical decision path is shown starting at the root node, and finally being collected in the F leaf node. Signal or background classification is based on the majority classification of events in each leaf node.

ally utilize kinematic variables that could be obtained from MEM based jet assignments. Thus, a Boosted Decision Tree was developed to implement this technique.

The choice to utilize a Boosted Decision Tree (BDT) technique, in particular, is predicated on the fact that the maximum-likelihood jet assignments for both the $t\bar{t}H$ and $t\bar{t}+b\bar{b}$ MEM calculations are often identical. This, in turn, results in identical 'signal max' and 'background max' kinematic variables driving strong linear correlations between them. The robustness of BDTs towards correlated inputs make it the natural choice. Finally, it is the opinion of the author that BDTs are generally less complex than other MVA techniques and thus provide an opportunity to decrease the overall complexity of an already exceedingly complex analysis while maintaining performance.

A decision tree, illustrated in Figure 9.7, classifies events by categorizing each event by some pre-defined criteria. This criteria would define the cuts, c_k, in the example above. Based on this categorization procedure, each terminal leaf node can be labeled as a signal region or a background region based on the majority of signal or background training events it collected. This process can be repeated for a different set of cuts until the classification purity in each leaf reaches some maximum. In practice however, this method results in significant overtraining pathologies as the tree-depth grows. Further, single decision trees aren't as accurate as other more robust methods.

To improve the performance and robustness of decision tree classifiers, the algorithm is extended by generating an ensemble of decision trees. This en-

semble is generated such that each tree's selection criteria is optimized using a randomly chosen subset of the input variables and is therefore independent of the other trees, allowing for a far more generalized collection of classification leaves. Thus, events classified by a such an ensemble are categorized by the average of leaf categories of all the trees in which they are collected. This adaptation on the simple decision tree algorithm vastly improves its susceptibility to overtraining, at the cost of computational complexity and some obfuscation of the inner workings of the classifier.

Boosted decision tree techniques are based on the general concept of *boosting* in which an ensemble of trees are generated as a part of a deeper learning procedure. There are multiple boosting algorithms that take a number of approaches, of which the AdaBoost [96] method was chosen. The AdaBoost method assigns weights to each of the input training events. Events that are mis-classified by previous trees have their weights increased, while correctly classified events have their weights diminished. Thus, subsequent trees are generated that are better at classifying the highly weighted 'hard' events. These more specialized trees are subsequently added to the existing ensemble, and the next iteration begins.

This BDT is trained in a similar fashion as the NNs described above, however, unlike the NNs, the $t\bar{t}H_{125}$ signal sample is trained against both the $t\bar{t}+b\bar{b}$ and the $t\bar{t}+c\bar{c}$ background samples from the the $(\geq 6\text{j}, \geq 4\text{b})$ region only. All training samples are simulated using the ATLAS Fast simulation utilizing approximated calorimeter response, whereas the test samples utilize the full ATLAS simulation. Additional information regarding the specific implementation of this BDT can be found in Appendix C

9.2.1 BDT Variable Selection

The selection of the variables for the BDT follows largely the same process of variable ranking and exclusion that was introduced in the prior discussion on the Neural Networks. However, due to the fact that MEM information can be leveraged to identify full jet-assignments for both $t\bar{t}H$ and $t\bar{t}+b\bar{b}$ processes, the number of possible input variables increases significantly. Thus, an intuitive mechanism is employed to decrease the input variable space down to a manageable size. This mechanism relies on the following heuristic: kinematic variables are prioritized based on separation power and 'ensemble diversity.' Here, the ensemble diversity is a qualitative description which prioritizes new variables that do not depend on objects already used by current variables in the ensemble. This heuristic leads to the identification of 16 variables from which non-performant variables are excluded leading to six final variables. These variables are listed in Table 9.5, and described in further detail below.

The first two variables come directly from the MEM calculations themselves. Described in Section 8.2, these are the averaged Neyman-Pearson ratio, R_{ME}^{Avg}, of the jet-assignment likelihoods, and the sum of the logged signal likelihoods,

Input Variables used for the Four Neural Networks

Variable	Definition
R_{ME}^{Avg}	Newman-Pearson ratio MEM discriminant
$SSLL$	Summed MEM signal log-likelihood
	Max L($t\bar{t}+b\bar{b}$)
M_{bb}	Invariant mass of the bb object
$\Delta\eta(bb, t_{\mathrm{Far}})$	Largest $\Delta\eta$ between the X_{bb} object and the reconstructed top quarks
	Max L($t\bar{t}H$)
$\Delta\eta(q, q')$	Delta η between the two light quarks assigned to the hadronically decaying W boson
$\lvert d\eta(bb, b^t b^t)\rvert$	Quadrature sum of the $\Delta\eta$ between the two top-associated b-tagged jets, and the $\Delta\eta$ between the two remaining b-tagged jets

Table 9.5: The variables considered for use within the ME-BDT analysis.

$SSLL$. These two variables serve two different purposes. First, R_{ME}^{Avg} is constructed specifically to discriminate between $t\bar{t}H$ and $t\bar{t}+b\bar{b}$ events. Whereas, $SSLL$, is used to discriminate against events which are clearly not $t\bar{t}H$-like. Here, $SSLL$ is used in the same way as in the NN analysis. The mono-modal structure of these distributions further make them amenable to the common variable transformations used in conjunction with most MVA techniques. Due to the common information between them, they do experience a moderate linear correlation of approximately 25%, which is the most significant correlation among the six input variables. These variables are plotted in Figure 9.9, showing both the separation power of the variables against the total summed backgrounds, as well as against the individual backgrounds themselves.

The second set of two inputs are derived from kinematic variables based on the jet-assignment with the largest $t\bar{t}+b\bar{b}$ matrix element likelihood. The first input is the invariant mass of the two b-tagged jets not assigned to the top quarks. In principle, the jet-assignment with the highest $t\bar{t}+b\bar{b}$ likelihood will be assigned such that the b-tagged jets remain consistent with the expected kinematics of the top quarks and the assumed gluon in the $t\bar{t}+b\bar{b}$ process. This requirement shapes this invariant mass to low values with a maximum near 50 GeV. Due to this behavior, in circumstances where the two b-tagged jets originating from the top quarks could be clearly matched, signal events with a H_{bb} mass in the vicinity of 125 GeV resist this shaping effect and subsequently form a bump consistent with a 125 GeV Higgs boson, resulting is significant separation.

The second maximum $t\bar{t}+b\bar{b}$ likelihood input variable contains kinematic information for the angular relationships between the reconstructed top quarks

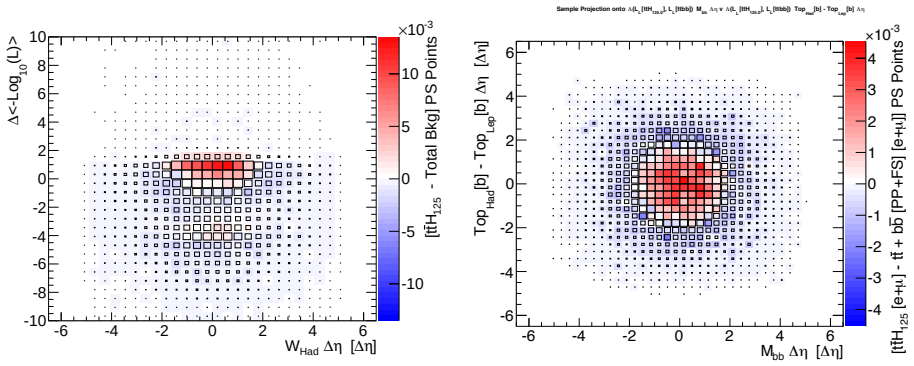
and the hypothesized gluon object which decays to the two additional b-quarks. Here, the $\Delta\eta$ between each top quark and the reconstructed gluon is determined, and the largest of the these two numbers is used, called $\Delta\eta_{\text{Far}}$. Distributions of this variable show considerable shape differences between $t\bar{t}H$ events which show largely consistent values of η among the parent objects, whereas background events show a multi-modal distribution with peaks at $|\Delta\eta_{\text{Far}}| = 2$. Both input variables under the maximum $t\bar{t} + b\bar{b}$ likelihood assignments used are presented in Figure 9.11, again showing the total background discrimination as well as individual background sample comparisons.

The final set of two inputs are defined by kinematic variables based on the jet-assignment with the largest $t\bar{t}H$ matrix element likelihood. The first of these is the $\Delta\eta$ between the two light-flavor jets assigned to the hadronically decaying W boson. The importance of this variable is somewhat hard to identify given the little apparent separation it provides. However, when compared in two dimensions with the accompanying likelihoods, shown in Figure 9.8(a), non-linear correlations appear which make the two variables together a potent combination. This comes from the fact that this assignment doesn't change throughout the 12 permutations in the $(\geq 6\text{j}, \geq 4\text{b})$ analysis, and changes slightly in the $(\geq 6\text{j}, 3\text{b})$ analysis. Thus, the quality of the light-flavor jet-assignments effectively set the overall likelihood scale for the maximally contributing assignments.

The final input is similarly based on the jet-assignemnt with the largest $t\bar{t}H$ matrix element likelihood. It is a complex parameter based on the η distributions of all the b-tagged jets. These parameters exhibit a correlation that emphasizes signal events with the top quarks and Higgs evolving through small $\Delta\eta$ windows more so than the background. Furthermore, this behavior exhibits a clustering effect wherein b-tagged jets associated with the Higgs boson prefer tighter $\Delta\eta$ separations when the b-tagged jets associated with the top quarks are measured with larger $\Delta\eta$ separations. This quasi-radial distinction between two variables was transformed into single input via a quadrature sum. The resulting variable is named, $|d\eta(bb, b^t b^t)|$. This behavior is shown in Figure 9.8(b). Finally, separation and distribution shapes of the signal and background sample for these variables are shown in Figure 9.13.

9.2.2 BDT Performance

The final results from the six variable BDT are shown in Figure 9.20, while the final separation values can be seen in Table 9.6. Though not as impressive a difference as the more complex NN analysis, the BDT out performed the NNs trained without the MEM information in the $(\geq 6\text{j}, \geq 4\text{b})$ region, but underperformed in the $(\geq 6\text{j}, 3\text{b})$ region. This is not surprising as the BDT is only trained on data from the $(\geq 6\text{j}, \geq 4\text{b})$ region to reduce complexity. It is entirely possible that independently training a BDT for $(\geq 6\text{j}, 3\text{b})$, with different variables, will result in a considerable improvement given the results



(a) $\Delta\eta(q, q')$ (b) $\Delta\eta(b^H, b^H)$ vs. $\Delta\eta(b^{t_L}, b^{t_H})$

Figure 9.8: Two dimensional discrimination plots for $\Delta\eta(q, q')$, (a), and $\Delta\eta(b^H, b^H)$ vs. $\Delta\eta(b^{t_L}, b^{t_H})$, (b), using the maximum $t\bar{t}H$ likelihood jet assignment, showing potential selection criteria for an MVA technique. Red bins denote higher signal purity and blue bins denote higher background purity. The black squares are proportional in area to the number of events in the bins.

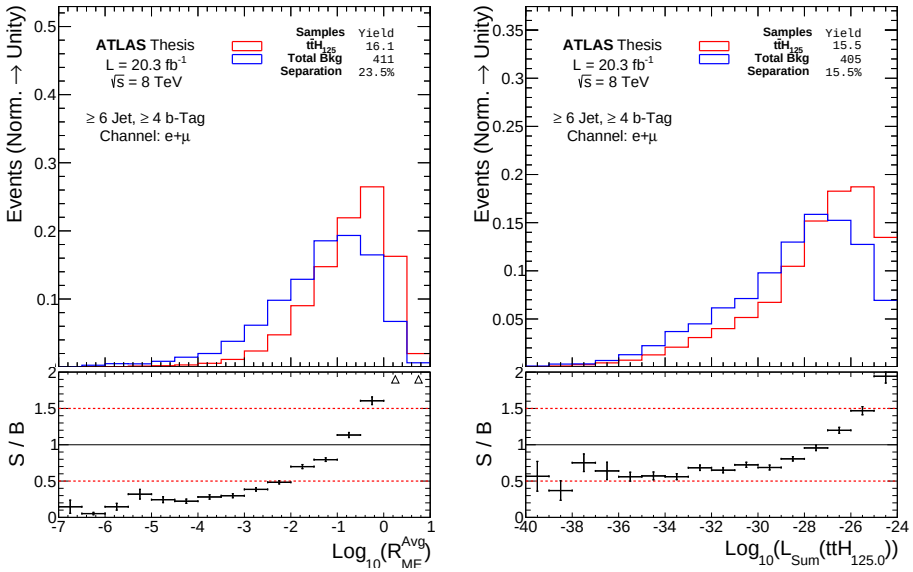
Separation Power Comparison of the Final BDT Discriminant

MVA Configuration	$(\geq 6\text{j}, 3\text{b})$	$(\geq 6\text{j}, \geq 4\text{b})$
Kinematic NN - Without MEM	26.3%	29.7%
Kinematic NN - With MEM	27.8%	32.7%
BDT - With MEM	**21.1%**	**30.5%**

Table 9.6: Final BDT separation power.

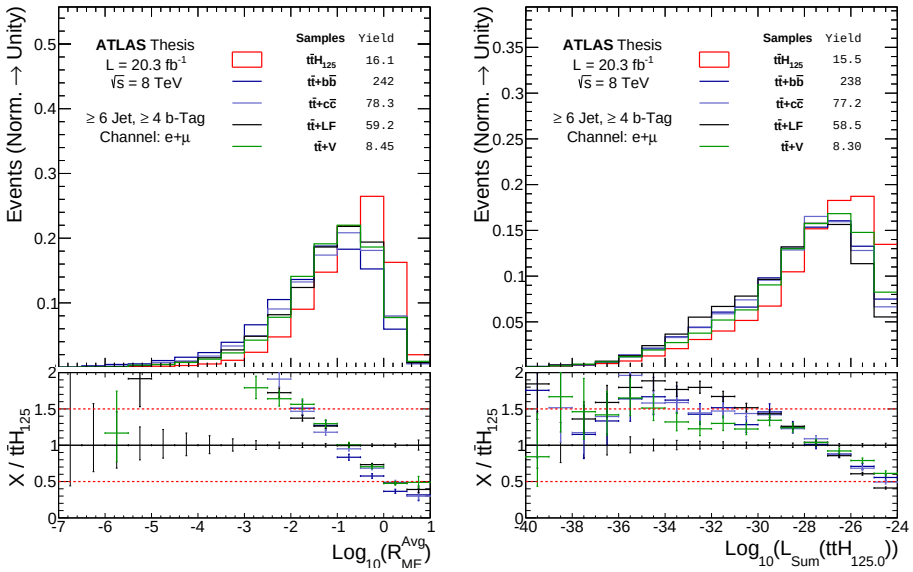
in $(\geq 6\text{j}, \geq 4\text{b})$.

Validation plots for the training and testing outputs are shown in Figure 9.19. Pre-fit data to Monte-Carlo agreement is shown in Figure 9.15 for the $(\geq 6\text{j}, 3\text{b})$ region and in Figure 9.16 for the $(\geq 6\text{j}, \geq 4\text{b})$ region. The post-fit data to Monte-Carlo agreement is shown in Figure 9.15 for the $(\geq 6\text{j}, 3\text{b})$ region and in Figure 9.16 for the $(\geq 6\text{j}, \geq 4\text{b})$ region. Finally, linear correlations between the input variables are shown in Figure 9.21 for signal and for background.



(a) R_{ME}^{Avg} Separation

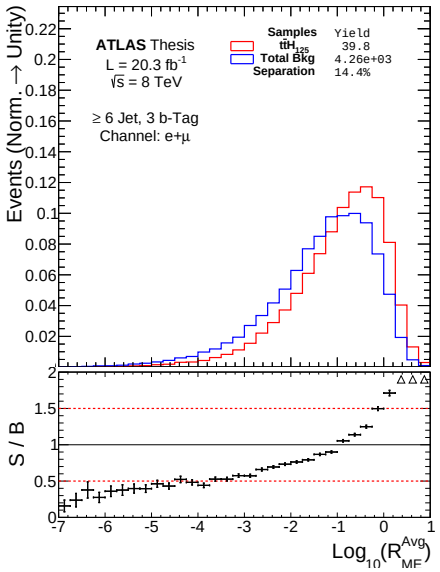
(b) $SSLL$ Separation



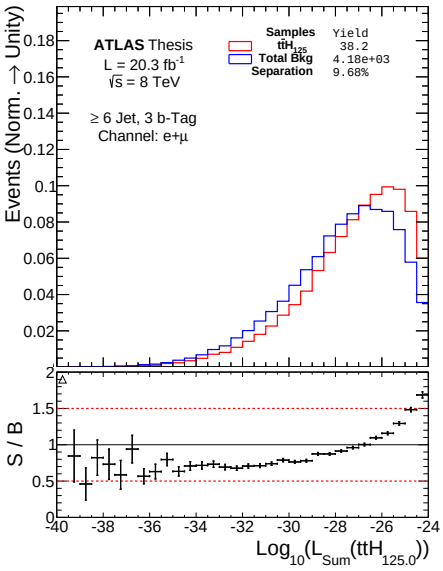
(c) R_{ME}^{Avg} Shapes

(d) $SSLL$ Shapes

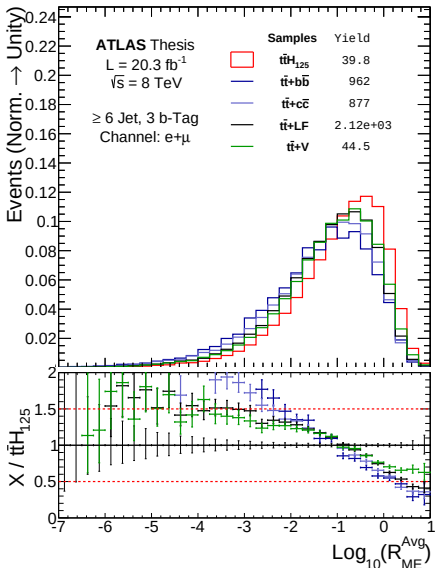
Figure 9.9: Expected BDT input variable separation for the two matrix element method likelihood-based inputs in the ($\geq$ 6j, $\geq$ 4b) region. These plots show the expected performance against the summed background, Figures (a) and (b), and against each background sample explicitly, Figures (c) and (d).



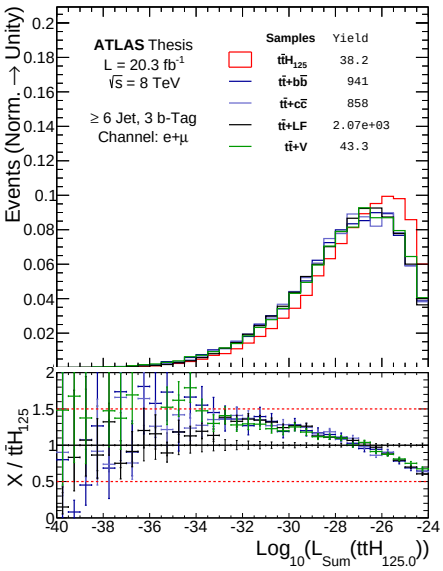
(a) R_{ME}^{Avg} Separation



(b) $SSLL$ Separation

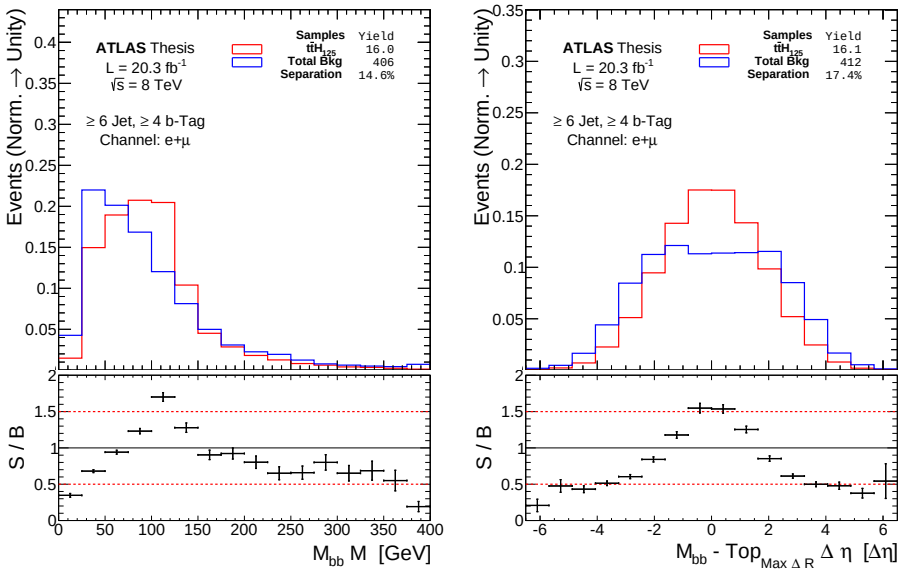


(c) R_{ME}^{Avg} Shapes



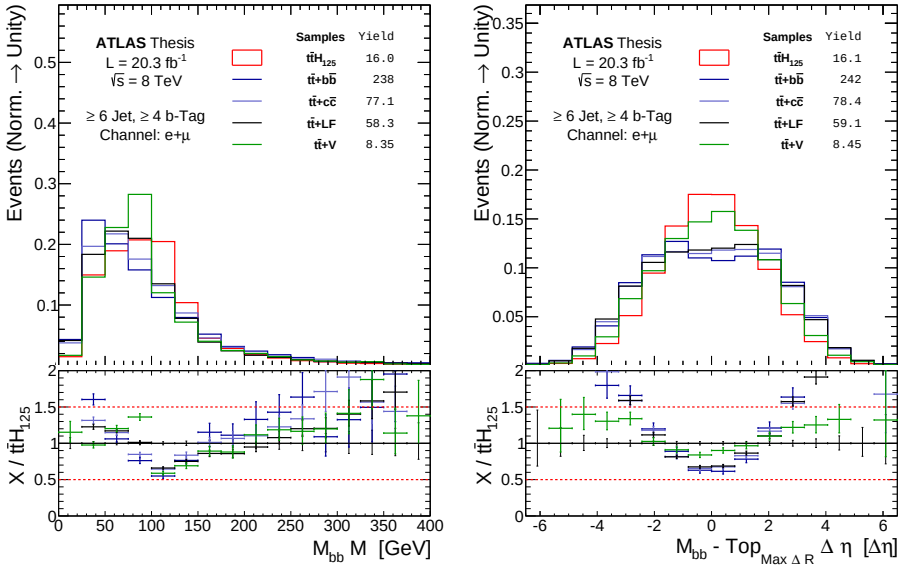
(d) $SSLL$ Shapes

Figure 9.10: Expected BDT input variable separation for the two matrix element method likelihood-based inputs in the ($\geq$ 6j, 3b) region. These plots show the expected performance against the summed background, Figures (a) and (b), and against each background sample explicitly, Figures (c) and (d).



(a) M_{bb} Separation

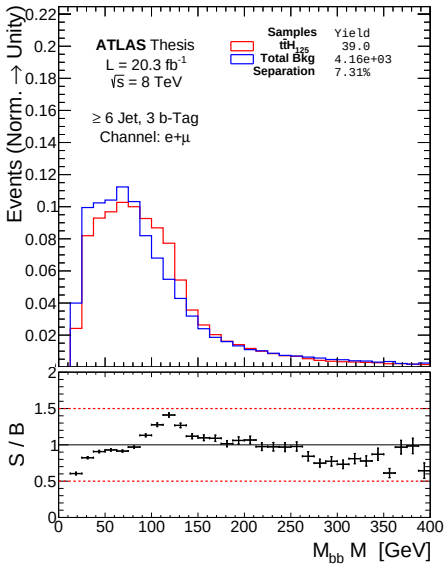
(b) $\Delta\eta(bb, t_{\mathrm{Far}})$ Separation



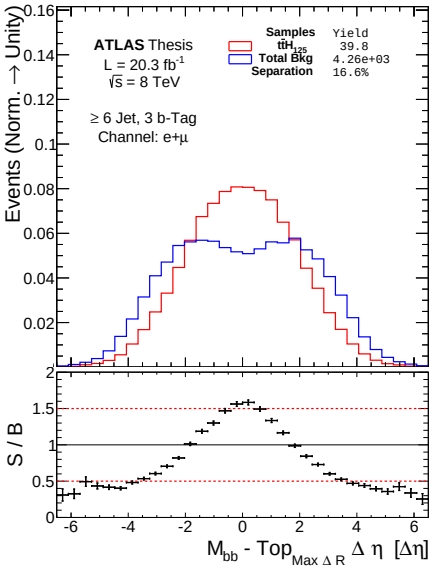
(c) M_{bb} Shapes

(d) $\Delta\eta(bb, t_{\mathrm{Far}})$ Shapes

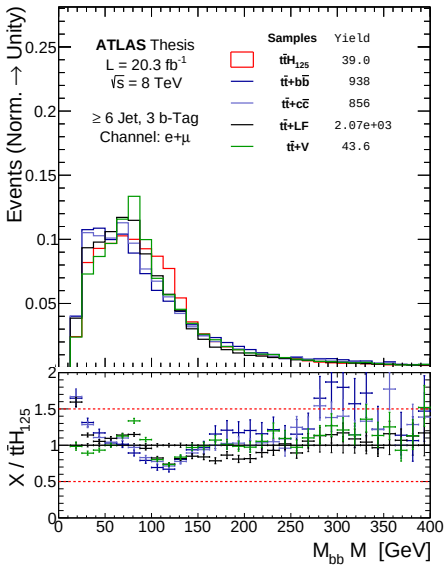
Figure 9.11: Expected BDT input variable separation for the two, maximum $t\bar{t} + b\bar{b}$ likelihood assignment kinematic variables in the ($\geq$ 6j, $\geq$ 4b) region. These plots show the expected performance against the summed background, Figures (a) and (b), and against each background sample explicitly, Figures (c) and (d).



