Observation of Four-Top Quark Production and Measurement of Off-shell Higgs Boson Interactions with Top Quarks with ATLAS

By

Benjamin Ryan Roberts

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Haichen Wang, Chair
Professor Yury Kolomensky
Professor Karl Van Bibber

Summer 2025

Observation of Four-Top Quark Production and Measurement of Off-shell Higgs Boson Interactions with Top Quarks with ATLAS

Benjamin Ryan Roberts

Abstract

Observation of Four-Top Quark Production and Measurement of Off-shell Higgs Boson Interactions with Top Quarks with ATLAS

by

Benjamin Ryan Roberts

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Haichen Wang, Chair

The ATLAS experiment at CERN measures proton–proton collisions produced by the Large Hadron Collider at energies up to 13.6 TeV. Analysis of the Run 2 dataset, which totals an integrated luminosity of 140 fb^{-1} at a center of mass energy of 13 TeV, has resulted in unprecedented precision in measurements of rare processes involving the top quark and the Higgs boson, the heaviest known fundamental particles. Measuring these processes provides direct probes of the top Yukawa coupling and tests of the Standard Model fermion mass generation mechanism. The total decay width of the Higgs boson can also be accessed by comparing the rates of on- and off-shell Higgs boson production.

The simultaneous production of four top quarks ($t\bar{t}t\bar{t}$) is a rare process sensitive to possible contributions from heavy new particles. Additionally, $t\bar{t}t\bar{t}$ production is sensitive to the top Yukawa coupling through contributions from off-shell Higgs bosons. Therefore, $t\bar{t}t\bar{t}$ production may be used to measure the top Yukawa coupling independently of the Higgs boson's total width.

The first observation of $t\bar{t}t\bar{t}$ production in ATLAS established the $t\bar{t}t\bar{t}$ process with a statistical significance of 6.1σ. The observation of $t\bar{t}t\bar{t}$ production was performed in the same-sign dilepton and multilepton final states and uses a graph neural network to separate the small $t\bar{t}t\bar{t}$ signal from background processes. The measured value of the $t\bar{t}t\bar{t}$ production cross-section was $\sigma = 22^{+6.6}_{-5.5}$ fb, which is 1.9 times the Standard Model prediction. Interpretations of the observed $t\bar{t}t\bar{t}$ signal were used to set limits on the coefficients of four-fermion SMEFT operators and the Higgs oblique parameter. Further interpretations set limits on the top Yukawa coupling and its CP properties. The top Yukawa coupling was constrained to $\kappa_t < 1.9$ at the 95% CL when considering the variation of the $t\bar{t}H$ background and $\kappa_t < 2.3$ when considering only the effect on $t\bar{t}t\bar{t}$ production. Additional limits were set on the cross-section of $t\bar{t}t$ production, which remains unobserved.

Following the observation of $t\bar{t}t\bar{t}$ production, the results were combined with measurements of on-shell Higgs boson production and decay to produce the first limits on the total width of the Higgs boson Γ_h based on measurements of Higgs boson couplings to fermions. The combination placed a limit of $\Gamma_h < 450$ MeV at the 95% CL when loop-induced interactions of the Higgs boson were profiled and $\Gamma_h < 160$ MeV when they were parameterized as functions of tree-level couplings to Standard Model Particles.

"It's the same old song." —The Four Tops

"Not all those who wander are lost." —J.R.R. Tolkien

To Mom and Dad.

Contents

List of Figures

List of Tables

Acknowledgments

This thesis represents, at most, an incremental contribution to human scientific knowledge. As ever, it builds atop the efforts of many great scientists who came before, and to them we must always be grateful. The large collaborations and long timelines of experimental particle physics emphasize research as a collective endeavor. Thanks to my past and present colleagues in the ATLAS collaboration who have designed, constructed, and operated a stunningly complex detector and the infrastructure to analyze the collected data. In the last six years, I have been able to work on many projects with colleagues from across ATLAS, many of whom have provided invaluable insight and feedback and who have contributed to an atmosphere of critical inquiry and collegiality within the collaboration.

Thanks to all the students, postdocs, staff scientists, and faculty members in the LBNL ATLAS group whom I have had the pleasure to know in the last six years. Their support and mentorship on matters administrative, technical, and scientific have been paramount in the pursuit of my degree and my education as a physicist. I have greatly enjoyed the opportunity to learn from so many experts, always willing to share what they know. I am grateful to my advisor, Professor Haichen Wang, who personally taught me not only physics and data analysis, but also how to be a professional scientist. Thank you for the many thoughtful explanations, constant feedback, and helpful advice, and thank you for giving me the foundation and eventually the space and opportunity to become an independent researcher. Haichen is the latest of numerous teachers and mentors who have encouraged and challenged me throughout my 23 years of education. I am forever grateful to all of them.

The Ph.D. is known to be an unusual and even difficult period in one's life. My Ph.D. studies have been marked by much of the external turmoil, life transitions, self-doubt, and research-related challenges that create this difficulty. That I have finished intact and improved is entirely due to the endless support of many friends who have been there when needed and pushed me to improve myself. Friends made at Cornell, in Berkeley, and during my time at CERN have all been part of this journey. Thank you to everyone who has gone camping, hiking, climbing, running, cycling, or skiing with me; who has played board games, softball, soccer, or cricket; who has shared a back-row burrito, Monday meal, or R1 beer. Together, you have added much more joy to my Ph.D. experience than there has been pain.

Thank you to Ashwin, Burhan, and Antonio, my roommates on Francisco Street, for endless laughs and precious memories. It was always a pleasure to come home to you. Thanks also to Ashwin and Kehang, whose dependable friendship in Berkeley and at CERN I have relied upon many times. You both care deeply about the people around you, and your passion for physics is a constant reminder of the great beauty and deep questions that brought us to this field. To pursue a Ph.D. and be able to study fundamental physics has required good fortune in many forms. In no way am I more fortunate than in my loving family. My parents have always supported me in every endeavor, academic and otherwise, have always valued learning and knowledge, and have always set a tremendous example in work ethic, curiosity, and kindness. Thanks, Mom and Dad.

Chapter 1

Introduction

Direct studies of the interactions of sub-atomic particles at high energies are only possible by intersecting energetic beams of particles and measuring the products of the resulting collisions. The Large Hadron Collider (LHC) is the largest and highest energy particle collider yet constructed and is capable of accelerating protons to energies of 6.8 TeV, over 7,000 times their rest energy, and colliding them at instantaneous rates in excess of four billion collisions per second. The four large experimental collaborations at the LHC, ALICE, ATLAS, CMS, and LHCb, have constructed large particle detectors to measure the outcomes of these proton–proton collisions and study the properties of the particles they produce. Together, the experiments use the shared platform of proton–proton collisions to address a wide range of topics in nuclear and particle physics. The ATLAS and CMS detectors are both designed as general-purpose detectors that capture nearly all energetic particles produced in the collisions and are capable of observing collisions at the highest rates to maximize the study of rare processes. Operating since 2008, ATLAS and CMS have performed world-leading precision measurements of the production, interactions, and decays of the heaviest previously discovered particles, the top quark and the W and Z bosons, and they have also performed numerous searches for previously undiscovered particles and interactions. In 2012, the ATLAS and CMS Collaborations announced the discovery of a new particle, later confirmed to have properties similar to those of the Higgs boson predicted in the Standard Model of Particle Physics. The Higgs boson discovery has cemented the Standard Model as the leading description of nature in high-energy physics (HEP), providing precise predictions of strong and electroweak interactions that have been verified by countless experiments.

Run 2 of the LHC lasted from 2015 to 2018 and delivered a large dataset at an increased center of mass energy $\sqrt{s} = 13$ TeV. Together with significant advances in data analysis techniques, this dataset has significantly extended the reach of searches for new physics outside of the Standard Model and the precision of measurements of Standard Model processes, including a number of rare production and decay modes of the Higgs boson. This marks the beginning of an era of precision measurements of the Higgs boson's properties, especially through the continued observation of increasingly rare processes. Measurements of the interaction of the Higgs boson with the top quark test the mass generation mechanism for the

fermions, one of the central theoretical motivations for the introduction of the Higgs boson. Direct measurements of this interaction have been performed in the observation of Higgs boson production in association with top quark pairs, known as $t\bar{t}H$ production. Alternative measurements based on interactions of an off-shell Higgs boson with top quarks evade assumptions about the properties of the Higgs boson decay, including its total decay width, which cannot be directly measured at the LHC. Such a measurement can be performed by studying the simultaneous production of four top quarks, $t\bar{t}t\bar{t}$ which can be mediated by the top-Higgs interaction.

This thesis presents the observation of $t\bar{t}t\bar{t}$ production and subsequent interpretations of the observed $t\bar{t}t\bar{t}$ signal in terms of possible modifications from physics beyond the Standard Model. A measurement of the top-Higgs coupling in $t\bar{t}t\bar{t}$ production is combined with measurements of Higgs boson production and decays to produce constraints on the total decay width of the Higgs boson. The $t\bar{t}t\bar{t}$ observation is done in final states with same-sign light leptons and uses a novel graph neural network algorithm to reject collision events from background processes.

Chapter 2 reviews the theoretical basis of the Standard Model and the Higgs boson, the phenomenology of hadron collisions, and the experimental and phenomenological status of Higgs boson and top quark physics at the LHC. Chapter 3 describes the Large Hadron Collider and the ATLAS detector. Chapter 4 describes the techniques used to reconstruct particles produced in proton collisions from their detector signatures and the performance of these techniques in ATLAS. Chapter 5 describes the Run 2 collision dataset and the large dataset of simulated collision events used to perform the $t\bar{t}t\bar{t}$ measurement. Chapter 6 details the statistical framework and techniques used for measurement, limit setting, and evaluation of expected sensitivity. The parameterizations used for re-interpretations of the $t\bar{t}t\bar{t}$ cross-section measurement are also presented. Chapter 7 provides background in machine learning and especially graph neural networks and transformers, which are used in the $t\bar{t}t\bar{t}$ observation and several results found in the appendices. Chapter 8 describes the same-sign dilepton and multilepton event selection and the estimation and modeling of background processes in this region. Chapter 9 describes the development and training of the graph neural network discriminant used to reject background processes and enhance sensitivity to $t\bar{t}t\bar{t}$ production. Chapter 10 describes the results of the $t\bar{t}t\bar{t}$ observation, validation of the observed signal, and interpretations in terms of the top-Higgs coupling and effective field theory parameters. Chapter 11 describes the combination of the $t\bar{t}t\bar{t}$ measurement with measurements of Higgs boson production and decays to produce constraints on the total decay width of the Higgs boson. Appendices A to C describe other work to which the author has contributed, but which does not directly pertain to the observation of $t\bar{t}t\bar{t}$ production. This work includes additional measurements of the top Yukawa coupling, studies of machine learning algorithms, and preparations for future measurements with upgrades to the detector and reconstruction algorithms.

Chapter 2

Theoretical Considerations

2.1 The Standard Model of Particle Physics

Particle physics is concerned with the behavior of individual or small groups of subatomic particles. The main theoretical tool for describing these particles and their interactions is quantum field theory (QFT). A result of the combined early twentieth-century developments of relativity and quantum mechanics, quantum field theory provides a fully quantum mechanical and fully relativistic description of subatomic particles as excited states of fields, which constitute the fundamental degrees of freedom. Quantum mechanically, the fields are operators defined as functions of the spacetime coordinates. The treatment here follows the excellent book by Peskin and Schroeder [1] and attempts to use the conventions adopted there unless otherwise noted. In particle physics, the choice of fields to add to a theory is limited by the imposition of a $3+1$ dimensional spacetime with a global Poincaré symmetry. This work describes quantum field theories in the Lagrangian formalism and ignores those that cannot be expressed by a Lagrangian. Poincaré symmetry demands that the Lagrangian be invariant under boosts, translations, and rotations. The requirement of translational invariance implies that the Lagrangian density should be a function only of the fields and their derivatives, while rotational and Lorentz invariance requirements imply that the fields should be representations of the Lorentz group and that the Lagrangian must be composed of Lorentz invariant combinations of those fields. The representations are labeled with half-integers $m/2$ called spin and correspond to particles with intrinsic angular momentum $\hbar m/2$.

The simplest fields to introduce are scalar fields, which transform identically under the Lorentz group and are thus spin-0. Spin-1/2 fields can be represented by solutions to the Weyl equations $i(\partial_0 - \vec{\sigma}\cdot\vec{\nabla})\psi_L = 0$, $i(\partial_0 + \vec{\sigma}\cdot\vec{\nabla})\psi_R = 0$, the solutions of which are two-component spinors. The introduction of continuous local symmetries into a quantum field theory requires the addition of a dynamical field in the spin-1 representation of the Lorentz group, known as a gauge field. This introduces a conserved charge carried by particles interacting with the gauge field. When the symmetry group of the local symmetry, called

the gauge group, is non-Abelian, the resulting theory is known as a Yang-Mills theory, and the associated gauge bosons become charged under the gauge interaction [2]. A typical case is a gauge field A^i_μ interacting with a Dirac fermion ψ^i transforming in the fundamental representation with the gauge index i and the Dirac index suppressed. The Lagrangian of such a theory is written in eq. (2.1). Summation over repeated indices is implied.

$$\begin{aligned} \mathcal{L} &= \bar{\psi}_i(i\not{D})\psi_i - \frac{1}{4}\left(F^i_{\mu\nu}\right)^2 - m\bar{\psi}_i\psi_i \\ \text{where } F^i_{\mu\nu} &= \partial_\mu A^i_\nu - \partial_\nu A^i_\mu + g f^{ijk} A^j_\mu A^j_\nu \\ \text{and } \not{D} &= \gamma^\mu\left(\partial_\mu - igA^i_\mu t^i\right) \end{aligned} \tag{2.1}$$

The structure function f^{ijk} is defined by the commutation relations of the group's generators and t^i are the generators in the fundamental representation.

Symmetries and Particle Content

The Standard Model of Particle Physics (SM) is the crowning achievement of particle physics in the twentieth century. The SM is a quantum field theory that explains nearly all observed fundamental particles, including their properties, interactions, and combinations into composite particles. Indeed, the SM can be considered to explain the majority of physical phenomena and has accurately predicted the outcomes of countless experiments; notable exceptions are described in section 2.3. The SM has a gauge group of $SU(3)_C \times SU(2)_L \times U(1)_Y$, where the subscripts refer to the charge associated with each gauge group described in table 2.1. The $SU(3)_C$ gauge group acting on the color charge describes the strong nuclear force, and the resulting theory of strongly interacting quarks and gluons is known as quantum chromodynamics (QCD). The $SU(2)_L \times U(1)_Y$ gauge group describes the electromagnetic and weak nuclear (together electroweak or EW) forces. The fermionic content of the SM is divided into quarks and leptons, including a left-handed $SU(2)$ doublet of each and accompanying right-handed $SU(2)$ singlets. This structure is repeated three times in groups known as generations. The final component of the SM is a scalar $SU(2)$ doublet known as the Higgs field. The representations and quantum numbers of all SM fields are summarized in table 2.2.

Group	Field	Charge	Coupling
$SU(3)$	G^μ	Color, C	g_s
$SU(1)$	W^μ	Weak Isospin, L	g
$U(1)$	B^μ	Hypercharge, Y	g'

Table 2.1: Gauge groups of the Standard Model and standard notation for the fields, charges, and coupling associated with each.

Field	Spin	Handedness	$SU(3)_C$ rep.	$SU(2)_L$ rep.	$U(1)_Y$ Hypercharge
ϕ	0		**1**	**2**	$1/2$
Q^i	$1/2$	L	**3**	**2**	$1/6$
L^i	$1/2$	L	**1**	**2**	$-1/2$
u^i_R	$1/2$	R	**3**	**1**	$2/3$
d^i_R	$1/2$	R	**3**	**1**	$-1/3$
l^i_R	$1/2$	R	**1**	**1**	-1
B	1		**1**	**1**	0
W	1		**1**	**3**	0
G	1		**8**	**1**	0

Table 2.2: Fields in the Standard Model before electroweak symmetry breaking. The index $i = 1..3$ runs over the fermion generations. Color, weak charge, Lorentz, and Dirac indices are omitted. The right-handed fermions are written in their charge conjugated (i.e. left-handed) forms.

The Standard Model is described by a Lagrangian density that includes every renormalizable (mass dimension less than or equal to 4) term allowed by the symmetries of the theory. A single term, $\frac{\theta g_s^2}{64\pi^2}\epsilon^{\mu\nu\lambda\sigma}F^a_{\mu\nu}F^a_{\lambda\sigma}$, is excluded from this prescription. Known as the θ term, this is a total derivative, but nonetheless would produce non-perturbative physical effects violating charge-parity (CP) symmetry in the strong force. Experimental constraints on baryon electric dipole moments strongly constrain this term to be near 0. This is known as the Strong CP Problem [3]. The full Lagrangian is rendered in eq. (2.2) with the gauge indices suppressed, the index i running over the different gauge fields, the index j running over the fermion multiplets, the indices a and b for the fermion generations, and the gauge covariant derivative defined as in eq. (2.1), but with the inclusion of all gauge fields.

$$\begin{aligned}\mathcal{L} = D_\mu\phi^\dagger D^\mu\phi + \bar{\psi}_j(i\not{D})\psi_j - \frac{1}{4}F^i_{\mu\nu}F^{i\mu\nu} \\ -\mu^2\phi^\dagger\phi + \lambda\left(\phi^\dagger\phi\right)^2 - \left(\lambda^{ab}_d\bar{Q}^a_L\phi d^b_R - \lambda^{ab}_u\bar{Q}^a_L\sigma_2\phi u^b_R - \lambda^{ab}_e\bar{E}^a_L\phi e^b_R + \text{h.c.}\right)\end{aligned} \tag{2.2}$$

The Lagrangian includes the kinematic terms, which also induce the gauge interactions, the scalar potential, which depends only on the Higgs field, and the Yukawa terms with couplings λ^{ab}_f. These terms consist of a left-handed $SU(2)$ doublet fermion with spinor and color indices contracted with a right-handed $SU(2)$ singlet fermion and with the $SU(2)$ indices contracted with the scalar Higgs field. The terms shown are the only possibilities that respect $U(1)_Y$. There is only one term for the leptons because there is no right-handed neutrino ν^i_R, which would be neutral under all other SM interactions.

2.2 Electroweak Symmetry Breaking

This section describes the mechanism and consequences of electroweak symmetry breaking in the Standard Model based on the ideas of Higgs, Brout, Englert, and others [4, 5, 6, 7, 8, 9]. The Higgs mechanism is introduced to explain the origin of massive vector gauge bosons and preserve unitarity via spontaneous breaking of the electroweak gauge symmetry [10]. In the SM, this proceeds through the scalar potential of the Higgs field and results in a single massive scalar boson, the Higgs boson, and gives masses to three electroweak gauge bosons, the quarks, and the charged leptons.

Vacuum Expectation Value and the Higgs Boson

The potential of the Higgs field in the Standard Model is described by eq. (2.3) and has a set of global minima when eq. (2.4) is satisfied.

$$V(\phi) = -\mu^2\phi^\dagger\phi + \lambda\left(\phi^\dagger\phi\right)^2 \tag{2.3}$$

$$|\phi|^2 = \phi^\dagger\phi = \frac{v^2}{2} = \frac{\mu^2}{2\lambda} \tag{2.4}$$

Perturbations from the vacuum must be described by an expansion around this minimum. Choosing an $SU(2)_L$ gauge transformation to rotate the field value to be a single real-valued component completely fixes the gauge. The resulting minimum is at

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

and the expansion of the field around this minimum is

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

This choice is known as the unitary gauge and is used throughout the text. The upshot of this is that the gauge symmetry is nonlinearly realized as a result of this spontaneous symmetry breaking. This is known as electroweak symmetry breaking (EWSB). The effect on the Higgs potential is

$$V(\phi) = -\mu^2\phi^\dagger\phi + \lambda\left(\phi^\dagger\phi\right)^2 = -\frac{1}{2}\mu^2v^2 + \frac{1}{4}\lambda v^4 + \left(\frac{3}{2}\lambda v^2 - \frac{1}{2}\mu^2\right)h^2 + \lambda v h^3 + \frac{1}{4}\lambda h^4. \tag{2.5}$$

Ignoring the constant terms, which do not affect the theory's dynamics outside of gravitational effects, the results are a mass term $m_h^2 = 2\mu^2$ and two self-interaction terms $\lambda v h^3 + \frac{1}{4}\lambda h^4$ for the remaining scalar boson, which is known as the Higgs boson.

Gauge Boson Masses and Interactions

The effects of symmetry breaking on the gauge bosons appear from the kinetic term of the Higgs Lagrangian

$$\begin{aligned}\mathcal{L}_{kin} = (D_\mu\phi)^\dagger D^\mu\phi &= \frac{1}{2}\left|\left(\partial_\mu - igW_\mu^a\frac{\sigma^a}{2} - i\frac{1}{2}g'B_\mu\right)\begin{pmatrix}0\\v+h\end{pmatrix}\right|^2 \\ &= \frac{1}{2}\partial_\mu h\partial^\mu h + \frac{(v+h)^2}{2}\left[\left|-igW_\mu^a\frac{\sigma^a}{2} - i\frac{1}{2}g'B_\mu\right|^2\right]_{22} \\ = \frac{1}{2}\partial_\mu h\partial^\mu h + \frac{(v+h)^2}{2}&\left[\frac{g^2}{4}W_\mu^1W^{1\mu} + \frac{g^2}{4}W_\mu^2W^{2\mu} + \frac{1}{4}\left(gW_\mu^3 - g'B_\mu\right)\left(gW^{3\mu} - g'B^\mu\right)\right].\end{aligned} \tag{2.6}$$

The choice of the unitary gauge ensures that all single derivative terms are removed, and the v^2 terms introduce masses for the gauge bosons. The last term is simplified by a change of basis to rotate the fields W_μ^3 and B_μ into $Z_\mu = (gW_\mu^3 - g'B_\mu)/(\sqrt{g^2+g'^2})$ and $A_\mu = (g'W_\mu^3 + gB_\mu)/(\sqrt{g^2+g'^2})$, resulting in independent, correctly normalized kinematic terms: $(W_{\mu\nu}^3)^2+(B_{\mu\nu})^2 = (Z_{\mu\nu})^2+(A_{\mu\nu})^2$. The three massless gauge bosons from $SU(2)_L$ and one from $U(1)_Y$ are transformed into three massive bosons, $W^\pm$ and Z^0, and one massless one, the familiar photon γ of electromagnetism. The photon interacts with electric charge $Q = Y + T^3$, where T^3 is the weak isospin, with coupling strength $gg'/\sqrt{g^2+g'^2}$. One effect of this electroweak mixing is that the heavy $W^\pm$ bosons carry electric charge even though in the theory with unbroken symmetry, there is no interaction between W^i and B. Similarly, the Z boson gains a coupling to right-handed fermions while the $W^\pm$ bosons still only couple to left-handed fermions. The gauge boson masses are proportional to v with a proportionality constant determined by the gauge couplings of $SU(2)_L \times U(1)_Y$.

$$m_W = \frac{gv}{2} \text{ and } m_Z = \frac{v}{2}\sqrt{g^2+g'^2} = \frac{\sqrt{g^2+g'^2}}{g}m_W = \frac{m_W}{\cos\theta_W}$$

The weak mixing angle θ_W describes the mixing between $SU(2)_L$ and $U(1)_Y$. Equation (2.6) also contains interaction terms between the Higgs boson and the heavy gauge bosons W and Z, but not the massless photon.

Fermion Masses and Interactions

Electroweak symmetry breaking also introduces mass terms for the fermions. The fermions are massless in the unbroken Lagrangian eq. (2.2) because the left- and right-handed fermions transform differently under $SU(2)_L$ and are therefore incompatible with a Dirac mass term. The fermion mass generation originates from the Yukawa interaction terms, which are greatly simplified by diagonalizing the matrices λ_f^{ab} from eq. (2.2). This diagonalization proceeds via a field redefinition, but requires that the upper and lower components of Q_L^a be rotated

independently to diagonalize both λ_u^{ab} and λ_d^{ab}. Thus, the matrices used in the redefinition will appear in off-diagonal terms wherever Q_L^i is contracted with another field or with a non-diagonal matrix in between. The former happens only in the Yukawa coupling, but any mixture is eliminated by the unitary gauge (this also allows the diagonalization to work), and the latter occurs only in the interaction term with the $W^\pm$ bosons. The result is that this interaction becomes

$$\bar{Q}_L^a(\not{W}^1\sigma_1 + \not{W}^2\sigma_2)Q_L^a = W_\mu^+ \bar{u}_L^a \gamma^\mu d_L^a + W_\mu^- \bar{d}_L^a \gamma^\mu u_L^a \to V_{ab} W_\mu^+ \bar{u}_L^a \gamma^\mu d_L^b + V_{ab}^\dagger W_\mu^- \bar{d}_L^a \gamma^\mu u_L^a,$$

where the $SU(2)_L$ doublet Q_L is separated into up and down type left-handed quarks u_L and d_L. The matrix V_{ab} with indices running over generations is the Cabibbo-Kobayashi-Maskawa (CKM) matrix and introduces mixing between quark generations in charged-current weak interactions. The Yukawa terms are now separated by generation after spontaneous symmetry breaking. For the first generation, these terms read

$$\begin{aligned}\mathcal{L}_{yuk} &= -\frac{\lambda_d}{\sqrt{2}}\begin{pmatrix}\bar{u}_L \\ \bar{d}_L\end{pmatrix}\cdot\begin{pmatrix}0 \\ v+h\end{pmatrix}d_R - \frac{\lambda_u}{\sqrt{2}}\begin{pmatrix}\bar{u}_L \\ \bar{d}_L\end{pmatrix}\cdot\begin{pmatrix}v+h \\ 0\end{pmatrix}u_R - \frac{\lambda_d}{\sqrt{2}}\begin{pmatrix}\bar{e}_L \\ \bar{\nu}_e\end{pmatrix}\cdot\begin{pmatrix}0 \\ v+h\end{pmatrix}e_R + \text{h.c.} \\ &= -v\frac{\lambda_d}{\sqrt{2}}\bar{d}_L d_R - v\frac{\lambda_u}{\sqrt{2}}\bar{u}_L u_R - v\frac{\lambda_e}{\sqrt{2}}\bar{e}_L e_R - h\frac{\lambda_d}{\sqrt{2}}\bar{d}_L d_R - h\frac{\lambda_u}{\sqrt{2}}\bar{u}_L u_R - h\frac{\lambda_e}{\sqrt{2}}\bar{e}_L e_R + \text{h.c.},\end{aligned} \tag{2.7}$$

with similar terms for the second and third generations. After EWSB, the quarks and charged leptons acquire masses and couplings to the Higgs boson with strengths proportional to the masses. The coupling strengths λ_i are known as the Yukawa couplings and result in masses $m_i = v\lambda_i/\sqrt{2}$.

The remaining particles after EWSB are categorized and named in the mass basis in fig. 2.1. It is now easy to count the free parameters in the SM and directly relate them to measurable quantities. There are three gauge couplings, four real parameters of the CKM matrix, nine fermion masses from Yukawa couplings, the Higgs vacuum expectation value (corresponding to the W and Z masses), and the Higgs boson mass for a total of 18 free parameters. A convenient basis following the historical development of the SM and emphasizing the low energy quantities that are best measured is to use the fine-structure constant α and the Weinberg angle θ_W for the $SU(2) \times U(1)$ gauge couplings. The CKM matrix can be parameterized by the magnitudes of three off-diagonal elements $|V_{us}|$, $|V_{ub}|$, and $|V_{cb}|$, or equivalently three mixing angles θ_{12}, θ_{13}, and θ_{23} and a single CP mixing angle δ. The best measured values of each are shown in table 2.3 together with the measured physical quantity from which the parameter is derived. Many of these quantities receive quantum corrections that result in their values running as a function of the energy scale. This does not affect the parameter counting, but requires that the determination of the parameters be associated with a particular scale.

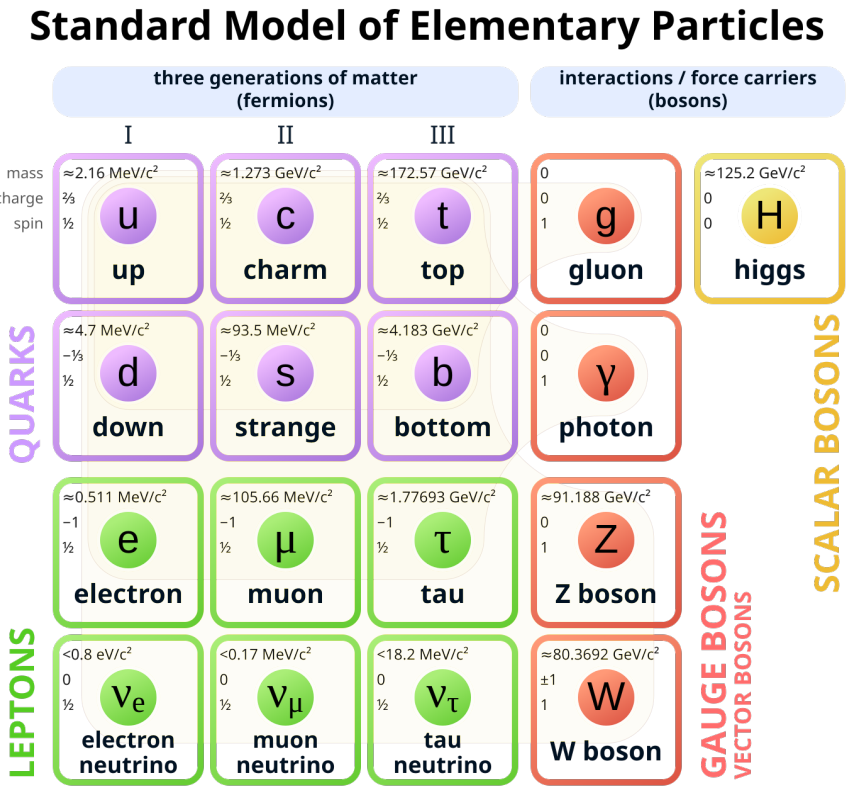


Figure 2.1: Particle content of the Standard Model of particle physics in the mass basis after electroweak symmetry breaking. The electric charge, spin, and mass of each particle are shown.

2.3 The Case for BSM Physics and Effective Field Theory

Many open problems in particle physics, if they are to be solved within the framework of QFT, call for the introduction of fields or interactions that are not a part of the Standard Model, known sometimes as New Physics (NP), but better and more often called Beyond the Standard Model (BSM) physics. These problems may stem from cosmological observations, such as dark matter, dark energy, and baryogenesis, or from particle physics observations such as neutrino masses, flavor physics anomalies, and the anomalous magnetic moment of the muon. Others are theoretical concerns like the hierarchy problem, the flavor problem, the strong CP problem, and the stability of the vacuum state of the SM. Many theoretical solutions to these problems have been proposed, ranging from simple modifications adding a new field to the SM Lagrangian to paradigmatic shifts involving entirely new sectors with multitudes of as-yet undiscovered states. While the exact nature of BSM physics,

Parameter	Value		Technique
α^{-1}	137.035 999 084(21)		
$\sin\theta_W$	0.231 29(4)		
α_s	0.1180(9)		Various. See [11].
v			Fermi coupling constant from μ decay
m_h	125.20 (11)	GeV	
m_e	0.510 998 950 00(15)	MeV	
m_μ	105.658 3755(22)	MeV	
m_τ	1776.93(9)	MeV	
m_u	2.16(7)	MeV	Pion masses
m_c	1.2730(46)	GeV	$e^+e^- \to$ Charmonium at threshold
m_t	172.57(29)	GeV	Direct measurement at hadron colliders
m_d	4.70(7)	MeV	Pion masses
m_s	93.5(8)	MeV	Kaon masses
m_b	4.183(7)	GeV	$e^+e^- \to$ Bottomonium at threshold
$\sin\theta_{12}$	0.225 01(68)		
$\sin\theta_{13}$	0.003 732(90)		
$\sin\theta_{23}$	0.041 83(79)		
δ	1.147(26)		

Table 2.3: Measured values of the 18 free parameters of the Standard Model. Digits in parentheses represent the uncertainty following the PDG convention. All values are in natural units $\hbar = c = 1$. For details such as the renormalization scheme and scale, see the PDG review tables [11].

including the energy scale at which it becomes observable and the complexity of its underlying structure, is not clear, it is clear that the SM is not a complete theory of nature.

The Higgs boson occupies a central role in the SM and naturally enters many theories of BSM physics as well. The Higgs sector in the SM consists only of a single $SU(2)_L$ doublet and may be thought of as a minimal way to break the electroweak symmetry; there is no strong *a priori* preference for the SM setup beyond its simplicity, though it conveniently restricts the possibilities for unobserved flavor changing and CP-violating effects. On the contrary, theories of electroweak symmetry breaking featuring more complicated scalar sectors or other dynamics often have their own theoretically favorable features. Theories in which the Higgs mass is protected by a new symmetry, including supersymmetric or composite Higgs models, necessarily introduce new states above the electroweak scale and avoid the hierarchy problem in which the Higgs mass is quadratically sensitive to new physics at high energies [12]. Additionally, alterations to the Higgs sector may allow the phase transition in the early

universe [13] to be a first-order transition. A first-order phase transition would create an out-of-equilibrium state in the early universe, fulfilling one of Sakharov's [14] conditions for baryogenesis to explain the observed excess of matter over antimatter in the universe. Another of Sakharov's conditions, violation of CP symmetry, is also linked to the Higgs sector because all observed sources of CP-violation are associated with the CKM matrix, which results from diagonalization of the quark Yukawa matrices. Additional CP-violation is possible in theories with expanded Higgs sectors via mixing between scalar and pseudoscalar bosons. The Higgs boson may also have couplings to a dark sector, including a candidate for dark matter. If the dark sector contains light particles, the Higgs boson may have a small decay width into BSM particles.

This thesis concerns measurements of the top Yukawa coupling, a fundamental parameter in the SM to which many theoretical calculations, including the vacuum stability of the universe [15], are sensitive. In BSM theories, the top Yukawa coupling may play a significant role in EWSB, and the coupling may be reduced in theories in which the top quark mass receives contributions from vacuum expectation values of additional scalar fields. Furthermore, the top-Higgs interaction is a possible source of tree-level CP-violation in the Higgs sector, and measurements of off-shell Higgs boson interactions with top quarks at the LHC allow indirect constraints on the total Higgs boson's total width and thus possible decays into BSM particles. The wealth of new physics scenarios with observable BSM effects in the Higgs sector motivates a thorough experimental program to verify the predictions for the SM Higgs boson. Experimentally and theoretically, the Higgs coupling to top quarks is afforded particular interest due to the large top quark mass and coupling to the SM Higgs boson.

Many of these issues invite solutions with new particles at or below the collision energy of the LHC, which may be produced and observed directly. However, many may also be solved with new physics entering at higher scales. The effective field theory (EFT) approach [16] allows the enumeration and parameterization of possible effects of heavy new physics through the introduction of non-renormalizable operators into the Lagrangian. These operators may be generated by integrating out massive degrees of freedom in a UV-complete theory. A full enumeration of the possible effects depends on the assumptions about new physics, and several popular frameworks are relevant in the Higgs sector. The most general is the Higgs Effective Field Theory (HEFT) [17], which considers all possible dimension six operators that can be added to the SM Lagrangian after EWSB. Adding the assumption that the Higgs mechanism proceeds with an $SU(2)_L$ doublet as in the SM, this reduces to the Standard Model Effective Field Theory (SMEFT) [18, 19], which considers all possible dimension 6, and more recently dimension 8 [20], operators that can be added before EWSB. The EFT approach introduces a large number of parameters and requires measurements to be made below the energy scale of BSM states, which has encouraged the widespread use of two other ways of parameterizing experimental Higgs results. The first is the Simplified Template Cross Section (STXS) [21], which divides the differential Higgs boson production phase space into a number of bins with emphasis on separating regions with additional sensitivity to BSM effects and ensuring that the cross-section in each bin is accessible both experimentally and theoretically. The second is the κ framework, which introduces coupling modifiers for each

interaction term of the Higgs boson. The modifiers are all unity in the SM and may receive corrections in some BSM models. In addition to the tree-level coupling modifiers, important loop-level modifiers are considered as well, to parameterize modifications to the loop-induced Higgs couplings to massless gauge bosons. The modifiers κ_g, κ_γ, and $\kappa_{Z\gamma}$ are introduced to parameterize the $H \to gg$, $H \to \gamma\gamma$, and $H \to Z\gamma$ loops, respectively.

Following the philosophy of considering all allowed renormalizable interactions, the κ framework may be extended in the Yukawa sector to include the CP-odd fermionic coupling terms $\tilde{\kappa}_i h\bar{\psi}_i\gamma^5\psi_i$ [22] where $\tilde{\kappa}_i$ is the CP-odd coupling strength. These couplings introduce CP-violation given that the Higgs boson is a CP-even state and can result from mixing with an additional pseudoscalar boson. This work is primarily concerned with the top-Higgs coupling and therefore focuses on the top Yukawa sector with the dimension four Lagrangian

$$\mathcal{L} = h\bar{t}(\kappa_t + i\tilde{\kappa}_t\gamma^5)t.$$

Another common parameterization is

$$\mathcal{L} = hk_t\bar{t}(\cos\alpha_t + i\sin\alpha_t\gamma^5)t, \tag{2.8}$$

where α_t, often rendered simply as α, is the known as the CP mixing angle. In the SM, $\alpha_t = 0$, or equivalently $\tilde{\kappa}_t = 0$. In some work, k_t of the second parameterization is referred to as κ_t; elsewhere, κ_t and $\tilde{\kappa}_t$ of the first parameterization are written as a_t and b_t. This work will occasionally switch between the two parameterizations as is convenient.

2.4 Phenomenology of High Energy Hadron Colliders

Proton–proton colliders offer the best direct measurements of particle physics at high energy scales because protons have the largest mass of all stable subatomic particles. Stability means that protons are easily produced and stored, charge makes them easy to accelerate and steer with electromagnetic fields, and high mass allows them to be accelerated to higher energies before losses from synchrotron radiation become significant. On the other hand, the proton's nature as a composite particle creates particular complications for understanding the physics of these collisions. The highest energy collider, the Large Hadron Collider (LHC), collides protons at a center of mass energy $\sqrt{s} = 7-13.6$ TeV and is the leading experimental tool for direct measurements of high energy electroweak, Higgs boson, and top quark physics.

Parton Distribution Functions

The quark model describes the proton as a bound state of three quarks (u, u, d). For reference, see the PDG Review [11]. The full QCD description of the proton is non-perturbative and requires numerical treatment on a lattice. While such studies are not yet able to describe proton behavior in high-energy collisions, much of the relevant physics may be understood with perturbative calculations. As a result of asymptotic freedom, deep inelastic scattering

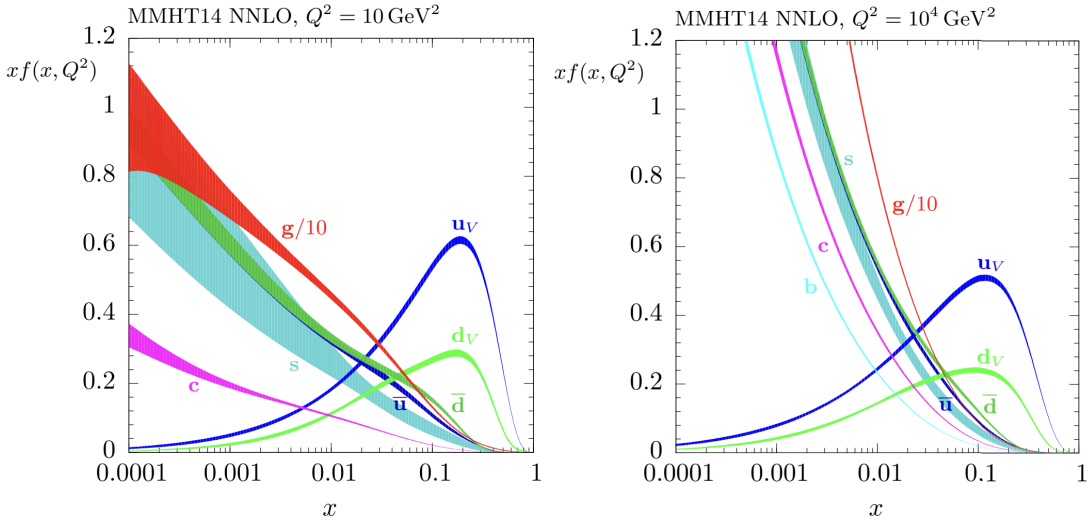


Figure 2.2: MMHT2014 NNLO PDFs at $Q^2 = 10$ GeV^2 and $Q^2 = 104$ GeV^2, with associated 68% confidence-level uncertainty bands. Reproduced from [23]

(scattering with a large invariant momentum transfer Q^2) of leptons and neutrinos off of hadrons may be approximated as interactions with a single quark q, which carries a fraction x of the proton's momentum. The quark is drawn from a distribution $f_q(x, Q)$, known as the parton distribution function (PDF), which describes the partonic content of protons as a function of x, Q. The PDF is independent of Q at zeroth order in α_s, but acquires corrections from quark and gluon splitting described by the DGLAP equations. Similarly, pp collisions with large momentum transfer can be treated as parton-parton collisions where the partons are drawn independently from the PDFs of the incoming protons. The differential cross-section for the hard scatter process producing final state Y is then written as a sum over initial states and an integral over the parton momentum fractions x_1, x_2.

$$\frac{d\sigma(pp \to Y + X)}{dQ} = \sum_{ij} \int dx_1 dx_2 f_i(x_1, Q) f_j(x_2, Q) \frac{d\sigma}{dQ}(i + j \to Y) \tag{2.9}$$

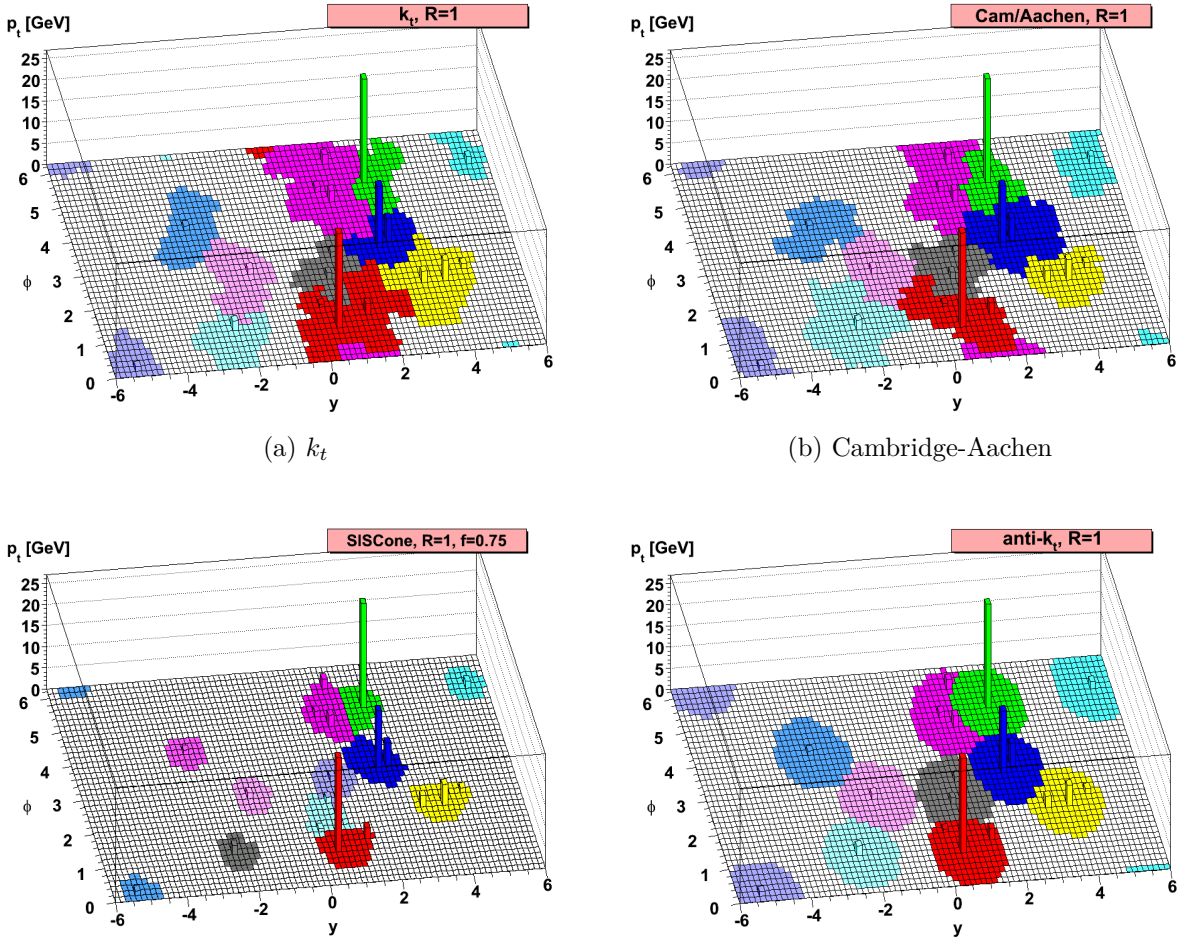
The proton PDF is a subject of intense study and global averages are produced by several groups, including the CTEQ [24], NNPDF [25], and MMHT [23] collaborations, for use in theoretical calculations of pp cross-sections. These combinations use a variety of deep inelastic scattering measurements and hadronic scattering experiments, as different scattering processes are sensitive to the PDFs of different combinations of partons and ranges of x. Figure 2.2 shows the results of one such fit, in which it is evident that gluon-initiated processes tend to have the highest cross-sections. As a point of reference, central $t\bar{t}$ production at threshold requires incoming partons with $x \times 6.5$ TeV $= m_t \approx 173$ GeV corresponding to $x = 0.027$ and an exchange of $Q^2 \approx m_t^2$. In this region, the gluon PDF is approximately

10 times that of the highest quark contribution and more than twice the sum of all quark contributions.

Development of Hard Scatter Partons

Quarks and gluons produced in the hard scatter collision or prompt decays of the final state particles do not directly reach the detector due to the confinement property of QCD. Instead, several aspects of QCD must be considered to understand the detector signatures of these particles and design suitable observables. The first aspect is infrared and collinear (IRC) divergences — perturbative QCD calculations suffer divergences in the emission of radiation with low momentum (IR) or at a small angle to the initial parton (collinear). These divergences cancel with virtual corrections when considering observables that are not sensitive to infrared or collinear radiation according to the KLN theorem [26, 27]. The second is hadronization — QCD has a Landau pole at a scale $\Lambda_{QCD} \approx 200$ MeV, near which the theory becomes non-perturbative. At low energies, the physical states of the theory are hadrons, and measurements of hadronic final states must consider the process by which quarks and gluons combine to form color-neutral bound states. Fortunately, both of these effects generally factorize from the hard-scatter collision. This motivates a standard treatment of QCD effects on the initial and final states as independent from the hard scatter that contains the process of interest. Many observables insensitive to IRC divergences have been proposed, but the most useful for electroweak scale physics are those that replace final state quarks and gluons with jets. Jets are collimated patterns of QCD radiation that appear at the detector level as hadrons and their decay products. A variety of jet definitions have been proposed and used for experiments past and present, including many that are "IRC safe," that is, insensitive to IRC divergences [28]. The preferred jet clustering algorithm for most measurements at the LHC is the anti$-k_t$ algorithm [29] with a radius parameter $R = 0.4$. It is chosen for a combination of IRC safety, which is theoretically desirable, and consistent geometric shapes, which are experimentally important for pileup rejection and calibration.

While measurements of final states including only jets, leptons, and photons are well-defined theoretically and may be calculated in perturbation theory, a full simulation of the QCD radiation within a jet and the subsequent hadronization is needed to allow calibration of the detector signatures and for use in identifying the parton that originates the jet, especially identifying jets from b-quarks. For this purpose, simulations of the hard scatter, which are generally based on scattering matrix elements calculated at next-to-leading-order (NLO) in α_s, are combined with parton shower calculations, which use resummed calculations to separate contributions from large logarithms by changing the parameter of the perturbative expansion to include the logarithm. These algorithms proceed via iterative splitting of partons into pairs according to splitting/radiation formulae obtained from perturbative QCD calculations. When parton shower and matrix element calculations proceed beyond leading-order, merging and matching schemes [30, 31, 32], are needed to avoid double-counting radiation that can be produced at either step. Unlike IRC divergences, divergences associated



(a) k_t (b) Cambridge-Aachen

(c) SISCone (d) anti-k_t

Figure 2.3: Parton level QCD event clustered with four different jet clustering algorithms showing the detector areas for which additional soft radiation would be included in the jet. [29]

with Λ_{QCD} are completely physical and are not canceled. Treating this region requires non-perturbative models of hadronization, such as the Lund string model [33] or cluster models [34]. These models are based on calculations of the QCD potential and scaling laws from perturbative QCD, but necessarily contain empirical inputs, in the form of the fragmentation functions that describe the likelihood for a parton at a given momentum scale to hadronize into a given species of hadron. These fragmentation functions must be taken from experiments, generally deep inelastic scattering measurements, and are assumed to be universal.

General-purpose Monte Carlo event generators simulate particle collisions from a chosen hard scatter process and provide as output a set of detector-stable particles. This process necessarily includes calculations of PDFs, underlying events, hard scattering matrix elements, parton showering, hadronization, and prompt and semi-prompt decays of heavy bosons, unstable hadrons, and τ-leptons.

The structure of a simulated proton–proton collision producing a pair of top quarks is shown schematically in fig. 2.4 as generated in the Pythia8 general-purpose Monte Carlo event generator. Concentric rings of increasing radius correspond to physics processes from both the initial state protons and the final state top quarks at lower energies, larger distances, and longer time scales, beginning with the hard-scattering process labeled by $d\hat{\sigma}_0$, which is modeled by perturbative scattering matrix element calculations, proceeding to the perturbative QCD and electroweak radiation, which includes showering of initial and final partons and prompt decays of heavy resonances, and finishing with color reconnection and hadronization. The objects in the outermost region are the initial state protons before the emission of a parton from the PDF, stable final state leptons and hadrons, and the decay products of promptly decaying hadrons.

2.5 Top Quark Physics at the LHC

The LHC is a powerful tool for top quark physics with a large $t\bar{t}$ production cross-section and high luminosities. This has been exploited to measure the top quark mass, width, decays, and production with record precision [36]. Figure 2.5 shows the cross-section predictions and ATLAS measurements of selected important SM scattering processes at the LHC, including the principal top quark production modes. In addition to the dominant $t\bar{t}$ production at the LHC, studies of the rarer tW, $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}\gamma$, tWZ, and $t\bar{t}t\bar{t}$ processes can constrain the top quark couplings to scalar and vector bosons. These processes may also receive enhancements from BSM particles with cascade decays into top quarks, including heavy colored particles, from heavy new physics that induces anomalous couplings of top quarks to electroweak bosons, or that generates four-fermion contact interactions when integrated out in an EFT framework.

The rare $t\bar{t}t$ process is somewhat different than other top processes. It consists of two related processes, $t\bar{t}tW$ and $t\bar{t}tq$ [38, 39, 40], with cross-sections of 1.02 and 0.65 fb, respectively, calculated at NLO in QCD. Representative Feynman diagrams for $t\bar{t}tW$ and $t\bar{t}tq$

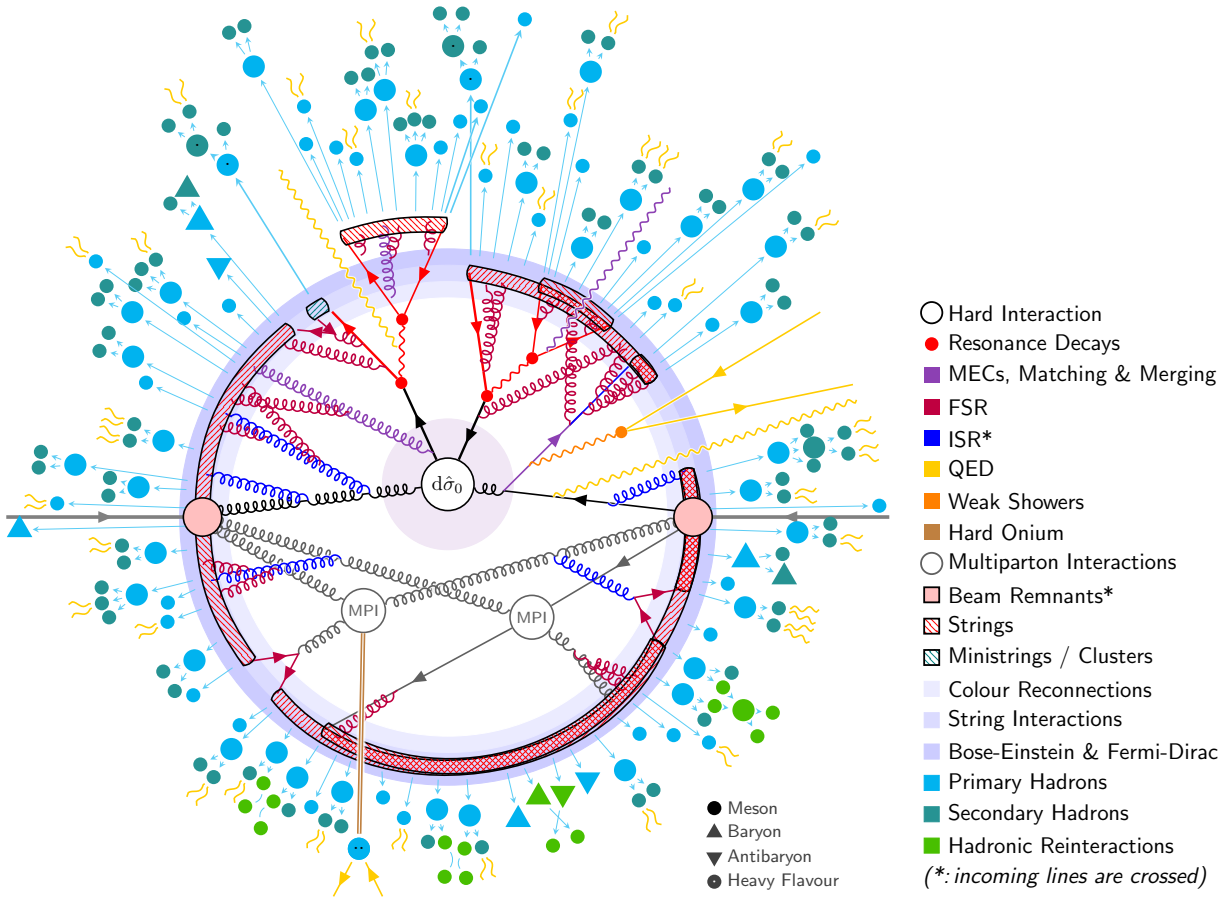


Figure 2.4: Schematic of the structure of a $pp \to t\bar{t}$ event, as modeled by the PYTHIA8 general purpose Monte Carlo event generator. Several minor simplifications have been made to improve readability. [35]

production are shown in fig. 2.6. Because of its small cross-section and the similarity with the $t\bar{t}t\bar{t}$ final state, $t\bar{t}t$ has not yet been experimentally observed. Because the overall cross-section is so small, $t\bar{t}t$ offers limited sensitivity to the top-philic new physics compared to $t\bar{t}t\bar{t}$ and other processes. However, $t\bar{t}tq$ production in particular may be significantly enhanced by interactions that induce flavor-changing neutral currents. Many BSM models allow such interactions unless additional symmetries are imposed to protect the flavor sector, including models with extended Higgs sectors. In these models, both the SM-like Higgs boson and additional scalar bosons may acquire flavor-changing interactions, which could be observed as an enhancement in $t\bar{t}tq$ production.

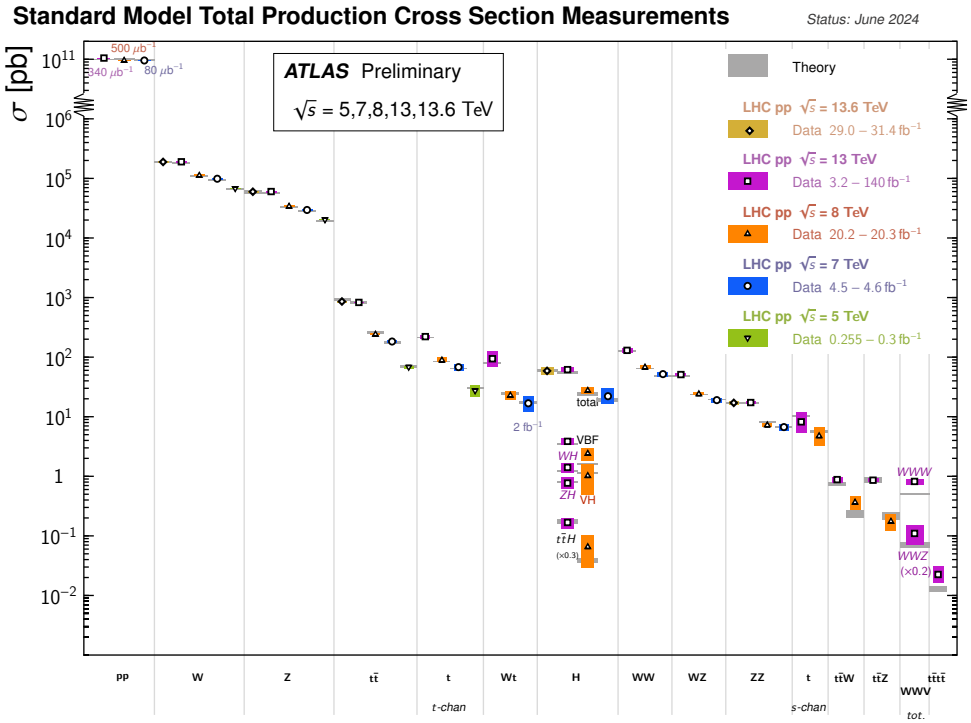


Figure 2.5: Summary of several Standard Model cross-section measurements (a) with associated references (b). The measurements are corrected for branching fractions, compared to the corresponding theoretical expectations. [37]

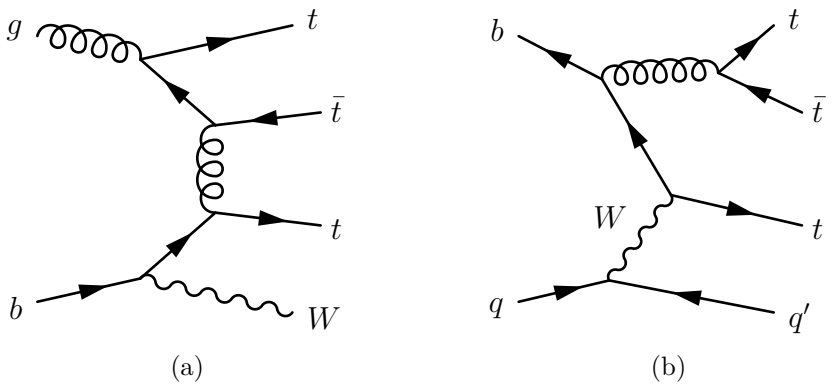


Figure 2.6: Leading-order tree-level diagrams for $t\bar{t}t$ production showing the dominant $t\bar{t}tW$ (a) and $t\bar{t}tq$ (b) modes.

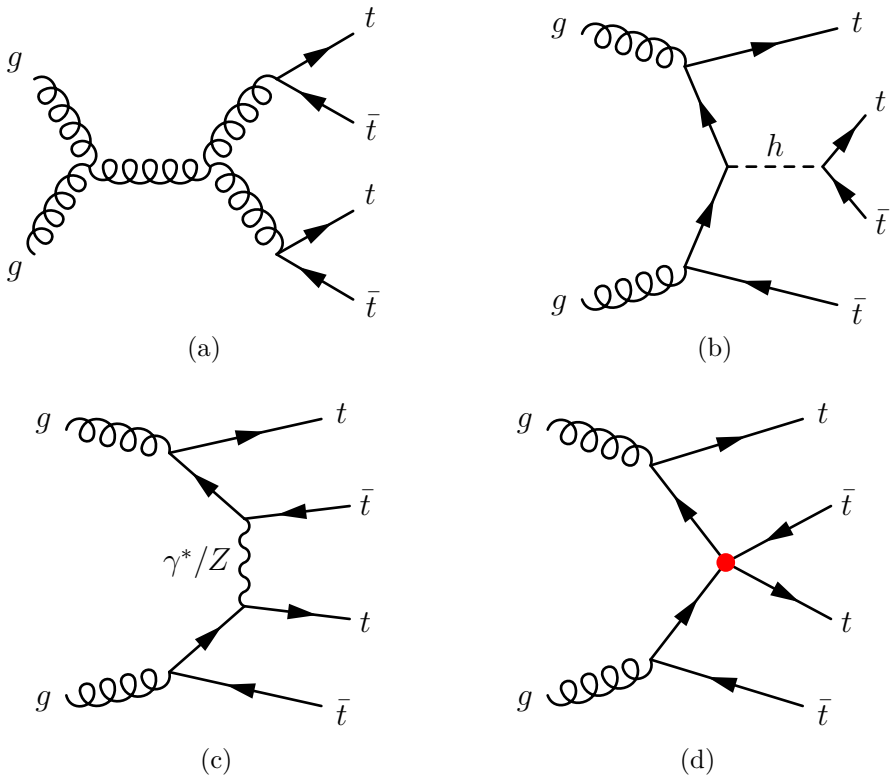


Figure 2.7: Feynman diagrams for $t\bar{t}t\bar{t}$ production showing pure QCD (a), Higgs mediated (b), and electroweak (c) production as well as a diagram representing BSM contributions through four fermion EFT operators (d).

$t\bar{t}t\bar{t}$ Production

The simultaneous production of four top quarks, known as $t\bar{t}t\bar{t}$ production, is a rare, previously unobserved process at the LHC with a high mass final state and sizable contributions from both pure QCD production and diagrams mediated by electroweak bosons. The cross-section is calculated to be $\sigma_{t\bar{t}t\bar{t}} = 12 \pm 2.4$ fb [41, 42, 43] at next-to-leading-order in QCD and including electroweak corrections. The uncertainty is evaluated as the variation in the calculated cross-section with the renormalization scale. A threshold resummation up to the next-to-leading-logarithm and including next-to-leading-order electroweak corrections gives a more precise calculation of $\sigma_{t\bar{t}t\bar{t}} = 13.4^{+1.0}_{-1.8}$ fb [44]. The effects of various corrections to the cross-section calculation on the predicted cross-section and the uncertainty from scale variation are shown in fig. 2.8. The combination of these corrections results in a cross-section over 100% above the leading-order prediction. Measuring the $t\bar{t}t\bar{t}$ production cross-section experimentally provides an important test of the SM in a complex regime with challenging theoretical calculations.

Calculations at leading-order [45] indicate that the cross-section contribution from Higgs-

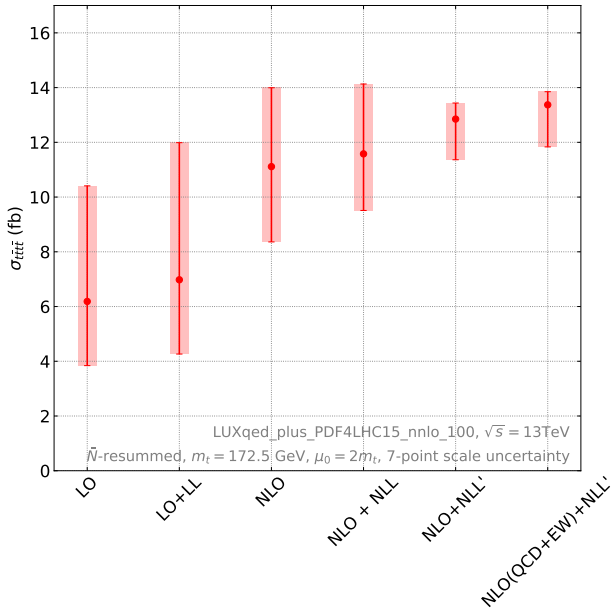


Figure 2.8: Predictions for the total $pp \to t\bar{t}t\bar{t}$ cross-section at $\sqrt{s} = 13$ TeV for fixed-order calculations and resummation-improved results, obtained using the 7-point scale variation. [44]

mediated diagrams is 15%, though that contribution is almost exactly canceled by the interference of the Higgs-mediated diagrams with the pure QCD production. An increase in the $t\bar{t}t\bar{t}$ production cross-section above the SM expectation is possible if the top quark couples more strongly to the Higgs boson or if it has BSM CP properties [46, 45]. This is one of many scenarios in which BSM physics may enhance $t\bar{t}t\bar{t}$ production, which is sensitive due to its low SM cross-section and high mass final state. The possibility of modifications from BSM physics that preferentially couples to the top quark is particularly interesting in light of the importance of the top quark in solutions to the hierarchy problem and many other well-motivated extensions of the Standard Model. Scenarios featuring enhanced $t\bar{t}t\bar{t}$ production include gluino pair production in supersymmetric [47, 48] theories, scalar gluon pair production [49, 50], contributions from a heavy scalar or pseudoscalar boson in models with extended Higgs sectors such as the two Higgs doublet model [51, 52, 53], or theories with composite top quarks [54]. The $t\bar{t}t\bar{t}$ cross-section may also be affected by modifications to the Higgs oblique parameter [55].

Heavy BSM physics that explains the flavor hierarchy or is Higgs-like and couples to mass is expected to couple preferentially to the third generation of fermions. This motivates the study of EFT operators involving the third generation. In the case of a heavy neutral boson (either a heavy Higgs boson or a Z') with a large top quark coupling, contributions to the simultaneous production of four top quarks ($t\bar{t}t\bar{t}$) may be expected via diagrams like fig. 2.7(d). Integrating out the heavy mediator generates four fermion contact interactions [56, 57, 58,

59]. This work considers four such operators accessible in $t\bar{t}t\bar{t}$ production.

$$\begin{aligned}
\mathcal{L}_{EFT} &= \frac{c^1_{QQ}}{\Lambda^2}\mathcal{O}^1_{QQ} + \frac{c^1_{Qt}}{\Lambda^2}\mathcal{O}^1_{Qt} + \frac{c^1_{tt}}{\Lambda^2}\mathcal{O}^1_{tt} + \frac{c^8_{Qt}}{\Lambda^2}\mathcal{O}^8_{Qt} \\
\mathcal{O}^1_{QQ} &= \left(\bar{Q}_L\gamma^\mu Q_L\right)\left(\bar{Q}_L\gamma_\mu Q_L\right) \\
\mathcal{O}^1_{tt} &= \left(\bar{t}_R\gamma^\mu t_R\right)\left(\bar{t}_R\gamma_\mu t_R\right) \\
\mathcal{O}^1_{Qt} &= \left(\bar{Q}_L\gamma^\mu Q_L\right)\left(\bar{t}_R\gamma_\mu t_R\right) \\
\mathcal{O}^8_{Qt} &= \left(\bar{Q}_L\gamma^\mu T^A Q_L\right)\left(\bar{t}_R\gamma_\mu T^A t_R\right)
\end{aligned} \tag{2.10}$$

Previous measurements of $t\bar{t}t\bar{t}$ production by the ATLAS [60, 61] and CMS [62] collaborations have yielded evidence of $t\bar{t}t\bar{t}$ production with a statistical significance of 4.7σ for ATLAS in a combination of the single lepton, opposite-sign dilepton, same-sign dilepton, and trilepton final states, and a significance of 4.0σ for CMS with these final states in addition to the fully hadronic final state. Both collaborations measure cross-sections above the SM expectation, but the measurements are consistent both with each other and with the SM.

2.6 Higgs Boson Physics at the LHC

Since the discovery of the Higgs boson in 2012 by the ATLAS and CMS experiments [63, 64], subsequent measurements from the ATLAS and CMS collaborations at the LHC have confirmed that the observed boson has spin 0 and is consistent with a state that is even under CP transformations [65, 66]. The latest measurements of the Higgs boson mass [67, 68] give masses of 125.11 ± 0.11 and 125.35 ± 0.15 GeV based on combinations of measurements in the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ decay channels using data from LHC Runs 1 and 2. The experiments have observed a broad range of production modes, including gluon–gluon fusion (ggF), vector boson fusion (VBF), associated production with a vector boson (VH), and top-associated production ($t\bar{t}H$), with cross-sections between 500 fb and 50 pb [21]. They have also established a number of decay modes, including decays into b-quarks, τ-leptons, photons, and pairs of massive electroweak bosons (WW^* and ZZ^*) with branching fractions ranging from 10^{-3} to 0.6. Many of these results from Run 2 are summarized and combined in each experiment's ten-year Higgs boson retrospective [69, 70, 71, 72]. Searches for the rare Higgs boson decays into $Z\gamma$ [73] and $\mu^+\mu^-$ [74] have resulted in evidence at the 3σ and 1.7σ levels, respectively, and searches for tH [75] production and Higgs boson decays into $c\bar{c}$ [76] have placed upper limits around 15 and 10 times the SM prediction, respectively. Interpreted in the κ framework, these measurements establish that the Higgs couplings to the five most massive SM particles (excluding the Higgs boson itself) are consistent with the SM expectation and that no other Higgs couplings are significantly larger than the SM expectation. Additional differential cross-section measurements, including measurements of the simplified template cross-sections, are all consistent with the SM expectation for Higgs

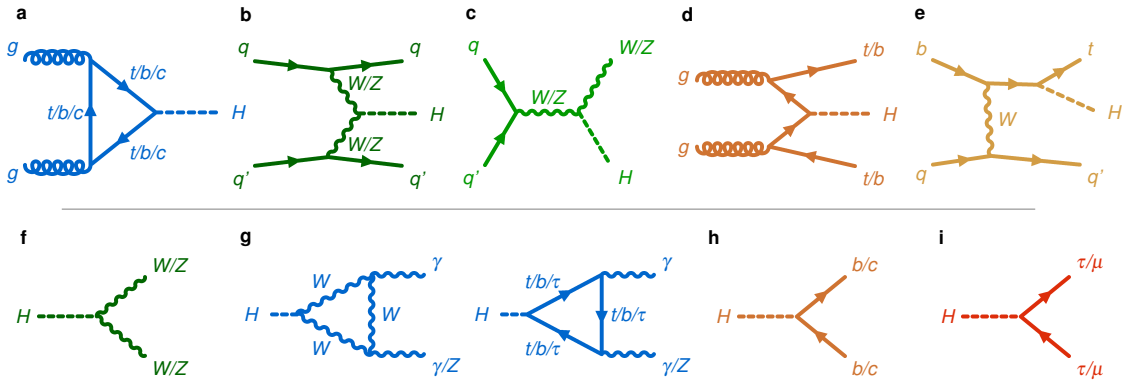


Figure 2.9: Examples of Feynman diagrams for Higgs boson production and decays. The Higgs boson is produced via gluon–gluon fusion (a), vector-boson fusion (VBF; b), and associated production with vector bosons (c), top- or b-quark pairs (d), or a single top quark (e). f-i, The Higgs boson decays into a pair of vector bosons (f), a pair of photons or a Z boson and a photon (g), a pair of quarks (h), and a pair of charged leptons (i). Loop-induced Higgs boson interactions with gluons or photons are shown in blue, processes involving couplings to W or Z bosons in green, to quarks in orange, and to leptons in red. Two different shades of green (orange) are used to separate the VBF and VH ($t\bar{t}H$ and tH) production processes. [69]

boson production at the LHC. The latest ATLAS combination of results is shown in figs. 2.10 and 2.11 interpreted in the κ and STXS frameworks.

CP Properties of the Higgs Boson

Measurements searching for CP-violation or deviations from the expected CP properties of the Higgs boson in the Higgs couplings to vector bosons have been performed in Higgs boson decays into vector boson pairs $H \to ZZ^*$ and $H \to WW^*$ and in vector boson fusion Higgs production [77, 78, 79, 80]. Because CP-odd interactions with vector bosons can only enter through non-renormalizable terms, these measurements are performed in an EFT framework. No deviations from the SM expectation that the Higgs boson is a CP-even state have been observed. Similar constraints have been placed on CP-violation in the Yukawa sector using interpretations based on eq. (6.7). The main candidates for these studies are top quarks, where CP-violating effects can show up in top-associated Higgs boson production, and τ-leptons, where the effects can be measured in the Higgs decay $H \to \tau^+\tau^-$. The former is treated in a dedicated section on the top Yukawa coupling. The latter requires reconstruction of the τ-lepton polarization and has been performed by the ATLAS [81] and CMS [79] experiments. The CP-odd hypothesis has been excluded, but CP mixing angles on the order of $\alpha_\tau \approx 30°$ are still well within allowed bounds.