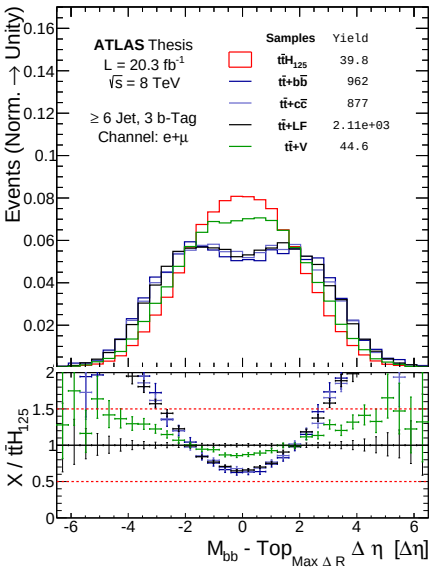
(a) M_{bb} Separation



(b) $\Delta\eta(bb, t_{\mathrm{Far}})$ Separation

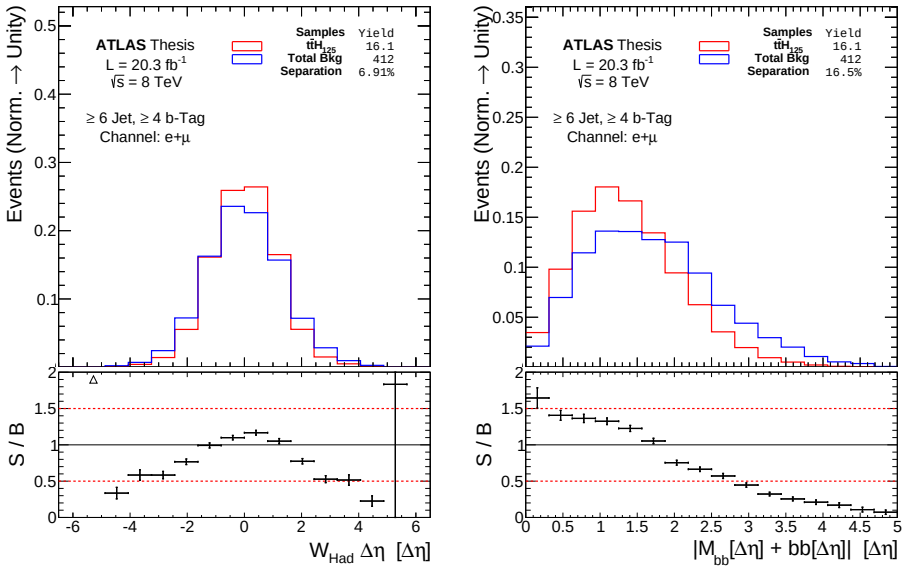


(c) M_{bb} Shapes



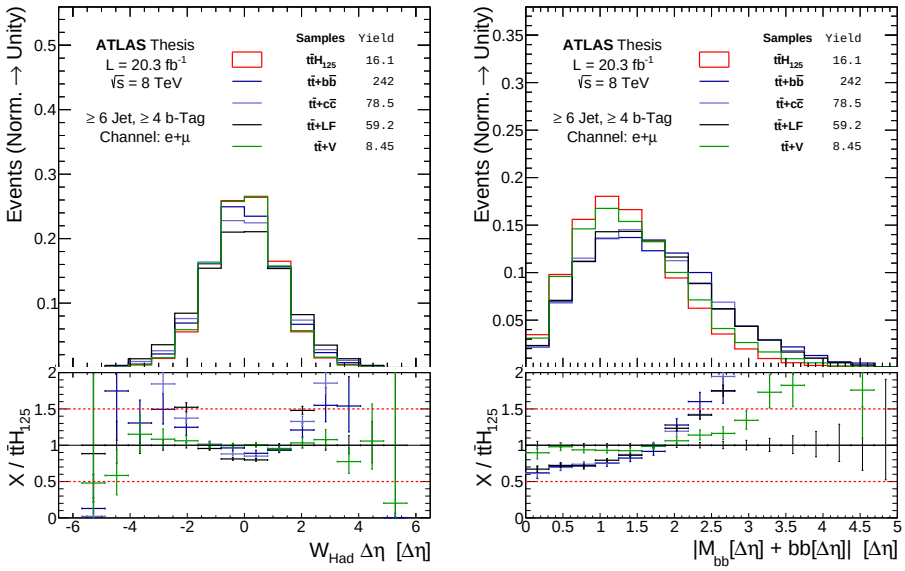
(d) $\Delta\eta(bb, t_{\mathrm{Far}})$ Shapes

Figure 9.12: Expected BDT input variable separation for the two, maximum $t\bar{t}+b\bar{b}$ likelihood assignment kinematic variables in the $(\geq 6\mathrm{j}, 3\mathrm{b})$ region. These plots show the expected performance against the summed background, Figures (a) and (b), and against each background sample explicitly, Figures (c) and (d).



(a) $\Delta\eta(q, q')$ Separation

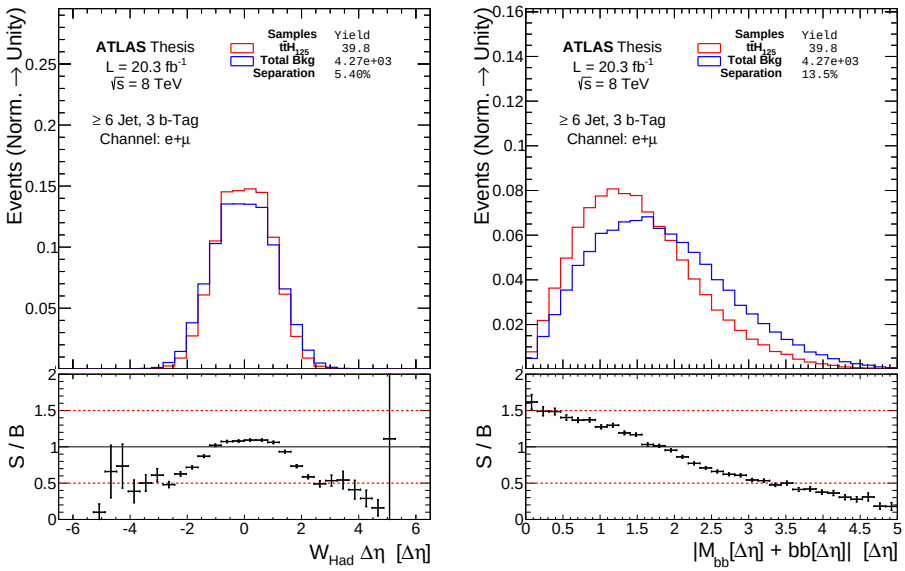
(b) $|d\eta(bb, b^t b^t)|$ Separation



(c) $\Delta\eta(q, q')$ Shapes

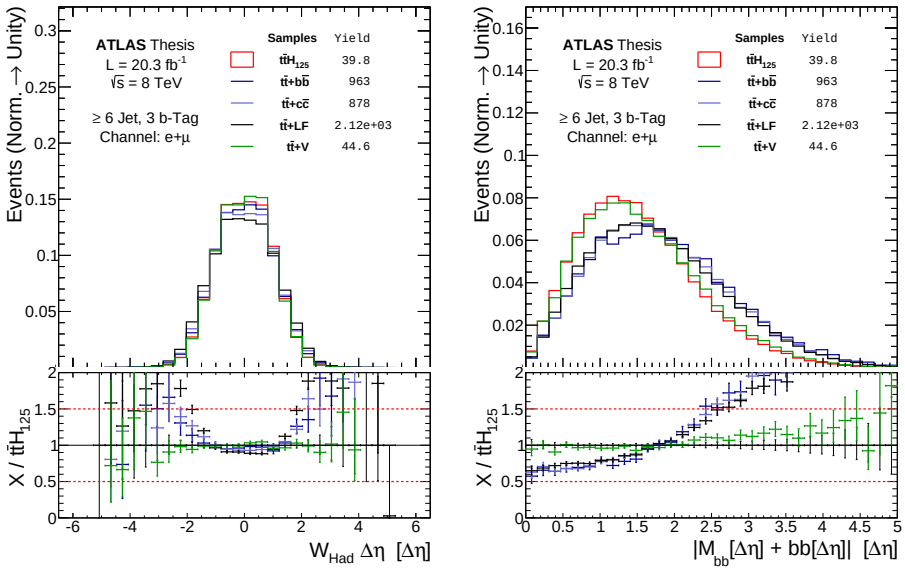
(d) $|d\eta(bb, b^t b^t)|$ Shapes

Figure 9.13: Expected BDT input variable separation for the two, maximum $t\bar{t}H$ likelihood assignment kinematic variables in the ($\geq$ 6j, $\geq$ 4b) region. These plots show the expected performance against the summed background, Figures (a) and (b), and against each background sample explicitly, Figures (c) and (d).



(a) $\Delta\eta(q, q')$ Separation

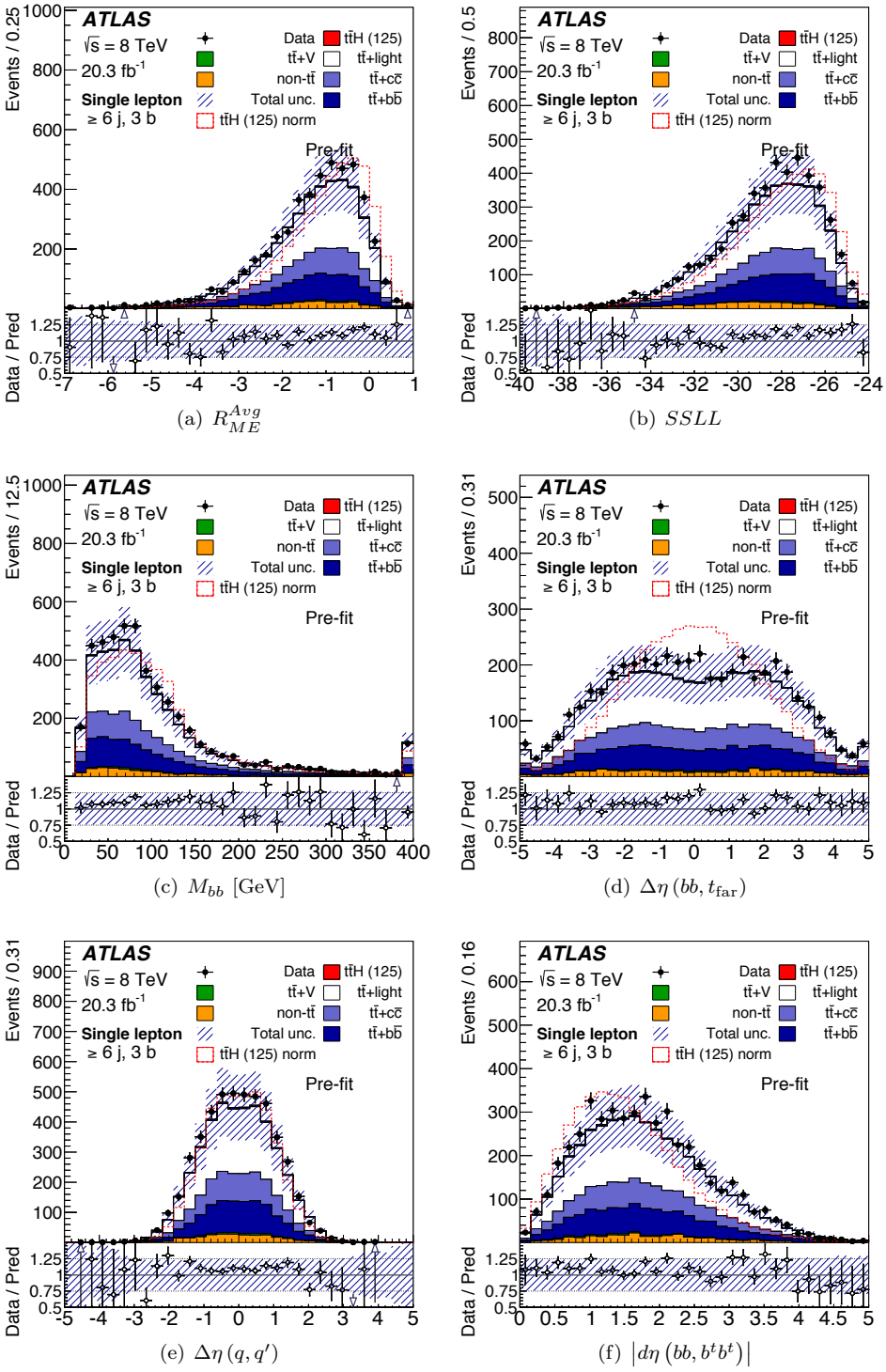
(b) $|d\eta(bb, b^t b^t)|$ Separation



(c) $\Delta\eta(q, q')$ Shapes

(d) $|d\eta(bb, b^t b^t)|$ Shapes

Figure 9.14: Expected BDT input variable separation for the two, maximum $t\bar{t}H$ likelihood assignment kinematic variables in the ($\geq$ 6j, 3b) region. These plots show the expected performance against the summed background, Figures (a) and (b), and against each background sample explicitly, Figures (c) and (d).

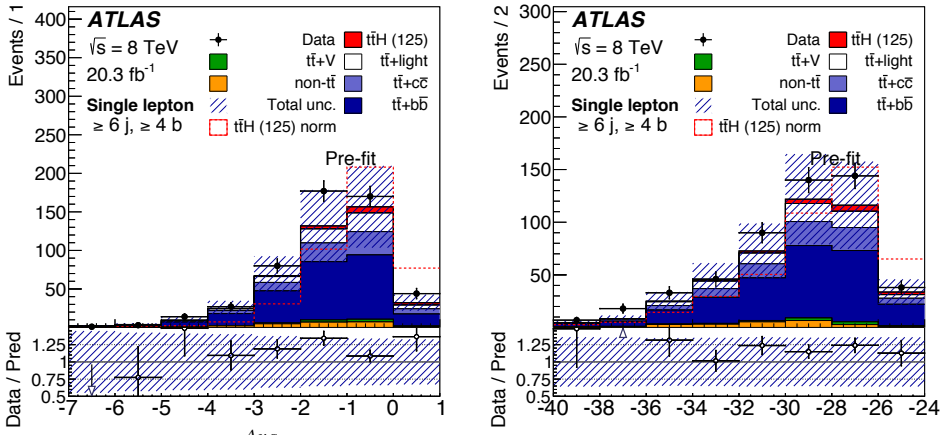


(a) R_{ME}^{Avg} (b) $SSLL$

(c) M_{bb} [GeV] (d) $\Delta\eta\left(bb, t_{\text{far}}\right)$

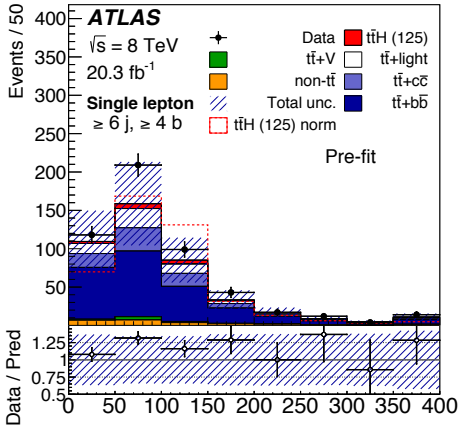
(e) $\Delta\eta\left(q, q'\right)$ (f) $\left|d\eta\left(bb, b^t b^t\right)\right|$

Figure 9.15: Pre-fit, data to Monte-Carlo agreement of the 6 BDT input variables for the ($\geq$ 6j, 3b) region. The variables are ordered by their order in Table 9.5.

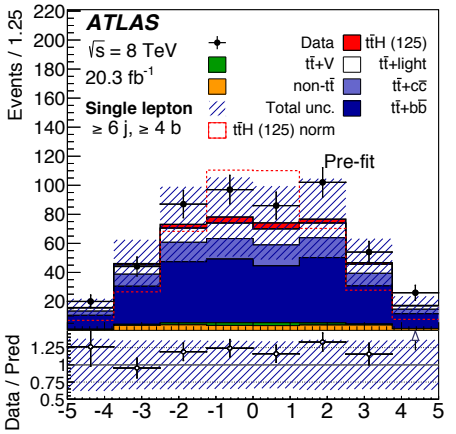


(a) R_{ME}^{Avg}

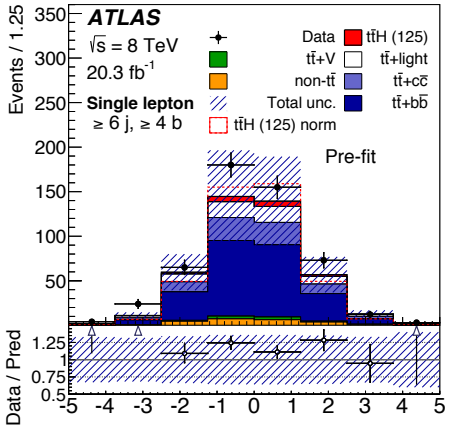
(b) $SSLL$



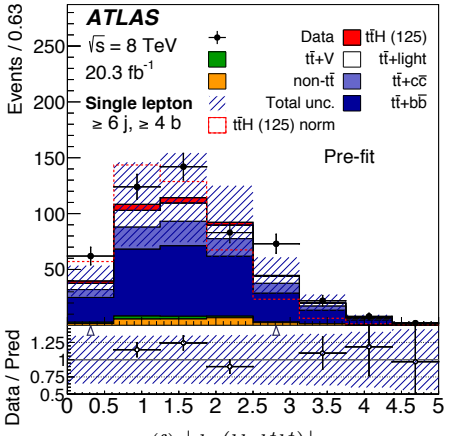
(c) M_{bb} [GeV]



(d) $\Delta\eta\,(bb, t_{\mathrm{far}})$

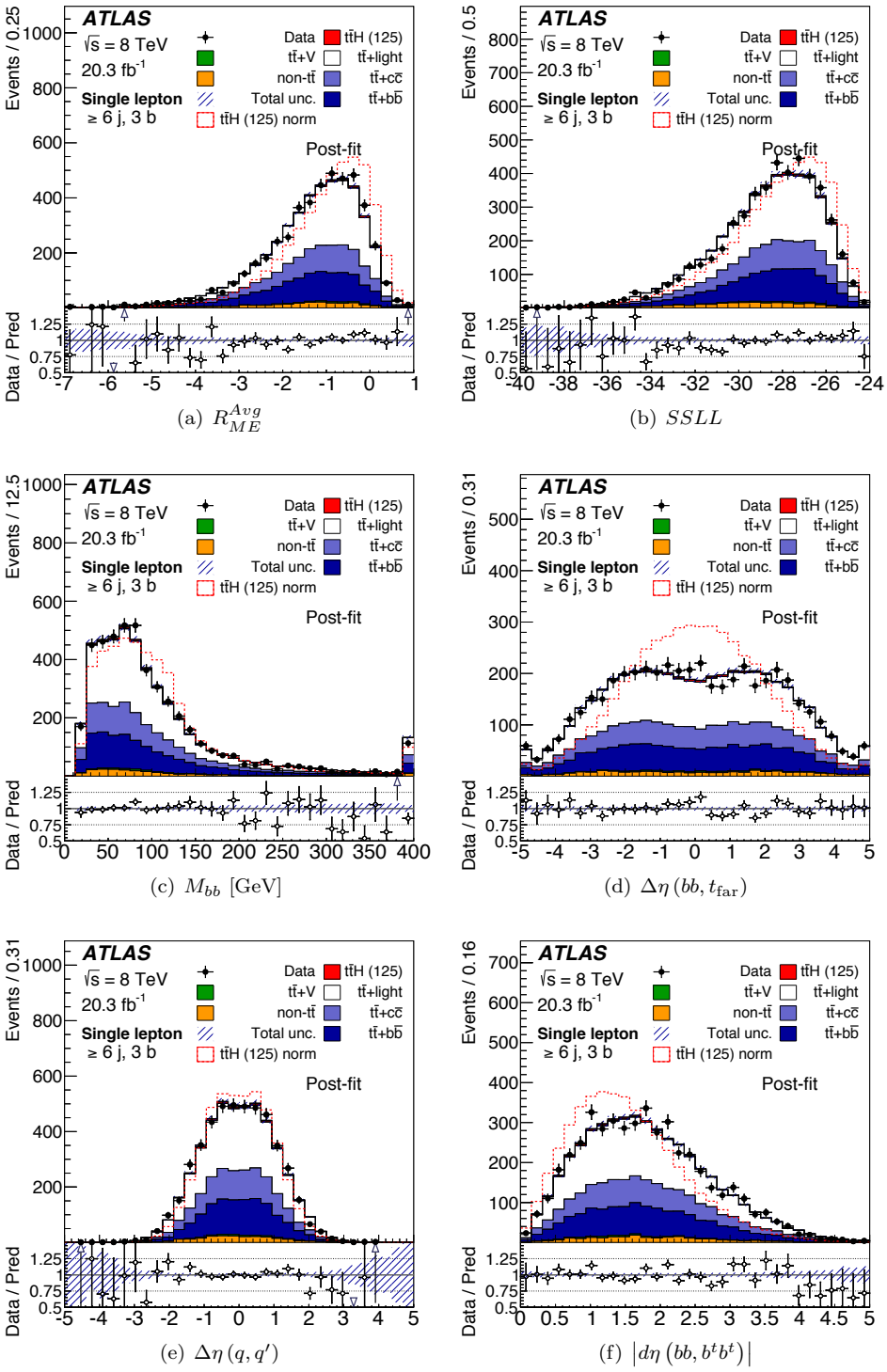


(e) $\Delta\eta\,(q, q')$



(f) $\left|d\eta\left(bb, b^t b^t\right)\right|$

Figure 9.16: Pre-fit, data to Monte-Carlo agreement of the 6 BDT input variables for the $(\geq 6\mathrm{j}, \geq 4\mathrm{b})$ region. The variables are ordered by their order in Table 9.5.

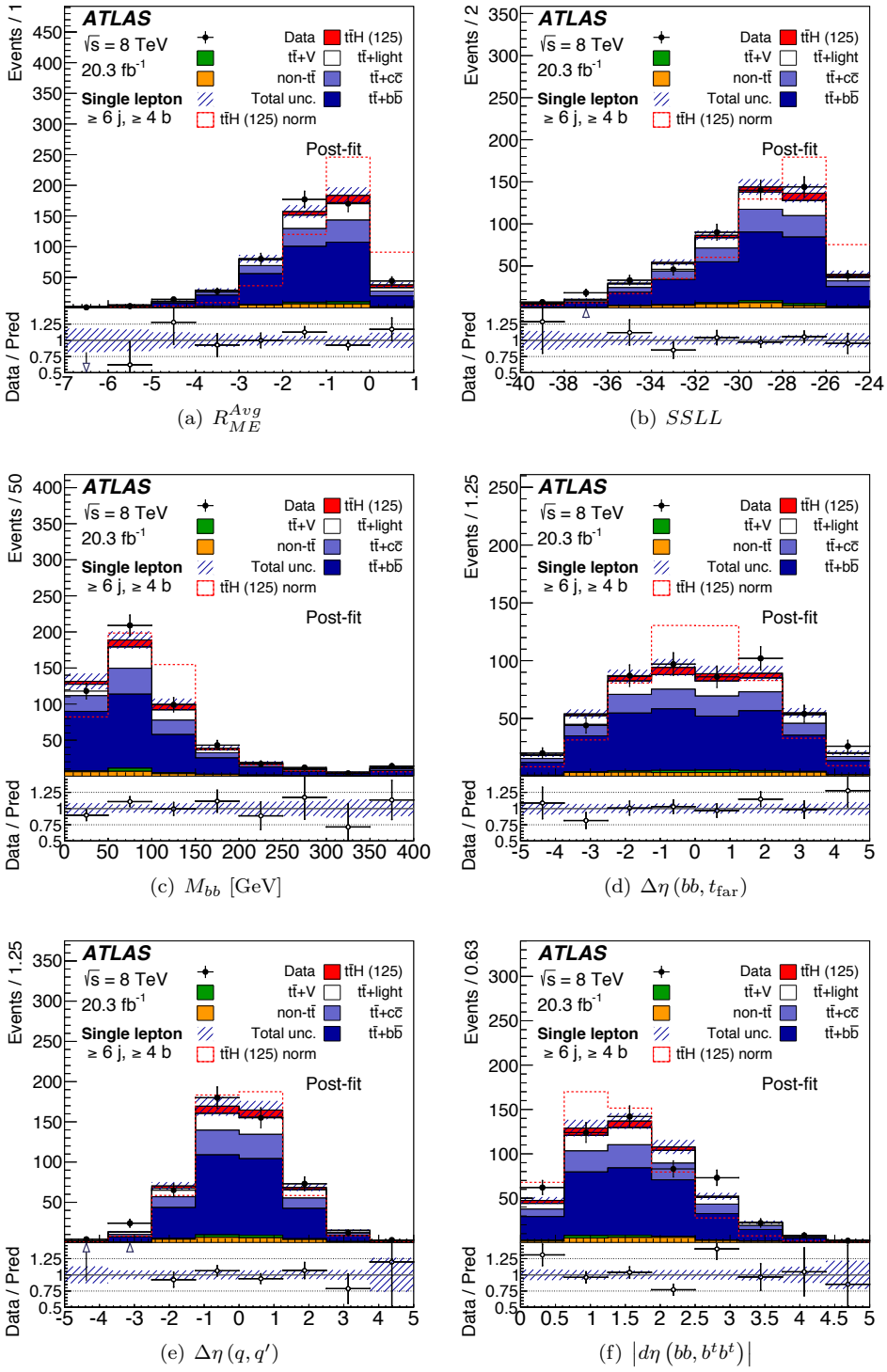


(a) R_{ME}^{Avg} (b) $SSLL$

(c) M_{bb} [GeV] (d) $\Delta\eta\left(bb, t_{\text{far}}\right)$

(e) $\Delta\eta\left(q, q'\right)$ (f) $\left|d\eta\left(bb, b^t b^t\right)\right|$

Figure 9.17: Post-fit, data to Monte-Carlo agreement of the 6 BDT input variables for the ($\geq$ 6j, 3b) region. The variables are ordered by their order in Table 9.5.