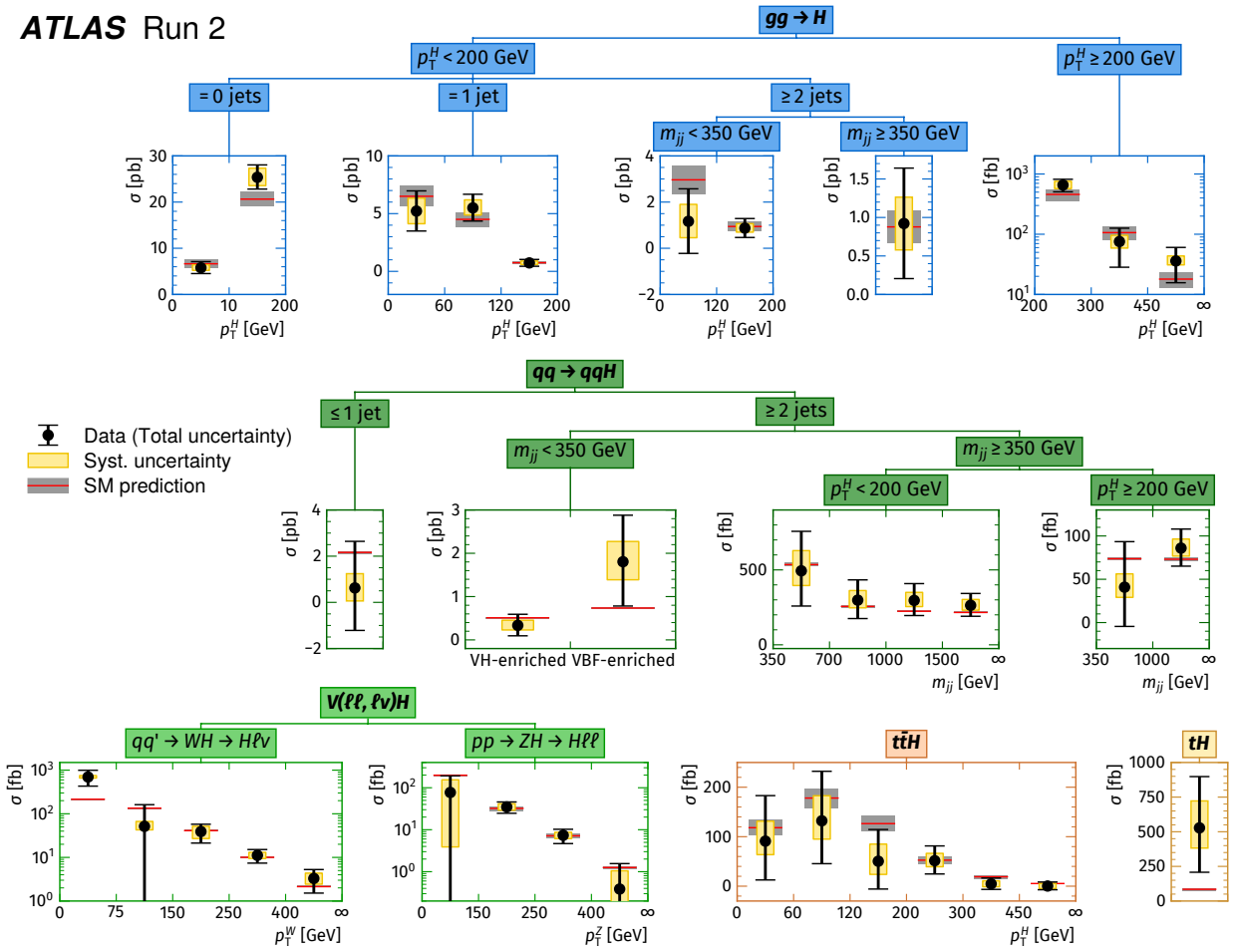


Figure 2.10: Measurements of Higgs boson production in the STXS framework. [69]

Higgs Boson Total Width

The total width Γ_h, or equivalently the lifetime τ_h, of the Higgs boson is not directly accessible at the LHC. The SM width prediction of 4.1 MeV [21] is much narrower than the approximately 1 GeV detector resolution [82, 83] in the "golden" $H \to \gamma\gamma$ and $H \to 4\ell$ channels, so lineshape measurements offer only a loose upper bound on the Higgs width. The lifetime is inversely proportional to the width and provides an alternative opportunity to measure the width, but the SM prediction of $\tau_h = 16 \times 10^{-8}$ fs $= 48$ fm/c leads to displacements much smaller than the vertexing resolution of the detectors. CMS has put an upper limit on the lifetime of $\tau_h < 190$ fs [83] by measuring the displacement of the Higgs boson decay vertex from the beamspot. This corresponds to a loose lower bound on the total width of the Higgs boson. Because the cross-sections for Higgs boson production by coupling to species i and decaying through coupling to species f are approximated by a Breit–Wigner

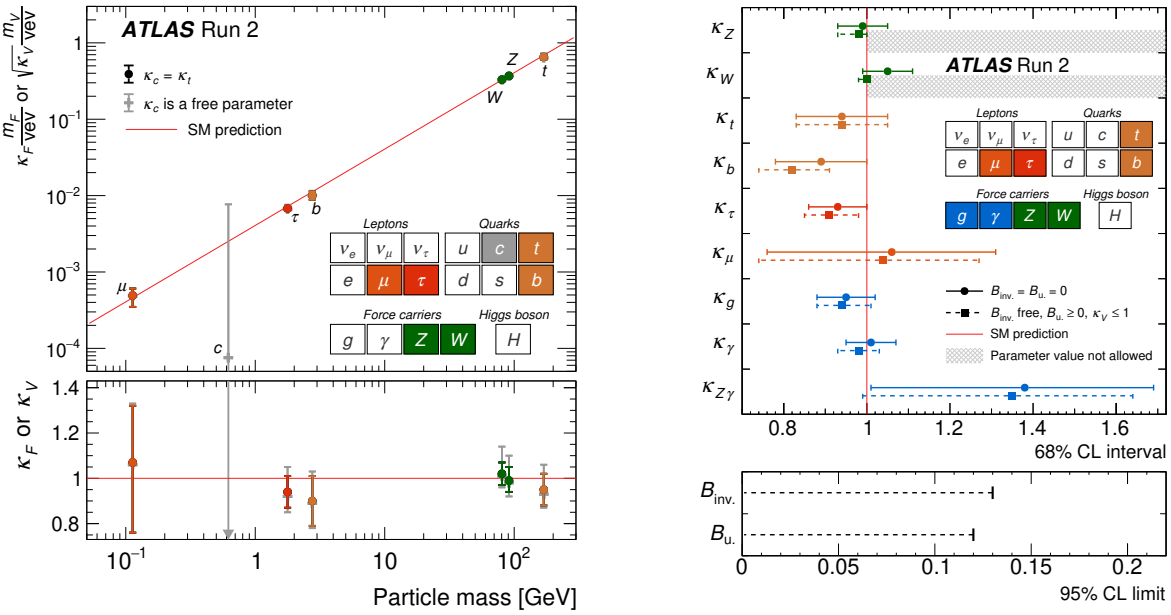


Figure 2.11: (Left) Higgs coupling strengths as a function of particle mass versus the SM prediction with loop-induced couplings parameterized in terms of tree-level couplings to SM particles. (Right) Measurements of Higgs coupling modifiers with loop-induced and tree-level couplings measured independently. [69]

distribution in the invariant mass m of the Higgs boson decay products

$$\frac{d\sigma_{i\to H\to f}}{dm} \propto \frac{g_i^2 g_f^2}{(m^2 - m_h^2)^2 + m_h^2\Gamma_h^2}, \tag{2.11}$$

the total on-shell cross-section depends on Γ_h as

$$\sigma_{i\to H\to f} \propto \frac{g_i^2 g_f^2}{m_h^2 \Gamma_h} \tag{2.12}$$

and the ratio to the Standard Model cross-section prediction, known as the signal strength μ, is given by

$$\mu_{i\to H\to f} = \frac{\sigma_{i\to H\to f}}{\sigma_{i\to H\to f}^{SM}} = \frac{g_i^2 g_f^2}{m_h^2\Gamma_h} \frac{m_h^2 \Gamma_h^{SM}}{g_{i,SM}^2 g_{f,SM}^2} = \kappa_i^2 \kappa_f^2 \frac{\Gamma_h^{SM}}{\Gamma_h}. \tag{2.13}$$

Measurements of on-shell Higgs boson production and decay depend on the total width, and the individual branching fractions cannot be identified with an inclusive cross-section

measurement. Without an assumption that the Higgs has negligible decays into BSM particles or rare unobserved SM decay modes, the total width, or equivalently the total production cross-section, remains an almost unconstrained direction in the Higgs sector. Equation (2.13) shows that the flat direction corresponds to scaling all Higgs coupling modifiers κ by $(\Gamma_h/\Gamma_h^{SM})^{1/4}$, whereas in the absence of BSM contributions to Γ_h, Γ_h scales as κ^2. The parameterization used in Ref. [69] for scenarios with no Higgs width into BSM decays is

$$\begin{aligned}\frac{\Gamma_h}{\Gamma_h^{SM}} &= 0.581\kappa_b^2 + 0.215\kappa_W^2 + 0.082\kappa_g^2 + 0.063\kappa_\tau^2 + 0.026\kappa_Z^2 \\ &+0.029\kappa_c^2 + 0.0023\kappa_\gamma^2 + 0.0015\kappa_{Z\gamma}^2 + 0.0004\kappa_s^2 + 0.00022\kappa_\mu^2.\end{aligned} \tag{2.14}$$

Thus, in the simplified picture where all Higgs boson couplings are scaled uniformly by κ, eq. (2.13) can be expressed in terms of the partial width associated with BSM Higgs boson decays Γ_h^{BSM}.

$$\mu_{i\to H\to f} = \kappa^4 \frac{\Gamma_h^{SM}}{\kappa^2\Gamma_h^{SM} + \Gamma_h^{BSM}} \implies \frac{\Gamma_h^{BSM}}{\Gamma_h^{SM}} = \kappa^2\left(\frac{\kappa^2}{\mu_{i\to H\to f}} - 1\right) \tag{2.15}$$

If all on-shell Higgs boson measurements are consistent with the SM predictions, undetected Higgs boson decays into BSM particles may be allowed if all Higgs couplings are increased by an appropriate factor. Note that $\Gamma_h^{BSM} < 0$ may be reasonable if BSM physics enters Higgs boson decays via destructive interference in the ggF or $H \to \gamma\gamma$ loops. The addition of off-shell measurements of Higgs boson production, which is accessible due to enhancements from kinematic thresholds [84, 85, 86], has provided considerable indirect constraining power. Equation (2.11) predicts vanishingly small cross-sections for off-shell Higgs boson production with $m \gg m_h$, but the Breit–Wigner behavior is altered when the fully on-shell ZZ, WW, and $t\bar{t}$ final states become kinematically accessible. Though the Breit–Wigner shape is no longer accurate, it is still the case that off-shell Higgs boson production does not depend on Γ_h.

$$\sigma_{i\to H^*\to f} = f(m_h)g_i^2 g_f^2, \quad \mu_{i\to H^*\to f} = \kappa_i^2\kappa_f^2 \tag{2.16}$$

Therefore, the ratio of the Higgs boson total width to its SM predictions is given by the ratio of the on- and off-shell Higgs boson production signal strengths.

$$\frac{\mu_{i\to H\to f}}{\mu_{i\to H^*\to f}} = \frac{\Gamma_h}{\Gamma_h^{SM}} \tag{2.17}$$

In the simplified model with a single modifier for Higgs boson couplings to SM particles, the observed signal strength ratio can be related to the width modification from BSM particles or to κ.

$$\frac{\mu_{i\to H\to f}}{\mu_{i\to H^*\to f}} = \frac{\Gamma_h^{SM}}{\kappa^2\Gamma_h^{SM} + \Gamma_h^{BSM}} = \frac{\Gamma_h^{SM}}{\sqrt{\mu_{i\to H^*\to f}}\Gamma_h^{SM} + \Gamma_h^{BSM}} \tag{2.18}$$

$$\frac{\Gamma_h^{BSM}}{\Gamma_h^{SM}} = \left(\frac{\mu_{i\to H^*\to f}}{\mu_{i\to H\to f}} - \sqrt{\mu_{i\to H^*\to f}} \right) \tag{2.19}$$

Positive contributions of BSM particles to the Higgs decay width are expected to induce decreases in the off-shell signal strength when the on-shell signal strength is fixed to unity. In practice, measurements of off-shell Higgs boson production contend with large interference from QCD and EW production and are better considered width-independent measurements of the Higgs coupling strengths. These off-shell measurements can still be combined with the width-dependent constraints from on-shell measurements to produce a value of Γ_h. Measurements have been made in the ZZ [87, 88] and WW [89] channels. The most precise results are from Ref. [87] and give $\Gamma_h = 4.3^{+2.7}_{-1.9}$ MeV. Each of these measurements is dependent on the assumption that no new BSM physics differentially affects on- and off-shell Higgs boson production rates. In the presence of such physics, including light new particles that may run in the gluon–gluon fusion loop [90, 91, 92], the off-shell measurements may be considered as a test of these modifications.

Top Yukawa Constraints

Direct measurements of the top Yukawa coupling are an important test of the electroweak symmetry breaking mechanism and the origin of fermion masses. Observation of this coupling has shown that the observed Higgs boson has properties consistent with the SM mass generation mechanism, and further, more precise measurements will tighten the constraints on well-motivated extensions of the Higgs sector. Due to the large masses of the Higgs boson and top quark, the LHC is the only experiment capable of directly producing these particles to study this coupling. LHC processes sensitive to the top Yukawa can be divided into three categories: loop-induced Higgs couplings, top-associated Higgs production, and Higgs-mediated electroweak corrections to top quark production. The former category is dominated by the Hgg and $H\gamma\gamma$ vertices, hence, the constraints come from the gluon–gluon fusion production cross-section and the diphoton decay rate. The second category consists of the production of a Higgs boson in association with one (tH) or two ($t\bar{t}H$) top quarks. The tH production consists of two separate processes, one with an external W boson (tWH) and the other with an internal W boson (tHq or $tHjb$). The SM tH cross-section is suppressed by destructive interference between diagrams in which the Higgs boson radiates from a top quark and those in which it radiates from a W boson. The much larger $t\bar{t}H$ cross-section dominates the direct on-shell determination of the top Yukawa coupling. The final category consists of loop-level modifications to $t\bar{t}$ production and tree-level contributions to $t\bar{t}t\bar{t}$ production.

Measurements of $t\bar{t}H$ production have been performed targeting Higgs bosons decaying into $b\bar{b}$ [93, 94], $\gamma\gamma$ [95], and $\tau\tau$ [96] and in the multilepton final states [97] which include decays into pairs of τ-leptons or electroweak bosons at least one of which decays leptonically. Many of these measurements also consider tH production. A few dedicated searches for tH production have also been performed [75]. These measurements achieve a precision on the

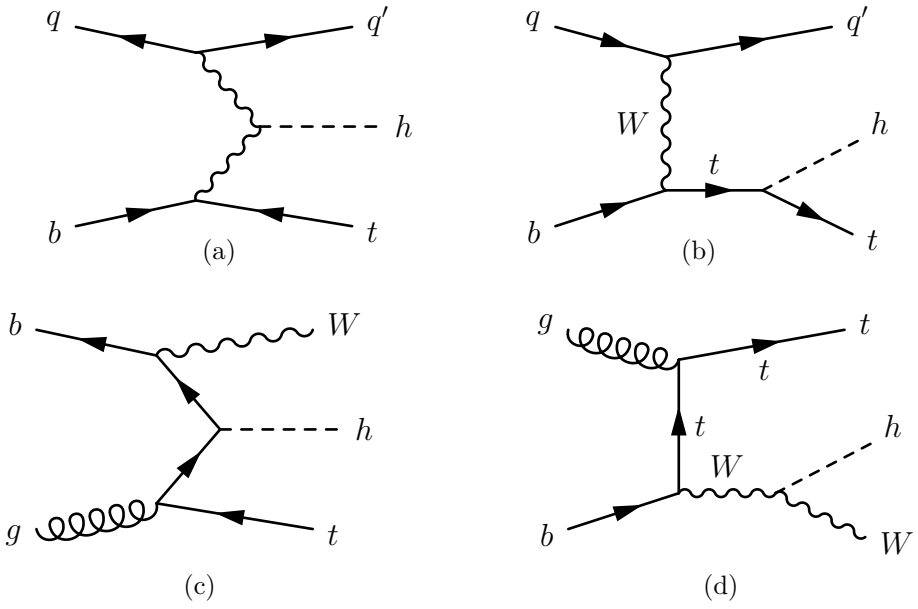


Figure 2.12: Feynman diagrams for tH production showing diagrams for tHq (a,b) and tWH (c,d) production and diagrams proportional to κ_t (b,c) and κ_W (a,d).

order of 10% in the measurement of the top Yukawa coupling, and none find a significant deviation from the SM expectation, though many measure $t\bar{t}H$ signal strengths 10%–30% below the SM expectation.

Top Yukawa constraints from off-shell contributions to top quark production have been studied in the context of $t\bar{t}$ production, where the Higgs loop introduces an especially large correction near the $t\bar{t}$ production threshold, and in $t\bar{t}t\bar{t}$ production, where the Higgs mediated diagrams and their interference with the dominant QCD production have effects on the order of 10% on the $t\bar{t}t\bar{t}$ production cross-section and additional effects on differential observables. The effect on $t\bar{t}$+jets has been studied by the CMS experiment [99] with 35.8 fb^{-1} resulting in a limit of $|\kappa_t| < 1.67$ and the effect on $t\bar{t}t\bar{t}$ has been studied [100] with 137 fb^{-1} resulting in a limit of $|\kappa_t| < 1.7$ excluding the uncertainty in the prediction of the SM $t\bar{t}t\bar{t}$ cross-section. An ATLAS measurement based on $t\bar{t}t\bar{t}$ production is one of the central results of this work and is presented in chapter 10.

CP-violation introduced in the top Yukawa sector is strongly constrained by low-energy measurements. The most stringent constraints come from the electric dipole moment of the electron [101, 102], which is disallowed without CP-violation and receives contributions from Barr–Zee diagrams of the type shown in fig. 2.14. This diagram contains a Higgs boson radiating from an electron and is hence proportional to the electron Yukawa coupling, which is too small to be observable at the LHC. With the assumption that all CP-even Yukawa couplings have their SM values and CP-violation is only present in the top Higgs coupling, the ACME limit [103] $d_e < 1.1 \times 10^{-29}$ e cm translates to a limit of $|\tilde{\kappa}_t| < 1.15 \times 10^{-3}$.

In addition to the assumption that the electron couples to the Higgs boson as predicted

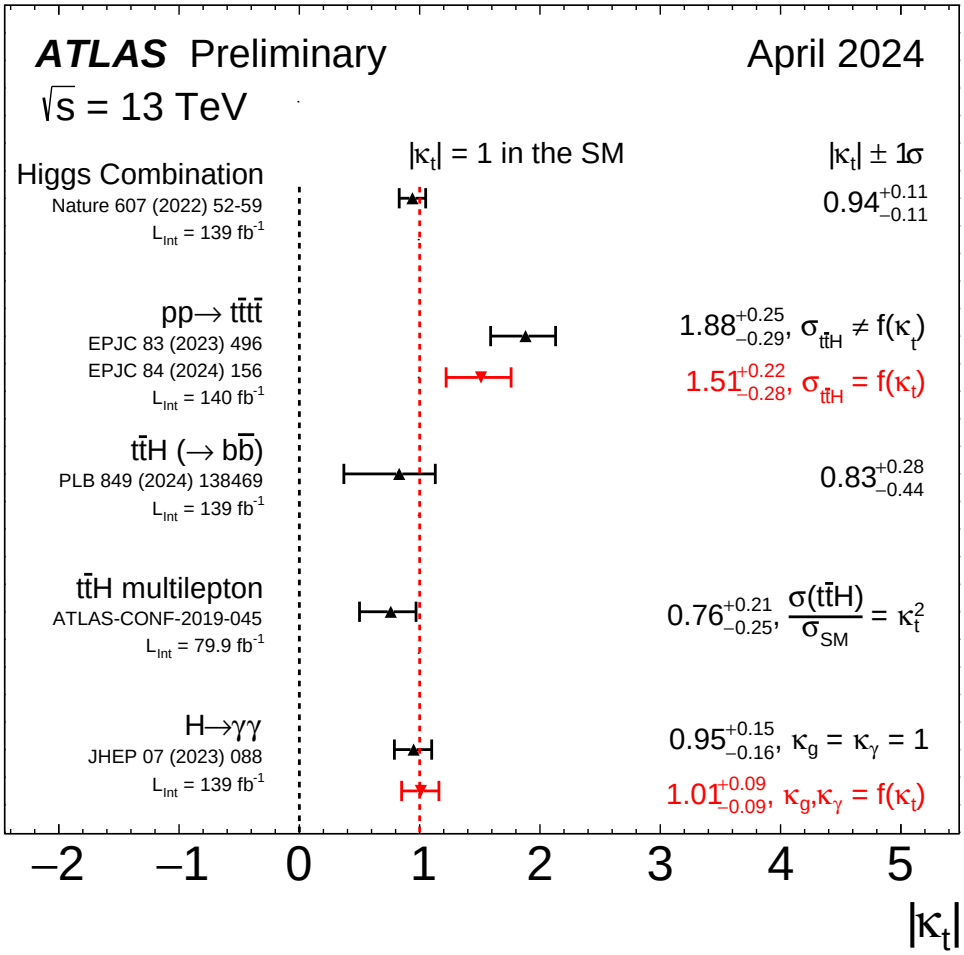


Figure 2.13: Summary of measurements of the top-Higgs Yukawa coupling modifier in the Kappa formalism, with the ATLAS experiment. The black and red points denote alternative parameterizations of κ_t in the measurements. For the combined Higgs property measurement, $\sigma_{t\bar{t}t\bar{t}}$ is fixed to the SM expectation, and other coupling modifiers are fit simultaneously. In the $t\bar{t}t\bar{t}$ measurement, values for $|\kappa_t|$ are extracted with $\sigma_{t\bar{t}H}$ either parameterized as a function of κ_t (in red) or profiled (in black). In the $t\bar{t}H$ multilepton measurement, $|\kappa_t|$ is derived from the best-fit value of $\sigma_{t\bar{t}H}$ assuming $\sigma_{t\bar{t}H} \propto \kappa_t^2$. In the $H \to \gamma\gamma$ measurement, $|\kappa_t|$ is determined either with κ_g and κ_γ parameterized as a function of $|\kappa_t|$ (in red), or with both set to their SM expectation values (in black). All other couplings in these measurements are fixed to the values predicted by the SM. This plot was modified in April 2024 to integrate updated $t\bar{t}t\bar{t}$ results and to update the references for the $t\bar{t}H$ and $H \to \gamma\gamma$ measurements. Status of figure: April 2024. [98]

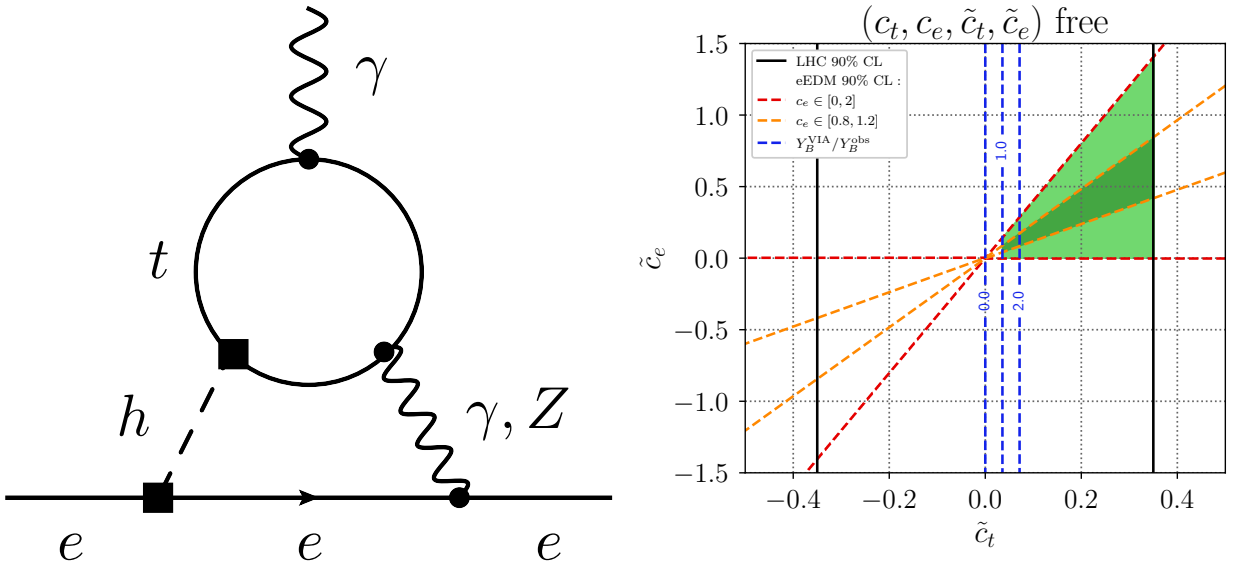


Figure 2.14: (Left) Feynman diagram showing the contribution of CP violating Yukawa couplings to the electron EDM. (Right) 2D exclusion contours for the CP-odd electron and top quark Yukawa couplings showing regions allowed by possible cancellations. [101, 102].

in the SM, this limit is sensitive to possible cancellations with CP-odd couplings to other fermions. Many of these scenarios are implausible because the top quark naturally produces the largest corrections, but the addition of a small CP-odd electron Yukawa coupling can yield significant cancellations proportional to the CP-even component of the top Yukawa coupling. This scenario, explored in Ref. [102], motivates direct measurements of the CP nature of the top Yukawa coupling at the LHC. The same high-energy processes sensitive to the CP-even coupling also show various levels of sensitivity to CP-violating effects [22, 46, 45]. Direct probes of the CP properties of the top Yukawa coupling have been performed for $t\bar{t}H$ and tH in the $H \to \gamma\gamma$ [104, 105], $H \to b\bar{b}$ [106, 94], and multilepton [107] final states with the most stringent limit on the CP mixing angle being $\alpha < 43°$ in the $H \to \gamma\gamma$ channel. While no significant deviation from the SM expectations for the strength and CP properties of the top Higgs coupling has yet been observed, the indirect measurements carry a range of model assumptions, and direct measurements leave room for up to 10% deviations in the coupling strength and substantial CP-violation. Given the crucial role of the top Yukawa coupling in electroweak and cosmological physics, these uncertainties call for further measurements to precisely determine the nature of the Higgs and Yukawa sectors.

Chapter 3

The ATLAS Experiment at the Large Hadron Collider

As described in chapter 2, the most powerful experimental probes of the electroweak scale are high-energy proton collisions. An accelerator capable of reaching the energies required for these experiments must be physically large and be integrated into a sophisticated supporting infrastructure, including a multi-stage chain of accelerators and injectors to feed the larger machine. Presently, the only such machine is the Large Hadron Collider (LHC), described in the first section of this chapter. The collisions themselves require large detectors with difficult operating conditions and stringent technical specifications to achieve the physics goals of the LHC program. ATLAS is one such detector, and data collected by the ATLAS detector have been used to perform the measurements presented in this work. The remainder of the chapter describes the LHC and the design of the ATLAS detector.

3.1 The Large Hadron Collider

The LHC is a circular synchrotron particle accelerator and collider sited at CERN and straddling the border between France and Switzerland [108]. The LHC uses the 26.7 km tunnel originally excavated for LEP. The accelerator has eight curved sectors with dipole magnets designed to reach 8.33 T, corresponding to a proton beam energy of 7 TeV, with protons being the primary species used in the accelerator. The LHC is also used to collide heavy ions to study quark-gluon plasma and other topics in nuclear physics and QCD. In the case of proton–proton collisions, the accelerator chain [109] begins with Linear Accelerator 2 (LINAC2), which reaches an energy of 50 MeV. Then, successive synchrotrons known as the BOOSTER, Proton Synchrotron, and Super Proton Synchrotron with energies reaching 1.4, 25, and 450 GeV, respectively. The final acceleration to 7 TeV is performed within the LHC ring itself. Between the curved sectors lie eight straight segments, four of which host interaction points and are occupied by the main LHC experiments, ALICE, ATLAS, CMS, and LHCb. The remaining straight sections host two sites for beam cleaning and collimation,

the radiofrequency accelerator complex, and the beam dumps.

Collision Conditions

Beginning in 2010, the LHC has delivered collisions at center of mass energies $\sqrt{s} = 7$–13.6 TeV over the course of three major runs. The luminosity $\mathcal{L}$ delivered to the two general-purpose experiments, ATLAS and CMS, at each energy is summarized in table 3.1.

Years	Run	$\sqrt{s}$ [GeV]	$\mathcal{L}$ [fb^{-1}]
2010–2011	1	7.0	5.5
2012	1	8.0	21.3
2015–2018	2	13.0	157.4
2022–2024	3	13.6	195.0

Table 3.1: *pp* Luminosity Delivered to ATLAS by the LHC, 2010–2024 [110][111][112][113].

The LHC beams are segmented into bunches with a bunch separation of 25 ns, corresponding to a bunch crossing (collision) frequency of 40 MHz. The nominal design specifications of the beams are listed in table 3.2

Beam Energy	7000GeV
Number of Bunches	2808
Bunch Spacing	24.95 ns
Intensity per Bunch	1.15×10^{11} p/b
Beam Current	0.58 A
Transverse Emittance	3.75 μm
Longitudinal Emittance	2.5 eVs
Bunch Length (4σ)	1.0 ns
Energy Spread (4σ)	0.45×10^{-3}

Table 3.2: Nominal design specifications of the LHC beams at collision.

The LHC has exceeded its design specifications and now produces bunches with over 1.6×10^{11} protons per bunch and an instantaneous luminosity of 2.1×10^{34} cm^{-1}s^{-1}, leading to an average number of collisions per bunch crossing (pileup) of over 60, determined by limitations of the detectors and their readout systems.

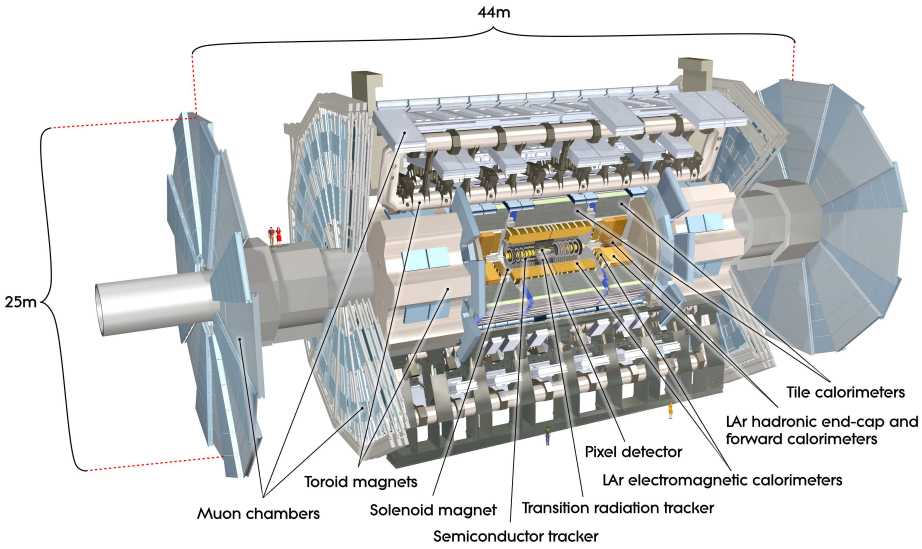


Figure 3.1: Cut-away schematic view of the ATLAS detector at the beginning of operation.

3.2 The ATLAS Experiment

The ATLAS (A Toroidal LHC ApparatuS) experiment [114] is a particle detector located at interaction point 1 at the Large Hadron Collider. It has been designed, built, and operated by the ATLAS Collaboration. The ATLAS detector is designed as a nearly hermetic general-purpose detector with a broad physics reach. The detector has a cylindrical geometry with concentric layers providing synergistic measurements of particles produced in collisions at the center of the detector. Figure 3.1 shows a diagram of the ATLAS detector. Closest to the interaction point, the Inner Detector (ID) performs tracking and momentum measurements of charged particles with a combination of silicon and gaseous tracking technologies. The tracking information is also used to identify vertices at which multiple tracks intersect. Vertexing information is used to distinguish particles originating from different pp collisions (pileup) and identify particles originating from in-flight decays of long-lived particles such as b-hadrons. The electromagnetic and hadronic calorimeters absorb electromagnetically and hadronically interacting particles, respectively, and measure their energies. Outside the calorimeters, the muon system identifies and measures the momenta of muons, which pass through the calorimeters with minimal energy loss.

ATLAS uses a coordinate system with the z-axis oriented along the beam direction, the x-axis pointing towards the center of the LHC ring, and the y-axis pointing upwards. For detector design, reconstruction, and data analysis, ATLAS adopts a convenient cylindrical coordinate system (r, η, ϕ) where (r, ϕ) are the polar coordinates in the $x - y$ plane and the pseudorapidity $\eta = -\ln\left(\tan\left[\frac{\theta}{2}\right]\right)$ is defined in terms of the polar angle θ and equal to the rapidity

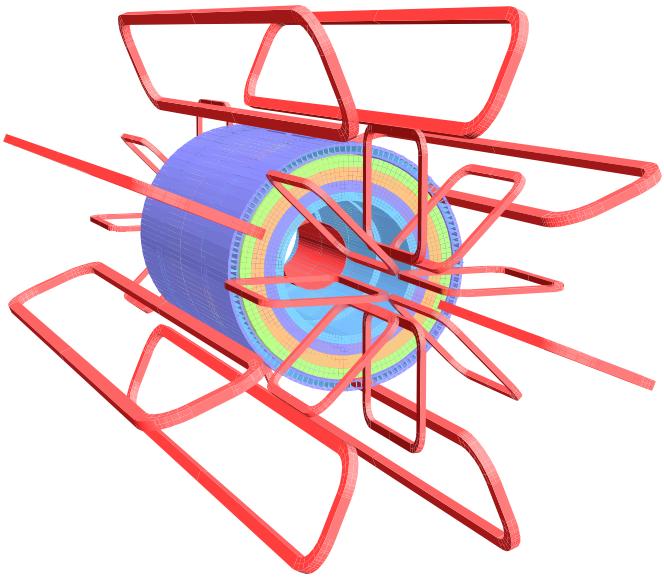


Figure 3.2: Schematic view of the ATLAS magnet system.

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

of a massless particle. Each detector system is divided into barrel and endcap components with surfaces generally oriented in the $\phi - z$ and $x - y$ planes, respectively. The barrel covers the central region, roughly corresponding to pseudorapidities $|\eta| < 1.3$, and the endcap covers the forward region with pseudorapidities of $1.3 < |\eta| < 3.2$.

Magnet Systems

ATLAS uses two large magnet systems, shown in fig. 3.2, to bend the trajectories of charged particles. Precise momentum measurement of tracks with high transverse momentum p_T requires a large magnetic field. The field in the inner detector is provided by a superconducting NbTi solenoid with an approximate field strength of 2 T. The solenoid is located between the ID and the electromagnetic calorimeter. The magnetic field in the muon system is provided by three large air-core superconducting toroidal magnets that produce a field predominantly in the ϕ direction. Separate toroid systems serve the barrel and endcap regions, reaching fields up to 2.5 and 3.5 T, respectively, in the magnet bores.

Inner Detector

The ATLAS Inner Detector [115], shown in fig. 3.3, consists of three sub-detectors, which combine to give efficient tracking and vertexing, manageable cost and data size, and a small

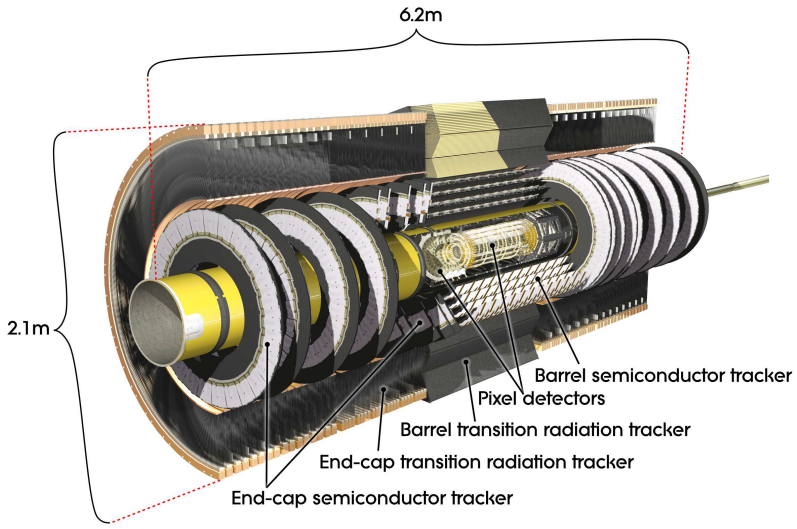


Figure 3.3: Schematic view of the ATLAS Inner Detector as installed for Run 1.

.

material budget. The detector provides tracking in the range $|\eta| < 2.5$ and typically provides at least 12 measurements for tracks in this region.

The innermost system is the silicon pixel detector [116], composed of three layers of silicon pixel sensors with a pitch of 50×400 μm^2. After Run 1, an additional layer of pixel modules, known as the Insertable B Layer (IBL) [117, 118] was added inside the innermost layer of the original Pixel system, providing a total of four measurements from the pixel system for most tracks. The IBL sensors are an average of 33.5 mm from the interaction point and about 10 mm from the beam pipe. The shorter distance to the interaction point provides improved vertex resolution at the cost of higher occupancy and radiation dose. Adding an extra layer also provides critical redundancy as the innermost layer of the original Pixel detector nears the end of its operational life due to the total ionizing dose (TID) it has received. To account for the increased particle density and improve the vertexing resolution, the IBL uses pixels with a pitch of 50×250 μm^2. In the high η regions, the IBL uses 3D sensors with reduced drift time and better performance after irradiation compared to the traditional 2D sensors used in the central region and the rest of the Pixel system.

Outside of the Pixel system sit four double layers of silicon micro-strip detectors [119] with a pitch of approximately 80 μm $\times$ 6 cm. This strip detector is known as the semiconductor tracker (SCT). Each double layer consists of pairs of modules with the sensors placed at a stereo angle of 40 mrad. Both the Pixel detector and the SCT cover the full range $|\eta| < 2.5$, and together they provide 12 position measurements for most tracks.

The outermost part of the ID is the Transition Radiation Tracker (TRT) [120], which is composed of straw tubes arranged in a honeycomb pattern. The tubes are filled with a gas mixture of 70% Xe, 27% CO_2, and 3% O_2. The straw tubes have an individual position

resolution of 130 μm from the drift time and provide discriminating power to identify tracks from electrons, which emit transition radiation when they pass through the wall of the drift tubes, resulting in a different pulse shape than minimally ionizing particles. In parts of the TRT, the noble gas has been changed from xenon to argon to mitigate the cost of leaks. This reduces the detector's electron discrimination power, but does not significantly affect the tracking performance. The TRT covers the region $|\eta| < 2.0$ and provides 30 measurements of a typical track in that region.

Calorimetry

The ATLAS calorimeter system, shown in fig. 3.4, uses sampling calorimeters with four combinations of active and absorber material to absorb and measure electromagnetic and hadronic showers. The electromagnetic calorimeter [121] sits just outside the central solenoid and covers the pseudorapidity region $|\eta| < 3.2$. The calorimeter uses a lead absorber in a novel accordion shape to ensure that all particle trajectories pass through the same fractions of active and absorber material. The active material is liquid argon (LAr) and is instrumented with copper electrodes to measure the ionization charge produced by high-energy showers. The electromagnetic calorimeter is instrumented in cells organized into three distinct layers. Each layer has differing granularity and depth, with layer 1 highly segmented to discriminate photons from $\pi^0 \to \gamma\gamma$ decays. Layer 2 is the deepest and contains the majority of the shower development, but has coarser granularity in the $\eta - \phi$ plane, whereas layer 3 is shallow and serves to estimate the energy escaping the back of the calorimeter. Electrons and photons in the electromagnetic calorimeter pass through at least 22 radiation lengths, and their showers produce minimal leakage into the hadronic calorimeter.

An additional thin layer of instrumented liquid argon with no absorber material, known as the presampler, covers the region $|\eta| < 1.8$. The presampler layer provides an estimate of energy loss in the inner detector and other material upstream of the calorimeter, including the inner detector and cryostat wall.

In the barrel region $|\eta| < 1.7$, the hadronic calorimeter [122] is composed of plates of steel absorber interleaved with plastic scintillators, which are instrumented with wavelength-shifting fibers and photomultiplier tubes. In the endcap region $1.7 < |\eta| < 3.2$, the hadronic calorimeter uses copper absorbers in a plate pattern with liquid argon active material. In the far forward region $3.1 < |\eta| < 4.9$, ATLAS uses a single calorimeter system, the Forward Calorimeter (FCal), rather than distinct electromagnetic and hadronic calorimeters. The first segment of FCal uses a copper absorber and acts as the electromagnetic calorimeter, while the final two use tungsten and largely contain hadronic showers, with all three using liquid argon as the active material.

Muon System

The ATLAS muon system [123] incorporates four distinct detector technologies to satisfy the needs of precision position and momentum measurement, fast readout and timing

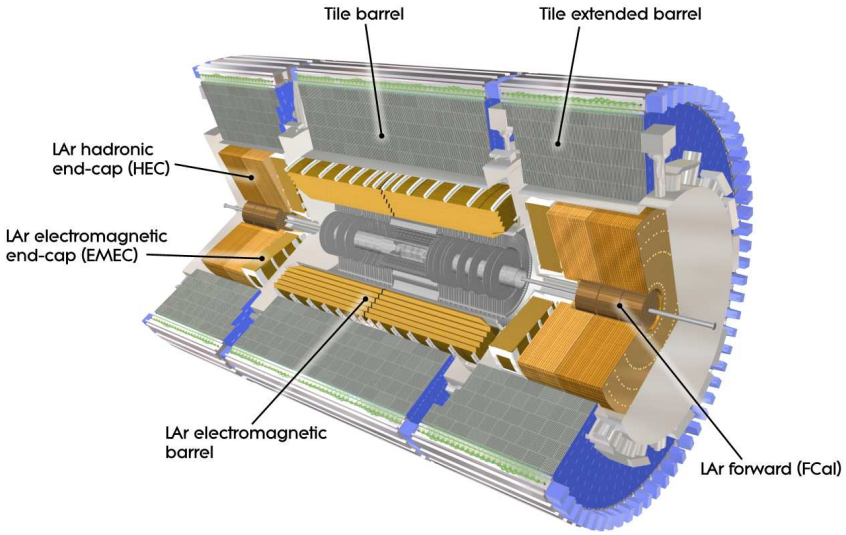


Figure 3.4: Schematic diagram of the ATLAS calorimeter systems.

information for the trigger, and robustness to high rates and backgrounds. In most of the detector, Monitored Drift Tubes (MDTs) provide precision muon tracking. In the first layer of the muon system for $|\eta| > 2.0$, this role is performed by the Cathode Strip Chambers (CSCs). Muon identification and tracking in the trigger is provided by Resistive Plate Chambers (RPCs) in the barrel ($|\eta| < 1.05$) and Thin Gap Chambers (TGCs) in the endcap ($1.05 < |\eta| < 2.7$).

The MDT chambers are composed of 30.0 mm aluminum tubes filled with an Ar/CO_2 gas mixture and a central tungsten/rhenium wire to collect ionization electrons drifting in the radial electric field. Each tube measures the distance of the track from the central wire with a precision of 80 μm. Typical MDT chambers are constructed with 1–6 m long tubes and are 1–2 m wide. They consist of one layer of three or four tubes each on the front and back sides of the chamber.

The CSCs are multiwire proportional chambers whose cathodes are segmented into strips. They use the same Ar/CO_2 gas mixture and provide a position resolution of 40 μm in the bending plane and 5 mm in the transverse plane, and are capable of handling rates of up to 1000 Hz/cm^2.

The RPCs are composed of resistive plastic plates instrumented to measure ionization in a 2 mm gap filled with an organic gas mixture allowing relatively low voltage operation. The RPCs are operated in avalanche mode, which increases the timing performance, and are oriented to provide measurements along the direction of the MDT tubes. Most muons in the central region pass through three layers of RPCs, with two layers inside the toroid system and the final layer outside it.

The TGCs are multi-wire proportional chambers with a small 1.4 mm gap between the

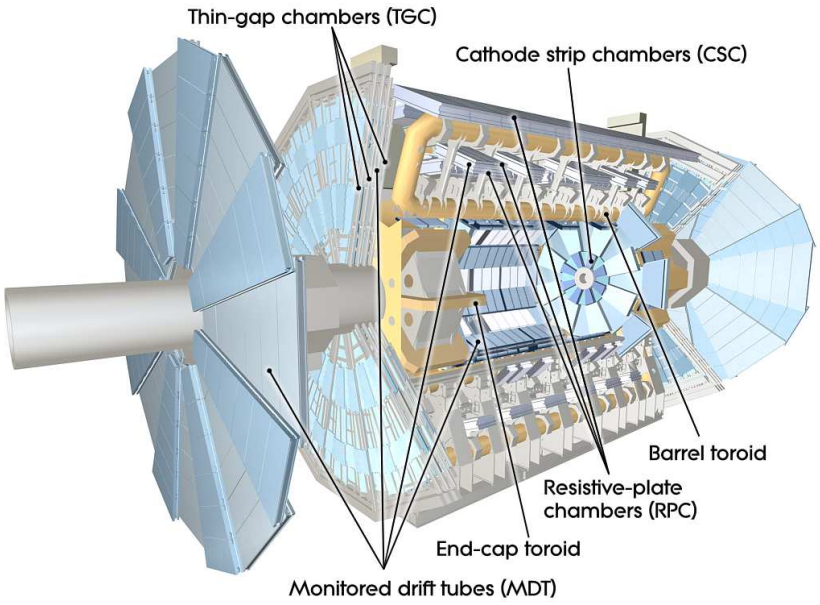


Figure 3.5: Schematic diagram of the ATLAS muon spectrometer as installed for Run 1.

wires and the cathode. This configuration gives sufficient timing performance to identify the bunch crossing in which a hit was produced with greater than 99% accuracy.

Luminosity Measurement

The luminosity is measured mainly by the LUCID–2 [124] detector that records Cherenkov light produced in the quartz windows of photomultipliers located close to the beampipe. The LUCID measurement uses an absolute calibration taken from van der Meer scans [125] performed at low pileup, measurements of overall activity in the inner detector and calorimeters, and additional corrections provided by other sources of luminosity monitoring throughout the run.

Trigger and Data Acquisition

A full collision event in ATLAS produces approximately 1.3 MB of data. Continuous readout at the 40 MHz bunch crossing frequency would therefore require a bandwidth of around 52 TB/s. This rate is infeasible for writing to disk and for the front-end electronics of the Inner Detector and the precision stations of the muon system. The elastic proton–proton cross-sections, along with the cross-sections for inelastic production of small numbers of low-energy QCD jets, are more than five orders of magnitude above those of electroweak boson or top quark production. This motivates the ATLAS trigger system, which is designed to selectively read out and store data from only those bunch crossings containing interesting

physics events. The design of the ATLAS trigger system features several levels, with each successive layer performing more complex processing and using additional readout while processing events at a lower rate and with higher latency.

The first level of the trigger system is the L1 trigger, which uses reduced granularity information from the calorimeter and muon systems to crudely reconstruct candidate muons, jets, electrons/photons, τ-leptons, missing transverse energy, and total transverse energy. This reconstruction is done by custom hardware to achieve the necessary throughput and latency. The results are sent to the central trigger processor, which compares reconstructed objects to a 'menu' of trigger selections chosen to produce a rate of up to 100 kHz with a maximum latency of 2.5 μs. This latency represents the maximum time that events must be stored in the on-detector read-out electronics before being overwritten or selected by the L1 trigger to be read out to the Readout Drivers (RODs).

The L1 trigger reconstruction is used to define regions of interest (ROIs), in which the full detector information is used by the L2 trigger. At L2, a more thorough reconstruction is performed in the ROIs from L1, taking an average 40 ms per event. The resulting objects are compared against a tighter trigger menu to reduce the event rate to 3.5 kHz. Events failing to satisfy the L2 trigger requirements are removed from the readout buffers, and those passing are reconstructed with a full-event reconstruction known as the event builder, which approximates the offline reconstruction used for data analysis. These fully reconstructed events are passed to the full software level reconstruction implemented in high-level trigger (HLT) [126], which reduces the rate to about 1.25 kHz and saves the events to disk for long-term storage.

Chapter 4

Object Reconstruction and Performance

The ATLAS Collaboration uses a common software suite [127, 128] in data simulation, in the reconstruction and analysis of real and simulated data, in detector operations, and the trigger and data acquisition systems of the experiment. Reconstruction of collision events begins with low-level detector information and produces intermediate detector objects such as tracks and clusters of energy deposits in the calorimeters before finally producing high-level physics objects representing the products of the hard-scatter collision of interest. At each stage of reconstruction, the quality of the reconstruction is studied in simulation and data to produce precise calibrations of the detector response and to model systematic uncertainties from the detector and reconstruction. The following presentation focuses on the reconstruction of the high-level physics objects used in the observation of $t\bar{t}t\bar{t}$ production described in chapters 8 to 10 and reported in Ref. [129, 130]. The reconstruction algorithms used for the Run 2 Higgs boson measurements described in chapter 11 and appendix A are largely the same, but some details vary. For a full description of those differences, see the references in the respective chapters.

4.1 Tracking and Vertex Finding

Charged particle tracking begins with the measurement of space points from clusters of hits in the pixel and SCT detectors and drift circles from hits in the TRT. Track seeds are formed from combinations of at least three space points in the pixel and innermost SCT layer, and are then extended in an iterative process that adds matching hits in successively more distant layers and refits the tracks, ending in matching to the TRT. A second pass of tracking starts with the TRT and works inwards, using only hits that have not yet been assigned to a track. Tracks are required to have a $p_T > 500$ MeV and $|\eta| < 2.5$. Additionally, they must have at least 9 (11) hits in the silicon tracker for tracks with $|\eta| \leq (\geq)1.65$, have at most one hit shared with another track, and have at most 0 (2) missing hits in pixel (SCT)

modules intersected by the track.

Reconstructed tracks are also used to identify the precise locations of the pp collisions in a bunch, known as primary vertices (PVs), within the region where the two beams overlap, known as the beamspot. The PVs are fit with an iterative procedure that assigns tracks to nearby vertices and refits the positions of those vertices using the assigned tracks and known beam-spot location before re-performing the assignment procedure. The primary vertex of the collision event of interest is identified as the vertex with the highest $\sum p_T^2$ [131], where the sum is over all tracks assigned to the vertex. This track-based vertexing algorithm can fail for $H \to \gamma\gamma$ final states, which produce low multiplicities of charged particles. Analyses targeting this final state, including the $t\bar{t}H,\ H \to \gamma\gamma$ measurement described in appendix A, instead use a multivariate approach [132] incoporating directional information for the photons measured by the electromagnetic calorimeter.

4.2 Muon Reconstruction

Muon candidates are reconstructed from a combination of charged particle tracks in the Inner Detector and muon segments reconstructed by a tracking procedure in the muon system [133]. Tracks in the muon system are reconstructed by first identifying track segments in the individual stations and using the constraint from the interaction point to combine segments into track candidates. The track candidates are then fit with a global χ^2 fit including the full effects of the magnetic field and detector material, as in the case of Inner Detector tracks. Muon candidates are divided into several categories depending on the signatures they contain, ranging from the cleanest reconstruction of a muon segment matched to an ID track to a calorimeter deposit characteristic of a minimally ionizing particle and matched to an ID track. Each of these is subject to different quality cuts used to define the universal identification thresholds.

Reconstructed muons are required to have $|\eta| < 2.5$ to match the acceptance of the inner detector and $p_T > 15$ GeV to reduce contamination from fake and non-prompt muons. They are additionally required to meet a 'medium' identification criterion [133] based on the track fit quality and number of missing hits in the precision muon stations. The total efficiency of the muon reconstruction and identification is around 90% in the $t\bar{t}t\bar{t}$ analysis phase space.

4.3 Electron and Photon Reconstruction

Electron and photon (e/γ) reconstruction begins with the identification of energy deposits in the electromagnetic (EM) calorimeter [134] that are not accompanied by a significant hadronic component. ATLAS uses the topo-cluster algorithm to build calorimeter energy clusters by identifying calorimeter cells with energy deposits more than four standard deviations above the level of the detector noise and successively adding cells with energies more than two standard deviations above the noise to define the core of the cluster. A

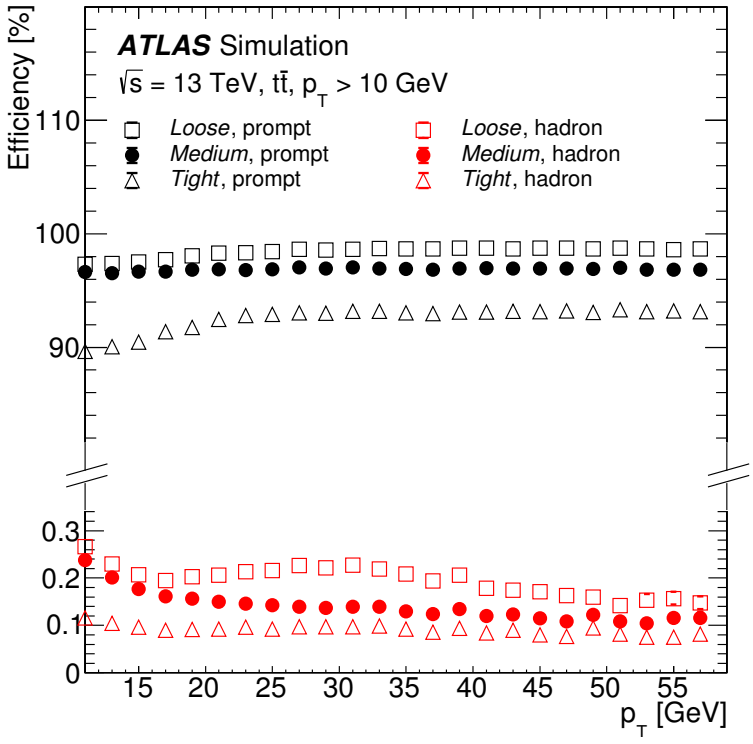
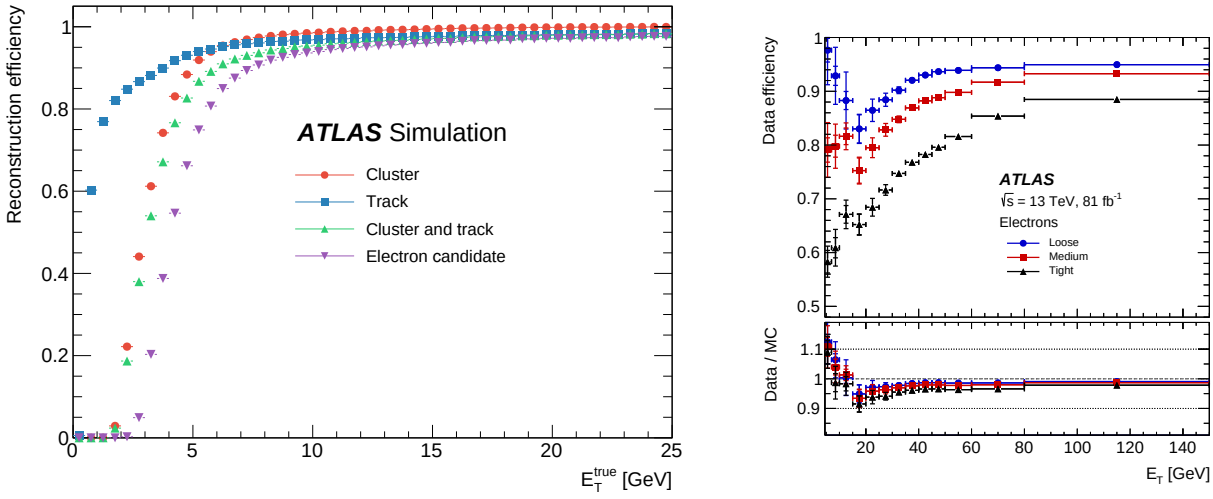


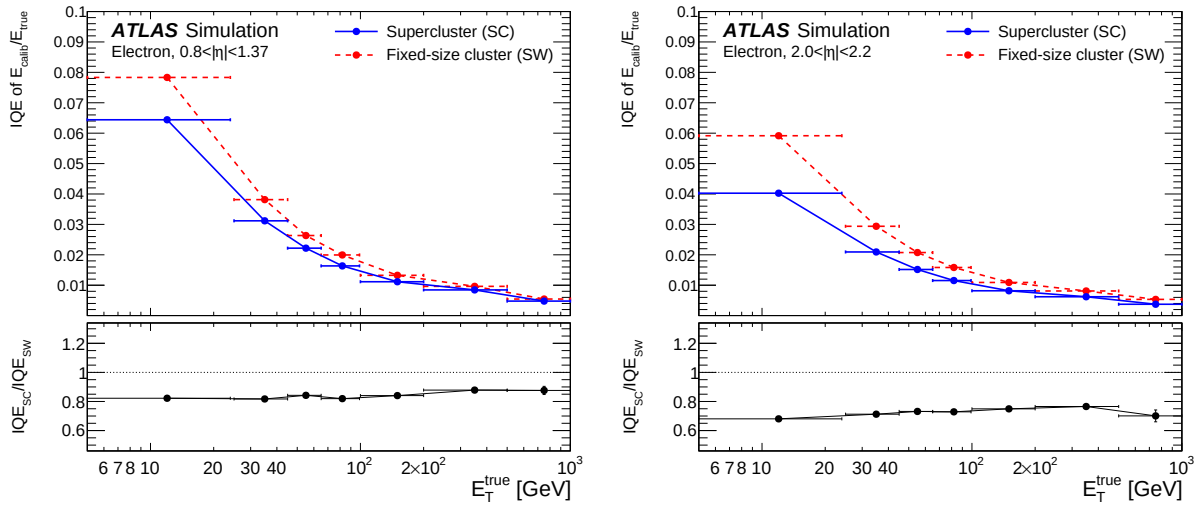
Figure 4.1: Muon reconstruction and identification efficiency as a function of p_T for muons in simulated $t\bar{t}$ events. [133].

final step adds all adjacent cells with positive energy deposition to the cluster. e/γ candidates are formed from topo-clusters in the electromagnetic calorimeter with barycenters in the acceptance of the inner detector, $|\eta| < 2.47$ for $e^{\pm}$ and $|\eta| < 2.37$ for γ, and outside of the transition region between the barrel and endcap calorimeters, $|\eta_{\text{cluster}}| \notin [1.37, 1.52]$. Electrons are identified by searching for ID tracks in a region of interest corresponding to the locations of EM topo-clusters. The matching includes a fit accounting for possible additional energy losses from electrons undergoing bremsstrahlung in the detector material. Nearby EM topo-clusters within a rectangular $(\Delta\eta, \Delta\phi)$ window of $(0.075, 0.125)$ are combined into a supercluster that is assumed to contain the majority of the electron shower. Additional EM topo-clusters in a wider $(0.125, 0.3)$ window are included if they match the electron track. These superclusters, together with the matched track, are the basis for the ambiguity resolution with the photon reconstruction algorithm, electron ID to discriminate deposits from hadrons, isolation requirements, and energy measurement. Each of these is described in Ref. [134], as are the differences in the techniques used to reconstruct photons and converted photons, which are omitted here.



(a) (b)

Figure 4.2: Efficiency for the reconstruction (left) and identification (right) of electrons as a function of p_T. [134].



(a) (b)

Figure 4.3: Electron energy resolution in simulated events with $\langle\mu\rangle = 0$ as a function of p_T for two different pseudorapidity regions. [134].

Electrons are required to have $p_T > 15$ GeV and to meet the 'tight' identification requirement based on the likelihood method [134]. The electron reconstruction has an efficiency of $> 90\%$, and the identification criteria have an efficiency of 70%–80% in the region relevant to the $t\bar{t}t\bar{t}$ analysis. The relative electron energy resolution ranges from 2%–6% in the p_T and η regions relevant to the $t\bar{t}t\bar{t}$ analysis.

New efforts to calibrate the electron and photon energies with graph neural networks or transformer algorithms show possibilities for significant improvement. The author of this thesis has led one such effort, which is detailed in appendix B.

4.4 Removal of Fake and Non-Prompt Leptons

Because the measurement of $t\bar{t}t\bar{t}$ production faces significant backgrounds from non-prompt leptons, additional requirements are placed on electron and muon candidates to reduce contamination from these backgrounds. To ensure that lepton candidates are consistent with prompt production at the primary vertex, the significance of their transverse impact parameter $|d_0|/\sigma(d_0)$ must be less than five (three) for electron (muon) candidates and the longitudinal impact parameter must satisfy $|z_0 \sin\theta| < 0.5$ mm. They are also required to satisfy the tight cut of a BDT known as the prompt lepton isolation veto (PLIV) [135] that has been trained to identify leptons from non-prompt sources. PLIV is trained using tracks and calorimeter energy deposits in a cone around the lepton candidate and a jet constructed from the tracks within the cone. The most discriminating variables are the matching of the lepton to the jet, the outputs of the b-tagging algorithms on the jet, and the isolation of the lepton track. Electrons are also required to pass a BDT discriminant trained to reject electron candidates with wrongly identified charges [134] based on the quality of the electron track and how well it matches the calorimeter energy cluster. The BDT, known as ECIDS, uses the electron's p_T, η, $q \times d_0$, E/p, the angle between the cluster and track positions, and the average charge deposited in each layer of the ID as inputs. An additional selection is implemented to identify electrons from unidentified photon conversions in material and low mass electron pairs produced from electromagnetic radiation γ^* on the hard scatter process. This selection, known as DFCAA, is used to define control regions and estimate these backgrounds. The DFCAA selection looks for a second track near the track associated with an electron candidate, but which does not satisfy the quality requirements to reconstruct material conversions. If there is a second track sharing a vertex with the electron track and the di-track mass is less than 100 MeV, the electron is identified as consistent with a material conversion or γ^*. If the vertex has a radius $r > 20$ mm, it is considered consistent with material conversion of a photon, otherwise, it is considered consistent with γ^*.

Electron and muon candidates satisfying these requirements are known as 'tight' leptons. An additional category, known as 'loose' leptons, uses relaxed identification and isolation requirements. Loose electrons use the 'medium' likelihood identification criteria [134] and the loose isolation criterion. Loose muons also use the relaxed isolation requirement. In the measurement of $t\bar{t}t\bar{t}$ production, tight leptons are generally used for positive selection of

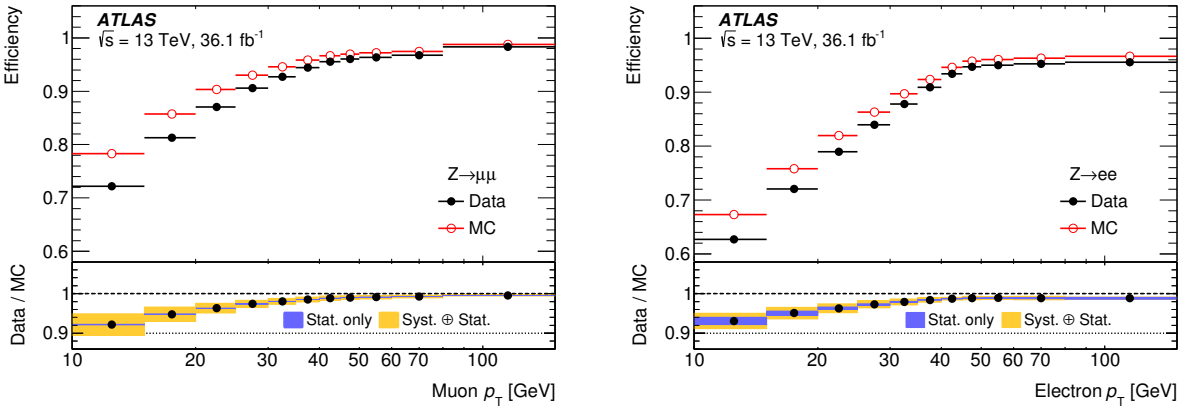


Figure 4.4: Efficiency to select well-identified prompt muons (left) and electrons (right) at the chosen PLIV working point, as a function of the lepton p_T. The measurements in data (simulation) are shown as full black (open red) circles. The bottom panel displays the ratio of data to simulation results, with the blue (yellow) band representing the statistical (total) uncertainty. This ratio is the scale factor applied to correct the simulation. [135]

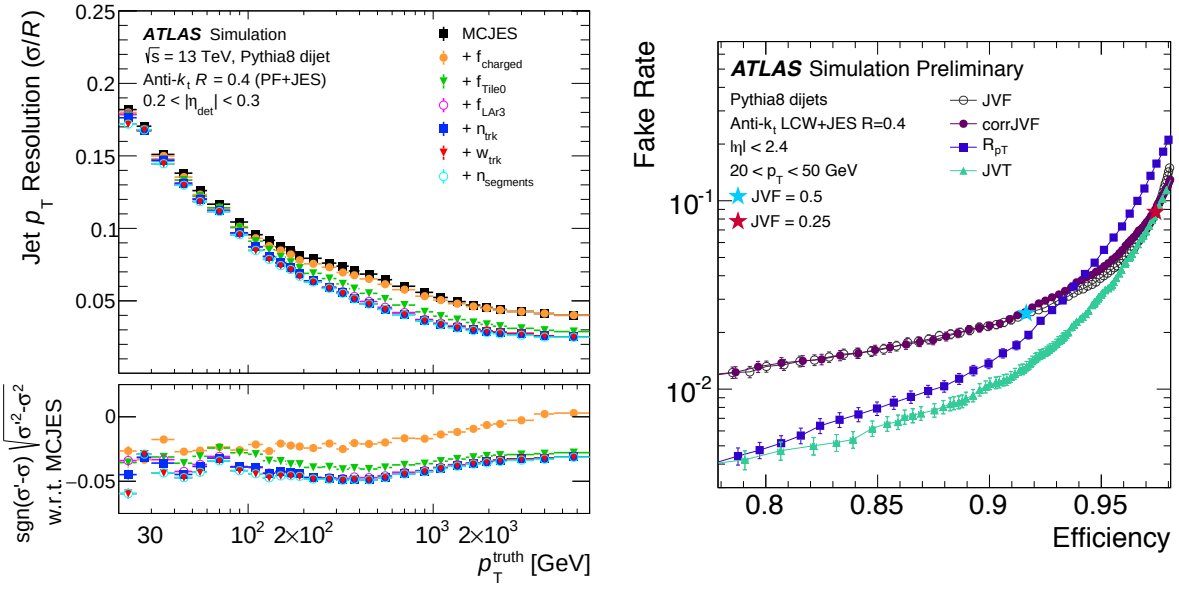
events with leptonic final states, and loose leptons are used as vetoes, including for overlap removal.

4.5 Jet Clustering and Flavor Tagging

ATLAS uses an anti-k_T [29, 136] jet clustering algorithm with a standard radius parameter $R = 0.4$ and performs the jet clustering using unified flow objects (UFOs) following a particle flow algorithm [137, 138]. In this algorithm, the four-vector inputs to the jet clustering algorithm consist of both tracks from the ID and calorimeter energy clusters reconstructed with the topo-cluster algorithm in both the hadronic and electromagnetic calorimeters. When tracks and calorimeter deposits overlap, a subtraction algorithm is applied.

The anti-k_T algorithm is chosen for IRC safety and its tendency to create compact jets with consistent sizes, which are easier to correct for effects such as energy deposits from pileup collisions. A jet-vertex tagger (JVT) is used to remove jets originating from pileup interactions with $|\eta| < 2.4$ and $p_T < 60$ GeV [139]. Jets are required to have pseudorapidity $|\eta| < 2.5$ and $p_T > 20$ GeV. The jet energy calibration is described in Ref. [138].

b-hadrons produced in the hadronization of b-quarks from top quark decays typically travel a few millimeters before decaying. Jets containing b-hadrons (b-jets) are identified (b-tagged) with a deep neural network algorithm known as DL1r [140, 141] that uses several features to identify b-jets, most notably the presence of a secondary vertex resulting from the displaced decay of the b-hadron. The DL1r algorithm takes as its inputs the outputs of several



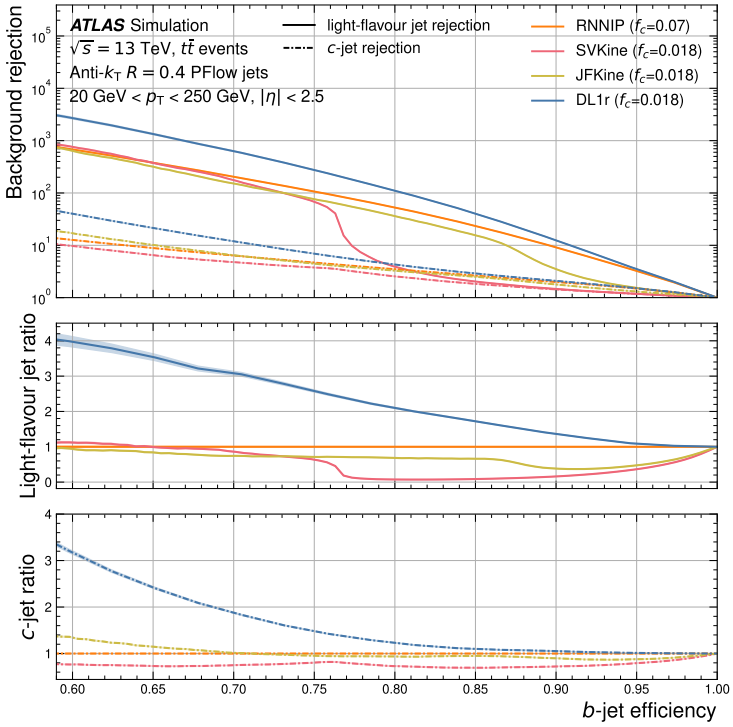
(a) (b)

Figure 4.5: Right: Jet energy resolution as a function of p_T after various corrections. Left: Fake rate versus efficiency for the pileup rejection algorithm JVT. The light blue star shows the working point used in this work. [138, 139]

low-level jet taggers, `IP2D`, `IP3D`, `SV1`, `JetFitter`, and `RNNIP`. `IP2D` and `IP3D` are classical likelihood ratio discriminants based on the impact parameters of tracks contained in the jet, `RNNIP` is a recurrent neural network trained on the track parameters of the constituent tracks, `SV1` is a classical algorithm for finding secondary vertices based on a χ^2 fit, and `JetFitter` is a topological secondary vertexing algorithm. The output of the DL1r algorithm is a real number representing the probability that the jet is a b-jet, but the calibration procedure is only performed for cuts at specific probabilities corresponding to b-tagging efficiencies of 85%, 77%, 70%, and 60% in $t\bar{t}$ events, resulting in a series of four progressively tighter cuts. A jet may be assigned an integer from one to four corresponding to the tightest b-tagging working point passed by the jet, or zero if it passes none. This integer is known as the pseudo-continuous b-tagging (PCBT) score.

4.6 Missing Transverse Energy

Neutrinos from top quark decays do not interact with the detector material and are reconstructed using the momentum imbalance in the transverse plane. The missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed as the negative vector sum of the transverse momenta of all fully reconstructed and calibrated objects associated with the primary vertex and an



(a)

Figure 4.6: Background rejection efficiency of the DL1r b-tagging algorithm and the low-level taggers used as inputs to DL1r as a function of the chosen b-jet efficiency. The ratio panels show the improvement with respect to using only the `RNNIP` tagger. [140]

additional term to account for ID tracks associated with the primary vertex but not to any physics object [142].

4.7 Overlap Removal

A sequential overlap removal procedure [60] is used to disambiguate objects reconstructed as candidates for multiple types of final state object from the same detector signatures. The order of the overlap removal is chosen to prefer the particle most consistent with the observed signature. First, electrons with an ID track matching a muon candidate are removed, as the existence of a matching track in the muon system makes the object unlikely to be an electron. The jet clustering uses all ID tracks and calorimeter energy clusters, and therefore isolated leptons are expected to be reconstructed both as jets and as leptons. Additional overlap removal seeks to remove jets that consist only of isolated leptons, as jets are meant to be proxies for hadronic activity, and to remove leptons that are part of hadronic jets and result from hadron decays. Jet-related overlap removal uses $\Delta R_y = \sqrt{\Delta\phi^2 + \Delta y^2}$, which is distinct

from the normal definition of ΔR based on the pseudorapidity η rather than the rapidity y. The next overlap removal step removes the jet nearest each electron if the electron and jet have $\Delta R_y < 0.2$. Any remaining electrons with $\Delta R_y < 0.4$ of a jet are then removed. Jets within $\Delta R_y < 0.2$ of a muon are removed if they contain fewer than three tracks. Finally, muons are removed if they are within $\Delta R_y < 0.4 + 10 \text{ GeV}/p_T^\mu$ of a jet.

Chapter 5

Data and Monte Carlo Simulation

This chapter presents the collision dataset and samples of simulated collision events used in the observation of $t\bar{t}t\bar{t}$ production [129]. Other measurements discussed in chapter 11 and appendix A are based on collisions from the same period, but rely on different triggers. The simulated data samples differ substantially, and the interested reader is encouraged to read about them in the respective publications detailing those analyses.

5.1 Run 2 Collision Dataset

The Run 2 *pp* collision dataset collected by the ATLAS detector from 2015 to 2018 amounts to an integrated luminosity of 140 fb^{-1} at a center of mass collision energy $\sqrt{s} =$ 13 TeV from periods with stable operation of all major detector systems. The uncertainty in the combined 2015–2018 integrated luminosity is 0.83% [125], obtained using the LUCID-2 detector [124] for the primary luminosity measurements, complemented by measurements using the inner detector and calorimeters. The difference from the delivered luminosity of 157.4 fb^{-1} comes from collisions not recorded during the high voltage ramp of the ID and during the trigger dead time introduced to account for limitations in detector readout electronics. This figure also accounts for the removal of periods with detector issues or other data defects. The integrated luminosity during which the detector was ready to record physics data, known as the recorded luminosity, is 145.5 fb^{-1}. The data quality efficiency is evaluated as the ratio of the recorded luminosity to the luminosity that produced data deemed good enough for physics analysis and is 95.6%. The data quality requirements are detailed in Ref. [143].

The average number of collisions per bunch crossing (pileup) is $\langle\mu\rangle = 13.4$ for 2015, $\langle\mu\rangle = 25.1$ for 2016, $\langle\mu\rangle = 37.8$ for 2017, $\langle\mu\rangle = 36.1$ for 2018, and $\langle\mu\rangle = 33.7$ for the whole dataset.

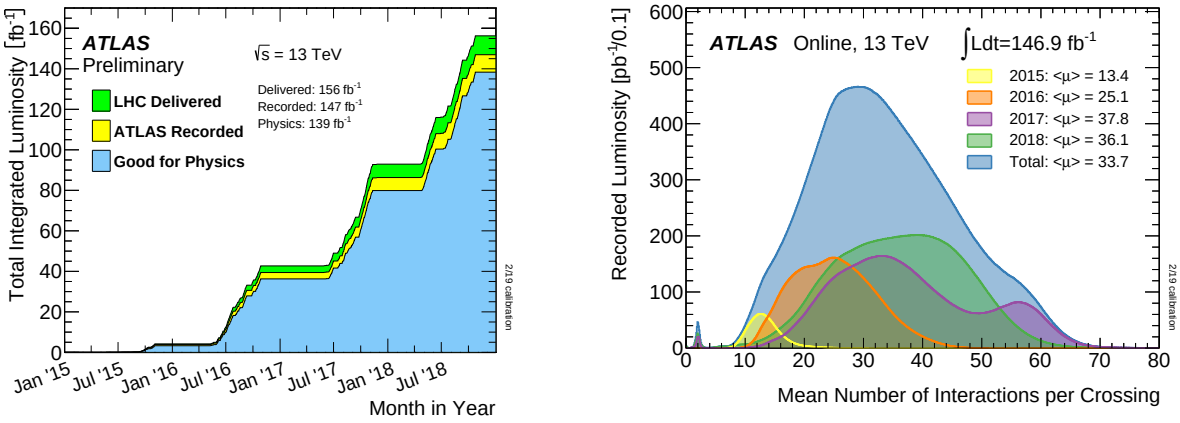


Figure 5.1: (Left) Integrated luminosity recorded by the ATLAS detector over time during Run 2 compared to the total delivered luminosity and the luminosity passing data quality requirements. The total luminosity shown in this plot differs from the text because it is based on a preliminary calibration of the luminosity measurement. The text shows the final value. (Right) Distribution of the mean number of interactions per bunch crossing for different data-taking periods and inclusively for the Run 2 dataset.

Triggers

Events for the SSML $t\bar{t}t\bar{t}$ measurement are triggered by a combination of single and dilepton triggers, which require one of one electron, two electrons, one muon, two muons, or one electron and one muon [126, 144, 145, 146]. Each trigger has its own identification and isolation requirements and thresholds for the p_T of the lepton(s), which vary somewhat between data-taking periods. The trigger thresholds for different data-taking periods, conditions, and identification and isolation criteria are detailed in table 5.1 and table 5.2. The criteria implemented in the trigger logic are similar, but not identical, to those described in chapter 4.

The trigger thresholds are generally higher in later data-taking years as higher pileup requires a higher trigger threshold to maintain the same trigger rate. The electron trigger reconstruction and efficiency are discussed in Ref. [144], and the muon trigger reconstruction and efficiency are discussed in Ref. [145]. In addition to passing the trigger, the offline reconstruction must return objects matching those that fired the trigger.

5.2 Monte Carlo Simulation of pp Collisions

The $t\bar{t}t\bar{t}$ signal and SM background processes are modeled with Monte Carlo (MC) simulation. In addition to the simulation of the hard scatter collision based on the perturbative matrix element, additional steps simulate the underlying event (UE), parton showering (PS),

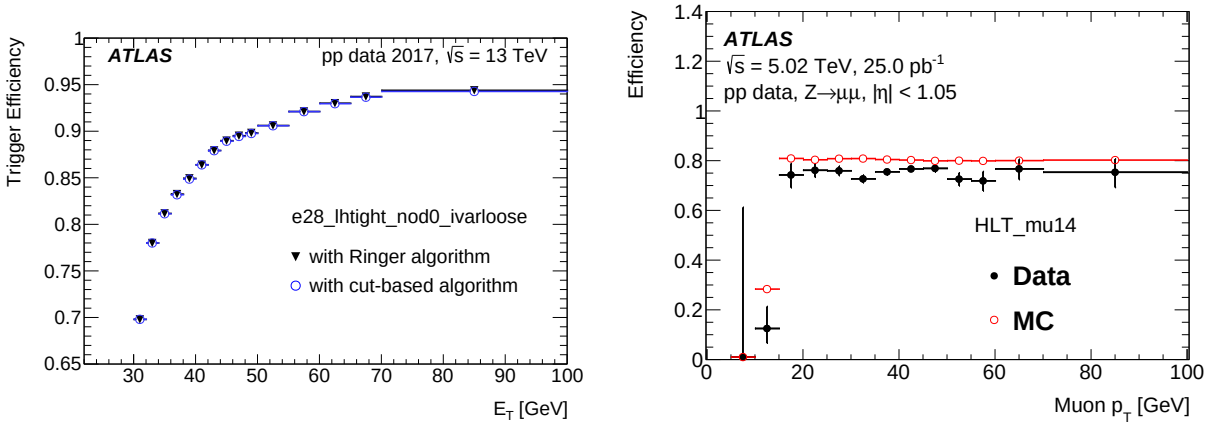


Figure 5.2: Left: Characteristic efficiency of the single electron triggers [144]. Right: Characteristic efficiency of the single muon triggers [145].

Year of data-taking	ID Working Point	HLT p_T Threshold [GeV]
single electron triggers		
2015	Medium ID	24
2015	Medium ID	60
2015	Loose ID	120
2016–2018	Tight ID + Loose Isolation	26
2016–2018	Medium ID	60
2016–2018	Loose ID	140
single muon triggers		
2015	Loose	20
2015	None	50
2016–2018	Medium	26
2016–2018	None	50

Table 5.1: List of single lepton triggers used in the analysis per data period.

and the interactions with the ATLAS detector and corresponding detector signals (detector simulation). Table 5.3 lists these tools. Pileup collisions are included by overlaying inclusive pp collisions simulated with PYTHIA8. The Pythia parameters are tuned to the A3 tune [147]. The simulations of the $t\bar{t}t\bar{t}$ signal and major background processes, including $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}H$, and $t\bar{t}$, are all performed at next-to-leading-order (NLO) in the strong and electroweak coupling constants. Matrix element calculations use the best available measurements of the

Year of data-taking	ID/Isolation Working Point	HLT p_T Threshold [GeV]
di-electron triggers		
2015	Loose	12 (e_1 and e_2)
2016	Loose	17 (e_1 and e_2)
2017–2018	Loose	24 (e_1 and e_2)
electron-muon triggers		
2015	Loose (e only)	17 (e) and 14 (μ)
2015	Medium (e only)	7 (e) and 24 (μ)
2016–2018	Loose (e only)	17 (e) and 14 (μ)
2016–2018	Medium (e only)	7 (e) and 24 (μ)
di-muon triggers		
2015	None	10 (μ_1 and μ_2)
2015	None	18 (μ_1) and 8 (μ_2)
2016–2018	None	14 (μ_1 and μ_2)
2016–2018	None	22 (μ_1) and 8 (μ_2)

Table 5.2: List of dilepton triggers used in the analysis per data period.

SM parameters. The Higgs boson and top quark masses are the most relevant as they are measured with less precision, and they set the relevant scales for both the signal and major background processes. They are set to $m_h = 125$ GeV and $m_t = 172.5$ GeV, respectively. Photon emission is modeled using a resummed calculation at the leading-logarithm level in either the parton shower generator or PHOTOS. The interactions of detector stable particles with the detector material and the resulting detector signals are simulated with GEANT4, though some alternative samples use a faster parameterized simulation of calorimeter showers [148] known as Atlfast II (AFII). Differences between the full and fast simulations are accounted for in the calibration of electron and jet resolutions, scales, and efficiencies. The remaining differences after these corrections are negligible.

Simulating collisions with the effects of BSM physics included requires the provision of information about the new particles and their interactions to the matrix element generator. This can be achieved with a Universal FeynRules Output (UFO) file. Of particular relevance to this work are the Higgs Characterization and SMEFT@NLO models, which are UFO implementations of theories tested by the $t\bar{t}t\bar{t}$ measurement. The Higgs Characterization Model allows matrix element generation with different hypotheses about the Higgs boson's properties, including different values of the top Yukawa coupling and associated CP mixing angle α. SMEFT@NLO implements the effects of dimension-six operators as enumerated in the SMEFT framework.

Tool	Purpose(s)	Version(s)	Reference(s)
Pythia8	Matrix Element, Parton Shower, and Fragmentation	8.186, 8.210, 8.230	[149]
PowhegBox	Matrix Element	—	[150, 151, 152, 153]
MadGraph5_aMC@NLO	Matrix Element	2.3.3, 2.6.2, and 2.8.1	[30, 154]
Sherpa	Matrix Element, Parton Shower	2.2.1, 2.2.2, 2.2.10, and 2.2.11	[155, 156]
Herwig7	Parton Shower	7.04	[157, 158]
EvtGen	Heavy Flavor Decays	1.2.0, 1.6.0	[159]
PHOTOS	Photon Emission	—	[160]
Geant4	Detector Simulation and Digitization	—	[161]
FeynRules	Interfacing Physics Models with Generators via UFO Format	—	[162, 163]
Higgs Characterization Model	UFO Model for Tree-Level Modifications to Higgs Couplings	—	[164]
SMEFT@NLO	UFO Model for BSM Effects Parameterized by Dimension 6 Operators	—	[165]
OpenLoops	Virtual QCD Corrections for $t\bar{t}W$	2	[166]
Comix	Matrix Elements with additional colored particles in the final state	—	[167]
MadSpin	Top Quark Decays including Full Spin Correlations	—	[168]

Table 5.3: Summary of software packages used in Monte Carlo simulations for [129].

Different combinations of tools are used in the simulation of each process according to the practice used in dedicated studies of that process. Large processes are simulated multiple times using different combinations of generators and parton showering tools to estimate the modeling uncertainties resulting from the different approximations made in these tools. To increase the accuracy of the simulation, different values of the renormalization

and factorization scale are chosen for each process according to the energy scale of the hard scatter, with variations included for major processes to provide an uncertainty estimate. Parton distribution functions are chosen to match the order of the matrix element calculation in α_s and include NNPDF3.1NNLO, NNPDF3.0NNLO, NNPDF2.3LO, and MMHT2014 LO.

Samples generated with POWHEGBOX and MADGRAPH5_aMC@NLO use PYTHIA8 for parton showering, fragmentation, and UE with the A14 tune [169] and samples generated with HERWIG7 use the H7UE tune [158]. Samples with fragmentation from PYTHIA8 or HERWIG7 use EvtGen to decay heavy-flavor hadrons. A summary of all MC samples used in Ref. [129], including the choices of generator, parton shower, and PDF set, is provided in section 5.4.

5.3 $t\bar{t}t\bar{t}$ Signal Monte Carlo Simulation

Standard Model $t\bar{t}t\bar{t}$

The SM production of $t\bar{t}t\bar{t}$ events is modeled with MADGRAPH5_aMC@NLO with renormalization and factorization scales $\mu_R, \mu_F = m_T/4$ [41] where $m_T = \sum \sqrt{p_T^2 + m^2}$ is the sum of transverse masses of all particles generated from the matrix element. Generator uncertainties are estimated from a SHERPA 2.2.11 sample with $\mu_R, \mu_F = \hat{H}_T/2$, and parton showering uncertainties are estimated from an alternative sample with MADGRAPH5_aMC@NLO interfaced to HERWIG7. A final alternate simulation of SM $t\bar{t}t\bar{t}$ production is performed at leading-order (LO) in the strong coupling constant. Unlike the NLO samples, the LO sample does not contain carry negatively weighted events and is therefore useful for the GNN training described in chapter 9.

BSM Signal Samples

Possible BSM modifications to $t\bar{t}t\bar{t}$ production are parameterized by Wilson coefficients or coupling modifiers in the case of EFT and top Yukawa interpretations, respectively. To accommodate the effects of these modifications on the cross-section and kinematics of $t\bar{t}t\bar{t}$ production, including changes to the acceptance and efficiency, alternate simulations are performed with the effects of BSM physics at several points in parameter space included. These samples use the fast AFII detector simulation and are generated at leading order due to the limitations of the underlying models. Each set of samples includes an SM reference point. The effects of BSM physics are then parameterized as linear modifications to the nominal NLO $t\bar{t}t\bar{t}$ sample derived from the ratios of yields in samples generated at LO with and without BSM effects. This procedure is described with the statistical methods in chapter 6.

Modifications to the top Yukawa coupling are modeled with the Higgs characterization model. Six samples are generated with coupling strength κ_t and CP mixing angle α listed

in table 5.4.

k_t	α	κ_t	$\tilde{\kappa_t}$	σ (fb)	Number of Events
1	0°	1	0	6.8	15M
2	0°	2	0	16.0	15M
1	90°	0	1	10.5	15M
0.5	0°	0.5	0	6.8	15M
2	90°	0	2	36.0	15M
1	45°	0.707	0.707	8.6	15M

Table 5.4: Alternate top Yukawa coupling values used for BSM $t\bar{t}t\bar{t}$ MC simulations.

Modifications from heavy BSM states parameterized by the Standard Model Effective Field Theory (SMEFT) are simulated using the SMEFT@NLO package, which interfaces with MadGraph5_aMC@NLO. The generation is performed only once, and the resulting events are reweighted to obtain samples for different values of the Wilson coefficients. The weights needed to reweight the sample to each of the chosen signal points are stored during the generation. The operators considered, $\mathcal{O}^1_{QQ}$, $\mathcal{O}^1_{Qt}$, $\mathcal{O}^1_{tt}$, and $\mathcal{O}^8_{Qt}$, are four-fermion operators affecting only the third generation of quarks and are defined in chapter 2. The chosen set of signal points is listed in table 5.5. The generated grid uses all possible combinations of Wilson coefficient values from each column, resulting in $5^4 = 625$ different signal points.

$\mathcal{O}^1_{QQ}$	$\mathcal{O}^1_{Qt}$	$\mathcal{O}^1_{tt}$	$\mathcal{O}^8_{Qt}$
0	0	0	0
1	0.8	0.5	1.5
2	1.5	2	3
4	3	6	6
8	6	4	12

Table 5.5: Wilson coefficient values used to generate signal samples with effects from four fermion SMEFT operators.

Modifications from changes to the Higgs oblique parameter $\hat{H}$ are modeled similarly to those from the top Yukawa coupling using the UFO model of Ref. [55] and MadGraph5_aMC@NLO. The Samples are generated at a grid of signal points with $\hat{H} \in [0, 0.05, 0.08, 0.12, 0.17]$.

5.4 Background Monte Carlo Simulation

$t\bar{t}W$ Simulation

The nominal modeling of $t\bar{t}W$ comes from Sherpa 2.2.10. The matrix element calculation includes diagrams with up to one additional final state parton at NLO and up to two additional partons at LO. The matrix element calculation is done with COMIX and OpenLoops and is merged with the Sherpa parton showering using the MEPS@NLO prescription [170] at a scale of 30 GeV. The renormalization and factorization scales are set to $\mu_R, \mu_F = m_T/2$. This sample includes terms up to $\mathcal{O}(\alpha\alpha_S^3)$. Corrections accounting for virtual electroweak corrections at $\mathcal{O}(\alpha^2\alpha_S^2)$ and LO electroweak production $\mathcal{O}(\alpha^3)$ are applied in the form of event-by-event weights [171, 155, 172].

An additional sample generated at LO in Sherpa 2.2.10 includes sub-leading NLO EW effects at $\mathcal{O}(\alpha^3\alpha_S)$. Samples at NLO in QCD and EW are combined according to the method of Ref. [173]. The resulting sample is normalized to a cross-section of 722 fb to match Ref. [173].

Uncertainties from approximations made in the MC generator are derived from an alternative sample generated with MadGraph5_aMC@NLO. A similar strategy is used, with separate samples at NLO in the strong and electroweak coupling constants. The NLO QCD sample is generated with different jet multiplicities and merged with the FxFx scheme [30] at a scale of 30 GeV.

$t\bar{t}Z$, $t\bar{t}H$, and $t/tW/t\bar{t}$ Simulation

The $t\bar{t}Z$ simulation includes off-shell $Z \to \ell^+\ell^-$ production in association with top quark pairs and begins at $m_{\ell\ell} > 1$ GeV. The cross-section normalization follows Ref. [21]. The nominal sample is generated with MadGraph5_aMC@NLO, and an alternative sample to estimate the generator uncertainty is generated with Sherpa 2.2.10.

$t\bar{t}H$ is simulated with PowhegBox as the nominal generator and alternative samples from MadGraph5_aMC@NLO as the generator to estimate the generator uncertainty, or Herwig7 for parton showering to estimate the uncertainty from the parton shower modeling. The alternate samples both use the AFII detector simulation. An additional alternate sample is generated for the BSM scenario where the top Yukawa coupling is purely CP Odd ($\kappa_t = 1$, $\alpha = 90°$). This sample is generated with MadGraph5_aMC@NLO using the Higgs characterization model [164].

$t\bar{t}$ is modeled with a combination of Powheg samples generated in slices of H_T to cover the whole phase space studied in the analysis with sufficient statistical power, including the high H_T $t\bar{t}t\bar{t}$ signal region. The cross-section is normalized to the next-to-next-to-leading-order (NNLO) calculation including the resummation of soft gluon radiation at the next-to-next-to-leading-logarithm (NNLL) level [174, 175, 176, 177]. The single top production generated at NLO in QCD using Powheg with the five-flavor scheme for tW and s-channel

production and the four-flavor scheme for t-channel production. Overlap with the $t\bar{t}$ simulation is removed according to the diagram removal scheme [178].

Other Backgrounds

$t\bar{t}t$, including both $t\bar{t}tq$ and $t\bar{t}tW$, is simulated at LO in MadGraph5. Use of the five flavor scheme prevents overlap with $t\bar{t}t\bar{t}$. The cross-section is normalized to the NLO calculation. Remaining small backgrounds with simulated datasets include tWZ, tZq, $t\bar{t}VV$, V, VV, VVV, and VH. The details of these samples, together with those described above, are listed in section 5.4. The LO simulations of $t\bar{t}WW, t\bar{t}ZZ, t\bar{t}WZ$, and $t\bar{t}HH$ production are normalized to the NLO cross-sections [154] [21]. To accurately capture vector boson association in conjunction with high jet multiplicity, LO simulations of V, VV, and VVV production with up to four, three, and two additional parton emissions are used to complement the NLO simulations with up to one additional parton emission.

Process	Generator	ME order	PDF	Parton shower	Tune
$t\bar{t}t\bar{t}$	MadGraph5_aMC@NLO	NLO	NNPDF3.1nlo	Pythia8	A14
	(MadGraph5_aMC@NLO)	(NLO)	(MMHT2014 LO)	(Herwig7)	(H7UE)
	(Sherpa 2.2.11)	(NLO)	(NNPDF3.0nnlo)	(Sherpa)	(Sherpa)
	(MadGraph5)	(LO)	(MMHT2014 LO)	(Pythia8)	(A14)
$t\bar{t}t\bar{t}$ κ_t	MadGraph5	LO	MMHT2014 LO	Pythia8	A14
$t\bar{t}t\bar{t}$ EFT	MadGraph5	LO	MMHT2014 LO	Pythia8	A14
$t\bar{t}t$	MadGraph5	LO	NNPDF2.3lo	Pythia8	A14
$t\bar{t}W$	Sherpa 2.2.10	MEPS@NLO	NNPDF3.0nnlo	Sherpa	Sherpa
	(MadGraph5_aMC@NLO)	(FxFx)	(NNPDF2.3lo)	(Pythia8)	(A14)
$t\bar{t}W$ EW	Sherpa 2.2.10	LO	NNPDF3.0nnlo	Sherpa	Sherpa
	(MadGraph5)	(LO)	(NNPDF2.3lo)	(Pythia8)	(A14)
$t\bar{t}Z$	MadGraph5_aMC@NLO	NLO	NNPDF3.0nlo	Pythia8	A14
	(Sherpa 2.2.11)	(MEPS@NLO)	(NNPDF3.0nnlo)	(Sherpa)	(Sherpa)
$t\bar{t}H$	PowhegBox	NLO	NNPDF3.0nlo	Pythia8	A14
	(PowhegBox)	(NLO)	(NNPDF3.0nlo)	(Herwig7)	(H7UE)
	(MadGraph5_aMC@NLO)	(NLO)	(NNPDF3.0nlo)	(Pythia8)	(A14)
$t\bar{t}$	PowhegBox	NLO	NNPDF3.0nlo	Pythia8	A14
Single-top	PowhegBox	NLO	NNPDF3.0nlo	Pythia8	A14
tWZ	MadGraph5_aMC@NLO	NLO	NNPDF3.0nlo	Pythia8	A14
tZq	MadGraph5	LO	NNPDF2.3lo	Pythia8	A14
$t\bar{t}VV$	MadGraph5	LO	NNPDF2.3lo	Pythia8	A14
V	Sherpa 2.2.1	MEPS@NLO	NNPDF3.0nnlo	Sherpa	Sherpa
VV,VVV	Sherpa 2.2.2	MEPS@NLO	NNPDF3.0nnlo	Sherpa	Sherpa
VH	Pythia8	LO	NNPDF2.3lo	Pythia8	A14

Table 5.6: Summary of the generation setups used for the signal and background simulation. Parentheses indicate alternate samples used to evaluate systematic uncertainties. V stands for any heavy electroweak vector boson. The matrix element (ME) order is the order in QCD of the perturbative calculation. Tune is the underlying-event tune of the parton shower (PS) generator. MEPS@NLO and FxFx refer to the methods used to match the ME to the PS in Sherpa [166, 167, 179, 170] and in MadGraph5_aMC@NLO [30], respectively.

Chapter 6

Statistical Methods and Model

6.1 Profile Likelihood Method

The analyses described in this thesis use the profile likelihood method for parameter estimation, hypothesis testing, and limit setting. This method has become standard for most LHC data analysis. This section summarizes the method, and the rest of the chapter deals with the specifics of the likelihood defined for the $t\bar{t}t\bar{t}$ measurement. The interested reader is invited to consult the treatments of Refs. [180, 181, 182] for a more comprehensive treatment of the statistical methods used at the LHC.

Likelihood Definition for a Binned Measurement

The profile likelihood method is based on a likelihood in the frequentist sense, i.e., a probability distribution function $\mathcal{L}(\mathbf{x}|\vec{\theta})$ representing the probability (density) of observing the data $\mathbf{x}$ in an experiment given a model of the experiment parameterized by the parameters $\vec{\theta}$. The motivation for introducing the likelihood is the maximum likelihood estimator. The maximum likelihood estimate of a parameter μ (written $\hat{\mu}$) based on observed data $\mathbf{x}$ is

$$\hat{\mu} = \text{argmax}_{\mu} \max_{\vec{\theta}} \mathcal{L}\left(\mathbf{x}|\vec{\theta}, \mu\right).$$

In this case, the remaining parameters $\vec{\theta}$, which do not include μ, are profiled. Under appropriate assumptions, including the validity of the assumed likelihood, the maximum likelihood estimator tends to the true value of μ in the limit of infinite data and asymptotically achieves the lowest possible variance, the Cramér-Rao bound. Furthermore, it is invariant under one-to-one reparameterizations of μ and corresponds to the value of μ which maximizes the Bayesian posterior distribution under a uniform prior. The construction of the likelihood is therefore the central work of data analysis, after which the optimization procedure (known as fitting) is used to determine the parameter estimates, and standard techniques are used to place confidence intervals or set limits on the parameters. As a technical detail, the fitting procedure generally minimizes the negative log likelihood $\text{NLL} = -\ln\mathcal{L}$ rather

than maximizing $\mathcal{L}$. These are functionally equivalent, but the former is more numerically stable.

The parameters $\vec{\theta}$ may be separated into two components, the parameters of interest (POIs) $\vec{\mu}$, which are the physics observables targeted by a measurement, and the nuisance parameters $\vec{\alpha}$, which parameterize effects from systematic uncertainties. Nuisance parameters are accompanied by constraint terms representing the *a priori* knowledge of the systematic effects from calibration measurements, previous experiments, or alternative theoretical calculations. In the case of the $t\bar{t}t\bar{t}$ measurement, some nuisance parameters are elevated to be normalization factors $\vec{\mathrm{NF}}$, which parameterize the overall normalizations of some difficult-to-model backgrounds. These normalization factors, together with the parameters of the parametric function used to model the $t\bar{t}W$ jet multiplicity distribution as described in chapter 8, differ from other nuisance parameters in that they are not accompanied by a constraint function. After this split, the likelihood function is written as $\mathcal{L}(\mathbf{x}|\vec{\theta}) = \mathcal{L}(\mathbf{x}|\vec{\mu}, \vec{\alpha}, \vec{\mathrm{NF}})$.

Specializing to the case of a binned measurement, the data $\mathbf{x}$ consists of the observed number of events x_{ij} in each bin i of the histogram from each region j in the measurement. Each x_{ij} is an independent counting experiment modeled by a Poisson distribution with the expected number of events ν_{ij} parameterized by $\vec{\theta}$.

$$f(x_{ij}|\nu_{ij}) = \frac{\nu_{ij}^{x_{ij}}}{x_{ij}!} e^{-\nu_{ij}} \tag{6.1}$$

The typical case is that the expected number of events is the sum of expectations from independent contributions from signal processes s^σ and background processes b^ρ. Each of these has its own expected contribution to each bin $s_{ij}^\sigma(\vec{\theta})$, $b_{ij}^\rho(\vec{\theta})$, and the total expectation for each bin may be written

$$\nu_{ij}(\vec{\theta}) = \sum_\sigma s_{ij}^\sigma(\vec{\theta}) + \sum_\rho b_{ij}^\rho(\vec{\theta}).$$

A further specialization is possible in the common case of measuring the signal strength of a single signal process with floating normalizations for some backgrounds. In this case, the expectation becomes

$$\nu_{ij} = \mu s_{ij}(\vec{\alpha}) + \sum_\rho \mathrm{NF}^\rho b_{ij}^\rho(\vec{\alpha}). \tag{6.2}$$

This is the canonical parameterization for a binned measurement; however, this work deals with cases with multiple signal processes, POIs other than the signal strength, and background processes that depend on the POIs, so the more general form is retained. The likelihood function for the measurement in a single bin is written

$$\mathcal{L}_{ij}\left(x_{ij}|\vec{\theta}\right) = \frac{\left(\sum_\sigma s_{ij}^\sigma(\vec{\theta}) + \sum_\rho b_{ij}^\rho(\vec{\theta})\right)^{x_{ij}}}{x_{ij}!} \exp\left(-\sum_\sigma s_{ij}^\sigma(\vec{\theta}) - \sum_\rho b_{ij}^\rho(\vec{\theta})\right). \tag{6.3}$$

The constraint term associated with the nuisance parameter α_i is denoted $f_i(a_i, \alpha_i)$ where a_i, known as the global observable, represents the nominal value of the nuisance parameter α_i. The global observable is treated as the result of an auxiliary measurement that constrains the nuisance parameter. Several choices for the functional form of f_i are possible depending on the nature of the uncertainty. The chosen functional form must have $f_i(a_i, a_i) > f_i(a_i, \alpha_i)$ for all $\alpha_i \neq a_i$ to ensure closure, must have a domain of validity in α_i corresponding to the physical meaning of the nuisance parameter, and should have a variance σ^2 approximating the uncertainty in the knowledge of α_i. In the most common case, the nuisance parameter is scaled to have a mean of zero and a variance of one, and a normal distribution is chosen for the constraint f_i. The complete likelihood used in the statistical analysis is then the product of a Poisson distribution for each bin and the constraint term for each nuisance parameter.

$$\mathcal{L} = \prod_{i,j} \mathcal{L}_{ij}\left(x_{ij}\middle|\vec{\mu}, \vec{\alpha}, \vec{\mathrm{NF}}\right) \prod_{k} f_k(a_k|\alpha_k) \tag{6.4}$$

The likelihood is defined as the probability to observe the observed event counts in each bin and the observed values of the global observable given the input values of the POIs, nuisance parameters, and normalization factors via constraint terms on the nuisance parameters and parameterized yields for each relevant signal and background process in each bin of the analysis. It is understood as a function that describes, for a given observed dataset, the relative consistency of different hypotheses with that dataset. As a matter of convenience, the remainder of the chapter will leave the likelihood's dependence on the observed data implicit where possible and work only with the conditional parameters. For the $t\bar{t}t\bar{t}$ measurement, the likelihood is implemented in HistFactory [183], which produces a RooFit [184] Workspace corresponding to the likelihood function.

Profile Likelihood Ratio as a Test Statistic

The profile likelihood ratio is a function of the POIs μ and is defined as the ratio of the likelihood $\min_\alpha \mathcal{L}(\mu, \alpha)$ maximized (profiled) over the nuisance parameters α to the likelihood value from the maximum likelihood estimate $\mathcal{L}(\hat{\mu}, \hat{\alpha})$ where $\hat{\alpha} = \mathrm{argmin}_{\vec{\alpha}} \mathcal{L}(\hat{\mu}, \vec{\alpha})$.

$$\lambda_{\vec{\mu}} = \frac{\min_{\vec{\alpha}} \mathcal{L}(\vec{\mu}, \vec{\alpha})}{\min_{\vec{\mu}', \vec{\alpha}} \mathcal{L}(\vec{\mu}', \vec{\alpha})} \leq 1. \tag{6.5}$$

The interpretation of $\lambda_{\vec{\mu}}$ is the relative consistency with the observed data $\mathbf{x}$ of the hypotheses that the POIs are $\vec{\mu}$ and $\hat{\hat{\mu}}$, for which $\mathbf{x}$ is most likely to be observed. This ratio is constructed using the values of the nuisance parameters that maximize the likelihood at each point rather than using those from the overall maximum likelihood fit, which would obscure some systematic effects, or by integrating out the nuisance parameters, which would require the likelihood to be well behaved over the whole region of integration. This condition is often not fulfilled, and integration would be impractical computationally.

Intuitively, the likelihood ratio is a measure of the consistency of the POI value $\vec{\mu}$ with the observed data. When the likelihood ratio is small, the observed data are more consistent

with $\hat{\vec{\mu}}$ than with $\vec{\mu}$, and when the likelihood ratio is near 1, the data are consistent with both $\hat{\vec{\mu}}$ and $\vec{\mu}$. This intuitive understanding explains the usage of the profile likelihood test statistic

$$t_{\vec{\mu}} = -2 \ln \lambda_{\vec{\mu}} = 2 \left(\ln \min_{\vec{\mu}', \vec{\alpha}} \mathcal{L}(\vec{\mu}', \vec{\alpha}) - \ln \min_{\vec{\alpha}} \mathcal{L}(\vec{\mu}, \vec{\alpha}) \right). \tag{6.6}$$

The prefactors have been chosen so that $t_{\vec{\mu}}$ has a natural interpretation in the case that $\mathcal{L}$ takes the form of a Gaussian distribution, which is the large statistics limit in most scenarios. In this case, using only a single POI and no nuisance parameters, the likelihood is written

$$\mathcal{L}(x|\mu) = \mathcal{N}(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$

with minimum $\hat{\mu} = x$. The test statistic becomes

$$t_\mu = \frac{(x-\mu)^2}{\sigma^2}.$$

In the Gaussian case, $\sqrt{t_\mu}$ is the number of standard deviations by which μ differs from $\hat{\mu}$. If the true value of μ is fixed to μ_{true} so that x is drawn from $\mathcal{N}(\mu_{true}, \sigma^2)$, the values of $t_{\mu_{true}}$ from repeated experiments follow a χ^2 distribution. A theorem due to Wilks [185] states that under appropriate assumptions, the likelihood ratio test statistic with n POIs follows a χ^2_n distribution in the asymptotic case of an infinite dataset. This is commonly used to approximate the p-value (probability of obtaining a larger value of the test statistic under the hypothesis) of a hypothesis test as

$$p = 1 - F_{\chi^2_n}(t_{\vec{\mu}}) \implies t_{\vec{\mu}} = F^{-1}_{\chi^2_n}(1-p).$$

Because of the natural connection to the Gaussian limit, it is customary to report p-values as exclusion in "units" of σ. For example, a 2σ exclusion corresponds to a p-value of 0.046, the probability for a normally distributed variable x to have $|x| > 2$, or equivalently, the probability for a χ^2 distributed variable y to have $y > 4$. This definition is conventional and valid even when Wilks' theorem does not apply, but under Wilks' theorem, the significance for a 1D hypothesis test in "units" of σ is expressed simply as $Z_\mu = \Phi^{-1}\left(2\Phi(\sqrt{t_\mu}) - 1\right)$ [181] where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution. The factor of two appears from a transition from a two-sided definition for the test statistic to a one-sided definition for the p-value. In the case that the test statistic is one-sided, as when measuring a signal strength, the distribution becomes $\frac{1}{2}\chi^2 + \frac{1}{2}\delta$ and the significance estimate becomes simply $Z_\mu = \sqrt{t_\mu}$. When the assumptions of Wilks' theorem do not hold, the rejection of the null hypothesis, also known as the significance of the result, must be evaluated by performing pseudo-experiments. Pseudo-experiments are generated by randomly sampling the likelihood function with the POIs and nuisance parameters fixed to a particular hypothesis, usually the null hypothesis for the POIs and the best fit values of the nuisance parameters. In addition to the experimentally measured data, the global observables must also

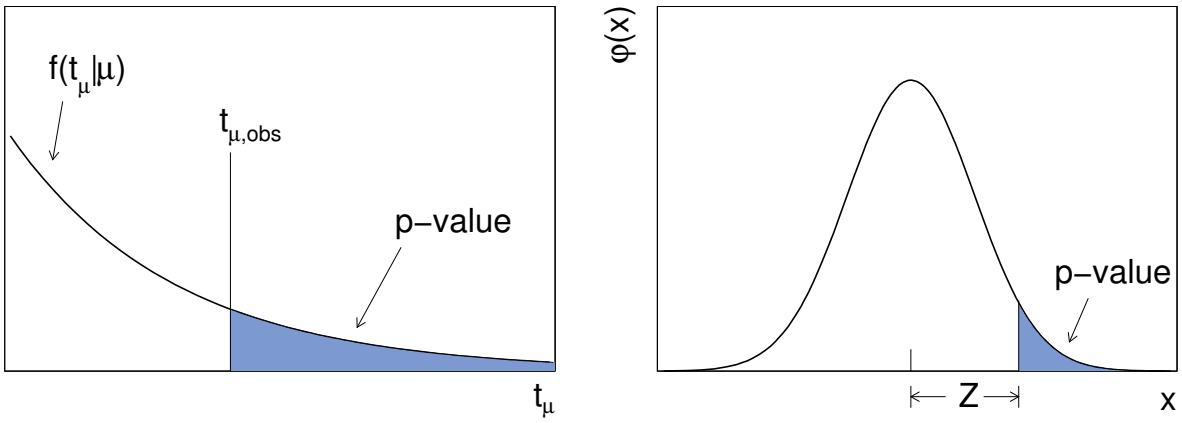


Figure 6.1: (Left) Illustration of the relation between the observed p-value and the test statistic t_μ. (Right) The standard normal distribution $\phi(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$ showing the relation between the significance Z and the p-value. Reproduced from [181].

be sampled from the constraint functions. The fit is performed on each pseudo-experiment, and a distribution of the resulting values of the test statistic is calculated. The p-value for the experiment is then evaluated by counting the fraction of pseudo-experiments for which the test-statistic is higher than the observed value.

Confidence Intervals

The purpose of defining the profile likelihood ratio test statistic is to move beyond the maximum likelihood estimate and place confidence intervals on the measured parameters. The frequentist interpretation of a confidence level with confidence p is that the interval will contain the true value of μ with probability p. Measurements in particle physics generally report the 1σ confidence interval corresponding in the one-dimensional case to a 68.3% confidence level. According to Wilks' theorem, this interval can be obtained in the 1D case by including all values of μ for which $t_\mu \leq 1$.

While this is a useful convention, in practical scenarios, confidence intervals are not expected to obtain this level of coverage. One reason for this is that estimates of systematic uncertainties are often deliberately conservative; the constraint on the nuisance parameters is weak, such that repetitions of the auxiliary experiments would more closely reproduce the original value than is suggested by the constraint term. This results in the reported confidence interval covering the true value of the parameter with probability higher than 68.3%.

CLs Limit Setting

When the measurement of a parameter is not able to reject the null hypothesis at a high level, the standard practice is to report exclusion limits, typically at the 95% confidence level, but sometimes also at the 68% and 90% confidence levels, denoting points that are excluded at the corresponding p-value. Naively, this is simply the inverse of the confidence interval problem, but there is one important practical complication. A likelihood such as that of eq. (6.3) is defined for $\mu < 0$, but $\mu < 0$ is unphysical. If the experiment is only sensitive to large values of μ, then a statistical fluctuation could lead to fewer observed events than the background expectation, and possibly to exclusion of signal strengths $\mu \leq 0$. A prescription for one-sided limit setting that avoids excluding arbitrarily small signals due to an observation below the expected background rate is required.

The standard prescription is the CLs method [186] replaces the p-value for the POI value μ (p_μ) with

$$CL_s = \frac{p_\mu}{1 - p_b}$$

where p_b is the p-value for the background-only hypothesis, i.e., the probability to observe a *lower* value of t_0 under the background-only hypothesis. Note that p_b uses an opposite convention to p_μ such that $p_{\mu=0} = 1 - p_b$. When observing background-like data, $1 - p_b \approx 0.5$. When there is a downward fluctuation such that $\hat{\mu} < 0$, p_μ for $\mu > 0$ decreases, but p_b increases, which has a compensating effect on CL_s.

The following description will specialize the discussion to the definition of the likelihood in the specific case of the $t\bar{t}t\bar{t}$ measurement. The regions and relevant background processes were discussed in chapter 8. The nominal yield expectation consists of the sum of signal and background estimations from the samples described there. What remains is the parameterization of the signal and background expectation in terms of the systematic uncertainties and the parameters of interest. In the measurement of the cross-section $\sigma_{t\bar{t}t\bar{t}}$ or signal strength $\mu_{t\bar{t}t\bar{t}}$, the POI is a simple multiplicative factor on the signal strength as in the case of eq. (6.2). For the parameterization used in interpretations of the $t\bar{t}t\bar{t}$ signal as a measurement of the top Yukawa coupling or Wilson coefficients of an effective field theory, see section 6.2.

6.2 Signal Parameterization for Interpretations

As described in chapter 5, possible BSM modifications of the $t\bar{t}t\bar{t}$ signal are modeled with alternate MC simulation at LO in QCD. To measure the underlying parameters of the BSM models, the change in the $t\bar{t}t\bar{t}$ yield in each bin is parameterized as a function of the underlying POIs using a fit to the alternate signal samples. This modification is applied multiplicatively to the nominal NLO $t\bar{t}t\bar{t}$ prediction denoted $\nu_{ij}^{t\bar{t}t\bar{t},SM}$ including the NLO cross-section calculation 12 ± 2.4 fb [41].

For the interpretation in terms of the top Yukawa coupling, the parameterization is based on the Lagrangian terms in eq. (2.8) following the functional form reported in Ref. [45].

$$\nu_{ij}^{t\bar{t}t\bar{t}}(\kappa_t, \tilde{\kappa}_t) = \nu_{ij}^{t\bar{t}t\bar{t},SM} \left(a_{ij} + b_{ij}\kappa_t^2 + c_{ij}\tilde{\kappa}_t^2 + d_{ij}\kappa_t^4 + e_{ij}\kappa_t^2\tilde{\kappa}_t^2 + f_{ij}\tilde{\kappa}_t^4\right) \tag{6.7}$$

To match the SM normalization, an additional constraint $a_{ij} + b_{ij} + d_{ij} = 1$ is required, allowing the replacement $a_{ij} = 1 - b_{ij} - d_{ij}$. This parameterization is implemented in a fit framework designed to fit EFT coefficients, which assumes that the SM values of the POIs are 0, unlike κ_t. It further assumes that the parameterization should take the form of a general polynomial up to order n, but there are no linear or cubic terms in the expansion of $\sigma_{t\bar{t}t\bar{t}}$ as a function of κ_t and $\tilde{\kappa}_t$, since each diagram has an even number of top-Higgs vertices. These problems are solved by a reparameterization $A_t = \kappa_t^2 - 1$, $B_t = \tilde{\kappa}_t^2$. Equation (6.7) then becomes

$$\nu_{ij}^{t\bar{t}t\bar{t}}(\kappa_t, \tilde{\kappa}_t) = \nu_{ij}^{t\bar{t}t\bar{t},SM} \left(1 + b'_{ij}A_t + c'_{ij}B_t + d'_{ij}A_t^2 + e'_{ij}A_tB_t + f'_{ij}B_t^2\right). \tag{6.8}$$

The coefficients are related by

$$\begin{aligned} a_{ij} &= 1 - b'_{ij} + d'_{ij} \\ b_{ij} &= b'_{ij} - 2d'_{ij} \\ c_{ij} &= c'_{ij} - e'_{ij} \\ d_{ij} &= d'_{ij},\ e_{ij} = e'_{ij},\ f_{ij} = f'_{ij} \end{aligned} \tag{6.9}$$

The primed coefficients in each bin $a'_{ij}...f'_{ij}$ are determined from a least squares fit to the bin yields taken from the BSM samples listed in table 5.4 to the functional form of eq. (6.8). The 20% uncertainty on the SM $t\bar{t}t\bar{t}$ cross-section prediction remains in place because the BSM modifications are modeled as multiplicative variations of the SM signal.

Reference [45] reports the cross-section dependence as

$$\sigma_{t\bar{t}t\bar{t}} = 9.998 - 1.522\kappa_t^2 + 2.883\tilde{\kappa}_t^2 + 1.173\kappa_t^4 + 2.713\kappa_t^2\tilde{\kappa_t}^2 + 1.827\tilde{\kappa_t}^4 \text{ fb} \tag{6.10}$$

and reports that the contribution of terms proportional to $\kappa_t\tilde{\kappa}_t$, $\kappa_t^3\tilde{\kappa}_t$, and $\kappa_t\tilde{\kappa_t}^3$, which arise from diagrams with one CP-even vertex and one CP-odd vertex are negligible as they enter only at NLO. In the parameterization used here, this becomes

$$\mu_{t\bar{t}t\bar{t}} = 1.04 - 0.16\kappa_t^2 + 0.30\tilde{\kappa_t}^2 + 0.12\kappa_t^4 + 0.28\kappa_t^2\tilde{\kappa_t}^2 + 0.19\tilde{\kappa_t}^4.$$

Table 6.1 compares this parameterization to results obtained from fits of eq. (6.7) to the alternate $t\bar{t}t\bar{t}$ MC samples produced with the Higgs characterizations model. The table shows results consistent with Ref. [45] even after the signal region selection defined in chapter 8 and a tight GNN > 0.9 selection on the GNN classifier described in chapter 9, corresponding to the most signal-like bin of the final measurement. These last two selections represent signal-enriched regions with potentially different behavior from inclusive $t\bar{t}t\bar{t}$ production. The fit in the last bin is also used to validate the results of the least squares fit in the fit

framework. The good agreement of these results shows limited dependence on the details of the MC generation setup or acceptance effects. The similarity between the last two rows also indicates that the GNN score shape has limited sensitivity to the top Yukawa coupling. Instead, the constraining power is dominated by the cross-section variation. The yield variation as a function of A_t and B_t is shown in fig. 6.2 for $t\bar{t}t\bar{t}$ and fig. 6.4 for $t\bar{t}H$. The parameters of the fit have minimal variation between bins in both cases.

Coefficient	a	b	c	d	e	f
Ref. [45]	1.04	−0.16	0.30	0.12	0.28	0.19
Higgs Characterization MC	1.03	−0.16	0.31	0.12	0.29	0.19
Signal Region Yield	1.04	−0.16	0.29	0.12	0.30	0.18
Signal Region GNN$\geq$ 0.9	1.03	−0.14	0.28	0.11	0.27	0.17

Table 6.1: Results of the parameterization of the $t\bar{t}t\bar{t}$ yield in different regions according to eq. (6.7). Apparent deviations from the closure criterion $a + b + d = 1$ are due to rounding.

For the interpretation in terms of the Wilson coefficients of four-fermion SMEFT operators, the parameterization is based on the Lagrangian terms in eq. (2.10), including both interference terms between diagrams for SM $t\bar{t}t\bar{t}$ production and EFT operators and terms arising solely from EFT operators, which are quadratic in the Wilson coefficients. The signal yield takes the form

$$\nu_{ij}^{t\bar{t}t\bar{t}}\left(\{C_\alpha\}\right) = \nu_{ij}^{t\bar{t}t\bar{t},SM}\left(1 + \sum_{\alpha} a_{ij}^{\alpha} C_\alpha + \sum_{\beta\gamma} a_{ij}^{\beta\gamma} C_\beta C_\gamma\right), \tag{6.11}$$

where Greek indices run over the EFT coefficients: $C_\alpha \in \{C_{tt}^1, C_{QQ}^1, C_{Qt}^1, C_{Qt}^8\}$. The coefficients in each bin a_i, a_{jk} are determined from a least squares fit to the bin yields taken from the BSM samples listed in table 5.5 to the functional form of eq. (6.11). The considered operators may also induce modifications to all major background processes via the introduction of higher order diagrams with a four top-quark vertex. Additionally, some operators induce a $t\bar{t}b\bar{b}$ vertex which modify $t\bar{t}b\bar{b}$ production at the tree level. In the signal region, these effects are expected be sub-dominant to the modifications to the $t\bar{t}t\bar{t}$ signal, so possible effects of the EFT operators on background processes are not considered.

For the measurement of the Higgs oblique parameter, the parameterization is performed in a similar manner using the functional form

$$\nu_{ij}^{t\bar{t}t\bar{t}}(\hat{H}) = \nu_{ij}^{t\bar{t}t\bar{t},SM}\left(1 + a\hat{H} + b\hat{H}^2\right). \tag{6.12}$$

following [55].

Special treatment of $t\bar{t}H$

In both the Higgs oblique and top Yukawa interpretations, modifications in the Higgs sector naturally affect the $t\bar{t}H$ background in addition to the $t\bar{t}t\bar{t}$ signal. Thus, the dependence of $t\bar{t}H$ on the POIs must also be considered. The simplest case is to assign a floating normalization to $t\bar{t}H$ as is done to $t\bar{t}W$ and the fake and non-prompt lepton backgrounds. This approach makes the $t\bar{t}t\bar{t}$ measurement independent of any additional BSM effects that may change the normalization of the $t\bar{t}H$ background, though it ignores any changes to the kinematics of $t\bar{t}H$ production that may be induced by BSM physics, particularly by a CP Odd component of the top Yukawa coupling.

One set of top Yukawa constraints is produced using this parameterization, which is hereafter called "$t\bar{t}H$ profiled" as the possible BSM effects are profiled in the fit in the same way as the nuisance parameters and normalization factors. This parameterization produces a top Yukawa constraint independent of modifications to the Higgs total width, Γ_h, the Higgs couplings to W and Z bosons, or the Higgs branch ratio to τ-leptons, which represent the dominant Higgs boson decay modes resulting in the SSML final state. Modifications to Γ_h would induce a pure normalization effect on the $t\bar{t}H$ background, but changes in the Higgs boson branching fractions could affect the shape of the $t\bar{t}H$ background as well. This possibility has not been studied in the context of the $t\bar{t}t\bar{t}$ analysis. Because $t\bar{t}H$ and $t\bar{t}W$ have similar shapes in the signal region, profiling $t\bar{t}H$ results in a significant anti-correlation between the $t\bar{t}H$ and $t\bar{t}W$ normalizations. Because the analysis is not designed for sensitivity to $t\bar{t}H$, it is difficult to separate from $t\bar{t}W$.

An alternative set of top Yukawa constraints uses a $t\bar{t}H$ generated with $\tilde{\kappa}_t = 1$ and $\kappa_t = 0$ to produce a similar bin-by-bin parameterization of the impact of the top Yukawa coupling as is done for $t\bar{t}t\bar{t}$. The parameterization takes the form

$$\nu_{ij}^{t\bar{t}H}(\kappa_t, \tilde{\kappa}_t) = \nu_{ij}^{t\bar{t}H,SM}\left(\kappa_t^2 + a_{ij}\tilde{\kappa}_t^2\right) \tag{6.13}$$

following [187], which indicates that the interference term between diagrams with CP-even and odd couplings is negligible. As with the $t\bar{t}t\bar{t}$ parameterizations, the coefficients a_{ij} are determined by a least squares fit to eq. (6.13) after a reparameterization

$$\left(\kappa_t^2 + a_{ij}\tilde{\kappa}_t^2\right) \rightarrow 1 + A_t + a_{ij}B_t.$$

This scenario is referred to as "$t\bar{t}H$ parameterized" and offers the best possible constraint on the top Yukawa coupling from this signal region, but requires the assumption that the Higgs total width and branching fractions to the SSML state are unchanged. As with $t\bar{t}t\bar{t}$, the uncertainty on the inclusive SM $t\bar{t}H$ cross-section remains.

Constraints on the Higgs oblique parameter use an inclusive cross-section dependence

$$\nu_{ij}^{t\bar{t}H} = \nu_{ij}^{t\bar{t}H,SM}\left(1 - \hat{H}\right)^2$$

because no shape dependence is expected [55].

6.3 Evaluation of Expected Results and Fit Quality

Asimov Data

When conducting blinded analysis, it is necessary to evaluate the expected result of an experiment under a given hypothesis to perform optimizations that maximize the sensitivity of the analysis to a chosen signature. For binned profile likelihood fits at the LHC, this is done with an Asimov dataset [181] constructed by setting the observed number of data events in each bin exactly equal to the expectation. A fit to the Asimov data returns the exact values of the parameter of interest and nuisance parameters used to generate the Asimov dataset, and statistical tests, such as the rejection of the null hypothesis, return the median result expected from data drawn from the likelihood with injected signal. Achieving this closure requires ensuring that the global observables are equal to the values of the nuisance parameters used to generate the Asimov data. The Asimov data, and thus derived figures of merit such as the expected significance, depend on choices of the nuisance parameters, which have reasonable *a priori* values, and the normalization factors, which may not. This is particularly relevant for the scaling factors a_0 and a_1 for the $t\bar{t}W$ jet multiplicity distribution in eq. (8.1). Thus, several versions of the Asimov data are used at different stages of the analysis development. The first is the "Plain" Asimov data, which uses the default MC predictions for all backgrounds, including setting all normalization factors to one and setting a_0 and a_1 to the values returned from a fit of eq. (8.2) to the $t\bar{t}W$ MC sample. The "Realistic" Asimov dataset is generated after performing a simultaneous fit to the data in all eight control regions with the $t\bar{t}t\bar{t}$ cross-section fixed to its SM value. The Realistic Asimov dataset is then generated from the background expectation produced by setting all normalization factors and nuisance parameters from their best fit values in the control region only fit. The analysis design is chosen to maximize the expected discovery significance of the $t\bar{t}t\bar{t}$ process under the SM hypothesis using the Realistic Asimov dataset before the signal region is unblinded. A final set of Asimov data is used to produce the expected significance figures and limits reported in Ref. [129]. This set of Asimov data is generated using the best fit values from the final fit to the data, which includes the signal region.

Fit Quality

A quantitative description of how well the best fit model matches the observed data across all signal and control regions, known as the goodness of fit, is obtained using a saturated likelihood model [188, 189]. A common and straightforward approach to evaluating the goodness of fit is to use the χ^2 metric formed by summing the one-dimensional χ^2 agreement of the data to the post-fit prediction in each bin of the measurement. The use of the χ^2 value is only valid when each bin yield is large enough to be in the Gaussian regime. In the saturated model approach, the χ^2 is replaced with a likelihood ratio

$$-2\ln\Lambda = -2\ln\frac{\mathcal{L}\left(\mathbf{x}\middle|\hat{\vec{\mu}},\hat{\vec{\alpha}},\hat{\vec{\mathrm{NF}}}\right)}{\mathcal{L}_{\mathrm{sat}}(\mathbf{x})},$$

where the saturated likelihood L_{sat} is a product of Poisson likelihoods with expectations set equal to the observed data, so that the observed data are the most likely result of the experiment.

$$\mathcal{L}_{\mathrm{sat}}(\mathbf{x}) = \prod_{i,j} f(x_{ij}|x_{ij}) = \prod_{i,j}\frac{x_{ij}^{x_{ij}}}{x_{ij}!}e^{-x_{ij}}$$

This likelihood gives the result obtained when fitting a model that perfectly matches the observed data, or equivalently, one with a floating degree of freedom for each bin of the measurement. The goodness of fit is evaluated using the asymptotic formula for the likelihood ratio to derive a p-value. The resulting p-value indicates the probability of observing worse post fit agreement under the assumption that the best fit parameters $\hat{\vec{\mu}}$, $\hat{\vec{\alpha}}$, and $\hat{\vec{\mathrm{NF}}}$ correspond to the true description of the experiment.

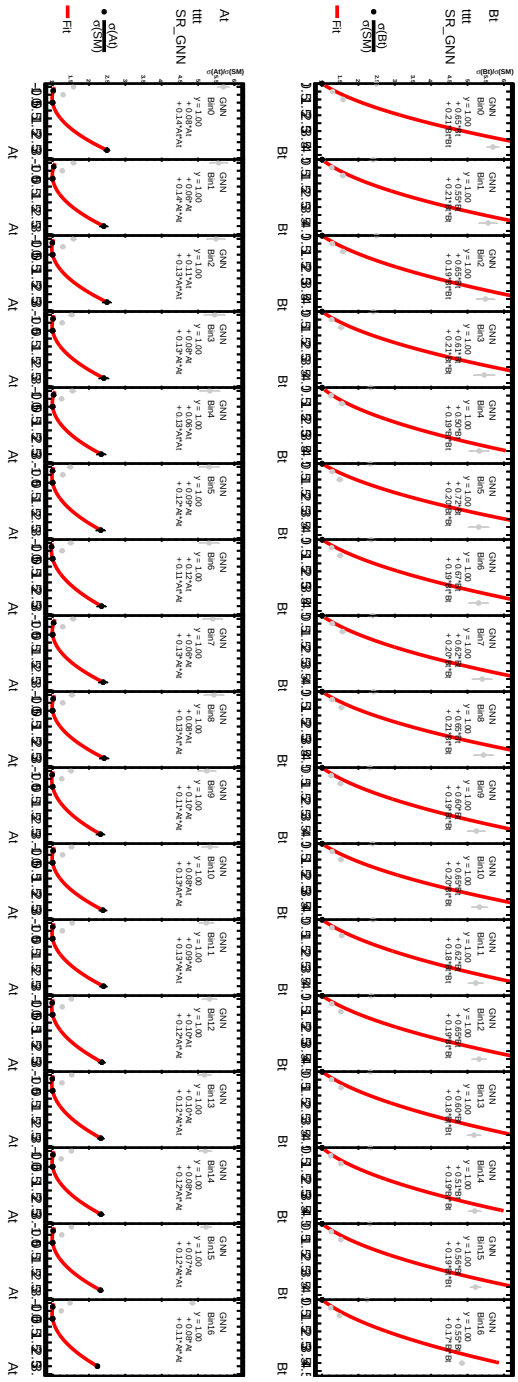
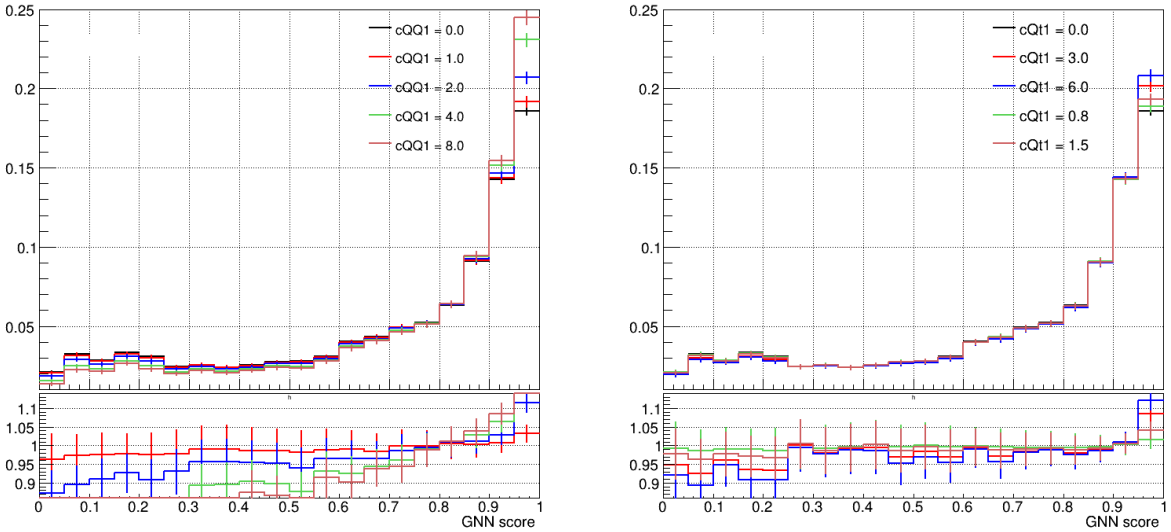
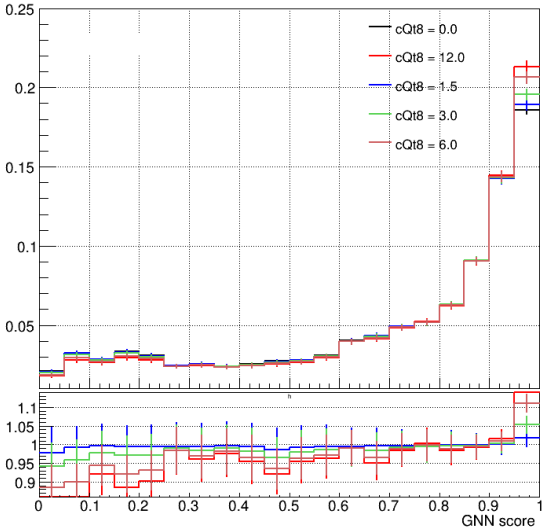


Figure 6.2: Best fit dependence of the $t\bar{t}t\bar{t}$ yield in each bin of the signal region on A_t and B_t for the top Yukawa parameterization. Each panel shows a different bin. Black points represent signal points for which the parameter not displayed is 0, while gray points represent points where both $A_t, B_t \neq 0$, including all points where $\kappa_t = 1$.

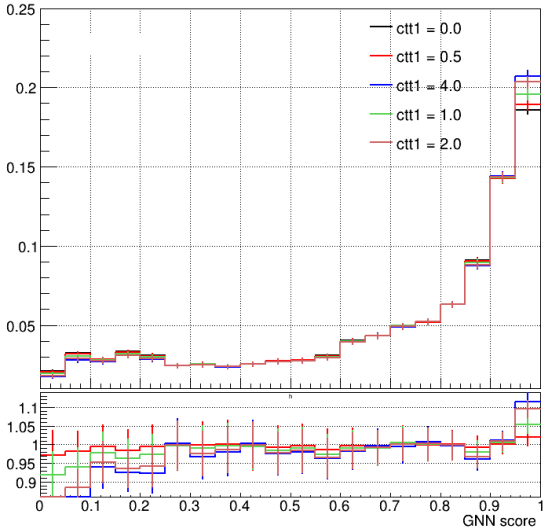


(a)

(b)



(c)



(d)

Figure 6.3: Variation of the GNN score distribution for the $t\bar{t}t\bar{t}$ signal as a function of the Wilson coefficients for four-fermion EFT operators. (a) C^1_{QQ} (b) C^1_{Qt} (c) C^8_{Qt} (d) C^1_{tt}

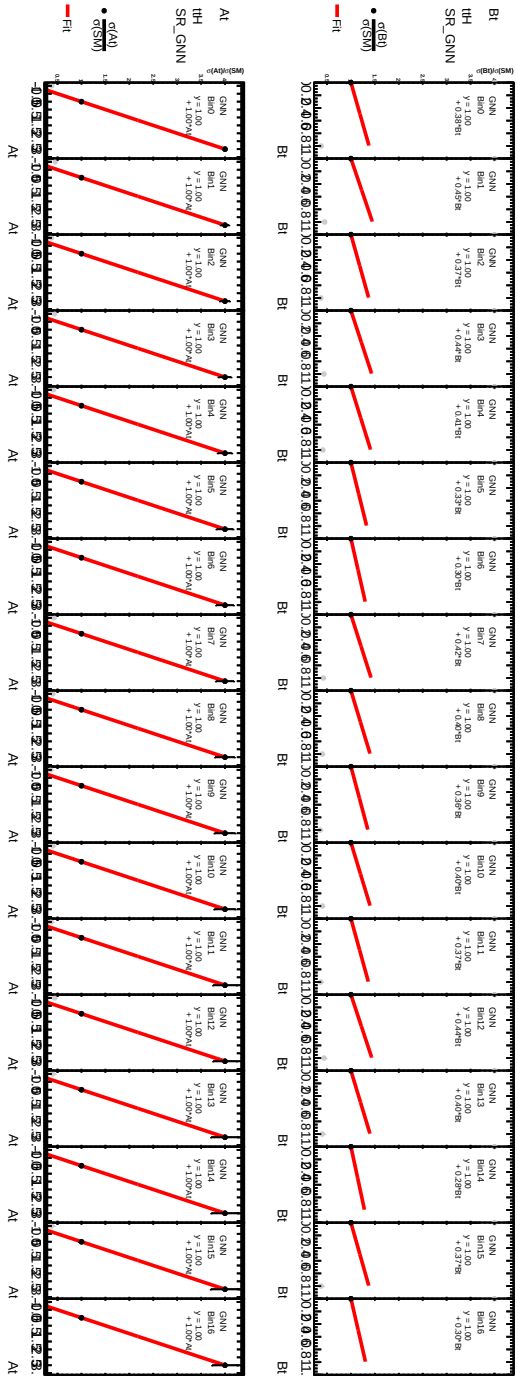


Figure 6.4: Best fit dependence of the $t\bar{t}H$ yield in each bin of the signal region on A_t and B_t for the top Yukawa parameterization. Each panel shows a different bin. Black points represent signal points for which the parameter not displayed is 0, while gray points represent points where both $A_t, B_t \neq 0$, including all points where $\kappa_t = 1$.

Chapter 7

Transformers and Graph Neural Networks

7.1 Machine Learning and Multivariate Analysis in HEP

Machine learning (ML), sometimes known by the overlapping but not interchangeable names multivariate analysis (MVA) or artificial intelligence (AI), broadly describes any algorithm for making predictions on input data that is first tuned (trained) on example data to optimize the parameters of the algorithm for the desired task. Algorithms tuned in this way can far outperform those hand-crafted by humans at many tasks, provided sufficient quantities of data are available, because they are capable of significantly more complexity and because modern techniques for optimizing these algorithms result in relatively fast convergence to nearly optimal solutions. ML algorithms, also known as models, can use many features of the data simultaneously and capture the correlations between those features in a way that hand-written cut-based algorithms cannot. Boosted decision trees (BDTs) [190] operate with a superposition of binary decision trees and are a common choice for HEP data analysis due to their simplicity, fast training time, and predictable behavior.

7.2 Neural Networks

Most ML tasks can be expressed in terms of a target probability distribution, with the ML model serving as an approximator of that distribution and the training dataset consisting of samples drawn from that distribution or a similar one. Across much of the ML industry, neural networks (NNs) are the default tool of machine learning. Neural networks are highly parametric functions $f(\mathbf{x}, \alpha)$ mapping a chosen input space $\mathcal{X}$ to a chosen output space $\mathcal{Y}$. Neural networks are trained with an optimization procedure that minimizes a loss function $\mathcal{L}(f(\mathbf{x}, \vec{\alpha}), \mathbf{x})$ with respect to the parameters $\vec{\alpha}$ using a given training dataset $\{\mathbf{x}_i^{train}\}$.

$$\hat{\vec{\alpha}} = \min_{\vec{\alpha}} \sum_i \mathcal{L}(f(\mathbf{x}_i, \vec{\alpha}), \mathbf{x}_i) \tag{7.1}$$

In the common case of classification with N classes, the model outputs in $\mathbb{R}^N$ are transformed to $[0,1]^N$ with the softmax function

$$\text{Softmax}(\mathbf{x})_i = \frac{e^{x_i}}{\sum_{j=1}^N e^{x_j}} \tag{7.2}$$

and the cross-entropy loss function [191]

$$\mathcal{L} = -\sum_i \log \hat{y}_{iy_i} \tag{7.3}$$

is preferred. The true class of sample i is $y_i = 1..N_c$ and $\hat{y}_{ij}$ is the model's prediction of the probability that sample i belongs to class j after the application of softmax. In the case of binary classification, the prediction may be simplified to a single real number $\hat{y}_{ij} \to \hat{y}_i$, the truth to a single binary number $y_i = 0, 1$, the softmax simplifies to a sigmoid activation, and the loss function becomes

$$\mathcal{L} = -\sum_i \left[y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i) \right]. \tag{7.4}$$

This loss function is chosen because a model optimized using it approximates the likelihood ratio. A short proof of the binary case follows. The straightforward generalization to the multi-class case [191] is omitted here.

Theorem 1. *Let $\rho_0(\mathbf{x}), \rho_1(\mathbf{x})$ be probability density functions with identical support defined on $\mathbb{R}^n$ and $f(x) : \mathbb{R}^n \to (0,1)$ be a function that minimizes the functional $\mathcal{L}[f] = -\int \left[\rho_0(\mathbf{x}) \log f(\mathbf{x}) + \rho_1(\mathbf{x})(1 - \log f(\mathbf{x}))\right] d\mathbf{x}$. Then $f(\mathbf{x}) = \frac{\rho_0(\mathbf{x})}{\rho_0(\mathbf{x}) + \rho_1(\mathbf{x})}$.*

Taking the functional derivative of the loss and setting it equal to zero gives

$$\delta\mathcal{L} = 0 = -\int \left[\frac{\rho_0(\mathbf{x})}{f(\mathbf{x})} \delta f(\mathbf{x}) - \frac{\rho_1(\mathbf{x})}{1 - f(\mathbf{x})} \delta f(\mathbf{x}) \right] d\mathbf{x} \tag{7.5}$$

There are no boundary terms, and the distributions have the same support, ensuring that f is never zero or one and there are no divergences. To minimize $\mathcal{L}$, the integrand must be zero everywhere.

$$\frac{\rho_0(\mathbf{x})}{f(\mathbf{x})} = \frac{\rho_1(\mathbf{x})}{1 - f(\mathbf{x})} \tag{7.6}$$

Rearranging to solve for f,

$$f(\mathbf{x}) = \frac{\rho_0(\mathbf{x})}{\rho_0(\mathbf{x}) + \rho_1(\mathbf{x})}. \tag{7.7}$$

QED.

Alternatively, eq. (7.6) can be rewritten to give the likelihood ratio as a function of f.

$$\frac{\rho_0(\mathbf{x})}{\rho_1(\mathbf{x})} = \frac{f(\mathbf{x})}{1 - f(\mathbf{x})} \tag{7.8}$$

Therefore, an ML model trained using the cross-entropy loss is equivalent to an approximator for the likelihood ratio and is therefore an optimal discriminator.

The principle of the optimization procedure is stochastic gradient descent [192], which iterates the model parameters towards an optimal point with a stepwise gradient descent method where each step uses only a subset of the training data.

$$\alpha \to \alpha' = \alpha + d\alpha = \alpha - \epsilon \frac{d}{d\alpha} \sum_{i=0}^{batch\ size} \mathcal{L}(f(\mathbf{x}_i, \alpha), \mathbf{x}_i) \tag{7.9}$$

The update shown in eq. (7.9) is a simple implementation of stochastic gradient descent, but most modern neural network trainings use the Adam optimizer [193], which incorporates the principles of momentum and variable step size per parameter to improve the convergence time.

The simplest neural networks are deep neural networks (DNNs), also known as multilayer perceptrons (MLPs). An MLP consists of alternating affine transformations, known as linear layers, and non-linear functions, known as activation functions. Typical choices of activation function include ReLU [194] and sigmoid functions [195] shown in eqs. (7.10) and (7.11), respectively.

$$\mathrm{ReLU}(x) = \max(x, 0) \tag{7.10}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{7.11}$$

A single layer of an MLP takes $\mathbb{R}^n \to \mathbb{R}^m$ and can be written as

$$f_i(x) = g(w_{ij} x_j + b_i) \tag{7.12}$$

where $w \in \mathbb{R}^{m \times n}$ is the weight matrix, $b \in \mathbb{R}^m$ is the bias vector, and $g(y_i)$ is the activation function. The universal approximation theorem [196] states that MLPs constructed this way can approximate any function subject to some smoothness requirements arbitrarily well as the width of the network becomes infinitely large. This theoretical result means that these functions are sufficiently general to approximate the likelihood ratio or other functions requiring numerical approximators. From an empirical perspective, neural networks owe their widespread use to their computational efficiency and tendency to interpolate well between data points in the training dataset.

Figures of Merit and the ROC AUC

The performances of machine learning algorithms are evaluated with many metrics. In the context of particle physics, the ultimate figure of merit is the physics knowledge gained by using the algorithm, often expressed as the sensitivity to a given signal, model, or point in parameter space. In practice, the expected sensitivity of a given data analysis is often expensive to compute and depends on many factors besides the machine learning algorithms, including event categorization, binning, systematic uncertainties, and fit procedure, which may not be finalized when optimizing the MVA. The neural network training procedure naturally produces the loss value, which has a one-to-one correspondence with the likelihood. However, the loss has drawbacks as a practical figure of merit because it is difficult to compare scenarios with different selections or signal-to-background ratios. Other common choices for binary classifiers, including the accuracy or even the maximum value of the Gaussian significance estimate $s/\sqrt{b}$ for a given cut on the score, depend on the choice of a single cut when realistic data analyses place multiple optimized cuts to divide the score distribution into bins and use the whole phase space. These considerations have led to the common use of the ROC AUC as a metric for the performance of an observable as a binary classifier. The receiver operator curve (ROC) is constructed by measuring the acceptance ϵ_s, ϵ_b of the signal and background datasets as a function of a variable x. The curve is constructed by plotting the acceptance for the signal on the vertical axis and the acceptance for the background on the horizontal axis. For a weighted set of discrete samples from the signal and background distribution, the curve may be drawn by sorting all samples by x and then measuring the signal and background acceptance for a cut placed at each sample. The ROC curve is independent of monotonic transformations of the variable x and thus can be used to compare any classifier variables.

$$\text{ROC AUC} = \int_0^1 \epsilon_s(x) d\epsilon_b(x) = \int \epsilon_s(x) \frac{d\epsilon_b(x)}{dx} dx = \int \epsilon_s(x) \rho_b(x) dx$$

The area under the ROC curve is known as the ROC AUC and can be calculated by using a sample dataset to approximate the ROC curve. An alternative definition of the ROC AUC is the probability that the classifier will correctly order a randomly chosen pair of signal and background events, i.e. that the signal event will receive a higher score. This can be shown by performing a transformation $x \to x' = \epsilon_b(x)$ such that $\epsilon_s(x')$ is the cumulative probability distribution for the signal and the background is uniformly distributed in $[0, 1]$.

$$\text{ROC AUC} = \int_0^1 \epsilon_s(x) \rho_b(x) dx = \int_0^1 dx \int_x^1 dy \rho_s(y) \rho_b(x) = \iint \rho_s(y) \rho_b(x) \Theta(y - x) dx dy$$

7.3 Graph Neural Networks for Particle Physics

Graph Structured Data

Many generalizations of DNNs have been applied to various tasks, often taking advantage of problems whose data does not naturally admit a representation in $\mathbb{R}^n$. This work concerns the case of graph-structured data [197, 198, 199, 200, 201, 202], described by a network of nodes connected by edges. In the context of ML, a graph g is defined as a triple (N, E, G) where

$$N = \{n_1..n_N\},\ n_i \in \mathbb{R}^m$$

is the set of nodes, each of which has m node features,

$$E = \{e_1..e_E\},\ e_i = (s_i, r_i, x_i), s_i, r_i \in N,\ x_i \in \mathbb{R}^k$$

is the set of edges, each of which has a sender and receiver node (s, r) and k edge features, and

$$G \in \mathbb{R}^l$$

contains the l global features of the graph. The graph structure is natural for many particle physics applications [203, 204] where data consists of an irregular array of entries with geometrical information, identifying features, and non-trivial relationships between each other. This is the case in detector reconstruction, where detector hits form the nodes of the graph and are connected by edges representing geometrical proximity or a physical relationship such as association with the same final state object. In this case, the graph structure naturally accommodates irregular detector geometry and event-to-event variability in the number of hits.

Applications to tracking [205, 206] and calorimetry [207, 208] have been explored in depth. The graph structure is also natural for reconstructed data, where final state particles or objects form the nodes of the graph, and edges represent kinematic proximity or a physical relationship such as a shared vertex or mother particle.

The fully connected graph is a common subclass where the set of edges contains every possible connection between pairs of distinct nodes.

$$E = \{(n_i, n_j, x_{ij}),\ 1 \leq i, j \leq N,\ i \neq j\}$$

The graph structure of a fully connected graph is fully symmetric between all nodes and edges, but the symmetry is broken by the node and edge features.

A related data structure is the point cloud, which has only nodes and their associated features and is equivalent to a graph with no edges or a fully connected graph with no edge features ($k = 0$). Neural networks that operate on point clouds or sets, such as Deep Sets [209, 210, 211], have been explored, but do not fully exploit the relationships between individual nodes of a graph.

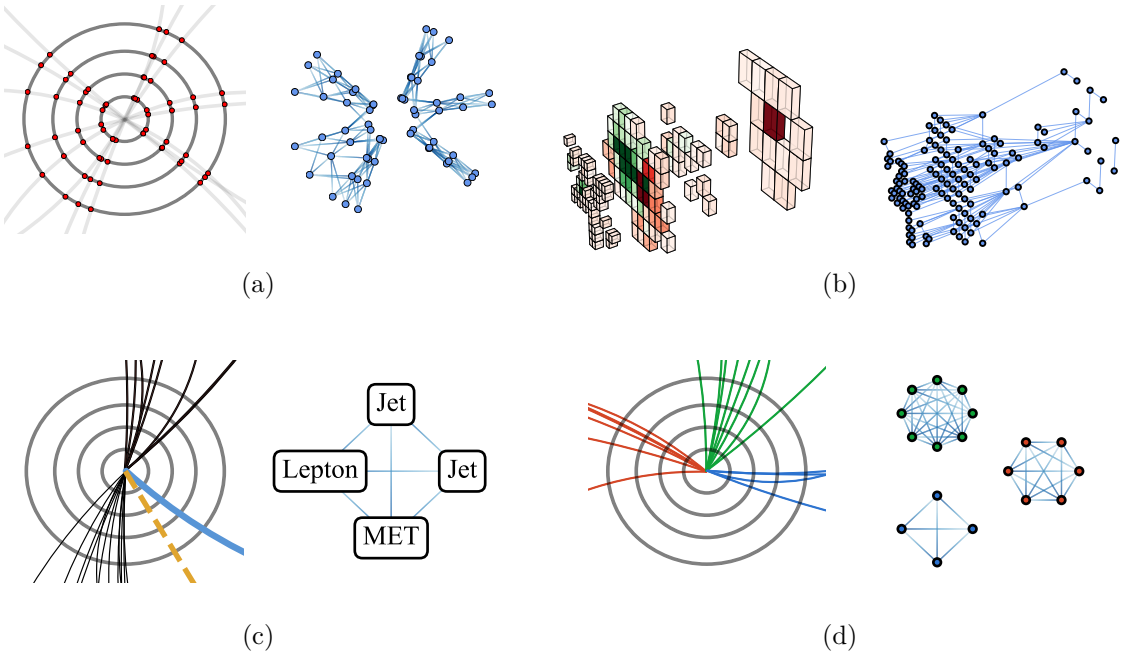


Figure 7.1: HEP data lend themselves to being represented as a graph for many applications: (a) clustering tracking detector hits into tracks, (b) segmenting calorimeter cells, (c) classifying events with multiple types of physics objects, (d) jet classification based on the particles associated wit the jet. Reproduced from [203].

Further generalizations of graphs, such as hypergraphs [212, 213] and heterogeneous graphs [214, 215] are outside the scope of this work, but may have interesting applications to particle physics.

Graph Neural Networks

DNNs cannot naturally handle graph-structured data because the number of features varies from graph to graph, and flattening the graph structure usually induces an ordering of the nodes. Graph neural networks (GNNs) are a broad class of machine learning models that act on graph-structured data while maintaining permutation invariance of the nodes and edges, accept variable-sized inputs, and respect the connection structure of the graph. The fundamental operating principles of GNNs are message passing and aggregation.

Message passing refers to one of several mechanisms for sending edge or node features to adjacent graph elements for use in re-computing the features of that element. Aggregation refers to a mechanism to combine a variable number of messages passed from adjacent graph elements into a fixed-length object, which can be the input to a feature update DNN. The

message passing structure considered here is the one defined by [202] and shown in fig. 7.2. The notation used here differs slightly, with nodes and the set of nodes referred to as n and N, and the global features as G.

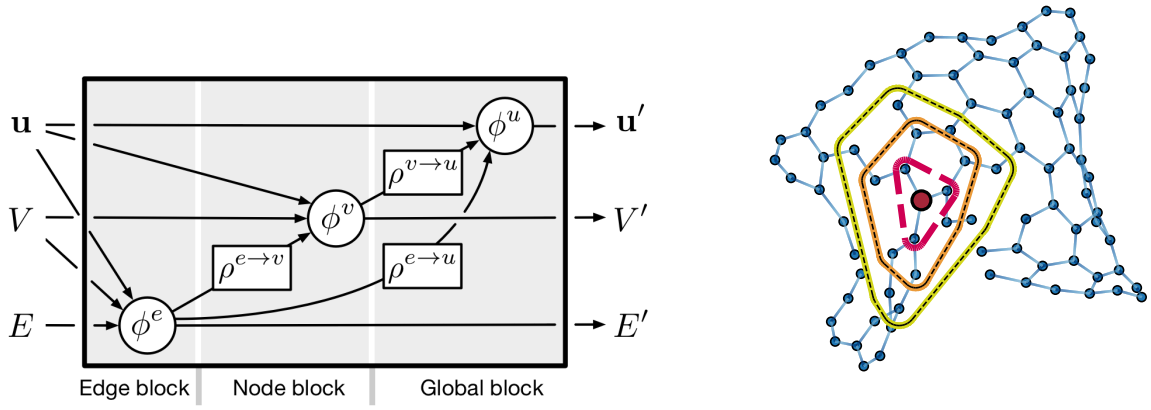


Figure 7.2: (Left) Schematic of the message passing block in the full GN model of [202]. (Right) Illustration of the propagation of information from a given node to an expanding neighborhood through successive iterations of message passing. Images from [202] and [203].

The update functions ϕ^e, ϕ^n, ϕ^g are general functions $\mathbb{R}^n \to \mathbb{R}^m$ and are typically taken to be MLPs or more generally deep neural networks which may contain additional normalization layers. To achieve permutation invariance and support arbitrary graph sizes, the aggregation functions $\rho^{en}, \rho^{ng}, \rho^{eg}$ must accept arbitrary length inputs and themselves be permutation invariant. Choices include sum, mean, max, and min, with the sum and mean being the most common. A single message passing step takes the graph (N, E, G) to a new graph with the same structure but modified features (N', E', G'). This step may be written as

$$\begin{aligned} e'_i &= \phi^e(s_i, r_i, x_i, G) \\ n'_i &= \phi^n\left(n_i, \rho^{en}(\{x \forall e = (s, r, x) \in E | r = n_i\}), G\right) \\ G' &= \phi^g\left(G, \rho^{eg}(E), \rho^{ng}(N)\right). \end{aligned} \tag{7.13}$$

The edge update is based only on the sender and receiver nodes that the edge connects, while the node update considers the aggregate of all edges for which the node is the receiver. All updates take in the original features of the graph element and the global features, and the global update considers the aggregates of all node features and all edge features. A typical implementation has repeated instances of the message passing step, begins with an encoder block that transforms the edge, node, and global features into a latent space usually of larger dimension than the input features, and ends with a decoder block that reduces the features down to the desired number of outputs. These may be thought of as alternate message

passing blocks where the dependence on element features for different types of elements is removed. Figure 7.3 illustrates this setup.

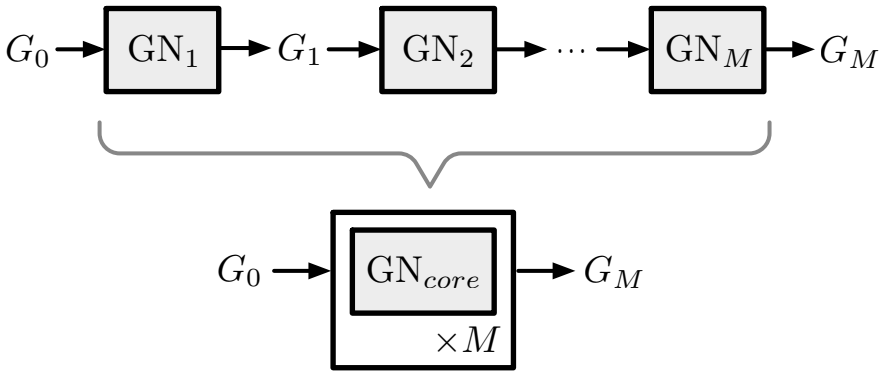


Figure 7.3: Typical GNN architecture with repeated message passing steps. Image from [202].

The input and output of the GNN described here are both graphs. This enables learning tasks with targets defined in the edge and node domains, rather than just the global domain. Such tasks are more difficult to define with a flat DNN or other NN architectures. Global learning targets allow classification and regression of features shared across the graph, while node and edge learning targets allow classification and regression of features of the graph's constituents, including identifying which nodes are of interest or which edges connect nodes sharing a common physical origin.

Transformers

Recent trends have favored transformer models, which have found use in commercial applications and elsewhere, over GNNs in applications of geometrical learning. It is therefore worth describing the characteristics of transformer models and the relationship to GNNs in the context of particle physics. Transformer architectures are primarily characterized by the use of an attention mechanism [216]. The idea of attention is to replace an aggregation with a weighted aggregation where the weights, also known as attention weights or scores, represent how important each element is to the quantity being calculated. The most common implementation is the scaled dot product attention of [216], which can be more naturally amenable to efficient computation than some GNN operations. The attention idea does not specify what the elements of the aggregation should be, and the idea has found use in many applications, including natural language processing [217], computer vision [218], and protein structure prediction [219]. The range of architectures under the transformer umbrella developed for these tasks is correspondingly large. This work will focus on architectures of the type commonly used in particle physics, such as the GNN architecture ParticleNet [220]

or the transformer architecture [221] developed for flavor tagging in ATLAS, for the sake of comparison to GNNs. The model implemented in appendix B is such a transformer. These models, and those of more recent efforts, contain many interesting developments, including Lorentz covariance [222], dynamic graph structures based on learned embedding [220], and integration of auxiliary training tasks [221], but all transformers models are unified in utilizing the attention mechanism as the message passing component of a (typically fully connected) graph.

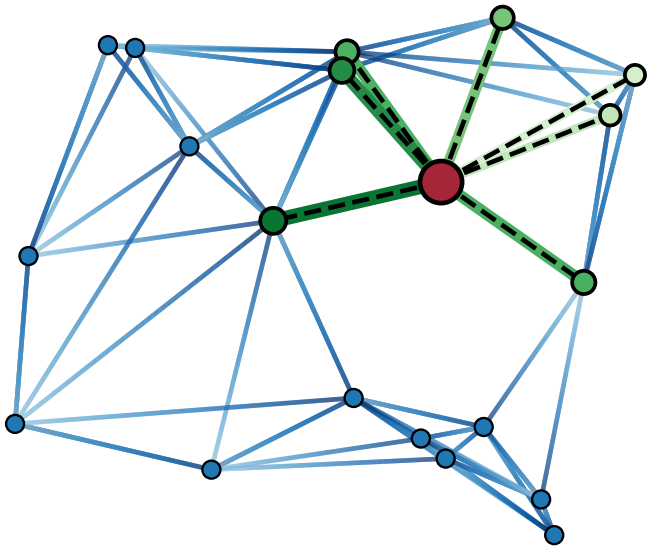


Figure 7.4: Illustration of the attention operation on a graph. Image from [203].

Transformers and GNNs are examples of geometric learning models, which seek to perform machine learning tasks on data with non-trivial geometric structure. Such models are largely differentiated by the mechanisms used to propagate information through the geometric structure, i.e., the message passing step. The message passing step consists of an update to the node features of each element based on that element and a combination of the other elements in the graph, $n_i \to n_i' = f(n_i, g(\{n_{j\neq i}\}))$. The update function f may be taken to be a simple MLP, and the form of g differentiates the different geometric learning architectures. The simplest option for g is the Deep Sets [209] approach of a simple sum:

$$g_{DS}(\{n_{j\neq i}\}) = \sum_{i\neq j} n_j$$

The transformer approach allows the contributions from the elements that are more relevant to be emphasized. The most commonly used attention mechanism is scaled dot product self-attention, which is written as

$$g_T(\{n_{j\neq i}\}) = \sum_{j\neq i} \text{softmax}\left(\frac{n_j^T M n_i}{\sqrt{d}}\right) n_j$$

where softmax is a normalization that forces the sum of weights to be one, d is the dimensionality, and M is a learnable matrix. The GNN further elaborates on the idea that the interaction between each pair of nodes should depend on the features of those nodes by promoting the role of the edges from a connectivity matrix to featured graph elements. In a GNN, even nodes with the same node features may interact differently if connected by an edge with different features. In this case, g_{GNN} may be thought of as the node update function ϕ^n of eq. (7.13).

The most common tasks for geometric learning models in HEP are jet classification, where the graph consists of a set of tracks associated with a jet, and event classification, where the graph consists of a set of final state objects in an event.

7.4 Development of the `root_gnn` and `root_gnn_dgl` Packages

This section describes the software landscape for machine learning packages enabling geometrical learning, including training, inference, and data pipelines for transformer and GNN models. This is a rapidly changing landscape, as these models are an active area of research in particle physics and the broader machine learning research community. The goal here is to provide context for the two GNN software packages, `root_gnn` and `root_gnn_dgl`, developed by the author, and describe the role they are intended to fill.

Commercially Available Packages for Geometric Neural Networks

Most modern machine learning development is based on open source Python packages such as Tensorflow [223], PyTorch [224], and JAX [225]. These packages support very general machine learning development and benefit from the support of major private technology companies. Additional software packages provide functionality to support graph neural networks and other geometrical learning models, including optimized graph data structures, graph operations, and message passing implementations of the type needed for GNNs and similar models. These packages are either components of the general machine learning frameworks or third-party implementations built on top of them. In the former category, PyG [226] is based on PyTorch, and TF-GNN [227] is based on Tensorflow. In the latter category, graph_nets [202] uses Tensorflow as a backend, and Deep Graph Library (DGL) [228] supports the use of PyTorch, Tensorflow, or MXNet [229] as a backend. Many of these frameworks allow export into the Open Neural Network Exchange (ONNX) format, which allows a model to be run from any of several supported backends, including a purely C++ based implementation suitable for incorporation with the LHC experiments' analysis software.

Particle Physics Focused Packages for Geometric Neural Networks

At the same time, the proliferation of geometrical learning models in high-energy physics has been accompanied by a number of publicly available code bases for training these models. These packages have largely been developed for a particular use case, and the products have varying levels of specificity in the model architectures and data types that they support. Some of these packages are built on top of existing commercially available GNN or transformer libraries, and all use Tensorflow or PyTorch as the underlying backend.