(a) R_{ME}^{Avg} (b) $SSLL$

(c) M_{bb} [GeV] (d) $\Delta\eta\left(bb, t_{\text{far}}\right)$

(e) $\Delta\eta\left(q, q'\right)$ (f) $\left|d\eta\left(bb, b^t b^t\right)\right|$

Figure 9.18: Post-fit, data to Monte-Carlo agreement of the 6 BDT input variables for the ($\geq$ 6j, $\geq$ 4b) region. The variables are ordered by their order in Table 9.5.

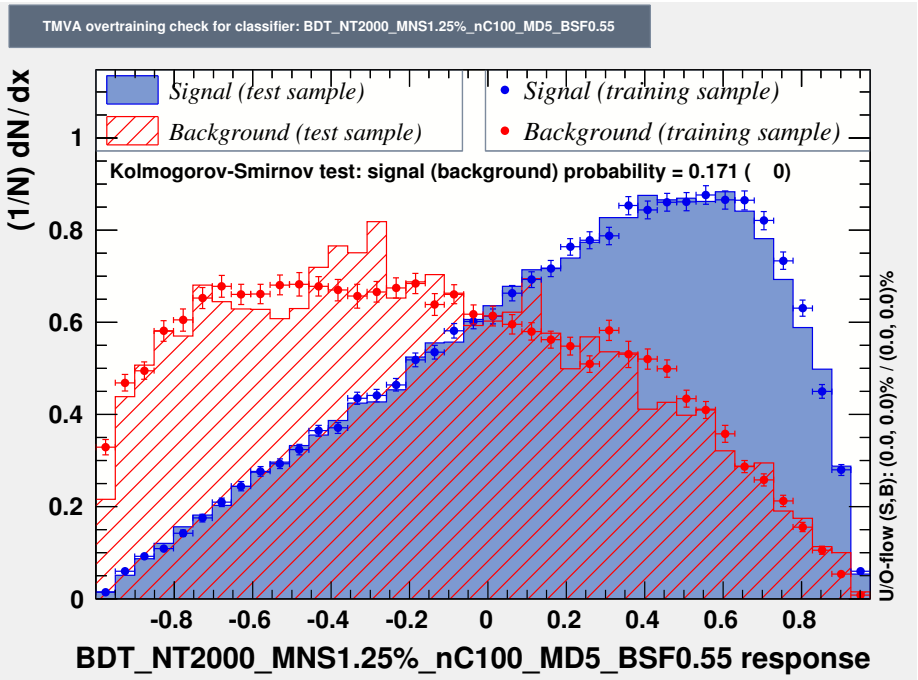


Figure 9.19: Overtraining tests for the BDT trained in the ($\geq$ 6j, $\geq$ 4b) region. Note the change in color convention, blue is signal, and red is background. This plot is generated using the TMVA toolkit. [97]

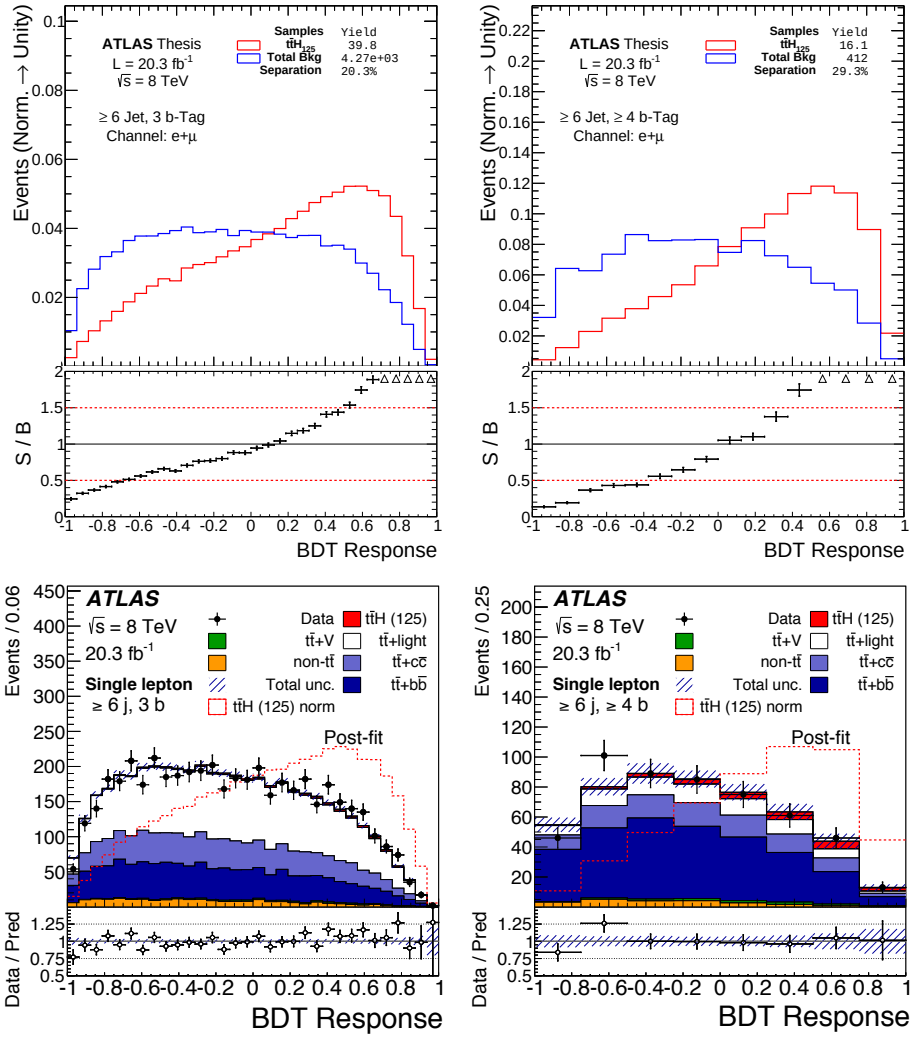


Figure 9.20: Separation and post-fit plots for the BDT discriminant in the $(\geq 6\text{j}, 3\text{b})$ and $(\geq 6\text{j}, \geq 4\text{b})$ regions Separation values can be seen in Table 9.6.

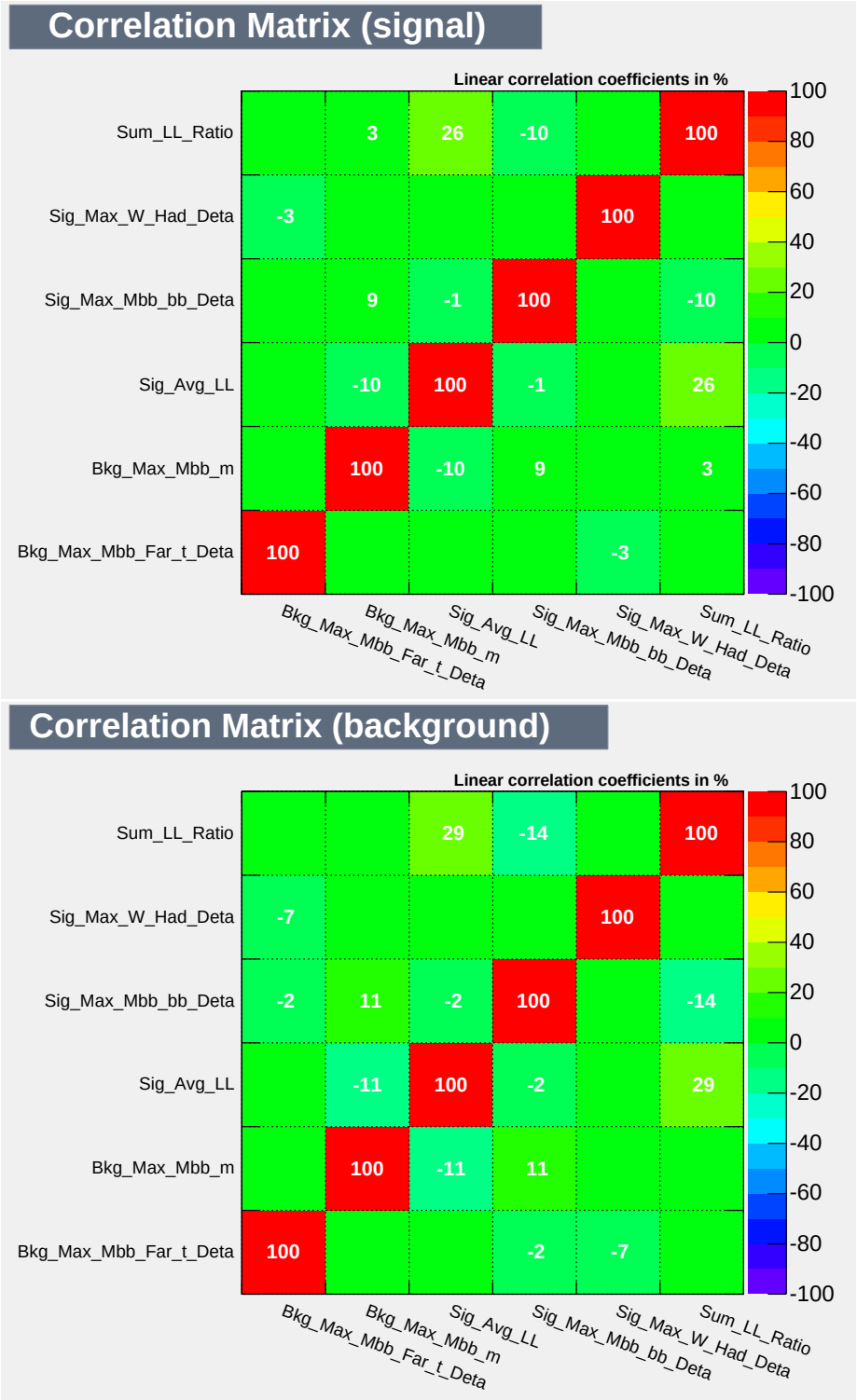


Figure 9.21: Linear correlations between BDT input variables with the signal and background input samples. These plots are generated with the TMVA toolkit, using the variable naming scheme used in the software. [97]

Chapter 10

Systematic Uncertainties

For this analysis, systematic uncertainties are considered which may influence the overall scale or the shape of the signal and background templates in the various analysis regions. This chapter proceeds to define the non-negligible sources of uncertainty, and quantify the expected effect they have on the final result. In many cases, these uncertainties are given conservative estimates which are then constrained by the available data. The process of constraining and quantifying the total uncertainty, given the observed data, is addressed in Chapter 11.

The systematic uncertainties considered in this analysis are appropriated in large part from the analysis performed in Reference [93]. This was done to profit from the extensive systematic studies performed for the referenced analysis which utilizes the exact same analysis model, jet and b-tag multiplicity regions, and $\sqrt{s}$=8 TeV data set. In addition, this systematics model provides a consistent apples-to-apples basis of comparison for quantifying the marginal benefit of the MEM. However, some improvements were made to the previous systematics model, and will be discussed in later sections in more detail, Section 10.2.4 in particular.

For this analysis, all evaluated sources of uncertainty are treated as independent and uncorrelated. Further, these sources are treated as correlated across jet and b-tag multiplicity regions. The independent sources of uncertainty are listed in Table 10.1 along with their influence as a normalization uncertainty, 'N,' or as a shape uncertainty, 'S,' or as both. In addition, the number of components related to a source of uncertainty is shown. In this analysis, these components are treated as independent sources of uncertainty.

10.1 Object Identification and Reconstruction

The evaluated uncertainties can be classified into two primary categories. The first category includes uncertainties related to the identification and reconstruction of the physics objects used by the analysis, defined in Chapter 6. These sources of uncertainty generally address differences in the resolution and scale effects between simulated and measured particle momenta in addition to uncertainties arising from the simulated reconstruction and identification efficiencies of the ATLAS detector.

Summary of Systematic Uncertainties

Systematic uncertainty	**Type**	**Components**
Luminosity	N	1
Physics Objects		
Electron	SN	5
Muon	SN	6
Jet energy scale	SN	22
Jet vertex fraction	SN	1
Jet energy resolution	SN	1
Jet reconstruction	SN	1
b-tagging efficiency	SN	6
c-tagging efficiency	SN	4
Light jet-tagging efficiency	SN	12
Background Model		
$t\bar{t}$ cross-section	N	1
$t\bar{t}$ modelling: p_T reweighting	SN	9
$t\bar{t}$ modelling: parton shower	SN	3
$t\bar{t}$+heavy-flavour: normalisation	N	2
$t\bar{t}$+$c\bar{c}$: HF reweighting	SN	2
$t\bar{t}$+$c\bar{c}$: generator	SN	4
$t\bar{t}$+$b\bar{b}$: NLO Shape	SN	8
W+jets normalization	N	3
W p_T reweighting	SN	1
Z+jets normalization	N	3
Z p_T reweighting	SN	1
Multijet normalisation	N	3
Single top cross-section	N	1
Single top model	SN	1
Diboson+jets normalization	N	3
$t\bar{t}V$ cross-section	N	1
$t\bar{t}V$ model	SN	1
Signal Model		
$t\bar{t}H$ scale	SN	2
$t\bar{t}H$ Generator	SN	1
$t\bar{t}H$ hadronization	SN	1
$t\bar{t}H$ PDF	SN	1

Table 10.1: List of systematic uncertainties considered and the number of orthogonal components attributed to each. Here an 'orthogonal component' is defined as a independent source of uncertainty related to the associated systematic classification. Uncertainties influencing the sample normalizations are denoted with an 'N' and uncertainties influencing the discriminant distribution's shapes are denoted with an 'S.'

Lepton Systematic Uncertainties

Uncertainty Component	**Relative Uncertainty** [%]
e Trigger	$\pm \leq 0.01$
e ID	± 1.0
e Reconstruction	$\pm \leq 0.01$
μ Trigger	± 0.95
μ ID	$\pm \leq 0.01$
μ Reconstruction	$\pm \leq 0.01$

Table 10.2: Lepton trigger, ID, and reconstruction efficiency uncertainties prior to the fit, in the $(\geq 6\text{j}, \geq 4\text{b})$ region over all samples.

Lepton Systematic Uncertainties Cont.

Uncertainty Component	**Relative Uncertainty** [%]
e Scale	± 0.01
e Resolution	$\pm \leq 0.01$
μ Scale	$\pm \leq 0.01$
μ MS Resolution	$\pm \leq 0.01$
μ ID Resolution	± 0.09

Table 10.3: Lepton resolution and scale uncertainties prior to the fit, in the $(\geq 6\text{j}, \geq 4\text{b})$ region over all samples.

10.1.1 Lepton Identification and Reconstruction

Systematic uncertainties related to the lepton object definition quantify the uncertainty with respect to the triggering, identification, and reconstruction efficiencies. These uncertainties also cover differences in these quantities between data and simulation. Scale factors used to correct the simulation for these efficiency differences are determined by tag-and-probe techniques using pure $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ samples. These uncertainties for the $(\geq 6\text{j}, \geq 4\text{b})$ region are shown in Table 10.3 prior to the fitting procedure.

In addition to these uncertainties, small discrepancies in the reconstructed momentum resolution and scale are accounted for with additional corrections defined by the ATLAS performance groups. These parameters are checked with measurements of the Z boson and J/ψ meson via their leptonic decay modes. For electrons in particular, $W \to e\nu$ events are used to study the E/p response. For muons, uncertainties associated with the muon spectrometer are kept separate from those calculated for the inner detector tracking. These uncertainties for the $(\geq 6\text{j}, \geq 4\text{b})$ region are shown in Table 10.3 prior to the fitting procedure.

10.1.2 Jet Reconstruction and Identification

Studies of the jet reconstruction efficiency show it to be approximately 0.2% lower for simulation than in data for jets with a p_T below 30 GeV, above this

threshold the agreement is consistent. Thus, to account for this mis-modeling, 0.2% of jets with a p_T below 30 GeV are randomly removed and all jet-related kinematic variables, including the $E_{\mathrm{T}}^{\mathrm{miss}}$, are recomputed. This is done prior to event selection. This sample is then treated as a systematic uncertainty to cover potential jet reconstruction modeling effects related to this mis-modeling.

Further, jets are removed from consideration that do not pass the JVF selection introduced in the object definitions in Section 6.2. The uncertainty introduced with the efficiency of this selection is first measured using $Z \to \ell^+\ell^- + 1$ jet events in simulation and in data, where $\ell = \{e, \mu\}$. These events are categorized into two samples, one for which the single jet originated from the hard process, and another for which the single jet originated from a different pp interaction in the same bunch crossing, also know as *pile-up*. Using this information, it was found to be sufficient for this analysis to simply vary the JVF cut up and down by 0.1, and to repeat the analysis with the varied cut value.

The uncertainty related to the jet energy scale (JES) is derived using a combination of test-beam data, LHC collision data, and simulation [71]. As a result of these analyses, the JES uncertainty is broken up into 22 orthogonal components that each have distinct η and p_T dependencies. Tools provided by the ATLAS collaboration for the express purpose of calculating these uncertainties are used for each of the 22 components.

The jet energy resolution (JER) uncertainties were similarly measured for data and simulation. Using *in-situ* techniques [71], the expected relative p_T resolution is determined as a function of a given jet's η and p_T. This lead to a systematic uncertainty defined as the magnitude of the difference between the calculated resolutions from data and simulation. To account for this, jets in simulation were smeared by this difference in resolution. The resulting changes in shape and normalization of the resulting distributions are then compared to the default, or nominal, prediction. Simulated jets are smeared by randomly drawing a number, x, from a Gaussian distribution with a mean of 0 and a width equal to the magnitude of the difference between the data and simulation resolutions. The jet energy is then scaled by $(1 + x)$. The resulting uncertainty profiles are then symmetrized about the nominal distribution.

All the uncertainties related to jet reconstruction and energy calibration are summarized for the $(\geq 6\mathrm{j}, \geq 4\mathrm{b})$ region in Table 10.4. Given the large number of jets in the $t\bar{t}H\left(H \to b\bar{b}\right)$ final-state, these components of the total systematic uncertainty play an important role.

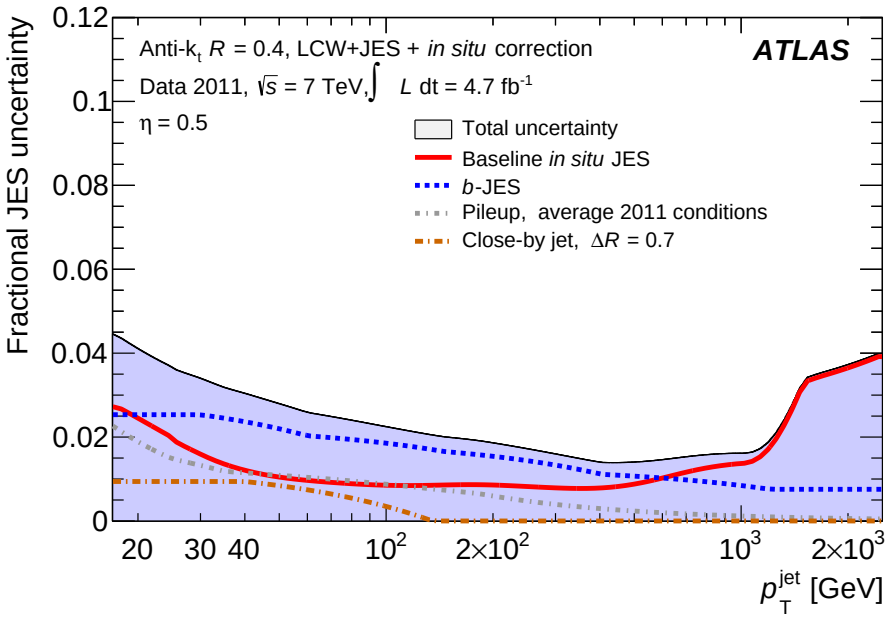
10.1.3 Jet Flavor Identification

The last subcategory of object related systematic uncertainties includes uncertainties related to jet flavor identification or flavor tagging. These uncertainties originate from the algorithms used to tag jets as b-jets, c-jets or light jets. These tagging algorithms are typically calibrated at different working points. In the

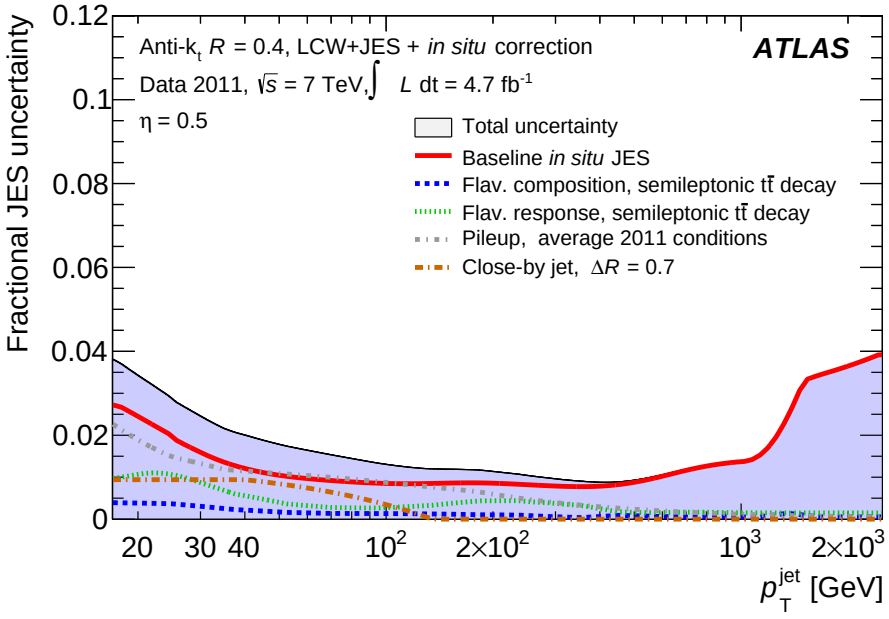
Jet Systematic Uncertainties

Uncertainty Component	**Relative Uncertainty** [%]
JVF	±2.9
Jet Energy Resolution	± ≤0.01
Jet Reconstruction	±1.7
Jet Detector 1	±1.2
Jet Detector 2	± ≤0.01
Jet Detector 3	±0.01
Jet Model 1	±6.3
Jet Model 2	±0.69
Jet Model 3	± ≤0.01
Jet Model 4	± ≤0.01
Jet Mixed 1	± ≤0.01
Jet Mixed 2	± ≤0.01
Jet Statistics 1	±1.4
Jet Statistics 2	±0.08
Jet Statistics 3	±0.52
Jet η Model	±2.7
Jet η Stat	±1.2
Jet Single Parton Response	± ≤0.01
Jet Flavor B	±3.5
Jet Flavor Composition	±3.9
Jet Flavor Response	±2.0
Jet μ	± ≤0.01
Jet n-Pileup-Vertexes	±0.75
Jet Pileup p_T	± ≤0.01
Jet Pileup ρ	±3.8

Table 10.4: Jet reconstruction and calibration uncertainties prior to the fit, in the ($\geq$ 6j, $\geq$ 4b) region over all samples. Uncertainties on the *in situ* assessments of the JES are categorized into four categories defined by the sources: the detector description, the physics modeling, mixed detector and modeling effects, and the statistical uncertainties. These are described in much more detail in Reference [71], pages 46-48.



(a) LCW+JES, anti-kt $R = 0.4$, b-jets



(b) LCW+JES, anti-kt $R = 0.4$, Semileptonic Top-decays

Figure 10.1: Fractional jet energy scale (JES) for selected jets. Figure (a) shows the fractional jet energy scale systematic uncertainty as a function of b-jet p_T for anti-kt jets with distance parameter of R = 0.4 calibrated using the LCW+JES calibration scheme. Figure (b) shows the sample-dependent fractional JES systematic uncertainty as a function of jet p_T for anti-kt jets with distance parameter of R = 0.4 calibrated using the LCW+JES calibration scheme. The uncertainty shown applies to semileptonic top-decays with average 2011 pile-up conditions, and does not include the uncertainty on the jet energy scale of b-jets. [71]

Jet Flavor Tagging Uncertainties

Uncertainty Class	Relative Uncertainty [%]
b-Tagging	± 8.2
c-Tagging	± 4.8
Light-Tagging	± 5.7

Table 10.5: Jet flavor tagging uncertainties prior to the fit, in the $(\geq 6\text{j}, \geq 4\text{b})$ region over all samples. Here, the quoted relative uncertainty is defined as the quadrature sum of the multiple independent components of each uncertainty classification.

case of this analysis, the 70% b-tagging efficiency working point was used. Due to small discrepancies between the simulation and data efficiencies, scale factors are used to maintain agreement. These scale factors subsequently translate into p_T and η dependent systematic uncertainties.

For the b-tagging efficiency, the total uncertainty is decomposed into 7 independent sources of uncertainty. For the c-tagging efficiency, there are 4 independent sources. For the light-tagging efficiency, there are 12 independent sources of uncertainty. In each of these cases, the independent sources are eigenvectors of their respective flavor and p_T dependence of their tagging efficiencies, and are considered uncorrelated between b-jets, c-jets, and light flavor jets. These uncertainties are then propagated through the TRF weighting procedure outlined in Appendix A.

All the uncertainties related to jet flavor tagging are summarized for the $(\geq 6\text{j}, \geq 4\text{b})$ region in Table 10.5. The uncertainties related to the light mis-tag rate are especially important due to the presence of $t\bar{t}+$ Light-Flavor contamination in the $(\geq 6\text{j}, 3\text{b})$ and $(\geq 6\text{j}, \geq 4\text{b})$ signal regions.

10.2 Model Uncertainties

10.2.1 Luminosity

All processes modeled utilizing Monte-Carlo (MC) simulation techniques are scaled to the total integrated luminosity of the $\sqrt{s} = 8$ TeV data set. Therefore, uncertainty on the total integrated luminosity results in associated systematic uncertainties on the normalizations of the constituent MC samples. This uncertainty is estimated to be 2.8% based on the same methods described in Reference [98].