The SPANet [230, 231, 232] package, originally developed to reconstruct heavy particle decays with jet assignment while preserving Lorentz covariance, is also capable of event- and particle-level regression and classification tasks. ParticleNet [220] has been developed for jet learning tasks, but has been explored for even classification and other areas.

The SALT [233] framework, initially developed for jet flavor tagging in ATLAS, has been generalized to support training multi-task transformer models for a range of HEP applications. SALT is used for the transformer-based electron and photon energy calibration described in appendix B.

The `root_gnn` Package

The `root_gnn` package began with GNN code developed for the Exa.Trkx [234] project and is based on the `graph_nets` [202] package. `root_gnn` aims to produce a simple, configurable, end-to-end pipeline for training GNNs on reconstructed collision events. The package includes tools for processing events saved in the ROOT TTree [235] data format into an intermediate TFrecord [223] file containing fully connected graphs constructed from the TTree entries and for performing inference on TTrees, including the combination of multiple trainings in a k-fold setup. The package implements the full GN block of [202] in a recurrent structure as shown in fig. 7.3. The aggregation functions ρ are sums while the update functions ϕ are all MLPs with an additional `LayerNorm` [236] layer after each MLP. The inputs to each update MLP are concatenated, and the output is always in a latent space with a dimension shared by the entire model. A final output MLP acts on the global features to produce the output score(s). The default training uses the Adam [193] optimizer and the binary cross-entropy loss.

The graph construction, model hyperparameters, and training are configured via a `.ini` file, which consists of a set of mandatory sections to define the training setup and an arbitrary number of object definition sections to define the object types used as the nodes of the graph. Additional optional sections allow the implementation of configurable weights on the training samples or subsets thereof.

The object type sections define the branch names used for the object's four-momentum (p_T, η, ϕ, E), tagging variables for b-tagging and lepton charge, scaling factor for dimensionful inputs, and whether to use the measured energy or to apply the massless assumption. The ordering and grouping of the object sections determine the particle type labels used to label objects in the graph's node features.

Mandatory sections for the graph definition list the global features and event tracking information, including the Monte Carlo event weight and the variable used to split events into testing and training datasets. An additional mandatory segment stores key arguments for the training and data preparation scripts, most notably the model and training hyperparameters. The model has three important hyperparameters: the latent size, which determines the dimension of all latent space representations in the model, the number of layers, which determines the depth of all MLPs in the model, and the number of times that the message passing step is iterated. Additional hyperparameters for the training include the batch size, learning rate, and the dropout rate [237] used to mitigate overfitting. The GNN classifier described on chapter 9 is trained using the `root_gnn` package.

The `root_gnn_dgl` Package

The `root_gnn_dgl` package was developed to match the functionality of `root_gnn` while placing the code in a framework that would allow more flexibility in the graph and model definition. The general design philosophy is similar, but the code is based on the DGL framework, eliminating dependencies on `graph_nets` and Sonnet, and uses `YAML` configuration, which allows for a more flexible nested configuration format.

The default model of `root_gnn_dgl` is the same as `root_gnn`, but more variations are available. Several GNN models trained with `root_gnn_dgl` are described in the Appendices, and a version of the package will be made publicly available circa Summer 2025. `root_gnn_dgl` aims to be a simple tool to allow rapid prototyping for analysis teams that would like to experiment with GNNs without impeding long-term tuning and innovation in scenarios better suited to non-standard techniques.

Chapter 8

Same-Sign Multilepton Selection and Background Modeling

The data analysis described in this and the following chapters is that of Ref. [129, 130], though many aspects of the analysis have been taken from a previous ATLAS search for $t\bar{t}t\bar{t}$ production in the same-sign dilepton and multilepton (SSML) channels [60] and are described more completely there. This thesis aims to present a complete description of the analysis with emphasis on the improvements made over the previous publication, which enabled the observation of $t\bar{t}t\bar{t}$ production in this work.

Top quarks decay via $t \to W^+b$ with a branching fraction of nearly 100%, and the W boson decays hadronically 67% of the time. The remaining 33% of W decays are $W^+ \to \ell^+\nu_\ell$ where $\ell = e, \mu, \tau$ with roughly equal probability. The $t\bar{t}t\bar{t}$ final state therefore decays into four b-quarks and 0–4 charged leptons from leptonic W boson decays, which are accompanied by the same number of neutrinos and 0–8 light jets from the remaining hadronic W boson decays. ATLAS and CMS measurements of $t\bar{t}t\bar{t}$ production [60, 61, 62, 238] are categorized by the number and charges of the semileptonic top-quark decays, with varying signal yields in each channel according to the W boson branching ratios and combinatorics. Each channel features significantly different background rates and compositions. The advantage of choosing events with leptonic W decays is the ability to suppress QCD backgrounds both in the trigger and the offline analysis. τ-leptons are more difficult to reconstruct than electrons or muons, so "lepton" here includes only light leptons, i.e., electrons and muons. Hadronically decaying τ-leptons are not considered within this definition, but τ leptons decaying leptonically are. The most sensitive channels are the same-sign dilepton and multilepton channels, which require two leptons with the same electric charge. These are referred to as same-sign leptons. This channel targets $t\bar{t}t\bar{t}$ events where the W bosons from either both tops or both anti-tops decay leptonically. The low branching ratio of just 12.6%, including cases with two same-sign leptons and events with at least three leptons (all including leptonic τ decays), is mitigated by the excellent background rejection afforded by the same-sign lepton requirement. Where requiring the presence of one or two isolated leptons significantly suppresses background from QCD multijet production, requiring same-sign leptons further suppresses backgrounds from

typical lepton sources, such as $W+$jets production, $t\bar{t}$ production, or Drell–Yan production, which have final states with single leptons or oppositely charged pairs of leptons.

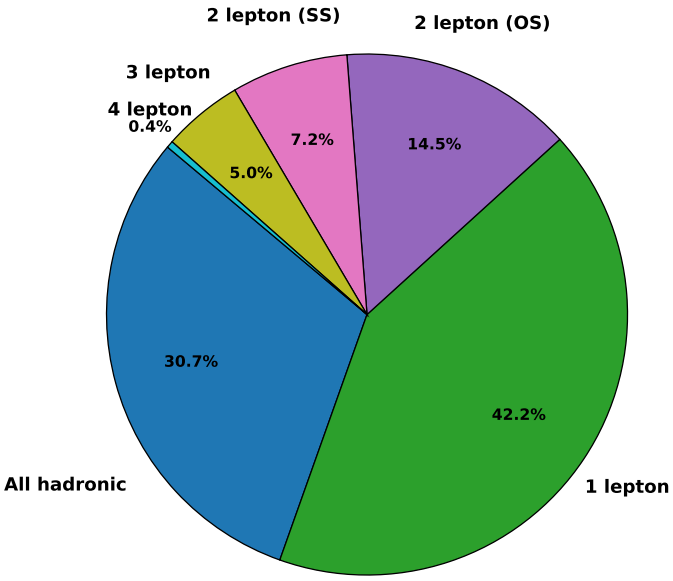


Figure 8.1: Branching ratios of $t\bar{t}t\bar{t}$ events to different final states. Leptonically decaying τ-leptons are considered leptons, but hadronically decaying τ-leptons are not.

8.1 Event Selection

The analysis considers reconstructed electrons, muons, jets, and $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstructed as described in chapter 4, with the event selection and observables constructed from those objects. The primary vertex is chosen as described in chapter 4 and is required to have at least two ID tracks with $p_T > 500$ MeV. The analysis uses single and dilepton triggers, which are detailed in chapter 5. Each of these has its own identification and isolation requirements and thresholds for the p_T of the lepton(s), all of which vary somewhat between data-taking periods. The leading lepton is required to have $p_T > 28$ GeV, which is always above the single lepton trigger threshold. Additionally, at least one reconstructed lepton must match the trigger-level lepton that triggered the event.

The SSML requirement is implemented for electrons and muons satisfying the 'tight' reconstruction criteria. Additional requirements are placed on the leptons to further reduce

background contamination. In events with three or more leptons, e^+e^- and $\mu^+\mu^-$ pairs are required to have invariant mass satisfying $|m_{\ell\ell} - 91 \text{ GeV}| > 10$ GeV to reject backgrounds from leptonic Z boson decays. Similarly, because electrons are more likely to have an incorrect charge assignment, events with two same-sign electrons must have $|m_{ee} - 91 \text{ GeV}| > 10$ GeV and $m_{ee} > 15$ GeV, which reduces backgrounds from charge mis-assignment in electronic Z decays, radiative production of low mass electron pairs, and photon conversions in the detector material.

An additional requirement of at least one b-tagged jet at the 77% working point is placed, as the $t\bar{t}t\bar{t}$ signal produces four bottom quarks and all relevant background processes also contain b-jets. All of these criteria constitute the preselection, which encompasses all events entering the final signal region of the analysis and those entering control regions used to constrain the background model in the fit. Data outside of this selection are used to measure the charge mis-identification rate and to validate the modeling of the $t\bar{t}Z$ background, both of which are detailed in section 8.2.

To further isolate the $t\bar{t}t\bar{t}$ signal from background processes, the signal region requires events to have at least six jets, two of which must be b-tagged at the 77% working point. Additionally, the scalar sum of all leptons and jets (H_T) is required to satisfy $H_T > 500$ GeV. This selection is motivated by the large multiplicity and activity of $t\bar{t}t\bar{t}$ events, which have an invariant mass of at least $4m_t = 692$ GeV and are expected to produce four b-jets and four additional jets when there are two semileptonic top-quark decays. This fully defines the signal region (SR) selection. Further separation of the $t\bar{t}t\bar{t}$ signal from backgrounds in the signal region is achieved with a graph neural network discriminant described in chapter 9. The signal region includes an additional requirement that the GNN output, referred to as "GNN score," satisfies GNN score > 0.1. With 140 fb^{-1} of data in Run 2, ATLAS is expected to have recorded 1,680 $t\bar{t}t\bar{t}$ events, of which 212 should satisfy the SSML requirements. The signal region has an expected $t\bar{t}t\bar{t}$ yield of 37.7 events, corresponding to an acceptance of 18% for SSML events and 2.2% for inclusive $t\bar{t}t\bar{t}$ production.

8.2 SSML Background Composition

The main backgrounds in the signal region are characterized by the requirements of same-sign leptons, large jet multiplicity, and the presence of b-jets. All major backgrounds have top quark pairs, which can produce leptons and b-jets. The same-sign lepton requirement is met either with the addition of a leptonically decaying heavy boson, a lepton with mis-assigned charge (QmisID), a non-prompt lepton from a hadronic decay, or a fake lepton candidate produced by hadronic activity. The first of these cases may be considered the irreducible background, while the latter three are experimental and reducible in nature. Nearly all background processes fail to produce the required multiplicity without additional jets resulting from initial state radiation, parton showering, or additional radiation in the hard scatter, which increases the background model's sensitivity to uncertainties from higher-order QCD effects. A summary of the expected signal region yield for the signal and most

background processes is presented in table 8.1.

Process	SR Yield
$t\bar{t}t\bar{t}$	37.7 ± 8.4
$t\bar{t}W^+$	74.0 ± 11.6
$t\bar{t}W^-$	43.4 ± 7.4
$t\bar{t}WW$	10.8 ± 5.5
$t\bar{t}\ell^+\ell^-{}_{high}$	69.0 ± 14.4
$t\bar{t}\ell^+\ell^-_{low}$, $t\bar{t}Z_{qq,\nu\nu}$	2.6 ± 1.4
$t\bar{t}H$	64.6 ± 10.6
QmisID	27.2 ± 3.7
Mat CO	16.5 ± 2.3
γ^*	14.1 ± 2.0
HFe	3.1 ± 1.0
HFμ	7.1 ± 1.2
LF	5.2 ± 5.2
Other fake	10.7 ± 3.5
$t(\bar{t})X$	7.8 ± 2.5
VV	7.0 ± 5.1
others	0.1 ± 0.1
$t\bar{t}t$	2.9 ± 0.9
rare $t\bar{t}VV$	5.1 ± 2.6
Total	408.9 ± 41.7

Table 8.1: Expected yields of the signal and dominant background processes in the signal region evaluated from MC simulation.

The contributions from different background processes are estimated using templates to describe the kinematic distribution of the background process. The shape and normalization of these templates are generally derived from MC simulation with systematic uncertainties on the shape and normalization representing uncertainties in the SM cross-section calculations, modeling of additional QCD radiation, and experimental effects. These systematic variations are detailed in section 8.8. Some background processes, including $t\bar{t}W$ production and backgrounds with fake and non-prompt leptons, are difficult to model in simulation or are known to differ from the best available SM prediction. These backgrounds' templates are given floating normalizations and extra degrees of freedom, which are constrained by simultaneously fitting dedicated control regions (CRs). The background estimate in these

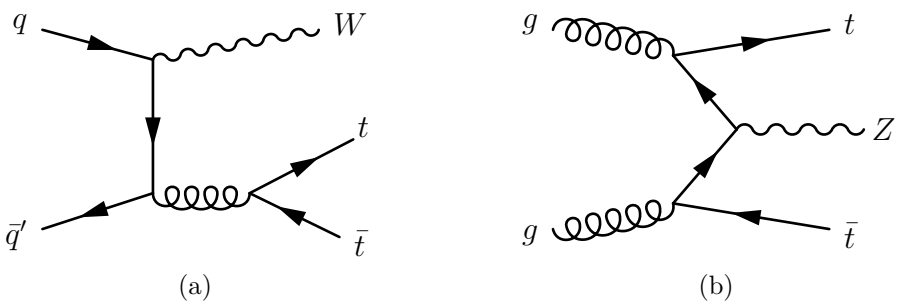


Figure 8.2: Leading-order tree-level Feynman diagrams for the leading backgrounds in the SSML channel, (a) $t\bar{t}W$ and (b) $t\bar{t}Z$.

CRs follows the same template-based approach as in the signal region. The following classification of background processes describes the estimation of background from each category, including the fake and non-prompt leptons, $t\bar{t}W$, and electrons with misassigned charge, for which a data-driven estimate of the misassignment rate is implemented. The background estimation strategy generally follows that of Ref. [60] except for the data-driven modeling of the $t\bar{t}W$ jet multiplicity distribution, which has been introduced following observation of mis-modeling in the simulation of $t\bar{t}W$ production with more than seven jets.

8.3 $t\bar{t}$ + Heavy Boson Backgrounds

The first class of backgrounds is top associated production $t\bar{t}X$, where $X = W, Z, H$ is a heavy boson, accounting for 68.3% of the expected background in the signal region. These backgrounds carry sizable theoretical uncertainties and have only been measured with limited precision. ATLAS [69, 70] and CMS [71, 72] combinations produced for the ten year anniversary of the Higgs boson discovery are compatible with the SM, but feature uncertainties of roughly 20% on the $t\bar{t}H$ signal strength, and more recent [93, 94] measurements in the $H \to b\bar{b}$ decay channel both show lower than expected signal strengths with similar uncertainties. The NLO prediction for the $t\bar{t}H$ cross-section [21] carries a relative uncertainty of 8%. The latest ATLAS and CMS measurements of $t\bar{t}Z$ and $t\bar{t}W$ are summarized along with the theoretical predictions in fig. 8.3.

$t\bar{t}Z$ and $t\bar{t}H$ measurements are consistent with the SM prediction, and the background rates from these processes are subdominant to $t\bar{t}W$, so their shapes and yields are taken from MC simulation. The $t\bar{t}Z$ modeling is verified in a dedicated validation region consisting of tri-lepton events matching the signal region selection but with the Z veto on $m_{\ell\ell}$ reversed. The post-fit $t\bar{t}Z$ background model agrees well with the data in this validation region as shown in fig. 8.4, indicating the validity of the background model.

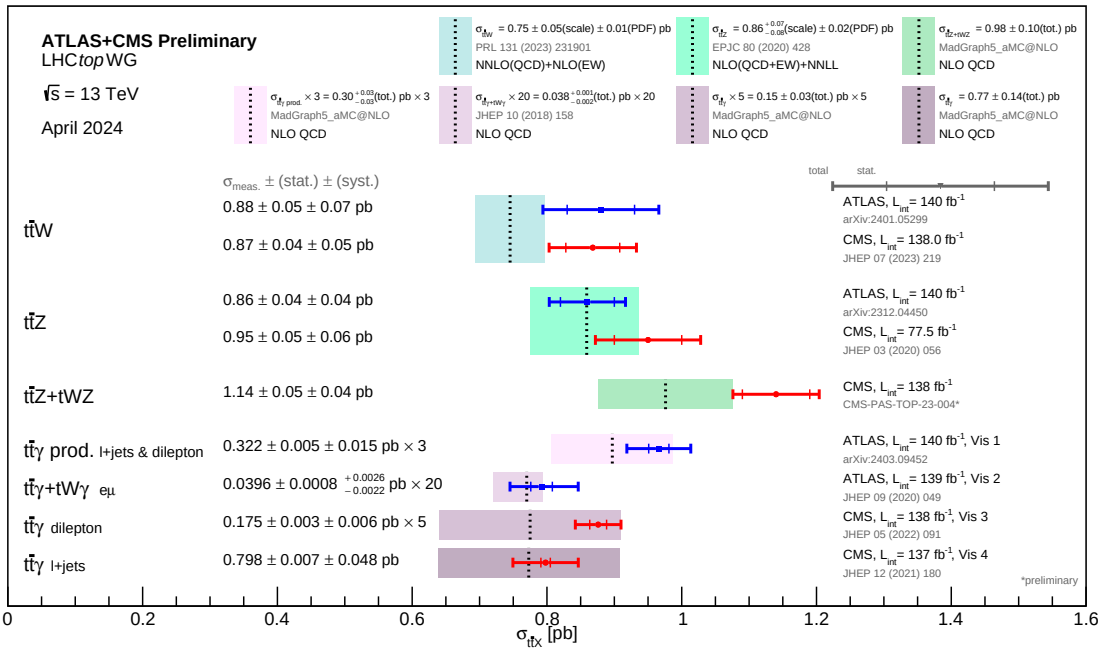


Figure 8.3: Predicted and measured cross-sections for top quark production in association with an electroweak boson at $\sqrt{s} = 13$ TeV. [98]

8.4 Special Treatment of $t\bar{t}W$

The most recent ATLAS and CMS measurements of $t\bar{t}W$ production [60, 239, 240, 241, 242] measure a cross-section about 17% higher than the SM prediction. The CMS measurement [240] also shows a discrepancy in the ratio of $t\bar{t}W^+$ and $t\bar{t}W^-$ production. This ratio differs from one due to the asymmetry between the u- and d-quark parton distribution functions. Studies in the previous ATLAS Run 2 $t\bar{t}t\bar{t}$ measurement [60] observed additional mis-modeling of $t\bar{t}W$ production in association with additional jets, resulting in large systematic uncertainties in a signal-like phase space. Both of these effects may be understood as artifacts of missing higher-order terms in calculations of $t\bar{t}W$ production, given that diagrams with new topologies open up at NNLO, and are consistent with theoretical difficulties in $t\bar{t}W$ modeling [41]. This has motivated a data-driven approach to modeling the $t\bar{t}W$ background and its jet multiplicity distribution. The overall normalization of the $t\bar{t}W^+$ and $t\bar{t}W^-$ backgrounds are floated separately with normalization factors $\mathrm{NF}_{t\bar{t}W^+(4\mathrm{jet})}$ and $\mathrm{NF}_{t\bar{t}W^-(4\mathrm{jet})}$ and the differential cross-section $\sigma^n_{t\bar{t}W} \equiv \frac{d\sigma_{t\bar{t}W}}{dn_j}(n)$ in jet multiplicity n_j for $n_j \geq 4$ is modeled by eq. (8.1) [243, 244, 245, 246]

$$R_n = \frac{\sigma^{n+1}_{t\bar{t}W}}{\sigma^n_{t\bar{t}W}} = a_0 + \frac{a_1}{n-3} \tag{8.1}$$

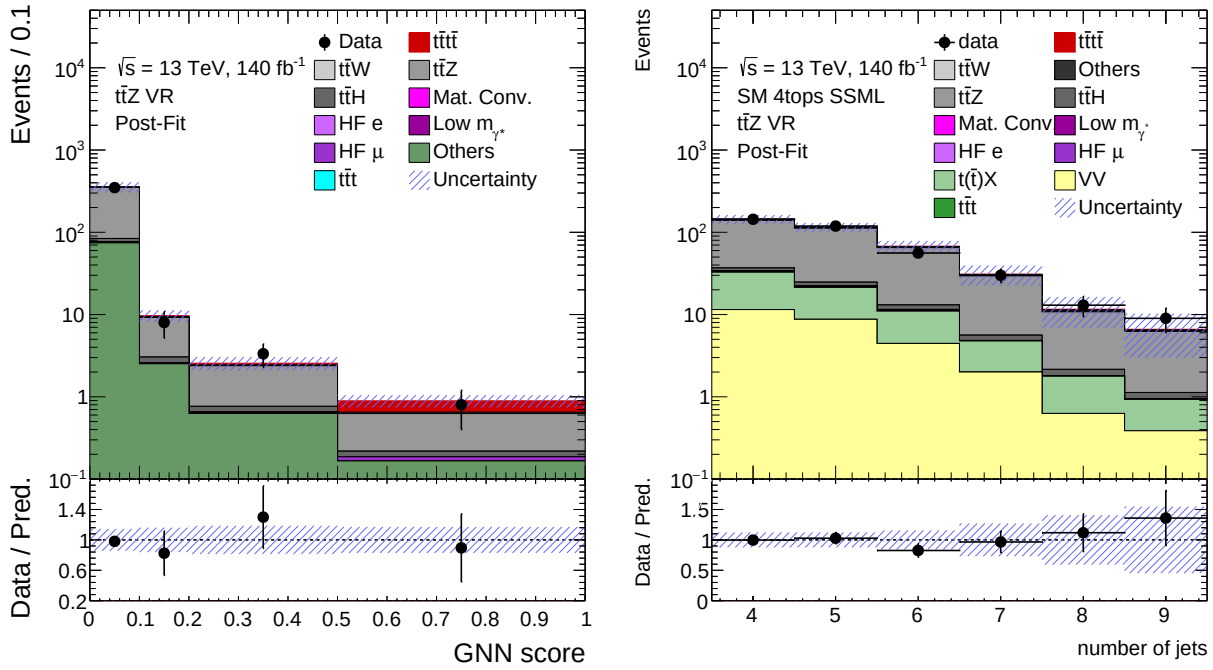


Figure 8.4: Postfit comparison of the data agreement with the background model in the $t\bar{t}Z$ validation region as a function of the GNN score (left) and N_{j} (right). "Others" is dominated by single top+X production, followed by VV.

where a_0, a_1 are free parameters in the final fit and the two terms represent exponential and Poisson scaling of the jet multiplicity distribution, respectively. The exponential term a_0 implies a fixed probability of additional QCD radiation after each additional jet and is a good approximation at large n [246]. The Poisson term is impactful at low n and is expected [246] from perturbative QCD calculations of multijet and Z + jets production, including effects from initial and final state radiation. The cut-off at $n = 4$ comes from the expectation that a dileptonic $t\bar{t}W$ event with no additional QCD radiation will produce four final-state quarks, two b-quarks from the top decays, and two light quarks from the hadronically decaying W boson. Equation (8.1) parameterizes the yield and jet-multiplicity variation, but the relative acceptance in different regions and the kinematic observables of $t\bar{t}W$ events are otherwise taken from MC simulation with a standard treatment of systematic effects detailed in section 8.8. $t\bar{t}W$ events with at least 10 jets are treated as a single template with a normalization term combining the contributions for $N_{\text{j}} = 10, 11, 12$ from eq. (8.1).

The full formula for the modeled $t\bar{t}W$ yield Y_{r_i} in bin r_i within fit region r in terms of the MC prediction $Y^{\pm\text{MC}}_{r_i(j)}$ for the j jet $t\bar{t}W^{\pm}$ yield in that bin and the inclusive MC prediction $Y^{\pm\text{MC}}_{(j)}$ for the j jet $t\bar{t}W^{\pm}$ yield in the pre-selection region is

$$Y_{r_i} = \sum_{j=4}^{9} \left[\left(Y_{(4)}^{+MC} \mathrm{NF}_{t\bar{t}W^+(4\mathrm{jet})} \frac{Y_{r_i(j)}^{+MC}}{Y_{(j)}^{+MC}} + Y_{(4)}^{-MC} \mathrm{NF}_{t\bar{t}W^-(4\mathrm{jet})} \frac{Y_{r_i(j)}^{-MC}}{Y_{(j)}^{-MC}} \right) \prod_{j'=4}^{j} \left(a_0 + \frac{a_1}{j'-3} \right) \right] + \left(Y_{(4)}^{+MC} \mathrm{NF}_{t\bar{t}W^+(4\mathrm{jet})} \frac{Y_{r_i(\geq 10)}^{+MC}}{Y_{(\geq 10)}^{+MC}} + Y_{(4)}^{-MC} \mathrm{NF}_{t\bar{t}W^-(4\mathrm{jet})} \frac{Y_{r_i(\geq 10)}^{-MC}}{Y_{(\geq 10)}^{-MC}} \right) \sum_{j=10}^{12} \prod_{j'=4}^{j} \left(a_0 + \frac{a_1}{j'-3} \right). \tag{8.2}$$

A fit of eq. (8.2) to the nominal MC simulation of $t\bar{t}W$ shown in fig. 8.5 demonstrates that the $t\bar{t}W$ model can capture the expected shape of the jet multiplicity distribution. Additional tests with alternate MC generators and other systematic variations show that variations in the distribution can be captured by the chosen scaling function.

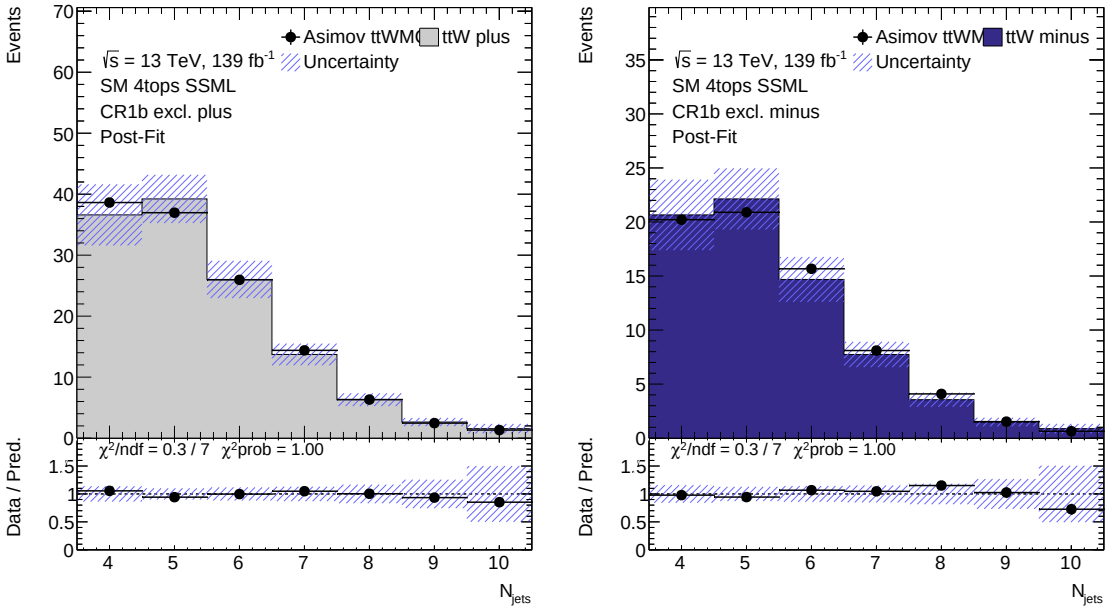


Figure 8.5: Post-fit N_j distribution estimated by Equation (8.2) from a stat-only fit to the $t\bar{t}W$ MC prediction in the CR 1b(+) and CR 1b(-) regions.

Four data control regions are defined to constrain the $t\bar{t}W$ background model, each using the jet multiplicity as an observable to fit the scaling parameters a_0, a_1. The CR $t\bar{t}W^+$+jets and CR $t\bar{t}W^-$+jets regions replace the SR H_T and N_j cuts with $H_T < 500$ GeV or $4 \leq N_\mathrm{j} < 6$ and include only $\mu\mu$ events or $e\mu$ events with $|\eta(e)| < 1.5$ to lessen contamination from reducible backgrounds. To reduce the signal contamination, events with $N_\mathrm{j} < 6$ but $N_b \geq 3$ are also excluded. The CR $t\bar{t}W^+$+jets (CR $t\bar{t}W^-$+jets) region requires the leptons to be

positively (negatively) charged. These regions constrain the jet multiplicity distribution in a region with high b-jet multiplicity, but lack statistical power at high N_{j}.

To complement this, CR 1b(+)and CR 1b(-)regions are similarly split by lepton charge and require at least four jets, of which exactly one is b-tagged, and $H_T > 500$ GeV. An additional cut requires that the two leading leptons are not consistent with a photon conversion or γ^* emission using the DFCAA classification described in chapter 4, which prevents overlap with CRs for fake and non-prompt lepton backgrounds. These regions provide coverage to constrain and verify the scaling relation up to 10 jets.

$t\bar{t}W$ background	a_0	a_1	$\mathrm{NF}_{t\bar{t}W^+(\mathrm{4jet})}$	$\mathrm{NF}_{t\bar{t}W^-(\mathrm{4jet})}$
Value	0.51 ± 0.10	$0.22^{+0.25}_{-0.22}$	$1.27^{+0.25}_{-0.22}$	$1.11^{+0.31}_{-0.28}$

Table 8.2: The $t\bar{t}W$ modeling parameters determined from the fit. The uncertainties include both the statistical and systematic uncertainties. The nominal pre-fit value for $\mathrm{NF}_{t\bar{t}W^+(\mathrm{4jet})}$ and $\mathrm{NF}_{t\bar{t}W^-(\mathrm{4jet})}$ is 1.

The agreement in each of the $t\bar{t}W$ control regions is shown in fig. 8.6, and the fitted normalizations and scaling parameters are shown in table 8.2. The $t\bar{t}W$ model, together with the modeling of sub-dominant backgrounds, captures the behavior of the data in these regions even with large numbers of additional jets. As $t\bar{t}W$ is the only major process in the analysis region with significant charge asymmetry, the yield difference between events with positive and negative total charge can be used to validate the $t\bar{t}W$ jet multiplicity distribution with minimal impact from other background processes. This validation, done inclusively in the combination of all regions used in the final fit, is shown in fig. 8.7 and demonstrates good agreement between the data and the model.

Additional tests show that the data agreement does not significantly change when considering $t\bar{t}W$ events with different numbers of b-jets or when changing the parameterization by choosing alternate forms of eq. (8.1) or decorrelating the scaling parameters a_0, a_1 for the positive and negatively charged regions. These tests show that the chosen treatment gives the best possible modeling of $t\bar{t}W$ in the control regions without adding unnecessary degrees of freedom that could impact the signal sensitivity.

8.5 Fake and Non-Prompt Lepton Backgrounds

The second major class of backgrounds consists of fake and non-prompt leptons and accounts for 15.3% of the expected signal region background. These categories are dominated by $t\bar{t}$ events, but also include single top and V + jets production. They are separated in simulation using truth-level information to identify the source of the lepton candidates. These backgrounds are categorized as follows:

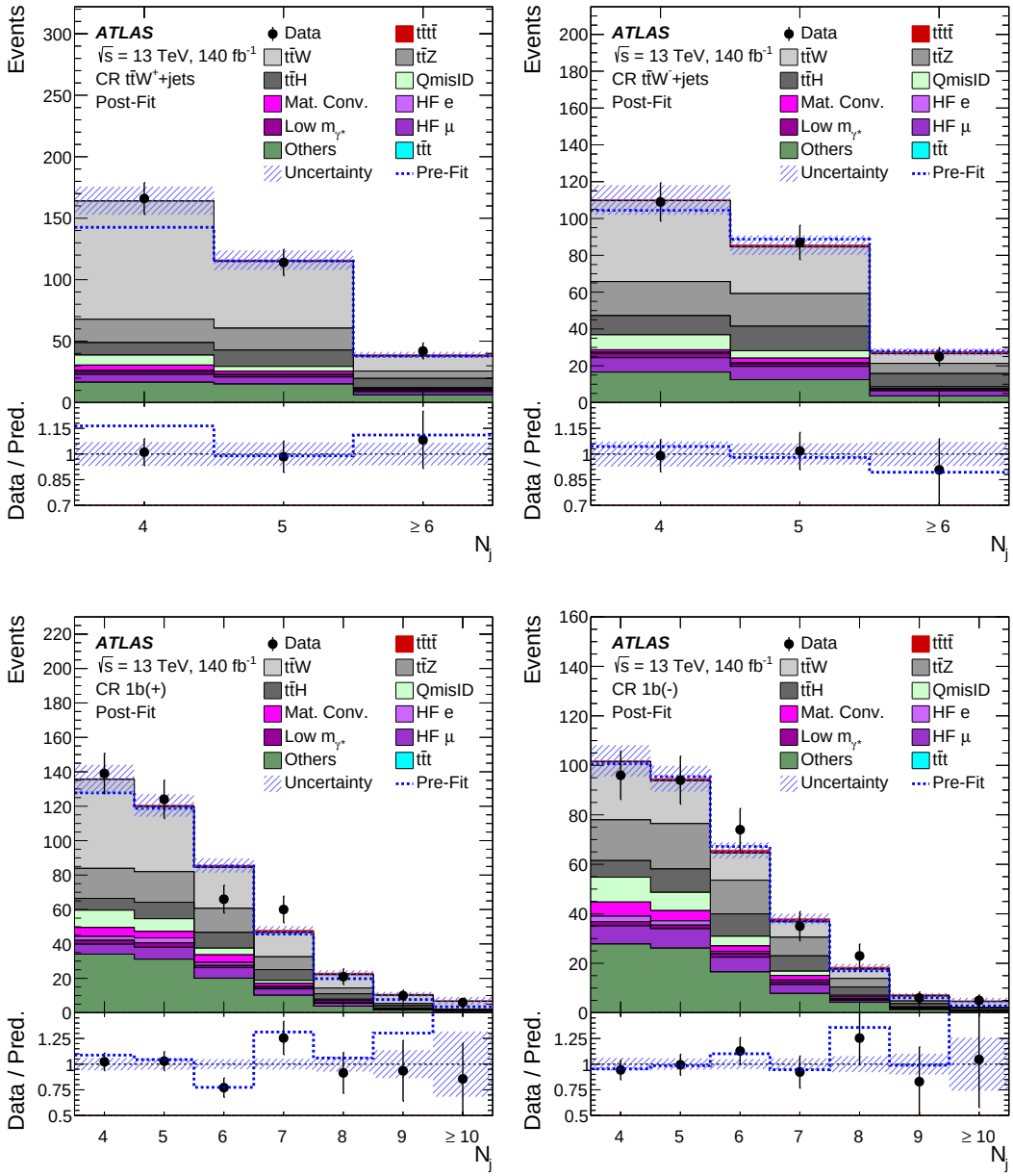


Figure 8.6: Post-fit distributions for the number of jets (N_j) in each of the four $t\bar{t}W$CRs: (a) CR $t\bar{t}W^+$+jets, (b) CR $t\bar{t}W^-$+jets, (c) CR 1b(+)and (d) CR 1b(-). The ratio of the data to the total post-fit prediction is shown in the lower panel. The dashed blue lines show the pre-fit prediction in the upper panel and the ratio of the data to the total pre-fit prediction in the lower panel. The shaded band represents the total post-fit uncertainty in the prediction. Reproduced from [129].

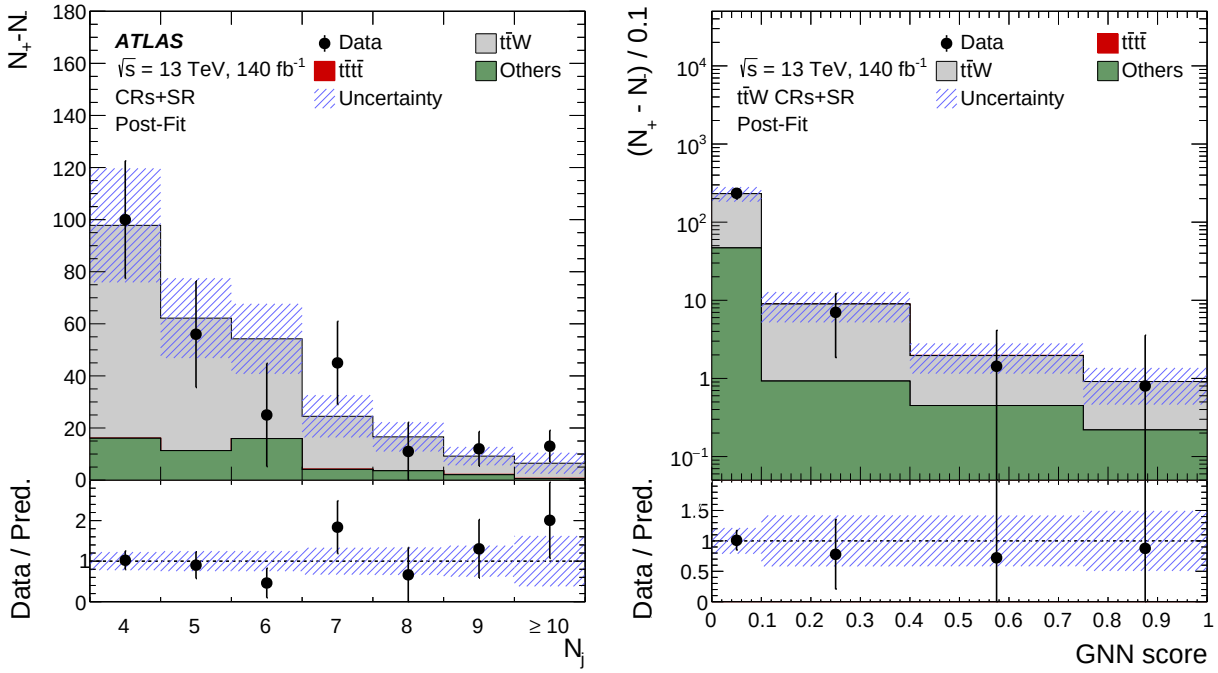


Figure 8.7: Post-fit distribution for the difference between the number of events positive and negative total lepton charge ($N_+ - N_-$) as a function of the N_j (left) and the GNN score (right) in the sum of four $t\bar{t}W$ CRs and the SR. The uncertainties on the normalization factors and the $t\bar{t}W$ modeling parameters are represented by the shaded band. The ratio of the data to the total post-fit prediction is shown in the lower panel. Reproduced from [129].

- Material Conversion (Mat. Conv.): Electrons originating from photon conversion into e^+e^- in the detector material.
- Low m_{γ^*}: Electrons originating from off-shell photon production of low mass e^+e^- pairs. Electrons produced this way are neither fake nor non-prompt, but are included because their signature is similar to that of material conversions at small radii.
- HF e: Electrons originating from the decays of b- or c-hadrons.
- HF μ: Muons originating from the decays of b- or c-hadrons.
- LF: Electrons or muons originating from the decays of light hadrons.
- Other fakes: All other $t\bar{t}$, single top, and V + jets events passing the SSML selection. Includes isolated hadrons faking electrons with large energy depositions in the EM calorimeter or faking muons with punch through into the muon system, among other sources.

The background template for each of these categories is derived from MC simulation of $t\bar{t}$, single top, and V + jets events with the corresponding uncertainties applied. Simulated events are categorized into these six categories based on the truth record. Additionally, material conversions, Low m_{γ^*}, HF e, and HF μ are given floating normalizations due to modeling difficulties with conversion rates, the acceptance of soft electrons, $t\bar{t}$ production in association with additional heavy flavor jets [247], and the rate at which heavy flavor hadrons are reconstructed as leptons. The normalization factors, $\mathrm{NF}_{\text{Mat. Conv.}}$, $\mathrm{NF}_{\text{Low } m_{\gamma^*}}$, $\mathrm{NF}_{\text{HF } e}$, and $\mathrm{NF}_{\text{HF } \mu}$, respectively, are each constrained in dedicated control regions. The method and region definitions largely follow those of Ref. [60].

To constrain $\mathrm{NF}_{\text{Mat. Conv.}}$, the control region CR Mat. Conv. is defined to include same-sign dilepton events in the ee and $e\mu$ channels with 4–6 jets, including at least one b-jet. Additionally, at least one electron must be consistent with a photon conversion according to the DFCAA criteria defined in chapter 4. This selection produces a relatively SR-like selection of photon conversion events to evaluate the rate of photon conversions relative to the MC simulation. This region also contains sizable contamination from QmisID backgrounds from trident production [248], $e^{\pm} \to e^{\pm}e^{+}e^{-}$, which mimics photon conversions.

To constrain $\mathrm{NF}_{\text{Low } m_{\gamma^*}}$, the control region CR Low m_{γ^*} is defined using the same selection as CR Mat. Conv., but with the requirement that one electron is consistent with γ^* production and that neither is consistent with material conversion, again using the DFCAA criteria. The targeted event topology is a $t\bar{t}$ event with a single leptonic W decay and a low mass $e^{+}e^{-}$ pair, one of which is not reconstructed.

To constrain $\mathrm{NF}_{\text{HF } e}$ ($\mathrm{NF}_{\text{HF } \mu}$), CR HF e (CR HF μ) is defined to target $t\bar{t}$ events with two leptonic top decays and an electron (muon) from heavy flavor decays. These regions require at least one jet and exactly one b-tagged jet with $100 < H_T < 300$ GeV and $E_{\mathrm{T}}^{\mathrm{miss}} > 50$ GeV. The regions are defined in the tri-lepton channels with eee and $ee\mu$ ($e\mu\mu$ and $\mu\mu\mu$) events contributing to CR HF e (CR HF μ) and also require the total lepton charge to be ± 1. The target event topology in these regions is a $t\bar{t}$ event decaying into $b\bar{b}\ell^{+}\ell'^{-}\nu\bar{\nu}$, where one b-quark is reconstructed as a lepton which results from a semi-leptonic b-hadron decay. Both regions, and especially CR HF e due to the smaller $b \to e + X$ branching ratios, have limited purity. This is mitigated somewhat by binning the regions in the p_T of the lowest p_T lepton ℓ_3, assumed to be the result of the heavy flavor decay. For HF e and HF μ events, the third lepton is expected to be soft, which is not true of contaminating $t\bar{t}W$, VV, or single-top events.

The agreement of the data with the fitted background model in each of the fake and non-prompt lepton background control regions is shown in fig. 8.8, and the fitted normalization factors are shown in table 8.3.

8.6 Backgrounds from Charge Misidentification

Backgrounds from leptons with mis-identified charges (QmisID) result from migration of common processes with opposite-sign lepton (OS) final states, primarily $t\bar{t}$, into the SSML

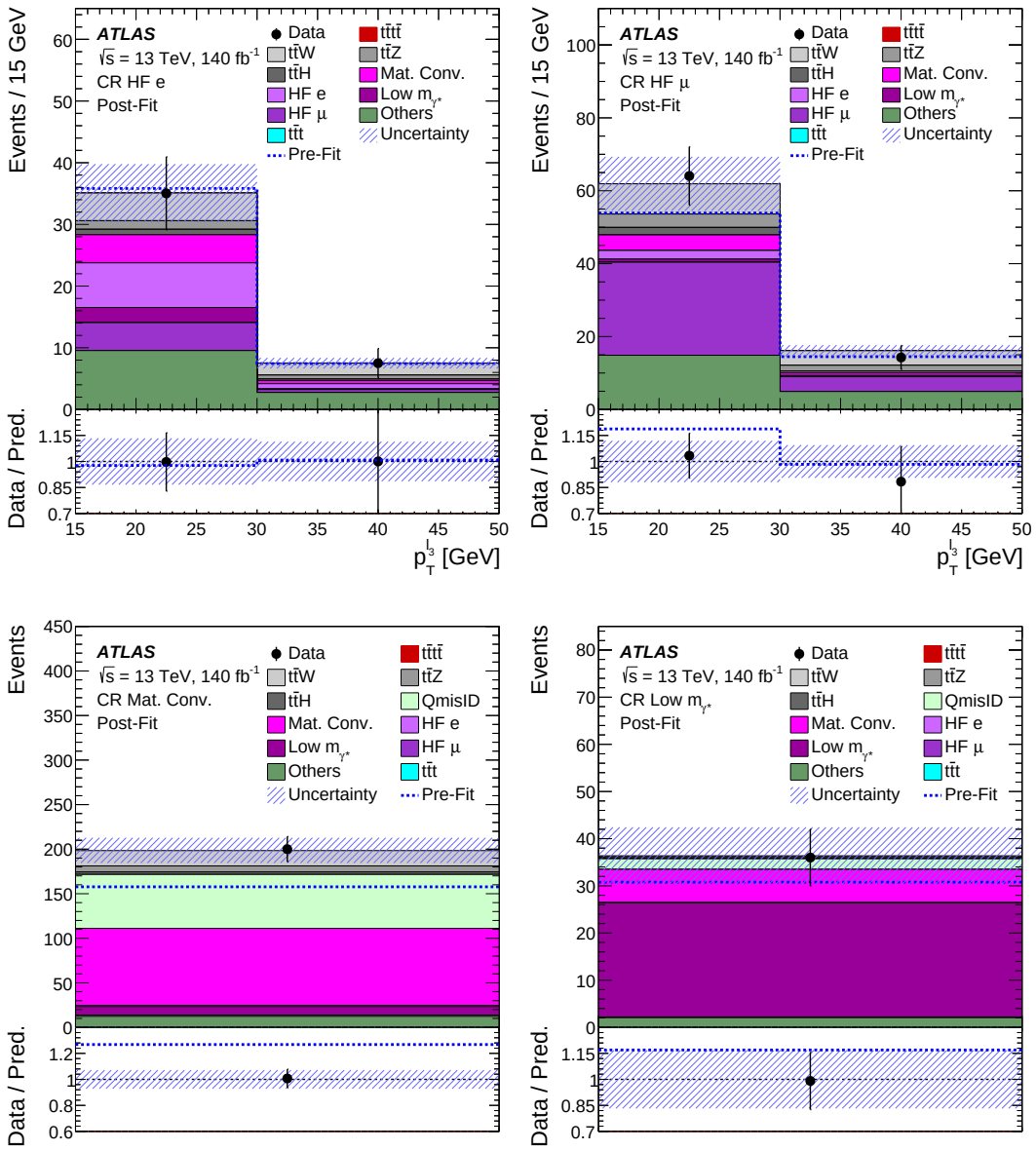


Figure 8.8: Post-fit distributions for the fitted variables in the CRs for the fake/non-prompt lepton background. For (a) CR HF e and (b) CR HF μ, the third-highest lepton p_T is fitted, while for (c) CR Mat. Conv. and (d) CR Low m_{γ^*}, the number of events is used. The ratio of the data to the total post-fit prediction is shown in the lower panel. The dashed blue lines show the pre-fit prediction in the upper panel and the ratio of the data to the total pre-fit prediction in the lower panel. The shaded band represents the total post-fit uncertainty in the prediction. In (a) and (b), the last bin contains the overflow. Reproduced from [129].

Fake/non-prompt background	$\mathrm{NF}_{\text{Mat. Conv.}}$	$\mathrm{NF}_{\text{Low } m_{\gamma^*}}$	$\mathrm{NF}_{\text{HF } e}$	$\mathrm{NF}_{\text{HF } \mu}$
Value	$1.80^{+0.47}_{-0.41}$	$1.08^{+0.37}_{-0.31}$	$0.66^{+0.75}_{-0.46}$	$1.27^{+0.53}_{-0.46}$

Table 8.3: The normalization factors for fake and non-prompt lepton background processes determined from the fit. The uncertainties include both the statistical and systematic uncertainties. The nominal pre-fit value for these factors is 1.

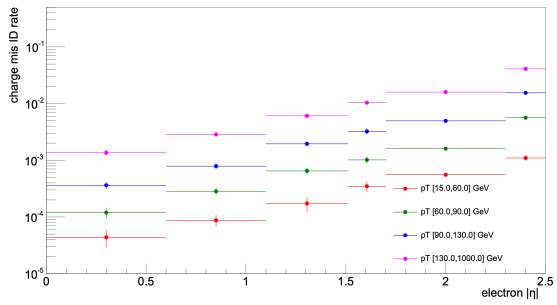
region. Additionally, the charge mis-assignment rate for muons is much smaller than for electrons, so the QmisID background is only relevant for the ee and $e\mu$ channels. These backgrounds amount to 7% of the total expected background in the signal region. The QmisID background contribution in the signal and control regions is estimated with a data-driven determination of the electron charge mis-assignment probability as a function of the electron p_T and η. These charge flip rates are then applied to MC simulations in the OS channel to produce the final background estimate.

The mis-assignment probability is determined for electrons in a $Z \to e^+e^-$ control sample by comparing the opposite-sign (OS) and same-sign (SS) regions near the Z mass peak. The estimation is performed separately for the signal region, CR Mat. Conv., CR Low m_{γ^*}, and CR $t\bar{t}W^{\pm}$+jets using each region's selection in the ee channel with the $m_{\ell\ell}$ cut and the same-sign requirement lifted. An additional estimation includes all other events used in the control and validation regions. A fit to a Breit–Wigner function determines the $m_{\ell\ell}$ window included, which differs between the OS and SS regions due to energy loss in the trident production process that leads to some QmisID events. Events within 4σ of the fitted mean are included in the final fit with a sideband-based subtraction to remove non-Z backgrounds. The expected yields $\bar{N}_{ij}^{SS}$ and $\bar{N}_{ij}^{OS}$ for regions with electrons in kinematics bins i and j are parameterized in terms of the individual charge mis-assignment probabilities ϵ_i and ϵ_j.

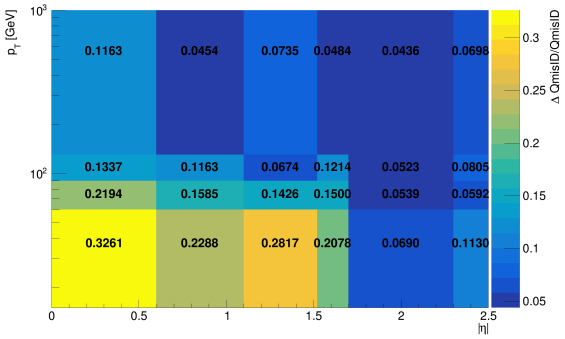
$$\begin{aligned} \bar{N}_{ij}^{SS} &= \left(\bar{N}_{ij}^{SS} + \bar{N}_{ij}^{OS}\right)\left(\epsilon_i(1-\epsilon_j) + \epsilon_j(1-\epsilon_i)\right) \\ \bar{N}_{ij}^{OS} &= \left(\bar{N}_{ij}^{SS} + \bar{N}_{ij}^{OS}\right)\left((1-\epsilon_i)(1-\epsilon_j) + \epsilon_i\epsilon_j\right) \end{aligned} \tag{8.3}$$

A likelihood is constructed as a product of Poisson distributions for the observed yields N_{ij}^{SS}

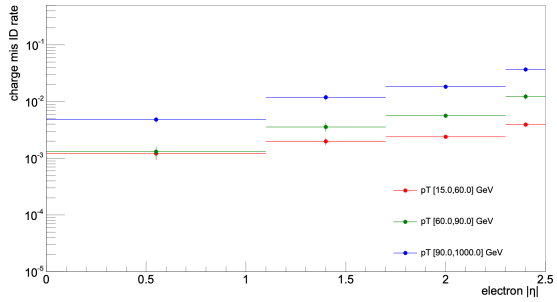
$$\begin{aligned} \mathcal{L}(\epsilon|N_{SS}, N_{OS}) &= \prod_{i,j} f(N_{SS}^{ij}|\bar{N}_{SS}^{ij}) \\ &= \prod_{i,j} \frac{\left[N_{ij}^{tot}\left(\epsilon_i(1-\epsilon_j) + \epsilon_j(1-\epsilon_i)\right)\right]^{N_{SS}^{ij}} \exp\left[N_{ij}^{tot}\left(\epsilon_i(1-\epsilon_j) + \epsilon_j(1-\epsilon_i)\right)\right]}{N_{SS}^{ij}!}, \end{aligned} \tag{8.4}$$



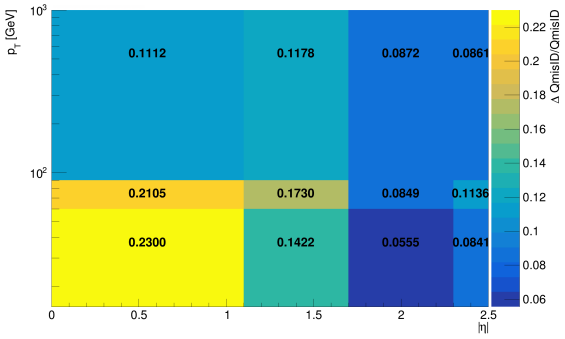
(a) Signal Region: QmisID rate



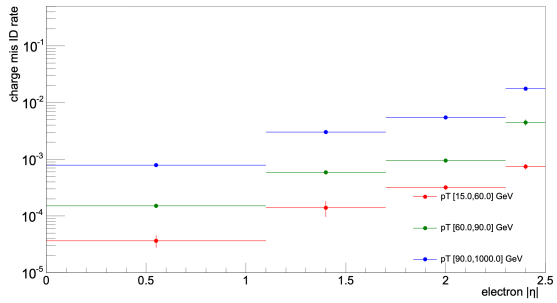
(b) Signal Region: Relative uncertainty on ϵ_i.



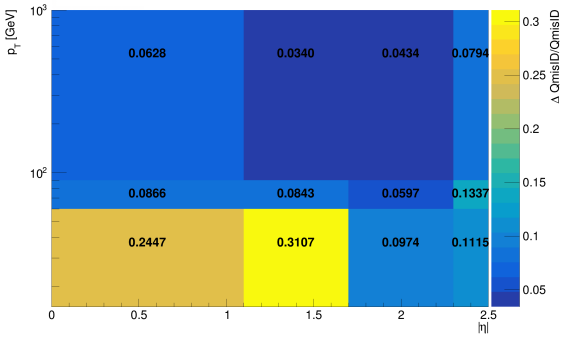
(c) CR Mat. Conv.: QmisID rate



(d) CR Mat. Conv.: Relative uncertainty on ϵ_i.



(e) CR $t\bar{t}W^{\pm}$+jets: QmisID rate



(f) CR $t\bar{t}W^{\pm}$+jets: Relative uncertainty on ϵ_i.

Figure 8.9: Fitted charge misidentification rate and total relative uncertainty for different regions.

where the expected total $OS + SS$ yield has been replaced with the observed yield. This likelihood is minimized to determine the mis-identification rates ϵ_i. Figure 8.9 shows the best fit values of ϵ_i, together with the uncertainties for the signal and control regions with significant QmisID background. The rates range from $< 10^{-4}$ to around 4% and are highest at high p_T, where the curvature used to identify the charge is less, and at high $|\eta|$, where there is more detector material. The QmisID background is then estimated by taking the OS simulation with a weight corresponding to the probability ratio for the event to enter the OS or SS region according to the fitted charge flip rates.

$$w = \frac{\bar{N}_{ij}^{SS}}{\bar{N}_{ij}^{OS}} = \frac{\epsilon_i + \epsilon_j - 2\epsilon_i\epsilon_j}{1 - \epsilon_i - \epsilon_j + 2\epsilon_i\epsilon_j} \tag{8.5}$$

8.7 Rare Backgrounds

The remaining background processes are all considered rare and account for 9.1% of the expected background in the signal region. All are modeled solely from MC simulation, and those that have not been measured previously are assigned large normalization uncertainties, described in section 8.8. The largest among them is $t\bar{t}WW$, followed by $t + X$ and then VV. The small background from three tops production ($t\bar{t}t$) is notable because the final state is very similar to that of the $t\bar{t}t\bar{t}$ signal. $t\bar{t}t$ is not given special treatment in the analysis optimization, but statistical interpretations of the observed signal as $t\bar{t}t$ production are presented in chapter 10. Outside of these interpretations, $t\bar{t}tW$ and $t\bar{t}tq$ are considered as a single background.

The full definition of all signal and control regions used in the final fit is in table 8.4, and the expected background composition in each of these regions is shown in fig. 8.10.

8.8 Systematic Uncertainties

Interpolation, Smoothing, and Pruning

The technical implementation of systematic uncertainties in the likelihood begins with systematic variations corresponding to a one standard deviation fluctuation in the underlying nuisance parameter. These systematic variations are available for each systematic uncertainty and each MC sample and are split into shape and normalization components. The shape variation of each MC template is parameterized with a piecewise linear function of the nuisance parameter α with the expected yield of process ρ in bin j of region i written in terms of normalization effects N_j^ρ and shape effects S_{ij}^ρ.

$$\nu_{ij}^\rho(\alpha) = N_i^\rho(\alpha) S_{ij}^\rho(\alpha) \nu_{ij}^\rho(0).$$

The normalization uncertainty is parameterized with a piecewise exponential function to ensure that the yield expectation is always positive.

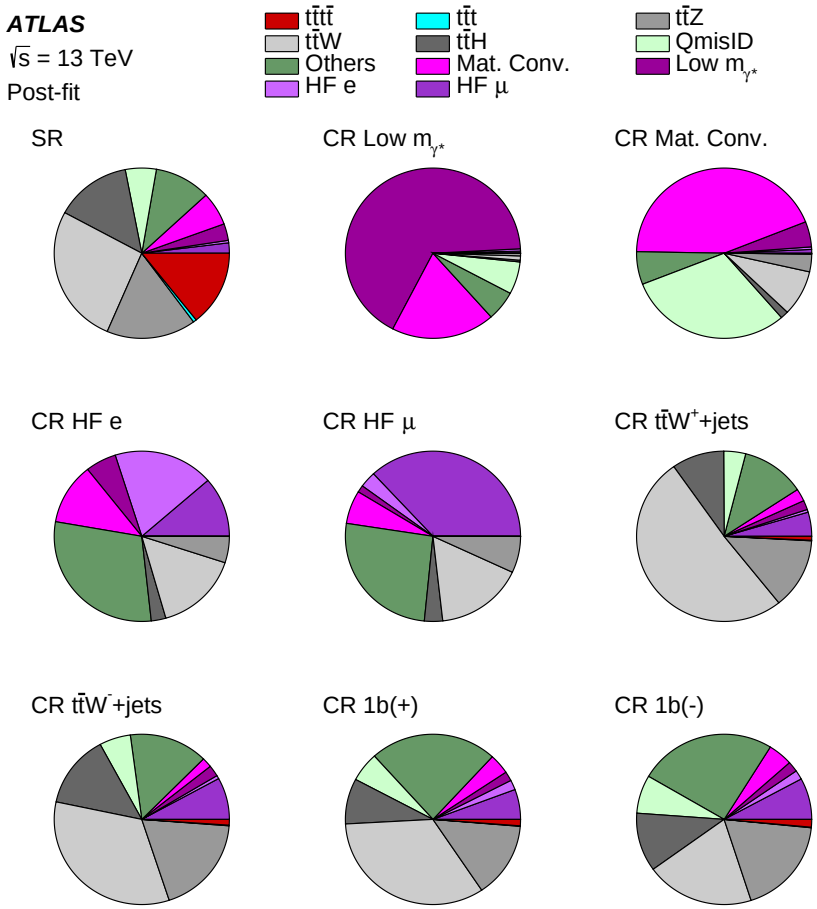


Figure 8.10: Post-fit pie chart for the background composition in each of the analysis regions.

$$N_i^\rho(\alpha) = \begin{cases} \left(\frac{\sum_j \nu_{ij}^\rho(1)}{\sum_j \nu_{ij}^\rho(0)}\right)^\alpha & \alpha \geq 0 \\ \left(\frac{\sum_j \nu_{ij}^\rho(-1)}{\sum_j \nu_{ij}^\rho(0)}\right)^{|\alpha|} & \alpha < 0 \end{cases} \tag{8.6}$$

The shape uncertainty is modeled with a piecewise linear function.

$$S_{ij}^\rho(\alpha) = \begin{cases} \nu_{ij}^\rho(0) + \left(\frac{\nu_{ij}^\rho(1)}{N_i^\rho(1)} - \nu_{ij}^\rho(0)\right)\alpha & \alpha \geq 0 \\ \nu_{ij}^\rho(0) - \left(\frac{\nu_{ij}^\rho(-1)}{N_i^\rho(-1)} - \nu_{ij}^\rho(0)\right)\alpha & \alpha < 0 \end{cases} \tag{8.7}$$

For uncertainties with no shape component, such as cross-section and luminosity uncertainties, the shape component is omitted. These parameterizations ensure that the template

Region	Channel	N_{j}	N_b	Other selection	Fitted variable
CR Low m_{γ^*}	SS, ee or eμ	$4 \leq N_{\mathrm{j}} < 6$	≥ 1	ℓ_1 or ℓ_2 from γ^* decay ℓ_1, ℓ_2 not from photon conversion	event yield
CR Mat. Conv.	SS, ee or eμ	$4 \leq N_{\mathrm{j}} < 6$	≥ 1	ℓ_1 or ℓ_2 from photon conversion	event yield
CR HF μ	e$\mu\mu$ or $\mu\mu\mu$	≥ 1	$= 1$	$100 < H_T < 300$ GeV $E_{\mathrm{T}}^{\mathrm{miss}} > 50$ GeV total charge $= \pm 1$	$p_{\mathrm{T}}^{\ell_3}$
CR HF e	eee or eeμ	≥ 1	$= 1$	$100 < H_T < 275$ GeV $E_{\mathrm{T}}^{\mathrm{miss}} > 35$ GeV total charge $= \pm 1$	$p_{\mathrm{T}}^{\ell_3}$
CR $t\bar{t}W^+$+jets	SS, eμ or $\mu\mu$	≥ 4	≥ 2	$\lvert\eta(e)\rvert < 1.5$ when $N_b = 2$: $H_T < 500$ GeV or $N_j < 6$ when $N_b \geq 3$: $H_T < 500$ GeV total charge > 0	N_{j}
CR $t\bar{t}W^-$+jets	SS, eμ or $\mu\mu$	≥ 4	≥ 2	$\lvert\eta(e)\rvert < 1.5$ when $N_b = 2$: $H_T < 500$ GeV or $N_j < 6$ when $N_b \geq 3$: $H_T < 500$ GeV total charge < 0	N_{j}
CR 1b(+)	2LSS+3L	≥ 4	$= 1$	ℓ_1, ℓ_2 not from photon conversion $H_T > 500$ GeV total charge > 0	N_{j}
CR 1b(-)	2LSS+3L	≥ 4	$= 1$	ℓ_1, ℓ_2 not from photon conversion $H_T > 500$ GeV total charge < 0	N_{j}
SR	2LSS+3L	≥ 6	≥ 2	$H_T > 500$ GeV	GNN score

Table 8.4: Definition of the signal and control regions. The control regions labeled as CR $t\bar{t}W^+$+jets, CR $t\bar{t}W^-$+jets, CR 1b(+) and CR 1b(-) are defined to determine the background modeling parameters of $t\bar{t}W^+$ and $t\bar{t}W^-$ events. The control regions labeled as CR HF e, CR HF μ, CR Mat. Conv. and CR Low m_{γ^*} are defined to determine the normalization of fake/non-prompt lepton background events. N_{j} (N_b) indicates the jet (b-tagged jet) multiplicity in the event. H_T is defined as the scalar sum of the transverse momenta of the isolated leptons and jets. ℓ_1, ℓ_2 and ℓ_3 refer to the highest p_T, second-highest p_T and third-highest p_T leptons, respectively. $\eta(e)$ is electron pseudorapidity. Total charge is the sum of charges for all leptons. GNN denotes the Graph Neural Network discriminant trained to separate $t\bar{t}t\bar{t}$ signal from the background described in chapter 9.

takes the correct form when $\alpha = \pm 1$ and provides a continuous variation of ν with α that is non-differentiable only at $\alpha = 0$ and only for asymmetric uncertainties.

A smoothing algorithm based on convolution of the systematic variation with a kernel is applied to the systematic variation templates of the VV scale variation, generator, parton shower, and certain experimental systematic uncertainties. To limit the computational complexity of the fitting procedure, systematic uncertainties with a maximum variation of

$< 0.5\%$ are removed from the likelihood (pruned). The pruning is performed independently for the normalization and shape components.

Experimental Systematic Uncertainties

Experimental uncertainties describe effects from the resolution, scale, and efficiency for the reconstruction and identification of certain objects, the efficiency of the trigger, and luminosity measurement. Each of these effects is modeled in the MC simulation based on calibration measurements performed with the detector. Statistical or systematic variations of the detector performance from the results of those calibration measurements are modeled with variations of the MC simulation under different conditions.

The most significant systematic uncertainties come from the measured efficiencies of the pseudocontinuous b-tagging working points because a large b-jet multiplicity is one of the primary distinguishing features of the $t\bar{t}t\bar{t}$ signal. These uncertainties enter the selection through requirements on the number of b-tagged jets and through the GNN, which heavily uses the pseudocontinuous b-tagging scores. These uncertainties are parameterized separately for b-, c-, and light jets in bins of p_T. Correlations between p_T bins are considered by performing a principal component analysis to decompose the uncertainty model into 44 (20) uncorrelated variations of the b-tagging efficiency as a function of p_T for b- (light or c-) jets. The derivation of those uncertainties is described in Ref. [249, 250, 251]. The calibration is only performed for $p_T < 400/300/250$ GeV for b-/c-/light jets, so additional extrapolation uncertainties are extracted for jets above these thresholds.

Other sources of experimental uncertainty include jet energy scale and resolution [138, 252, 253], jet-vertex-selection efficiency [254], luminosity [125], electron and muon efficiencies, energy scale, and resolution, including in the trigger [133, 134], $E_{\mathrm{T}}^{\mathrm{miss}}$ [142], and pileup reweighting [255]. These sources of uncertainty are subdominant, and their implementation follows standard calibration procedures developed by the ATLAS Collaboration. The author leaves their description to the provided references.

Signal Modeling Systematic Uncertainties

Uncertainties on the theoretical modeling of $t\bar{t}t\bar{t}$ production are the largest source of systematic uncertainty. The largest contribution comes from the 20% relative uncertainty on the $t\bar{t}t\bar{t}$ cross-section [41] $\sigma_{\mathrm{t\bar{t}t\bar{t}}} = 12.0 \pm 2.4$ fb, which approximates the error in the cross-section calculation from truncating the expansion in α_s and α. The approximation is made following the standard procedure of varying the renormalization scale μ_R. This uncertainty remains dominant when considering the more recent calculation with threshold resummation at NNLL [44] of $\sigma_{t\bar{t}t\bar{t}} = 13.4^{+1.0}_{-1.8}$ fb, though it only affects the measurement of the $t\bar{t}t\bar{t}$ signal strength and the BSM interpretations of the $t\bar{t}t\bar{t}$ signal. Additional uncertainties corresponding to the matching procedure and the combination of parton showering, hadronization, and fragmentation models come from the Sherpa and MadGraph5_aMC@NLO+Herwig samples described in chapter 5. These uncertainties are implemented with the $+1\sigma$ varia-

tion templates defined by the alternate samples and the -1σ variation defined as an anti-symmetric variation, that is, a variation of equal size and opposite sign. All $t\bar{t}t\bar{t}$ samples are normalized to the cross-section prediction of Ref. [41], though acceptance effects may change the expected number of $t\bar{t}t\bar{t}$ events in each region. Shape uncertainties from truncation are estimated by varying the renormalization and factorization scales μ_R and μ_F together by a factor of 2. The envelope of these variations is taken as the uncertainty. An overall normalization uncertainty of 1% is applied to $t\bar{t}t\bar{t}$ production to account for uncertainties in the PDFs. This is estimated from the variation produced by 100 ensembles from the `NNPDF30_nlo_as_0118` PDF set according to the PDF4LHC prescription [256].

Background Modeling Systematic Uncertainties

Uncertainties on the background modeling are subdominant to those on the signal modeling, and effects modeled with the floating normalization factors and $t\bar{t}W$ jet multiplicity scaling parameters are considered part of the statistical uncertainty. Those uncertainties would shrink if the measurement were performed with a larger dataset. Of the remaining background uncertainties, those from $t\bar{t}H$ and $t\bar{t}W$ are largest.

$t\bar{t}H$ and $t\bar{t}Z$ receive cross-section uncertainties of 10% and 12%, respectively, following the recommendation of Ref. [21], and an additional 1% normalization uncertainty from the PDFs is added as in the case of $t\bar{t}t\bar{t}$. There are no normalization uncertainties for $t\bar{t}W$ as they can be absorbed in the floating normalization. The generator uncertainties for $t\bar{t}W$ using the MadGraph5_aMC@NLO FxFx sample, for $t\bar{t}Z$ using the Sherpa sample, and for $t\bar{t}H$ using the MadGraph5_aMC@NLO sample, are all implemented in the same way as the $t\bar{t}t\bar{t}$ generator uncertainties, as is the $t\bar{t}H$ parton showering uncertainty using the comparison to Herwig. Other processes receive cross-section uncertainties $\Delta\sigma$ outlined in table 8.5.

Process	$t\bar{t}t\bar{t}$ [41]	$t\bar{t}Z$ [21]	$t\bar{t}H$ [21]	Single Top	tZq+tWZ [257, 258, 259]
$\Delta\sigma/\sigma$ [%]	20	12	30	30	30

Process	$VV(\leq 3j)$ [260]	$VV(4j)$ [260]	$VV(\geq 5j)$ [260]
$\Delta\sigma/\sigma$ [%]	20	50	60

Process	$t\bar{t}t$	$t\bar{t}WW$ [154]	$t\bar{t}$	$t\bar{t}$ + Fake/Non-prompt lepton from light jet [60, 261]
$\Delta\sigma/\sigma$ [%]	35	50	30	99

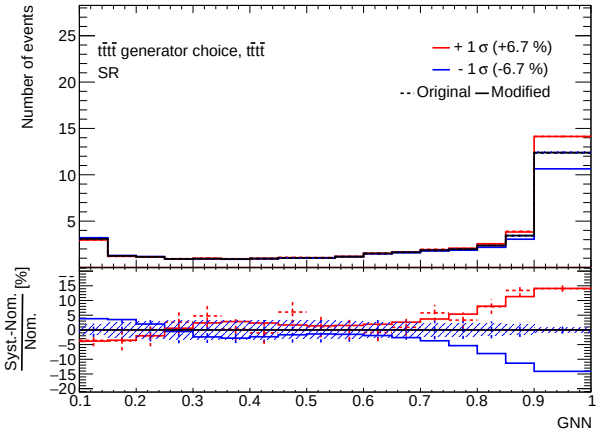
Process	$VH + VVV + t\bar{t}ZZ + t\bar{t}WZ + t\bar{t}HH + t\bar{t}WH$ [60, 261, 154]
$\Delta\sigma/\sigma$ [%]	50

Table 8.5: Cross-section uncertainties for various processes used in the $t\bar{t}t\bar{t}$ analysis. The provided citations show the theoretical calculations or experimental measurements that form the basis of the cross-section uncertainties.

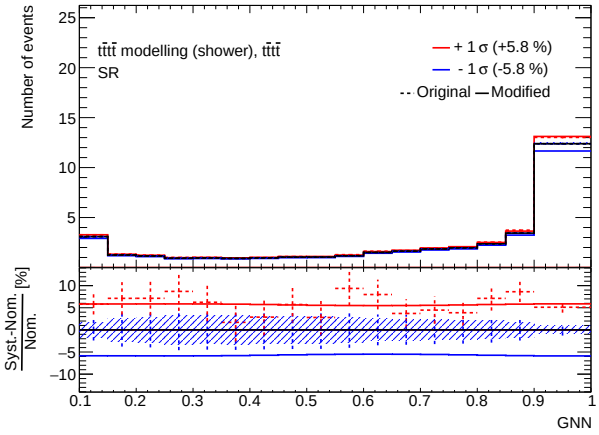
Major backgrounds ($t\bar{t}W/Z/H$, VV, and $t\bar{t}t$) additionally receive uncertainties corresponding to the variation of the renormalization and factorization scales μ_R and μ_F following the prescription used for $t\bar{t}t\bar{t}$. The most signal-like region of the GNN distribution is enhanced in backgrounds with additional b-jets either from mis-tagging of light jets or from the production of additional heavy flavor. All backgrounds receive an additional uncertainty to account for modifications to the production of additional b-quarks. These uncertainties are motivated by observations of mis-modeling in $t\bar{t}b\bar{b}$ [262]. Separate uncertainties of 50% each are applied to events with exactly three or > 3 truth jets containing a b-hadron. Backgrounds with similar topologies are grouped as the production of additional b-quarks should behave similarly for similar topologies. The groups are $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}H$, $t\bar{t}VV$+Single Top, $VV + VVV + VH$, $t\bar{t}$+jets, and $t\bar{t}t$. In the case of $t\bar{t}t$, only the $> 3b$ uncertainty is applied as three b-jets are expected from the top quarks.

A single uncertainty on the data-driven estimate of the charge flip rate used to model the QmisID background is derived from the combination of the statistical uncertainty from the data-driven fit, variations from the choice of Z mass window, and non-closure measured by applying the method to simulated $Z \to e^+e^-$ events instead of data. This uncertainty ranges from 5%–30% depending on the electron kinematics. Uncertainties on the extrapolation of the Material Conversion and Low m_{γ^*} backgrounds from the CR Mat. Conv. and CR Low m_{γ^*} control regions defined using the DFCAA selection for non-prompt leptons to the signal region, which is inclusive in the DFCAA category, are evaluated by comparing the extrapolation efficiency in simulated and recorded events with a Z boson decaying into two muons and an additional photon from initial state radiation. A flat uncertainty of 33% (21%) is applied for events with electrons from Material Conversion (γ^*) identified as non-prompt by the DFCAA selection. A single relative uncertainty is sufficient to cover the extrapolation because both the overall normalizations of the backgrounds are floated. Shape uncertainties on the predicted HF e and HF μ backgrounds are derived from a region with loosened lepton isolation cuts, which is enhanced in leptons from heavy flavor decays. The uncertainty is derived from the ratio of the data in this region to the MC prediction and ranges from 20%–100% in different bins of the analysis. All other backgrounds with fake and non-prompt leptons are minor and are modeled directly from the MC simulation. They are collectively given a flat 30% normalization uncertainty. In addition to these systematic uncertainties in the data-driven estimation of backgrounds from fake and non-prompt lepton backgrounds, the standard theoretical and experimental uncertainties on the MC simulation of these processes, including on $t\bar{t}$+jets production with additional b-jets, also affect these backgrounds.

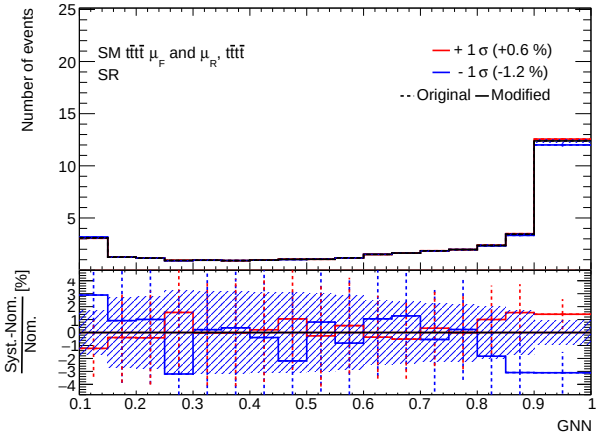
The systematic variations for the theoretical uncertainties with the largest impact on the $t\bar{t}t\bar{t}$ cross-section measurement are shown in figs. 8.11 to 8.15. The most impactful uncertainties are generally those affecting the signal and background yields in the most signal-like bins of the GNN score distribution. Thus, the $t\bar{t}t\bar{t}$ modeling uncertainties and the uncertainties placed on background processes with the radiation of additional b-quarks are all highly ranked, since $t\bar{t}t\bar{t}$ events and background events with additional b-jets are among the most signal-like. The ranking of the uncertainties with the largest impact on the



(a) $t\bar{t}t\bar{t}$ generator systematics

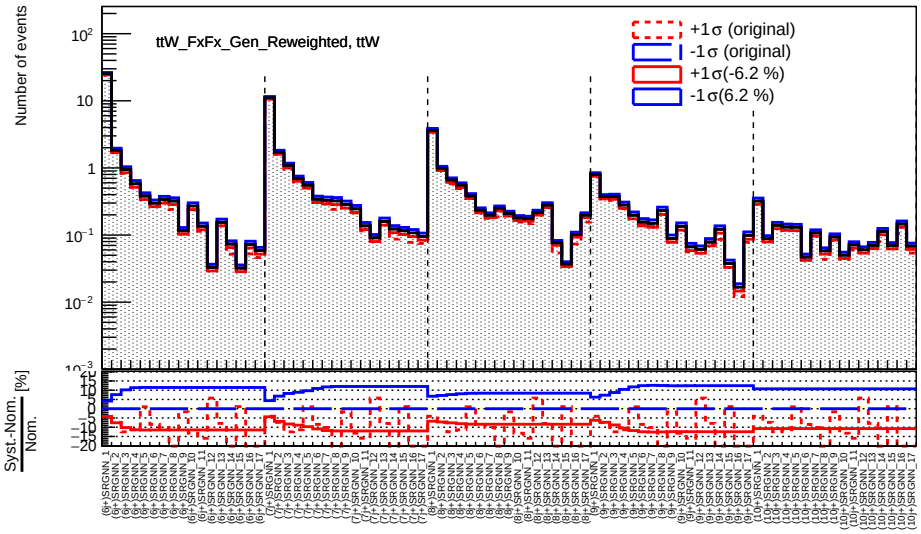


(b) $t\bar{t}t\bar{t}$ parton shower systematics

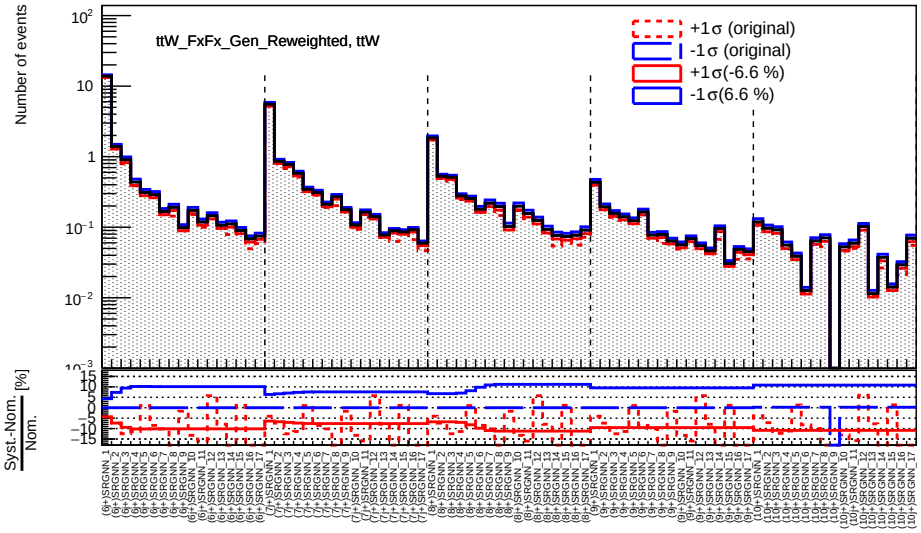


(c) $t\bar{t}t\bar{t}$ scale variation systematics

Figure 8.11: $t\bar{t}t\bar{t}$ generator (a), parton shower (b), and scale variation (c), systematic variations in the signal and control regions. The dashed lines are variations derived directly from the alternate samples' yields, and the solid lines are the variations after smoothing.