10.2.2 Small Background Modeling and Estimation

W and Z+Jets Modeling

The W+jets background is determined by MC simulation normalized to the inclusive NNLO theoretical prediction [61], as described in Section 5.4. A cor-

Uncertainty Component	**Relative Uncertainty [%]**
W+Jets Cross-Section	±48.0%
W+Jets Cross-Section Ex. 5-Jet	±24.0%
W+Jets Cross-Section Ex. ≥6-Jet	±24.0%
W+Jets p_T Correction	±31.1%
Z+Jets Cross-Section	±48.0%
Z+Jets Cross-Section Ex. 5-Jet	±24.0%
Z+Jets Cross-Section Ex. ≥6-Jet	±24.0%
Z+Jets p_T Correction	±19.4%

Table 10.6: W and Z+jets modeling uncertainties prior to the fit, in the ($\geq$ 6j, $\geq$ 4b) region over the W+jets and Z+jets samples respectively.

rection is then applied to account for measured mis-modeling of the W boson p_T. Thus, a 48% uncertainty is applied to the sample in the 4 jet multiplicity regions and an additional extrapolation uncertainty of 24% is applied to events in the 5 and $\geq$ 6 jet multiplicity regions. Further, to account for uncertainty associated with the W boson p_T correction, a 31.3% uncertainty is applied, which is consistent the size of the correction.

The Z+jets background is similarly determined by MC simulation and is normalized to the NLO theoretical prediction [61]. Discrepancies in the measured versus modeled Z boson p_T are also corrected, resulting in an additional source of uncertainty. Taken together, a baseline uncertainty of 48% is applied to the background normalization with an additional 24% extrapolation uncertainty for events in the 5 and $\geq$ 6 jet multiplicity regions. The size of the Z boson p_T correction is similarly covered with a 19.4% systematic uncertainty. The W and Z+jets modeling uncertainties are summarized for the ($\geq$ 6j, $\geq$ 4b) region in Table 10.6.

Electroweak Backgrounds Modeling

Diboson backgrounds are handled in much the same way as the W+jets and Z+jets backgrounds. Normalizing the MC diboson contribution to the NLO theoretical prediction [62], results in a 5% systematic uncertainty added in quadrature with 24% extrapolation uncertainties for events in the 5 and $\geq$ 6 jet multiplicity regions.

Backgrounds resulting from associated $t\bar{t}$+V production are given a systematic uncertainty of ±30% on the theoretical prediction [99, 100] in addition to contributions from initial state radiation.

Finally, single top production is also modeled using MC techniques utilizing s, t, and Wt-channel production mechanisms [101, 102]. The weighted average of the theoretical prediction uncertainties for these components to the single top background result in an applied systematic uncertainty of $+5\%/-4\%$. An addi-

Uncertainty Component	Relative Uncertainty [%]
Diboson Cross-Section	±24.5%
Diboson Cross-Section Ex. 5-Jet	±24.0%
Diboson Cross-Section Ex. ≥6-Jet	±24.0%
$t\bar{t}+\mathrm{V}$ Cross-Section	±30.0%
$t\bar{t}+\mathrm{V}$ Initial State Radiation Scale	±1.4%
Single Top Cross-Section	±4.2%
Single Top Model	±24.0%

Table 10.7: Diboson, $t\bar{t}+\mathrm{V}$ and single top modeling uncertainties prior to the fit, in the $(\geq 6\mathrm{j}, \geq 4\mathrm{b})$ region over the Diboson, $t\bar{t}+\mathrm{V}$, and Single Top samples respectively.

tional modeling uncertainty of 24% is added to account for extrapolation to high jet multiplicity regions. The electroweak background modeling uncertainties are summarized for the $(\geq 6\mathrm{j}, \geq 4\mathrm{b})$ region in Table 10.7.

Multijet Background Modeling

The estimation of the multijet background from the "Matrix Method" [65] gives rise to uncertainties due to limited statistics in high jet and b-tag multiplicity regions and to uncertainties on the estimated fake rates in the measured control regions. Taking these effects together, a conservative uncertainty of 50% is applied to the multijet normalization which is correlated across jet and b-tag multiplicity regions. This uncertainty is uncorrelated between the muon and electron channels.

10.2.3 Inclusive $t\bar{t}$ Background Modeling and Estimation

Due to the importance of the $t\bar{t}$ backgrounds, multiple modeling systematic uncertainties are taken into account. Theoretical uncertainties, generation uncertainties, and uncertainties on the sequential top-p_T reweighing correction are all taken into account.

Theoretical uncertainties on the inclusive $t\bar{t}$ production cross-section are evaluated using the NNLO and NNLL calculations described in Section 5.3. These result in an overall inclusive $t\bar{t}$ uncertainty of $+6\%/-6.5\%$, which includes uncertainty on the top quark mass.

Further, uncertainties due to the choice of parton shower and hadronization models are also included. These uncertainties are evaluated by comparing results from the Pythia 6.425 [54] and Herwig 6.520 [59] shower models when interfaced with events produced with the Powheg [44] generator. Shape uncertainties are assessed by characterizing the response of these different shower models with respect to the nominal configuration and then symmeterizing the

Inclusive $t\bar{t}$ Background Modeling Uncertainties

Uncertainty Component	Relative Uncertainty [%] $t\bar{t}+b\bar{b}$	$t\bar{t}+c\bar{c}$	$t\bar{t}$ + Light-Flavor
$t\bar{t}$ Cross-Section	±6.2%	±6.2%	±6.3%
$t\bar{t}$ Parton Shower	±10.7%	±16.0%	±12.8%
$t\bar{t}$-p_T Correction - b-tag Efficiency	–	±0.68%	±0.62%
$t\bar{t}$-p_T Correction - Fragmentation Model	–	±2.4%	±1.9%
$t\bar{t}$-p_T Correction - Initial/Final State Radiation	–	±11.8%	±10.6%
$t\bar{t}$-p_T Correction - Jet Energy Resolution	–	±4.0%	±3.5%
$t\bar{t}$-p_T Correction - Close By Jet (JES)	–	±1.3%	±1.2%
$t\bar{t}$-p_T Correction - Jet Detector 1 (JES)	–	±1.4%	±1.2%
$t\bar{t}$-p_T Correction - Jet Eta Calibration (JES)	–	±1.5%	±1.3%
$t\bar{t}$-p_T Correction - Jet Flavor B (JES)	–	–	–
$t\bar{t}$-p_T Correction - Generator Model	–	±2.8%	±2.8%
$t\bar{t}$-p_T Correction - No Top p_T Correction	–	±10.7%	–
$t\bar{t}$-p_T Correction - No $t\bar{t}$-p_T Correction	–	±1.5%	–

Table 10.8: Inclusive $t\bar{t}$ background modeling uncertainties in the in the ($\geq$ 6j, $\geq$ 4b) region for the respective flavor components.

resulting differences. The nature of altering the parton shower model necessarily results in a different heavy-flavor response. As a result, three components are included and considered uncorrelated, one for each flavor component, $t\bar{t}+b\bar{b}$, $t\bar{t}+c\bar{c}$, and $t\bar{t}$ + Light-Flavor. These uncertainties are assessed following the sequential top-p_T reweighing procedure for both parton-shower samples.

Finally, systematic uncertainties are assessed for the sequential top-p_T reweighing correction introduced in Section 5.3. Though this correction is essential for consistency between simulated inclusive $t\bar{t}$ samples and data, its associated systematic uncertainties represent 95% of this measurements total experimental uncertainty. To quantify this, the uncertainty was broken down into 9 components each representing the largest sources of systematic uncertainty for the $t\bar{t}$ differential cross-section measurement which provided the weights for the correction.

It is important to note that the uncertainties related to the $t\bar{t}$-p_T correction are assessed specifically for the $t\bar{t}+c\bar{c}$ and $t\bar{t}$ + Light-Flavor components only, as the $t\bar{t}+b\bar{b}$ component is corrected separately to the NLO Sherpa + OpenLoops [55, 56] description. Furthermore, to address uncertainty regarding the applicability of this evaluation for the $t\bar{t}+c\bar{c}$ component in particular, an additional uncertainty has been evaluated that simply removes the top-p_T and $t\bar{t}$-p_T corrections altogether. A summary of these components, in addition to the other assessed uncertainties, can be found in Table 10.8.

10.2.4 $t\bar{t}$+Heavy-Flavor Content Modeling and Estimation

The nature of the $t\bar{t}+b\bar{b}$ and $t\bar{t}+c\bar{c}$ components of the inclusive $t\bar{t}$ background present large systematic uncertainties primarily due to the complexity of the theoretical calculations necessary for computing beyond-leading-order predictions for these individual components. Fortunately, the NLO SHERPA + OPENLOOPS [55, 56] prediction for $t\bar{t}+b\bar{b}$ provides a solid basis for assessing the impact of various systematic uncertainties for $t\bar{t}+b\bar{b}$ in particular and for both heavy-flavor components in general. As the comparison between the SHERPA + OPENLOOPS prediction and the default POWHEG +PYTHIA prediction agree to within 50%, a conservative 50% systematic uncertainty is applied to the normalization of both of the $t\bar{t}+b\bar{b}$ and $t\bar{t}+c\bar{c}$ components.

As NLO calculations are not available for $t\bar{t}+c\bar{c}$ production, the $t\bar{t}+c\bar{c}$ contribution is additionally assessed systematic uncertainties based on variations in charm quark mass and LO scale variations. It is important to note here that the $t\bar{t}+c\bar{c}$ events are produced via parton shower models using POWHEG [44] , therefore, MADGRAPH 5 LO [63, 64] $t\bar{t}+c\bar{c}$ events are generated and varied. The magnitude of these variations are then applied as systematics to the default $t\bar{t}+c\bar{c}$ samples. The MADGRAPH sample is varied by changing the factorization and renormalization scales up and down by factors of 2, varying the matching threshold and changing the mass of the charm quark, leading to three additional, uncorrelated uncertainties.

As opposed to the $t\bar{t}+c\bar{c}$ component, the $t\bar{t}+b\bar{b}$ component does have an NLO prediction for use in establishing generator level uncertainties. As a result, three scale uncertainties are applied, two PDF uncertainties are applied, one uncertainty is applied on the shower recoil model, and finally two uncertainties are applied accounting for $t\bar{t}+b\bar{b}$ production from multiparton interactions and final-state radiation, for a total of eight uncorrelated systematic uncertainties.

The first scale uncertainty is assessed by varying the default renormalization scale up and down by a factor of two. The second uncertainty replaces the renormalization, factorization, and resummation scale with the the quantity $\prod_{i=t,\bar{t},b,\bar{b}} E_{T,i}^{1/4}$. The final scale uncertainty replaces the renormalization scale with the quantity $(m_t m_{b\bar{b}})^{1/2}$. These different scalings adapt better to the differing production mechanisms of the extra b-quark pairs. The two PDF uncertainties compare the default CT10NLO PDF set [48, 49] against the MSTW2008 68% NNLO PDF set [40, 41] and the NNPDF 2.3 PDF set [103] implementations. A summary of these components, in addition to the other assessed uncertainties, can be found in Table 10.9.

10.2.5 Signal Modeling and Estimation

The modeling of the $t\bar{t}H$ signal samples give rise to five similar sources of systematic uncertainty. First, NLO POWHEL samples were generated specifically

$t\bar{t}$+Heavy Flavor Background Modeling Uncertainties

Uncertainty Component	**Relative Uncertainty [%]**
$t\bar{t} + c\bar{c}$ Cross-Section	$\pm$50.0
$t\bar{t} + c\bar{c}$ Scale	$\pm$2.1
$t\bar{t} + c\bar{c}$ Match Threshold	$\pm$1.5
$t\bar{t} + c\bar{c}$ Charm Mass	$\pm \leq$0.01
$t\bar{t} + b\bar{b}$ Cross-Section	$\pm$50.0
$t\bar{t} + b\bar{b}$ Scale 1	$\pm$1.1
$t\bar{t} + b\bar{b}$ Scale 2	$\pm$2.1
$t\bar{t} + b\bar{b}$ Scale 3	$\pm$5.4
$t\bar{t} + b\bar{b}$ MSTW PDF	$\pm \leq$0.01
$t\bar{t} + b\bar{b}$ NNPDF PDF	$\pm$2.0
$t\bar{t} + b\bar{b}$ Shower Recoil Model	$\pm$3.1
$t\bar{t} + b\bar{b}$ FSR	$\pm$4.6
$t\bar{t} + b\bar{b}$ MPI	$\pm \leq$0.01

Table 10.9: $t\bar{t}$+Heavy Flavor background modeling uncertainties in the in the ($\geq$ 6j, $\geq$ 4b) region for the respective flavor components.

to evaluate the impact of the factorization and renormalization scales on the $t\bar{t}H$ kinematic distributions. These samples varied the default scales by a factor of two, up and down. The effects of these variations were then applied to the default PowHel samples by weighting the nominal sample to match the varied samples. This same procedure was used for the second scale variation in which the default static scale was replaced with a dynamic scale.

In addition to the scale uncertainties, PDF uncertainties are assessed using the PDF4LHC Working Group recommendations found in Reference [104]. The recommendation establishes a single PDF systematic uncertainty by utilizing the CT10NLO PDF set [48, 49] , the MSTW2008 68% NNLO PDF set [40, 41] and the NNPDF 2.3 PDF set [103] . First, an envelope enclosing variations of parameters for each PDF set is evaluated. Then, variations between the PDF sets are evaluated. The resulting uncertainty is obtained by reweighting the signal MC events to the different PDF sets then evaluating the change in acceptance as a function of the final discriminating distributions used in the final fit model.

Finally, uncertainties related to the choice of generator and parton shower models are included. The generator uncertainties are established by comparing the nominal $t\bar{t}H$ PowHel model with a MC@NLO [105–107] model. The difference in Higgs boson p_T is taken as the correction and subsequent systematic uncertainty. For the parton shower model, the uncertainty is evaluated by showering the PowHel samples with Pythia 8.1 [47] and Herwig 6.520 [59] . Differences between the Herwig samples and the nominal samples are then taken as systematic uncertainties. It is important to note that in the case of using Herwig as the parton shower model, corrections are introduced to match

$t\bar{t}H$ Signal Modeling Uncertainties

Uncertainty Component	**Relative Uncertainty** [%]
$t\bar{t}H$ Static Scale Variation	± 0.51
$t\bar{t}H$ Dynamic Scale	± 1.6
$t\bar{t}H$ PDF	± 1.4
$t\bar{t}H$ Generator	± 0.81
$t\bar{t}H$ Parton Shower	± 1.3

Table 10.10: $t\bar{t}H$ signal modeling uncertainties in the in the $(\geq 6\text{j}, \geq 4\text{b})$ region.

the Higgs boson branching fractions with the NLO calculations used to generate the PowHel +Pythia sample [42]. A summary of these uncertainties, in addition to the other assessed uncertainties, can be found in Table 10.10.

Chapter 11

Results

Over the preceding chapters, multiple techniques have been introduced for the purpose of efficiently utilizing the available data with which to measure the potential values of the $t\bar{t}H$ production cross-section. Utilizing the binned discriminating distributions provided by these techniques, statistical methods can be brought to bear to identify and quantify the presence of any signal contribution. The strength of any identified signal contribution will be parameterized by the quantity μ, defined by Eq. 4.14 in Chapter 4.

11.1 Multi-Region Fit Model

In Chapter 6, the foundations of the multi-region fit model were laid with the definition of nine analysis regions based on the jet and b-tag multiplicity of their constituent events. Each of these regions provide a binned distribution on which to fit the model.

The statistical analysis used in conjunction with the multi-region fit is based on a binned, profiled likelihood technique. First, a Poisson probability, Eq. 11.1, is constructed for a given bin which is a function of the expected number of events from contributing processes in the bin, x, and the total number of observed events in the bin n. Then, a likelihood function, Eq. 11.2, is defined as the product of individual Poisson probabilities for each of the considered bins, where $\vec{\mathbf{n}}$ denotes the vector of observed events in each bin and $\vec{\mathbf{x}}$ denotes the vector of expected events in each bin under a given hypothesis:

$$P(n|x) = \frac{x^n e^{-x}}{n!} \tag{11.1}$$

$$L(\vec{\mathbf{x}}|\vec{\mathbf{n}}) = \prod_{i=1}^{N_{\text{bins}}} P(n_i|x_i). \tag{11.2}$$

To incorporate the effects of systematic uncertainties on the number of expected events, nuisance parameters (NPs) are utilized to augment the bin expectations in accordance with their related systematic uncertainty. Equation 11.6 shows how the inclusion of a set of N_θ nuisance parameters, $\boldsymbol{\theta} = \{\theta_1, \theta_2, \cdots, \theta_n\}$, changes the number of nominally expected events, x', to the final expectation x. In this analysis, these NPs are constrained by priors determined from ancillary

measurements. These constraints take the form of Gaussian or log-normal priors, $N(x|\hat{\theta}_{ij}, \sigma^{\theta}_{ij})$, parametrized by the means, $\hat{\theta}_{ij}$, and widths, σ^{θ}_{ij}, of each prior, j, per bin, i. Now we can more explicitly define 'shape' and 'normalization' uncertainties. Shape uncertainties are categorized as NPs which are characterized by bin dependent prior distributions, whereas normalization uncertainties are characterized by bin-independent priors.

$$x_i = x'_i + \sum_{j=1}^{N_\theta} x'_i \theta_j \sigma^{\theta}_{ji} \tag{11.3}$$

$$P(n_i|x_i, \boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\sigma}^{\theta}_i) = \frac{x_i^{n_i} e^{-x_i}}{n_i!} \prod_{j=1}^{N_\theta} N(x|\hat{\theta}_{ji}, \sigma^{\theta}_{ji}) \tag{11.4}$$

$$L(\vec{\mathbf{x}}, \boldsymbol{\theta}|\vec{\mathbf{n}}) = \prod_{i=1}^{N_{\text{bins}}} P(n_i|x_i, \boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\sigma}^{\theta}_i) \tag{11.5}$$

It is here, that the power of the multi-region fit model becomes evident, as the fit seeks to maximize some test statistic by effectively maximizing this likelihood. The fit accomplishes this by identifying values for the NPs which best reflect the available data while penalizing the fit for NPs pulled far from their central values. The benefit of the multi-region fit model to constrain systematic uncertainties comes from the ability of the fit to identify likely values of the NPs in high statistics regions with a higher degree of confidence than in the low statistics regions. This confidence is then transmitted to the low statistics signal regions by constraining the allowed variances of the associated NPs. In other words, the background regions are used to measure the the values of the NPs, generating information which is then used to diminish uncertainty in the signal regions which are used to measure μ. This process of including the NPs into a global, simultaneous fit to data is known as profiling.

The final piece of the likelihood comes in the form of the unconstrained signal strength, μ. This is the same μ as the one introduced earlier in this chapter, which is defined as the ration of the measured to predicted Standard Model cross-sections. This parameter only acts as a multiplicative factor on the signal processes in the bin occupancy expectation and therefore requires a more explicit definition of our hypothesis, x. The resulting change replaces x with the sum of signal, s, and background, b, components, modified by the signal strength parameter, μ. The result of this change gives an updated bin

occupation probability defined by Eq. 11.7:

$$x_i = (\mu s_i + b_i) + \sum_{j=1}^{N_\theta} (\mu s_i + b_i)\theta_j \sigma_{ji}^{\theta} \tag{11.6}$$

$$P(n_i|\mu s_i, b_i, \boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_i, \boldsymbol{\sigma}_i^{\theta}) = \frac{x_i^{n_i} e^{-x_i}}{n_i!} \prod_{j=1}^{N_\theta} N(x|\hat{\theta}_{ji}, \sigma_{ji}^{\theta}). \tag{11.7}$$

We can now simplify the above equations by suppressing the parameters that are known or defined *a priori*. In the case of this analysis, s_i and b_i, are estimated according to Standard Model Monte-Carlo predictions, and as mentioned before, $\hat{\theta}_{ij}$ and σ_{ij}^{θ} are determined by ancillary measurements or by conservative estimates. Thus, the above likelihood can be restated as follows leaving the likelihood defined by Eq. 11.9:

$$P(n_i|\mu, \boldsymbol{\theta}) = \frac{x_i^{n_i} e^{-x_i}}{n_i!} \prod_{j=1}^{N_\theta} N(x_i|\hat{\theta}_{ji}, \sigma_{ji}^{\theta}) \tag{11.8}$$

$$L(\mu, \boldsymbol{\theta}|\vec{\mathbf{n}}) = \prod_{i=1}^{N_{\text{bins}}} P(n_i|\mu, \boldsymbol{\theta}). \tag{11.9}$$

With a likelihood definition in hand, Eq. 11.9, we can now define the final test statistic. This analysis uses the test statistic described by Eq. 11.10. Here, $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ represent fitted values which maximize the likelihood function subject to the constraint that, $0 \leq \hat{\mu} \leq \mu$. $\tilde{\boldsymbol{\theta}}$ denotes the set of nuisance parameters which maximize the likelihood function for a given μ:

$$q_\mu = -2 \log \frac{L(\mu, \tilde{\boldsymbol{\theta}}|\vec{\mathbf{n}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}}|\vec{\mathbf{n}})}. \tag{11.10}$$

This test statistic can be conveniently computed in a number of modes relevant to this analysis. Allowing μ to float unconstrained allows for the identification of a signal strength most consistent with data. Further, constraining μ to equal 0 allows for the determination of the background only prediction. Finally, constraining μ to 1 allows for the determination of the Standard Model prediction. In each of these cases, the RooFit and RooStats frameworks [108] are used for fitting, profiling, and for establishing upper limits on μ using the CLs method [109–111].

11.2 Fit Results

Utilizing the fit model above, multiple parameters are calculated and utilized to quantify our understanding of the $t\bar{t}H_{125}$ process, given our understanding of the present Standard Model backgrounds, our understanding of the ATLAS

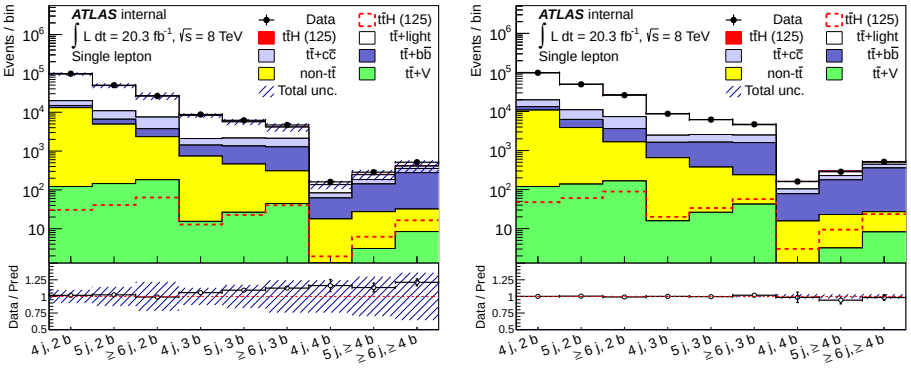


Figure 11.1: Pre-fit (top) and post-fit (bottom) yields and normalization uncertainties for each of the nine analysis regions. [1]

detector, and the constraining power of the 20.3 fb^{-1}, $\sqrt{s}$=8 TeV data set.

First, no significant excess of events is observed in the combined regions of this analysis. Thus, 95% confidence exclusion limits, based on the CLs method, are determined and presented. For the fit based on output distributions using the the *Matrix Element NN* technique, a signal 3.5 times the SM is excluded by the data. If we assume that there is no signal, an assumption consistent with $\mu = 0$, we would expect to exclude signal hypotheses greater than 2.6.

Second, an unconstrained signal-plus-background hypothesis fit allows the most likely μ to be determined with corresponding total and statistical uncertainties. Using the default method, with which the *Matrix Element NN* technique was employed, an excess of events is identified which results in a fitted signal strength of $\mu = 1.2 \pm 1.3(\text{total}) \pm 0.7(\text{stat.})$. The corresponding fitted yields for this result are shown in Table 11.1.

Figure 11.1 shows the yields and their associated uncertainties for the nine analysis regions. The yields and their uncertainties are shown before and after the multi-region fit so as to show the constraining power the fit has on the uncertainty in the various analysis regions. The large reduction in total uncertainty is attributed to the constraints provided by the data in conjunction with the correlations between the different nuisance parameters introduced by the fit. The pre-fit and post-fit effects of the leading systematic uncertainties are shown for the $(\geq 6\text{j}, 3\text{b})$ and $(\geq 6\text{j}, \geq 4\text{b})$ regions in Table 11.2 and in Table 11.3 respectively.

Finally, Figures 11.2-11.4 show all the pre-fit and post-fit characteristic distributions from from all the analysis regions, which were used as inputs to the profiled likelihood fit. For these plots, data is shown as black points with statistical error bars superimposed over the Monte-Carlo based SM predictions. In the ratio subplots, data to Monte-Carlo agreement is given with the total uncertainty per bin given by the hashed area.