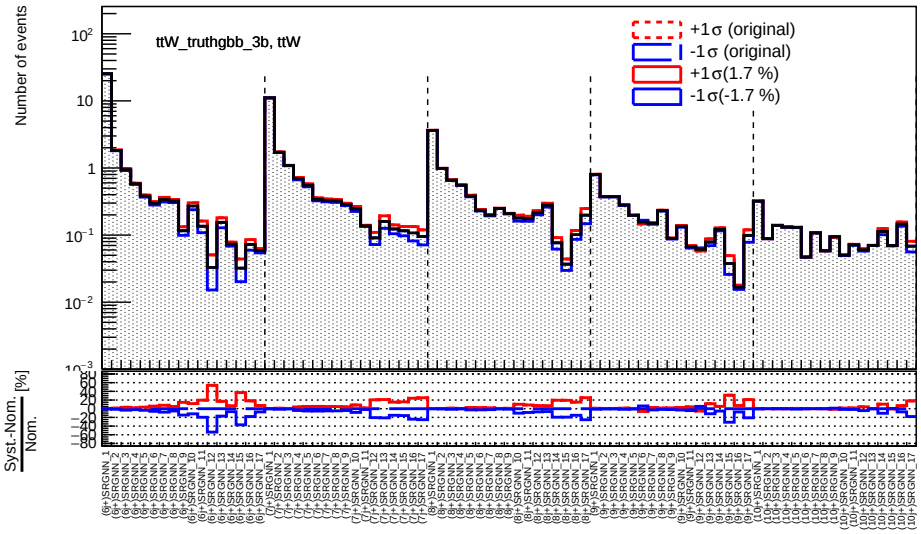


(a) $t\bar{t}W^+$ Generator

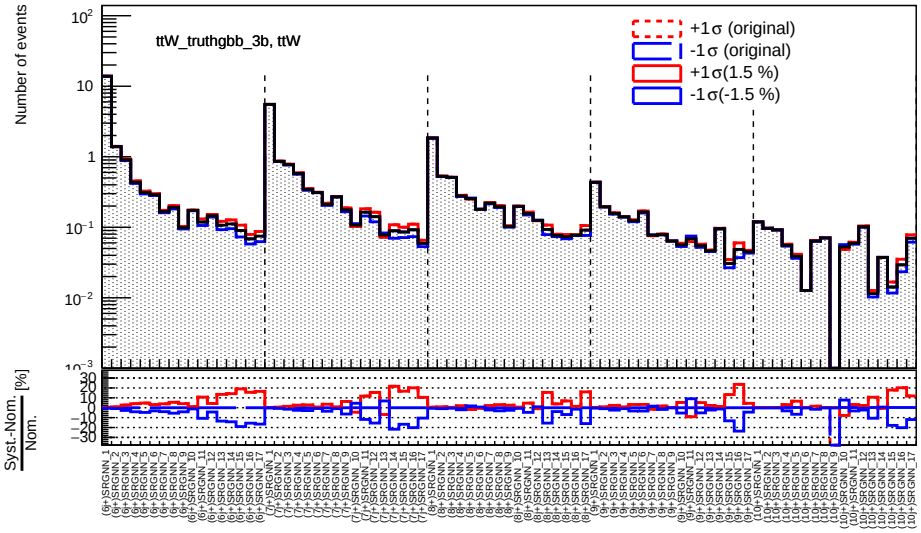


(b) $t\bar{t}W^-$ Generator

Figure 8.12: $t\bar{t}W^+$ (a) and $t\bar{t}W^-$ (b) systematic variation from uncertainties for Generator choice in the signal region. Dashed lines are systematic variations derived from the yields, and solid lines are the systematic variations after smoothing. The brackets in the bin labels indicate N_{j} and the sign of the total charge.

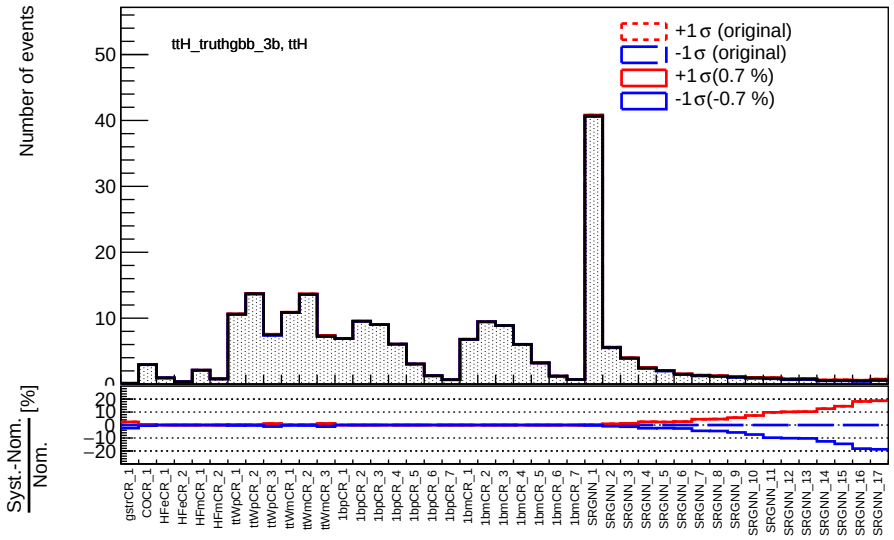


(a) $t\bar{t}W^+$ 1b

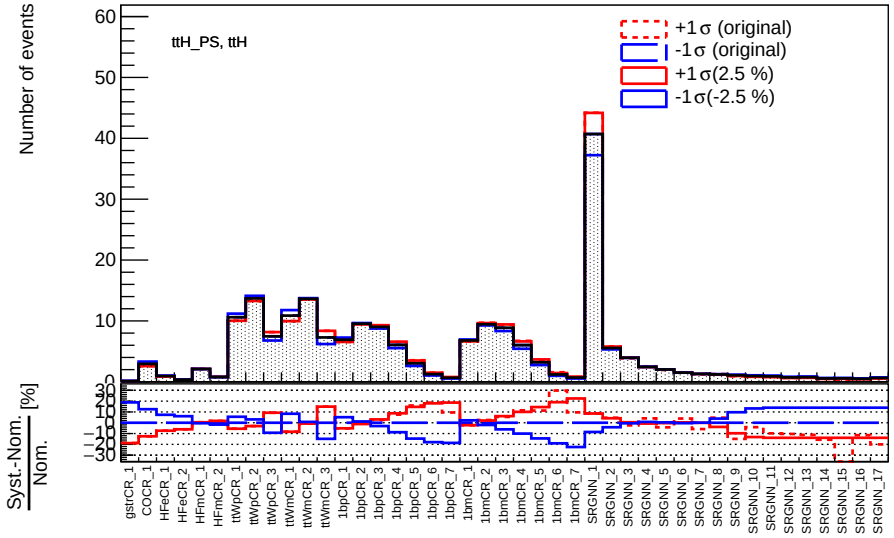


(b) $t\bar{t}W^-$ 1b

Figure 8.13: $t\bar{t}W^+$ (a) and $t\bar{t}W^-$ (b) systematic variation from uncertainties on $t\bar{t}W$ production with one additional b-jet in the signal region. Dashed lines are systematic variations derived from the yields, and solid lines are the systematic variations after smoothing. The brackets in the bin labels indicate N_{j} and the sign of the total charge.

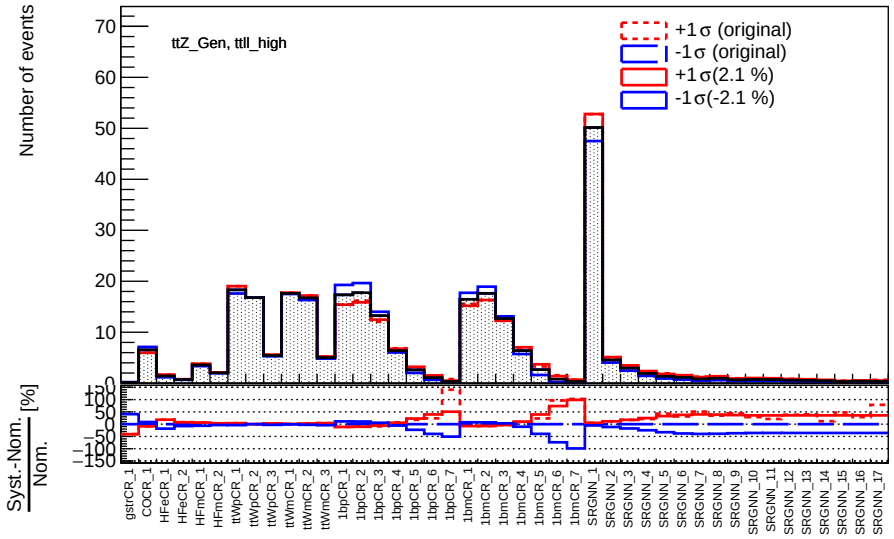


(a) $t\bar{t}H$ + 1 true b-jet systematics



(b) $t\bar{t}H$ parton shower systematics

Figure 8.14: $t\bar{t}H$ additional b-jet and $t\bar{t}H$ parton shower systematic variations in the signal and control regions. The dashed lines are variations derived directly from the alternate samples' yields, and the solid lines are the variations after smoothing.



(a) $t\bar{t}Z$ generator systematics

Figure 8.15: $t\bar{t}Z$ generator systematic variations in the signal and control regions. The dashed lines are variations derived directly from the alternate samples' yields, and the solid lines are the variations after smoothing.

signal strength measurement is shown in fig. 8.16, and the contributions of various sources of uncertainty on the cross-section measurement are listed in section 8.8.

Monte Carlo Statistical Uncertainty

As each template is composed of a histogram filled with Monte Carlo events, the expectation in each bin carries an uncertainty corresponding to the statistical power of the MC samples used to fill the histograms. This uncertainty is estimated following the method of Ref. [263] adapted as described in Ref. [183], with the signal and background expectations in each bin receiving a normalization factor $\gamma^s_{ij}, \gamma^b_{ij}$. The γ factors are nuisance parameters constrained by a gamma distribution with standard deviation corresponding to the statistical uncertainty of the MC prediction $\sigma^s_{MC,ij}$ or $\sigma^b_{MC,ij}$.

$$f(\nu^b_{ij}|\gamma^b_{ij}) = \text{Pois}\left(\frac{{\sigma^b_{ij}}^2}{{\nu^b_{ij}}^2}\middle|\gamma^b_{ij}\frac{{\sigma^b_{ij}}^2}{{\nu^b_{ij}}^2}\right)$$

Here, the statistical power of an MC sample with events weighted by weights w_i is written ${\sigma^b_{MC,ij}}^2 = \sum_i w_i^2$. This constraint is maximized when $\gamma^b_{ij} = 1$ and has a width corresponding

to the relative MC statistical uncertainty $\sigma_{ij}^{b}/\nu_{ij}^{b}$. This treatment treats the value of the MC prediction as the result of an auxiliary measurement, i.e., the simulation, and the uncertainty as a constant chosen by the experimenter. The contribution of these uncertainties is negligible except in the high GNN score region, where the background is sparsely populated. Even in this region, the MC statistical uncertainty is highly subdominant. Note that the MC statistical uncertainties on systematic variation samples and BSM signals are not included in the analysis, but are checked to be small.

Uncertainty source	$\Delta\sigma$ [fb]		$\Delta\sigma/\sigma$[%]	
Signal modeling				
$t\bar{t}t\bar{t}$ generator choice	+3.7	−2.7	+17	−12
$t\bar{t}t\bar{t}$ parton shower model	+1.6	−1.0	+7	−4
Other $t\bar{t}t\bar{t}$ modeling	+0.8	−0.5	+4	−2
Background modeling				
$t\bar{t}H$+jets modeling	+0.9	−0.7	+4	−3
$t\bar{t}W$+jets modeling	+0.8	−0.8	+4	−3
$t\bar{t}Z$+jets modeling	+0.5	−0.4	+2	−2
Other background modeling	+0.5	−0.4	+2	−2
Non-prompt leptons modeling	+0.4	−0.3	+2	−2
$t\bar{t}t$ modeling	+0.3	−0.2	+1	−1
Charge misassignment	+0.1	−0.1	+0	−0
Instrumental				
Jet flavor tagging (b-jets)	+1.1	−0.8	+5	−4
Jet uncertainties	+1.1	−0.7	+5	−3
Jet flavor tagging (light-flavor jets)	+0.9	−0.6	+4	−3
Jet flavor tagging (c-jets)	+0.5	−0.4	+2	−2
Simulation sample size	+0.4	−0.3	+2	−1
Other experimental uncertainties	+0.4	−0.3	+2	−1
Luminosity	+0.2	−0.2	+1	−1
Total systematic uncertainty	+4.6	−3.4	+20	−16
Statistical				
Intrinsic statistical uncertainty	+4.2	−3.9	+19	−17
$t\bar{t}W$+jets normalization and scaling factors	+1.2	−1.1	+6	−5
Non-prompt leptons normalization (HF, Mat. Conv., Low m_{γ^*})	+0.4	−0.3	+2	−1
Total statistical uncertainty	+4.7	−4.3	+21	−19
Total uncertainty	+6.6	−5.5	+29	−25

Table 8.6: List of the uncertainties in the cross-section $\sigma_{t\bar{t}t\bar{t}}$, grouped in categories. The quoted values are obtained by repeating the fit, fixing a set of nuisance parameters of the sources corresponding to each category, and subtracting in quadrature the resulting uncertainty from the total uncertainty of the nominal fit presented in the last row. The total uncertainty is different from the sum in quadrature of the components due to correlations among nuisance parameters. [129]

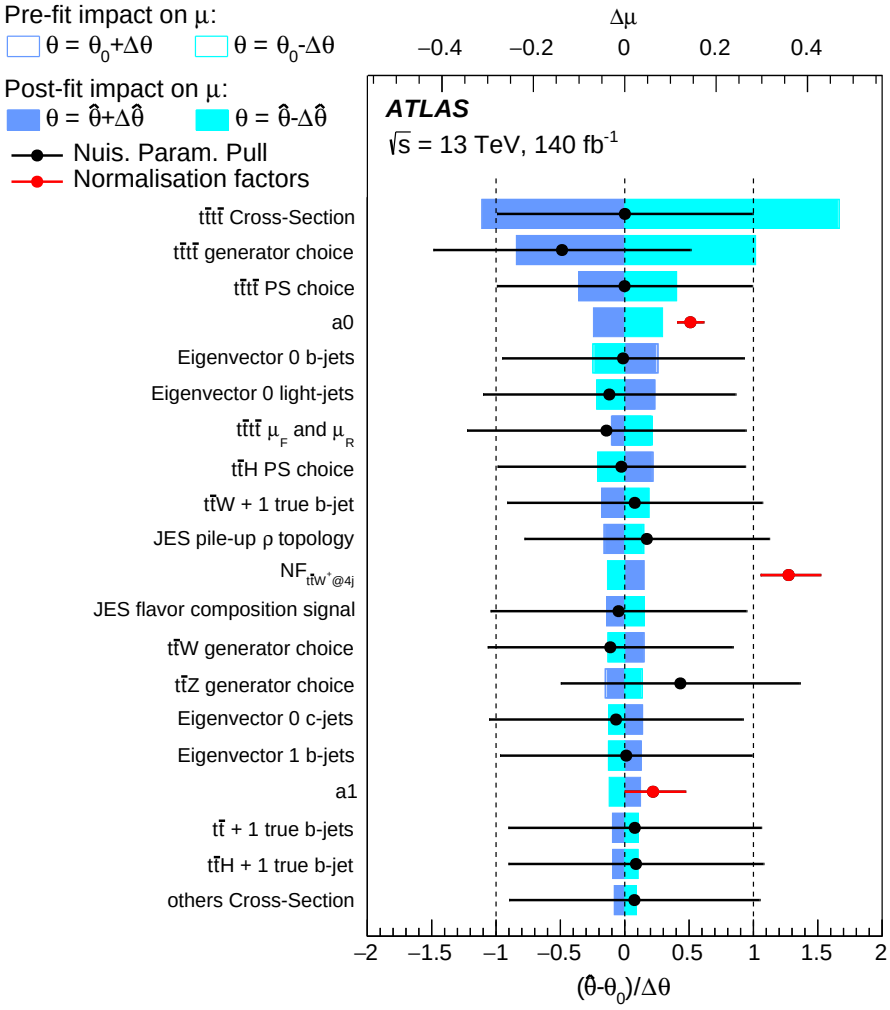


Figure 8.16: Ranking of the nuisance parameters included in the fit according to their impact on the $t\bar{t}t\bar{t}$ signal strength μ. Only the 20 most highly ranked nuisance parameters are shown. The empty blue rectangles correspond to the pre-fit impact on μ and the filled blue ones to the post-fit impact on μ, both referring to the upper x-axis scale. The impact of each nuisance parameter, $\Delta\mu$, is computed by comparing the nominal best-fit value of μ with the result of the fit when fixing the nuisance parameter to its best-fit value, $\hat{\theta}$, shifted by its pre-fit (post-fit) uncertainties $\pm\Delta\theta$ ($\pm\Delta\hat{\theta}$). For most of the nuisance parameters shown in this figure, the pre- and post-fit impacts coincide since the corresponding nuisance parameter is not constrained. The points show the pulls of the free parameters of the fit (red) and those constrained by a penalty factor (black), relative to their nominal values, θ_0. The nominal value for the $\mathrm{NF}_{t\bar{t}W^{+}(4\mathrm{jet})}$is $\theta_0 = 1$ while it is $\theta_0 = 0$ for all other parameters. The pulls of the most important parameters and their relative post-fit errors, $\Delta\hat{\theta}/\Delta\theta$, are referenced to the lower x-axis scale. Reproduced from [129].

Chapter 9

Graph Neural Network Discriminator for $t\bar{t}t\bar{t}$

This chapter describes studies of the application of GNNs to the measurement of $t\bar{t}t\bar{t}$ production in the SSML channel and especially the development of a GNN classifier used to discriminate $t\bar{t}t\bar{t}$ events from background processes in the observation of $t\bar{t}t\bar{t}$ production. The decay of a same-sign lepton $t\bar{t}t\bar{t}$ event is expected to produce four b-jets, four light flavor jets, two leptons, and an $E_{\mathrm{T}}^{\mathrm{miss}}$ signature from the combination of two neutrinos. These objects are not distributed randomly, but instead are arranged into the decays of four top quarks and the subsequent decays of four W bosons. Background events with a similar multiplicity typically contain two top quarks, three or more electroweak boson decays, and additional jets from QCD radiation. The differences between the $t\bar{t}t\bar{t}$ signal and the backgrounds are difficult to extract from object counting or simple kinematic observables. A GNN classifier is better suited to exploiting the complex geometrical structure of these events in momentum space.

9.1 GNN Implementation for Signal/Background Separation

The $t\bar{t}t\bar{t}$ signal/background separation GNN is implemented in the `root_gnn` package described in chapter 7. The graph definition uses all jets, electrons, and muons satisfying the reconstruction and selection criteria described in chapter 4 as nodes. An additional node represents $E_{\mathrm{T}}^{\mathrm{miss}}$. The node features are $(p_T, \eta, \phi, E_0, \mathrm{PCBT}, q, \mathrm{PID})$, where $E_0 = p_T \cosh \eta$ is the energy under the assumption that the object is massless or highly relativistic, and PCBT is the pseudocontinuous b-tagging score indexed so that a jet with PCBT of one passes no b-tagging threshold and a jet with PCBT of five passes the tightest b-tagging threshold. PID is an integer indicating the type of object represented by the node. $q = \pm 1$ is the reconstructed charge of an electron or muon, which is expected to help as the large $t\bar{t}W$ background is not charge-symmetric, and same-sign same flavor lepton pairs are more likely to result from

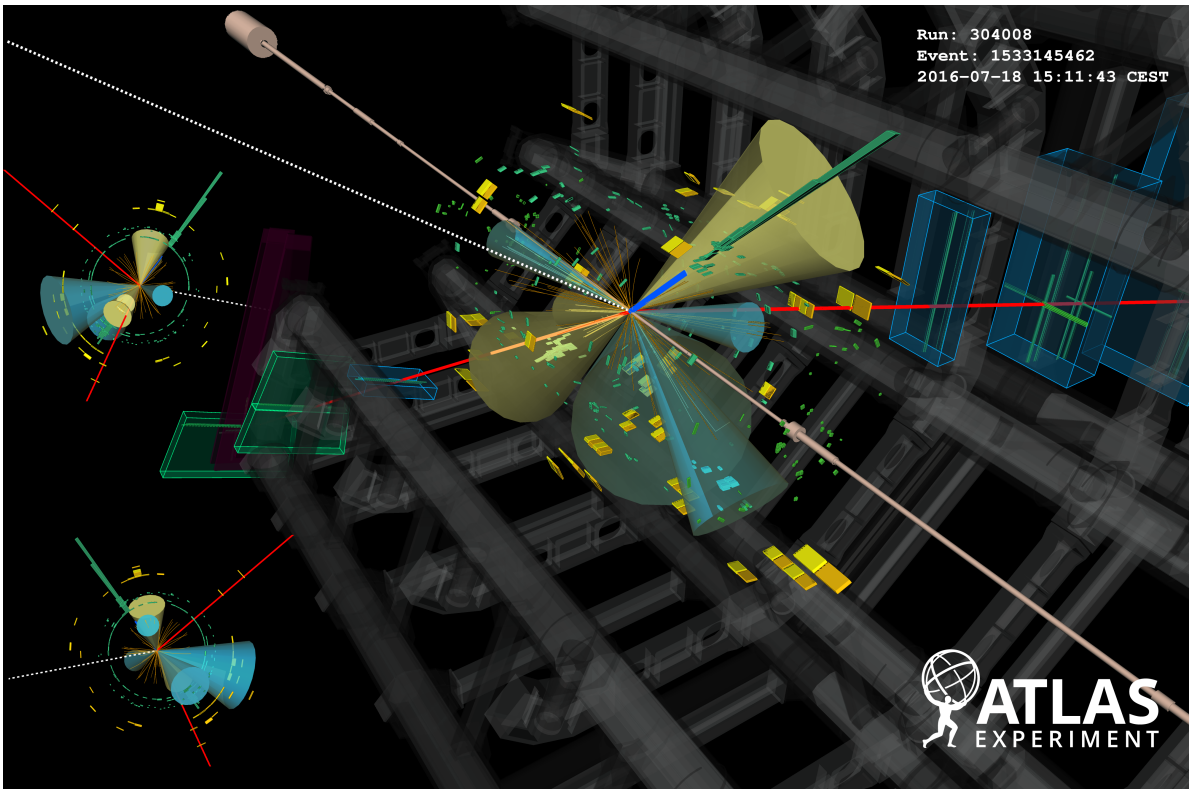


Figure 9.1: Event display of a candidate four-top-quark event from data collected in 2016 (Run 304008, Event 1533145462). The event contains seven jets (four of them are b-tagged); three of the top quarks produce leptons in their decays (two muons, shown in red, and one electron, shown in blue), and the fourth top quark decays into jets. Tracks of charged particles in the inner detector are visualized as orange lines. Green rectangles correspond to energy deposits in cells of the electromagnetic calorimeter, while yellow rectangles correspond to energy deposits in cells of the hadron calorimeter. Also shown are the muon chambers associated with the two muon tracks (blue and green boxes). The jets (b-tagged jets) are shown as yellow (azure) cones. The direction of the missing transverse momentum is indicated by a dotted line.

backgrounds with Z/γ^* bosons. Placeholder values of zero are assigned for features that are not defined for a given object. In the case of $E_{\mathrm{T}}^{\mathrm{miss}}$, this includes setting $\eta = 0$, which further implies that $E_0 = p_T$. The edges are assigned edges features of $(\Delta\eta, \Delta\phi, \Delta R)$ where $\Delta\phi$ is signed and in $(-\pi, \pi]$ and $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$. The number of jets N_{j} is used as a global feature. An example graph of a $t\bar{t}t\bar{t}$ candidate event from 2016 collision data is illustrated in fig. 9.2. The event, which is also shown in the detector display in fig. 9.1, has four reconstructed b-jets and three leptons, e^-, μ^+, and μ^-. The GNN assigns the event a score of 0.988, corresponding to a high likelihood of being a $t\bar{t}t\bar{t}$ event.

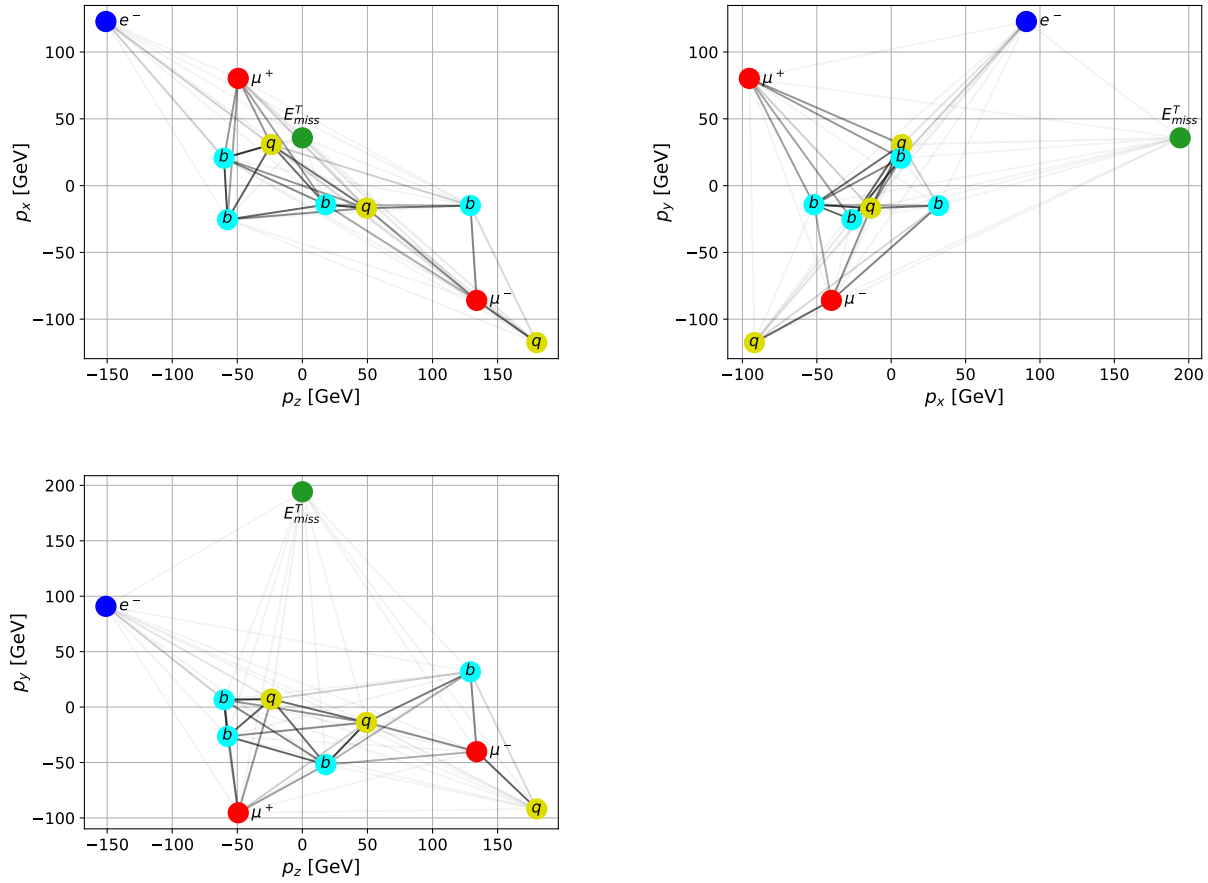


Figure 9.2: Graph constructed from a four-top-quark collision event from data collected in 2016. The graph nodes are plotted in momentum space with the node types indicated by the colors and labels. The edges connecting each pair of nodes are shaded with lightness proportional to the momentum difference between the two nodes.

The latent dimension of the model is 64, the number of layers in each MLP is 4, and the number of message passing steps is 4. These hyperparameters result in a model with 200,321 trainable parameters, most of which correspond to the weight matrices in the MLPs. The typical weight matrix contributes 64×64 trainable parameters, and the typical MLP with four layers contributes 16,384 parameters from weight matrices. The training uses a dropout rate of 5%, a batch size of 256, and a learning rate of 10^{-4}. Each of these is found to be a reasonable default value in the case of the $t\bar{t}t\bar{t}$ signal/background separation problem. Optimization efforts used to arrive at this setup are described later in this section.

The signal in the training consists of the leading-order $t\bar{t}t\bar{t}$ sample described in chapter 5, while the background training dataset is composed of all nominal simulated MC background samples shuffled together. The size and composition of the signal and background training samples are shown in table 9.1. The Monte Carlo event weights are not used in training, but an additional weight of 3.0 is applied in the loss function for $t\bar{t}W$ background events to minimize the impact of $t\bar{t}W$ modeling uncertainties, including the statistical uncertainty associated with the $t\bar{t}W$ jet multiplicity scaling. A k-fold training strategy splits the signal and background datasets into six equal pieces with identical distributions based on the sequentially assigned event numbers. Each of the six trainings excludes one of these pieces, which is held aside as the testing dataset. The remainder of the analysis is performed with each piece using the GNN training for which it is the testing set.

For each of the six folds, the signal and background datasets are preprocessed in a step that randomly shuffles the datasets, including shuffling different background processes together, and constructing batches of 256 signal events and batches of 256 background events. During training, the list of signal and background batches is not reordered. Because the datasets do not have uniform sizes, the signal and background datasets wrap around to the beginning at different points and are generally not synchronized. The training length is measured in epochs of the signal datasets, i.e., the number of times the whole signal dataset has been used in training. The nominal trainings are performed for 78 to 81 epochs, depending on the fold, with the stopping point determined by the wall time, totaling 128 hours per fold across several jobs with checkpointing between jobs. Most of the training is performed on an NVIDIA V100 GPU. Dropout is found to effectively prevent overfitting and improve epoch-to-epoch consistency of the model performance, and this training length is found to be sufficient for the model to converge. These factors eliminate the need for an early stopping condition or a dedicated testing set used to choose the best checkpoint and thus allow the use of a larger training dataset without increasing the number of folds.

These parameters are used for the final GNN training reported in Ref. [129] and are generally found to be safe defaults. The next portion of this section describes the performance and validation of this GNN model, but also reports additional studies of GNN training and optimization for $t\bar{t}t\bar{t}$ measurement in the SSML channel and may use slightly different training setups. Important differences are pointed out where they are relevant.

Process	Number of Events	Fraction of Events
$t\bar{t}W$	37,111	14.96%
$t\bar{t}WW$	1,262	0.51%
$t\bar{t}Z$	94,568	38.13%
$t\bar{t}H$	70,597	28.46%
QmisID	2,048	0.83%
Mat. Conv.	1,068	0.43%
Low m_{γ^*}	931	0.38%
HF e	149	0.06%
HF μ	327	0.13%
LF	301	0.12%
Other $t\bar{t}$+jets	594	0.24%
Single Top	1,359	0.55%
Diboson	3,209	1.29%
$t\bar{t}t$	10,909	4.40%
Other $t\bar{t}VV$	20,261	8.17%
Others	3,345	1.35%
Total Background	248,039	100.00%
LO $t\bar{t}t\bar{t}$	368,578	100.00%

Table 9.1: Size of the signal and background samples used in training before the division into six folds. The training is performed on a combination of all of these samples.

9.2 GNN Top Reconstruction

Reconstructing top-quark decays in $t\bar{t}t\bar{t}$ events is significantly more difficult than in $t\bar{t}$ events due to the larger combinatoric factors, the need to disentangle two neutrinos, and the high probability that at least one final state parton is outside the detector acceptance or overlaps with another. Development of GNN-based reconstruction of hadronic top decays originally devised for the $t\bar{t}t\bar{t}$ measurement is tested in the context of the background rejection GNN. The top reconstruction is defined as a binary edge classification task with edges connecting two jets truth matched to the same top-quark decay given a label of one, and all other edges given a label of 0. The GNN training uses the `root_gnn` package with modifications to allow edge classification. The GNN training achieves an edge classification ROC AUC of 0.88 for $t\bar{t}t\bar{t}$ events.

Top-quark decays can then be reconstructed by finding triplets of jets for which the sum of the three edge scores, known as the top score, meets a given threshold. The leading top candidate is required to have a top score > 1.4, and the sub-leading top to have a top score

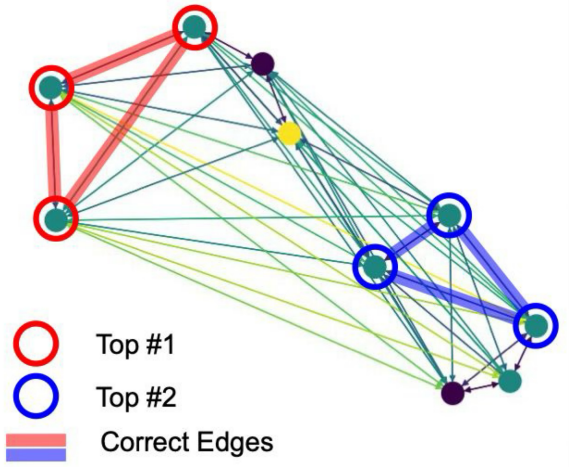


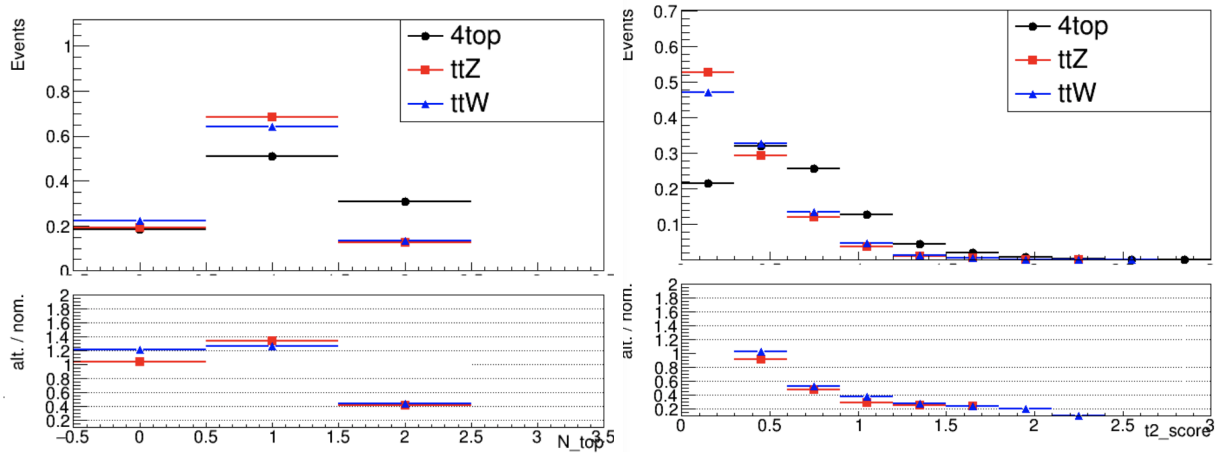
Figure 9.3: Diagram of the edge-classification scheme for top reconstruction. The red and blue highlighted nodes represent jets resulting from the hadronic decays of top quarks. The red and blue highlighted edges are those connecting jets resulting from the same hadronic top-quark decay and therefore receive a label of 1.

Expected rejection of the $\mu_{t\bar{t}t\bar{t}} = 0$ Hypothesis $[\sigma]$	Stat. Only	Stat.+Syst.
No Top Reconstruction	4.64	3.37
With Top Reconstruction	4.65	3.36

Table 9.2: Expected significance of the $t\bar{t}t\bar{t}$ signal with respect to the null hypothesis with signal/background separation GNNs trained with and without inputs from the top reconstruction GNN.

> 0.55 to be considered top-quark candidates. The technical efficiency, that is, the efficiency for correctly identifying the triplet associated with both hadronic top decays in $t\bar{t}t\bar{t}$ events where this is possible, is 65%.

The number of reconstructed top quarks and the score of the second top quark are shown in fig. 9.4. The $t\bar{t}Z$ and $t\bar{t}W$ backgrounds are expected to have only one hadronic top decay in the dilepton channel, and thus have fewer reconstructed top quarks. The score of the second top candidate also shows significant separation power between the $t\bar{t}t\bar{t}$ signal and the $t\bar{t}W$ and $t\bar{t}Z$ backgrounds. This can be leveraged in the signal/background classification GNN by either incorporating the scores of the top candidates as global features or by directly providing the edge scores produced by the top reconstruction GNN as edge features for the graph input to the signal/background separation GNN. The latter approach has been tried, resulting in a negligible increase in the separation power as shown in table 9.2.



(a) (b)

Figure 9.4: The distribution of the number of reconstructed top candidates (fig. 9.4a) and the top score for the sub-leading (fig. 9.4b) top candidate.

This may be because the signal/background GNN already fully exploits all useful information about the topology of hadronic top-quark decays in rejecting the background. Because the performance improvement is negligible and performing top reconstruction adds significant complexity and an additional round of GNN inference, top reconstruction is not used in the final analysis.

9.3 Optimization of the Signal/Background Separation GNN

The MLPs in the nominal GNN training use the ReLU activation function, but empirical searches [264] for better activation functions have indicated that the swish (also known as sigmoid linear unit or SiLU)

$$f(x) = \frac{x}{1 + \exp(-x)}$$

or Leaky ReLU

$$f(x) = \max(0, x) + \min(\epsilon x, 0), \ 0 < \epsilon < 1$$

activation functions may provide better performance. Table 9.3 shows that the AUC did not significantly improve when using these activation functions.

The nominal GNN training does not use the event weights from the Monte Carlo Simulation or those that correct the trigger and object reconstruction efficiencies and the pileup profile. Instead, all signal events receive a weight of 1, as do all background events except

	ReLU	Swish	Leaky ReLU
AUC on Testing Dataset	0.883	0.883	0.884

Table 9.3: ROC AUC achieved by the $t\bar{t}t\bar{t}$ background rejection GNN with different activation functions. These trainings were performed during the optimization process and do not include all improvements to the training procedure.

for $t\bar{t}W$ events, which receive a weight of 3. This choice of weighting leads to a difference between the background composition of the training dataset reported in table 9.1 and the expected background composition of the signal region reported in table 8.1. The use of weights may also cause the kinematic distributions of the backgrounds to vary from those expected from the full background estimation. Using the fully weighted signal and background prediction in training would thus result in a more accurate estimate of the probability densities associated with signal and background in the real experiment. However, the use of event weights also reduces the statistical power of the sample as events with large weights dominate the gradients used to update the model, since events with small weights provide negligible contributions to the weighted gradient. Events with negative weights, which result from the use of NLO samples for some backgrounds, can have especially harmful effects. Meanwhile, many background processes have similar topologies, and changes to the composition may be tolerable in terms of the final separation power. To find the optimal balance between the effective statistics of the training dataset and the accuracy of the underlying probability density function, several weighting schemes are tried. The results of these studies are shown in fig. 9.5.

This plot shows the separation power, measured by the ROC AUC using the testing datasets and the fully weighted nominal signal and background predictions, of various setups against the effective statistics N_{eff} of the training dataset. N_{eff} is defined as

$$N_{eff} = \frac{\left(\sum_i w_i\right)^2}{\sum_i w_i^2},$$

where $i = 1..N$ runs over all events in the dataset and $\{w_i\}$ are the event weights. This formula has the desired properties that if all w_i are equal and non-zero, then $N_{eff} = N$ and otherwise $0 < N_{eff} < N$. The group of points in the top right corner represents tests in which events are applied only at the sample level rather than the event level. In the legend, "None" means that no events receive weights in the training, "$w > 0$" means that events with negative weights are excluded from the training, and "Majors Only" indicates that the background consists only of $t\bar{t}W$, $t\bar{t}Z$, and $t\bar{t}H$. "X xY" indicates that background process X is given an event weight of Y. The weights used for $t\bar{t}H$ and $t\bar{t}Z$ are chosen to match the effective statistics of the nominal $ttW \times 3$ setup. Finally, "x-sec only" indicates that each background process receives a weight that results in the background composition matching the expected yields in table 8.1 while keeping the average weight of background events equal

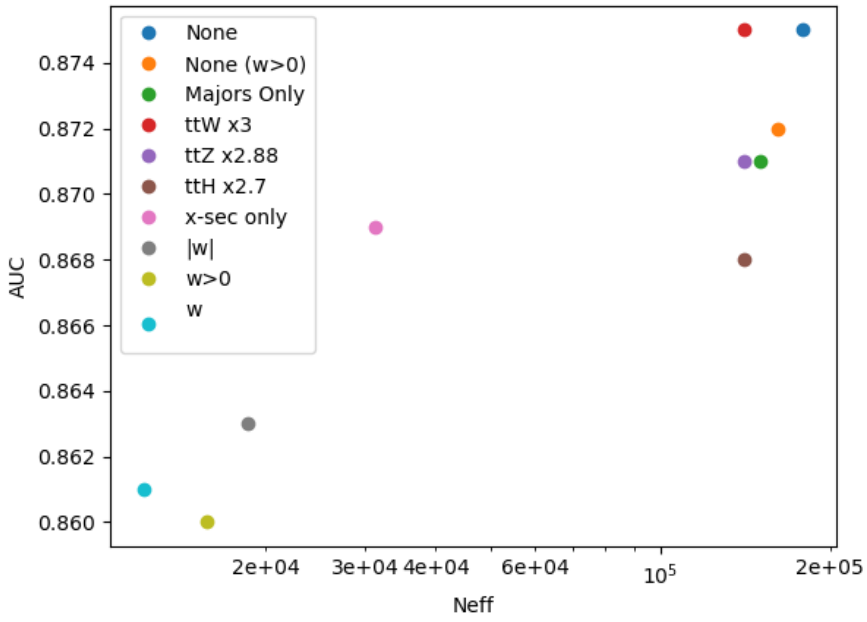


Figure 9.5: ROC AUC versus effective training statistics of various methods of weighting events during GNN training. The cluster at the top right does not use event weights and applies flat weights to at most one background. The cluster at the bottom left uses MC event weights.

to 1. This setup has an intermediate level of statistical power because some relatively large background samples receive large or small weights.

The worse-performing trainings in the lower left corner all use the full event weights used in the background prediction with different treatment of events with negative weights, and suffer from significantly lower effective statistics. $|w|$ indicates that the magnitude of the weight is used, while the other trainings use all weights as is, either including or excluding events with negative weights. Setups with individual event weights or with many different processes receiving weights show worse separation power than those with minimal reweighting, and the separation power is strongly correlated with the effective statistics of the training dataset. The inclusion of an extra weight for $t\bar{t}W$ notably does not lose performance because $t\bar{t}W$ is significantly underrepresented in table 9.1 and is the dominant background in the signal region. The setup with the $t\bar{t}W$ weight is chosen as nominal because improved separation of the $t\bar{t}W$ background reduces sensitivity of the $t\bar{t}t\bar{t}$ measurement to the large $t\bar{t}W$ modeling uncertainties, including the degrees of freedom introduced by the data-driven $t\bar{t}W$ background estimation. These effects are not accounted for when using the AUC as a figure of merit and require a full evaluation of the expected sensitivity using a fit to the

Number of Folds	$t\bar{t}W$ Weight	ROC AUC	Expected Significance [σ]
2	1.0	0.896	3.58
2	3.0	0.893	3.63
4	1.0	0.899	—
4	3.0	0.902	—
6	3.0	0.903	3.66

Table 9.4: ROC AUC and sensitivity comparison of training with different choices of $t\bar{t}W$ weight and different numbers of folds. The ROC AUC is evaluated using the full NLO $t\bar{t}t\bar{t}$ sample and the full MC background estimate in the signal region. The expected significance is evaluated using the profile likelihood ratio with the Realistic Asimov data and an injected $t\bar{t}t\bar{t}$ signal strength of one as described in chapter 6.

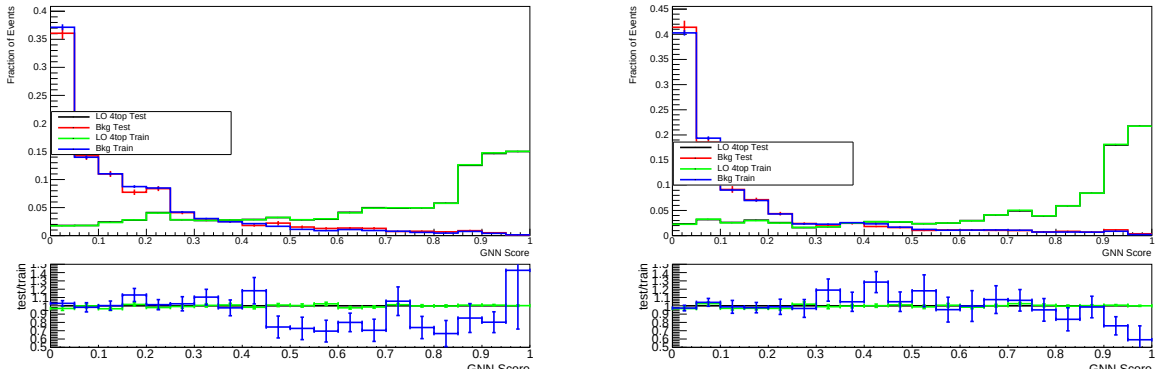
likelihood function, including the relevant systematic uncertainties.

A scan of possible values of the $t\bar{t}W$ training weight and the number of folds is used to determine the importance and optimal values of these parameters. The results of the scan are shown in table 9.4. The best background rejection, measured by the ROC AUC, and the best expected significance are both found with the nominal setup. A larger scan including more values of the $t\bar{t}W$ weight and the possibility of weighting only $t\bar{t}W$ events with at least seven jets has been performed at an earlier stage of the analysis optimization and using a BDT, which can be trained more quickly. The results of the additional GNN tests are shown in table 9.5. All three studies indicate that the nominal choice of six folds with a $t\bar{t}W$ training weight of three is the best choice to maximize the expected sensitivity to the $t\bar{t}t\bar{t}$ signal after considering systematic effects. In some cases, the ROC AUC decreases while the expected significance increases. This happens when the rejection of the $t\bar{t}W$ background improves and the impact of the $t\bar{t}W$ modeling uncertainties is reduced.

The final performance of the nominal GNN is summarized in table 9.6, which shows the ROC AUC scores for the six folds of the GNN, figs. 9.6 and 9.7, which show the GNN output distributions for the MC simulated signal and background samples, and fig. 9.8, which shows the ROC curves using the nominal signal and background predictions in the signal region. Variations of the score distributions and AUCs between folds are associated with differences in the b-jet multiplicity of the training datasets, consistent with a statistical fluctuation. Similar variations are observed in the BDT training and GNN trainings with different numbers of folds. All six folds show consistent results between the testing and training samples, indicating minimal over-fitting of the training dataset.

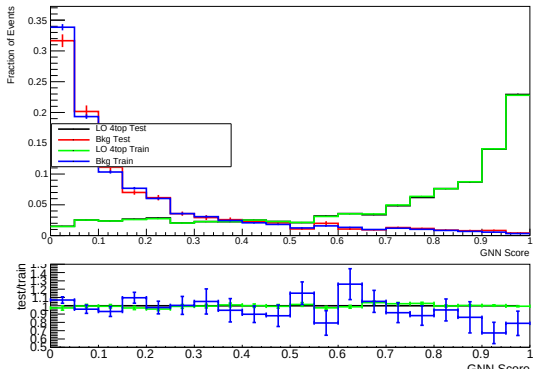
The distribution of the GNN score for the observed collision data is shown in fig. 9.9 with a comparison to the expected yields determined from the post-fit signal plus background expectation. A normalized comparison of the shape of the $t\bar{t}t\bar{t}$ signal, inclusive background, and $t\bar{t}t$ background is shown in fig. 9.10.

A BDT trained to separate signal and background following the strategy of Ref. [60] but

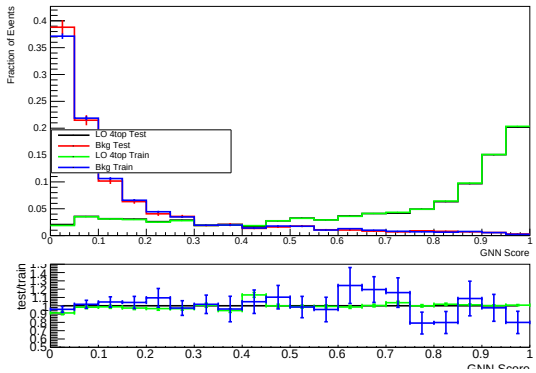


(a)

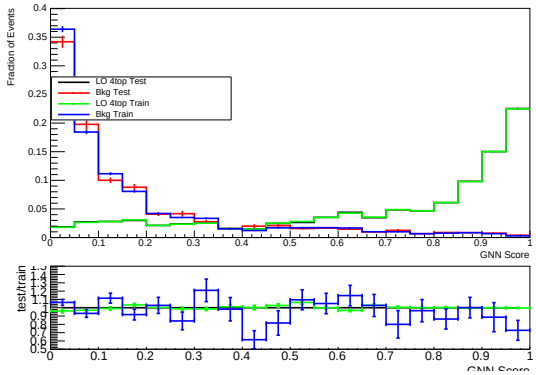
(b)



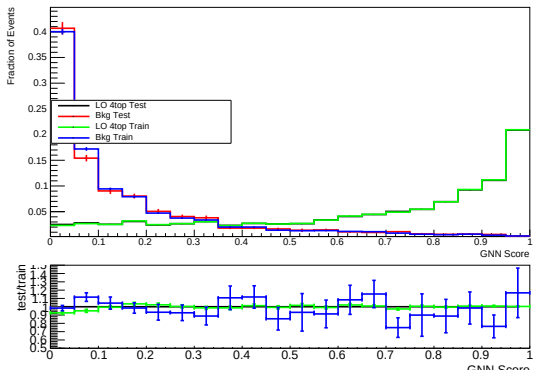
(c)



(d)

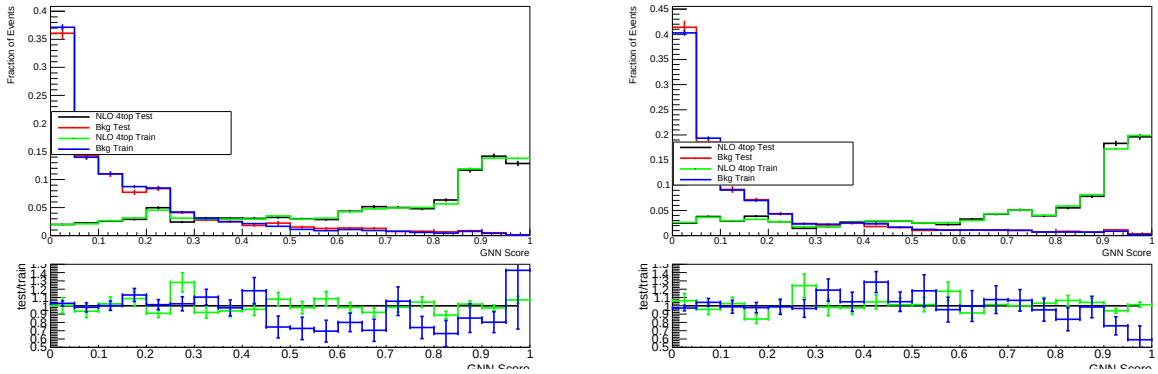


(e)



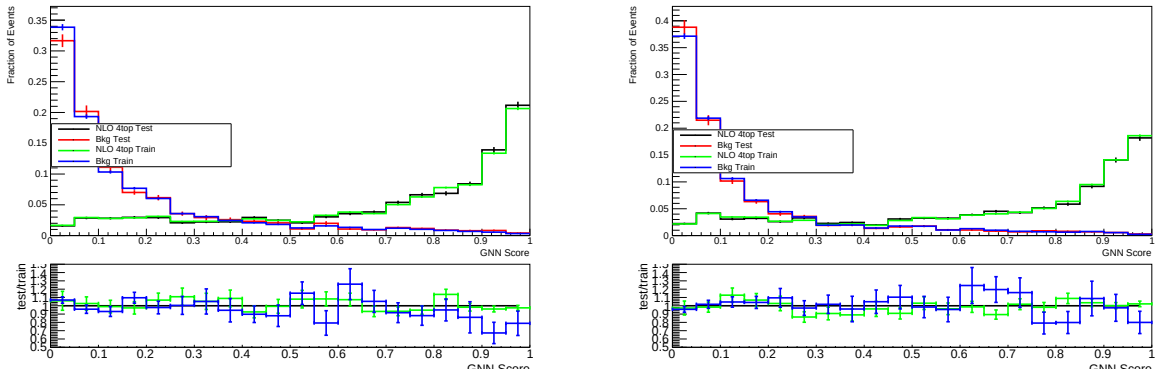
(f)

Figure 9.6: GNN Score distributions on the testing and training datasets for each fold of the six-fold training. The signal sample is the LO $t\bar{t}t\bar{t}$ sample, which is used in training.



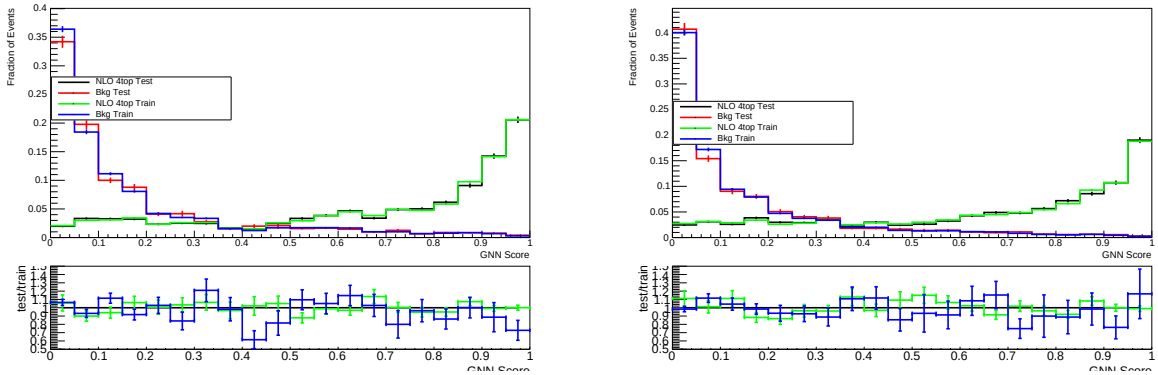
(a)

(b)



(c)

(d)



(e)

(f)

Figure 9.7: GNN Score distributions on the testing and training datasets for each fold of the six-fold training. The signal sample is the NLO $t\bar{t}t\bar{t}$ sample, which is not used in training.

$t\bar{t}W$ Weight	$t\bar{t}W \geq 7$ jets Weight	ROC AUC	Expected Significance [σ]
1.0	1.0	0.880	2.88
3.0	1.0	0.878	3.09
9.0	1.0	0.869	3.08
1.0	1.5	0.876	2.93
1.0	3.0	0.876	2.99

Table 9.5: ROC AUC and sensitivity comparison of training with different choices of $t\bar{t}W$ weight performed at an early stage of analysis optimization. All trainings have two folds. The ROC AUC is evaluated using the full NLO $t\bar{t}t\bar{t}$ sample and the full MC background estimate in the signal region. The expected significance is evaluated using the profile likelihood ratio with the Realistic Asimov data and an injected $t\bar{t}t\bar{t}$ signal strength of one as described in chapter 6.

GNN Fold	Test AUC	Train AUC
0	0.909	0.912
1	0.913	0.917
2	0.910	0.914
3	0.914	0.919
4	0.909	0.915
5	0.909	0.916
Combined	0.910	N/A

Table 9.6: ROC AUC of the GNN on testing and training datasets. The LO signal sample is used for the AUC calculation.

reoptimized and retrained using the improved object reconstruction, selection, and tagging. The BDT serves as a cross-check on the final results and a comparison to evaluate the improvement offered by the GNN. Additionally, because the BDT can be trained much faster than the GNN, it serves as a useful tool during optimization and has been used in the determination of the optimal weighting scheme during training and in studying the effects of varying the number of folds in the k-fold training. The BDT hyperparameters are optimized with a grid search, and the sum of jet PCBT scores is replaced with N_{j} and the sum of the leading four PCBT scores. The ultimate separation power of the re-optimized BDT is an ROC AUC of 0.873 compared with an AUC of 0.903 for the GNN, both evaluated using the fully weighted nominal signal and background prediction in the signal region and using the entire simulated dataset following the k-fold method.

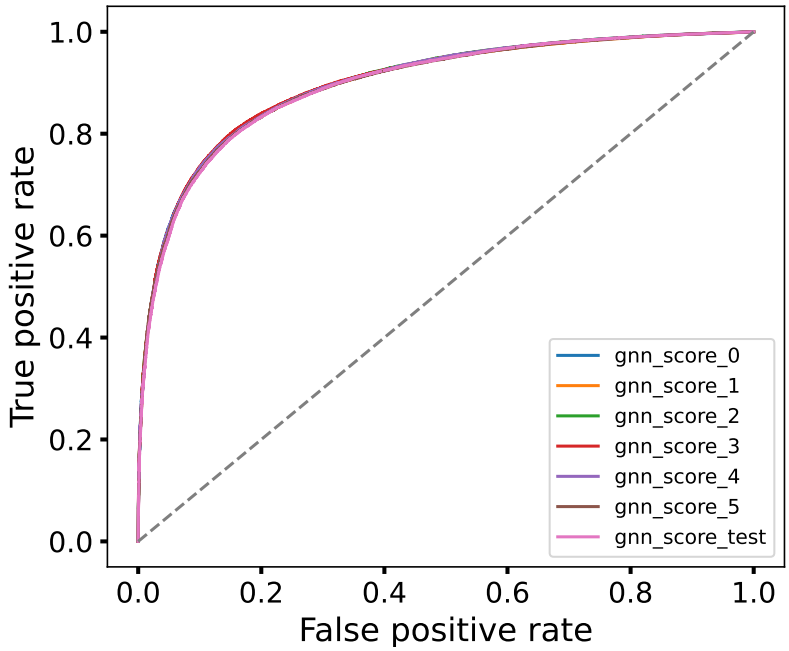


Figure 9.8: ROC Curve for each fold of the GNN and for the combined GNN evaluated on the testing set using the nominal signal and background samples.

9.4 Validation of the Signal/Background Separation GNN

Because the output of the GNN discriminant for events from MC simulations is used to construct the expected data yield, the robustness of the GNN output to mis-modeling in the simulation must be validated before unblinding the signal region and fitting the observed GNN distribution. Because the GNN is trained using all reconstructed objects, it may be more sensitive to mis-modeling in parton showering, pileup modeling, or other effects than traditional methods. To verify that the GNN score distribution in data matches that in simulation, the agreement between the data and the MC prediction is tested in several validation regions. The ideal characteristics of a validation region are low signal contamination, good coverage of the whole GNN score region, and event topologies similar to the signal region. The first requirement is necessary so that the validation regions can be studied independently of any measurement of the $t\bar{t}t\bar{t}$ process, while the others prefer a region with $t\bar{t}t\bar{t}$-like events to accurately assess the expected agreement between data and MC in the signal region. One region useful for this validation is the $t\bar{t}Z$ validation region shown in fig. 8.4. This validation region shows good agreement over the whole range of GNN scores; however, it includes only

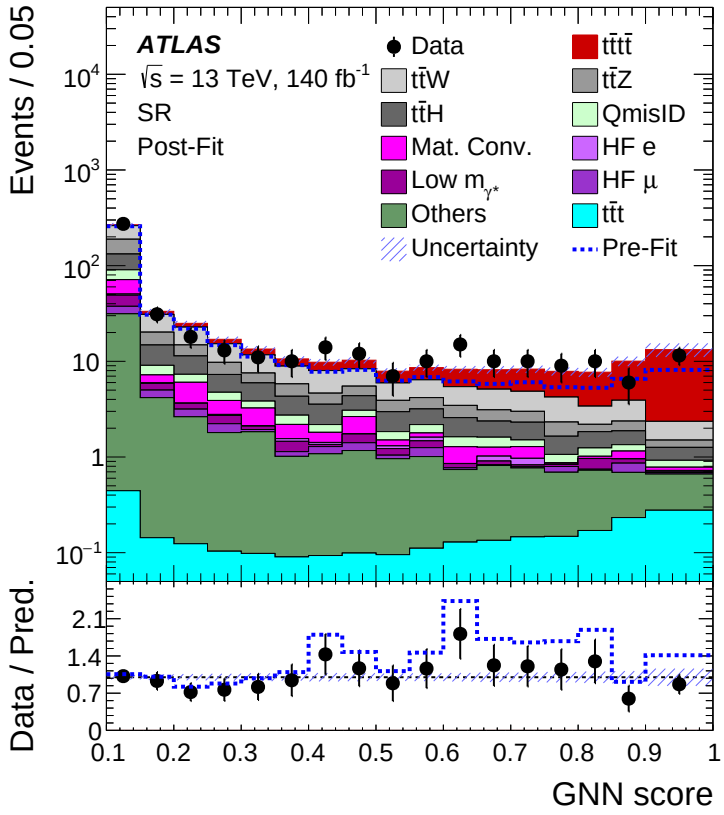


Figure 9.9: Comparison between data and the predictions after a fit to data for the GNN distribution in the SR. The first bin contains underflow events. The ratio of the data to the total post-fit prediction is shown in the lower panel. The dashed blue lines show the pre-fit prediction in the upper panel and the ratio of the data to the total pre-fit prediction in the lower panel. The shaded band represents the total post-fit uncertainty in the prediction. Reproduced from [129].

four data events with GNN score > 0.5 and provides limited information about the modeling in that region.

An alternative approach using events from CR $t\bar{t}W^{\pm}$+jets, the $t\bar{t}Z$ validation region, and CR Mat. Conv. creates events with higher multiplicity that are more likely to receive higher GNN scores. This is done with a method known as "jet reflection." When constructing the graph for an event, a copy of each jet is included with the sign of η and ϕ reversed, effectively including the jet's reflection about the detector origin. This doubles the event's jet and b-jet multiplicities, making control region events much more signal-like without introducing any $t\bar{t}t\bar{t}$ signal. The results of this test are shown in fig. 9.11. The results are also produced with the optimized BDT to compare the modeling quality to the strategy of Ref. [60].

The jet reflection method generally results in good agreement with the data. Interest-

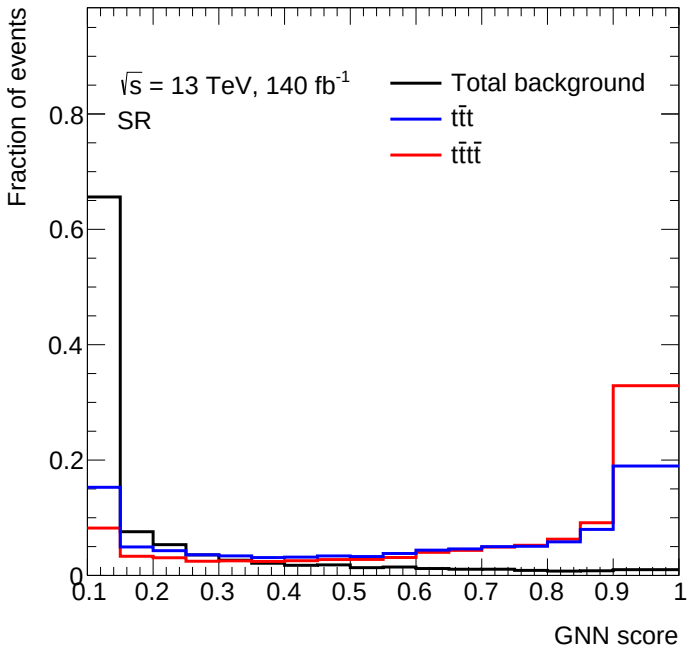


Figure 9.10: Comparison of the shape of the GNN distribution in the SR between the total background (black), the $t\bar{t}t$ background (blue), and the $t\bar{t}t\bar{t}$ signal (red).

ingly, the GNN distributions after jet reflection tend to be less signal-like than the BDT distributions, indicating that the GNN exploits information about top-quark decays in $t\bar{t}t\bar{t}$ events and is less sensitive to variations in the multiplicity of background processes. One drawback to the jet reflection method is that deviations from the modeling of additional QCD radiation in the underlying event, matrix element, or parton showering may not be detected, as the CR events used for jet reflection have fewer jets and hence do not contain this additional radiation. To supplement the jet reflection method, an additional validation region is constructed using events with two opposite-sign leptons (2LOS). The definition uses the $e^{\pm}\mu^{\mp}$ channel with $N_{\mathrm{j}} > 4$ and at least two b-jets and is dominated by $t\bar{t}$ events. The different flavor channel is chosen to reduce Z + jets production, and therefore study events with top quark pairs, which are more comparable to the expected backgrounds in the $t\bar{t}t\bar{t}$ signal region. The comparison is made to two different $t\bar{t}$ simulations, the nominal PowhegBox sample described in chapter 5, and a MadGraph5_aMC@NLO sample, which is used to provide a comparison between the data and a simulation using the same setup as the nominal $t\bar{t}t\bar{t}$ simulation. This validation is performed with an earlier version of the GNN, which only used two folds for training rather than six. This version is referred to as GNN_91 in the plots. The results of this comparison are shown in figs. 9.12 to 9.15. The region is well populated with over 200 events in the most signal-like GNN bin, despite

the dominant $t\bar{t}$ production having a strongly background-like GNN shape. All comparisons show good agreement of the data with the simulation within the uncertainty on the MC prediction.

9.5 Studies of a Dedicated GNN for the Top Yukawa Interpretation

Possible GNN applications to increase the sensitivity of BSM interpretations of the $t\bar{t}t\bar{t}$ signal, particularly the top Yukawa measurement, were also explored. The first of these possibilities is a binary GNN classifier to separate events generated with top Yukawa coupling strengths and CP mixing angles. In principle, such a training could exploit the different momentum and angular distributions of the top quarks in the two scenarios to increase the sensitivity to BSM physics in the top-Higgs coupling. Kinematic differences between the $t\bar{t}t\bar{t}$ signal in the different top Yukawa samples are shown fig. 9.16 for $t\bar{t}t\bar{t}$ and fig. 9.17 for $t\bar{t}H$.

Minimal variation is observed in the background rejection GNN or the object-level observables studied here, though there may be larger discrepancies in variables sensitive to the angular separation of the top quarks. The largest variation is seen for the extreme $\kappa_t = 0$, $\tilde{\kappa}_t = 2$ scenario. The variation between SM $t\bar{t}t\bar{t}$ samples generated at NLO and LO is frequently larger than that between samples with BSM modifications to the top-Higgs coupling. A measurement overly reliant on these variations may therefore not be reliable without consideration of NLO effects. These observations are also true of the variations in the $t\bar{t}H$ background with α.

Two sets of GNN trainings, one to distinguish the $\kappa_t = 2$ scenario from SM $t\bar{t}t\bar{t}$ production and one to distinguish the case of a fully $CP-$odd top Yukawa coupling, are tried. The maximum ROC AUC achieved between any two $t\bar{t}t\bar{t}$ samples generated with different values of the top Yukawa coupling is 0.535, indicating negligible separation. This AUC is also less than the separation power between samples generated at leading and next-to-leading-order in the QCD coupling constant, indicating that the separation power of samples with different top Yukawa couplings may not be reliable because it does not include next-to-leading-order effects. This approach is thus not pursued.

The second possible application of the GNN to the top Yukawa measurement is the implementation of a multi-class signal/background discrimination training with separate background categories for $t\bar{t}H$ and other backgrounds, including $t\bar{t}W$. This would allow the $t\bar{t}H$ normalization to be constrained independently of the floating $t\bar{t}W$ normalization and increase the top Yukawa constraint provided by the $t\bar{t}H$ process. A multi-class training is performed following this idea using the `root_gnn` model with the cross entropy loss and three outputs passed through a softmax normalization to ensure they sum to one and can be treated as probabilities. The results of the multi-class training are shown in fig. 9.18. The signal region is then divided into three categories as listed in table 9.8. The expected yields in the resulting background-like signal regions are shown in fig. 9.19. The multi-

class training maintains the same ROC AUC for overall signal/background separation as the nominal training when using the $t\bar{t}t\bar{t}$ score as the discriminant and achieves better separation of $t\bar{t}H$ and $t\bar{t}W$ than the nominal region definitions.

Class Name	Samples Used	Number of Events in Training
$t\bar{t}t\bar{t}$	LO $t\bar{t}t\bar{t}$ for training,	368,578
	NLO $t\bar{t}t\bar{t}$ for validation	
$t\bar{t}H$	Standard Model $t\bar{t}H$	70,597
Background	All backgrounds excluding $t\bar{t}H$	178,307

Table 9.7: Class definitions for the multi-class training.

Region Name	$t\bar{t}t\bar{t}$ Score Cut	$t\bar{t}H$ and Bkg. Score Cut
$t\bar{t}t\bar{t}$ SR	$t\bar{t}t\bar{t}$ score > 0.5	None
$t\bar{t}H$ SR	$t\bar{t}t\bar{t}$ score < 0.5	$t\bar{t}H$ score $\geq$ Bkg. score
$t\bar{t}W$ SR	$t\bar{t}t\bar{t}$ score < 0.5	$t\bar{t}H$ score $<$ Bkg. score

Table 9.8: SR definitions based on multi-class training.

The resulting sensitivity to the top Yukawa coupling is tested using the methods described in chapter 6 and is significantly improved when $t\bar{t}H$ is parameterized, with the expected 95% limit improving from $|\kappa_t| < 1.63$ to $|\kappa_t| < 1.38$ based only on the cross-section variations of $t\bar{t}H$ and $t\bar{t}t\bar{t}$ production with κ_t. Similar improvements are seen in the 2D $(\kappa_t, \tilde{\kappa}_t)$ exclusion and, as expected, there is no significant change when profiling $t\bar{t}H$ since the degeneracy between $t\bar{t}H$ and $t\bar{t}W$ does not affect these results.

This approach is ultimately not pursued, as the multi-class training shows lower expected sensitivity to the SM $t\bar{t}t\bar{t}$ signal by about 2% when using the nominal region definition and by about 3.5% when using the split signal region of table 9.8. Using different MVA discriminants for the $t\bar{t}t\bar{t}$ cross-section measurement and the top Yukawa interpretations would significantly increase the overall complexity of the validation procedure and statistical analysis, and would not improve the more model-independent results with $t\bar{t}H$ profiled and is therefore not pursued. Additional MVA discriminants to constrain the EFT couplings are likewise not considered due to the large complexity induced by the number of EFT operators and combinations thereof.

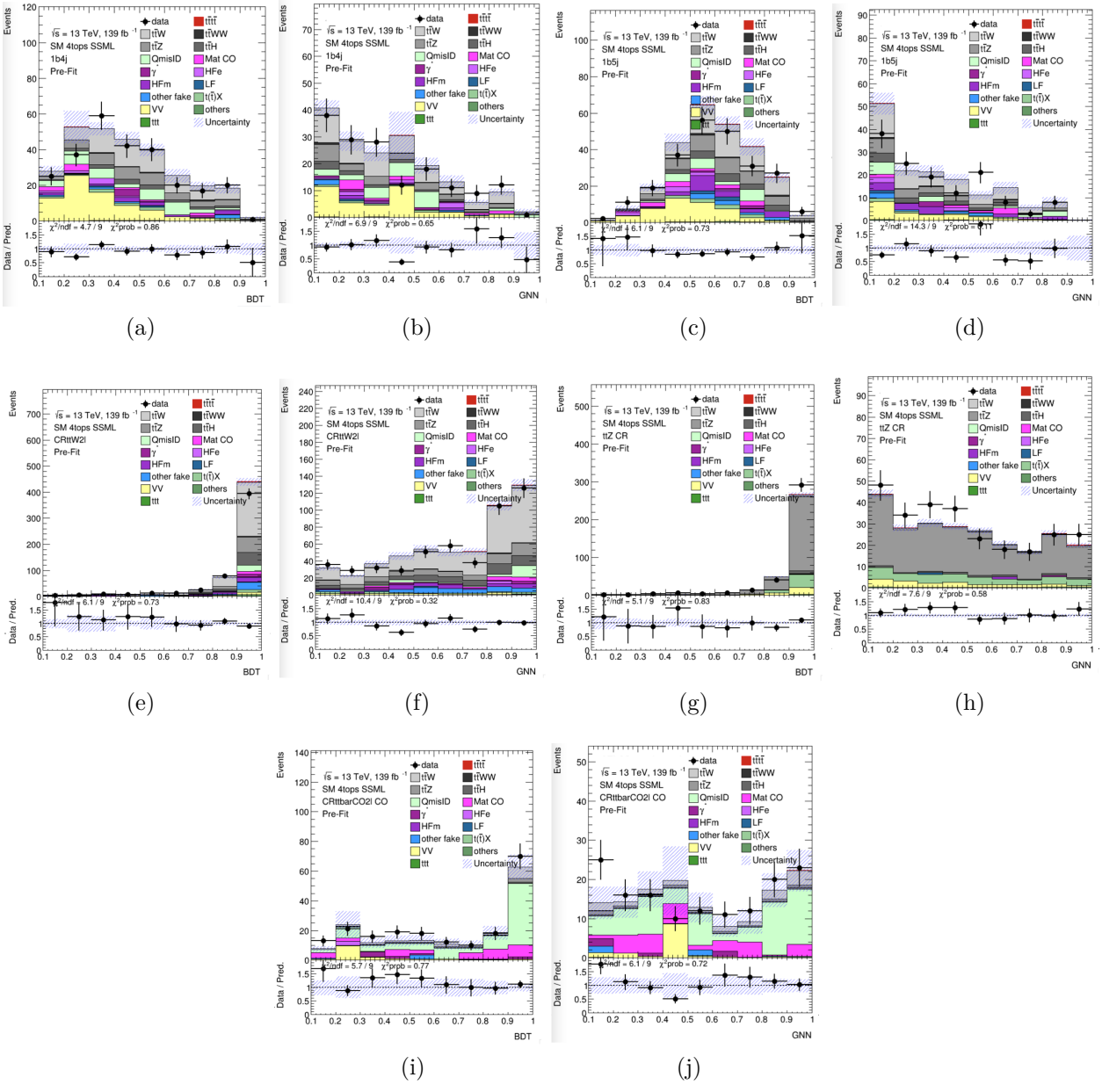
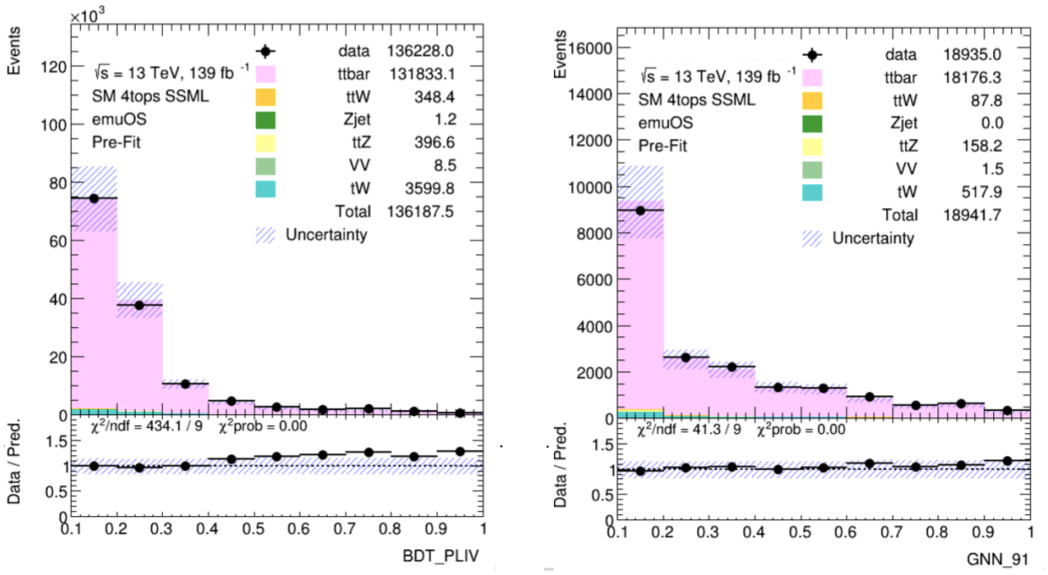
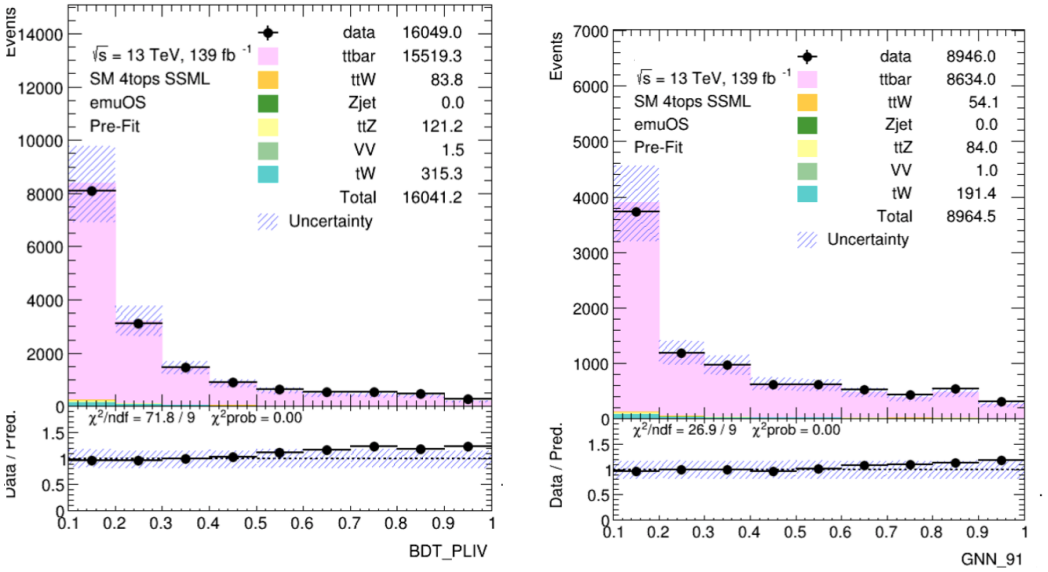


Figure 9.11: Data-MC comparison on the GNN/BDT scores with artificially injected jets from the jet reflection method in the analysis control regions and validation regions.



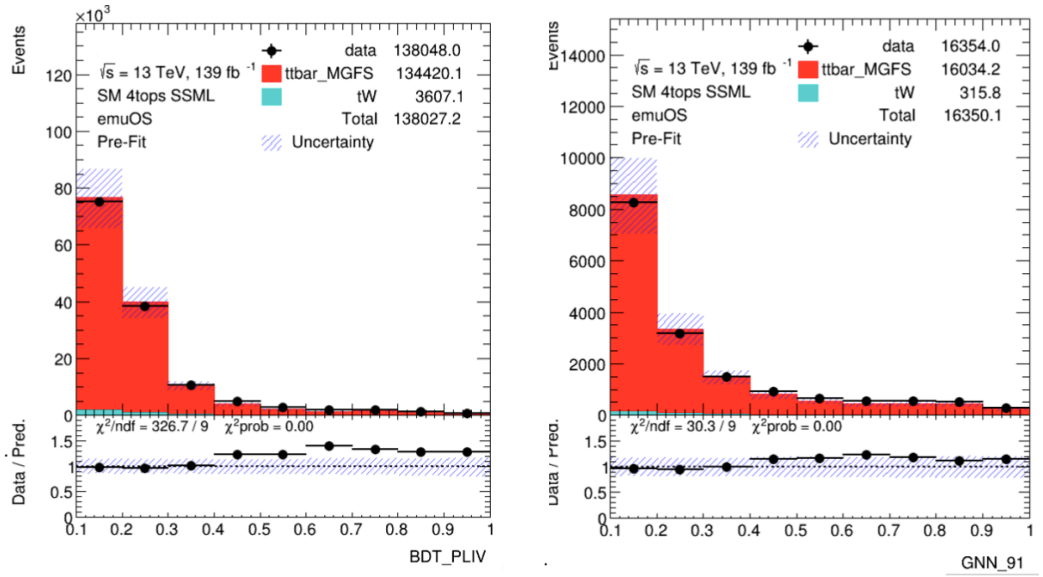
(a) (b)

Figure 9.12: Data-MC comparison for the GNN/BDT scores in the 2LOS region, with PowhegBox $t\bar{t}$+jets MC and $N_{\mathrm{j}} \geq 4$.



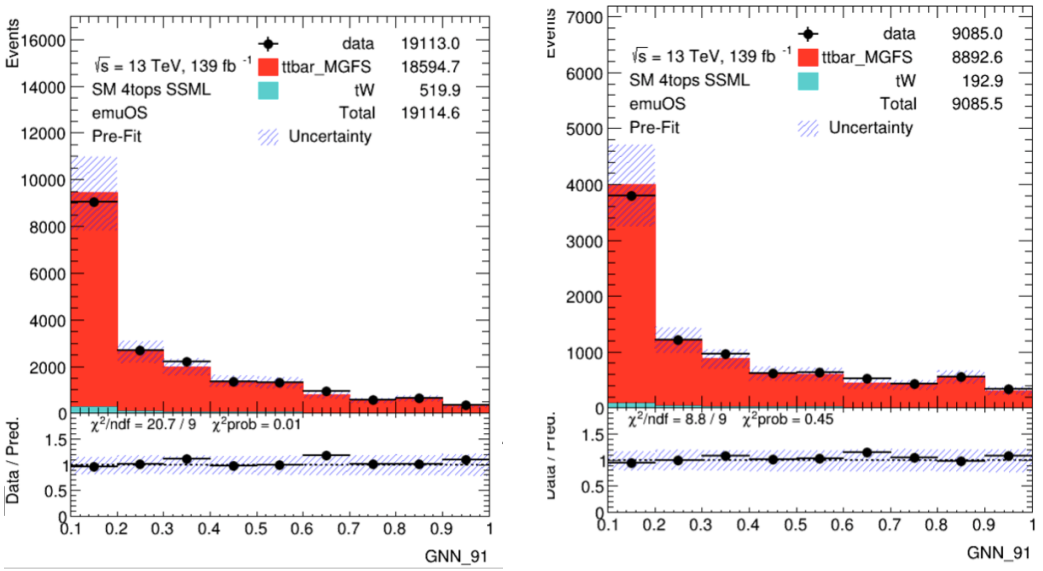
(a) (b)

Figure 9.13: Data-MC comparison on the GNN/BDT scores in the 2LOS region, with PowhegBox $t\bar{t}$+jets MC and $N_{\mathrm{j}} \geq 6$.



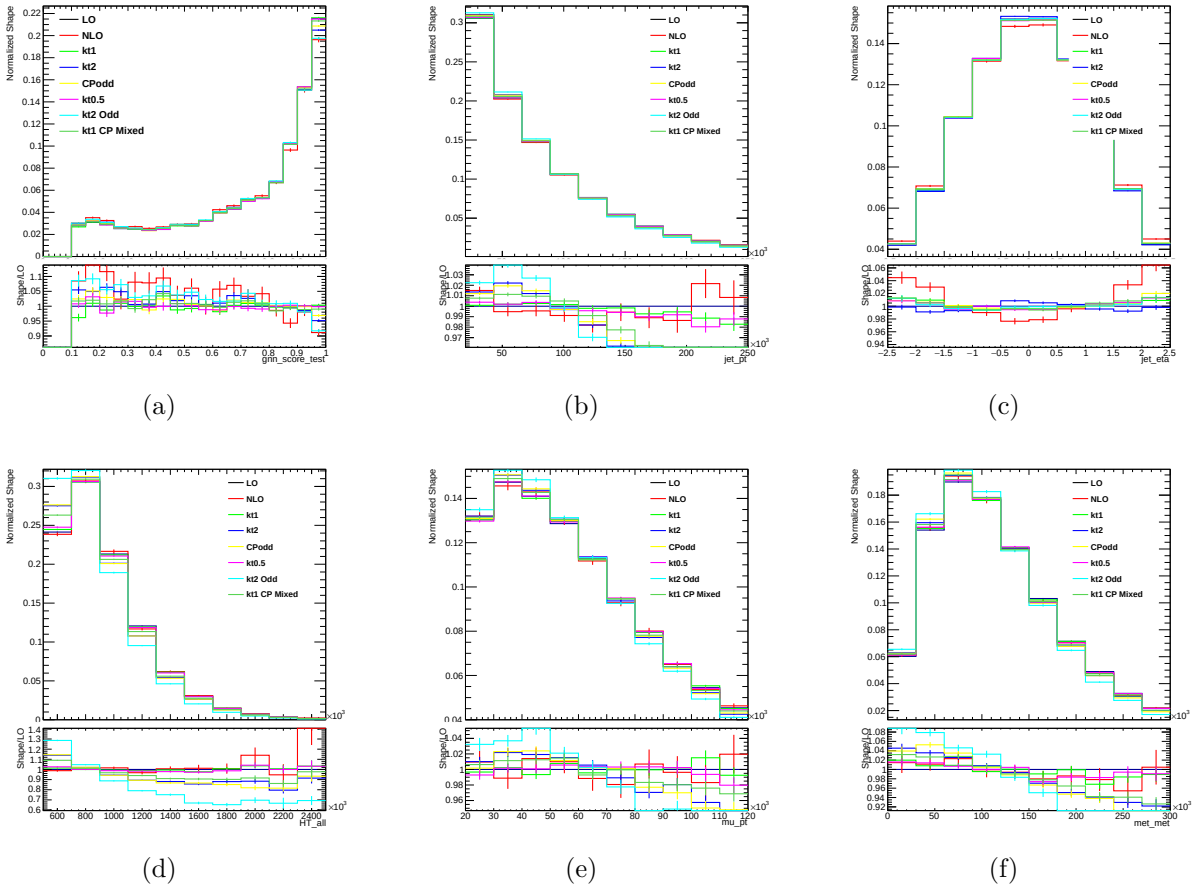
(a) (b)

Figure 9.14: Data-MC comparison on the GNN/BDT scores in the 2LOS region, with MadGraph5_aMC@NLO $t\bar{t}$+jets MC and $N_{\mathrm{j}} \geq 4$.



(a) (b)

Figure 9.15: Data-MC comparison on the GNN/BDT scores in the 2LOS region, with MadGraph5_aMC@NLO $t\bar{t}$+jets MC and $N_{\mathrm{j}} \geq 6$.



(a) (b) (c)

(d) (e) (f)

Figure 9.16: Kinematic shape variations of $t\bar{t}t\bar{t}$ among samples with BSM top Yukawa couplings and the nominal NLO signal sample. (a) GNN score distribution. (b) p_T of all jets. (c) η of all jets. (d) H_T. (e) p_T of all muons. (f) $E_{\mathrm{T}}^{\mathrm{miss}}$. In the legend, 'LO' refers to the leading-order MadGraph5_aMC@NLO sample, 'NLO' is the nominal $t\bar{t}t\bar{t}$ sample, 'kt1' is the SM sample generated with the Higgs characterization model, and the remainder are $t\bar{t}t\bar{t}$ samples generated with alternate values of the top Yukawa coupling. The number after 'kt' is the overall coupling strength $k_t = \sqrt{\kappa_t^2 + \tilde{\kappa}_t^2}$ where $k_t = 1$ is implied if not otherwise specified, 'Odd' or 'CPodd' indicates a sample generated with $\alpha = 90°$, and 'CP Mixed' indicates $\alpha = 45°$.

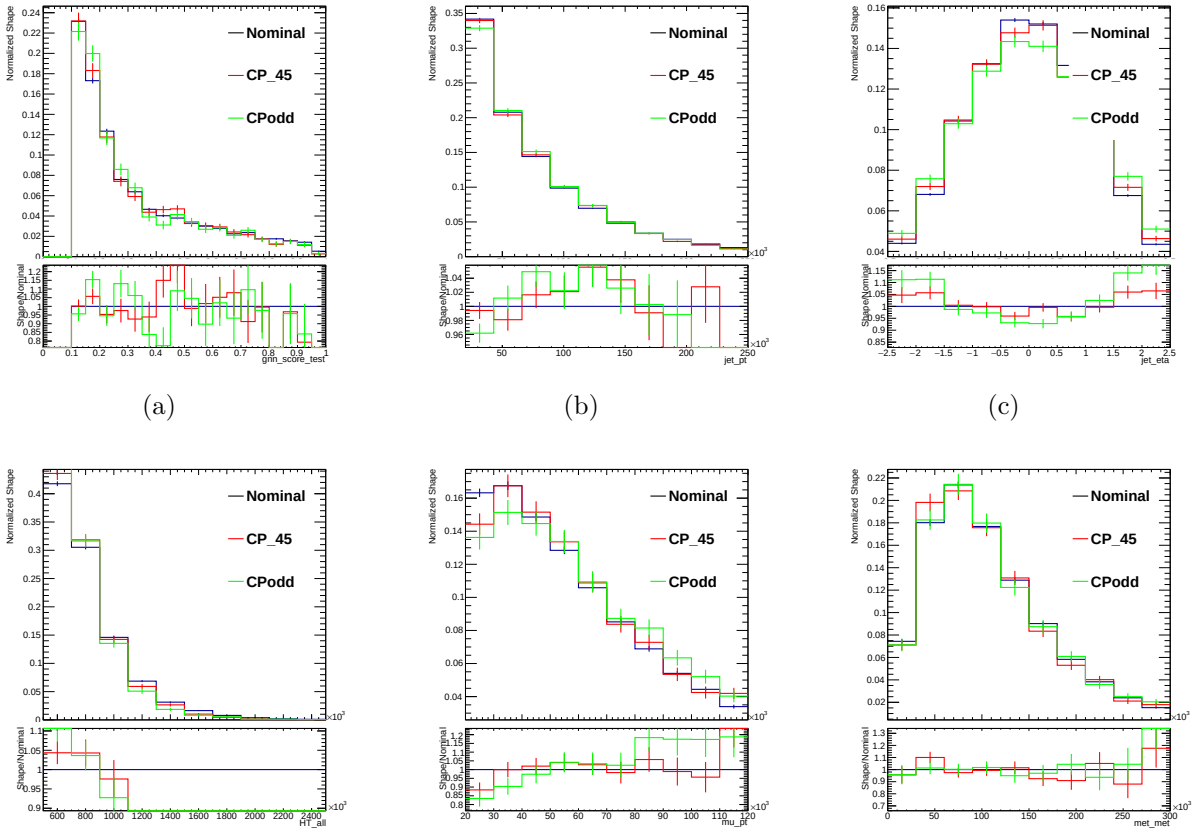
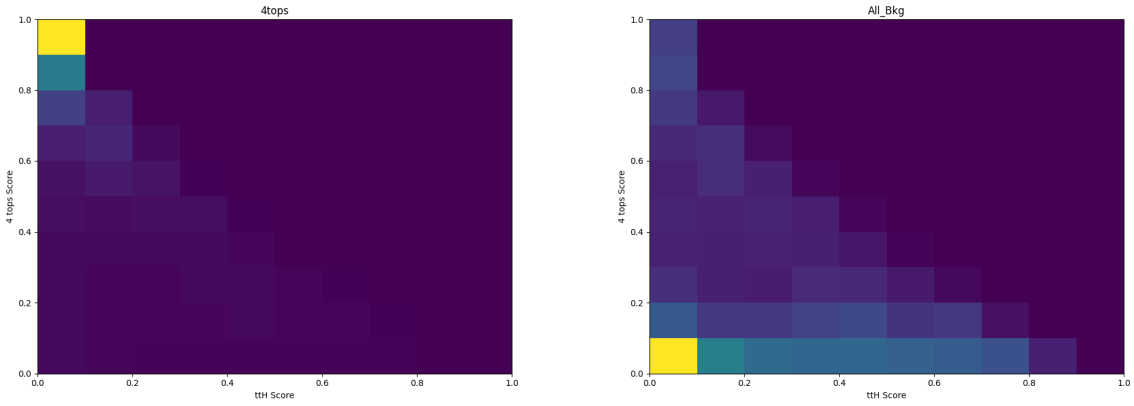
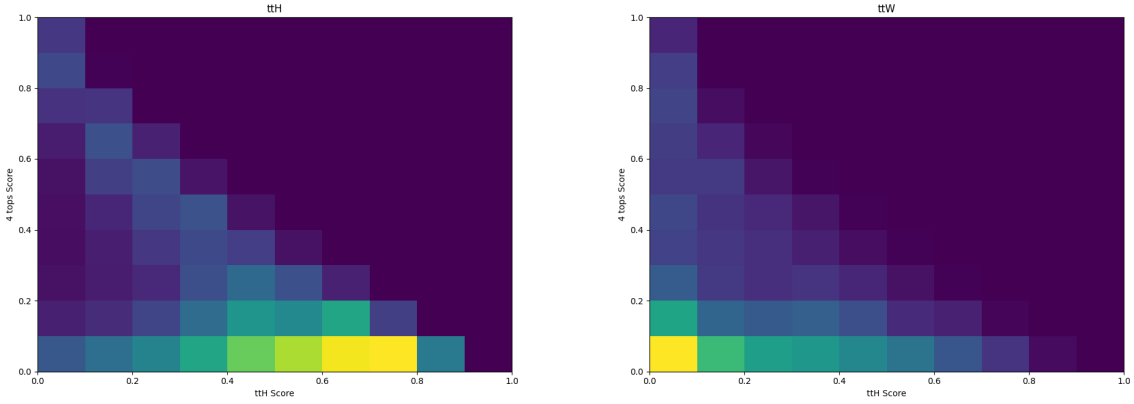


Figure 9.17: Kinematic shape variations of $t\bar{t}H$ between samples $\alpha = 0°$ or $90°$. (a) GNN score distribution. (b) p_T of all jets. (c) η of all jets. (d) H_T. (e) p_T of all muons. (f) $E_{\mathrm{T}}^{\mathrm{miss}}$. In the legend, 'Nominal' refers to the nominal sample used to model the $t\bar{t}H$ background while 'CP 45' and 'CPodd' refer to samples generated with $\alpha = 45°$ and $\alpha = 90°$, respectively.



(a) (b)



(c) (d)

Figure 9.18: Score distributions for the multi-class training in the ($t\bar{t}t\bar{t}$ score, $t\bar{t}H$ score) plane. The background score is equal to 1 − $t\bar{t}t\bar{t}$ score − $t\bar{t}H$ score. (a) $t\bar{t}t\bar{t}$. (b) All backgrounds, including $t\bar{t}H$. (c) $t\bar{t}H$. (d) $t\bar{t}W$.

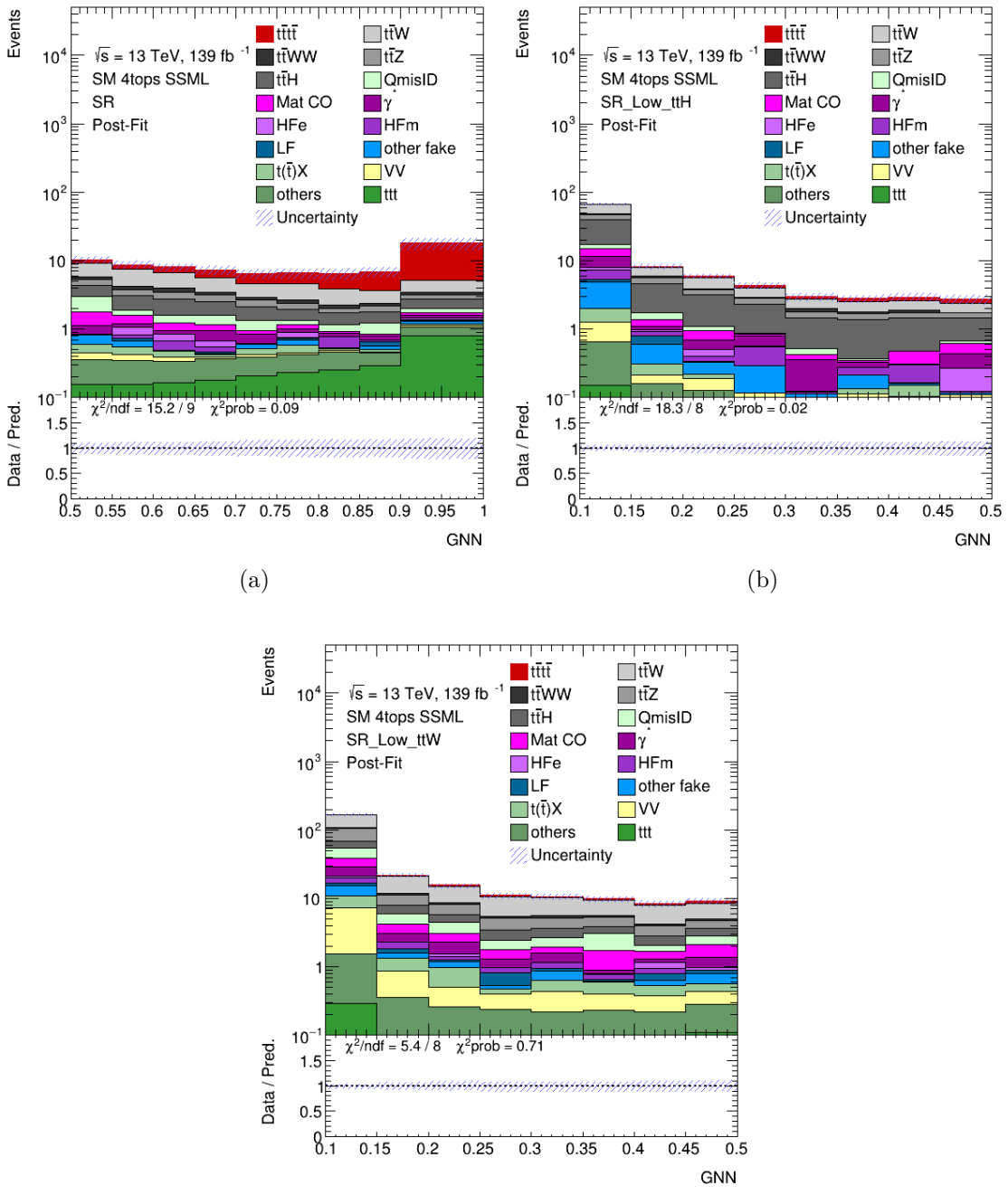


Figure 9.19: Expected distributions in the signal regions of the multi-class setup after a control region only fit. (a) $t\bar{t}t\bar{t}$ Signal Region. (b) $t\bar{t}H$ Signal Region. (c) Signal Region.

Chapter 10

Observation of Four-Top-Quark Production and Interpretations

10.1 Signal Strength and Cross-section Measurements

Cross-section and Signal Strength

The $t\bar{t}t\bar{t}$ production cross-section is measured using a maximum likelihood fit to the signal region and eight control regions described in chapter 8. The fit profiles the systematic uncertainties as described in chapter 6 and includes a single parameter of interest $\mu_{t\bar{t}t\bar{t}}$ scaling the overall cross-section of the $t\bar{t}t\bar{t}$ process. The expected signal and background yields before the fit to data and from the best fit model are shown in table 10.1 for the inclusive $t\bar{t}t\bar{t}$ signal region and the signal-like GNN ≥ 0.6 region. The data yields are also included and show an excess of 32 events relative to the pre-fit expectation in the GNN ≥ 0.6 region, which is reduced in the post-fit results by a 20 event increase in the expected $t\bar{t}t\bar{t}$ yield and a six event increase in the background expectation in this region. The profile likelihood ratio test statistic is shown as a function of the $t\bar{t}t\bar{t}$ production cross-section in fig. 10.1.

The best fit value of the cross-section is $\sigma_{t\bar{t}t\bar{t}} = 22^{+6.6}_{-5.5}$ fb, corresponding to a signal strength of $\mu_{t\bar{t}t\bar{t}} = 1.9 \pm 0.4(\text{stat.})^{+0.7}_{-0.4}(\text{syst.}) = 1.9^{+0.8}_{-0.5}$. The statistical significance of the observed signal is evaluated using the profile likelihood ratio to a null hypothesis in which $\sigma_{t\bar{t}t\bar{t}} = 0$. The significance estimated from the asymptotic method is 6.1σ, making this the first observation of $t\bar{t}t\bar{t}$ production with ATLAS. The expected significance when observing SM $t\bar{t}t\bar{t}$ production is 4.3σ when using the nominal SM $t\bar{t}t\bar{t}$ cross-section prediction of [41] and 4.7σ when using the prediction of [44]. The expected significance when observing $\mu = 1.89$ (2) is 6.82σ (7.08σ). Both are evaluated using Realistic Asimov datasets as described in chapter 6.

The compatibility of this result with the Standard Model expectation for $t\bar{t}t\bar{t}$ production is evaluated using the profile likelihood ratio with a null hypothesis of $\mu = 1$ and the nominal cross-section calculation [41]. The asymptotic approximation yields a statistical significance of 1.8σ, corresponding to a p-value of 0.06 to reject the Standard Model hypothesis. When

	Pre-fit		Post-fit	
	SR	GNN≥0.6	SR	GNN≥0.6
$t\bar{t}W$	130 ± 40	9 ± 4	127 ± 35	12 ± 4
$t\bar{t}Z$	72 ± 15	3.4 ± 1.8	79 ± 15	4.4 ± 2.0
$t\bar{t}H$	65 ± 11	4.6 ± 1.3	68 ± 10	5.0 ± 1.4
QmisID	27 ± 4	1.78 ± 0.26	27 ± 4	1.80 ± 0.24
Mat. Conv.	16.5 ± 2.3	0.73 ± 0.25	30 ± 8	1.4 ± 0.5
HF e	3.1 ± 1.0	0.4 ± 0.5	2.3 ± 2.4	0.3 ± 0.4
HF μ	7.1 ± 1.2	0.31 ± 0.15	9 ± 4	0.41 ± 0.22
Low m_{γ^*}	14.1 ± 2.0	0.52 ± 0.19	15 ± 5	0.56 ± 0.22
Others	47 ± 11	3.9 ± 1.2	50 ± 10	4.3 ± 1.2
$t\bar{t}t$	2.9 ± 0.9	1.5 ± 0.5	2.9 ± 0.9	1.5 ± 0.5
Total bkg	390 ± 50	26 ± 5	412 ± 21	32 ± 4
$t\bar{t}t\bar{t}$	38 ± 4	25.2 ± 3.2	69 ± 15	45 ± 10
Total	430 ± 50	51 ± 7	480 ± 19	77 ± 8
Data	482	83	482	83

Table 10.1: Pre-fit and post-fit background and signal yields in the signal region and for events with GNN score larger than 0.6. The total systematic uncertainty differs from the sum in quadrature of the different uncertainties due to correlations.

considering the updated cross-section calculation of [44], this becomes 1.7σ. There are competing effects between the smaller uncertainties of the more precise calculation and the higher central value.

Fit Quality and Signal Shape

The goodness of fit is measured using a saturated model [188, 189], yielding a probability of 76% and indicating good agreement with the post-fit signal plus background model. To further investigate the consistency of the observed signal with the $t\bar{t}t\bar{t}$ expectation, several sensitive observables are checked in the signal region with an additional cut requiring the GNN score to be at least 0.6 to see if the distributions in data match those of the expected $t\bar{t}t\bar{t}$ signal. These distributions are shown in fig. 10.2.

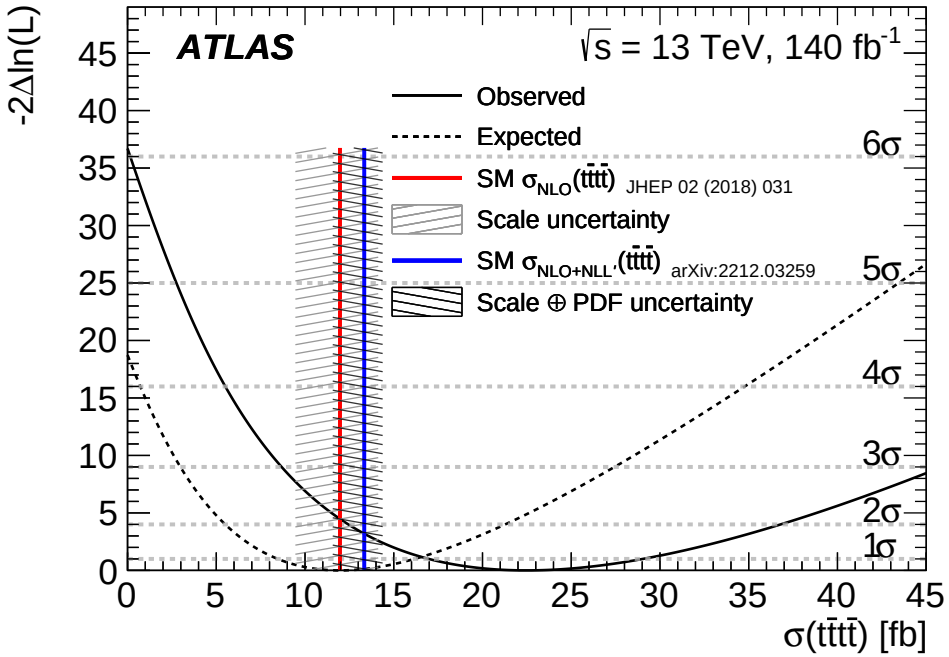


Figure 10.1: Profile likelihood ratio as a function of the $t\bar{t}t\bar{t}$ signal cross-section. The continuous line represents the observed likelihood while the dashed line corresponds to the expected one. The red line shows the SM prediction computed at NLO in QCD and EW from Ref. [41] with its scale uncertainty (dashed light gray region) and the blue line shows the resummed $\sigma_{t\bar{t}t\bar{t}}$ calculation of Ref. [44] with its scale and PDF uncertainty (dashed dark gray region).

The cut of 0.6 is chosen to include the bin $0.6 <$ GNN score < 0.65, which shows the most significant excess of the data over the signal plus background prediction. These same distributions are additionally checked inclusively in the signal region (fig. 10.3), in the background dominated region GNN score < 0.6 (fig. 10.4), and in the most signal-like bin, GNN score > 0.9 (fig. 10.5).

Each of these sets of distributions shows good agreement in the overall multiplicity, number of b-jets, and overall hardness of the event measured by H_T, with all of these variables tending to be significantly higher in the signal-like regions. The good overall agreement indicates that the background is well-modeled and the observed signal is similar to the SM expectation for $t\bar{t}t\bar{t}$ production.

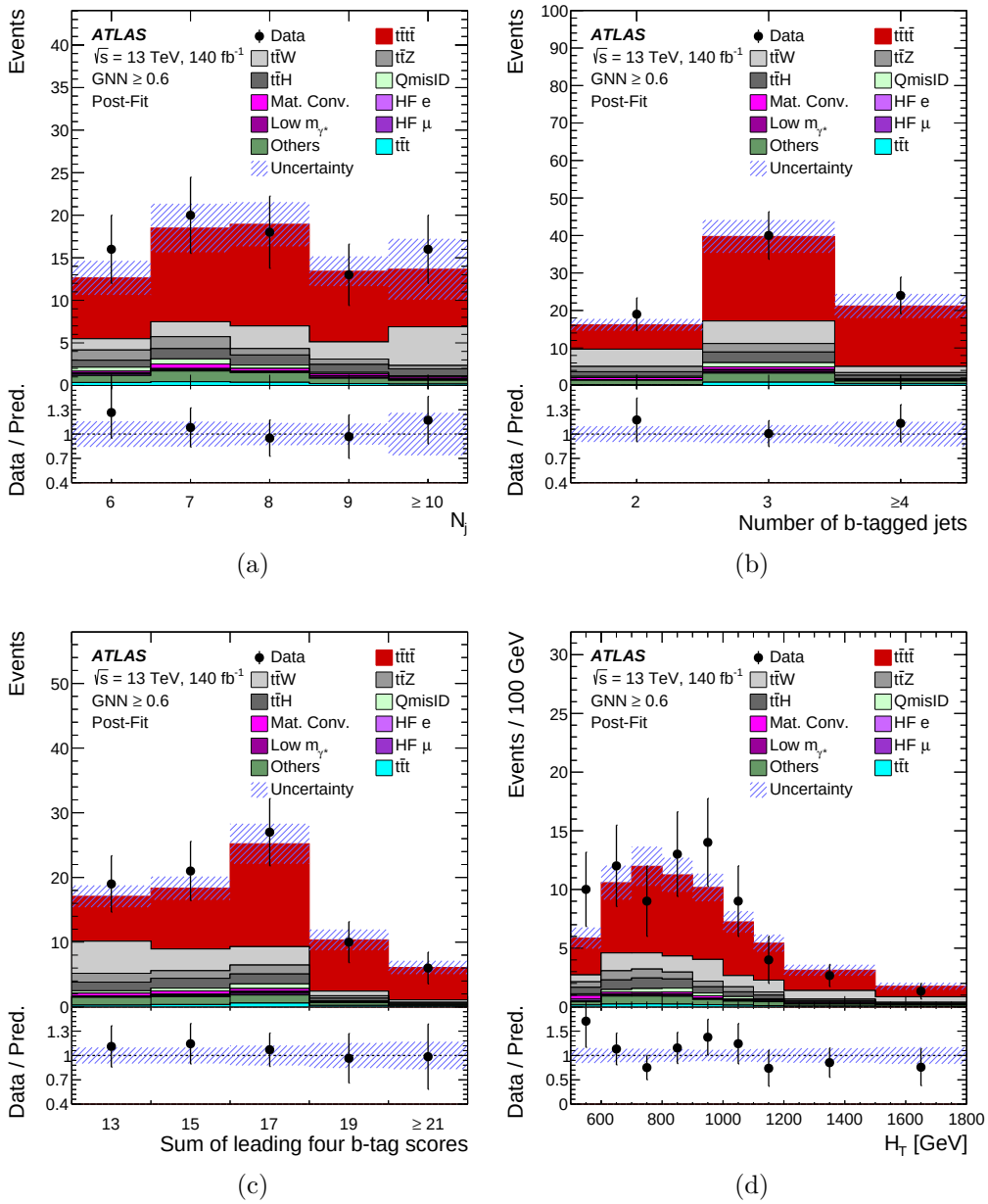


Figure 10.2: Comparison between data and prediction in the signal-enriched region with GNN ≥ 0.6 after the fit to data for the distributions of (a) the number of jets, (b) the number of b-jets, (c) the sum of the four highest PCBT scores of jets in the event, and (d) the sum of transverse momenta over all jets and leptons in the event (H_T). The ratio of the data to the total post-fit computation is shown in the lower panel. The shaded band represents the total post-fit uncertainty in the prediction. The first and last bins contain underflow and overflow events, respectively. Reproduced from [129].

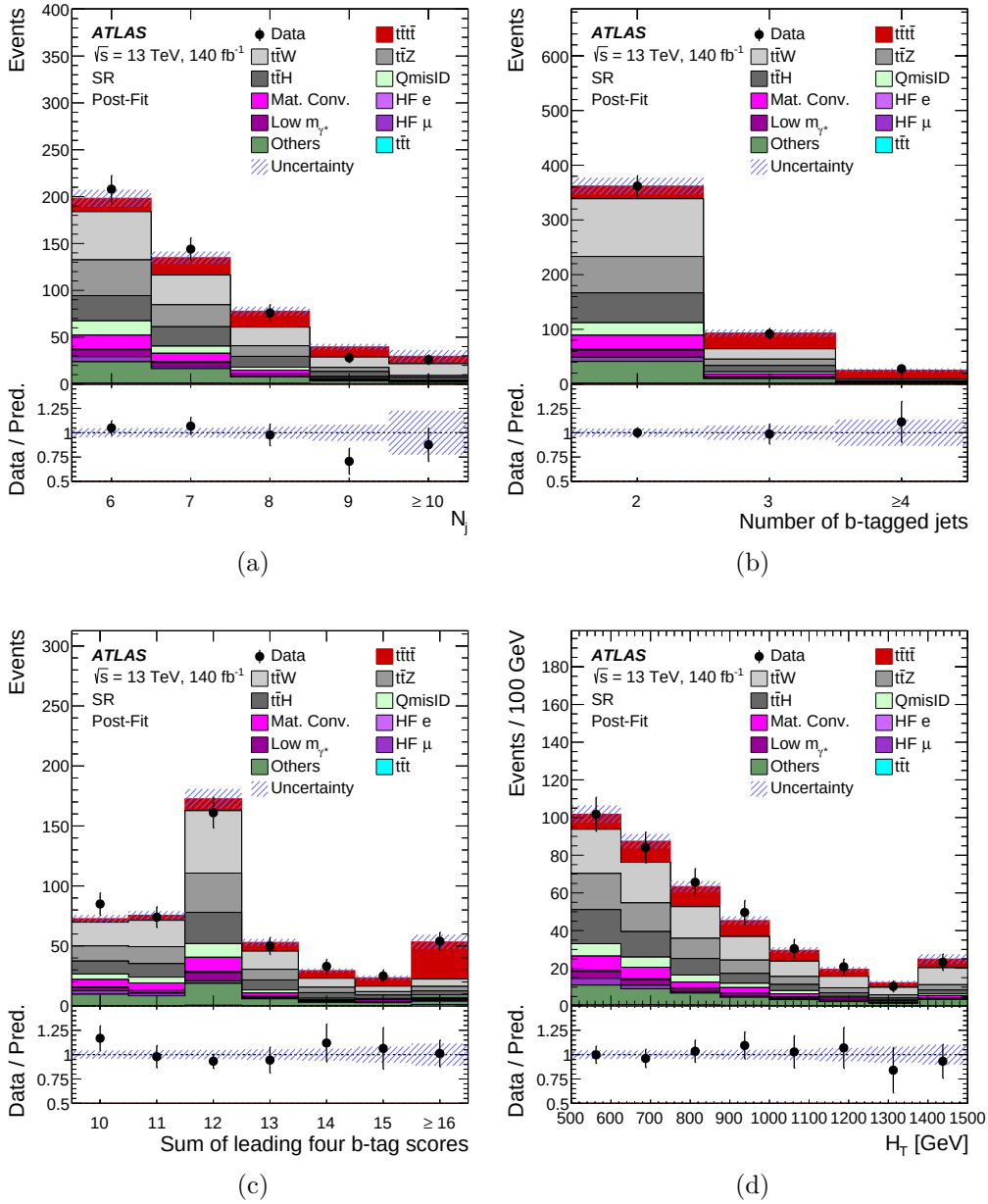


Figure 10.3: Comparison between data and prediction in the signal region after the fit to data for the distributions of (a) the number of jets, (b) the number of b-jets, (c) the sum of the four highest PCBT scores of jets in the event, and (d) the sum of transverse momenta over all jets and leptons in the event (H_T). The ratio of the data to the total post-fit computation is shown in the lower panel. The shaded band represents the total post-fit uncertainty in the prediction. The first and last bins contain underflow and overflow events, respectively.

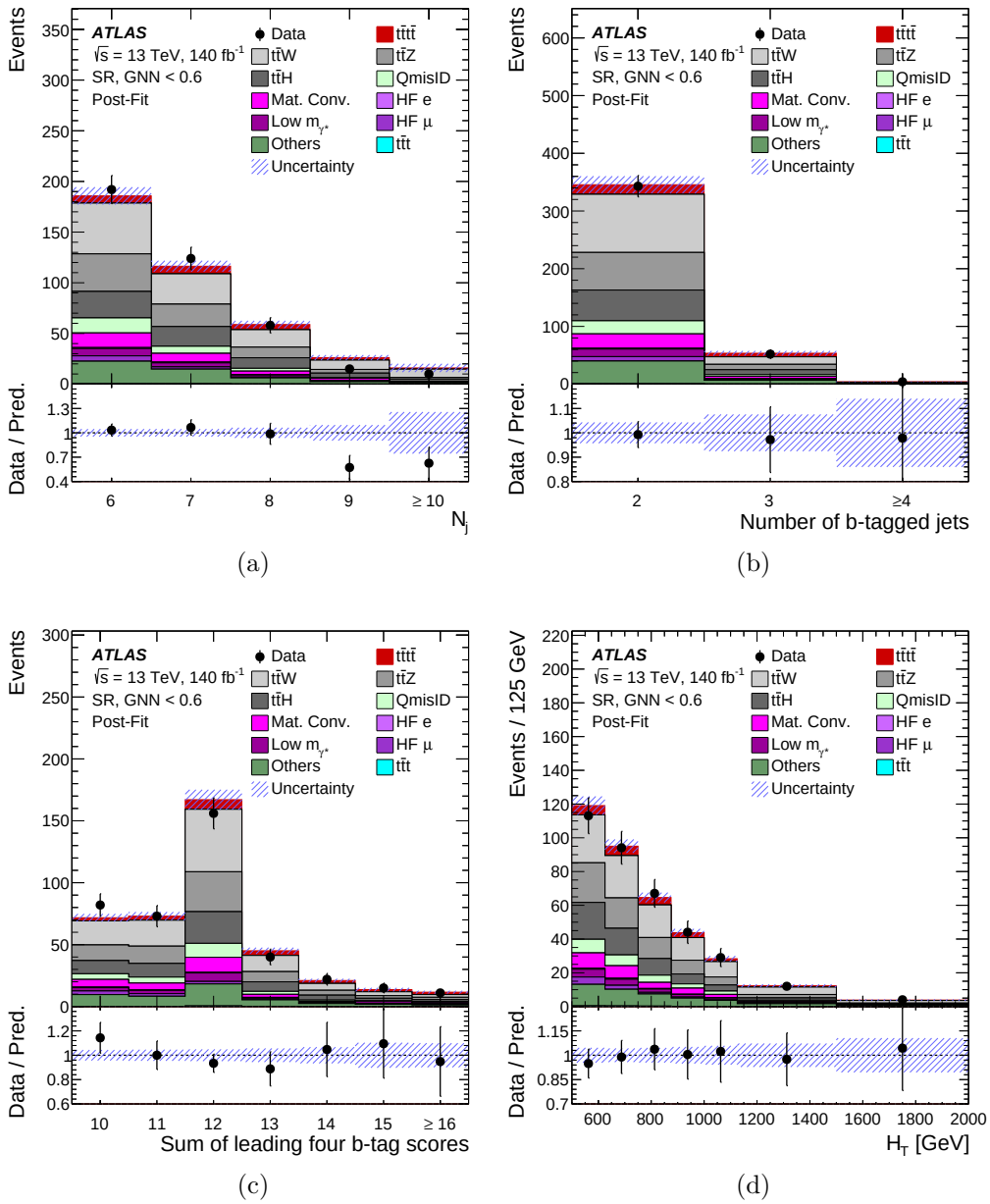


Figure 10.4: Comparison between data and prediction in the background dominated GNN < 0.6 region after the fit to data for the distributions of (a) the number of jets, (b) the number of b-jets, (c) the sum of the four highest PCBT scores of jets in the event, and (d) the sum of transverse momenta over all jets and leptons in the event (H_T). The ratio of the data to the total post-fit computation is shown in the lower panel. The shaded band represents the total post-fit uncertainty in the prediction. The first and last bins contain underflow and overflow events, respectively.

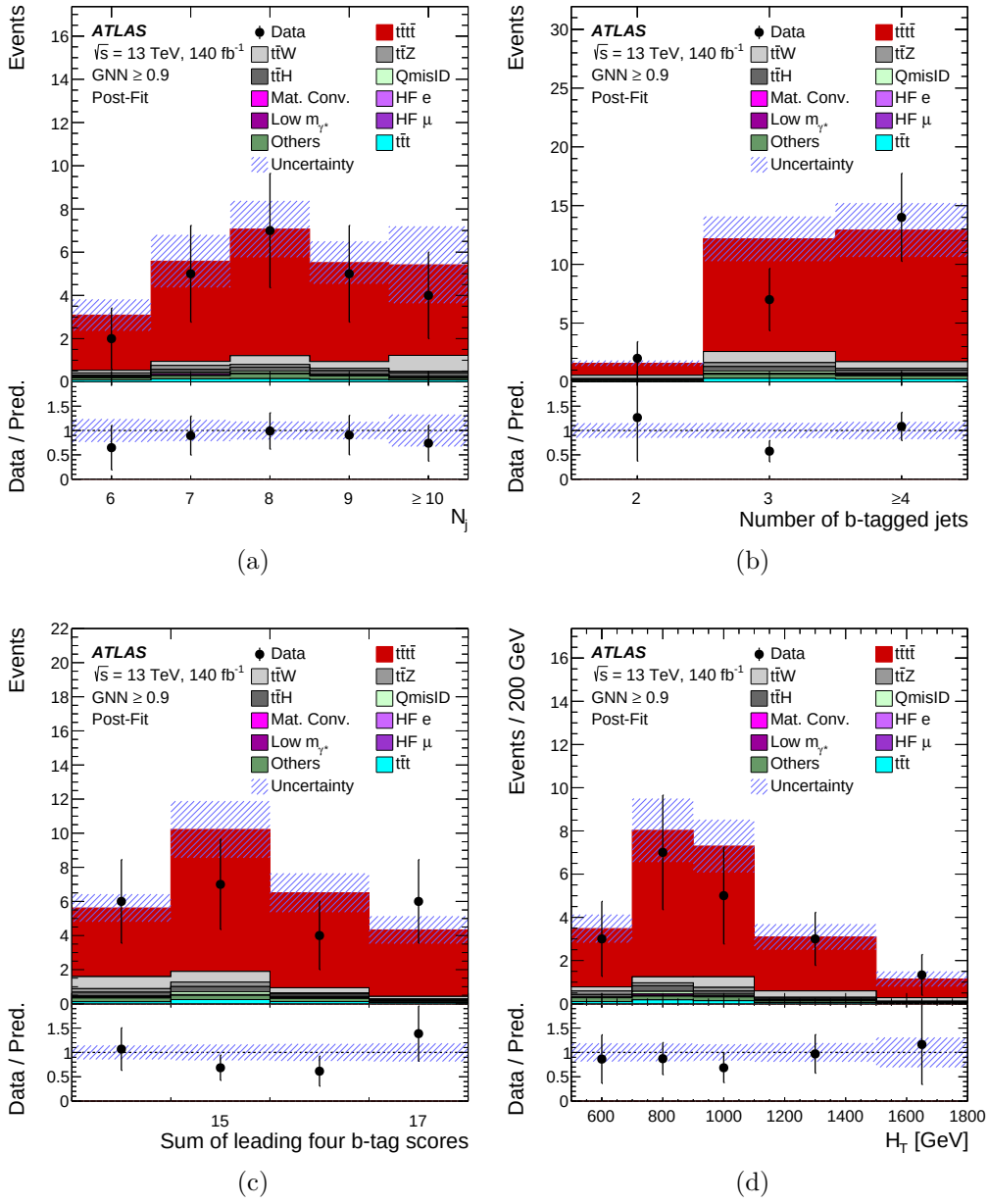


Figure 10.5: Comparison between data and prediction in the signal-enriched region with GNN ≥ 0.9 after the fit to data for the distributions of (a) the number of jets, (b) the number of b-jets, (c) the sum of the four highest PCBT scores of jets in the event, and (d) the sum of transverse momenta over all jets and leptons in the event (H_T). The ratio of the data to the total post-fit computation is shown in the lower panel. The shaded band represents the total post-fit uncertainty in the prediction. The first and last bins contain underflow and overflow events, respectively.

Verification with BDT and H_T Fits

As additional checks on the observed $t\bar{t}t\bar{t}$ signal, the fit is performed with the GNN discriminant replaced either by a BDT trained analogously to that of [60] but re-optimized for the improved object selection and flavor tagging, or by the scalar sum of transverse momenta H_T. The BDT fit has an expected (observed) significance of 6.0σ (3.9σ) and gives a best fit signal strength of $1.9^{+0.7}_{-0.5}$, while the H_T fit has an expected (observed) significance of 5.0σ (2.8σ) and gives a best fit signal strength of $2.1^{+0.8}_{-0.6}$. These results are qualitatively consistent with those of the fit to the GNN score distribution. The compatibility is not tested quantitatively, as such a test would require throwing large numbers of toys to account for the correlations between the GNN score and the BDT and H_T discriminants.

10.2 Interpretations of the Three-Top Background

One contributor to the sensitivity increase in this analysis compared to that of Ref. [60] is the use of an NLO prediction for the $t\bar{t}t$ cross-section, which is accompanied by a substantially reduced systematic uncertainty on the $t\bar{t}t$ normalization. The sensitivity to this small process arises from a near degeneracy in the GNN score shape of $t\bar{t}t\bar{t}$ and $t\bar{t}t$, as shown in fig. 9.10 and fig. 10.6. A similar degeneracy is present in the BDT discriminant of Ref. [60] and the BDT and H_T fits performed in this work as cross-checks of the GNN result. These degeneracies result from the underlying similarity between the $t\bar{t}t\bar{t}$ and $t\bar{t}t$ production processes. Because no measurement has placed limits on SM $t\bar{t}t$ production or provided limits on the $t\bar{t}t$ cross-section, several interpretations are performed considering the possible contribution of BSM enhancements of $t\bar{t}t$ production to the observed signal.

Simultaneous Fit of $t\bar{t}t$ and $t\bar{t}t\bar{t}$

The first test is a simultaneous measurement of the $t\bar{t}t$ and $t\bar{t}t\bar{t}$ cross-sections where the ratio of $t\bar{t}tq$ production to $t\bar{t}tW$ is kept at the SM value. The constraints derived from this fit are shown in fig. 10.7. The $t\bar{t}t$ and $t\bar{t}t\bar{t}$ cross-sections are strongly anti-correlated in the fit because both processes have similar shapes. The correlation coefficient derived from the covariance matrix is 93%. The fit slightly prefers to have most of the signal come from $t\bar{t}t$ production, but is still consistent with the $t\bar{t}t\bar{t}$ only hypothesis at the 1σ level.

Limits on the $t\bar{t}t$ cross-section are derived by making assumptions on the $t\bar{t}t\bar{t}$ cross-section. These are equivalent to producing a one-dimensional likelihood scan by taking a horizontal slice of fig. 10.7. The 95% CL intervals for the $t\bar{t}t$ cross-section are shown table 10.2 with the assumption of SM $t\bar{t}t\bar{t}$ production or using the best fit value of $\sigma_{t\bar{t}t\bar{t}}$. $t\bar{t}tW$ and $t\bar{t}tq$ production are sensitive to different BSM modifications, with $t\bar{t}tW$ having similar effects as $t\bar{t}t\bar{t}$ and $t\bar{t}tq$ being particularly sensitive to flavor changing neutral currents [265, 266]. Additional limits are placed on the individual $t\bar{t}tW$ and $t\bar{t}tq$ cross-sections with the other $t\bar{t}t$ production mode fixed to the SM prediction.

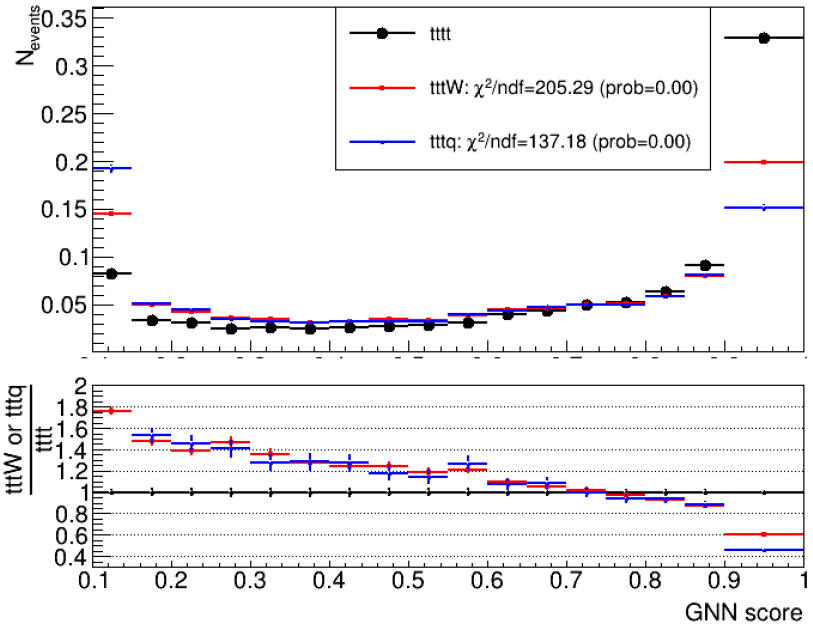


Figure 10.6: GNN shape comparison of $t\bar{t}tW$, $t\bar{t}tq$, $t\bar{t}t\bar{t}$ in signal region

Processes	95% CL cross-section interval [fb]	
	$\mu_{t\bar{t}t\bar{t}} = 1$	$\mu_{t\bar{t}t\bar{t}} = 1.9$
$t\bar{t}t$	[4.7, 60]	[0, 41]
$t\bar{t}tW$	[3.1, 43]	[0, 30]
$t\bar{t}tq$	[0, 144]	[0, 100]

Table 10.2: Observed 95% CL intervals for the $t\bar{t}t$, $t\bar{t}tq$, and $t\bar{t}tW$ cross-sections assuming a $t\bar{t}t\bar{t}$ signal strength of 1.0 or 1.9. To derive the $t\bar{t}tW$ ($t\bar{t}tq$) cross-section interval, the $t\bar{t}tq$ ($t\bar{t}tW$) cross-section is fixed to its SM prediction. Reproduced from [129].

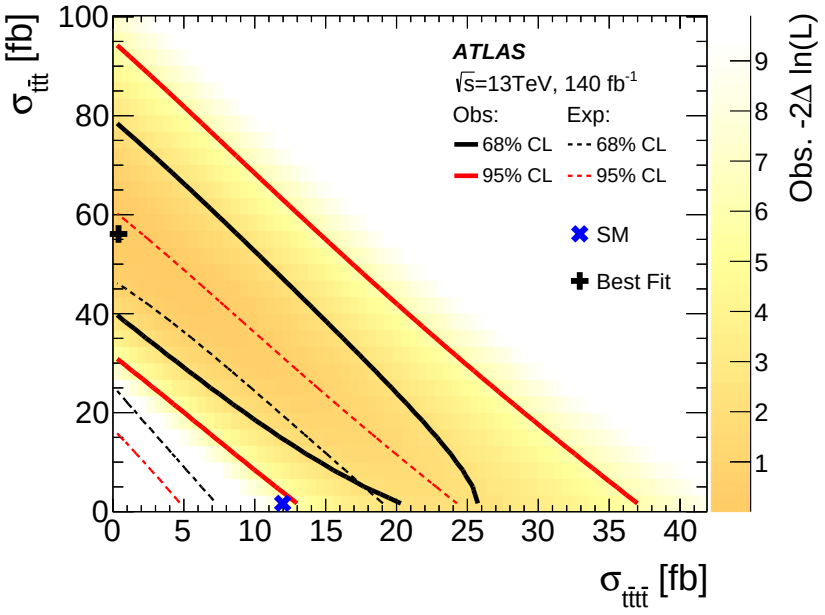


Figure 10.7: Two-dimensional negative log-likelihood contour for the $t\bar{t}t$ cross-section ($\sigma_{t\bar{t}t}$) versus the $t\bar{t}t\bar{t}$ cross-section ($\sigma_{t\bar{t}t\bar{t}}$) when the normalizations of both processes are treated as free parameters in the fit. The blue cross shows the SM expectation of $\sigma_{t\bar{t}t\bar{t}} = 12$ fb from Ref. [41] and $\sigma_{t\bar{t}t} = 1.67$ fb, both computed at NLO, while the black cross shows the best-fit value. The observed (expected) exclusion contours at 68% (black) and 95% CL (red) are shown in solid (dashed) lines. The gradient-shaded area represents the observed likelihood value as a function of $\sigma_{t\bar{t}t}$ and $\sigma_{t\bar{t}t\bar{t}}$. Reproduced from [129].

When $\sigma_{t\bar{t}t\bar{t}}$ is fixed to its best fit value, $\sigma_{t\bar{t}t}$ is consistent with zero, but when $\sigma_{t\bar{t}t\bar{t}}$ is fixed to the SM prediction, the 95% CL limits on $t\bar{t}t$ and $t\bar{t}tW$ exclude both zero and the SM predictions for those processes because the remaining observed signal is absorbed by the floating $t\bar{t}t$ normalization.

10.3 Effective Field Theory Constraints

Constraints on the Wilson coefficients of effective field theory operators are derived using the signal parameterization described in chapter 6. Other than the bin-by-bin parameterization of $t\bar{t}t\bar{t}$ production, the fit is the same as that used to measure $\sigma_{t\bar{t}t\bar{t}}$.

Four Fermion Operators

The 95% CL intervals for the four fermion Wilson coefficients when considering the variation of only one coefficient at a time are shown in table 10.3. The results are consistent

with the SM prediction for all four operators, though the observed limits are wider than expected because the observed $t\bar{t}t\bar{t}$ signal is larger than the SM expectation. The expected effect from most EFT contributions to $t\bar{t}t\bar{t}$ production is an increase in $\sigma_{t\bar{t}t\bar{t}}$ and an especially large excess in the high H_T region. The former effect is observed, but the latter is not visible in figs. 10.2 and 10.5. The observed limits are consistent with expected limits evaluated with an injected $t\bar{t}t\bar{t}$ signal strength of 2.

Operators	Expected C_i/Λ^2 [TeV^{-2}]	Observed C_i/Λ^2 [TeV^{-2}]
$\mathcal{O}^1_{QQ}$	[-2.5, 3.2]	[-4.0, 4.5]
$\mathcal{O}^1_{Qt}$	[-2.6, 2.1]	[-3.8, 3.4]
$\mathcal{O}^1_{tt}$	[-1.2, 1.4]	[-1.9, 2.1]
$\mathcal{O}^8_{Qt}$	[-4.3, 5.1]	[-6.9, 7.6]

Table 10.3: Expected and observed 95% CL intervals on EFT coupling parameters assuming one EFT parameter variation in the fit.

An additional set of limits is derived ignoring the quadratic contributions in eq. (6.11). Here, the limits are set on the absolute value of the Wilson coefficient. This interpretation gives a sense of whether the EFT constraints are dominated by the linear terms, which are $\mathcal{O}(\Lambda^{-2})$ and receive no modifications from dimension eight EFT operators, or by the quadratic terms, which contain $\mathcal{O}(\Lambda^{-4})$ effects entering at the same order of the expansion in Λ^{-1} as dimension eight operators.

Operators	Expected $\lvert C_i/\Lambda^2\rvert$ [TeV^{-2}]	Observed $\lvert C_i/\Lambda^2\rvert$ [TeV^{-2}]
$\lvert\mathcal{O}^1_{QQ}\rvert$	2.3	6.6
$\lvert\mathcal{O}^1_{Qt}\rvert$	1.4	4.0
$\lvert\mathcal{O}^1_{tt}\rvert$	1.0	2.8
$\lvert\mathcal{O}^8_{Qt}\rvert$	3.6	10.8

Table 10.4: Expected and observed 95% CL upper limits on the magnitude of EFT coupling parameters assuming one EFT parameter variation in the fit and considering effects up to order $1/\Lambda^2$ in the cross-section.

Each operator shows comparable but weaker limits than those of table 10.3, indicating that some constraint power comes from the quadratic terms. This is generally expected because the quadratic terms do not rely on interference with SM diagrams and can populate a less SM-like phase space. All these results are obtained by varying a single Wilson coefficient at a time. The expected limits obtained on each coefficient from fitting all four coefficients

simultaneously, including the quadratic effects, and profiling the remaining three coefficients, are quite similar to those obtained by varying only a single coefficient at a time. This is expected because all variations increase the yield in the $t\bar{t}t\bar{t}$ yield in the most signal-like bins. At the limit point, the profiling therefore prefers small values of the profiled coefficients even when they are not fixed to 0. The observed limits derived from fits including cross terms are not produced.

Higgs Oblique Parameter

Limits on the Higgs oblique parameter are set following the same procedure, albeit with only a single parameter of interest. The likelihood scan for this interpretation is shown in fig. 10.8.

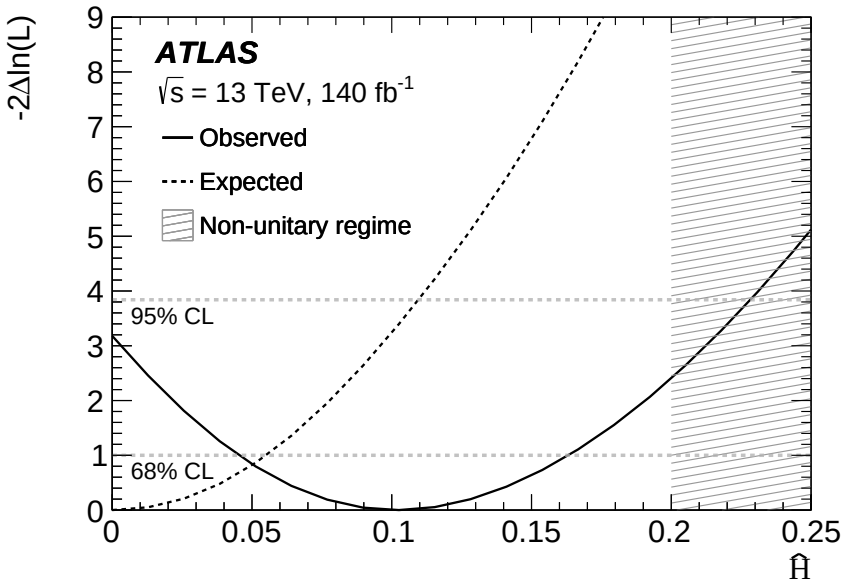


Figure 10.8: The negative log-likelihood values as a function of the Higgs oblique parameter $\hat{H}$. The solid line represents the observed likelihood while the dashed line corresponds to the expected one. The dashed region shows the non-unitary regime.

The observed upper limit according to the parameterization described in chapter 6 is $\hat{H} < 0.23$, while the expected limit is $\hat{H} < 0.11$. The observed excess in the $t\bar{t}t\bar{t}$ signal results in an increase in the limit above the expected limit and moves it outside the region of perturbative validity $\hat{H} < 0.2$ defined by unitarity, shown by the shaded region of fig. 10.8. Previous limits on $\hat{H}$ have been reported in Refs. [55, 100, 267].

10.4 Off-Shell Measurement of the Top Yukawa Coupling

Limits on the top Yukawa coupling are extracted following the same procedure as those on EFT coefficients. The underlying parameter space is two-dimensional, and several one-dimensional limits are also produced assuming that the top Yukawa coupling is CP-even. The results are divided into the two treatments of $t\bar{t}H$, parameterized and profiled, as described in chapter 6.

$t\bar{t}H$ Parameterized

When both $t\bar{t}t\bar{t}$ and $t\bar{t}H$ are parameterized as a function of the top Yukawa coupling and $\tilde{\kappa}_t$ is fixed to 0, the observed (expected) 95% CL limit on κ_t is $|\kappa_t| < 1.9$ (1.6). The likelihood scan associated with this result is shown in fig. 10.10. The movement of the central value is associated with the observed excess in the $t\bar{t}t\bar{t}$ signal. The large improvement in the exclusion near $\kappa_t = 0$ is due to the small decrease of the $t\bar{t}t\bar{t}$ cross-section when $|\kappa_t| < 1$. The two-dimensional limit plot in fig. 10.9 shows that the fit prefers a large CP mixing angle. Increasing the CP mixing angle allows the fit to accommodate a larger $t\bar{t}t\bar{t}$ cross-section without significantly increasing the $t\bar{t}H$ cross-section.

The best fit values in the two dimensional fit are $|\kappa_t| = 1.11^{+0.47}_{-1.11}$ and $|\tilde{\kappa}_t| = 1.03^{+0.38}_{-0.65}$, which correspond to signal strengths of $\mu_{t\bar{t}t\bar{t}} = 1.9$ and $\mu_{t\bar{t}H} = 1.7$. The $t\bar{t}t\bar{t}$ signal strength is consistent with the measured value in the $t\bar{t}t\bar{t}$ cross-section measurement. When $\tilde{\kappa}_t$ is fixed to 0, the best fit value of κ_t becomes $|\kappa_t| = 1.51^{+0.22}_{-0.26}$, which corresponds to $\mu_{t\bar{t}t\bar{t}} = 1.3$ and $\mu_{t\bar{t}H} = 2.3$. This parameterization does not fully accommodate the observed $t\bar{t}t\bar{t}$ excess because such a variation induces a large increase in the $t\bar{t}H$ background inconsistent with the observed data.

$t\bar{t}H$ Profiled

When $t\bar{t}H$ is kept to its SM shape but given a floating normalization, the observed (expected) 95% limit in the pure CP-even interpretation becomes $|\kappa_t| < 2.3$ (1.9). The limits are somewhat relaxed without the constraint from the $t\bar{t}H$ background. The difference between the observed and expected limits increases because the $t\bar{t}H$ process does not constrain κ_t from adopting the value preferred by the $t\bar{t}t\bar{t}$ signal strength of 1.9. The two-dimensional limit in this case, shown in fig. 10.11, has a near degeneracy in the CP mixing angle. This degeneracy arises because the top Yukawa sensitivity is dominated by $\sigma_{t\bar{t}t\bar{t}}$ and therefore the two-dimensional constraint mostly comes from one measured parameter.

The best fit values in the two dimensional fit are $|\kappa_t| = 1.40^{+1.11}_{-1.40}$ and $|\tilde{\kappa}_t| = 0.90^{+0.81}_{-0.90}$, which corresponds to a signal strength of $\mu_{t\bar{t}t\bar{t}} = 2.0$, and the fitted $t\bar{t}H$ normalization is $\mu_{t\bar{t}H} = 1.53 \pm 0.89$. The large uncertainties on κ_t and $\tilde{\kappa}_t$ result from the mostly flat direction, which keeps $\mu_{t\bar{t}t\bar{t}}$ constant. When $\tilde{\kappa}_t$ is fixed to 0, the best fit value of κ_t be-

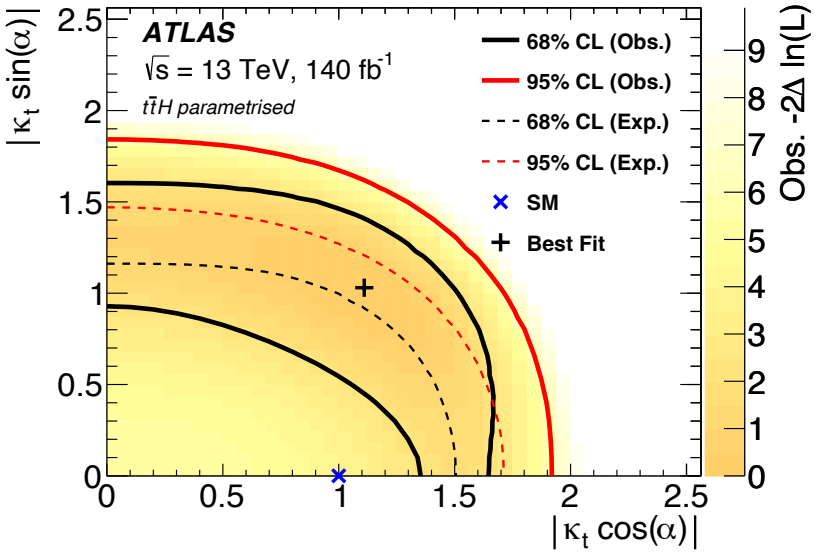


Figure 10.9: Two-dimensional negative log-likelihood contours for $|\kappa_t \cos(\alpha)|$ versus $|\kappa_t \sin(\alpha)|$ at 68% and 95%, where κ_t is the top-Higgs Yukawa coupling strength parameter and α is the mixing angle between the CP-even and CP-odd components. The gradient-shaded area represents the observed likelihood value as a function of κ_t and α. Both the $t\bar{t}t\bar{t}$ signal and $t\bar{t}H$ background yields in each fitted bin are parameterized as a function of κ_t and α. The blue cross shows the SM expectation, while the black cross shows the best fit value.

comes $|\kappa_t| = 1.88^{+0.24}_{-0.28}$, which corresponds to $\mu_{t\bar{t}t\bar{t}} = 2.0$, and the fitted $t\bar{t}H$ normalization is $\mu_{t\bar{t}H} = 1.13 \pm 1.13$. With $t\bar{t}H$ profiled, a single degree of freedom in the top Yukawa sector can accommodate the observed $t\bar{t}t\bar{t}$ signal. This fit cannot significantly constrain the $t\bar{t}H$ normalization due to the large correlations with $t\bar{t}W$.

Table 10.5 shows a comparison of the best fit values of κ_t and $\tilde{\kappa}_t$ for the different top Yukawa scenarios with the SM cross-section measurement. Also shown are the $t\bar{t}t\bar{t}$ and $t\bar{t}H$ signal strengths calculated from κ_t and $\tilde{\kappa}_t$ or retrieved directly from the fit, as appropriate, and the $t\bar{t}W$ normalization factors. All four top Yukawa fits result in $t\bar{t}H$ normalizations higher than the SM expectation, which are generally compensated by lower normalizations for $t\bar{t}W$. The highest $t\bar{t}H$ normalization occurs when $t\bar{t}H$ is parameterized as a function of κ_t and $\tilde{\kappa}_t$ is fixed to 0, in which case $\mathrm{NF}_{t\bar{t}W-(4\mathrm{jet})}$ reaches its lowest value, 0.7. The $t\bar{t}t\bar{t}$ dependence on the top Yukawa coupling is mostly determined by the cross-section variation and the best fit values always correspond to a $t\bar{t}t\bar{t}$ signal strength close to that of the SM cross-section measurement except when $t\bar{t}H$ is parameterized as a function of κ_t and $\tilde{\kappa}_t$ is fixed to 0, in which case there are not enough degrees of freedom for $t\bar{t}t\bar{t}$ and $t\bar{t}H$ to take on their preferred normalizations.

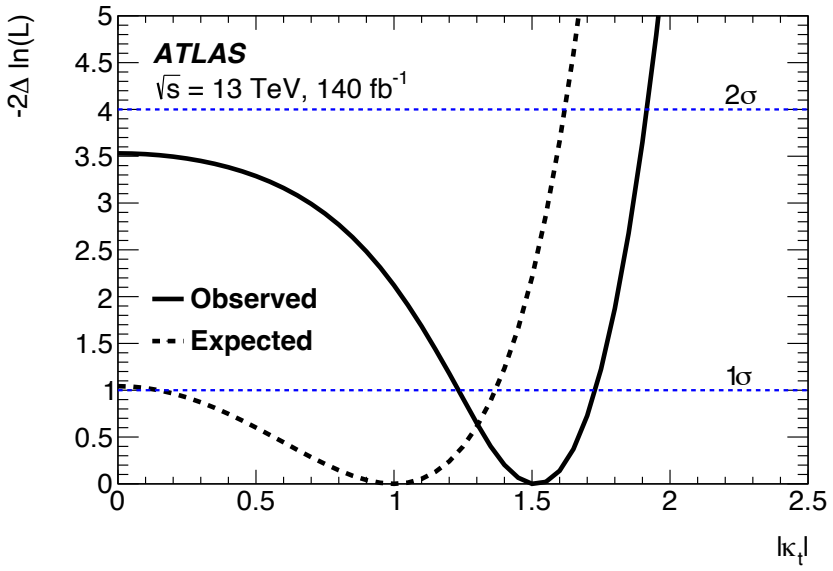


Figure 10.10: Negative log-likelihood curves for the measurement on κ_t. The top-Higgs Yukawa coupling is assumed to be purely CP-even ($\alpha = 0$). The yields of both the $t\bar{t}t\bar{t}$ signal and the $t\bar{t}H$ background in each fitted bin are parameterized as a function of κ_t. The observed (expected) distribution is shown with the solid (dashed) line.

Consistency with other Top Yukawa Constraints

The top Yukawa constraints reported here are generally less stringent than those reported in Ref. [100] using the same method and prefer larger values of κ_t. This difference is largely due to the higher observed $t\bar{t}t\bar{t}$ signal strength in this work. A comparison to other measurements of the strength and CP properties of the top-Higgs coupling using $t\bar{t}H$ and tH production in ATLAS is shown in fig. 10.13. In contrast to the $t\bar{t}t\bar{t}$based measurement, measurements of $t\bar{t}H$ production are more precise, but rely on assumptions about the total width of the Higgs boson and must input information about the Higgs boson branch ratios by either assuming the SM values or using inputs from measurements of other Higgs boson production modes. The $t\bar{t}t\bar{t}$based measurement instead requires no assumptions about the Higgs boson's couplings to other particles, including its total width, but instead assumes no other BSM physics significantly affects $t\bar{t}t\bar{t}$ production. The comparison illustrates the complementarity of measurements of different final states, which vary in their assumptions and sensitivity to different scenarios.

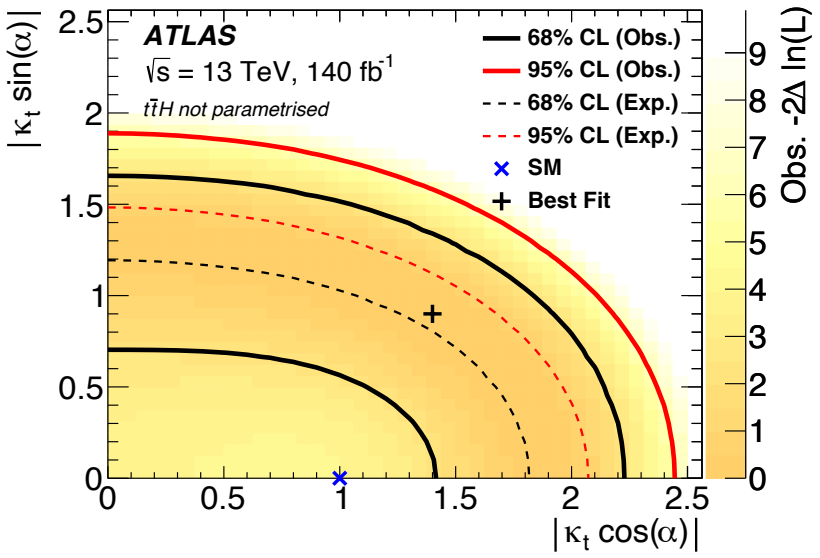


Figure 10.11: Two-dimensional negative log-likelihood contours for $|\kappa_t \cos(\alpha)|$ versus $|\kappa_t \sin(\alpha)|$ at 68% and 95%, where κ_t is the top-Higgs Yukawa coupling strength parameter and α is the mixing angle between the CP-even and CP-odd components. The gradient-shaded area represents the observed likelihood value as a function of κ_t and α. The $t\bar{t}t\bar{t}$ signal yield in each fitted bin is parameterized as a function of κ_t and α, and the $t\bar{t}H$ background is given a floating normalization. The blue cross shows the SM expectation, while the black cross shows the best fit value.

Fit Scenario	κ_t	$\tilde{\kappa}_t$	$\mu_{t\bar{t}t\bar{t}}$	$\mu_{t\bar{t}H}$	$\mathrm{NF}_{t\bar{t}W^+(\mathrm{4jet})}$	$\mathrm{NF}_{t\bar{t}W^-(\mathrm{4jet})}$
$\sigma(t\bar{t}t\bar{t})$	—	—	1.9	1.0	$1.27^{+0.25}_{-0.22}$	$1.11^{+0.31}_{-0.28}$
$t\bar{t}H$ Parameterized	$1.11^{+0.47}_{-1.11}$	$1.03^{+0.38}_{-0.65}$	1.9	1.7	1.20 ± 0.25	0.89 ± 0.42
$t\bar{t}H$ Profiled	$1.40^{+1.11}_{-1.40}$	$0.90^{+0.81}_{-0.90}$	2.0	1.53	1.22 ± 0.25	0.94 ± 0.43
$t\bar{t}H$ Parameterized, $\tilde{\kappa}_t = 0$	$1.51^{+0.22}_{-0.26}$	0	1.3	2.3	1.10 ± 0.24	0.70 ± 0.34
$t\bar{t}H$ Profiled, $\tilde{\kappa}_t = 0$	$1.88^{+0.24}_{-0.28}$	0	2.0	1.1	1.01 ± 0.21	0.93 ± 0.43

Table 10.5: Best fit values of κ_t, $\tilde{\kappa}_t$, $\mathrm{NF}_{t\bar{t}W^+(\mathrm{4jet})}$, and $\mathrm{NF}_{t\bar{t}W^-(\mathrm{4jet})}$ together with the corresponding values of the $t\bar{t}t\bar{t}$ and $t\bar{t}H$ signal strengths for the SM $t\bar{t}t\bar{t}$ cross-section fit and the four top Yukawa fit scenarios.

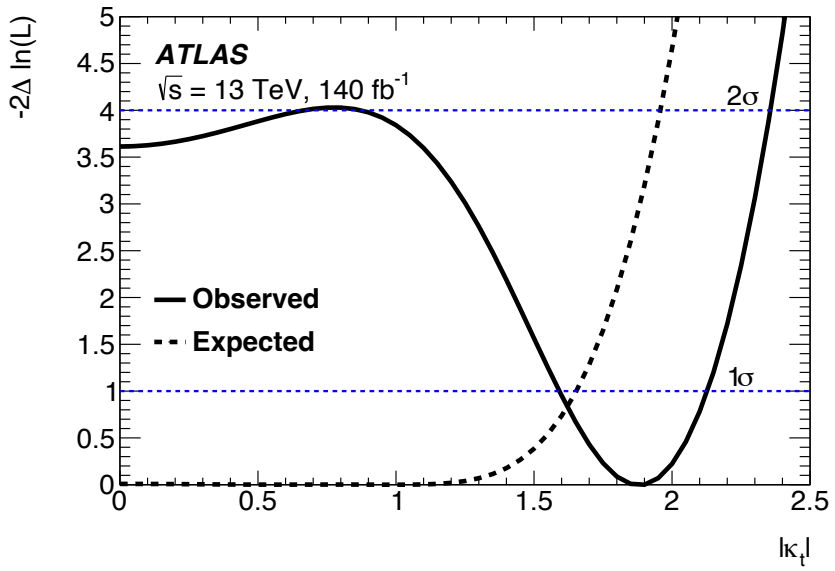


Figure 10.12: Negative log-likelihood contours for the measurement on κ_t. The top-Higgs Yukawa coupling is assumed to be purely CP-even ($\alpha = 0$). The $t\bar{t}t\bar{t}$ yield in each fitted bin is parameterized as a function of κ_t, and the normalization of the $t\bar{t}H$ background is profiled. The observed (expected) distribution is shown with the solid (dashed) line.