	4 j, 2 b	4 j, 3 b	4 j, 4 b
$t\bar{t}H$ (125)	47.6 ± 35	20.0 ± 15	3.0 ± 2.2
$t\bar{t}$+ light	78 200 ± 1600	6260 ± 160	56.5 ± 4.7
$t\bar{t} + c\bar{c}$	6430 ± 1800	845 ± 220	25.5 ± 6.5
$t\bar{t} + b\bar{b}$	2480 ± 490	969 ± 150	62.5 ± 8.5
W+jets	3650 ± 1100	166 ± 51	4.0 ± 1.2
Z+jets	1060 ± 540	49.1 ± 25	1.1 ± 0.6
Single top	4710 ± 320	333 ± 28	6.8 ± 0.7
Diboson	216 ± 65	11.3 ± 3.7	0.3 ± 0.1
$t\bar{t}V$	120 ± 38	15.8 ± 4.9	0.9 ± 0.3
Lepton misID	1080 ± 370	78.4 ± 26	2.6 ± 1.0
Total	98 000 ± 340	8750 ± 82	163 ± 5.6
Data	98 049	8752	161

	5 j, 2 b	5 j, 3 b	5 j, ≥ 4 b
$t\bar{t}H$ (125)	60.4 ± 44	33.7 ± 25	9.4 ± 6.9
$t\bar{t}$+ light	38 400 ± 1000	3610 ± 120	65.3 ± 5.6
$t\bar{t} + c\bar{c}$	4800 ± 1200	935 ± 230	50.7 ± 12
$t\bar{t} + b\bar{b}$	2380 ± 360	1260 ± 180	155 ± 20
W+jets	1210 ± 420	86.6 ± 31	4.0 ± 1.5
Z+jets	368 ± 200	27.9 ± 16	1.4 ± 0.8
Single top	1730 ± 150	185 ± 18	8.2 ± 0.7
Diboson	93.8 ± 35	7.96 ± 3.1	0.5 ± 0.2
$t\bar{t}V$	138 ± 43	26.1 ± 8.1	3.2 ± 1.0
Lepton misID	343 ± 110	43.5 ± 16	5.7 ± 2.2
Total	49 500 ± 220	6220 ± 54	303 ± 9.5
Data	49 699	6199	286

	≥ 6 j, 2 b	≥ 6 j, 3 b	≥ 6 j, ≥ 4 b
$t\bar{t}H$ (125)	89.0 ± 65	57.0 ± 42	23.6 ± 17
$t\bar{t}$+ light	18 900 ± 700	2080 ± 87	57.9 ± 5.3
$t\bar{t} + c\bar{c}$	3730 ± 890	888 ± 210	85.4 ± 21
$t\bar{t} + b\bar{b}$	1980 ± 310	1360 ± 190	331 ± 37
W+jets	455 ± 170	50.7 ± 19	4.43 ± 1.9
Z+jets	152 ± 86	15.6 ± 8.9	1.2 ± 0.7
Single top	734 ± 83	111 ± 14	11.4 ± 1.6
Diboson	44.7 ± 20	5.58 ± 2.6	0.5 ± 0.2
$t\bar{t}V$	166 ± 52	42.3 ± 13	8.2 ± 2.5
Lepton misID	117 ± 41	13.8 ± 5.3	1.1 ± 0.5
Total	26 400 ± 160	4620 ± 55	525 ± 18
Data	26 185	4701	516

Table 11.1: Event yields after the signal-plus-background multi-region fit for signal, backgrounds and data in all 9 analysis regions. The uncertainties quoted are the quadrature sums of the statistical and systematic uncertainties on the yields. These are computed taking into account correlations among nuisance parameters and among processes.

$\geq$ 6 j, 3 b								
	Pre-fit				Post-fit			
	$t\bar{t}H$ (125)	$t\bar{t}$ + light	$t\bar{t}+c\bar{c}$	$t\bar{t}+b\bar{b}$	$t\bar{t}H$ (125)	$t\bar{t}$ + light	$t\bar{t}+c\bar{c}$	$t\bar{t}+b\bar{b}$
Luminosity	±2.8	±2.8	±2.8	±2.8	±2.6	±2.6	±2.6	±2.6
Lepton efficiencies	±1.4	±1.5	±1.4	±1.5	±1.3	±1.3	±1.3	±1.3
Jet energy scale	±5.8	±13	±10	±9.2	±2.2	±5.1	±4.3	±3.5
Jet efficiencies	±1.8	±4.8	±2.8	±2.6	±0.8	±2.1	±1.3	±1.2
Jet energy resolution	±0.1	±3.0	±2.1	±1.4	±0.1	±1.6	±1.1	±0.7
b-tagging efficiency	±4.1	±5.2	±5.0	±5.5	±2.2	±2.9	±2.7	±2.9
c-tagging efficiency	±0.8	±4.7	±6.0	–	±0.5	±4.1	±5.1	–
Light jet-tagging efficiency	–	±5.2	±1.8	–	–	±3.0	±1.0	–
High p_{T}tagging efficiency	–	–	–	–	–	–	–	–
$t\bar{t}$ modelling: reweighting	–	±5.1	±5.9	–	–	±4.6	±5.2	–
$t\bar{t}$ modelling: parton shower	–	±9.0	±16	±10	–	±2.6	±10	±5.6
$t\bar{t}$ heavy-flavour: normalisation	–	–	±50	±50	–	–	±28	±14
$t\bar{t}$ heavy-flavour: modelling	–	±10	±15	±12	–	±3.5	±8.1	±10
Theoretical cross sections	–	±6.2	±6.2	±6.3	–	±4.1	±4.1	±4.1
$t\bar{t}H$ modelling	±2.8	–	–	–	±2.7	–	–	–
Total	±8.5	±23	±57	±54	±4.8	±4.3	±23	±14

Table 11.2: Pre-fit and post-fit normalization uncertainties for the principal uncertainties effecting the signal and $t\bar{t}$ background contributions for the ($\geq 6\mathrm{j}, 3\mathrm{b}$) region. Note that the total uncertainty can be different from the sum in quadrature of individual sources when taking into account the anti-correlations between them. All uncertainties are quoted in terms of % yield. [1]

≥ 6 j, ≥ 4 b								
	Pre-fit				Post-fit			
	$t\bar{t}H$ (125)	$t\bar{t}$ + light	$t\bar{t}+c\bar{c}$	$t\bar{t}+b\bar{b}$	$t\bar{t}H$ (125)	$t\bar{t}$ + light	$t\bar{t}+c\bar{c}$	$t\bar{t}+b\bar{b}$
Luminosity	±2.8	±2.8	±2.8	±2.8	±2.6	±2.6	±2.6	±2.6
Lepton efficiencies	±1.4	±1.4	±1.4	±1.5	±1.3	±1.3	±1.3	±1.3
Jet energy scale	±6.4	±13	±11	±9.2	±2.3	±5.3	±4.7	±3.6
Jet efficiencies	±1.7	±5.2	±2.7	±2.5	±0.7	±2.3	±1.2	±1.1
Jet energy resolution	±0.1	±4.4	±2.5	±1.6	±0.1	±2.3	±1.3	±0.8
b-tagging efficiency	±9.2	±5.6	±5.1	±9.3	±5.0	±3.1	±2.9	±5.0
c-tagging efficiency	±1.7	±6.0	±12	±2.4	±1.4	±5.1	±10	±2.1
Light jet-tagging efficiency	±1.0	±19	±5.2	±2.1	±0.6	±11	±3.0	±1.1
High p_{T}tagging efficiency	±0.6	–	±0.7	±0.6	±0.3	–	±0.4	±0.3
$t\bar{t}$ modelling: reweighting	–	±5.4	±6.1	–	–	±4.7	±5.4	–
$t\bar{t}$ modelling: parton shower	–	±13	±16	±11	–	±3.6	±10	±6.0
$t\bar{t}$ heavy-flavour: normalisation	–	–	±50	±50	–	–	±28	±14
$t\bar{t}$ heavy-flavour: modelling	–	±11	±16	±8.3	–	±3.6	±9.1	±7.1
Theoretical cross sections	–	±6.3	±6.3	±6.3	–	±4.1	±4.1	±4.1
$t\bar{t}H$ modelling	±2.7	–	–	–	±2.6	–	–	–
Total	±12	±32	±59	±54	±6.9	±9.2	±23	±12

Table 11.3: Pre-fit and post-fit normalization uncertainties for the principal uncertainties effecting the signal and $t\bar{t}$ background contributions for the ($\geq$ 6j, $\geq$ 4b) region. Note that the total uncertainty can be different from the sum in quadrature of individual sources when taking into account the anti-correlations between them. All uncertainties are quoted in terms of % yield. [1]

11.2.1 Technique Comparison

The default results presented above have been submitted for publication to the *European Physical Journal C* [1]. What remains in this chapter are the results from the *Matrix Element BDT* technique described in Section 9.2, as compared to the *Matrix Element NN* technique. As a reminder, these techniques differ only by the characteristic distributions used for the $(\geq 6\text{j}, 3\text{b})$ and $(\geq 6\text{j}, \geq 4\text{b})$ signal regions. Furthermore, to provide context to these two sets of results, they are compared with the earlier result published on the same data which utilizes roughly the same fit model and neural networks as the first configuration but without the use of the MEM likelihood variables. [93]

Table 11.4 summarizes the results from the two techniques defined in this analysis in addition to the prior result presented in [93]. Here, the *Kinematic Only NN* analysis technique refers to the earlier form of the $(\geq 6\text{j}, 3\text{b})$ and $(\geq 6\text{j}, \geq 4\text{b})$ Neural Networks which did not utilize any MEM based variables. The *Matrix Element NN* refers to the analysis technique defined in Section 9.1 that utilizes two MEM based variables in addition to the ten kinematic variables used the *Kinematic Only NN* analysis. Finally, the *Matrix Element BDT* analysis technique refers to the much simpler Boosted Decision Tree analysis defined in Section 9.2 that utilizes only six MEM based variables. As can be seen in this table, the *Matrix Element NN* technique results in a fitted signal strength of $\mu = 1.2 \pm 1.3(\text{total}) \pm 0.7(\text{stat.})$, representing a 19% improvement in total uncertainty as compared to the *Kinematic Only NN* technique. Further, the *Matrix Element BDT* technique results in a fitted signal strength of $\mu = 2.9 \pm 1.4(\text{total}) \pm 0.8(\text{stat.})$, representing a 13% improvement in total uncertainty as compared to the *Kinematic Only NN* technique. Regarding the significantly larger fitted μ, this is traced to the inferior separation in the $(\geq 6\text{j}, 3\text{b})$ region in particular, as compared with the *Matrix Element NN* technique. These results can be seen side by side in Figure 11.7.

Table 11.5 summarizes the observed upper limits for $\mu = \sigma(t\bar{t}H)/\sigma^{SM}(t\bar{t}H)$ assuming $m_H = 125$ GeV. In addition, the expected limits for the background only hypothesis are shown with corresponding $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands about the expected median limit. Here we see that the *Matrix Element NN* technique improves the expected limit from the *Kinematic Only NN* technique by 15% to 2.6 times the SM. Similarly, the *Matrix Element BDT* technique improves the expected limit from the *Kinematic Only NN* technique by 12% to 2.7 times the SM. However, unlike the *Matrix Element NN* technique, the *Matrix Element BDT* observes a 2.2σ excess. These results can also be seen in Figure 11.8 showing the different analysis techniques side by side.

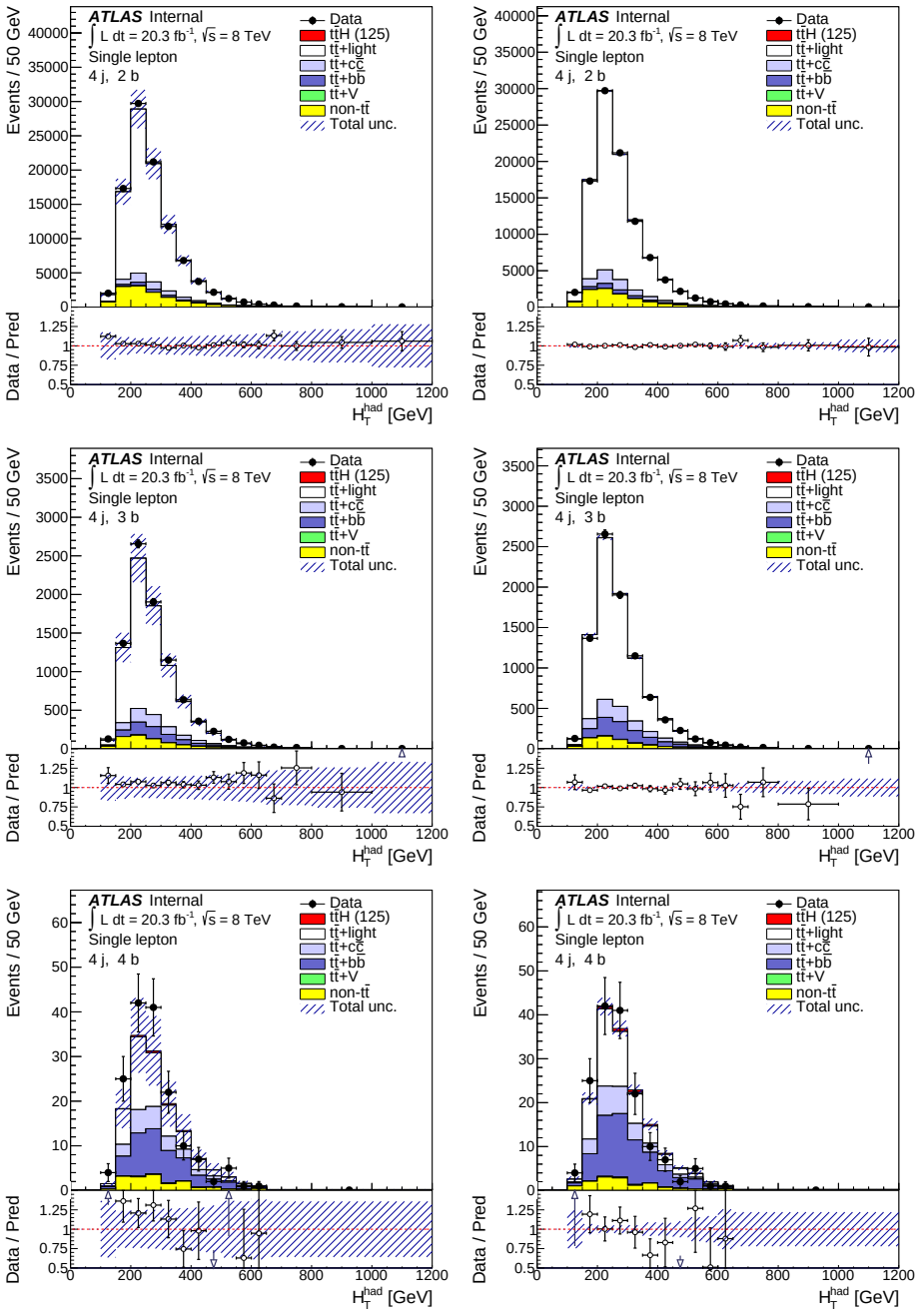


Figure 11.2: Pre-fit (left) and post-fit (right) characteristic distributions for the $(4\text{j}, 2\text{b})$, $(4\text{j}, 3\text{b})$, and $(4\text{j}, \geq 4\text{b})$ regions. [1]

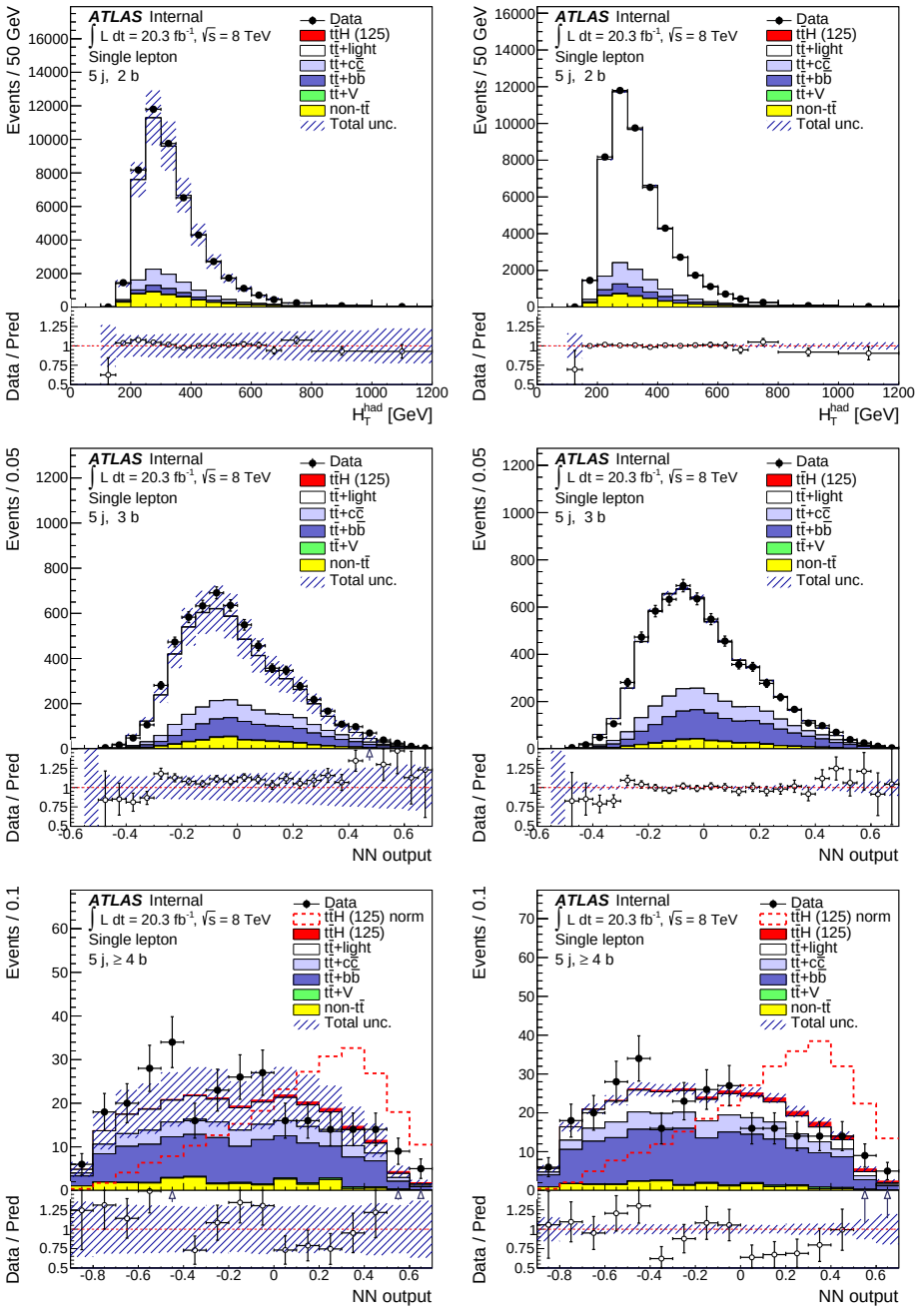


Figure 11.3: Pre-fit (left) and post-fit (right) characteristic distributions for the $(5\text{j}, 2\text{b})$, $(5\text{j}, 3\text{b})$, and $(5\text{j}, \geq 4\text{b})$ regions. [1]

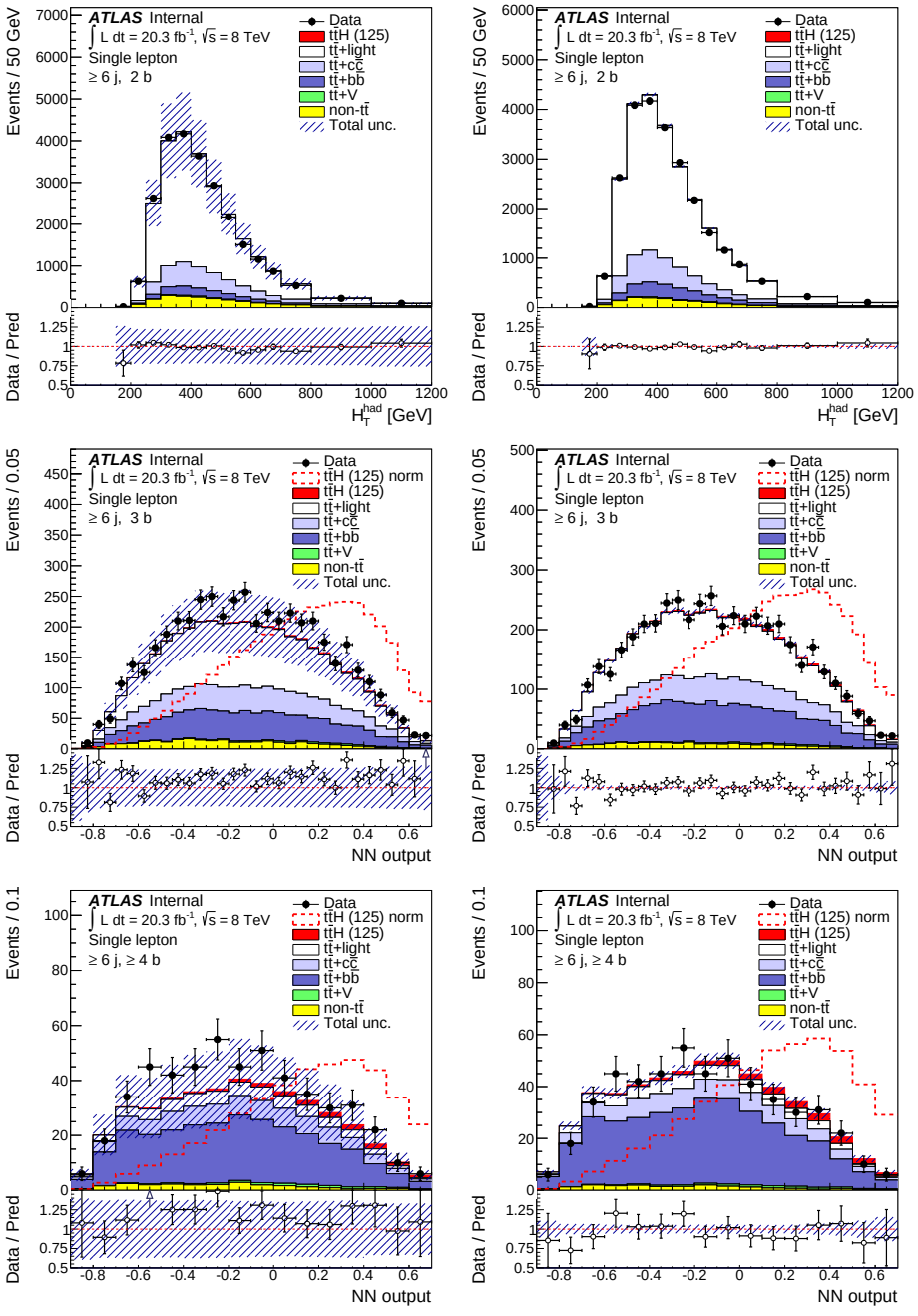


Figure 11.4: Pre-fit (left) and post-fit (right) characteristic distributions for the $(\geq 6\text{j}, 2\text{b})$, $(\geq 6\text{j}, 3\text{b})$, and $(\geq 6\text{j}, \geq 4\text{b})$ region. [1]

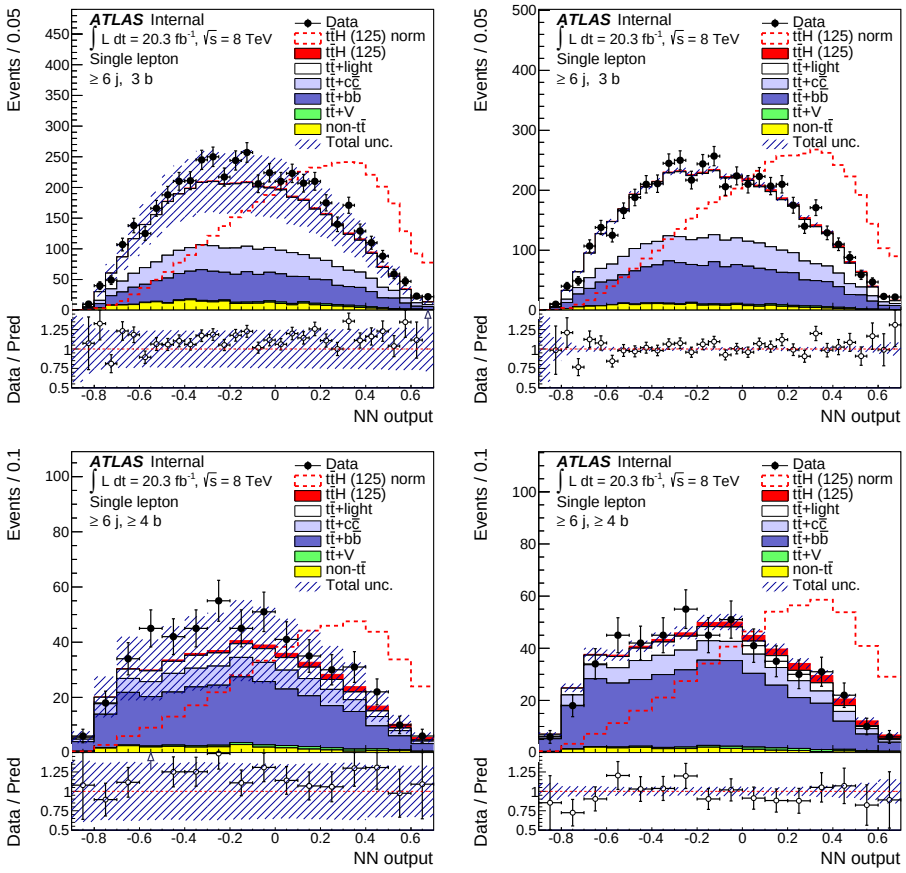


Figure 11.5: Pre-fit (left) and post-fit (right) Neural Network (with MEM Info) characteristic distributions for the $(\geq 6\text{j}, 3\text{b})$ and $(\geq 6\text{j}, \geq 4\text{b})$ regions. [2]

Fitted Values for the Signal Strength

Analysis Technique	**Fitted μ**	**Total Uncertainty**	**Stat. Uncertainty**
Kinematic Only NN†	1.3	$\pm$1.6	$\pm$0.8
Matrix Element NN	1.2	$\pm$1.3	$\pm$0.7
Matrix Element BDT	2.9	$\pm$1.4	$\pm$0.8

Table 11.4: The fitted values for the signal strength, μ, and their associated uncertainties assuming $m_H = 125$ GeV for the various analysis techniques considered by this analysis. † Result from Reference [93].