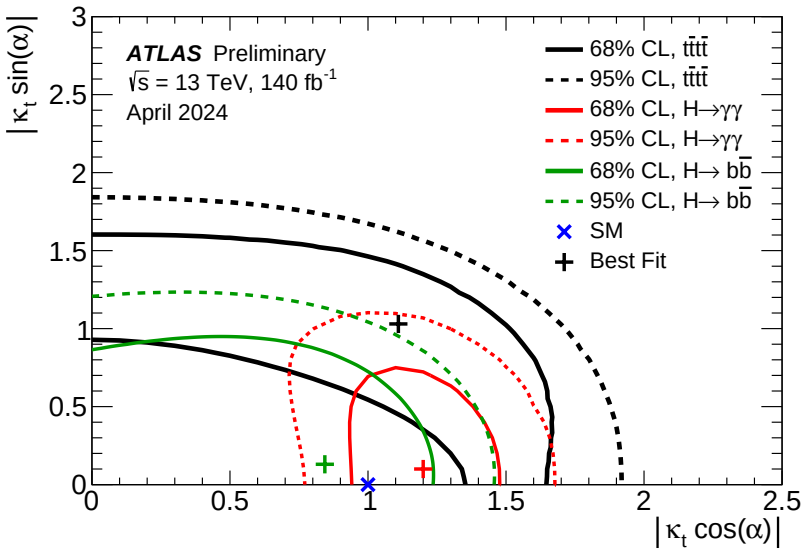


Figure 10.13: Summary of the two-dimensional limit contours for $\kappa_t \cos\alpha$ and $\kappa_t \sin\alpha$ at $\sqrt{s} = 13$ TeV with the ATLAS experiment. Here, κ_t represents the top-Higgs Yukawa coupling strength modifier in the Kappa formalism, while α describes the admixture of CP-even and CP-odd components in this coupling. The solid lines depict the 68% CL limits, while the dashed lines indicate the 95% CL limits on $\kappa_t \cos\alpha$ and $\kappa_t \sin\alpha$ from the $t\bar{t}t\bar{t}$ observation, Higgs CP property measurements with $H \to \gamma\gamma$, and Higgs CP property measurements with $t\bar{t}H$, $H \to b\bar{b}$. The '+' markers denote the observed best-fit values from these three measurements, while the '×' marker represents the SM expectation ($\kappa_t = 1, \alpha = 0$). In the $H \to \gamma\gamma$ measurement, the rates of the ggF and $H \to \gamma\gamma$ processes are constrained by results from combined Higgs boson coupling measurements. All other couplings in these measurements are fixed to the values predicted by the SM. This plot was modified in April 2024 to integrate updated $t\bar{t}t\bar{t}$ results. Status of figure: April 2024. [98]

Chapter 11

Higgs Total Width Constraint from $t\bar{t}t\bar{t}$ Production

As described in chapter 2, on-shell measurements of Higgs boson cross-sections require an assumption about the Higgs total width. Combination with off-shell measurements, which do not make the same assumption, gives an indirect probe of the Higgs width, which is not directly accessible at the LHC. In analogy to the constraints from Higgs-mediated production of W^+W^- and ZZ pairs [89, 87, 88, 268], the off-shell constraint on the top-Yukawa coupling in $t\bar{t}t\bar{t}$ production can be used in combination with top-Yukawa measurements in on-shell $t\bar{t}H$ and tH production to probe the Higgs total width. Unlike the VV measurements, where the production and decay vertices are the same in the on- and off-shell measurements, the $t\bar{t}H$ and tH measurements feature different decays since the decay into two off-shell top quarks is extremely suppressed. To provide a constraint without outside assumptions about the Higgs couplings to other particles, the $t\bar{t}t\bar{t}$ measurement must be combined with the full set of Higgs measurements from ATLAS to simultaneously constrain all combinations of couplings in Higgs boson production and decay. The $t\bar{t}t\bar{t}$ constraint on κ_t provides the off-shell constraint needed to access the Higgs width. This method is valid even if new physics affects the loop-induced gluon–gluon fusion process to change the running of κ_g, which is assumed to be SM-like in the VV width measurements.

11.1 Overview of the On-shell Higgs Combination

The bases of the combination are the $t\bar{t}t\bar{t}$ measurement [129], described in chapter 10 and previous chapters, and the 2022 combination of ATLAS Higgs measurements [69]. The latter consists of a full statistical combination at the likelihood level of measurements in six Higgs boson production modes and nine Higgs boson decay modes, totaling 29 independent measurements.

Each Higgs boson production mode and branching fraction is parameterized as a function of one or more coupling modifiers following the κ framework described in chapter 2. The

Target processes		$\mathcal{L}$ [fb^{-1}]	Reference
Off-shell measurement			
$pp \to t\bar{t}t\bar{t}$		140	[129]
On-shell measurement			
Production	Decay		
ggF, VBF, WH, ZH, $t\bar{t}H$, tH	$H \to \gamma\gamma$	140	[95]
$t\bar{t}H$+ tH	$H \to b\bar{b}$	140	[269]
WH, ZH	$H \to b\bar{b}$	140	[270, 271]
VBF	$H \to b\bar{b}$	126	[272]
ggF, VBF, WH + ZH, $t\bar{t}H$+ tH	$H \to ZZ$	140	[273]
ggF, VBF	$H \to W^+W^-$	140	[274]
WH, ZH	$H \to W^+W^-$	36.1	[275]
ggF, VBF, WH + ZH, $t\bar{t}H$+ tH	$H \to \tau^+\tau^-$	140	[276]
ggF + $t\bar{t}H$+ tH, VBF + WH + ZH	$H \to \mu\mu$	140	[74]
Inclusive	$H \to Z\gamma$	140	[277]

Table 11.1: Summary of on-shell and off-shell measurements used as input for the total width measurement. For the on-shell analyses, this table specifically breaks down the processes targeted by the measurement into production and decay modes. The off-shell measurement is included but not broken down in this manner. [278]

parameterizations used follow Ref. [69] and are shown in table 11.2. The signal strength for Higgs bosons with production mode i and decay mode f is then parameterized as

$$\mu_{i\to H\to f} = \mu(i)\frac{\Gamma^f}{\Gamma_h^{SM}}\frac{\Gamma_h^{SM}}{\Gamma_h} \tag{11.1}$$

where $\mu(i)$ and Γ^f/Γ_h^{SM} are parameterized as quadratic functions of the coupling modifiers κ. Thus, the signal strength is the sum of several terms quartic in the coupling modifiers and inversely proportional to the ratio of the Higgs boson's total width to the SM expectation. This ratio is written

$$R_\Gamma = \frac{\Gamma_h}{\Gamma_h^{SM}}, \tag{11.2}$$

and in Ref. [69] is expanded in terms of the coupling modifiers as in eq. (2.14) with a possible additional term accounting for BSM Higgs boson decays. For the combination with the $t\bar{t}t\bar{t}$ measurement, the assumptions of eq. (2.14) are relaxed and R_Γ is promoted to a freely floating parameter of interest.

Production	Loops	Main interference	Effective modifier	Resolved modifier
$\sigma(\text{ggF})$	Yes	t–b	κ_g^2	$1.040\,\kappa_t^2 + 0.002\,\kappa_b^2$ $- 0.038\,\kappa_t\kappa_b - 0.005\,\kappa_t\kappa_c$
$\sigma(\text{VBF})$	–	–	–	$0.733\,\kappa_W^2 + 0.267\,\kappa_Z^2$
$\sigma(qq \to ZH)$	–	–	–	κ_Z^2
$\sigma(gg \to ZH)$	Yes	t–Z	–	$2.456\,\kappa_Z^2 + 0.456\,\kappa_t^2 - 1.903\,\kappa_Z\kappa_t$ $- 0.011\,\kappa_Z\kappa_b + 0.003\,\kappa_t\kappa_b$
$\sigma(WH)$	–	–	–	κ_W^2
$\sigma(t\bar{t}H)$	–	–	–	κ_t^2
$\sigma(tWH)$	–	t–W	–	$2.909\,\kappa_t^2 + 2.310\,\kappa_W^2 - 4.220\,\kappa_t\kappa_W$
$\sigma(tHq)$	–	t–W	–	$2.633\,\kappa_t^2 + 3.578\,\kappa_W^2 - 5.211\,\kappa_t\kappa_W$
$\sigma(b\bar{b}H)$	–	–	–	κ_b^2
Partial decay width				
Γ^{bb}	–	–	–	κ_b^2
Γ^{WW}	–	–	–	κ_W^2
$\Gamma^{\tau\tau}$	–	–	–	κ_τ^2
Γ^{ZZ}	–	–	–	κ_Z^2
$\Gamma^{\gamma\gamma}$	Yes	t–W	κ_γ^2	$1.589\,\kappa_W^2 + 0.072\,\kappa_t^2 - 0.674\,\kappa_W\kappa_t$ $+0.009\,\kappa_W\kappa_\tau + 0.008\,\kappa_W\kappa_b$ $-0.002\,\kappa_t\kappa_b - 0.002\,\kappa_t\kappa_\tau$
$\Gamma^{Z\gamma}$	Yes	t–W	$\kappa_{(Z\gamma)}^2$	$1.118\,\kappa_W^2 - 0.125\,\kappa_W\kappa_t$ $+0.004\,\kappa_t^2 + 0.003\,\kappa_W\kappa_b$
$\Gamma^{\mu\mu}$	–	–	–	κ_μ^2

Table 11.2: Parameterizations of Higgs boson production cross-sections σ_i and partial decay widths Γ^f, normalized to their SM values, as functions of the coupling-strength modifiers κ. For effective κ parameters associated with loop processes, the resolved scaling in terms of the modifications of the Higgs boson couplings to the fundamental SM particles is given. The coefficients are derived following the methodology in Ref. [279].

11.2 Combination of $t\bar{t}t\bar{t}$ Observation with On-shell Higgs Measurements

The statistical combination of the $t\bar{t}t\bar{t}$ measurement with the on-shell Higgs measurements is performed at the likelihood level. The combined likelihood is constructed as the product of the likelihoods for the individual input measurements, with the parameters of interest reparameterized according to the desired interpretation. For systematic uncertainties corresponding to the same systematic effects, correlations are introduced by replacing the separate nuisance parameters of the input likelihoods with a single nuisance parameter and removing any resulting duplicate constraint terms from the combined likelihood. The resulting combined likelihood may be used with the common techniques described in chapter 6 to produce estimates and confidence intervals for the POIs. The correctness of this approach relies on the assumption that the combined likelihoods are independent, i.e., that the overlap in data entering each analysis is negligible.

Parameterization

The parameterization of on-shell Higgs boson production and decay presented in table 11.2 presents two options for the parameterization of processes occurring through loops, namely ggF production and $H \to \gamma\gamma$ and $H \to Z\gamma$ decays. The first option is to use effective coupling modifiers κ_g, κ_γ, and $\kappa_{Z\gamma}$ for the loop-induced couplings. The rare loop-induced process $gg \to ZH$ cannot be constrained by current measurements and is thus parameterized as a function of κ_t and κ_Z. This approach does not require assumptions about new physics entering the loops. In contrast to the ZZ and WW based width measurements, these loop couplings enter only in the on-shell Higgs measurements, and no assumptions are made about their running. The second option is to resolve the loops by calculating the contributions from SM particles and replacing the loop-induced couplings with parameterizations in terms of the tree-level couplings. This approach affords more sensitivity at the cost of model dependence, as measurements of ggF Higgs boson production and $H \to \gamma\gamma$ decay provide significant on-shell constraints on κ_t. Due to limited constraining power for second generation Yukawa couplings, equality with the third generation coupling modifiers is assumed, i.e. $\kappa_c = \kappa_t$ and $\kappa_s = \kappa_b$.

Special Considerations for $t\bar{t}t\bar{t}$ and $t\bar{t}H$ Processes

Because $t\bar{t}t\bar{t}$ and $t\bar{t}H$ both result in large multiplicity final states with similar characteristics, many measurements of one feature the other as a prominent background. In the $t\bar{t}t\bar{t}$ measurement likelihood, the $t\bar{t}H$ normalization is floated independently of κ_t as in the $t\bar{t}H$ profiled scenario presented in chapter 10. This approach is preferred because the composition of Higgs boson decay modes for the $t\bar{t}H$ background in the $t\bar{t}t\bar{t}$ signal region is not known, which precludes a parameterization like that of eq. (11.1). The impact of this choice

is evaluated in two separate tests using the $t\bar{t}H$ parameterized approach with the assumption that all of the $t\bar{t}H$ background arises from either $H \to W^+W^-$ or $H \to \tau^+\tau^-$ decays. These decay modes are expected to dominate in the SSML region. The expected limit on R_Γ with loop processes profiled improves by 1.8% in both scenarios, while the limit with loops resolved changes by less than 1%. In the $t\bar{t}H$, $H \to b\bar{b}$ measurement, the normalization of $t\bar{t}t\bar{t}$ production is fixed to the SM value. Correlating the $t\bar{t}t\bar{t}$ cross-section in the $t\bar{t}H$ $H \to b\bar{b}$ measurement with that of the $t\bar{t}t\bar{t}$ measurement assuming the cross-section dependence [46] has an impact of less than 1% on the expected limit on R_Γ. This correlation is not included because it does not account for possible acceptance and shape variations of the $t\bar{t}t\bar{t}$ process with κ_t in the signal regions of the $t\bar{t}H$, $H \to b\bar{b}$ measurement.

Overlap

The overlap between on-shell Higgs measurements has been checked for the combination [69] with overlaps up to a few percent, largely concentrated in control regions. The impact of these overlaps has been deemed negligible. Overlap between these measurements and the $t\bar{t}t\bar{t}$ measurement is expected to be minimal due to the same-sign lepton requirement, which is orthogonal to the lepton selections of all on-shell Higgs measurements except for the measurement of $t\bar{t}H$ production in the multilepton final state [135]. Overlap with the other on-shell measurements can only occur through the small differences in the lepton reconstruction requirements used in each measurement. The $t\bar{t}H$ multilepton measurement is excluded from the combination to avoid complications from overlapping signal regions. Removing the $t\bar{t}H$ multilepton measurement from the on-shell combination changes the reported value of $\kappa_t = 0.94 \pm 0.11$ [69] with loop processes profiled to $\kappa_t = 0.86 \pm 0.13$. When loop processes are resolved, the change is from a reported value of $\kappa_t = 0.95 \pm 0.07$ to $\kappa_t = 0.94 \pm 0.07$. The decreased impact is because of the dominant role of the loop-induced ggF and $H \to \gamma\gamma$ processes in constraining κ_t in this scenario. In both scenarios, the maximum impact on the central value or uncertainty of coupling modifiers other than κ_t is 0.02, with most results changing by less than 0.01.

Correlation of Systematic Uncertainties

Systematic uncertainties from different measurements that parameterize the same physical effects are given correlated nuisance parameters wherever possible, following the methods used for the on-shell Higgs combination. Uncertainties in the $t\bar{t}t\bar{t}$ measurement overlapping with the on-shell Higgs measurements are incorporated into this correlation scheme. When a group of uncertainties intended to cover the same effects is derived via different methods, such as when the calibration procedure changes, the meanings of individual nuisance parameters cannot be matched one-to-one, and the resulting systematic uncertainties are not correlated. In cases where these systematic uncertainties contributed significantly to one or both measurements, possible correlation schemes are tested to evaluate the impact on the expected results. In the following description, the notation (x) indicates how many nuisance parame-

ters x are correlated. Nuisance parameters relating to pileup reweighting (1), the soft term of the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement (3), jet vertex tagger (1), electromagnetic calorimeter energy scale and resolution (2), electron reconstruction (4), muon reconstruction (8), are fully correlated between the on-shell Higgs measurements and the $t\bar{t}t\bar{t}$ measurement. Nuisance parameters relating to the jet energy scale (27), jet energy resolution (14), are correlated between the $t\bar{t}t\bar{t}$ measurement and those components of the on-shell Higgs combination which use the same uncertainty scheme for the respective systematic effects. The $t\bar{t}t\bar{t}$ measurement uses a more complete set of flavor tagging uncertainties than any of the on-shell Higgs measurements due to the crucial role of flavor tagging in the analysis and a different flavor tagging algorithm than many of the earlier measurements. Thus, no correlations are introduced for the flavor tagging uncertainties.

Systematic uncertainties on theoretical predictions are also correlated when the systematic variations are evaluated the same way across measurements. For cross-section or other normalization uncertainties, this is true when the normalized distributions correspond to the same physics process(es). Generator or parton shower uncertainties based on a two-point comparison are correlated when the nominal and alternate MC generators are the same between the two analyses. The correlation of theoretical systematic uncertainties is most significant for uncertainties in $t\bar{t}H$ production, which is prominent in both the on-shell and $t\bar{t}t\bar{t}$ measurements. The $t\bar{t}H$ generator and parton shower uncertainties are all correlated, while the treatment of the overall $t\bar{t}H$ normalization uncertainties is irrelevant due to the floating $t\bar{t}H$ normalization in the $t\bar{t}H$ profiled interpretation of the $t\bar{t}t\bar{t}$ measurement. Likewise, $t\bar{t}Z$ appears as a background in measurements of $t\bar{t}H$ and $t\bar{t}t\bar{t}$ and the $t\bar{t}Z$ generator, PDF, and cross-section uncertainties are correlated. The $t\bar{t}t\bar{t}$ cross-section uncertainty is correlated and applied as a normalization uncertainty to both the κ_t parameterized yields in the $t\bar{t}t\bar{t}$ measurement and the SM $t\bar{t}t\bar{t}$ yield in the $t\bar{t}H$, $H \to b\bar{b}$ measurement. $t\bar{t}W$ is modeled with different MC generators across measurements and given a floating normalization in the $t\bar{t}t\bar{t}$ measurement, so no $t\bar{t}W$ modeling uncertainties are correlated.

Of the correlated theoretical uncertainties, only the $t\bar{t}H$ parton showering uncertainty contributes significantly to the overall uncertainty in both the $t\bar{t}t\bar{t}$ measurement and at least one (namely $t\bar{t}H$, $H \to b\bar{b}$) on-shell measurement. The correlation of this uncertainty is found to have a 1% impact on the expected limit on R_Γ. The second most relevant uncertainty is the $t\bar{t}t\bar{t}$ cross-section uncertainty, which has a 0.7% impact from correlation. All other correlated theoretical uncertainties have an impact below 0.05% from correlation.

The on-shell Higgs combination [69] used a preliminary calibration of the Run 2 luminosity that measured 139 fb^{-1} while the $t\bar{t}t\bar{t}$ measurement uses the updated luminosity measurement [125] of 140 fb^{-1}. For this combination, all on-shell measurements are updated to the new luminosity determination, and the luminosity uncertainty is adjusted to the more precise value. The luminosity uncertainty is decomposed into four components from different sources, each correlated differently for measurements based on a partial Run 2 dataset.

The impact of systematic uncertainties on the 95% CL limit on Γ_h is shown in table 11.3. The impact is evaluated as the change in the limit when removing the uncertainties while fixing the underlying nuisance parameter(s) to their best fit values. The systematic uncer-

Systematic uncertainty	Impact on 95% CL upper limit on Γ_h	
	Expected [%]	Observed [%]
Theory	37	33
$t\bar{t}t\bar{t}$ production	25	13
Higgs boson production/decay	5	6
Other processes	10	16
Experimental	2	2
Jet flavor tagging	2	1
Jet and missing transverse energy	< 1	< 1
Leptons and photons	< 1	< 1
All other systematic uncertainties	< 1	< 1

Table 11.3: Impact of the main sources of systematic uncertainties in the expected and observed 95% CL upper limit on the total width of the Higgs boson Γ_h when loop-induced couplings are profiled. The impact is quantified as the reduction of the upper limit when the corresponding systematic uncertainties are removed from consideration by fixing the associated nuisance parameters at the best-fit values. The impacts shown here are not a breakdown of uncertainties, and the sum in quadrature of individual impact values is not expected to equal the total impact value. [278]

tainty in the measurement is dominated by theoretical uncertainties. When evaluated on the observed rather than expected data, the importance of $t\bar{t}t\bar{t}$ modeling uncertainties is reduced while the impact of background modeling uncertainties, dominated by $t\bar{t}+b$ backgrounds in the $t\bar{t}H$, $H \to b\bar{b}$ measurement, increases. This is due to the non-linear relationship between the $t\bar{t}H$ and $t\bar{t}t\bar{t}$ cross-sections and the parameter of interest Γ_h. These trends are expected from a simplified description of the likelihood model in which $\mu_{t\bar{t}t\bar{t}} \propto \kappa_t^4$ and $\mu_{t\bar{t}H} \propto \kappa_t^2/R_\Gamma$. In this model, the impact of uncertainties affecting $\mu_{t\bar{t}H}$ increases when $\mu_{t\bar{t}H} < 1$ and the impact of uncertainties on $\mu_{t\bar{t}t\bar{t}}$ decreases when $\kappa_t > 1$. Table 11.4 shows the same impacts of systematic uncertainties for the case where the loop-induced couplings are profiled. The impacts are generally reduced, and the increase in the impact of $t\bar{t}+b$ uncertainties is not present because the $t\bar{t}H$measurement no longer provides the dominant on-shell measurement of κ_t.

11.3 Constraints on the Higgs Width

One-dimensional constraints on Γ_h and R_Γ derived from a profile likelihood test statistic on the combined likelihood function. Additionally, two-dimensional constraints in the Γ_h, κ_t plane are derived to evaluate the parameter space allowed by the on- and off-shell measurements. The expected results are derived from an Asimov dataset constructed with nuisance parameters derived from a fit to data with the normalizations of all Higgs boson production

Systematic uncertainty	Impact on 95% CL upper limit on Γ_h	
	Expected [%]	Observed [%]
Theory	34	31
$t\bar{t}t\bar{t}$ production	29	25
Higgs boson production/decay	3	4
Other processes	3	3
Experimental	2	2
Jet flavor tagging	2	1
Jet and missing transverse energy	< 1	< 1
Leptons and photons	< 1	< 1
All other systematic uncertainties	< 1	< 1

Table 11.4: Impact of the main sources of systematic uncertainties in the expected and observed 95% CL upper limit on the total width of the Higgs boson Γ_h when loop-induced couplings are resolved. The impact is quantified as the reduction of the upper limit when the corresponding systematic uncertainties are removed from consideration by fixing the associated nuisance parameters at the best-fit values. The impacts shown here are not a breakdown of uncertainties, and the sum in quadrature of individual impact values is not expected to equal the total impact value. [278]

modes and $t\bar{t}t\bar{t}$ production fixed to the SM expectation.

Profiled Loop Couplings

When the loop couplings are profiled, the best fit value of the total width is $\Gamma_h = 86^{+110}_{-49}$ MeV, and the observed (expected) 95% CL upper limit on Γ_h is 450 MeV (75 MeV). The observed limit is pushed upwards by both the observed excess of $t\bar{t}t\bar{t}$ signal over the SM expectation and the observed deficit of $t\bar{t}H$ production in the on-shell measurements, as expected from eq. (2.17). The observed and expected limits correspond to 110 and 18 times the SM expectation, respectively. The likelihood shape shown in fig. 11.1 shows a 2σ tension with the SM expectation, which is slightly higher than the 1.8σ discrepancy observed in the $t\bar{t}t\bar{t}$ signal due to the addition of the deficit in $t\bar{t}H$. With no systematic uncertainties, the observed (expected) limit would become 280 MeV (44 MeV).

The best fit value of the coupling modifiers with loop couplings profiled are shown in table 11.5. The values are compared to the results from the on-shell Higgs combination [69] with and without the inclusion of the $t\bar{t}H$ multilepton measurement [135], which is excluded from the combination with the $t\bar{t}t\bar{t}$ measurement. Equation (2.14) provides an estimate of the Higgs boson total width purely from the on-shell measurements. This estimate relies on the assumption that the Higgs boson has neither large couplings to BSM particles nor large unobserved couplings to first and second generation fermions. These assumptions are

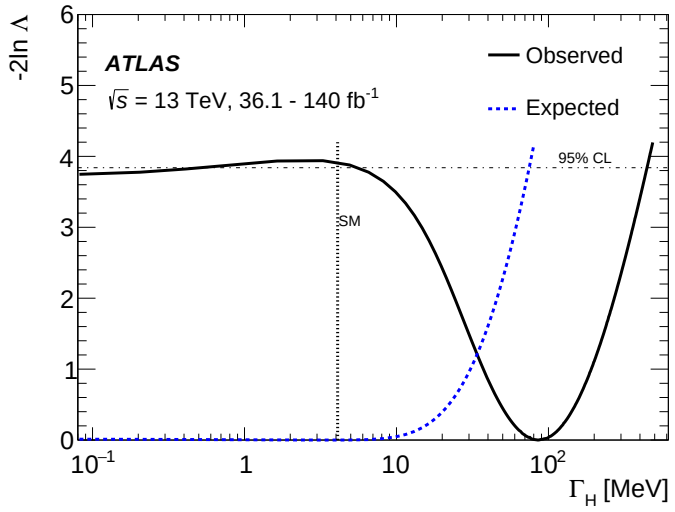


Figure 11.1: The observed (expected) profile likelihood ratio test statistic, $-2\ln\Lambda$, as a function of Γ_h is shown as a solid (dashed) line. The Higgs-top Yukawa coupling strength κ_t and other Higgs boson coupling strength parameters are profiled. The 95% CL interval is indicated by the intersections of the horizontal line with the $-2\ln\Lambda$ curves. The vertical line indicates the SM prediction. [278]

relaxed in the combination with the off-shell measurement, which allows R_Γ to be a free parameter independent of the coupling modifiers. The combination with $t\bar{t}t\bar{t}$ results in most coupling modifiers taking values around two and R_Γ taking a much larger value than the value obtained using the on-shell measurements with eq. (2.14). The consistency of these results is understood by considering the quantity $\kappa/R_\Gamma^{1/4}$, which is equal to the on-shell signal strength in eq. (2.13). These values closely match the results of the on-shell combination, indicating that the combination with $t\bar{t}t\bar{t}$ pulls all coupling modifiers up together in a way that is compensated by the increase in R_Γ to keep the on-shell signal strengths consistent with their measured values. The increase in R_Γ accommodates the observed excess in the $t\bar{t}t\bar{t}$ signal.

The two-dimensional exclusion contours in the Γ_h, κ_t plane are shown in fig. 11.2. With Γ_h promoted to a free parameter, κ_t cannot be constrained by the $t\bar{t}H$ measurements alone, so the constraint on κ_t instead comes from off-shell Higgs boson production in the $t\bar{t}t\bar{t}$ process, and the best fit value $\kappa_t = 1.9$ corresponds to the result presented in chapter 10 based on the $t\bar{t}t\bar{t}$ measurement. The best fit value of Γ_h follows from the scaling relation $R_\Gamma \sim \kappa_t^4$ as described in chapter 2.

	Full On-Shell	On-Shell Without $t\bar{t}H$ ML		On-Shell+$t\bar{t}t\bar{t}$	
POI	κ	κ	$\kappa/R_\Gamma^{1/4}$	κ	$\kappa/R_\Gamma^{1/4}$
R_Γ	0.88	0.90		21 ± 14	
κ_t	$0.94^{+0.11}_{-0.11}$	$0.86^{+0.13}_{-0.13}$	0.88	1.88 ± 0.24	0.88
κ_W	$1.05^{+0.06}_{-0.06}$	$1.05^{+0.06}_{-0.06}$	1.08	2.31 ± 0.40	1.08
κ_Z	$0.99^{+0.06}_{-0.06}$	$1.00^{+0.05}_{-0.06}$	1.03	2.18 ± 0.39	1.02
κ_b	$0.89^{+0.11}_{-0.11}$	$0.91^{+0.12}_{-0.11}$	0.93	2.00 ± 0.39	0.93
κ_τ	$0.93^{+0.07}_{-0.07}$	$0.93^{+0.07}_{-0.07}$	0.95	2.04 ± 0.37	0.95
κ_μ	$1.06^{+0.25}_{-0.30}$	$1.07^{+0.25}_{-0.30}$	1.10	2.36 ± 0.72	1.10
κ_γ	$1.01^{+0.06}_{-0.06}$	$1.02^{+0.06}_{-0.06}$	1.05	2.23 ± 0.40	1.04
κ_g	$0.95^{+0.07}_{-0.07}$	$0.96^{+0.08}_{-0.07}$	0.99	2.10 ± 0.37	0.98
$\kappa_{Z\gamma}$	$1.38^{+0.31}_{-0.37}$	$1.39^{+0.31}_{-0.37}$	1.43	3.04 ± 0.89	1.42

Table 11.5: Comparison of the best fit values for the κ modifiers with loop couplings profiled between the on-shell combination and the combination with off-shell production in $t\bar{t}t\bar{t}$. The last column shows the κs scaled by $R_\Gamma^{-1/4}$, which gives the effective κ that would be seen by the on-shell combination given the assumption of the SM Higgs width. These numbers are all consistent with the nature results within the uncertainties. The width values from the Nature fit are provided for reference and have been extracted using the parameterization in eq. (2.14).

Resolved Loop Couplings

When the loop couplings are resolved, an improved upper limit is placed because of the contribution of loop-induced processes to the on-shell κ_t measurement. The observed (expected) limit in this scenario is 160 MeV (55 MeV). The likelihood curve is shown in fig. 11.3, and the two-dimensional exclusion contours are shown in fig. 11.4.

The results follow the same pattern as when the loop couplings are profiled, but the best fit value of Γ_h is slightly lower as the preferred value of $\kappa_t/R_\Gamma^{1/4}$ is closer to one. This is due to the influence of κ_g and κ_γ, which replace the $t\bar{t}H$ measurements as the dominant on-shell constraints of the top-Higgs coupling. Additionally, the two-dimensional constraint is narrower in the direction constrained by the on-shell measurements due to the precision added by resolving the loop couplings and the flat direction $\kappa_t^4 = R_\Gamma$, which is only constrained by the $t\bar{t}t\bar{t}$ measurement, is emphasized. The comparison of the best fit coupling modifiers in table 11.6 also follows a familiar pattern and shows consistency with the on-shell

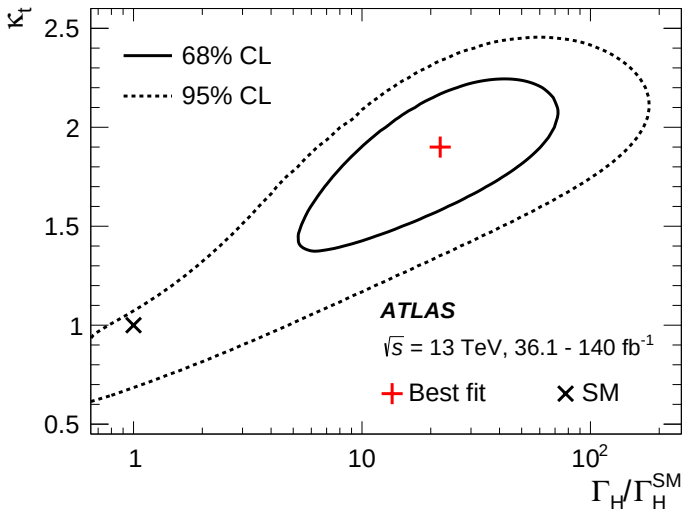


Figure 11.2: The 68% CL (solid line) and 95% CL (dashed line) contours for a simultaneous measurement of Γ_h normalized to the SM prediction and κ_t. The best-fit value and the SM prediction are also indicated in the figure. [278]

measurements.

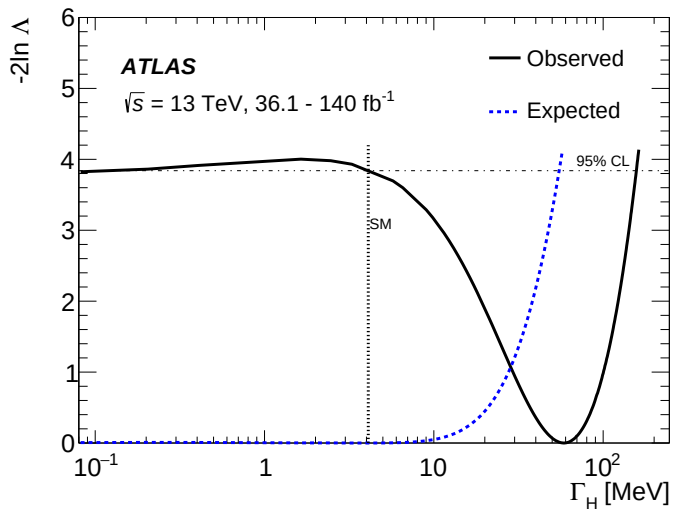


Figure 11.3: The observed (expected) profile likelihood ratio, $-2\ln\Lambda$, as a function of Γ_h is shown as a solid (dashed) line. The Higgs-top Yukawa coupling strength κ_t and other Higgs boson coupling strength parameters are profiled. Loop-induced couplings are parameterized as a function of tree-level couplings. The 95% CL interval is indicated by the intersections of the horizontal line with the $-2\ln\Lambda$ curves. The vertical line indicates the SM prediction.

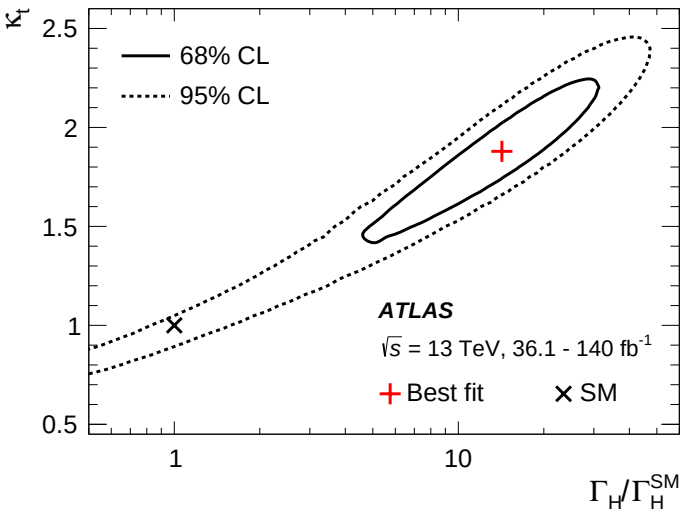


Figure 11.4: The 68% CL (solid line) and 95% CL (dashed line) contours for a simultaneous measurement of Γ_h normalized to the SM prediction and κ_t. Loop-induced couplings are parameterized as a function of tree-level couplings. The best-fit value and the SM prediction are also indicated in the figure.

	Full On-Shell	On-Shell Without $t\bar{t}H$ ML		On-Shell+$t\bar{t}t\bar{t}$	
POI	κ	κ	$\kappa/R_\Gamma^{1/4}$	κ	$\kappa/R_\Gamma^{1/4}$
R_Γ	0.89	0.89		14 ± 7	
κ_t	$0.95^{+0.07}_{-0.07}$	$0.94^{+0.07}_{-0.07}$	0.97	1.88 ± 0.23	0.97
κ_W	$1.02^{+0.05}_{-0.05}$	$1.02^{+0.05}_{-0.05}$	1.05	2.05 ± 0.28	1.06
κ_Z	$0.99^{+0.06}_{-0.05}$	$0.99^{+0.06}_{-0.05}$	1.02	2.00 ± 0.29	1.03
κ_b	$0.90^{+0.11}_{-0.11}$	$0.90^{+0.11}_{-0.11}$	0.93	1.81 ± 0.27	0.94
κ_τ	$0.94^{+0.07}_{-0.07}$	$0.93^{+0.07}_{-0.07}$	0.96	1.87 ± 0.28	0.97
κ_μ	$1.07^{+0.25}_{-0.31}$	$1.08^{+0.25}_{-0.31}$	1.11	2.17 ± 0.61	1.12

Table 11.6: Comparison of the best fit values for the κ modifiers with loop couplings resolved between the on-shell combination and the combination with off-shell production in $t\bar{t}t\bar{t}$. The last column shows the κs scaled by $R_\Gamma^{-1/4}$, which gives the effective κ that would be seen by the on-shell combination given the assumption of the SM Higgs width. These numbers are all consistent with the nature results within the uncertainties. The width values from the Nature fit are provided for reference and have been extracted using the parameterization in eq. (2.14).

Chapter 12

Conclusions and Outlook

A combination of the large Run 2 dataset with 140 fb^{-1} of collisions at $\sqrt{s} = 13$ TeV recorded by ATLAS and significantly improved object reconstruction and tagging and applications of deep learning techniques to data analysis have drastically increased the sensitivity to $t\bar{t}t\bar{t}$ production and allowed the first observation of $t\bar{t}t\bar{t}$ production with the ATLAS detector. The observation is performed in the same-sign dilepton and multilepton channels, taking advantage of the large background rejection offered in final states with same-sign leptons. Re-optimized object selection, improved background modeling, and the implementation of a powerful GNN classifier for background rejection contribute to the sensitivity increase. The measured $t\bar{t}t\bar{t}$ cross-section is $\sigma_{t\bar{t}t\bar{t}} = 22^{+6.6}_{-5.5}$ fb and the observed (expected) statistical significance of the signal is 6.1σ (4.3σ). The signal strength is measured to be $\mu_{t\bar{t}t\bar{t}} = 1.9^{+0.8}_{-0.5}$ and is compatible with the Standard Model expectation $\mu = 1$ at the $1.7 - 1.8\sigma$ level, depending on which calculation of the Standard Model $t\bar{t}t\bar{t}$ production cross-section is used. Alternative interpretations indicate that the present analysis is also consistent with an enhancement in the small $t\bar{t}t$ background, which is not well separated from $t\bar{t}t\bar{t}$ events.

Interpretations of the $t\bar{t}t\bar{t}$ signal are used to place limits on BSM modifications to $t\bar{t}t\bar{t}$ production from four-quark operators in the SMEFT context, from the Higgs oblique parameter, or from deviations of the top Yukawa coupling from the SM prediction. In the latter case, a 95% CL limit of $|\kappa_t| < 1.9$ is placed when also considering the effects of varying κ_t on the $t\bar{t}H$ background and a limit of $|\kappa_t| < 2.3$ is placed when de-correlating the effects in $t\bar{t}t\bar{t}$ and $t\bar{t}H$ production. Two-dimensional constraints are placed on the top Higgs coupling strength and CP mixing angle when the coupling is allowed to vary from a pure CP-even interaction. Compared to limits derived from on-shell ggF and $t\bar{t}H$ Higgs boson production, these limits are less stringent but less dependent on assumptions about Higgs boson decay branching fractions and the impact of BSM particles on loop-induced couplings. Furthermore, the direct measurement of κ_t in the off-shell regime complements the on-shell measurement in that any deviations may hint at BSM effects in the running of the top Yukawa coupling.

The measurement of κ_t using the off-shell Higgs boson contribution to $t\bar{t}t\bar{t}$ production is combined with measurements of on-shell Higgs boson production and decays to produce the first ever constraint on the total width of the Higgs boson based on measurements of

κ_t. The observed (expected) 95% CL upper limit is $\Gamma_h < 450$ (75) MeV when profiling the loop-induced Higgs boson production and decay modes and $\Gamma_h < 160$ (55) MeV when parameterizing the loop-induced Higgs boson production and decay modes as a function of tree-level couplings to Standard Model particles. Compared to limits derived in the $ZZ^{(*)}$ and $WW^{(*)}$ channels, these are less stringent but do not rely on assumptions about the running of the loop-induced ggF vertex from the on-shell scale m_h to the off-shell energy scale.

Analysis of the Run 2 dataset has brought much higher precision to the measurement of the Higgs boson than expected because of significant advances in analysis techniques, especially the incorporation of sophisticated machine learning techniques into many aspects of data analysis. In Run 3, the LHC has already delivered a higher integrated luminosity than in Run 2 at an increased center of mass energy of $\sqrt{s} = 13.6$ TeV, and the subsequent upgrade to the high luminosity LHC (HL-LHC) will result in a total integrated luminosity of up to 3000 fb^{-1}. With a ten-fold improvement in the size of the dataset, continued innovation in techniques for analysis and reconstruction, and advancements in the theoretical modeling of Higgs boson production in proton–proton collisions, measurements of the top Yukawa coupling may reach a precision of a few percent by the end of the LHC program around 2041. At the same time, the possibility of observing tH production, the production of Higgs boson pairs, and rare Higgs boson decays into muons or $Z\gamma$ all promise a significant improvement in the understanding of the Higgs boson and numerous opportunities to discover hints of physics beyond the Standard Model.

Appendix A

CP Measurements of the Top Yukawa Coupling in $t\bar{t}H$ and tH

The role of measurements of $t\bar{t}H$ and tH production in constraining the top-Higgs coupling has been established in chapter 2, and the possibility of combining such measurements with measurements of $t\bar{t}t\bar{t}$ production to constrain the total width of the Higgs boson has been explored in chapter 11. This appendix presents two such measurements to which the author of this thesis has contributed. Both measure the CP properties of the top-Higgs coupling by using both the cross-section variation and kinematic observables of $t\bar{t}H$ and tH production. The first measurement targets $t\bar{t}H$ production in the $H \to \gamma\gamma$ decay mode [105]. The author's contribution focused on the optimization of the BDT boundaries used to define the analysis regions and the visualization of the BDT distributions and top quark and Higgs boson mass peaks. The second targets tH production in the $H \to b\bar{b}$ decay mode with a semi-boosted topology with enhanced sensitivity to CP-odd couplings and is still in preparation. The author's contribution focused on the development and training of the GNN classifiers and the development of a data-driven background estimation strategy.

A.1 $t\bar{t}H(H \to \gamma\gamma)$ Measurement

The ATLAS Run 2 measurement of the CP properties of the top-Higgs interaction in the $t\bar{t}H$ final state with $H \to \gamma\gamma$ [105] uses the full Run 2 dataset of 140 fb^{-1} and is based on single and di-photon triggers [134]. The analysis separates events into leptonic and fully hadronic categories based on the presence of a light lepton. Hadronic top-quark decays are reconstructed using a BDT trained to identify triplets of jets associated with the same top-quark decay, including the correct assignment of two jets as the products of a W boson decay. The BDT takes as input the mass of the resulting top quark candidate and kinematic variables describing the W boson and b-jet candidates. In the leptonic channel, a leptonic W boson decay is reconstructed from the lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ using m_W and m_t as constraints to determine the longitudinal momentum of the neutrino. Then, a top quark candidate is

reconstructed by matching the W with the jet that gives the highest top reconstruction BDT score. In the hadronic channel, the triplet with the largest top reconstruction BDT score is taken. In both channels, a second top candidate is reconstructed from the remaining jets if there are at least two.

A two-dimensional MVA strategy is used to separate $t\bar{t}H$ and tH events from background processes from continuum di-photon production, and to separate $t\bar{t}H$ and tH events originating from the CP-even top-Higgs vertex of the SM and those originating from the CP-odd vertex with coupling $\tilde{\kappa}_t$. A background rejection BDT [280] is trained to separate $t\bar{t}H$ events from $\gamma\gamma$+jets and $t\bar{t}\gamma\gamma$ backgrounds while a CP sensitive BDT is trained to separate $t\bar{t}H$ and tH events generated in the SM case $\kappa, \alpha = (1, 0°)$ from those generated in the case of a CP-odd coupling $\kappa, \alpha = (1, 90°)$. The CP BDT uses the p_T and η of the reconstructed top quark Higgs boson candidates together with the azimuthal angles ϕ of the top quarks measured relative to the Higgs candidate, the angular separation $\Delta\eta$, $\Delta\phi$ between top candidates, the invariant mass of the top-Higgs system, and the top reco BDT scores. In addition to these variables describing the reconstructed top and Higgs candidates, event-level variables describing the number of b-jets, the significance of the measured $E_{\mathrm{T}}^{\mathrm{miss}}$ and overall activity are included. The background rejection and CP BDTs are trained separately in the hadronic and leptonic channels, and the two-dimensional BDT space, shown in figs. A.1 and A.2, is divided into a total of 20 categories as sketched in fig. A.3.

The region boundaries are chosen by an iterative scan over possible values, which first fixes the four (three) boundaries of the background rejection BDT in the hadronic (leptonic) channel and then selects up to two boundaries in the CP BDT score for each region of the background rejection BDT. The boundary optimization uses two figures of merit, $Z_{t\bar{t}H+tH}$ and $Z_{CP}(90)$, defined in eq. (A.1), which estimate the expected discovery sensitivity to the combined $t\bar{t}H$ and tH production and to the CP-odd coupling scenario, respectively.

$$\begin{aligned} Z_{t\bar{t}H+tH} &= \sqrt{2(S_0 + B)\ln\frac{S+B}{B} - 2S} \\ Z_{CP}(90) &= \sqrt{2(S_0 + B)\ln\frac{S_0+B}{S_{90}+B} + 2(S_{90} - S_0)} \end{aligned} \tag{A.1}$$

S_0, S_{90}, and B are the yields of $t\bar{t}H + tH$ in the SM scenario and in the CP-odd scenario, and the background expectation, respectively. The figures of merit are evaluated under the Higgs boson mass peak $m_{\gamma\gamma} \in [123, 127]$ GeV and describe the sensitivity of a single-bin number counting analysis performed in this region. The Z scores derived from different categories are added in quadrature to estimate the total sensitivity of an analysis conducted by dividing the events into categories based on their SB and CP BDT scores. These figures of merit are not optimized by the same choice of boundaries because they measure the sensitivity to different signal hypotheses against different background hypotheses. The final bin boundaries are chosen to maximize $Z_{t\bar{t}H+tH}$ subject to the requirement that $Z_{CP}(90)$ is within 0.15σ of its maximum value. Additionally, each region is required to have at

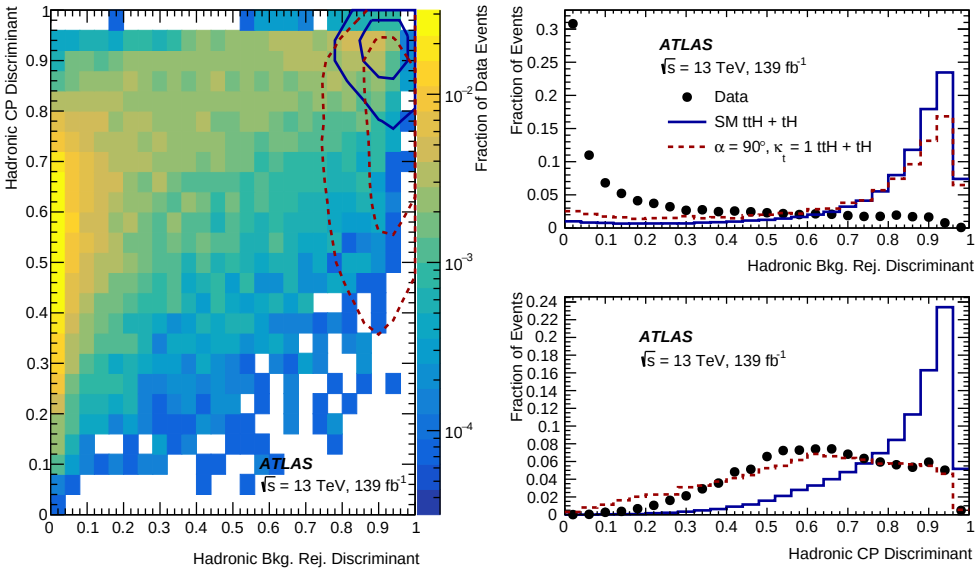


Figure A.1: (Left) Two-dimensional BDT distribution in the selected data events ($m_\gamma\gamma$ in [105, 160] GeV) from the Had region showing the Background Rejection BDT and CP BDT. The inner (outer) contours capture 25% (50%) of the $t\bar{t}H$ and tH signal events for CP-even (blue) and CP-odd (red) hypotheses. (Right) Projections onto the background rejection and CP BDT axes. Contributions from CP-even (blue) and CP-odd (red) $t\bar{t}H/tH$ processes and the data (black) are shown and normalized to unit area. The error bars on the data are statistical. [281]

least 0.8 background events in the Higgs mass window to ensure that the background-only fit can converge. For the purposes of this requirement, the background is estimated by extrapolating the background contribution from the mass sidebands into the mass window using an extrapolation derived from events with reversed photon identification and isolation criteria.

A parametric fit to the invariant mass of the di-photon system $m_{\gamma\gamma}$ is performed in each region for events with $m_{\gamma\gamma} \in [105, 160]$ GeV. The parametric fit uses a smoothly falling function chosen empirically from a range of options to model the background and a double-sided Crystal Ball function to fit the resonant $H \to \gamma\gamma$ signal. The parameters of the signal function are fixed by a fit to simulated data. The yields of $t\bar{t}H$, $tHjb$, and tWH are parameterized separately in each category using a χ^2 fit to the yield of each process in MC signal samples generated at a range of points in the (κ, α) plane similar to the fit done in each bin of the $t\bar{t}t\bar{t}$ measurement described in chapter 6. $t\bar{t}H$ production is parameterized with the functional form

$$y_i^{t\bar{t}H}(\kappa_t, \alpha) = A_i\kappa_t^2 \cos^2\alpha + B_i\kappa_t^2 \sin^2\alpha + E_i\kappa_t^2 \cos\alpha \sin\alpha, \tag{A.2}$$

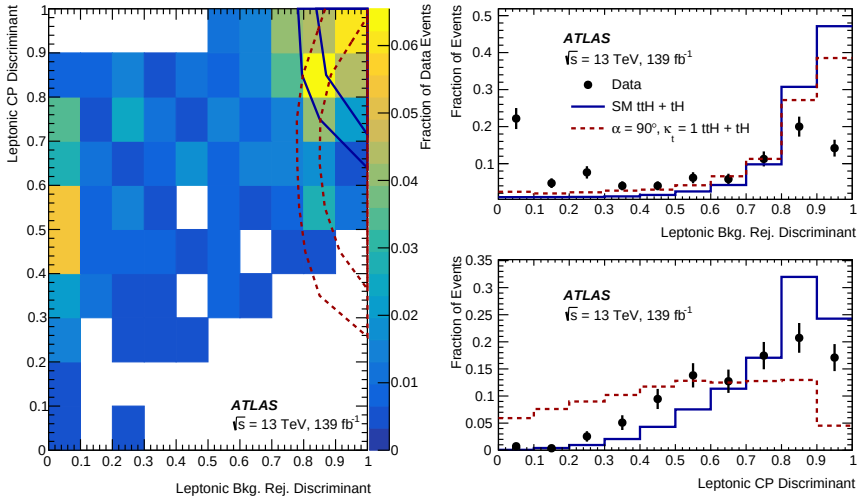
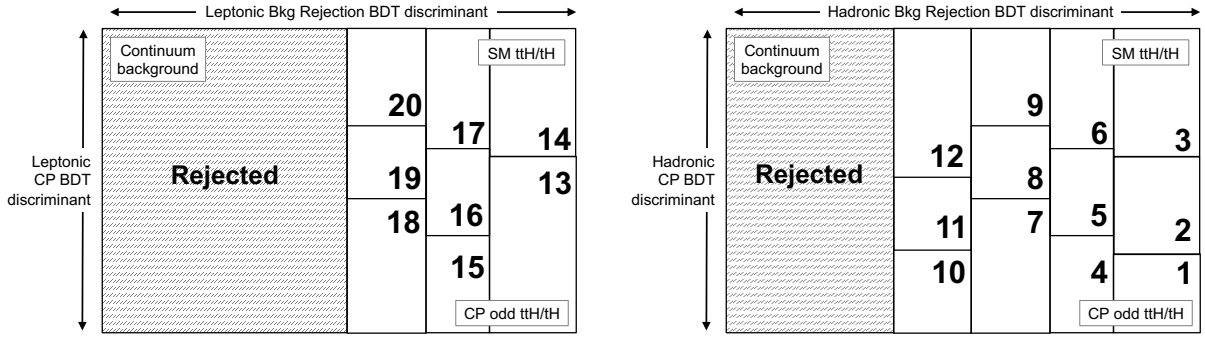


Figure A.2: (Left) two-dimensional BDT distribution in the selected data events ($m_\gamma\gamma$ in [105, 160] GeV) from the Lep region showing the Background Rejection BDT and CP BDT. The inner (outer) contours capture 25% (50%) of the $t\bar{t}H$ and tH signal events for CP-even (blue) and CP-odd (red) hypotheses. (Right) projections onto the background rejection and CP BDT axes. Contributions from CP-even (blue) and CP-odd (red) $t\bar{t}H/tH$ processes and the data (black) are shown and normalized to unit area. The error bars on the data are statistical.



(a) (b)

Figure A.3: Schematic of the leptonic (left) and hadronic (right) categories used in the BDT analysis. The backgrounds are largest on the left of the figures. The pure CP-even processes are largest on the top, and the pure CP-odd processes are largest on the bottom.

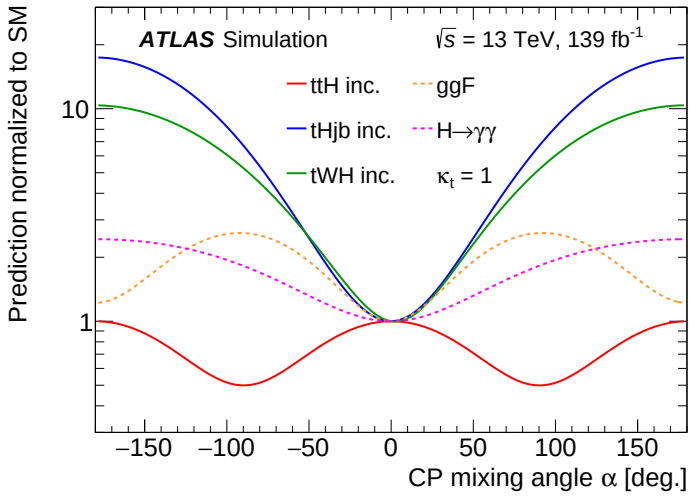


Figure A.4: Dependence of the expected rates on the mixing angle for $\kappa_t = 1$. The rates shown are normalized to the Standard Model expectation ($\alpha = 0$) for $t\bar{t}H$, $tHjb$, and tWH within the acceptance of the $t\bar{t}H(H \to \gamma\gamma)$ measurement. Also, the relative change in the branching ratio for $H \to \gamma\gamma$ and ggF production is shown. The production rates are unchanged under $\alpha \to -\alpha$ at leading-order in perturbation theory, and the tH rates are smallest at $\alpha = 0$.

where the final term proportional to E_i denotes the interference between diagrams with the CP-even and odd couplings and is negligible. This term is identically zero at leading order. The parameterizations of $tHjb$ and tWH require additional terms to account for diagrams in which the Higgs boson radiates from a W boson and interference of the top Yukawa diagrams with those. The tH processes are parameterized using the functional form

$$y_i^{tH}(\kappa_t, \alpha) = A_i\kappa_t^2 \cos^2\alpha + B_i\kappa_t^2 \sin^2\alpha + C_i\kappa_t \cos\alpha + D_i\kappa_t \sin\alpha + E_i\kappa_t^2 \cos\alpha \sin\alpha + F_i. \quad \text{(A.3)}$$

The resonant $H \to \gamma\gamma$ background from other Higgs boson production modes, predominantly ggF, is fixed to the SM prediction and subtracted from the fitted $H \to \gamma\gamma$ yield in each region to determine the $t\bar{t}H + tH$ yield. The measured cross-section times branching ratio is $\sigma_{t\bar{t}H} \times B_{\gamma\gamma} = 1.64^{+0.38}_{-0.36}(\text{stat.})^{+0.17}_{-0.14}(\text{syst.})$ corresponding to a signal strength $\mu_{t\bar{t}H} = 1.43^{+0.33}_{-0.31}(\text{stat.})^{+0.21}_{-0.15}(\text{syst.})$. The observed (expected) statistical significance of the $t\bar{t}H$ signal is 5.2σ (4.4σ). No observation of tH production is possible due to the small tH cross-section. A 95% CL upper limit of $\mu_{tH} < 12$ is placed on the combination $tHjb + tWH$. Figure A.5 shows events in the two dimensional plane of reconstructed m_h and m_t along with the one dimensional projection of both distributions. Events are weighted according to the signal-to-background ratio of the category they are assigned to to enhance the contribution from $H \to \gamma\gamma$ signal events.

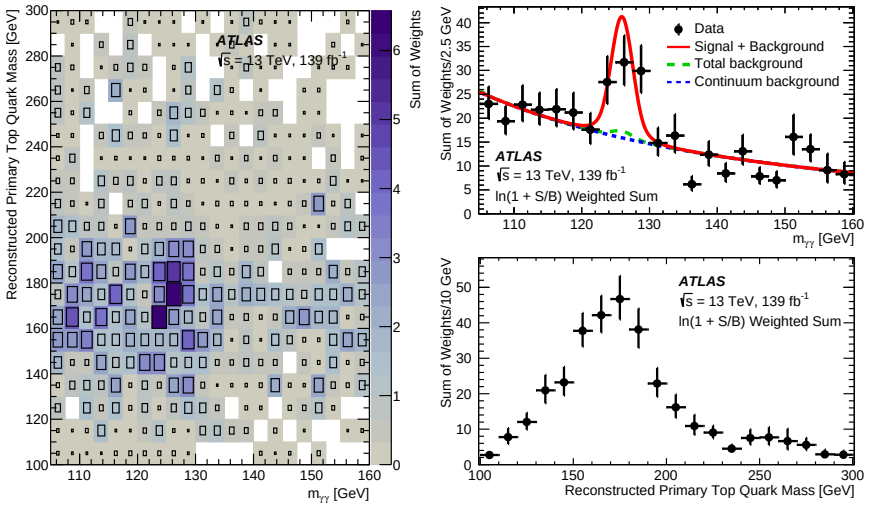


Figure A.5: Distribution of reconstructed primary top quark mass versus reconstructed Higgs boson mass in the data events. The right panels show the projections onto the Higgs boson mass and primary top quark mass axes. In the upper panel, the fitted continuum background (blue), the total background including non-$t\bar{t}H/tH$ Higgs boson production (green), and the total fitted signal plus background (red) are shown. The error bars on the data are statistical. [105]

Two-dimensional constraints are placed in the κ_t, α plane under two assumptions on the loop-induced ggF and $H \to \gamma\gamma$ processes. The first assumption uses a constraint from measurements of the ggF production cross-section and $H \to \gamma\gamma$ branching ratio. The second scenario parametrizes ggF production and the $H \to \gamma\gamma$ branching fraction as functions of $\kappa_t$and α. The two-dimensional limit contours in both scenarios are shown in Figure A.6.

When letting κ_t float freely, the observed (expected) 95% upper limit of $|\alpha| < 43°$ (63°) and the CP-odd scenario is rejected at the 3.9σ level. The $\alpha = 180°$ scenario in which the coupling is CP-even but has the opposite sign as the SM expectation is excluded at the 2.5σ level.

A.2 $tH(H \to b\bar{b})$ Measurement with a Boosted Higgs

A similar measurement of the CP properties in the top-Higgs interactions using the ATLAS Run 2 dataset and focusing on tH production with high p_T Higgs bosons decaying into $b\bar{b}$ is in preparation. The measurement aims to use a similar strategy with a two-dimensional MVA distribution constructed from ML algorithms trained to perform background rejection and to separate signals in the CP-odd and CP-even scenarios. This analysis uses graph neural networks (GNNs) similar to the background rejection GNN used for the $t\bar{t}t\bar{t}$ measurement described in chapter 9. The measurement seeks to exploit the large enhancement in

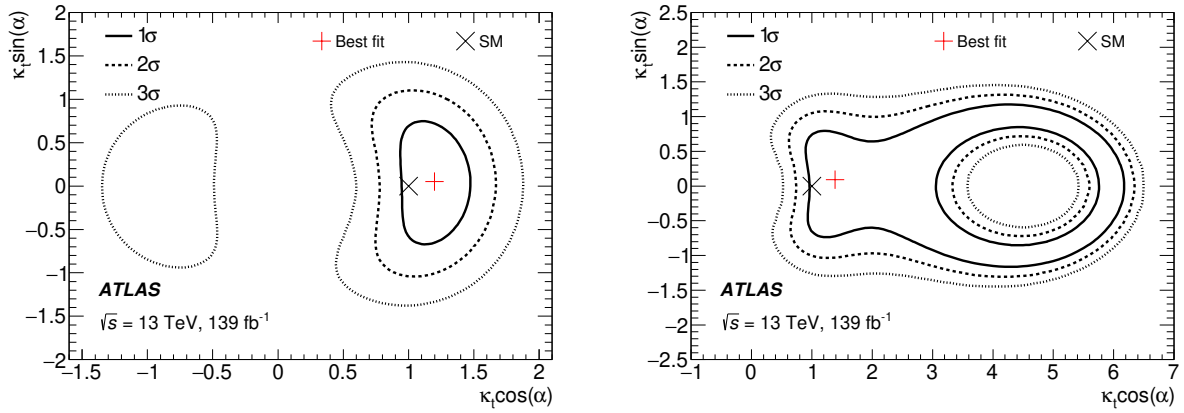


Figure A.6: Two-dimensional likelihood contours for $\kappa_t \cos\alpha$ and $\kappa_t \sin\alpha$ with ggF and $H \to \gamma\gamma$ constrained by the Higgs boson coupling combination [267] (Left) and varied as a function of κ_t and α (Right).

the tH cross-section when the CP properties of the top Yukawa coupling are altered. This enhancement, visualized in the phase space of the $t\bar{t}H(H \to \gamma\gamma)$ measurement in fig. A.4, arises because the SM features a large cancellation between tH diagrams proportional to κ_t and those proportional to κ_W. The enhancement is larger in the boosted regime, with the tH cross-section for $p_T^H > 200$ GeV attaining over eight times its SM value when the top-Higgs coupling is CP-odd.

Analysis Design

The analysis is performed in a final state with a single electron or muon, $E_{\mathrm{T}}^{\mathrm{miss}} > 50$ GeV, and exactly three b-tagged jets at the 77% working point of the DL1r b-tagger described in chapter 4. Two b-jets are expected from the decay of the Higgs boson, while the other is associated with a leptonic top-quark decay, which also produces a light lepton and a neutrino, which is reconstructed as $E_{\mathrm{T}}^{\mathrm{miss}}$. The Higgs candidate is reconstructed from the pair of b-tagged jets with the smallest angular separation ΔR_{bb}. To target the boosted regime, which is particularly sensitive to the introduction of the CP-odd term, and to reduce the large $t\bar{t}$ backgrounds, the transverse momentum of the dijet system assigned to be the Higgs candidate is required to satisfy $p_T^{bb} > 200$ GeV and $\Delta R_{bb} < 2.0$. Two GNNs are trained using the `root_gnn_dgl` package introduced in chapter 7.

The first, known as the signal/background separation (SB) GNN, plays the same role as the background rejection BDT of the $t\bar{t}H(H \to \gamma\gamma)$ CP measurement. The GNN is trained using the combination of $tHjb$ and tWH samples generated with $\kappa_t = 1$ and $\alpha = 0°, 15°, 30°, 45°, 60°, 75°, 90°$ as the target signal. This choice ensures that the SB GNN separates all signal processes from the background and allows the maximum possible training

statistics during training. The background training sample consists of simulated events from the dominant background process $t\bar{t}$ and the sub-dominant process tW, which is difficult to separate from the tH signal and becomes significant in the most signal-like regions. $t\bar{t}H$ is not included in the SB GNN training because the $t\bar{t}H$ and tH cross-sections move in opposing directions as $|\alpha|$ increases, so a large $t\bar{t}H$ contamination in the signal-like regions could reduce the sensitivity to BSM scenarios. Additionally, including a large fraction of $t\bar{t}H$ in the signal sample decreases the separation power of the targeted tH signal. The second GNN, known as the CP GNN, plays the same role as the CP BDT of the $t\bar{t}H(H \to \gamma\gamma)$ CP measurement. The CP GNN is trained with $tHjb + tWH$ events simulated with $\kappa_t, \alpha = (1, 90°)$ as the signal sample and $\kappa_t, \alpha = (0, 0°)$ as the background sample. Again, $t\bar{t}H$ is excluded from the training. In this case, because the $t\bar{t}H$ yield is much higher than tH in the SM, the GNN with $t\bar{t}H$ included in training effectively becomes a $t\bar{t}H$ versus tH discriminator rather than an observable related to the CP properties under study.

As in the $t\bar{t}H(H \to \gamma\gamma)$ CP measurement, the two-dimensional GNN output distribution will be divided into bins, the $t\bar{t}H$, $tHjb$, and tWH yields in each bin will be parameterized by eqs. (A.2) and (A.3). The stability of fit is improved by setting $E_i = 1$ for both fits and $D_i = 1$ for tH, as all of these terms are zero at leading order and are expected to have negligible contributions. A simultaneous binned fit of the parameterized signal yields and background expectation will be used to extract the best fit values and two-dimensional exclusion contours on κ_t and α. Because the analysis is still under development, details, including the background estimation strategy and GNN region definitions, are subject to further optimization before the analysis strategy is finalized and the results are unblinded.

GNN Optimization

The SB and CP GNN trainings use the same input features and GNN architecture as the $t\bar{t}t\bar{t}$ GNN. The GNNs are trained using a two-fold strategy to reduce the complexity of the training setup. As in the case of the $t\bar{t}t\bar{t}$ GNN, MC event weights are not used in training, and the ratio of the tW to $t\bar{t}$ backgrounds is controlled by fixing the number of entries of each process into each batch of the SB GNN training. During the GNN optimization procedure, candidate classifiers are evaluated using the ROC AUC to measure the separation between individual physics processes and between the combined background ($t\bar{t} + tW$) and the combined signal ($t\bar{t}H + tHjb + tWH$) under the CP-even and CP-odd hypotheses. Additionally, the sum in quadrature of the significance of a counting experiment performed in each bin is used as in eq. (A.1).

The choice of training samples is dictated by a number of factors, including the training statistics available from the MC samples and tradeoffs in the sensitivity of the analysis to different scenarios. Several scenarios are explored to ensure the best possible training setup. In the case of the SB GNN, the most important quantity to optimize is the fraction of tW background used in the training. A set of trainings scanning over possible tW fractions is performed to identify the optimal setup. The results of this scan are shown in table A.2.

Sample	Unweighted MC Events	Expected Yield
$tHjb$ $\alpha = 0°$	51,063	8.6
$tHjb$ $\alpha = 15°$	58,467	10.7
$tHjb$ $\alpha = 30°$	77,005	16.5
$tHjb$ $\alpha = 45°$	93,486	27.0
$tHjb$ $\alpha = 60°$	100,809	42.9
$tHjb$ $\alpha = 75°$	104,854	62.3
$tHjb$ $\alpha = 90°$	101,266	85.1
tWH $\alpha = 0°$	144,574	11.7
tWH $\alpha = 15°$	160,243	13.8
tWH $\alpha = 30°$	189,706	19.5
tWH $\alpha = 45°$	222,492	29.2
tWH $\alpha = 60°$	241,848	42.1
tWH $\alpha = 75°$	252,493	57.7
tWH $\alpha = 90°$	256,636	75.7
$t\bar{t}H$ $\alpha = 0°$	332,475	297.0
$t\bar{t}H$ $\alpha = 90°$	378,196	145.9
$t\bar{t}$	169,140	20,719.4
tW	8,048	935.4

Table A.1: Training statistics and expected yields for various processes relevant for training and evaluating the SB and CP GNNs. All signal samples are generated with $\kappa_t = 1$.

The central training with 4.3% tW achieves the best separation power for the overall background in the CP-odd scenario and nearly optimal separation power in the CP-even case and is therefore chosen as the nominal training setup. As expected, the separation power of the signal from the pure tW background improves dramatically with increasing fractions of tW in the training, but the separation of $t\bar{t}$ suffers as a result. For tW fractions of 2.1% and 4.3%, the SB GNN score distribution of the CP-even signal, dominated by $t\bar{t}H$, is less signal-like than that of tW as indicated by the ROC AUC values < 0.5. The tW fraction of 4.3% chosen as nominal is equal to the SM expectation for the $tW/t\bar{t}$ ratio in the analysis phase space and is also close to the ratio of the MC statistics of each process entering the training. Therefore, this choice gives the background composition closest to the physical expectation and maximizes the statistical power of the training dataset. Trainings performed at an earlier stage of the analysis optimization with a tW fraction of 0% performed significantly worse than trainings with tW included when considering the more complete background composition.

The analysis is optimized to maximize the discovery sensitivity to the CP-odd scenario.

tW Training Fraction [%]	ROC AUC			
Signal Hypothesis	$\alpha = 90°$	$\alpha = 0°$	$\alpha = 90°$	$\alpha = 0°$
Background Hypothesis	$t\bar{t} + tW$	$t\bar{t} + tW$	tW	tW
2.1% tW	0.845	0.719	0.628	0.462
4.3% tW	0.857	0.730	0.666	0.493
8.6% tW	0.855	0.731	0.683	0.513

Table A.2: ROC AUC of SB GNN trainings using different fractions of tW in the background sample. The signal hypothesis is varied between CP-even and CP-odd $t\bar{t}H + tHjb + tWH$, and the background hypothesis is varied between $t\bar{t}$ and tW. All AUCs are evaluated using the testing dataset with the full set of MC weights, and all processes are normalized to their SM cross-sections.

Signal Point	45° Training	60° Training	90° Training
15	0.530	0.541	0.550
30	0.645	0.633	0.640
45	0.677	0.670	0.681
60	0.694	0.699	0.697
75	0.689	0.685	0.697
90	0.687	0.685	0.688

Table A.3: ROC AUC of CP GNN trainings using signal points with different CP mixing angles as the signal training dataset. The signal hypothesis for the AUC calculation is varied, and the background hypothesis is fixed to SM $tHjb + tWH$. All signal points have $\kappa_t = 1$. All AUCs are evaluated using the testing dataset with the full set of MC weights, and all processes are normalized to their SM cross-sections.

However, the CP-odd scenario is ruled out at the 95% level by the $t\bar{t}H(H \to \gamma\gamma)$ measurement. An analysis optimized for sensitivity to the CP-odd scenario may not have optimal sensitivity to alternate scenarios with CP mixing because of linear interference terms in the tH parameterization. To verify the optimality of training the CP GNN with the CP-odd scenario as the signal, a set of trainings is performed with alternate CP mixing angles chosen as the signal hypothesis. The results of these trainings are shown in tables A.3 to A.5.

These trainings are performed at an earlier stage of the analysis optimization and use the jet removal mass decorrelation technique described in the next section. This results in lower overall values of the ROC AUC. The training performed with the CP-odd signal as the target achieves the best separation of every BSM signal point except for $\alpha = 60°$ from the SM. Once converted to the rejection power of the SM hypothesis, the $\alpha = 90°$ training performs the best. The difference is significant for larger mixing angles than the alternate

α_{obs}	0	15	30	45	60	75	90
45°	0.71	0.80	0.95	1.53	2.20	3.94	5.19
60°	0.69	0.77	0.93	1.43	2.08	3.66	4.77
90°	0.69	0.78	0.92	1.46	2.15	3.80	5.02

Table A.4: Expected significance in σ to reject the background-only hypothesis when observing a CP mixing angle α_{obs} using the number counting method for CP GNN trainings performed with alternate choices of signal sample.

α_{obs}	0	15	30	45	60	75	90
45°	0	0.01	0.06	0.28	0.62	1.64	2.39
60°	0	0.01	0.05	0.27	0.65	1.64	2.32
90°	0	0.01	0.06	0.29	0.68	1.83	2.57

Table A.5: Expected significance in σ to reject the SM hypothesis when observing a CP mixing angle α_{obs} using the number counting method for CP GNN trainings performed with alternate choices of signal sample.

training targets and negligible for angles at or below the alternate training targets. The rejection power of the background-only hypothesis does not follow this trend in part because of the variation in scores given to $t\bar{t}H$ by the different CP GNNs. CP GNNs trained at smaller values of α tend to admit more $t\bar{t}H$ into the most signal-like regions, which helps reject the background-only hypothesis, but is harmful in rejecting the SM.

Figures A.7 and A.8 show the distributions of the SB and CP GNN outputs for events in the signal region. The SB and CP GNNs tend to be positively correlated, and events may generally be deemed signal-like if they have high scores for both. The region where both scores are large is significantly enriched in CP-odd tH and is expected to be the most sensitive. tH events are generally more signal-like than $t\bar{t}H$ events, and signal events under the CP-odd hypothesis are generally more signal-like than events under the CP-even hypothesis. The tW background is significantly more signal-like than $t\bar{t}$, especially in the SB GNN distribution. This results in relatively large $tW/t\bar{t}$ ratios in the most sensitive regions.

GNN Mass Decorrelation

A data-driven background technique based on extrapolation from a two b-jet control region to the three b-jet signal region is under study. The extrapolation is performed by randomly increasing the PCBT scores of jets that are not b-tagged. The technique requires a fit to the m_{bb} mass spectrum in each region of the two-dimensional GNN plane used in the final analysis to constrain the overall normalization and relative contributions of the background templates derived from the control regions. In the fit, the sideband regions

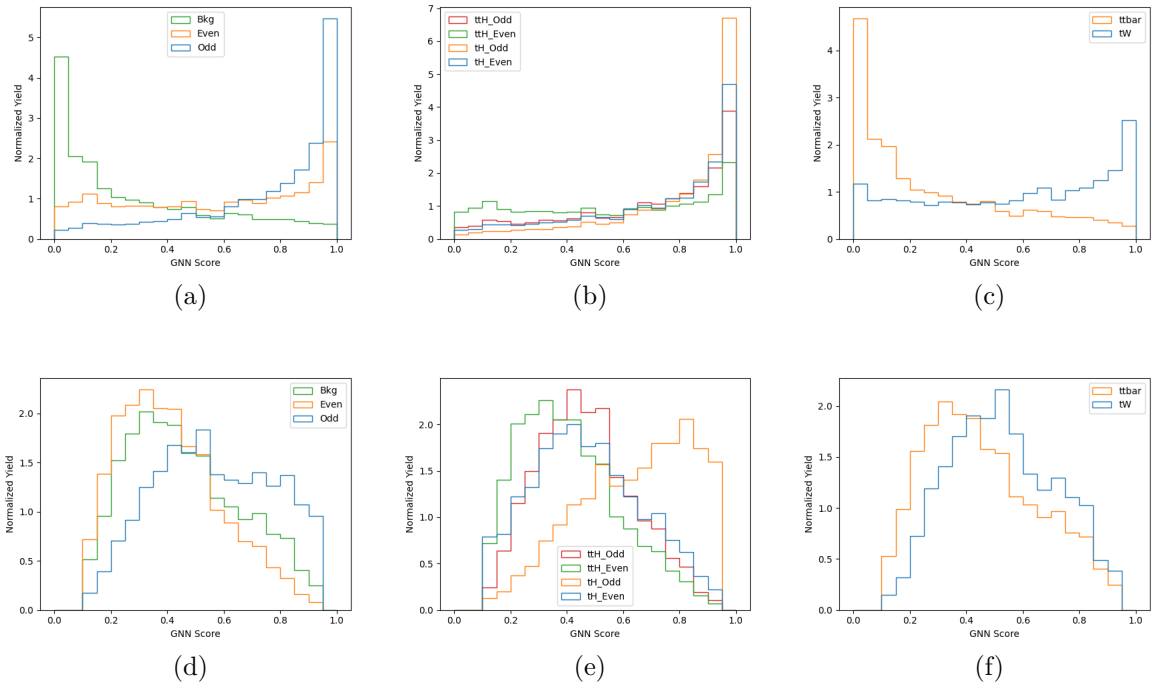


Figure A.7: Score distributions for the SB GNN (a-c) and CP GNN(d-f). Each histogram is normalized to have an equal area. (a,d) compare the $t\bar{t}+tW$ background with the $t\bar{t}H+tHjb+tWH$ signal under the CP-even and odd hypotheses. (b,e) show the $t\bar{t}H$ and tH shapes under different hypotheses. (c,f) show the $t\bar{t}$ and tW shapes.

constrain the background, allowing the bins in the mass window to constrain the signal normalization. This technique requires that the background does not peak sharply at $m_{bb} = m_h$ in the most signal-like regions. The naive SB GNN implementation creates such a peak because the p_T, η, and ϕ of the two b-jets are included as inputs to the GNN, and m_{bb} can be constructed from these observables.

Several techniques have been studied to mitigate the sculpting of the background mass spectrum induced by the GNN. Many studies requiring independence between a neural network output and an invariant mass reweight the training sample to have a flat mass distribution. This approach is not feasible in this case because the training datasets lack sufficient statistics to be useful after a significant reweighting. The most straightforward approach is to alter the GNN inputs so that there is not enough information to calculate m_{bb}. This ensures that the GNN does not learn to use the mass as a discriminating variable. Two attempts to achieve this by scaling the p_T of all jets in the event, either by H_T or by m_{bb}, fail to mitigate the mass sculpting. This may be related to the fact that ΔR_{bb} is highly correlated with m_{bb}, especially in the boosted regime. The most successful information

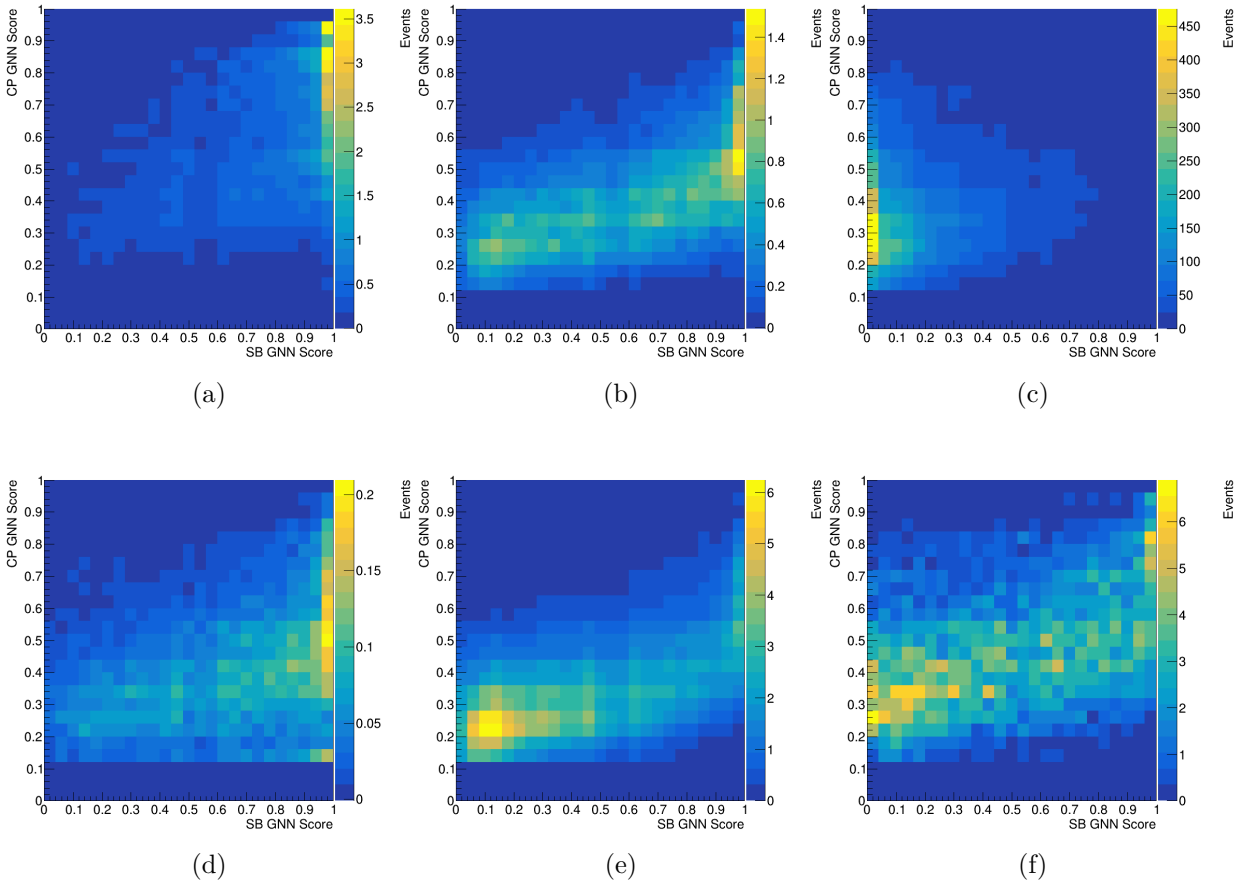


Figure A.8: Distributions in the 2D SB GNN versus CP GNN plane for CP-odd (a) and even (d) tH events, CP-odd (b) and even (e) $t\bar{t}H$ events, and the $t\bar{t}$ (c) and tW (f) backgrounds.

removal technique is to remove the two b-tagged jets assigned to the Higgs candidate from the graph and replace them with a single node representing the Higgs candidate. The node features for this node are given the four momentum $(p_T, \eta, \phi, E) = (p_T^{bb}, \eta_{bb}, \phi_{bb}, p_T^{bb} \cosh \eta_{bb})$ so that the mass information is not preserved. This training strategy is called "dijet removal."