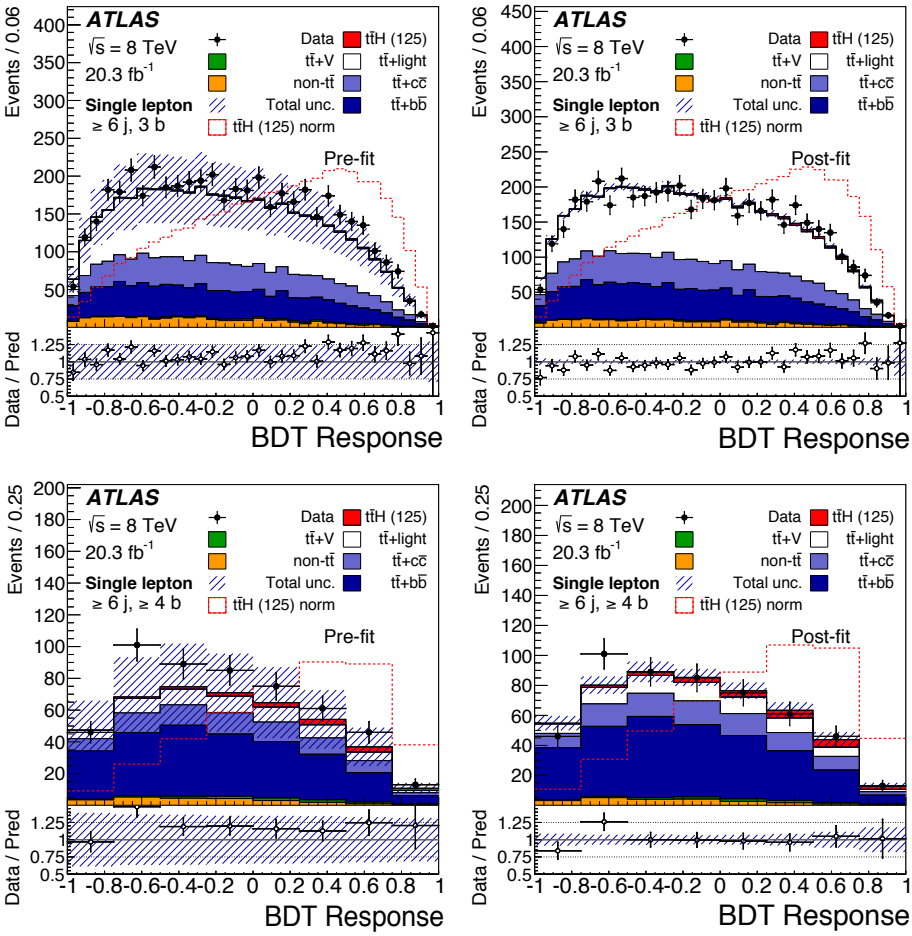


Figure 11.6: Pre-fit (left) and post-fit (right) Boosted Decision Tree characteristic distributions for the ($\geq$ 6j, 3b) and ($\geq$ 6j, $\geq$ 4b) regions.

Upper Exclusion Limits at 95% CL

Analysis Technique	**Observed**	-2σ	-1σ	**Median**	$+1\sigma$	$+2\sigma$
Kinematic Only NN†	4.2	1.7	2.2	3.1	6.0	8.4
Matrix Element NN	3.5	1.4	1.9	2.6	3.7	4.9
Matrix Element BDT	5.2	1.5	2.0	2.7	3.8	5.1

Table 11.5: Observed and expected (median for the background only hypothesis) 95% CL upper exclusion limits on $\sigma(t\bar{t}H)$ with $\pm1\sigma$ and $\pm2\sigma$ uncertainty bands assuming $m_H = 125$ GeV for the various analysis techniques considered by this analysis. † Result from Reference [93].

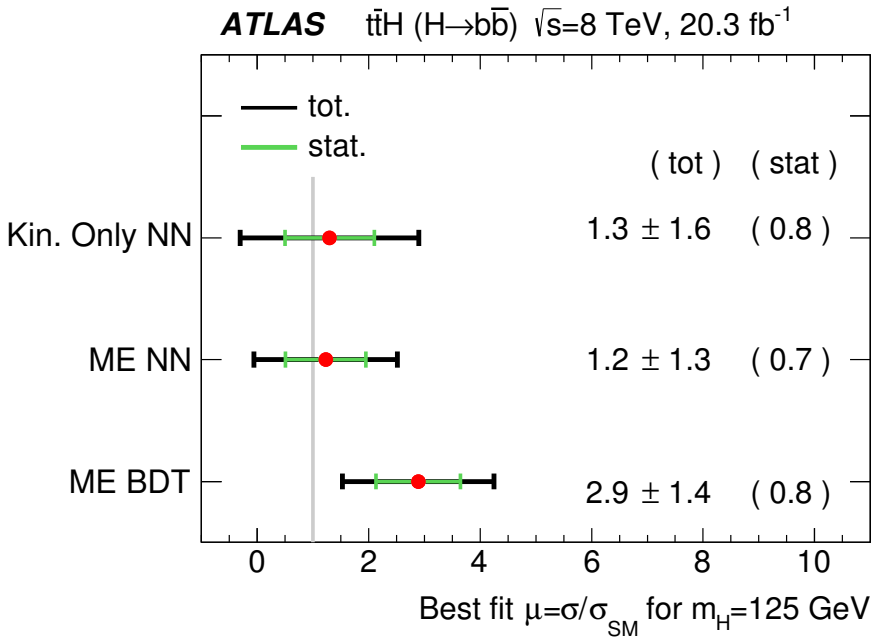


Figure 11.7: Comparisons of the fitted-μ under the signal-plus-background hypothesis using the three competing analysis techniques described in Section 11.2.1.

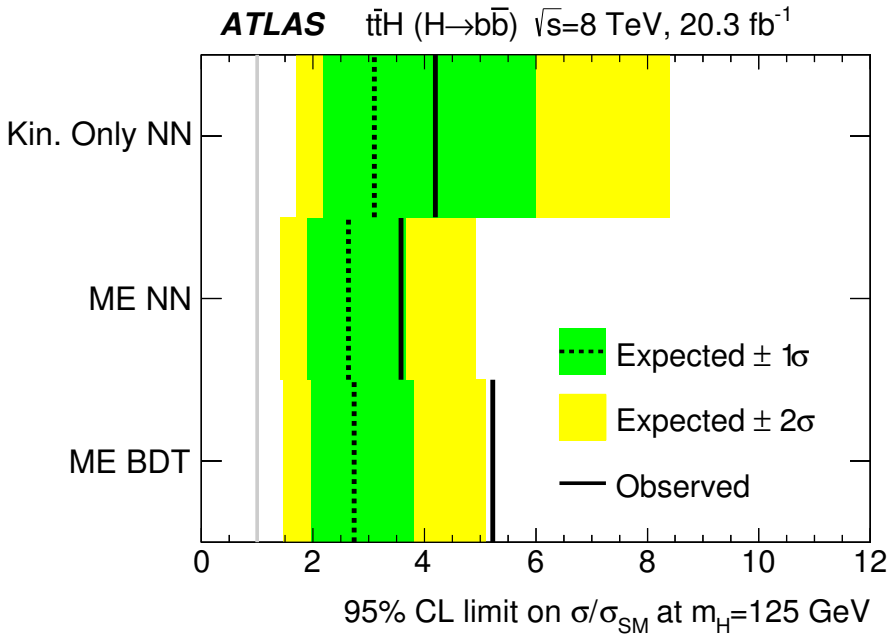


Figure 11.8: Comparisons of the 95% CLs exclusion limits using the three competing analysis techniques described in Section 11.2.1.

Chapter 12

Conclusions and Outlook

This dissertation has presented the development of multiple tailored techniques and methods for identification and quantification of the presence of a SM $t\bar{t}H\left(H \to b\bar{b}\right)$ process in data collected by the ATLAS detector at the Large Hadron Collider. I have presented a complex multi-region analysis with which we were able to significantly inhibit the impact of the present systematic uncertainties. And finally, in spite of considerable technical challenges, the Matrix Element technique has been shown to improve upon this highly-optimized baseline measurement by more than 15%. In addition, it has been shown that the typical likelihood-based outputs of the Matrix Element Method (MEM) can be enhanced via the application of downstream multivariate techniques.

12.1 Summary of Results

Utilizing the 20.1 fb^{-1} of $\sqrt{s} = 8$ TeV data accumulated in 2012 by the ATLAS detector, we have identified an excess of events consistent with the SM single lepton $t\bar{t}H\left(H \to b\bar{b}\right)$ final-state signature. When assessed in the framework of the multi-region fit, this excess is determined to have signal strength equal to 1.2 times the SM prediction with an associated uncertainty of ± 1.3 times the SM prediction. Of insufficient significance to claim an observation in this channel, we place 95% confidence exclusion limits on signal strengths in excess of 3.5 times the standard model. [1]

After adapting the fit model to utilize outputs from a specially designed Boosted Decision Tree, we similarly identify an excess which is determined to have signal strength equal to 2.9 times the SM prediction with an associated uncertainty of ± 1.4 times the SM prediction. Of insufficient significance to claim an observation in this channel, we place 95% confidence exclusion limits on signal strengths in excess of 5.2 times the standard model, with the modified analysis.

12.2 Future Outlook

The MEM has been shown to provide a significant degree of improvement over current state-of-the-art methods for analyzing the complex final-states of the

$t\bar{t}H\left(H \to b\bar{b}\right)$ channels. Furthermore, the MEM implementation reveals a very fundamental application of theoretical information to the problem of classifying measured events, something that will become more and more uncommon as other learning based multivariate techniques become more mainstream. However, it is clear that the significant computational overhead required by the MEM stands as an obstacle to further use.

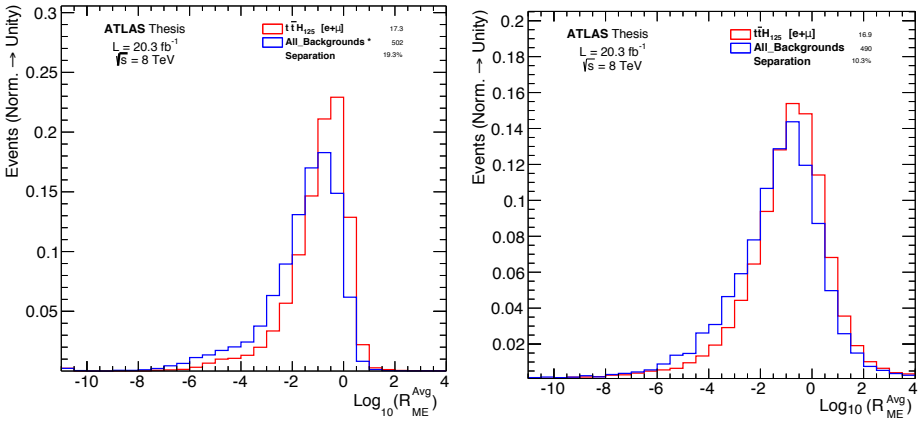
As we look forward, it is clear that ever increasing luminosities at the LHC will result in substantial increases in the number of recorded events. The number of events to be processed gets substantially higher when we take into account the required simulated events used in a given analysis. When these increases in the number of events are taken in the context of the presented MEM implementation, it is clear that substantial improvements will be required to keep computational overhead down to typical analysis timescales. Based on this fact, some effort was expended to identify future avenues of improvement. It is the opinion of the author that a possible way forward involves the use of appropriate machine learning techniques in addition to simply leveraging massively parallel high-performance computing (HPC) and graphical processing units (GPUs).

The principal technique envisioned involves the concept of a 'surrogate model' wherein a deep learning method is used to approximate some non-trivial function. I believe that the matrix element integral at the heart of the MEM can be approximated with such a surrogate model with sufficient accuracy to serve as a classifying discriminant identical to the one used in this analysis.

With this hypothetical technique, a small training set of representative Monte-Carlos would be generated, and subsequently processed with the full MEM, representing some small fraction of the total statistics. The results would then be passed to a regression-based machine-learning algorithm which would be trained to approximate the MEM integral. Finally, the trained algorithm would then attempt to approximate the value of the MEM integral for the remainder of the full dataset without requiring the time-consuming integrations typically required by the MEM.

In fact, this technique was attempted parallel to the development of this analysis. A nearest neighbors (kNN) method was trained on a small subset of this analysis' MEM output. The method then utilized a multi-dimensional linear-regression over selected nearest-neighbors to approximate the MEM output for the remainder of the dataset. The result was a discriminant with 50% less discriminating ability than the fully-integrated MEM, shown in Figure 12.1. However, it was six orders of magnitude faster in its calculation.

Only time will tell how exactly machine learning techniques will finally supplant traditional methods used in experimental high energy physics. However, it is my belief that it is only a matter of when and how, not if. As such, it seems that if the machine learning black-boxes are to take over, we might as well leash them to our full theoretical understanding encoded in the Matrix Element Method.



(a) Fully Integrated

(b) Surrogate Model

Figure 12.1: Figure (a) shows the fully integrated separation power of the MEM at an earlier stage in the development of this analysis. As such, the full statistics shown required over 100,000 CPU-hours to compute, resulting in a separation of 19.3%. Figure (b) shows the surrogate model, trained on a small fraction of events from (a). As such, the full statistics shown required less than 6 minutes to compute, resulting in a separation of 10.3%.

Appendix A

Tag Rate Function Method

Due to the high b-jet multiplicities associated with our signal and background final states, we stand to loose a substantial fraction of the original events simply due to the tagging efficiency of the b-jet identification algorithm. Using the $(\geq 6\text{j}, \geq 4\text{b})$, region as an example, we expect to lose just over 75% of the events due to the 70% efficiency working point of the MV1 b-tagging algorithm.

This loss in events has a significant negative impact on the analysis by substantially increasing the statistical uncertainty in our most signal rich analysis region. Neglecting the diminished sample upon which to train the Neural Network, fewer events introduce statistical fluctuations in both the nominal samples used to validate the simulation as well as in the samples used to estimate the various systematic uncertainties. Thus, there is a tremendous incentive to recover these events, especially considering that their rejection is due to a largely stochastic processes in b-tagging and not the underlying physics of the objects themselves or of the final state taken as a whole.

The TRF method attempts recover the lost events by mapping the true b-jet multiplicity contained in the simulation record to the potential outcomes corresponding to the studied b-jet multiplicity regions. This mapping will then be assigned a weight proportional to the probability of that outcome given the parameterized response of the MV1 algorithm.

In general terms, for a given final state of 6 jets, 4 of which are known to be jets originating from a b-quark via the simulation truth record, a probability will be computed for that event for each b-jet multiplicity region. In this case, it is likely that the probability under the 4 b-tag possibility is greater than the probability it has 0 or 2 b-tags as it is, in truth, a 4 b-tag event. As the MV1 algorithm is somewhat dependent on the kinematic properties of the jet, each jet can subsequently be defined as passing a particular working point under the MV1 algorithm which is then substituted for the actual MV1 weight in all subsequent aspects of the analysis.

It is important to note that this method is, as it requires access to the simulation's truth record, limited to being applied only on simulated data and is obviously not defined for real data. This caveat then carries with it an implicit requirement that the method not introduce a bias into the simulated data such that a potential signal or deficit could be faked. To this point, many studies

have been done to quantify potential bias in the replication of the kinematic distributions using TRF cuts verses MV1 cuts and to identify any scale factors and systematic uncertainties needed to account for this potential source of bias. These studies are presented here, but it is clear that that the tremendous improvement in statistics far outweigh the introduction of the well behaved nuisance parameters introduced by the method.

Below are the validation plots generated for the $t\bar{t}+b\bar{b}$, $t\bar{t}+c\bar{c}$, $t\bar{t}+$ Light-Flavor, and $t\bar{t}H$ samples. Validation plots are generated using the H_T^{had} distributions as well as the NN distributions, all of which are given to the profiled likelihood fit. In all cases observed differences were smaller than the statistical uncertainties from the direct-tagging-based samples. Figures A.1 and A.2 show the $t\bar{t}+b\bar{b}$ validation plots. Figures A.3 and A.4 show the $t\bar{t}+c\bar{c}$ validation plots. Figures A.5 and A.6 show the $t\bar{t}+$ Light-Flavor validation plots. Finally, Figures A.7 and A.8 show the $t\bar{t}H$ validation plots.

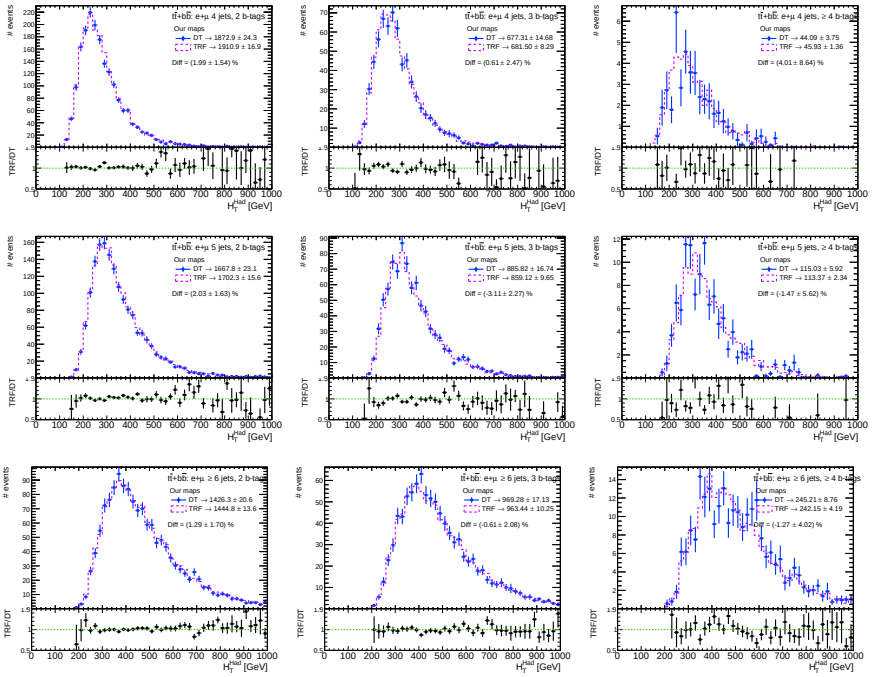


Figure A.1: Agreement between the TRF (violet) and direct MV1 (blue) tagging for the H_T^{had} distribution of the $t\bar{t} + b\bar{b}$ sample. All analysis regions are shown. [2]

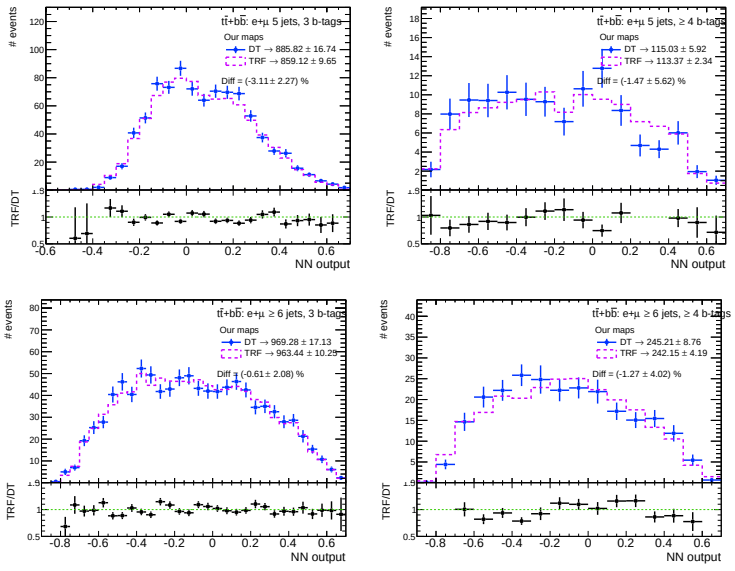


Figure A.2: Agreement between the TRF (violet) and direct MV1 (blue) tagging for the Neural Network output distribution of the $t\bar{t} + b\bar{b}$ sample. All analysis regions are shown. [2]

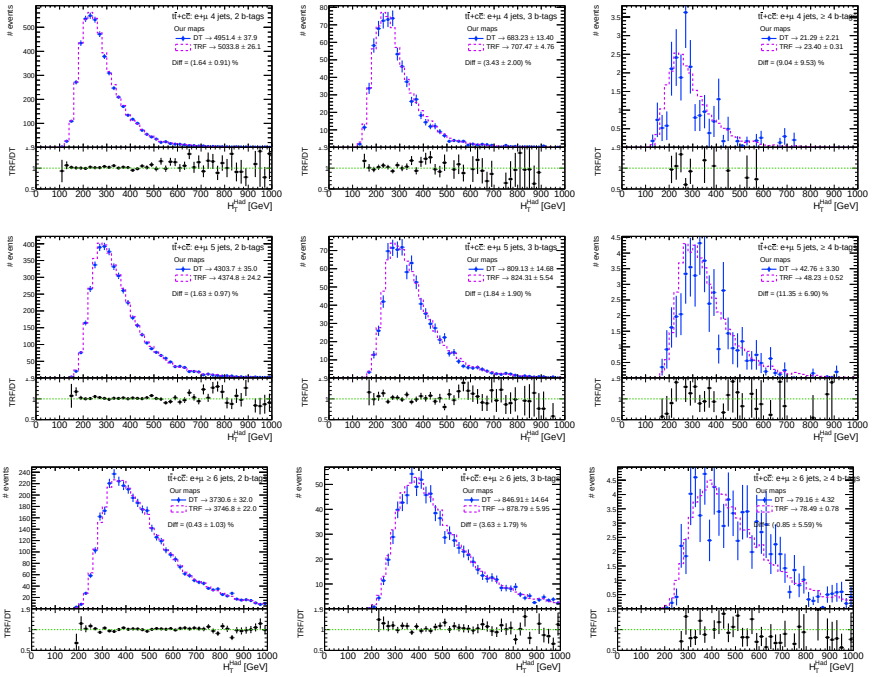


Figure A.3: Agreement between the TRF (violet) and direct MV1 (blue) tagging for the H_T^{had} distribution of the $t\bar{t} + c\bar{c}$ sample. All analysis regions are shown. [2]

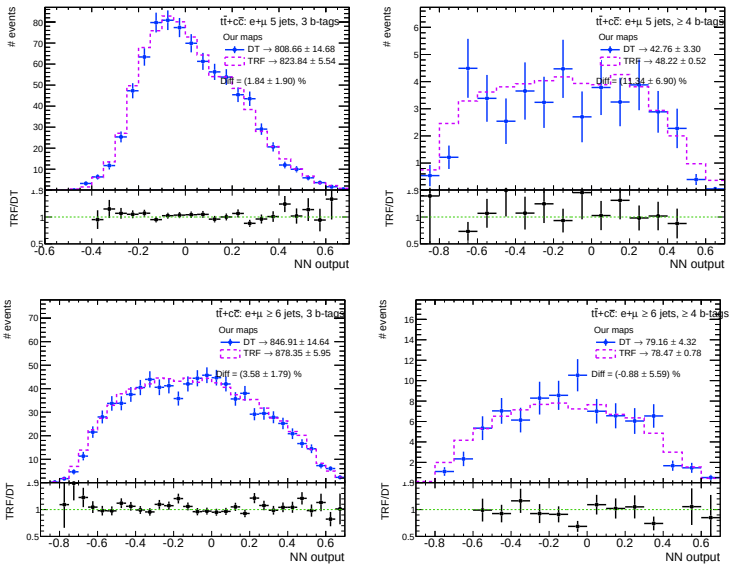


Figure A.4: Agreement between the TRF (violet) and direct MV1 (blue) tagging for the Neural Network output distribution of the $t\bar{t} + c\bar{c}$ sample. All analysis regions are shown. [2]

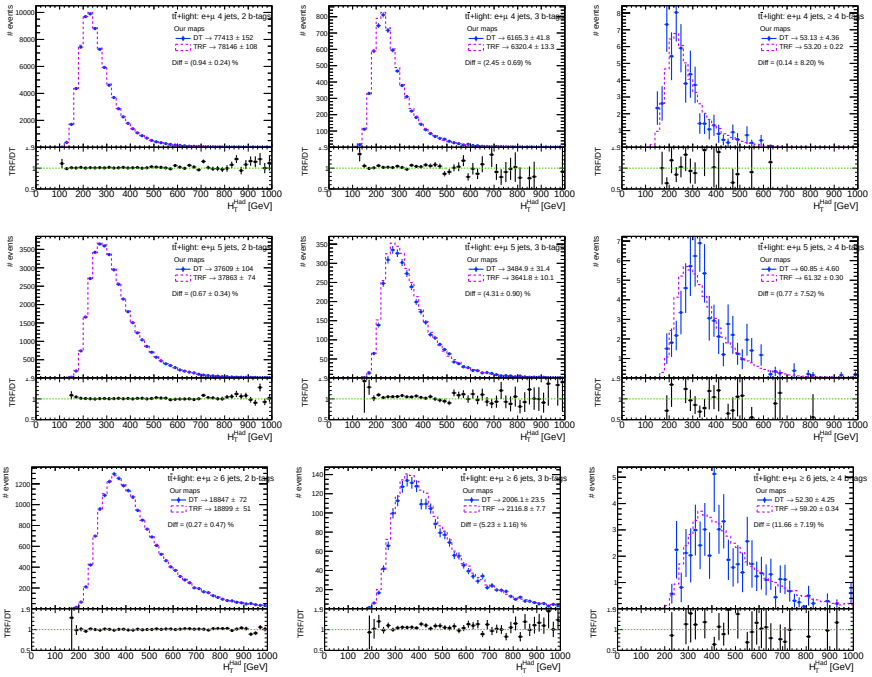


Figure A.5: Agreement between the TRF (violet) and direct MV1 (blue) tagging for the H_T^{had} distribution of the $t\bar{t}$ + Light-Flavor sample. All analysis regions are shown. [2]

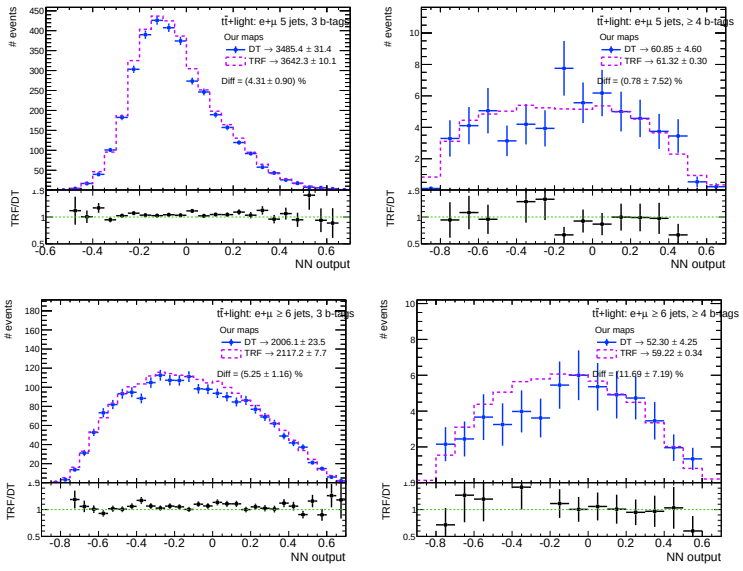


Figure A.6: Agreement between the TRF (violet) and direct MV1 (blue) tagging for the Neural Network output distribution of the $t\bar{t}$ + Light-Flavor sample. All analysis regions are shown. [2]

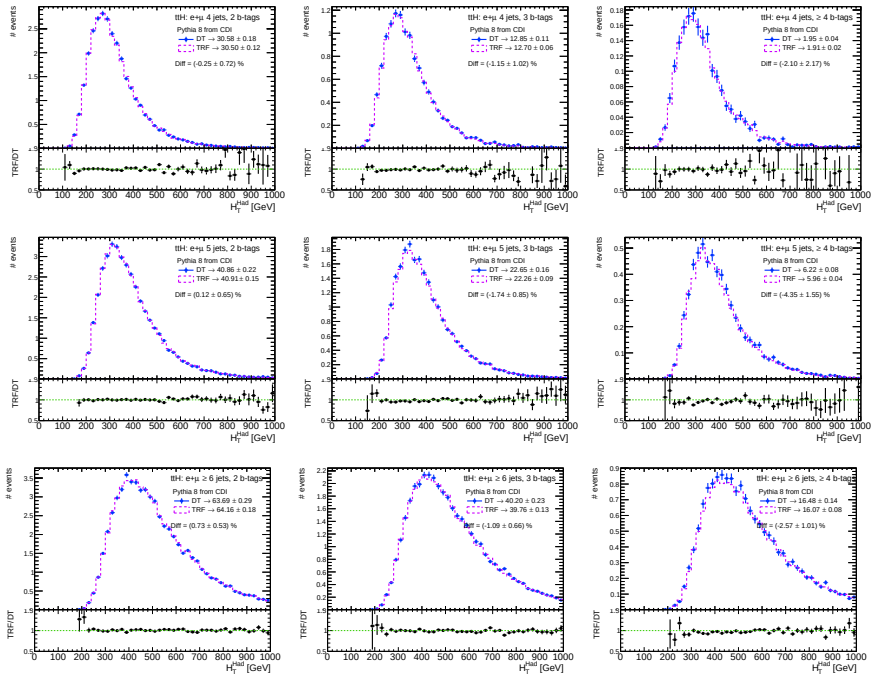


Figure A.7: Agreement between the TRF (violet) and direct MV1 (blue) tagging for the H_T^{had} distribution of the $t\bar{t}H$ sample. All analysis regions are shown. [2]

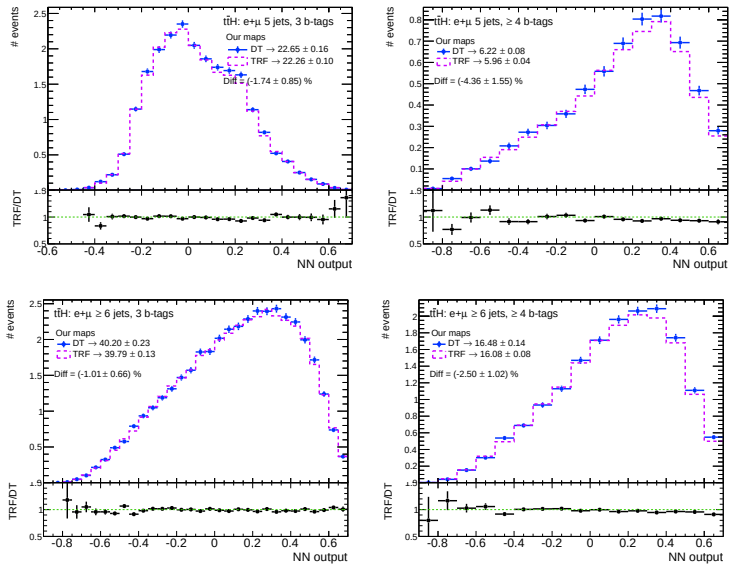


Figure A.8: Agreement between the TRF (violet) and direct MV1 (blue) tagging for the Neural Network output distribution of the $t\bar{t}H$ sample. All analysis regions are shown. [2]

Appendix B

Extended Truth Studies for the MEM

B.1 Truth Sample Preparation

Generated objects in the truth record are compared with every permutation of reconstructed and selected objects. At the time of each comparison, a χ^2 parameter is calculated. The χ^2 is defined as, $\chi^2 = \sum \Delta R_{\eta,\phi}(\text{Reco}_i, \text{Truth}_i)^2$, where the sum is over all true final-state objects. The permutation which minimizes the χ^2, is then chosen as the 'true' permutation. Following the selection of the true permutation, each matched pair of reconstructed and truth objects are further tested. Reconstructed objects which lie outside a $\Delta R_{\eta,\phi}$ cone of 0.4 centered on its matched truth object are vetoed. In this circumstance, the truth object is considered unmatched. For these efficiencies, the denominator includes all generated $t\bar{t}H\left(H \to b\bar{b}\right)$ events in the appropriate channels, and the numerator represents the subset where a selected object was successfully matched to the correct truth object. Tables B.1, B.2, and B.3, define the expected selection efficiencies given losses due primarily to identification, reconstruction and selection effects.

B.2 Truth-Matched Object Assignment Efficiency

The sample of completely truth-matched events is defined such that for every event, each final-state object from the hard process is matched using the process described above. The resulting truth-matched objects are then compared with the maximum likelihood permutation determined by the MEM. The MEM's assignment efficiency is therefore determined by the number of matched objects between the MEM permutation and the truth-matched reconstructed objects, divided by the total number of events in the sample. This efficiency is helpful as the detector response is effectively removed from consideration as it similarly effects the truth-matching and MEM-assignment phases.

Tables B.4, B.5, B.6, and B.7, present the assignment efficiencies of the $t\bar{t}H\left(H \to b\bar{b}\right)$ Matrix Element Method (MEM) given a sample of completely truth-matched events.

Tables B.8, B.9, B.10, and B.11, present the assignment efficiencies of the

Baseline Selection

Full Final State Selection Efficiency[%]

Channel	A H_{125} b	B H_{125} b	Top_H b	Top_L b	A W_H Jet	B W_H Jet	l	ν
≥ 6 J, $= 3$ b, e	84.2	66.3	78.5	80.0	83.3	59.6	100	100
≥ 6 J, $= 3$ b, μ	84.3	66.4	78.2	79.2	83.4	59.6	100	100
≥ 6 J, $= 3$ b, $e + \mu$	84.3	66.3	78.4	79.6	83.3	59.6	100	100
≥ 6 J, ≥ 4 b, e	88.9	80.0	86.8	87.9	87.4	61.6	100	100
≥ 6 J, ≥ 4 b, μ	88.8	80.0	86.7	87.8	87.0	61.9	100	100
≥ 6 J, ≥ 4 b, $e + \mu$	88.9	80.0	86.7	87.8	87.2	61.7	100	100

Table B.1: Full final state selection efficiency, in percent. Reconstructed objects are selected using the object selection criteria defined in 6. This misreconstruction and selection criteria can potentially result in the exclusion of final-state objects from the initial hard process, resulting in a reconstruction and selection efficiency less than 100%. This table shows the total selection efficiency as a function of the various final-state objects from the hard process.

Baseline Selection

Composite Selection Efficiency[%]

Channel	1 W	2 W	1 H_{125}	2 H_{125}	2 Top	Full Top Had	$t\bar{t}$ Sys.	Total Sys.
≥ 6 J, $= 3$ b, e	89.3	53.6	92.5	58.1	64.4	45.7	40.0	29.7
≥ 6 J, $= 3$ b, μ	89.4	53.6	92.7	57.9	63.5	45.0	39.0	29.0
≥ 6 J, $= 3$ b, $e + \mu$	89.3	53.6	92.6	58.0	64.0	45.3	39.5	29.4
≥ 6 J, ≥ 4 b, e	91.3	57.7	96.2	72.7	77.4	53.6	50.5	43.5
≥ 6 J, ≥ 4 b, μ	91.3	57.5	96.3	72.6	77.4	53.0	49.9	43.2
≥ 6 J, ≥ 4 b, $e + \mu$	91.3	57.6	96.3	72.6	77.4	53.3	50.2	43.3

Table B.2: Composite state selection efficiency, in percent. Reconstructed objects are selected using the object selection criteria defined in 6. This misreconstruction and selection criteria can potentially result in the exclusion of final-state objects from the initial hard process, resulting in a reconstruction and selection efficiency less than 100%. This table shows the total selection efficiency as a function of the various final-state objects from the hard process.

Baseline Selection

Jet Final State Selection Efficiency[%]

Channel	A H_{125} b	B H_{125} b	Top_H b	Top_L b	A W_H Jet	B W_H Jet
≥ 6 J, $= 3$ b, e	84.2	66.3	78.5	80.0	83.3	59.6
≥ 6 J, $= 3$ b, μ	84.3	66.4	78.2	79.2	83.4	59.6
≥ 6 J, $= 3$ b, $e + \mu$	84.3	66.3	78.4	79.6	83.3	59.6
≥ 6 J, ≥ 4 b, e	88.9	80.0	86.8	87.9	87.4	61.6
≥ 6 J, ≥ 4 b, μ	88.8	80.0	86.7	87.8	87.0	61.9
≥ 6 J, ≥ 4 b, $e + \mu$	88.9	80.0	86.7	87.8	87.2	61.7

Table B.3: Jet assignment selection efficiency, in percent. Reconstructed objects are selected using the object selection criteria defined in 6. This misreconstruction and selection criteria can potentially result in the exclusion of final-state objects from the initial hard process, resulting in a reconstruction and selection efficiency less than 100%. This table shows the total selection efficiency as a function of the various final-state objects from the hard process.

$t\bar{t} + b\bar{b}$ Matrix Element Method (MEM) given a sample of completely truth-matched events.

B.3 Direct Truth Object Assignment Efficiency

These truth objects are compared with the maximum likelihood permutation determined by the MEM. A match is defined as when the reconstructed object assigned by the MEM lies within a $\Delta R_{\eta,\phi}$ cone of 0.4 centered on its assigned truth object. The MEM's assignment efficiency is therefore determined by the number of matched objects assigned by the MEM, divided by the number of events in the sample. This efficiency is different than the reconstruction-match efficiency in that it does include detector response effects.

Tables 8.1, 8.2, and8.3 present the assignment efficiencies of the $t\bar{t}H\left(H \rightarrow b\bar{b}\right)$ Matrix Element Method (MEM) given a sample of events containing generated truth objects.

Tables B.12, B.13, B.14, and B.15, present the assignment efficiencies of the $t\bar{t} + b\bar{b}$ Matrix Element Method (MEM) given a sample of events containing generated truth objects.

Reco. Object Match [$L^{\max}(t\bar{t}H_{125})$]

Intermediate State Match Efficiency[%]

Channel	W_{H}	W_{L}	$\mathrm{Top_H}$	$\mathrm{Top_L}$	H_{125}
≥ 6 J, $= 3$ b, e	70.3	100	61.9	65.2	51.5
≥ 6 J, $= 3$ b, μ	70.6	100	62.5	65.4	51.0
≥ 6 J, $= 3$ b, $e + \mu$	70.4	100	62.2	65.3	51.2
≥ 6 J, ≥ 4 b, e	80.0	100	65.3	67.7	56.0
≥ 6 J, ≥ 4 b, μ	79.3	100	65.4	68.1	55.5
≥ 6 J, ≥ 4 b, $e + \mu$	79.7	100	65.3	67.9	55.7

Table B.4: Intermediate state matching efficiency, in percent, to truth-matched reconstructed objects, using the maximum signal likelihood jet permutation. Truth-matched reconstructed objects are measured final state particles that have been matched to the generated particles in teh truth record. The matching proceedure requires that the generated particle be within a cone of $\Delta R_{\eta,\phi} = 4.0$. The events used to determine these efficiencies were required to have all truth objects be matched to reconstructed jets which successfully passed the initial object selection.

Reco. Object Match [$L^{\max}(t\bar{t}H_{125})$]

Full Final State Match Efficiency[%]

Channel	A H_{125} b	B H_{125} b	$\mathrm{Top_H}$ b	$\mathrm{Top_L}$ b	A W_{H} Jet	B W_{H} Jet	l	ν
≥ 6 J, $= 3$ b, e	59.7	61.2	53.5	57.7	76.4	79.7	100	100
≥ 6 J, $= 3$ b, μ	59.8	61.3	53.5	58.0	75.4	79.8	100	100
≥ 6 J, $= 3$ b, $e + \mu$	59.8	61.3	53.5	57.8	75.9	79.7	100	100
≥ 6 J, ≥ 4 b, e	63.5	65.6	60.4	61.0	83.2	85.2	100	100
≥ 6 J, ≥ 4 b, μ	62.8	65.1	58.6	61.2	83.3	85.5	100	100
≥ 6 J, ≥ 4 b, $e + \mu$	63.1	65.3	59.5	61.1	83.2	85.4	100	100

Table B.5: Full final state matching efficiency, in percent, to truth-matched reconstructed objects, using the maximum signal likelihood jet permutation. Truth-matched reconstructed objects are measured final state particles that have been matched to the generated particles in teh truth record. The matching proceedure requires that the generated particle be within a cone of $\Delta R_{\eta,\phi} = 4.0$. The events used to determine these efficiencies were required to have all truth objects be matched to reconstructed jets which successfully passed the initial object selection.

Reco. Object Match [$L^{\max}(t\bar{t}H_{125})$]

Composite Match Efficiency[%]

Channel	1 W	2 W	1 H_{125}	2 H_{125}	2 Top	Full Top Had	$t\bar{t}$ Sys.	Total Sys.
≥ 6 J, $= 3$ b, e	88.4	67.7	74.3	46.7	36.9	47.6	34.8	34.8
≥ 6 J, $= 3$ b, μ	87.7	67.5	75.3	45.9	37.3	47.7	35.2	35.2
≥ 6 J, $= 3$ b, $e + \mu$	88.0	67.6	74.8	46.3	37.1	47.6	35.0	35.0
≥ 6 J, ≥ 4 b, e	90.7	77.7	77.7	51.4	44.4	55.9	43.0	43.0
≥ 6 J, ≥ 4 b, μ	92.0	76.8	76.5	51.4	44.2	55.4	42.8	42.8
≥ 6 J, ≥ 4 b, $e + \mu$	91.3	77.3	77.1	51.4	44.3	55.7	42.9	42.9

Table B.6: Composite state matching efficiency, in percent, to truth-matched reconstructed objects, using the maximum signal likelihood jet permutation. Truth-matched reconstructed objects are measured final state particles that have been matched to the generated particles in teh truth record. The matching proceedure requires that the generated particle be within a cone of $\Delta R_{\eta,\phi} = 4.0$. The events used to determine these efficiencies were required to have all truth objects be matched to reconstructed jets which successfully passed the initial object selection. 1 X, denotes the matching of *at least* one of X's daughters, and 2 X, denotes the matching of both of X's daughters.

Reco. Object Match [$L^{\max}(t\bar{t}H_{125})$]

Jet Final State Match Efficiency[%]

Channel	A H_{125} b	B H_{125} b	Top_{H} b	Top_{L} b	A W_{H} Jet	B W_{H} Jet
≥ 6 J, $= 3$ b, e	59.7	61.2	53.5	57.7	76.4	79.7
≥ 6 J, $= 3$ b, μ	59.8	61.3	53.5	58.0	75.4	79.8
≥ 6 J, $= 3$ b, $e + \mu$	59.8	61.3	53.5	57.8	75.9	79.7
≥ 6 J, ≥ 4 b, e	63.5	65.6	60.4	61.0	83.2	85.2
≥ 6 J, ≥ 4 b, μ	62.8	65.1	58.6	61.2	83.3	85.5
≥ 6 J, ≥ 4 b, $e + \mu$	63.1	65.3	59.5	61.1	83.2	85.4

Table B.7: Jet assignment matching efficiency, in percent, to truth-matched reconstructed objects, using the maximum signal likelihood jet permutation. Truth-matched reconstructed objects are measured final state particles that have been matched to the generated particles in teh truth record. The matching proceedure requires that the generated particle be within a cone of $\Delta R_{\eta,\phi} = 4.0$. The events used to determine these efficiencies were required to have all truth objects be matched to reconstructed jets which successfully passed the initial object selection. Jets labeled with an A are the leading jets in p_T, and jets labeled B are sub-leading.

Reco. Object Match $\left[L^{\max}(t\bar{t}b\bar{b})\right]$

Intermediate State Match Efficiency[%]

Channel	W_{H}	W_{L}	$\mathrm{Top_H}$	$\mathrm{Top_L}$	H_{125}
≥ 6 J, $= 3$ b, e	68.7	100	55.7	56.9	35.2
≥ 6 J, $= 3$ b, μ	69.7	100	56.2	58.1	35.3
≥ 6 J, $= 3$ b, $e + \mu$	69.2	100	56.0	57.5	35.2
≥ 6 J, ≥ 4 b, e	80.0	100	60.5	59.2	39.8
≥ 6 J, ≥ 4 b, μ	79.3	100	59.4	58.6	39.1
≥ 6 J, ≥ 4 b, $e + \mu$	79.7	100	60.0	58.9	39.5

Table B.8: Intermediate state matching efficiency, in percent, to truth-matched reconstructed objects, using the maximum background likelihood jet permutation. Truth-matched reconstructed objects are measured final state particles that have been matched to the generated particles in teh truth record. The matching proceedure requires that the generated particle be within a cone of $\Delta R_{\eta,\phi} = 4.0$. The events used to determine these efficiencies were required to have all truth objects be matched to reconstructed jets which successfully passed the initial object selection.

Reco. Object Match $\left[L^{\max}(t\bar{t}b\bar{b})\right]$

Full Final State Match Efficiency[%]

Channel	A H_{125} b	B H_{125} b	$\mathrm{Top_H}$ b	$\mathrm{Top_L}$ b	A W_{H} Jet	B W_{H} Jet	l	ν
≥ 6 J, $= 3$ b, e	46.0	51.1	49.1	49.6	75.2	78.0	100	100
≥ 6 J, $= 3$ b, μ	48.4	51.0	50.0	50.3	75.2	78.6	100	100
≥ 6 J, $= 3$ b, $e + \mu$	47.2	51.1	49.5	50.0	75.2	78.3	100	100
≥ 6 J, ≥ 4 b, e	50.9	55.9	55.7	52.6	83.2	85.2	100	100
≥ 6 J, ≥ 4 b, μ	50.8	56.2	53.8	51.5	83.3	85.5	100	100
≥ 6 J, ≥ 4 b, $e + \mu$	50.8	56.0	54.7	52.0	83.2	85.4	100	100

Table B.9: Full final state matching efficiency, in percent, to truth-matched reconstructed objects, using the maximum background likelihood jet permutation. Truth-matched reconstructed objects are measured final state particles that have been matched to the generated particles in teh truth record. The matching proceedure requires that the generated particle be within a cone of $\Delta R_{\eta,\phi} = 4.0$. The events used to determine these efficiencies were required to have all truth objects be matched to reconstructed jets which successfully passed the initial object selection.

Reco. Object Match $\left[L^{\max}(t\bar{t}b\bar{b})\right]$

Composite Match Efficiency[%]

Channel	1 W	2 W	1 H_{125}	2 H_{125}	2 Top	Full Top Had	$t\bar{t}$ Sys.	Total Sys.
≥ 6 J, $= 3$ b, e	87.0	66.2	67.5	29.7	26.0	42.7	23.5	23.5
≥ 6 J, $= 3$ b, μ	87.0	66.8	69.4	30.0	26.9	43.1	24.0	23.9
≥ 6 J, $= 3$ b, $e + \mu$	87.0	66.5	68.4	29.9	26.5	42.9	23.7	23.7
≥ 6 J, ≥ 4 b, e	90.7	77.7	72.0	34.7	32.2	50.7	30.4	30.4
≥ 6 J, ≥ 4 b, μ	92.0	76.8	72.1	34.9	30.7	50.2	29.4	29.3
≥ 6 J, ≥ 4 b, $e + \mu$	91.3	77.3	72.1	34.8	31.5	50.4	29.9	29.9

Table B.10: Composite state matching efficiency, in percent, to truth-matched reconstructed objects, using the maximum background likelihood jet permutation. Truth-matched reconstructed objects are measured final state particles that have been matched to the generated particles in teh truth record. The matching proceedure requires that the generated particle be within a cone of $\Delta R_{\eta,\phi} = 4.0$. The events used to determine these efficiencies were required to have all truth objects be matched to reconstructed jets which successfully passed the initial object selection. 1 X, denotes the matching of *at least* one of X's daughters, and 2 X, denotes the matching of both of X's daughters.

Reco. Object Match $\left[L^{\max}(t\bar{t}b\bar{b})\right]$

Jet Final State Match Efficiency[%]

Channel	A H_{125} b	B H_{125} b	Top_{H} b	Top_{L} b	A W_{H} Jet	B W_{H} Jet
≥ 6 J, $= 3$ b, e	46.0	51.1	49.1	49.6	75.2	78.0
≥ 6 J, $= 3$ b, μ	48.4	51.0	50.0	50.3	75.2	78.6
≥ 6 J, $= 3$ b, $e + \mu$	47.2	51.1	49.5	50.0	75.2	78.3
≥ 6 J, ≥ 4 b, e	50.9	55.9	55.7	52.6	83.2	85.2
≥ 6 J, ≥ 4 b, μ	50.8	56.2	53.8	51.5	83.3	85.5
≥ 6 J, ≥ 4 b, $e + \mu$	50.8	56.0	54.7	52.0	83.2	85.4

Table B.11: Jet assignment matching efficiency, in percent, to truth-matched reconstructed objects, using the maximum background likelihood jet permutation. Truth-matched reconstructed objects are measured final state particles that have been matched to the generated particles in teh truth record. The matching proceedure requires that the generated particle be within a cone of $\Delta R_{\eta,\phi} = 4.0$. The events used to determine these efficiencies were required to have all truth objects be matched to reconstructed jets which successfully passed the initial object selection. Jets labeled with an A are the leading jets in p_T, and jets labeled B are sub-leading.

Truth Object Match $\left[L^{\max}(t\bar{t}b\bar{b})\right]$

Intermediate State Match Efficiency[%]

Channel	W_{H}	W_{L}	$\mathrm{Top}_{\mathrm{H}}$	$\mathrm{Top}_{\mathrm{L}}$	H_{125}
≥ 6 J, $= 3$ b, e	62.3	49.5	50.8	44.7	30.5
≥ 6 J, $= 3$ b, μ	63.6	47.6	50.5	45.9	30.0
≥ 6 J, $= 3$ b, $e + \mu$	63.0	48.6	50.6	45.3	30.3
≥ 6 J, ≥ 4 b, e	69.7	49.9	54.6	46.7	35.1
≥ 6 J, ≥ 4 b, μ	69.7	48.0	54.3	46.0	33.8
≥ 6 J, ≥ 4 b, $e + \mu$	69.7	49.0	54.4	46.4	34.5

Table B.12: Intermediate state matching efficiency, in percent, to truth objects, using the maximum background likelihood jet permutation. Truth objects are the original generated particles, prior to the detector simulation. The events used to determine these efficiencies were required to have all truth objects be matched to reconstructed jets which successfully passed the initial object selection.

Truth Object Match $\left[L^{\max}(t\bar{t}b\bar{b})\right]$

Full Final State Match Efficiency[%]

Channel	A H_{125} b	B H_{125} b	$\mathrm{Top}_{\mathrm{H}}$ b	$\mathrm{Top}_{\mathrm{L}}$ b	A W_{H} Jet	B W_{H} Jet	l	ν
≥ 6 J, $= 3$ b, e	42.5	45.0	48.7	49.3	68.0	69.5	98.4	21.7
≥ 6 J, $= 3$ b, μ	44.5	44.9	49.4	49.3	68.1	70.0	98.3	22.3
≥ 6 J, $= 3$ b, $e + \mu$	43.5	44.9	49.1	49.3	68.1	69.8	98.3	22.0
≥ 6 J, ≥ 4 b, e	46.7	50.0	55.2	52.2	74.5	75.0	98.6	21.9
≥ 6 J, ≥ 4 b, μ	46.6	50.6	53.4	50.7	73.5	74.8	98.4	23.0
≥ 6 J, ≥ 4 b, $e + \mu$	46.6	50.3	54.3	51.5	74.0	74.9	98.5	22.4

Table B.13: Full final state matching efficiency, in percent, to truth objects, using the maximum background likelihood jet permutation. Truth objects are the original generated particles, prior to the detector simulation. The events used to determine these efficiencies were required to have all truth objects be matched to reconstructed jets which successfully passed the initial object selection.

Truth Object Match $\left[L^{\max}(t\bar{t}b\bar{b})\right]$

Composite Match Efficiency[%]

Channel	1 W	2 W	1 H_{125}	2 H_{125}	2 Top	Full Top Had	$t\bar{t}$ Sys.	Total Sys.
≥ 6 J, $= 3$ b, e	79.4	58.1	61.9	25.6	26.1	38.4	21.4	18.1
≥ 6 J, $= 3$ b, μ	79.6	58.6	63.8	25.6	26.3	38.0	21.2	18.2
≥ 6 J, $= 3$ b, $e + \mu$	79.5	58.3	62.9	25.6	26.2	38.2	21.3	18.2
≥ 6 J, ≥ 4 b, e	82.2	67.4	66.8	29.9	32.0	44.9	27.1	23.6
≥ 6 J, ≥ 4 b, μ	82.6	65.7	67.1	30.1	30.5	43.7	25.9	22.7
≥ 6 J, ≥ 4 b, $e + \mu$	82.4	66.5	66.9	30.0	31.3	44.3	26.5	23.1

Table B.14: Composite state matching efficiency, in percent, to truth objects, using the maximum background likelihood jet permutation. Truth objects are the original generated particles, prior to the detector simulation. The events used to determine these efficiencies were required to have all truth objects be matched to reconstructed jets which successfully passed the initial object selection. 1 X, denotes the matching of *at least* one of X's daughters, and 2 X, denotes the matching of both of X's daughters.

Truth Object Match $\left[L^{\max}(t\bar{t}b\bar{b})\right]$

Jet Final State Match Efficiency[%]

Channel	A H_{125} b	B H_{125} b	Top_{H} b	Top_{L} b	A W_{H} Jet	B W_{H} Jet
≥ 6 J, $= 3$ b, e	42.5	45.0	48.7	49.3	68.0	69.5
≥ 6 J, $= 3$ b, μ	44.5	44.9	49.4	49.3	68.1	70.0
≥ 6 J, $= 3$ b, $e + \mu$	43.5	44.9	49.1	49.3	68.1	69.8
≥ 6 J, ≥ 4 b, e	46.7	50.0	55.2	52.2	74.5	75.0
≥ 6 J, ≥ 4 b, μ	46.6	50.6	53.4	50.7	73.5	74.8
≥ 6 J, ≥ 4 b, $e + \mu$	46.6	50.3	54.3	51.5	74.0	74.9

Table B.15: Jet assignment matching efficiency, in percent, to truth objects, using the maximum background likelihood jet permutation. Truth objects are the original generated particles, prior to the detector simulation. The events used to determine these efficiencies were required to have all truth objects be matched to reconstructed jets which successfully passed the initial object selection. Jets labeled with an A are the leading jets in p_T, and jets labeled B are sub-leading.

Appendix C

The AdaBoost and the BDT Implementation

Like all boosting algorithms, the AdaBoost method seeks to construct a function, $\hat{F}(\mathbf{i})$, which accurately reproduces the the response of some target function, $f(\mathbf{i})$. As is the case with binary classifiers, this target function maps some input vector or tuple onto the binary label space defined as, $f(\mathbf{i}) = \{-1, 1\}$. In general, the input vectors are pulled from some combination of input distributions from the different classes according to the cumulative distribution $\mathscr{P}$. This distribution is typically fixed *a priori* which the boosting algorithm will attempt learn over the evolution of the boosting algorithm. To learn to approximate, $f(\mathbf{i})$ on $\mathscr{P}$, AdaBoost iteratively constructs $\hat{F}(\mathbf{i})$ out of 'weak learners, (i.e., simple MVAs such as a shallow decision tree), $h(\mathbf{i})$, well suited to minimizing the error $\epsilon = \sum_{j=1}^{N} e^{-h(\mathbf{i}_j) f(\mathbf{i}_j)}$, where N is the number of input vectors. What is unique about AdaBoost is that it builds a parallel distribution to $\mathbf{i} \in \mathscr{P}$ by manipulating weights assigned to each input vector. This weighting is updated after each epoch allowing new 'weak learners' to be produced which are adept at classifying previously 'hard' cases. This algorithm is shown in pseudo-code in Algorithm 1.

Finally, in addition to boosting, this analysis also includes the use of three additional regularization schemes designed to reduce potential overtraining pathologies. First, *bagging* is used with the sample fraction parameter set to 55%. This setting causes each training epoch to utilize a 55% subsample of training events. This subsample is chosen, with replacement, from the training data at the beginning of each epoch, maintaining the AdaBoost weights from prior evaluations. Second, a *shrinkage* factor of 0.2 is utilized to retard the learning process and further increase the robustness of the final classifier. Finally, this implementation is configured to utilize randomized trees which utilize a random subset of input variables with which to construct a particular decision tree. All remaining settings and hyper-parameters are listed in Table C.1.

TMVA Implementation of the AdaBoost Algorithm Data:

- Some target binary classification function, $f(\mathbf{i}) = \{-1, 1\}$
- N labeled samples, $\mathbf{i}$, from $\mathscr{P}$: $\langle (\mathbf{i}_1, f(\mathbf{i}_1)), (\mathbf{i}_2, f(\mathbf{i}_2)), \cdots, (\mathbf{i}_N, f(\mathbf{i}_N)) \rangle$
- Prior weight distribution D over the N samples
- Integer T specifying the number of training epochs
- Identify **WeakLearner**, $h(\mathbf{i})$, for which $\langle |h(\mathbf{i}) - f(\mathbf{i})| \rangle \leq 1/2$

Result: A function, $\hat{F}(\mathbf{i})$, which closely approximates the target function, $f(\mathbf{i})$.

Initialization:

- Initialize sample weight vector, $w_j^1 = D^1(j)$ for $j = \{1, 2, \cdots, N\}$
- If no prior information, set $w_j^1 = \frac{1}{N}$

for $t = \{1, 2, 3, \cdots, T\}$ **do**

1. Generate normalized *emphasis vector* $\mathbf{p}^t$:

$$\mathbf{p}^t = \frac{\mathbf{w}^t}{\sum_{j=1}^{N} w_j^t}$$

2. Optimize **WeakLearner**, $h^t(\mathbf{i})$, on samples $\mathbf{i} \in \mathscr{P}$, weighted with the *emphasis vector* $\mathbf{p}^t$, such that $h^t : \mathbf{i} \mapsto [-1, 1]$
3. Calculate **WeakLearner** error over the training sample:

$$\epsilon^t = \sum_{j=1}^{N} p_j^t e^{-h^t(\mathbf{i}_j) f(\mathbf{i}_j)}$$

4. Calculate weight attributed to this **WeakLearner**:

$$\alpha^t = \frac{1}{2} \log \left(\frac{1 - \epsilon^t}{\epsilon^t} \right)$$

5. Add weighted **WeakLearner** to ensemble:

$$F^t(\mathbf{i}) = F^{t-1}(\mathbf{i}) + \alpha^t h^t(\mathbf{i})$$

6. Update weights for next iteration:

$$w_j^{t+1} = w_j^t e^{-\alpha^t |h^t(\mathbf{i}_j) - f(\mathbf{i}_j)|}$$

end

Return:

$$\hat{F}(\mathbf{x}) = \frac{1}{T} F^T(\mathbf{x}) = \frac{1}{T} \sum_{t=1}^{T} \alpha^t h^t(\mathbf{x})$$

Algorithm 1: This pseudocode shows the TMVA implementation of the AdaBoost algorithm used for this analysis. The power of this method is two fold and is largely encapsulated in steps two and four. First, in step two, the **WeakLearner** is optimized on pre-weighted samples were misclassified or 'hard' samples are *emphasized* with higher than average weights (step six). Thus, new **WeakLearners** are developed specifically to handle hard cases. Second, in step four, the resulting **WeakLearners** are added to then ensemble in proportion to their minimization of the error, thus stronger **WeakLearners** impact the final classification more heavily than than weaker ones. [96, 97]

TMVA BDT Hyper-Parameters

Hyper-parameter	**Value**
# of variables	6
# of variable cut locations (granularity)	100
Maximum depth (layers)	5
Boost type	AdaBoost
AdaBoost learning rate	0.01
Learning rate shrinkage	0.2%
# of training epochs	2000
Bagged sample fraction	55%
Minimum node occupancy	1.25%

Table C.1: Full details of the TMVA BDT settings BDT utilized in the two signal regions. [97]

Bibliography

[1] ATLAS Collaboration, G. Aad et al., *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, arXiv:1503.05066 [hep-ex].

[2] S. Guindon, E. Shabalina, J. Adelman, M. Alhroob, A. Basye, J. Bouffard, M. Casolino, I. Connelly, V. Dao, T. Doyle, F. Filthaut, S. Henkelmann, V. Jain, A. Juste Rozas, A. Knue, T. Liss, E. Le Menedeu, J. Montejo Berlingen, M. Moreno Llacer, O. Nackenhorst, T. J. Neep, M. Owen, M. Pinamonti, Y. Qin, A. Quadt, C. Schwanenberger, L. Serkin, J. Thomas-Wilsker, and T. Vazquez Schroeder, *Search for the Standard Model Higgs boson produced in association with top quarks in **pp** collisions at 8 TeV with the ATLAS detector at the LHC*, ATL-COM-PHYS-2014-1471 (2014). https://cds.cern.ch/record/1969527.

[3] M. E. Peskin and D. V. Schroeder, *An introduction to quantum field theory*. Advanced book program. Westview Press Reading (Mass.), Boulder (Colo.), 1995. http://opac.inria.fr/record=b1131978. Autre tirage : 1997.

[4] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. Wiley, 1984.

[5] D. Griffiths, *Introduction to Elementary Particles*. Physics textbook. Wiley, 2008. http://books.google.com/books?id=w9Dz56myXm8C.

[6] Particle Data Group Collaboration, K. Olive et al., *Review of Particle Physics*, Chin.Phys. **C38** (2014) 090001.

[7] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, *Experimental Test of Parity Conservation in Beta Decay*, Phys. Rev. **105** (1957) 1413–1415. http://link.aps.org/doi/10.1103/PhysRev.105.1413.

[8] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Physical Review Letters **13** (1964) 508–509.

[9] G. S. Guralnik, C. R. Hagen, and T. W. Kibble, *Global Conservation Laws and Massless Particles*, Physical Review Letters **13** (1964) 585–587.

[10] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Physical Review Letters **13** (1964) 321–323.

[11] A. Collaboration, *Summary plots from the ATLAS Standard Model physics group*, Mar, 2015. https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/.

[12] F. Bezrukov and M. Shaposhnikov, *Why should we care about the top quark Yukawa coupling?*, arXiv:1411.1923 [hep-ph].

[13] N. Arkani-Hamed, A. G. Cohen, E. Katz, and A. E. Nelson, *The Littlest Higgs*, Journal of High Energy Physics **2002** no. 07, (2002) 034. http://stacks.iop.org/1126-6708/2002/i=07/a=034.

[14] N. Arkani-Hamed, A. G. Cohen, and H. Georgi, *Electroweak symmetry breaking from dimensional deconstruction*, Physics Letters B **513** no. 1–2, (2001) 232 – 240. http://www.sciencedirect.com/science/article/pii/S0370269301007419.

[15] M. Carena, E. Pontón, J. Santiago, and C. Wagner, *Light Kaluza–Klein states in Randall–Sundrum models with custodial*, Nuclear Physics B **759** no. 1–2, (2006) 202 – 227. http://www.sciencedirect.com/science/article/pii/S0550321306007991.

[16] R. Contino, L. Da Rold, and A. Pomarol, *Light custodians in natural composite Higgs models*, Phys. Rev. D **75** (2007) 055014. http://link.aps.org/doi/10.1103/PhysRevD.75.055014.

[17] A. Collaboration, *Summary plots from the ATLAS Higgs physics group*, Jan, 2015. https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/HIGGS/.

[18] M. Brice, *Installing the ATLAS calorimeter*, Image CERN-EX-0511013-01, Nov, 2005. https://cds.cern.ch/record/910381.

[19] *ECFA-CERN Workshop on Large Hadron Collider in the LEP Tunnel Proceedings*. 1984.

[20] The ATLAS Collaboration and LHC Experiments Committee, *ATLAS : letter of intent for a general-purpose pp experiment at the large hadron collider at CERN*, Detectors and Experimental Techniques (1992). https://cds.cern.ch/record/291061?ln=en.

[21] CERN, *ATLAS: Detector and physics performance technical design report. Volume 1*. No. CERN-LHCC-99-14, ATLAS-TDR-14. 1999. http://inspirehep.net/record/511648?ln=en.

[22] The ATLAS Collaboration, *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, arXiv:0901.0512. http://arxiv.org/abs/arXiv:0901.0512.

[23] J. Osborne, *CERN Civil Engineering*, presentation, CERN, London, November, 2012. https://edms.cern.ch/document/1251453/1.

[24] ATLAS Collaboration, ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008).

[25] T. C. Collaboration, *The CMS experiment at the CERN LHC*, Journal of Instrumentation **3** no. 08, (2008) S08004. http://stacks.iop.org/1748-0221/3/i=08/a=S08004.

[26] T. A. Collaboration, *The ALICE experiment at the CERN LHC*, Journal of Instrumentation **3** no. 08, (2008) S08002. http://stacks.iop.org/1748-0221/3/i=08/a=S08002.

[27] T. L. Collaboration, *The LHCb Detector at the LHC*, Journal of Instrumentation **3** no. 08, (2008) S08005. http://stacks.iop.org/1748-0221/3/i=08/a=S08005.

[28] D. Boussard and T. P. R. Linnecar, *The LHC Superconducting RF System*, Tech. Rep. LHC-Project-Report-316. CERN-LHC-Project-Report-316, CERN, Geneva, Dec, 1999.

[29] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008).

[30] E. Ciapala, L. Arnaudon, P. Baudrenghien, O. Brunner, A. Butterworth, T. Linnecar, P. Maesen, J. Molendijk, E. Montesinos, D. Valuch, and F. Weierud, *Commissioning of the 400 MHz LHC RF System,*.

[31] S. Dailler, *Cross section of LHC dipole.. Dipole LHC: coupe transversale.*, tech. rep., Apr, 1999. AC Collection. Legacy of AC. Pictures from 1992 to 2002.

[32] J. Pequenao, *Event Cross Section in a computer generated image of the ATLAS detector.*, tech. rep., Mar, 2008.

[33] ATLAS Collaboration, *Expected Performance of the ATLAS Experiment: Detector, Trigger and Physics*. CERN, Geneva, 2009.

[34] ATLAS Collaboration, ATLAS Collaboration, *ATLAS Muon Spectrometer: Technical Design Report,*. CERN-LHCC-97-22.

[35] ATLAS Collaboration Collaboration, *ATLAS Computing: technical design report*. Technical Design Report ATLAS. CERN, Geneva, 2005.

[36] N. Garelli, *ATLAS DAQ Public Results*, Jan, 2013. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsDAQ.

[37] G. Unal and L. Fiorini, *ATLAS Luminosity Public Results*, Sep, 2013. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[38] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv:1207.7214 [hep-ex].

[39] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv:1207.7235 [hep-ex].

[40] A. D. Martin et al., *Parton distributions for the LHC*, Eur. Phys. J. C **63** (2009) 189, arXiv:0901.0002 [hep-ph].

[41] A. D. Martin et al., *Uncertainties on alpha$_S$ in global PDF analyses and implications for predicted hadronic cross sections*, Eur. Phys. J. C **64** (2009) 653, arXiv:0905.3531 [hep-ph].

[42] LHC Higgs Cross Section Working Group Collaboration, S. Dittmaier et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, arXiv:1101.0593 [hep-ph].

[43] G. Bevilacqua, M. Czakon, M. Garzelli, A. van Hameren, A. Kardos, C. Papadopoulos, R. Pittau, and M. Worek, HELAC-NLO, Comput. Phys. Commun. **184** (2013) 986, arXiv:1110.1499v2 [hep-ph].

[44] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040.

[45] S. Frixione, P. Nason, and C. Oleari JHEP **11** (2007) 070, arXiv:0709.2092 [hep-ph].

[46] S. Alioli, P. Nason, C. Oleari, and E. Re JHEP **06** (2010) 040, arXiv:1002.2581 [hep-ph].

[47] T. Sjöstrand, S. Mrenna, and P. Skands, *A Brief Introduction to Pythia 8.1*, arXiv:0710.3820 [hep-ph].

[48] H.-L. L. et al., *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024, arXiv:1007.2241 [hep-ph].

[49] J. G. et al., *The CT10 NNLO Global Analysis of QCD,*, arXiv:1302.6246 [hep-ph].

[50] P. Golonka and Z. Wąs, *PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays*, Eur. Phys. J. C **45** (2006) 97, arXiv:0506026 [hep-ph].

[51] S. Jadach, J. H. Kühn, and Z. Waąs, *TAUOLA - a library of Monte Carlo programs to simulate decays of polarized τ leptons*, Comput. Phys. Commun. **64** (1991) 275.

[52] S. Agostinelli et al., *Geant4: a simulation toolkit*, Nucl. Instr. Meth. A **506** no. 3, (2003) 250.

[53] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, R. Tanaka, J. Baglio, P. Bolzoni, R. Boughezal, O. Brein, C. Collins-Tooth, S. Dawson, S. Dean, A. Denner, S. Farrington, M. Felcini, M. Flechl, D. de Florian, S. Forte, M. Grazzini, C. Hackstein, T. Hahn, R. Harlander, T. Hartonen, S. Heinemeyer, J. Huston, A. Kalinowski, M. Krämer, F. Krauss, J. S. Lee, S. Lehti, F. Maltoni, K. Mazumdar, S. O. Moch, A. Mück, M. Mühlleitner, P. Nason, C. Neu, C. Oleari, J. Olsen, S. Palmer, F. Petriello, G. Piacquadio, A. Pilaftsis, C. T. Potter, I. Puljak, J. Qian, D. Rebuzzi, L. Reina, H. Rzehak, M. Schumacher, P. Slavich, M. Spira, F. Stöckli, R. S. Thorne, M. V. Acosta, T. Vickey, A. Vicini, D. Wackeroth, M. Warsinsky, M. Weber, G. Weiglein, C. Weydert, J. Yu, M. Zaro, and T. Zirke, *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, arXiv:1101.0593. http://arxiv.org/abs/1101.0593.

[54] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, Journal of High Energy Physics **2006** no. 05, (2006) 026–026, arXiv:0603175 [hep-ph]. http://xxx.lanl.gov/abs/hep-ph/0603175.

[55] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, et al., *Event generation with SHERPA 1.1*, JHEP **0902** (2009) 007, arXiv:0811.4622 [hep-ph].

[56] F. e. a. Cascioli, *NLO matching for ttbb production with massive b-quarks*, arXiv:1309.5912 [hep-ph].

[57] M. L. Mangano et al., *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **07** (2003) 001, arXiv:0206293 [hep-ph].

[58] P. M. Nadolsky et al., *Implications of CTEQ global analysis for collider observables*, Phys. Rev. D **78** (2008) 013004, arXiv:0802.0007 [hep-ph].

[59] G. Corcella et al., *HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, JHEP **01** (2001) 010.

[60] M. L. Mangano et al., *Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b} + n$ jets as a case study*, Nucl. Phys. **B 632** (2002) 343, arXiv:0108069 [hep-ph].

[61] K. Melnikov and F. Petriello, *Electroweak gauge boson production at hadron colliders through $\mathcal{O}(\alpha_s^2)$*, Phys. Rev. D **74** (2006) 114017, arXiv:0609070 [hep-ph].

[62] J. Campbell and R. Ellis, *An update on vector boson pair production at hadron colliders*, Phys. Rev. D **60** (1999) 113006, arXiv:9905386 [hep-ph].

[63] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **1407** (2014) 079, arXiv:1405.0301 [hep-ph].

[64] J. Alwall, P. Artoisenet, S. de Visscher, C. Duhr, R. Frederix, M. Herquet, O. Mattelaer, P. Ko, and D. Ki Hong, *New Developments in MadGraphMadEvent*, pp. , 84–89. AIP, Sept., 2008. arXiv:0809.2410. http://arxiv.org/abs/0809.2410.

[65] ATLAS Collaboration, *Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV*, Eur. Phys. J. C **71** (2011) 1577, arXiv:1012.1792 [hep-ex].

[66] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063.

[67] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. **B641** (2006) 57, arXiv:0512210v2 [hep-ph].

[68] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, Eur. Phys. J. **C72** (2012) 1896, arXiv:1111.6097 [hep-ph]. http://fastjet.fr/.

[69] C. Cojocaru et al., *Hadronic calibration of the ATLAS liquid argon end-cap calorimeter in the pseudorapidity region $1.6 < |\eta| < 1.8$ in beam tests*, Nucl. Instr. Meth. A **531** (2004) 481, arXiv:physics:0407009v1 [physics].

[70] T. Barillari et al., *Local hadronic calibration*, ATL-LARG-PUB-2009-001 (2009).

[71] ATLAS Collaboration Collaboration, ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s}$=7 TeV with the ATLAS detector*, Eur. Phys. J. C **75** no. arXiv:1406.0076. CERN-PH-EP-2013-222, (2014) 17. 100 p.

[72] ATLAS Collaboration Collaboration, *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data,*.

[73] *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment*, ATLAS-CONF-2014-004 (2014).

[74] *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data*, Tech. Rep. ATLAS-CONF-2014-032, CERN, Geneva, Jun, 2014.

[75] ATLAS Collaboration Collaboration, G. Aad et al., *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data*, Eur.Phys.J. **C74** no. 11, (2014) 3130, `arXiv:1407.3935 [hep-ex]`.

[76] *Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton-Proton Collisions recorded in 2012 at 8 TeV*, Tech. Rep. ATLAS-CONF-2013-082, CERN, Geneva, Aug, 2013.

[77] K. Kondo, *Dynamical Likelihood Method for Reconstruction of Events With Missing Momentum. 1: Method and Toy Models*, J.Phys.Soc.Jap. **57** (1988) 4126–4140.

[78] CDF Collaboration Collaboration, C. Collaboration, *Precision measurement of the top-quark mass from dilepton events at CDF II*, Phys. Rev. D **75** (2007) 031105. `http://link.aps.org/doi/10.1103/PhysRevD.75.031105`.

[79] D0 Collaboration Collaboration, D. Collaboration, *A Precision Measurement of the Mass of the Top Quark*, Nature **429** (2004) 638–642.

[80] CDF Collaboration Collaboration, C. Collaboration, *Observation of Electroweak Single Top-Quark Production*, Phys. Rev. Lett. **103** (2009) 092002. `http://link.aps.org/doi/10.1103/PhysRevLett.103.092002`.

[81] D0 Collaboration Collaboration, D. Collaboration, *Observation of Single Top-Quark Production*, Phys. Rev. Lett. **103** (2009) 092001. `http://link.aps.org/doi/10.1103/PhysRevLett.103.092001`.

[82] F. M. Pierre Artoisenet, Priscila de Aquino and O. Mattelaer, *Unravelling $t\bar{t}h$ via the matrix element method*, `arXiv:1304.6414 [hep-ph]`.

[83] ATLAS Collaboration, *Measurement of the Top-Quark Mass using the Template Method in pp Collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector,*. ATLAS-CONF-2011-033.

[84] E. P. Jerzy Neyman, *Unravelling $t\bar{t}h$ via the matrix element method*, Phil. Trans. R. Soc. Lond. A **231** (1933) 694–706 289–337.

[85] M. Galassi et al., *GNU Scientific Library Reference Manual (3rd Ed.)*, ISBN 0954612078 (2009).

[86] G. P. Lepage, *A New Algorithm for Adaptive Multidimensional Integration*, J.Comput.Phys. **27** (1978) 192.

[87] J. Butterworth, G. Dissertori, S. Dittmaier, D. de Florian, N. Glover, et al., *Les Houches 2013: Physics at TeV Colliders: Standard Model Working Group Report*, `arXiv:1405.1067 [hep-ph]`.

[88] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl.Instrum.Meth. **A389** (1997) 81–86.

[89] J. Alwall et al., *MadGraph/MadEvent v4: the new web generation*, JHEP **09** (2007) 028, arXiv:0706.2334 [hep-ph].

[90] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1, arXiv:1207.7214v2 [hep-ex].

[91] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30, arXiv:1207.7235v1 [hep-ex].

[92] W. H. Press and G. R. Farrar, *Recursive Stratified Sampling for Multidimensional Monte Carlo Integration*, Computers in Physics **4** no. 2, (1990).

[93] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying to $b\bar{b}$ in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector at the LHC*, ATLAS-COM-CONF-2014-004. ATLAS-COM-CONF-2014-004.

[94] Phi-T GmbH, *NeuroBayes package*, http://neurobayes.phi-t.de/.

[95] R. Fletcher, *Practical Methods of Optimization; (2Nd Ed.)*. Wiley-Interscience, New York, NY, USA, 1987.

[96] Y. Freund and R. E. Schapire, *A Decision-theoretic Generalization of On-line Learning and an Application to Boosting*, J. Comput. Syst. Sci. **55** no. 1, (1997) 119–139. http://dx.doi.org/10.1006/jcss.1997.1504.

[97] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov, D. Dannheim, K. Danielowski, S. Henrot-Versille, M. Jachowski, K. Kraszewski, A. Krasznahorkay, Jr., M. Kruk, Y. Mahalalel, R. Ospanov, X. Prudent, A. Robert, D. Schouten, F. Tegenfeldt, A. Voigt, K. Voss, M. Wolter, and A. Zemla, *TMVA - Toolkit for Multivariate Data Analysis*, ArXiv Physics e-prints (2007), physics/0703039.

[98] ATLAS Collaboration, G. Aad et al., *Improved luminosity determination in pp collisions at sqrt(s) = 7 TeV using the ATLAS detector at the LHC*, Eur.Phys.J. **C73** no. 8, (2013) 2518, arXiv:1302.4393 [hep-ex].

[99] J. M. Campbell and R. K. Ellis, *$t\bar{t}W$ production and decay at NLO*, arXiv:1204.5678 [hep-ph].

[100] M. V. Garzelli, A. Kardos, C. G. Papadopoulos, and Z. Trocsanyi, *$t\bar{t}W$ and $t\bar{t}Z$ Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects*, JHEP **1211** (2012) 056, arXiv:1208.2665 [hep-ph].

[101] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. D **83** (2011) 091503, arXiv:1103.2792 [hep-ph].

[102] N. Kidonakis, *Next-to-next-to-leading logarithm resummation for s-channel single top quark production*, Phys. Rev. D **81** (2010) 054028.

[103] R. D. B. et al., *Parton distributions with LHC data*, Nucl. Phys. **B867** (2013) 244, arXiv:1207.1303 [hep-ph].

[104] M. B. et al., *The PDF4LHC Working Group Interim Recommendations*, arXiv:1101.0538 [hep-ph].

[105] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP **06** (2002) 029, arXiv:0204244 [hep-ph].

[106] S. Frixione, E. Laenen, P. Motylinski, and B. R. Webber, *Single-top production in MC@NLO*, JHEP **03** (2006) 092, arXiv:0512250 [hep-ph].

[107] S. Frixione, E. Laenen, P. Motylinski, C. White, and B. R. Webber, *Single-top hadroproduction in association with a W boson*, JHEP **07** (2008) 029, arXiv:0805.3067 [hep-ph].

[108] W. Verkerke and D. Kirkby, *RooFit Users Manual v2.91*,. http://rootfit.sourceforge.net.

[109] T. Junk, *Confidence level computation for combining searches with small statistics*, Nucl. Instr. Meth. A **434** (1999) 435, arXiv:9902006 [hep-ex].

[110] A. L. Read, *Presentation of search results: the CL_s technique*, J. Phys. G **28** (2002) 2693.

[111] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur.Phys.J. **C71** (2011) 1554, arXiv:1007.1727 [physics.data-an].