Figure A.9 shows the effect of dijet removal on the mass sculpting in the SB GNN. Without dijet removal, the $t\bar{t}$ background receives a strong peak at $m_{bb} = m_h$ in the region of high SB GNN scores. Likewise, the signal peak becomes narrower in this region and is quite broad in the low score region. With dijet removal, the mass sculpting effect is significantly reduced in all samples, and the effect in $t\bar{t}$ is to enhance the region near $m_{bb} = m_t$ rather than generate a sharp Higgs mass peak. Table A.6 shows the effect of dijet removal on the separation power of the SB GNN. The ROC AUC shows reduced separation power, as is expected from the removal of a powerful discriminating variable. However, when the expected

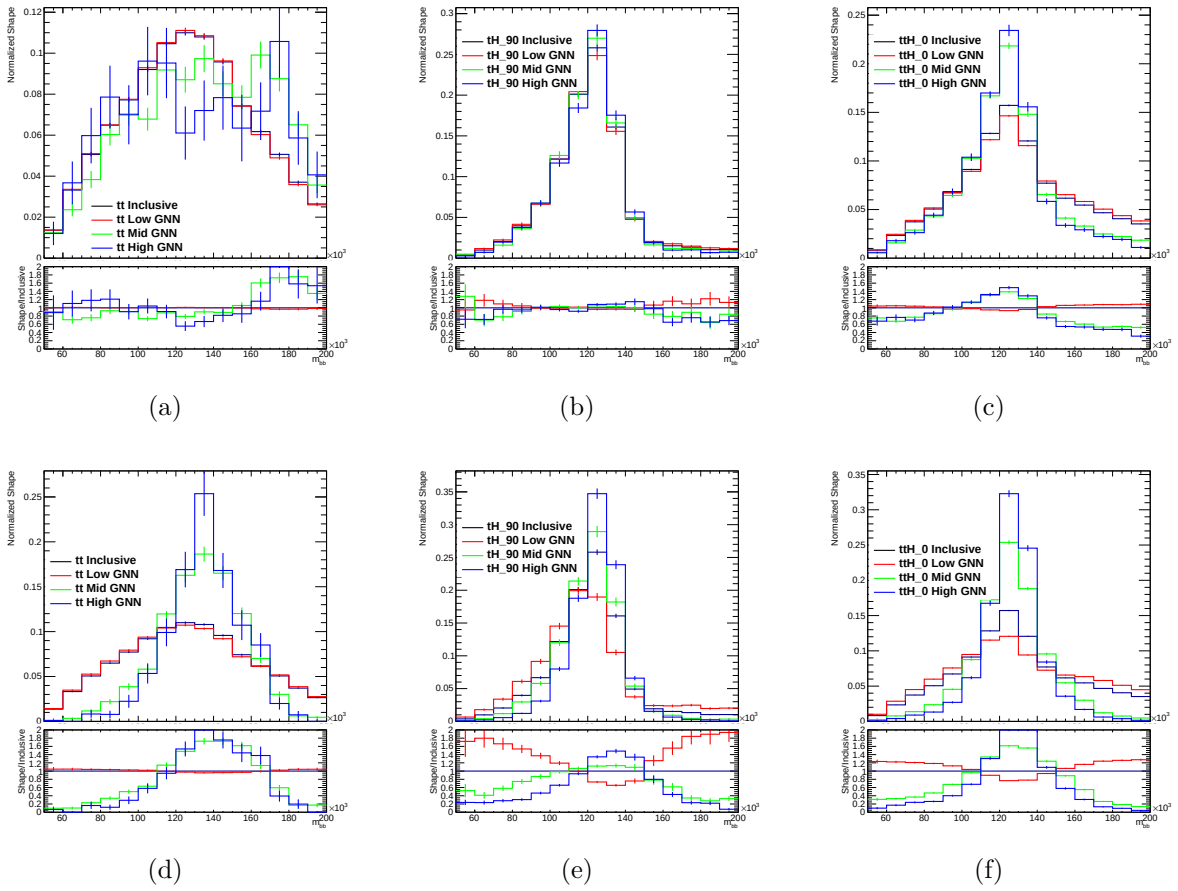


Figure A.9: Comparison of the m_{bb} spectrum of the $t\bar{t}$ background (a,d), CP Odd tH signal (b,e), and CP Even $t\bar{t}H$ signal (c,f) with (a-c) and without (d-f) removing the dijet system reconstructed as the Higgs boson candidate from the graph. Each plot compares the mass distribution inclusively in the analysis region to the distribution in regions with high, medium, and low SB GNN scores determined by cuts placed at a fixed efficiency for the $t\bar{t}$ background of 97.26% and 99.67%.

Metric	ROC AUC		Expected Significance [σ]		
Signal Hypothesis	CP Even	CP Odd	CP Even	CP Odd	CP Odd
Background Hypothesis	Bkg. Only	Bkg. Only	Bkg. Only	Bkg. Only	SM
Default SB GNN	0.730	0.857	3.39	2.15	0.61
Dijet Removal	0.648	0.796	2.27	2.16	0.94

Table A.6: ROC AUC and expected significance for the separation of $t\bar{t}H$+$tHjb$+tWH signal from $t\bar{t}$+tW background by the SB GNN trained with and without mass sculpting mitigation by removing the dijet system from the graph. The Expected significance is evaluated using three bins each in the SB and CP GNNs and applying the number counting formula in the mass window $m_{bb} \in [120, 130]$ GeV

sensitivity to the CP Odd scenario is evaluated using a method that separately exploits the Higgs mass information, as would be done in the case of the data-driven background estimation, the training with dijet removal achieves a better sensitivity. An alternative strategy using an MC-based background estimate and replacing the fit to the m_{bb} spectrum with a more finely binned fit to the GNN distribution is still under study, so a comparison to the ultimate sensitivity of this approach is not yet possible.

Alternate approaches to reduce the mass sculpting through the addition of a penalty term in the loss function used to train the SB GNN have also been studied. Though these approaches were ultimately not used in favor of the simpler information approval approach, they are documented here for the sake of completeness. The principle of these approaches is to add a term to the loss function, which is evaluated on a batch of events and grows when the GNN classifier output y is correlated with m_{bb}. To be effective as a loss term, the quantity must also be differentiable with respect to the GNN output. The first possible choice is the Pearson correlation coefficient

$$\rho_{y,m_{bb}} = \frac{\sum_{i=1}^{N}(m_{bb}^{i} y_i - \langle y \rangle \langle m_{bb} \rangle)}{N \sigma_y \sigma_{m_{bb}}}.$$

$\rho_{y,m_{bb}}$ is bounded between -1 and 1 and is 0 when there is no linear correlation between m_{bb} and y. Combining this with the standard binary cross-entropy loss, the loss function becomes

$$\mathcal{L} = \alpha |\rho_{y,m_{bb}}| - \sum_i \left(y_i^{\text{true}} \log(y_i) + (1 - y_i^{\text{true}}) \log(1 - y_i) \right)$$

where α is a hyperparameter controlling the size of the penalty term. An initial choice of $\alpha = 1/40$, chosen to roughly match the size of the loss terms based on the correlation evaluated on a training without the penalty, resulted in the post-training correlation coefficient measured in the testing dataset reducing from -0.25 to -0.20. Additional tests with $\alpha = 1/10$ and $\alpha = 1/4$ resulted in $\rho_{y,m_{bb}} = -0.10$ and $\rho_{y,m_{bb}} = -0.13$, respectively,

with indications of significant overtraining in the latter case. The second major drawback to this approach is the ability of the GNN to sculpt m_{bb} spectrum without inducing a linear correlation between m_{bb} and y. Qualitatively, this occurs if the Higgs mass peak is enhanced in high score regions by equally rejecting events with lower and higher m_{bb}. This motivates the development of an alternative penalty term that captures nonlinear correlations between m_{bb} and y. Given the joint probability density function for two variables $\rho(x, y)$, the statistical independence of the random variables x and y may be guaranteed by asserting that the joint probability distribution is equal to the product of the marginalized distributions, i.e., $\rho(x, y) = \rho_x(x)\rho_y(y)$ where $\rho_x(x) = \int \rho(x, y)dy$ and similarly for y. A loss term may be constructed using a difference metric between the true probability density and the product of marginalized distributions. In this case, the L^2 distance is used, yielding

$$\mathcal{L}_{\text{pen}} = \iint \left(\rho(x, y) - \rho_x(x)\rho_y(y)\right)^2 dxdy.$$

Use in the loss function requires a differentiable estimate of the probability density from a finite batch of samples, which may be obtained by kernel density estimation (KDE). In this case, a simple Gaussian KDE is applied.

$$\rho_x(x) = \frac{1}{N\sigma\sqrt{2\pi}} \sum_{i=1}^{N} e^{-\frac{(x-x_i)^2}{2\sigma^2}} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{N}(x|x_i, \sigma^2)$$

$$\rho_y(y) = \frac{1}{N\sigma\sqrt{2\pi}} \sum_{i=1}^{N} e^{-\frac{(y-y_i)^2}{2\sigma^2}} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{N}(y|y_i, \sigma^2)$$

$$\rho(x, y) = \frac{1}{N\sigma\sqrt{2\pi}} \sum_{i=1}^{N} e^{-\frac{(x-x_i)^2+(y-y_i)^2}{2\sigma^2}} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{N}(x|x_i, \sigma^2)\mathcal{N}(y|y_i, \sigma^2)$$

where $\mathcal{N}(x|\mu, \sigma^2)$ represents the PDF for a normal distribution with mean μ and variance σ^2. To simplify the calculation, x and y are scaled without loss of generality to have mean zero and variance one. The optimal choice of the bandwidth σ is then related to the number of samples, which determines the average space between data points. The choice $\sigma = N^{-1/3}$ is found to give a good estimate of the PDF. For simplicity, the shared variance σ^2 is suppressed in the arguments to $\mathcal{N}$ from here. After inserting the estimated densities into the formula for $\mathcal{L}_{\text{pen}}$, the integrals may be evaluated.

$$\mathcal{L}_{\text{pen}} = \iint \left(\rho(x, y)^2 - 2\rho(x, y)\rho_x(x)\rho_y(y) + \rho_x(x)^2\rho_y(y)^2\right) dxdy \tag{A.4}$$

Each term is evaluated separately. The last is shown as an example because it is the most complex.

$$\begin{aligned}\iint \rho_x(x)^2\rho_y(y)^2 dxdy &= \iint \left(\frac{1}{N^4}\sum_{i,j,k,l=1}^{N}\mathcal{N}(x|x_i)\mathcal{N}(x|x_j)\mathcal{N}(y|y_k)\mathcal{N}(y|y_l)\right)dxdy \\ &= \frac{1}{N^4}\sum_{i,j,k,l=1}^{N}\int\left(\mathcal{N}(x|x_i)\mathcal{N}(x|x_j)\right)dx\int\left(\mathcal{N}(y|y_k)\mathcal{N}(y|y_l)\right)dy \\ &= \frac{1}{4\pi\sigma^2 N^4}\sum_{i,j,k,l=1}^{N} e^{-\frac{(x_i-x_j)^2}{4\sigma^2}}e^{-\frac{(y_k-y_l)^2}{4\sigma^2}}\end{aligned} \tag{A.5}$$

The other calculations are similar but with some indices matched.

$$\iint \rho(x,y)^2 dxdy = \frac{1}{4\pi\sigma^2 N^2}\sum_{i,j=1}^{N} e^{-\frac{(x_i-x_j)^2}{4\sigma^2}}e^{-\frac{(y_i-y_j)^2}{4\sigma^2}} \tag{A.6}$$

$$\iint \rho(x,y)\rho_x(x)\rho_y(y) dxdy = \frac{1}{4\pi\sigma^2 N^3}\sum_{i,j,k=1}^{N} e^{-\frac{(x_i-x_j)^2}{4\sigma^2}}e^{-\frac{(y_i-y_k)^2}{4\sigma^2}} \tag{A.7}$$

The full integral may then be written as a single sum using the Kronecker delta δ_{ij}.

$$\mathcal{L}_{\text{pen}} = \frac{1}{4\pi\sigma^2 N^2}\sum_{i,j,k,l}^{N}\left(\delta_{ik}\delta_{jl} - \frac{2}{N}\delta_{ik} + \frac{1}{N^2}\right)e^{-\frac{(x_i-x_j)^2+(y_k-y_l)^2}{4\sigma^2}} \tag{A.8}$$

This loss function, called the KDE loss, is differentiable, positive semi-definite, and is guaranteed to provide the desired property if it is equal to zero. Written in this form, the KDE loss may be computed efficiently by precomputing the exponentiated differences separately for x and y and performing the sum using efficient vectorized operations. An implementation of the KDE loss is available in the `root_gnn_dgl` package. The full loss function used to train the SB GNN with KDE loss penalty included is

$$\mathcal{L} = \alpha\mathcal{L}_{\text{pen}} - \sum_i \left(y_i^{\text{true}}\log(y_i) + (1-y_i^{\text{true}})\log(1-y_i)\right),$$

where $\alpha = 4$ is found to be a good value of the weight. A final strategy studied is adversarial training in which a small adversary DNN is trained to predict m_{bb} using the output of the GNN. The adversary is trained with the mean absolute error loss function

$$\mathcal{L}_{\text{adv}} = \sum_i |m_i^{\text{pred}} - m_i^{\text{true}}|,$$

and the SB GNN classifier is trained with a penalty proportional to the loss achieved by the adversary DNN.

$$\mathcal{L} = -\alpha\mathcal{L}_{\text{adv}} - \sum_i \left(y_i^{\text{true}}\log(y_i) + (1-y_i^{\text{true}})\log(1-y_i)\right),$$

Signal Hypothesis Background Hypothesis	CP Even $t\bar{t}$	CP Odd $t\bar{t}$
Baseline	0.698	0.817
Dijet Removal	0.637	0.771
Adversarial Training	0.680	0.785
KDE Loss	0.662	0.781
Dijet Removal and KDE Loss	0.676	0.760
Dijet Removal and Adversarial Training	0.639	0.754

Table A.7: ROC AUC comparison for trainings with different mass sculpting strategies implemented.

where again $\alpha = 4$ is found to be an appropriate weight. Results of trainings performed at an earlier stage of the analysis optimization using different combinations of these strategies are shown in table A.7 and fig. A.10.

The results show that dijet removal is the most powerful mass sculpting mitigation technique, which may be attributed to the fact that the loss terms introduced by the other techniques all compete with the cross entropy loss, which is significantly reduced by using m_{bb} as a discriminating variable in the high score regions. As a result, these methods tend to reduce sculpting in the medium and low regions, but still create a mass peak in the highest GNN region. A quantitative measure of the mass sculpting is obtained by measuring the χ^2 value between the inclusive mass spectrum and the mass spectrum in the medium and high GNN regions. This metric indicates a minor improvement in the mass sculpting when adding the KDE loss or adversarial training in addition to the dijet removal, but the small improvement comes with a noticeable drop in separation power.

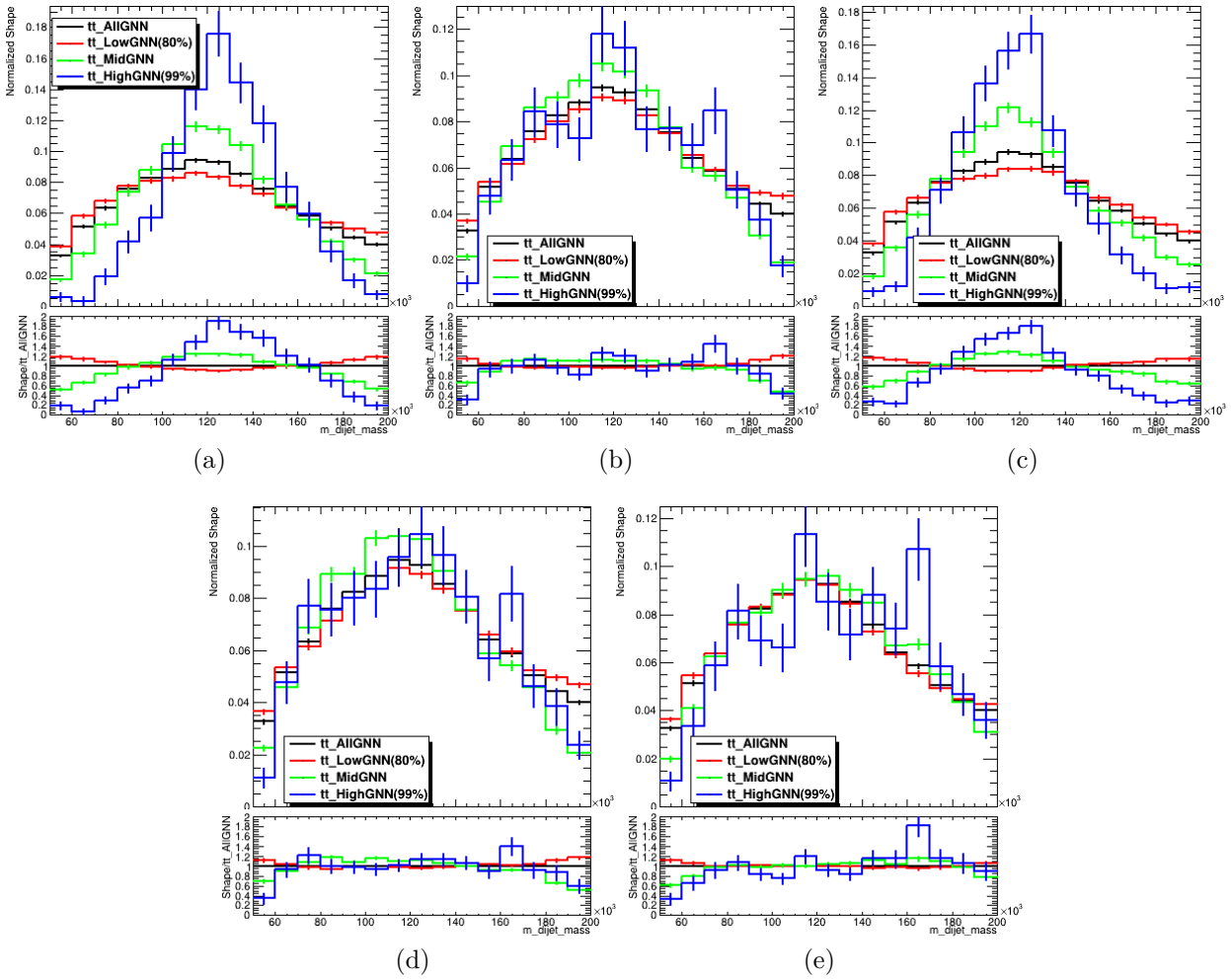


Figure A.10: Mass spectrum comparison for trainings with various mass sculpting mitigation methods. (a) Default training. (b) Dijet removal. (c) KDE loss. (d) KDE loss and dijet removal. (e) Adversarial training and dijet removal. Each comparison is performed using cuts chosen to have a $t\bar{t}$ acceptance of 1% for the high GNN region and 80% for the low GNN region.

Appendix B

Transformer Energy Calibration of Electrons and Photons

The accurate and efficient reconstruction of electrons and photons (hereafter e/γ) is fundamental to many LHC measurements because isolated high-energy electrons and photons are rare in pure QCD interactions, which constitute the majority of the inelastic pp cross-section. Additionally, the detector signatures of e/γ are relatively easy to identify both at trigger level and at reconstruction level. This makes e/γ powerful tools to reject large QCD backgrounds, including in the trigger, with many LHC data analyses relying on lepton or photon triggers to collect collision events of interest. Additionally, e/γ final states are measured more precisely than hadronic final states, which allows the reconstruction of resonances in important electroweak decays such as $Z \to e^+e^-$, $H \to \gamma\gamma$, $H \to ZZ^* \to 4\ell$, and $H \to Z\gamma \to \ell\ell\gamma$. Maximizing the detector performance in the measurement of these states with state-of-the-art machine learning methods has a broad physics impact on the LHC physics program.

As described in chapter 4, electrons and photons are reconstructed in ATLAS by finding large, isolated energy deposits in the electromagnetic calorimeters using the topo cluster algorithm [134]. These clusters are combined by a superclustering algorithm that merges nearby clusters and clusters matched to nearby tracks. The electron and photon energies are then calibrated to correct the total energy of the supercluster for energy losses in passive detector material and outside of the calorimeter. Additional corrections are applied for detector effects that differentially affect the calorimeter layers, and simulated data is corrected to match the observed performance of the real detector. Objects reconstructed as both electron and photon candidates are passed through an ambiguity resolution algorithm to determine their identity. Additional cuts on the quality of the reconstructed e/γ candidate, known as identification criteria, are placed to reject fake e/γ candidates from hadronic activity or from muons. Rejection of e/γ candidates with nearby activity in the hadronic calorimeter or muon system, known as isolation requirements, further removes real electrons and photons originating from hadronic decays rather than from the hard scatter process. The efficiency of the identification and isolation criteria must then be calibrated for both real electrons and

photons and for fake and non-prompt e/γ candidates. The various algorithms deployed in the process of e/γ reconstruction and calibration take different combinations of inputs from the electromagnetic calorimeter and other detectors and range from traditional cut-based algorithms using a few observables to DNNs.

This appendix describes the development and performance of a transformer algorithm that uses the individual cells of the liquid argon electromagnetic calorimeter as inputs and could, in principle, be used for many of these tasks. The development thus far has focused on the MC-based e/γ energy calibration for analyzing data taken during Run 3 of the LHC. Improvements to the e/γ energy resolution are expected to benefit most measurements using these final states, but the impact can be understood most easily in the case of measurements of narrow resonances such as $H \to \gamma\gamma$. In an analysis like the $t\bar{t}H(H \to \gamma\gamma)$ measurement described in appendix A, the significance of the observed signal is approximated by the Gaussian number counting estimate $z = s/\sqrt{b}$ evaluated under the $m_{\gamma\gamma}$ mass peak. The background estimate in the mass window is evaluated by an extrapolation from the signal sidebands, which carries additional statistical and systematic uncertainties from the background fit, choice of background fit function, and spurious signal test. These uncertainties may be summarized by a single uncertainty σ_b. Including the uncertainty in the background estimate added in quadrature with the inherent statistical uncertainty in the background, the significance is $z = s/\sqrt{b + \sigma_b^2}$. The $m_{\gamma\gamma}$ resolution is proportional to the photon energy resolution σ_E. When the resolution improves by a factor α, the mass window shrinks by a corresponding factor with a constant signal yield s while the background yield and uncertainty both decrease by the same factor α. Therefore, the simplified estimate of the signal sensitivity z becomes

$$z' = \frac{s}{\sqrt{\alpha b + \alpha^2 \sigma_b^2}} = z\sqrt{\frac{b + \sigma_b^2}{\alpha b + \alpha^2 \sigma_b^2}}. \tag{B.1}$$

In this simplified model, the sensitivity of measurements dominated by statistical uncertainties is proportional to $\sigma_E^{-1/2}$ and that of measurements dominated by uncertainties on non-resonant backgrounds is proportional to σ_E^{-1}. This model is overly simplistic, but it illustrates that improvement in the energy resolution can immediately bring increased precision to measurements of resonant final states by reducing both the statistical and systematic uncertainties. These improvements are relevant to searches for production of resonances from BSM particles and to measurements of SM processes, though the situation is more complicated for precision measurements such as the W and Higgs boson mass measurements, which are dominated by systematic uncertainties on the signal. The remainder of this appendix will describe the purpose and current implementation of the MC-based e/γ energy calibration, the implementation of the transformer model, and the performance of the transformer-based calibration in MC simulation and data.

B.1 MC e/γ Energy Calibration

The MC energy calibration is a correction to the measured energy of a cluster of calorimeter energy deposits assigned to an e/γ candidate to the true energy of the electron or photon that created the cluster. The MC-based energy calibration correction is the most significant part of the broader e/γ energy calibration procedure, which is shown in fig. B.1. The correction is derived from the simulated response of the calorimeter using MC simulations in which a single e/γ is generated and its interactions with the detector are simulated. This correction accounts for the effects of energy losses upstream of the calorimeter in the ID, lateral leakage not captured in the cluster, and longitudinal leakage into the hadronic calorimeter. It may also correct for contributions from pileup interactions and other effects from the calorimeter response. The current MC calibration [134] uses a BDT trained on information describing the longitudinal shower development, namely the energy deposited in each layer of the calorimeter, and angular information describing the location of the shower within the detector and the alignment of the cluster with the boundaries of the calorimeter cells in η and ϕ. To better capture the behavior of electromagnetic showers that are largely contained in the gap between the barrel and endcap calorimeters, the energy deposited in scintillators placed in the crack region is also used as an input feature to serve as a proxy for the energy lost in the passive detector material. The BDT is trained separately for electrons, photons that convert to e^+e^- in the ID (converted photons), and photons that are not associated with any hits in the ID (unconverted photons). Additionally, because the detector geometry and response vary significantly across the kinematic region of interest, the BDT is trained separately in 117 bins of E^T and η, resulting in a total of 351 individual BDT models. In the case of converted photons, additional input variables describing the reconstructed tracks and the conversion vertex associated with the photon candidate are used. The training target of the BDT is regression of the measured energy of the cluster to the true energy of the simulated e/γ taken from the generation.

B.2 Transformer Model

Transformers and graph neural networks are naturally suited to machine learning tasks on calorimeter data because the cell-level inputs are irregular, unordered, and sparse. They carry geometric relationships, but the geometry is irregular, especially at boundaries between detector regions, and long-range correlations within the cluster are also relevant. All of these features make these geometric learning algorithms suitable for use in calorimeter tasks, where using the cell-level inputs can provide performance improvement. Compared with the simpler observables used in the BDT, cell-level inputs are expected to provide additional information about the shape of the electromagnetic shower, which may help the algorithm to better estimate various sources of energy losses. The algorithm may also be able to identify energy deposits from pileup contributions or detector noise and more effectively remove these from the calibrated energy.

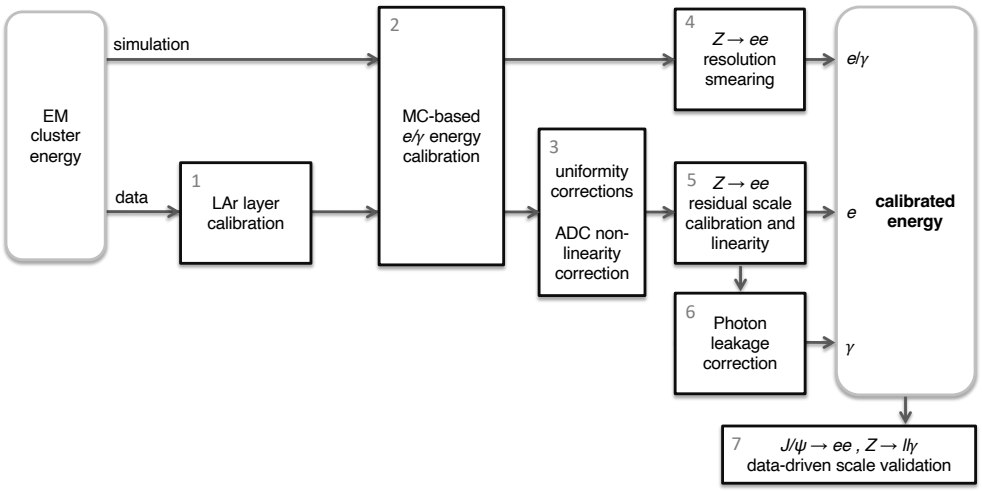


Figure B.1: Schematic overview of the electron and photon energy calibration procedure in ATLAS. [282]

The transformer model implemented for the e/γ energy calibration uses the same architecture and software framework [233] as the GN2 jet flavor tagging algorithm [283], also developed within ATLAS. A schematic of the model is shown in fig. B.2. The nodes of the graph input to the transformer are the individual cells of all calorimeter cells assigned to the cluster, including those in the pre-sampler. Each cell carries as cell features its energy E, angular position (η, ϕ), Cartesian coordinates (x, y, z), and layer number. The energies include corrections for the sampling fraction in each cell, and all positional features are measured in the lab frame. Each cell carries additional global features, which are shared among all cells and provide context about the cluster as a whole. These features are the total cluster energy, η, and ϕ. Additionally, a selection of BDT inputs not captured by the cell inputs or found not to be fully exploited by the transformer is provided. These are the energy fraction in the presampler, the energy fraction in the scintillators, and the ratio of the energies in the first two layers of the EM calorimeter. When training on converted photons, the same variables related to the conversion vertex used in the BDT are included. The typical cluster contains between 100 and 200 cells, but the largest clusters have over 600 cells.

The cell inputs are encoded into a 128-dimensional latent space by a DNN with a single

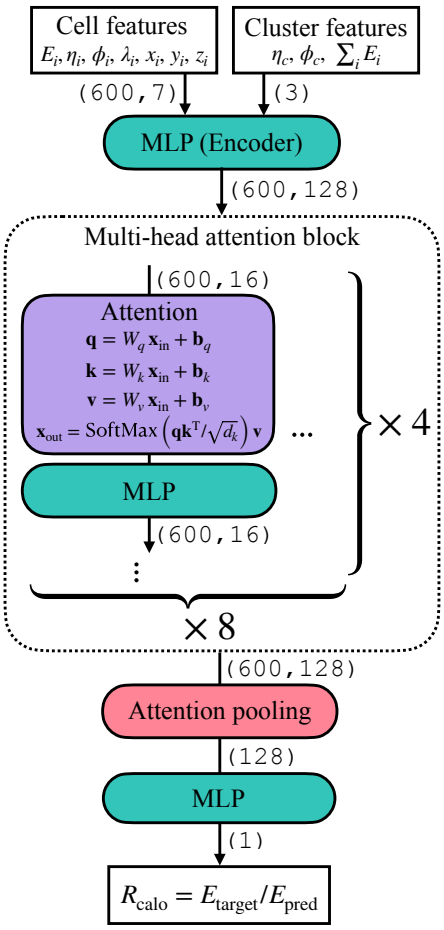


Figure B.2: Schematic of the transformer model architecture used for the MC e/γ energy calibration.

hidden layer of width 256. The encoded node features are then passed through a multi-headed scaled dot product self-attention [216] block with eight heads and four attention layers. LayerNorm [236] is applied before each layer. The queries, keys, and values are calculated using linear transforms of the embedded cell features, where the parameters of the linear transforms are independently learnable for the queries, keys, and values of each head of each layer. The cell features are combined via global attention pooling [284], which computes weights for the pooling by passing the cell features through a linear layer. The aggregated features g_i are computed from the cell features n_{ij} where the first index runs over the cells.

$$g_i = \sum_j \frac{\exp\left(B_k n_{kj}\right)}{\sum_l \exp(B_k n_{kl})} n_{kj}$$

The vector B_k is a learned linear layer used as the input to the softmax function to

compute weights for the cell feature aggregation.

A final DNN with three layers of width 128, 64, and 32 acts on the pooled cell features to produce the model output, which is the predicted ratio of the true e/γ energy to the total energy in the calorimeter energy cluster. All NN layers in the model use the swish activation function.

The transformer is trained using the mean absolute error (MAE) loss function, which in this case may be written

$$\mathcal{L}_{\text{MAE}} = \sum_i \left| x_i - \frac{E^i_{true}}{E^i_{clus}} \right| = \sum_i \left| \frac{E^i_{pred} - E^i_{true}}{E^i_{clus}} \right|$$

where the index i runs over the training batch, E^i_{true} is the target truth energy, E^i_{clus} is the sum of the energy deposited in all calorimeter cells in the cluster, and x_i is the transformer output corresponding to a predicted energy $E^i_{pred} = x_i E^i_{clus}$. The choice to predict the calibration factor for the cluster energy rather than the cluster energy itself simplifies the desired behavior at high energies, where the model output is expected to approach a constant value. This ensures that the model extrapolates better to e/γ energies higher than those in the training dataset. The choice of MAE as the loss function is found to result in better energy resolution than the mean squared error. The training is performed with the Adam optimizer and a learning rate schedule that quickly warms up from an initial value of 10^{-7} to a maximum of 5×10^{-4} before eventually decaying to 10^{-5}. Overfitting is mitigated with the use of a weight decay of 10^{-5}.

The training is performed separately for electrons, converted photons, and unconverted photons, as in the case of the BDT. The trainings are based on samples of 60 million simulated electrons and 60 million simulated photons with pileup collisions overlaid. The pileup and detector conditions match those observed in 2023 data taking, with an average of 50.9 interactions per bunch crossing. Each training is performed over 48 hours using a total of 16 Nvidia A100 GPUs. Each GPU processes a mini-batch of 256 events during each iteration of the training. The e/γ used in training are required to satisfy basic quality criteria to ensure that they are well reconstructed and are consistent with the true energies used as training targets. These cuts include the Loose Likelihood ID working point for electrons and the Loose ID working point for photons [134], a requirement of $p_T > 5$ GeV, and a requirement that the cluster have at least five cells. Additionally, the reconstructed e/γ must be matched to the true e/γ via a truth matching algorithm and must be within $\Delta R < 0.1$ of the true e/γ. Finally, an energy matching requirement is placed to exclude e/γ that are not properly truth matched or which emit hard radiation not captured in the cluster. This requirement is

$$\left| \frac{E_{clus} - E_{true}}{E_{clus} + E_{true}} \right| < \frac{1}{3},$$

which ensures that the ratio of the cluster energy to the true energy is between 0.5 and 2.0. The kinematic distributions of the electrons in the training sample after all selections are shown in fig. B.3.

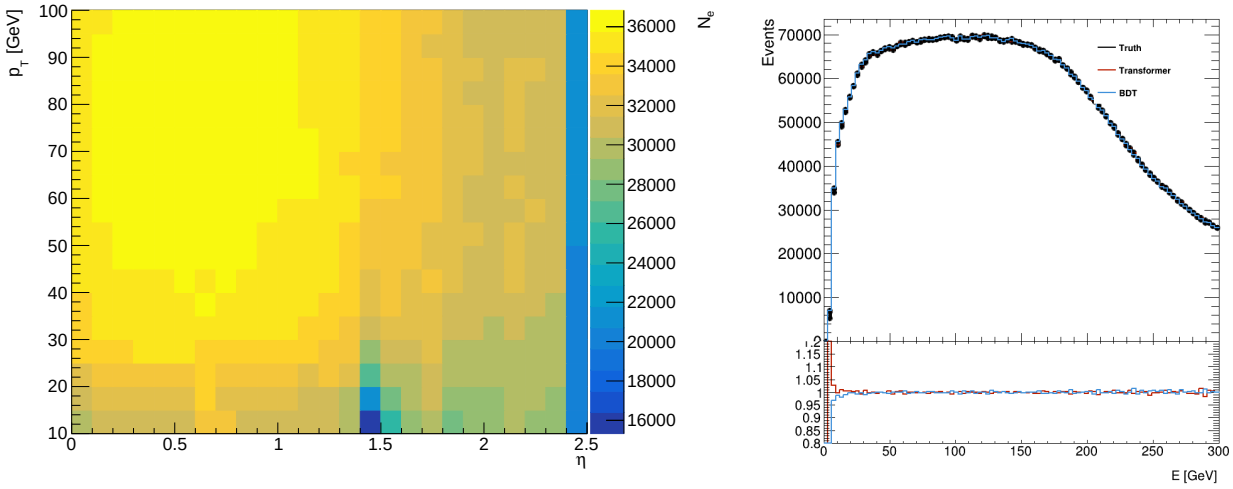


Figure B.3: (Left) Distribution in the pseudorapidity and p_T plane for electrons in the training sample. (Right) Energy distribution for electrons in the training sample, comparing the true energy to the energy calibrated by the transformer and the BDT. The bottom panel shows the ratio of the calibrated energy spectra to the true energy spectrum.

B.3 Performance of the Transformer-Based Calibration

This section presents best performance achieved by the transformer calibration to date and shows significant improvements over the previous results shown in Ref. [285]. The e/γ energy calibration transformer is under active development, and some results have been produced with different versions of the model. The qualitative features of the performance are consistent through various model optimizations and changes to the training setup, but some plots do not reflect the best overall performance of the transformer.

Performance in Single e/γ Simulation

The transformer learns an accurate correction to the cluster energy for all three species. The distribution of the relative error E_{pred}/E_{true} for each training is shown in fig. B.4.

The calibration should accurately correct the center of the distribution to one and should minimize the width of the peak, or equivalently, the resolution of the calibrated energy. The peak generated by the transformer is narrower and better centered on one than the BDT, indicating an improvement in the calibration. A complete evaluation of the quality of the calibration requires the use of several figures of merit to evaluate the energy resolution, bias, agreement between data and MC simulation, and size of the systematic uncertainties that must be placed on the calibration. These effects must be evaluated across the large range

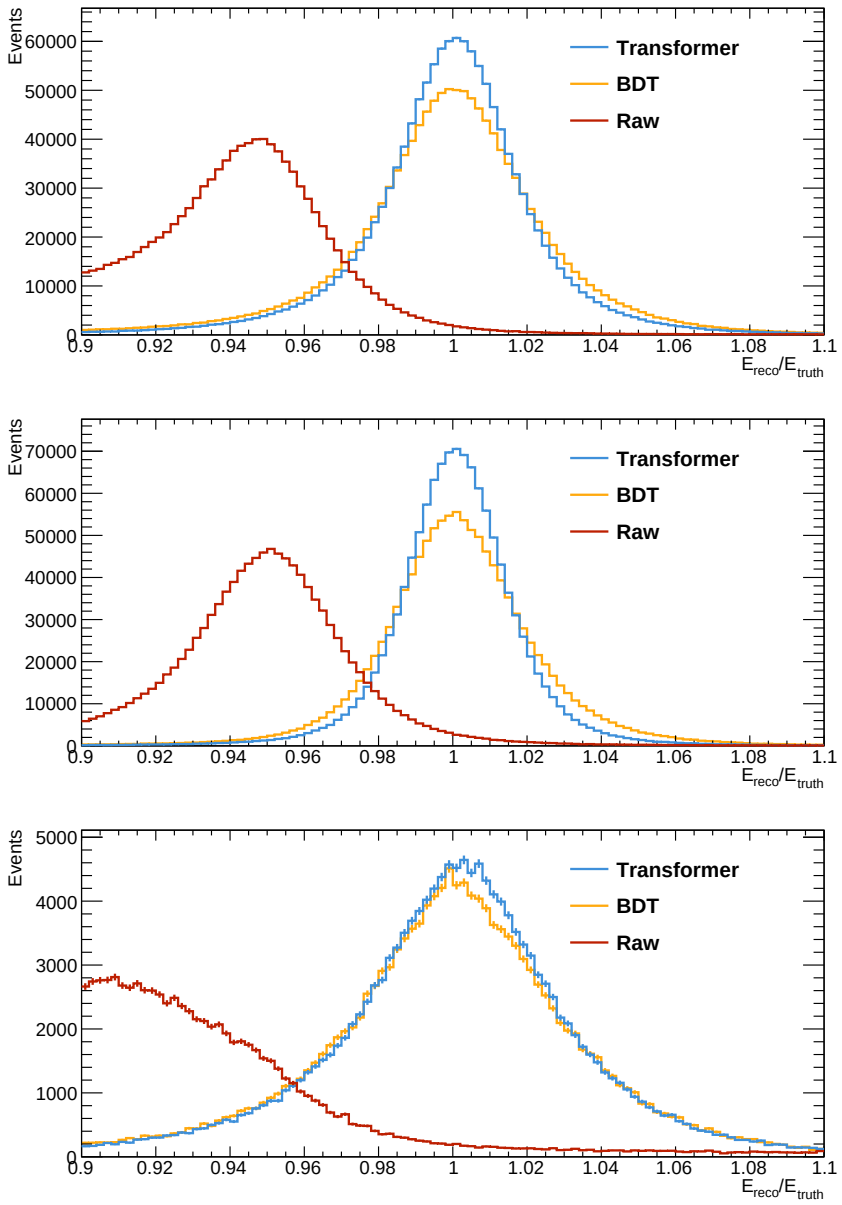


Figure B.4: Distribution of E_{pred}/E_{true} for various energy calibrations of electrons (top), unconverted photons (middle), and converted photons (bottom). Shown for e/γ with $30 < E < 100$ GeV and $|\eta| < 1.37$.

of energies relevant for LHC measurements and across different values of pseudorapidity for which the detector geometry varies significantly. The primary figure of merit used in the development is the resolution achieved in the single-particle simulation used for the training. The distribution of the relative error has long non-Gaussian tails in some regions, and the standard deviation may be overly sensitive to outlier effects, so the principal measure of resolution, denoted IQE or σ^E_{rel}, is based on the interquartile range (IQR).

$$\text{IQE} = \sigma^E_{rel} = \frac{\text{IQR}(E_{reco}/E_{true})}{1.349\langle E_{reco}/E_{true}\rangle} \tag{B.2}$$

The factor 1.349 is chosen so that a normal distribution with mean and variance one will have $\sigma^E_{rel} = 1$, and the mean position of the relative error in the denominator corrects possible bias in the resolution estimate from systematic underestimation. Without this factor, a calibration that consistently underpredicts the energy will appear to have a small resolution, even though the resolution may be large once the predictions are scaled to correct the bias. Figure B.5 shows a comparison of the relative resolution of the transformer and the BDT as a function of $|\eta|$ for e/γ in three different energy ranges.

The transformer shows a consistent improvement in the resolution of around 20% relative to the BDT across much of the relevant phase space. A more granular measurement of the resolution and bias of the BDT- and transformer-based calibrations is shown in fig. B.6. The improvement is generally larger in the forward regions in part because the transformer is more effective at compensating for the effects of pileup. This can be seen in fig. B.7, which instead shows the resolution as a function of the number of interactions per bunch crossing.

The transformer performs better at all values of pileup, and the improvement is particularly pronounced at large pileup, indicating that the transformer can compensate for the effects of pileup interactions on the cluster energy without any direct input about the pileup conditions of the event. This decreased susceptibility to pileup may be increasingly important for later data-taking years in Run 3, which tend to have more pileup collisions, and will be crucial for operations at the HL-LHC.

Comparison Between Data and Simulation at the Z Peak

Some properties of the electromagnetic showers are known not to be well modeled in the MC simulation and may therefore result in discrepancies in the calibration between simulated events and recorded data. While the performance in simulation can be corrected to match the data with the so-called *in situ* correction [134], achieving the increased sensitivity to key physics processes expected from a 20% improvement in the e/γ energy resolution requires ensuring that this resolution improvement is present in data and not an artifact of the simulation. The transformer is expected to be more susceptible to these effects because it uses the full granularity of the calorimeter and exploits details of the longitudinal shower shape, which exhibit poor modeling in the simulation.

The performance of the transformer in data is verified by using $Z \to e^+e^-$ decays as a standard candle. The performance of the transformer can be measured by performing a fit to

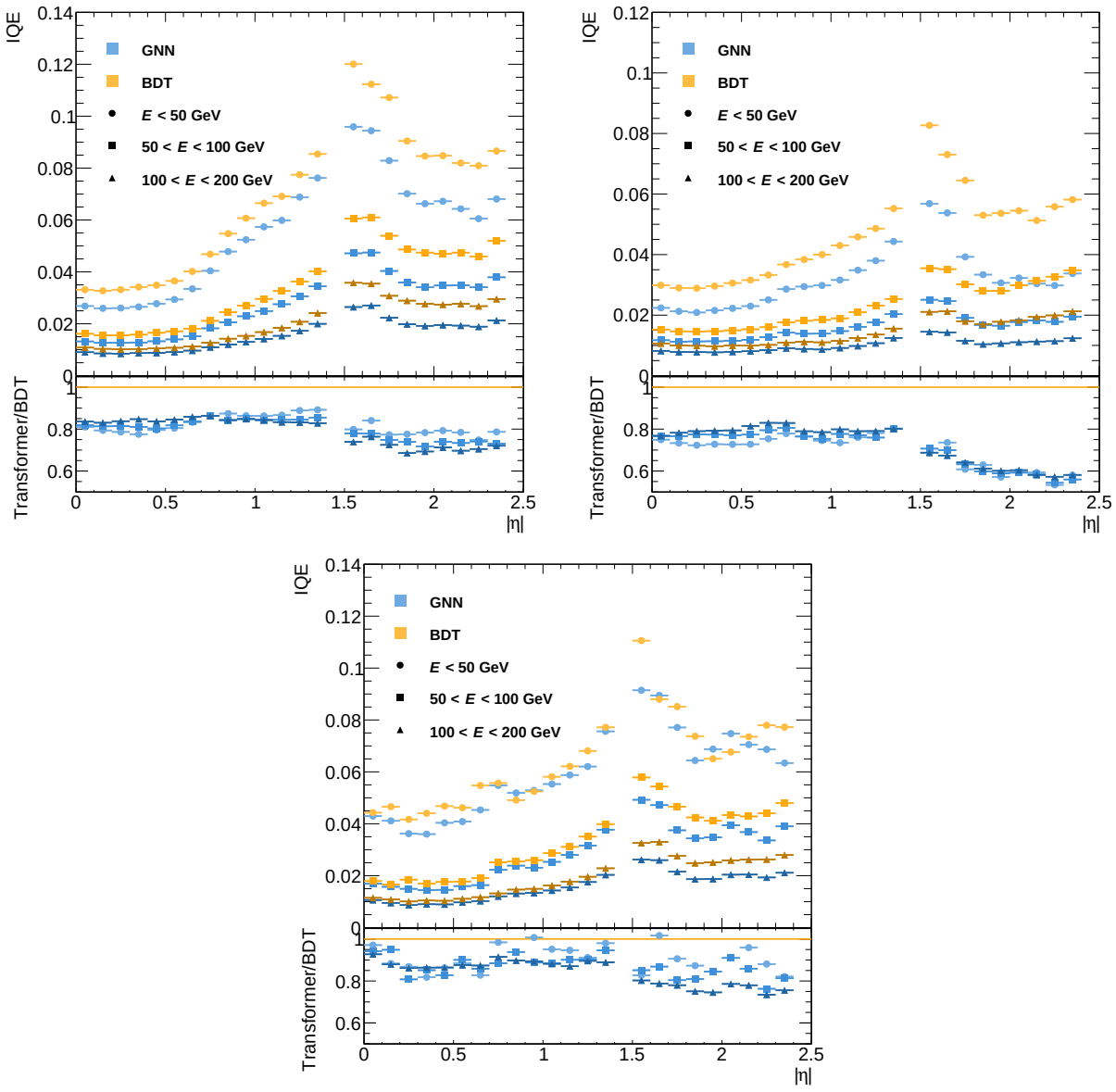


Figure B.5: Relative resolution σ_{rel}^{E} of electrons (top left), unconverted photons (top right), and converted photons (bottom) calibrated with the transformer and with the BDT in three different energy ranges as a function of $|\eta|$.

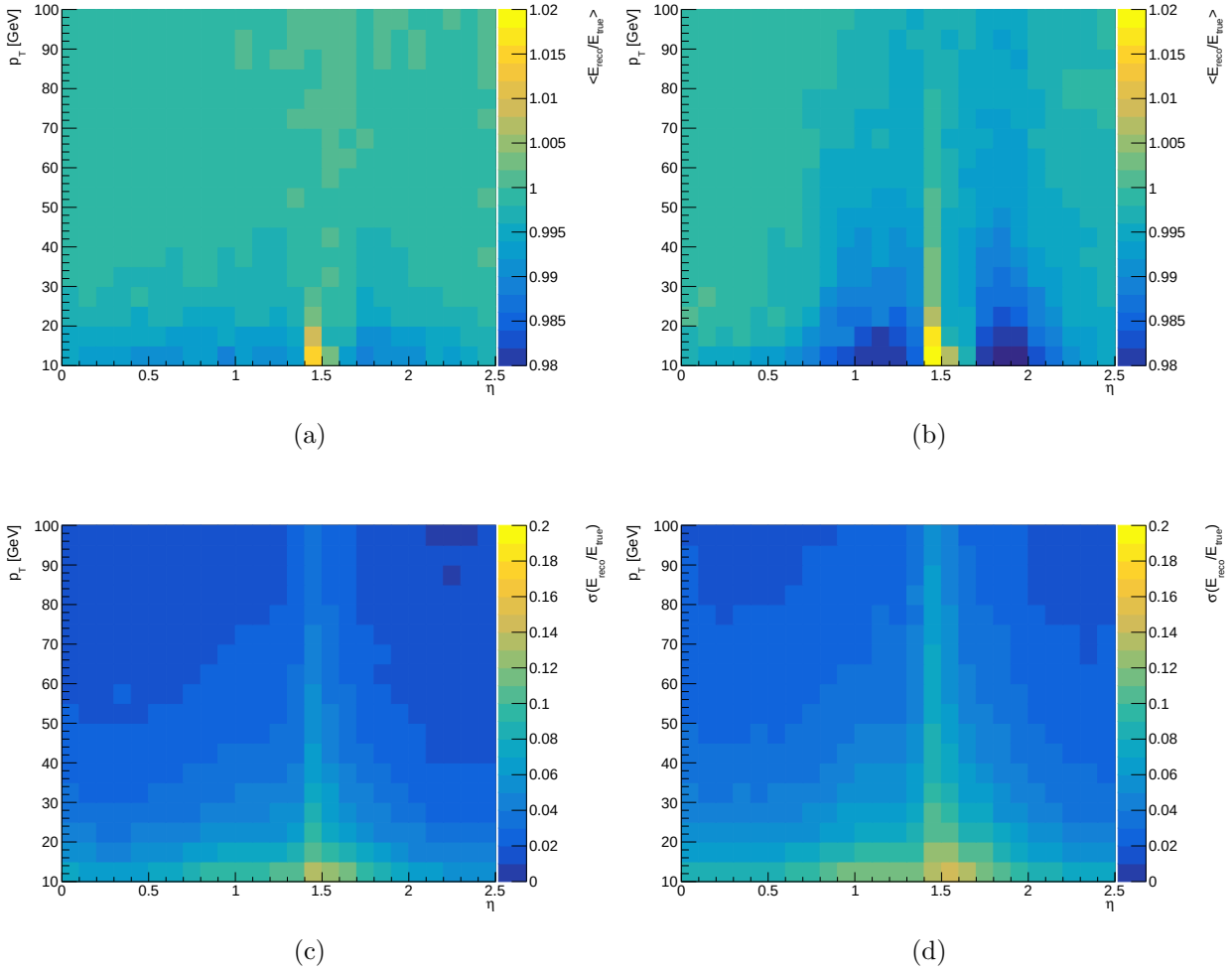


Figure B.6: Average bias (a,b) and resolution (c,d) in the transformer (a,c) and BDT (b,d) energy calibrations in bins of p_T and $|\eta|$. The resolution metric here is the standard deviation of the relative error E_{reco}/E_{true}.

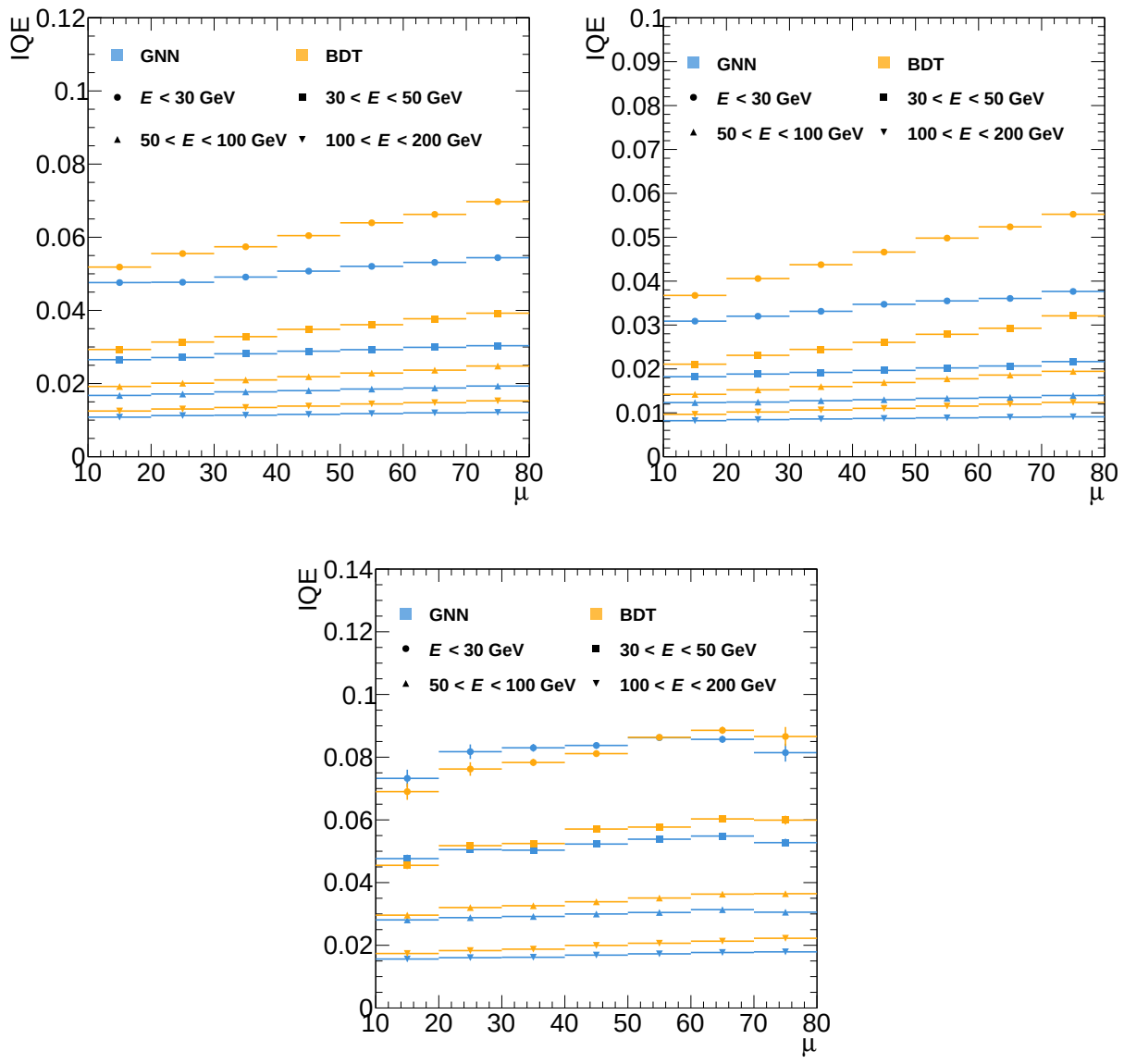


Figure B.7: Relative resolution σ_{rel}^{E} of electrons (top left), unconverted photons (top right), and converted photons (bottom) calibrated with the transformer and with the BDT in four different energy ranges as a function of the number of interactions per bunch crossing μ.

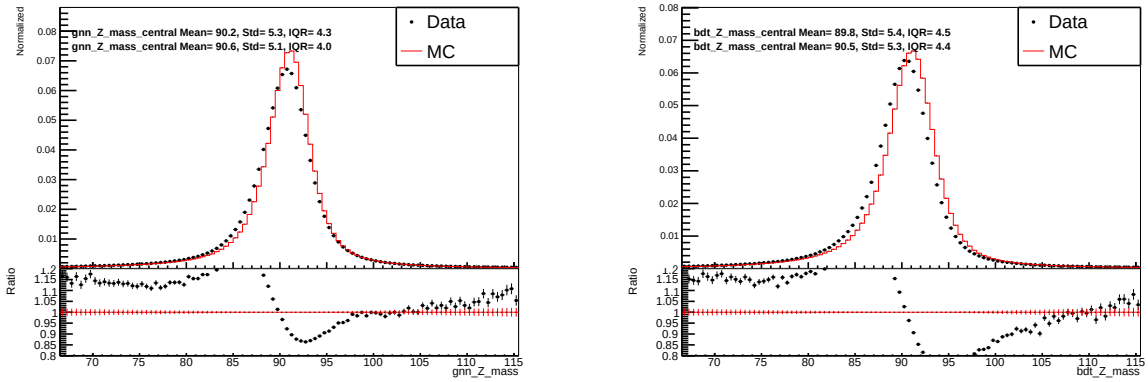


Figure B.8: Comparison of the m_{ee} spectra for the transformer (left) and BDT (right) between data and MC simulation for events with both electrons with $|\eta| < 1.37$.

the m_{ee} mass peak at $m_Z = 91$ GeV to extract the electron energy resolution and the bias in the energy calibration. This fit is performed in data and MC simulation for the transformer and the BDT to provide a full comparison of the algorithms' performance. Additionally, the fit is performed with different kinematic cuts to test the performance in different regions of the detector and different energy ranges. The $Z \to e^+e^-$ fits are performed on events from 2022 data taking containing two electrons reconstructed with $p_T > 27$ GeV and $|\eta| < 2.47$. The electrons are required to satisfy the medium ID and isolation working points. The tracks associated with the electrons must have a maximum longitudinal displacement from the primary vertex of 0.5 mm, and the statistical significance of this displacement must be less than 5σ. Events are required to pass at least one of the di-electron triggers, which have p_T thresholds well below 27 GeV. Additionally, the invariant mass of the di-electron system m_{ee} must be less than 180 GeV. The MC sample used for comparison is generated with PowhegBox for the matrix element generation and Pythia8 parton showering with the CTEQ6L1 PDF set and a pileup profile and detector conditions similar to those of 2022 data taking. The performance of the transformer is validated in this region using the simulated $Z \to e^+e^-$ MC, showing consistent resolution and bias with the single particle sample used in training across the relevant phase space. The calibrated m_{ee} spectra are shown in fig. B.8

Each m_{ee} spectrum is fit to the convolution of a non-relativistic Breit–Wigner distribution with a Crystal Ball function. The Breit–Wigner function is given a fixed pole mass of 91.19 GeV and a width of 2.50 GeV, corresponding to the mass and natural linewidth of the Z boson. The parameters of the Crystal Ball function are freely floated. The Crystal Ball is parameterized as

$$f(x;\alpha,n,\mu,\sigma) = N \begin{cases} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) & \frac{x-\mu}{\sigma} > -\alpha \\ A\left(B - \frac{x-\mu}{\sigma}\right)^{-n} & \frac{x-\mu}{\sigma} \leq -\alpha \end{cases}, \tag{B.3}$$

	$\sigma(m_{ee})$ [GeV]		$\langle m_{ee} \rangle - m_Z$ [GeV]	
Calibration	MC	Data	MC	Data
$\lvert\eta_e\rvert < 1.37$				
BDT	1.76	1.87	−0.09	−0.79
Transformer	1.54	1.80	−0.15	−0.51
$\lvert\eta_e\rvert > 1.52$				
BDT	1.91	2.23	−0.26	−0.90
Transformer	1.50	2.21	−0.17	−0.52

Table B.1: Resolution and bias in the measurement of m_{ee} for electrons in the Z peak evaluated in MC and data for the BDT and transformer energy calibrations extracted from fits to the m_{ee} spectrum separately considering events with both electrons in the barrel region and with both in the forward region.

where μ and σ describe the Gaussian core, α is the start point of the exponential tail, n is the exponent, and A, B, and N are calculated as functions of the floating parameters to ensure that the function is continuous, has a continuous derivative, and is properly normalized.

The results of these fits in the barrel and endcap regions are shown in figs. B.9 and B.10 and summarized in table B.1. As expected, the transformer shows a larger discrepancy in the resolution in data and MC than the BDT. However, the transformer generally has less bias and still improves the resolution when evaluated on data. The transformer results do not include the effects of a data-driven correction of the relative size of the energy deposited in each layer. Including this correction substantially reduces the discrepancies between data and MC simulation and improves the resolution achieved by the transformer in data.

B.4 Optimization of the Calibration Transformer

Several options have been explored to mitigate the observed discrepancy between the data and the MC simulation and maximize the transformer's performance in the real data. These efforts may be divided into attempts to change the simulated training dataset to better represent the real collision data and attempts to change the training procedure to reduce the observed discrepancy. In the former category, both classical and machine learning based approaches have been considered. One such approach based on machine learning uses the fact that the output of a binary classifier approximates the likelihood ratio as demonstrated in eq. (7.6). A classifier trained to discriminate calorimeter showers generated by MC simulation from those recorded in data is used to define a reweighting. The reweighting reduces the differences between showers in data and simulation, but the reweighting procedure reduces

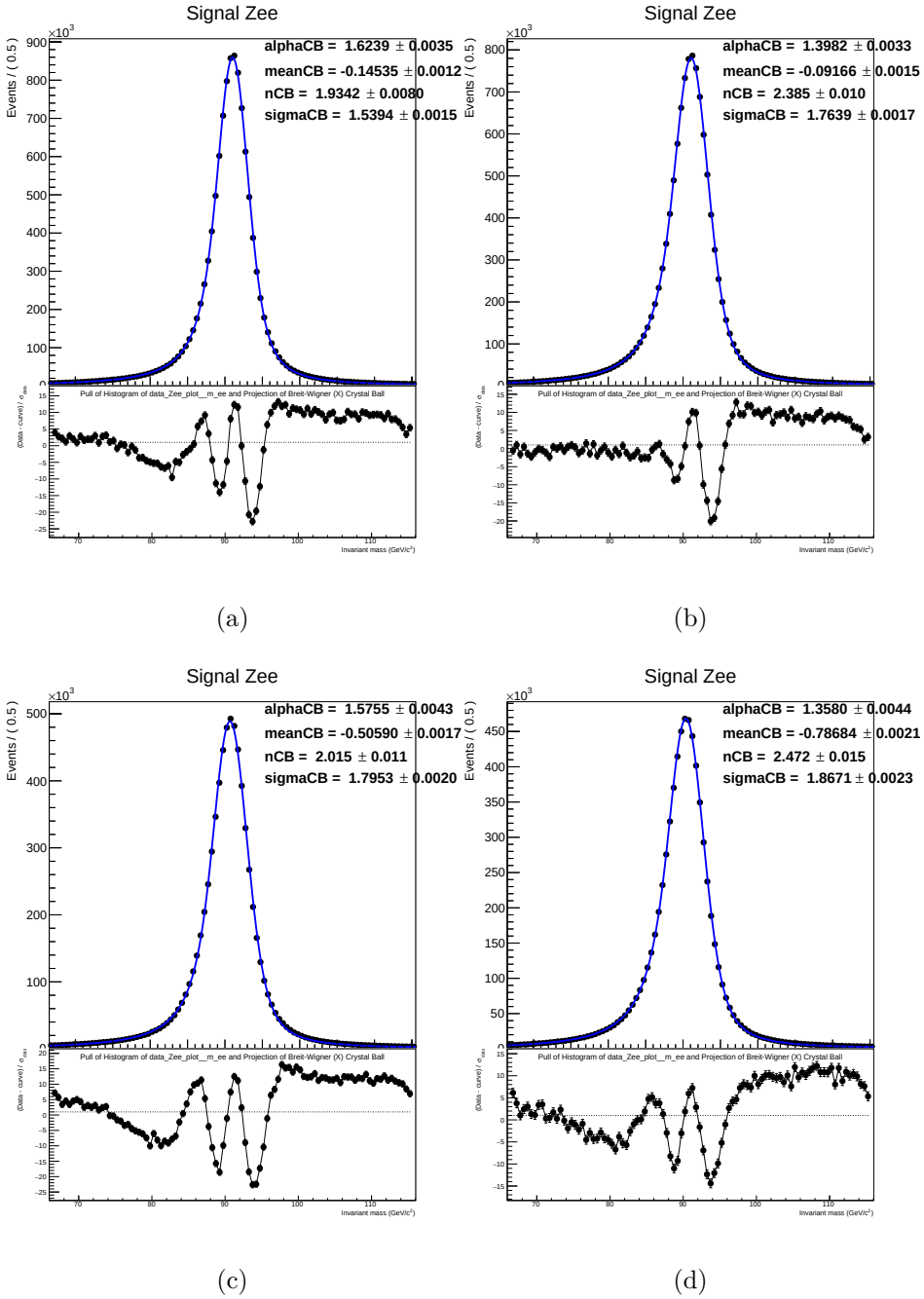


Figure B.9: Fits of the Z peak for electrons in MC simulation (a,b) and data (c,d) calibrated with the BDT (b,d) and the transformer (a,c) in the barrel region. Both electrons are required to have $|\eta| < 1.37$. The fitting function is the convolution of a Breit–Wigner function representing the natural lineshape of the Z boson with a Crystal Ball function representing the effects of the detector resolution on the measurement. The 'meanCB' and 'sigmaCB' represent the shift in the m_{ee} peak relative to $m_Z = 91.19$ GeV and the broadening of the lineshape in addition to the natural linewidth of 2.5 GeV.

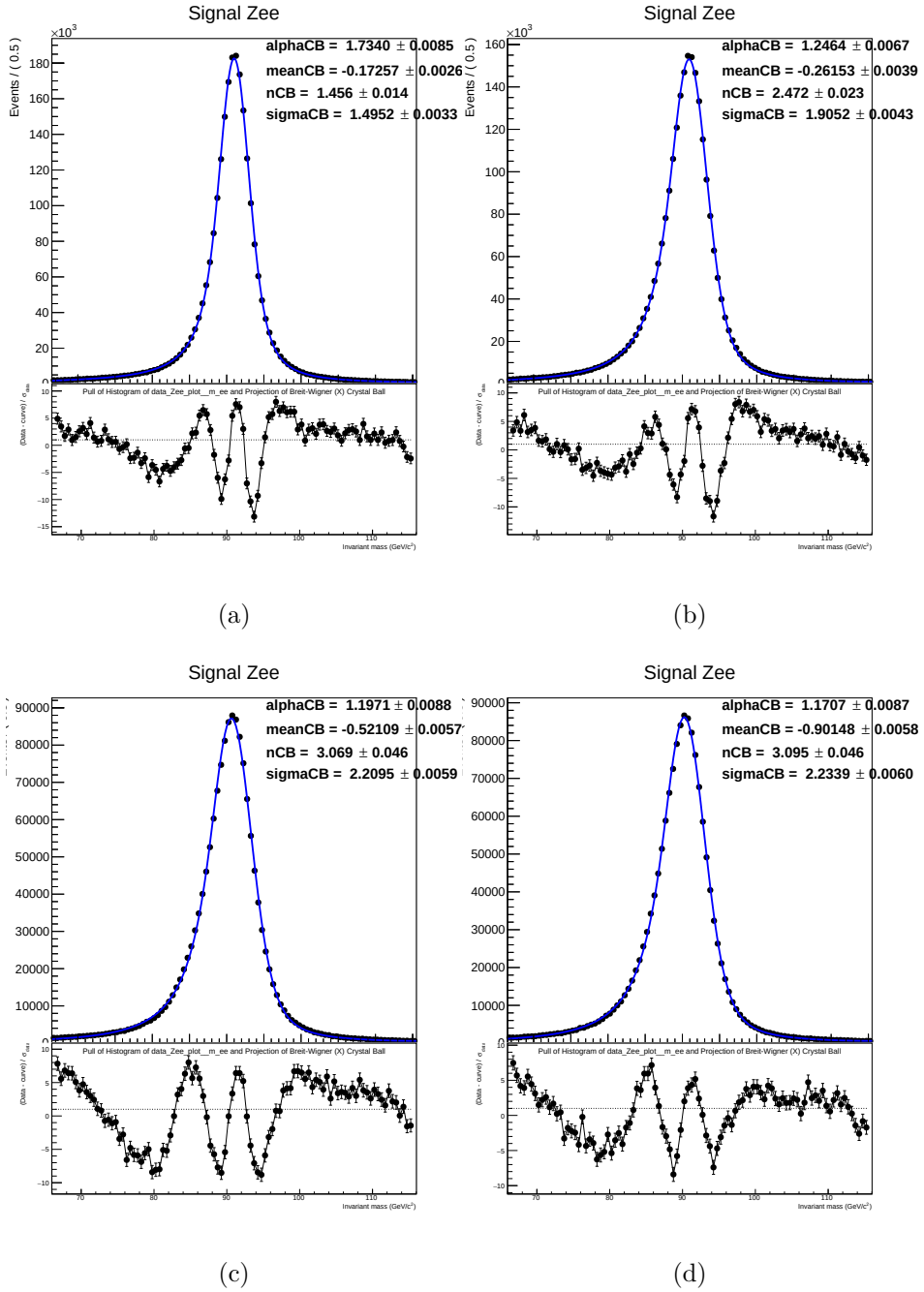


Figure B.10: Fits of the Z peak for electrons in MC simulation (a,b) and data (c,d) calibrated with the BDT (b,d) and the transformer (a,c) in the endcap region. Both electrons are required to have $|\eta| > 1.80$. The fitting function is the convolution of a Breit–Wigner function representing the natural lineshape of the Z boson with a Crystal Ball function representing the effects of the detector resolution on the measurement. The displayed parameters are those of the Crystal Ball function. The 'meanCB' and 'sigmaCB' represent the shift in the m_{ee} peak relative to $m_Z = 91.19$ GeV and the broadening of the lineshape in addition to the natural linewidth of 2.5 GeV.

the statistical power of the training dataset by several orders of magnitude, which renders the approach impractical for training the energy calibration.

In the second class of solutions, two approaches have been tried to produce a model robust to differences in the underlying distributions between data and simulation. The first generates adversarial examples using the gradient of the loss function to generate small perturbations of the training data that maximize the loss. This is done with the fast gradient sign method (FGSM) [286]. In principle, training the transformer on adversarial examples generated with the FGSM should result in a training that is robust to perturbations in the underlying distribution; however, in practice, trainings performed this way have not achieved competitive energy resolutions. The second adversarial approach directly incorporates the data into the training process using an adversarial network with gradient reversal as described in Ref. [287]. A second decoder of the pooled cell-level features with the same dimensions as the energy regression DNN is added to perform binary classification of data versus MC simulation. This network is trained with the binary cross-entropy loss using data and MC events satisfying the $Z \to e^+e^-$ selection criteria simultaneously with the regression DNN, which is still trained using the single particle simulation. Gradients from both networks are used to train the feature encoder and transformer networks, but the gradients from the adversarial classifier are given a factor of $-\alpha$ for $\alpha > 0$ so that the representation produced by the transformer is one in which data and simulation have minimal differences for the adversarial network to learn. This approach has been tried with a range of hyperparameters, but no choice achieves a meaningful difference in the comparison between data and MC over what is shown in table B.1. The setups resulting in the best resolution in data are generally those with the best resolution in MC, regardless of the size of the adversarial term.

B.5 Outlook for the e/γ Energy Calibration Transformer

The transformer-based e/γ energy calibration can bring a resolution improvement of at least 20% across the relevant kinematic phase space compared to the existing BDT-based calibration procedure. Realizing the improvement in physics capabilities implied by this development requires understanding complicated detector effects and mitigating the differences observed in MC simulations of the detector and real collision data. In-depth exploration of possible mitigations is ongoing. Before the transformer can be implemented as the default e/γ energy calibration algorithm in ATLAS, the studies of data/MC agreement are expected to be extended beyond the phase space accessible in $Z \to e^+e^-$ events to include low energy electrons for J/ψ decays and photons from radiative $Z \to \ell\ell\gamma$ decays.

Appendix C

ITk Tracker Upgrade for the HL-LHC

C.1 HL-LHC and Phase 2 Detector Upgrades

The success of the LHC program in discovering the Higgs boson and probing physics up to the TeV scale has illustrated the power of high-energy proton collisions as a tool for particle physics. An ambitious upgrade program for both the LHC and the detectors is underway. The LHC upgrade, known as the high luminosity LHC (HL-LHC) [288] will increase the number of protons per bunch and decrease the characteristic beam size at the interaction point β^*. The peak luminosity is expected to exceed 5×10^{34} cm^{-2}s^{-1} and the nominal pileup setting for ATLAS and CMS will be $\langle\mu\rangle = 140$ interaction per bunch crossing during luminosity leveling with the possibility of operating at up to $\langle\mu\rangle = 200$. The program is expected to run from 2029–2042 and result in a total integrated luminosity of 3000 fb^{-1} each delivered to ATLAS and CMS at a center of mass energy $\sqrt{s} = 14$ TeV.

This large dataset is expected to significantly expand the physics reach of most areas of the LHC physics program beyond the current frontiers. Naturally, many measurements of rare electroweak and Higgs boson processes currently limited by the statistical power of

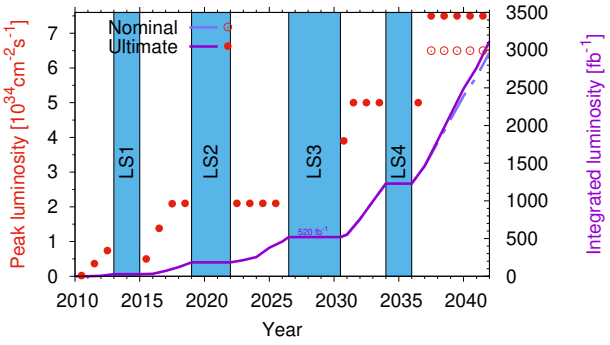


Figure C.1: Baseline run schedule and expected luminosity for the HL-LHC program.

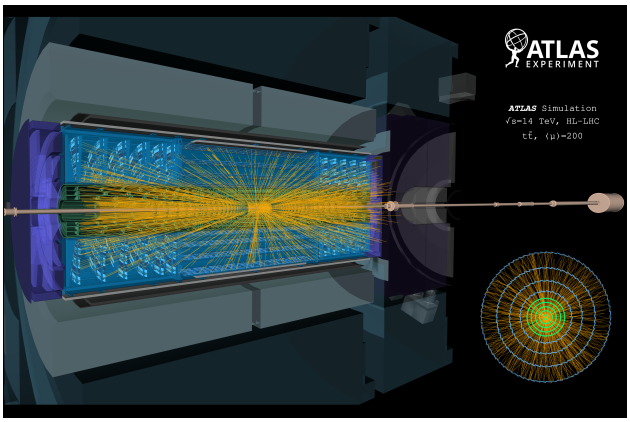


Figure C.2: Event display of the ITk detector for a simulated $t\bar{t}$ event in ATLAS during HL-LHC operations at $\langle\mu\rangle = 200$.

the collected data stand to improve significantly from the more than tenfold increase in integrated luminosity. ATLAS and CMS have prepared summaries of the expected physics reach of the full HL-LHC dataset [289]. The relative uncertainty on the measurement of $\sigma_{t\bar{t}t\bar{t}}$, for example, is expected to reach 6%. Such precision will definitively verify or reject the currently observed excess, and, if the discrepancy with the SM prediction remains, allow for a differential measurement of $t\bar{t}t\bar{t}$ production which could distinguish between possible BSM sources of the anomaly. Similar cases may be made for many other rare SM processes and theoretically motivated BSM scenarios.

The collision conditions of the HL-LHC, the expected data rates, and the long duration of the experimental program all far surpass the capabilities of the current detectors. In response, all four major LHC collaborations have prepared a series of detector upgrades, known as the Phase 2 upgrades, to prepare the detectors to handle the high data rates and occupancies, large instantaneous and integrated radiation doses, and dense population of particles from pileup interactions. These upgrades affect every detector subsystem and the trigger and DAQ systems. ATLAS has also prepared reports describing the detector upgrades [290] and software and computing needs [291] for the HL-LHC program. This work will focus on the replacement of the ATLAS Inner Detector with a new all-silicon inner tracker known as ITk and the author's contributions to the preparation for the production and operation of the silicon strips component of the ITk detector.

ITk: A New Tracker for ATLAS

The inner tracker must be completely replaced as the large quantity of radiation damage received in Runs 1–3 will put parts of the detector near the ends of their lifetimes. No part of the detector can survive the significantly increased radiation damage of the HL-LHC, and the front-end electronics cannot handle the occupancy or readout rate required for the

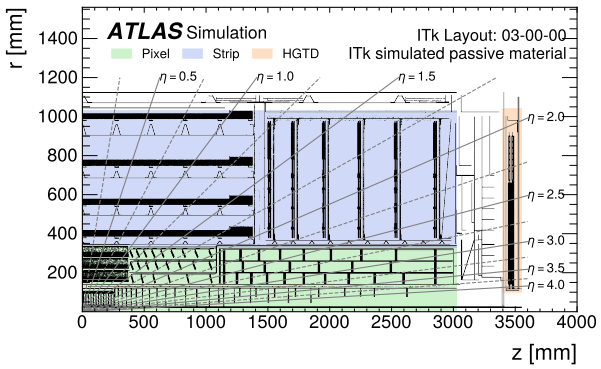


Figure C.3: Layout of the new all-silicon tracking system for the ATLAS experiment. This figure shows the location of passive material projected onto the longitudinal plane. The location of the ITk Pixel (in green) and the ITk Strip detectors (in blue) are shown, alongside the High-Granularity Timing Detector (in orange).

Phase 2 program. The gaseous detector technology of the TRT is not suited to the high occupancy of the HL-LHC and will therefore be replaced with an extension of the silicon tracking detector. Additionally, the tracking and vertexing performance needed to perform pileup mitigation and reconstruct tracks in dense environments surpass the capabilities of the current detector. The design specifications for the ITk include physics-motivated targets for tracking performance, engineering constraints on size, thermal and mechanical properties, and radiation tolerance, and DAQ-motivated requirements on the readout rate and latency.

The ITk is designed with nine layers of silicon pixel and strip sensors arranged at radii from 33–100 mm from the beam pipe. The design, shown schematically in fig. C.3, will extend tracking capabilities to $|\eta| < 4.0$ from the $|\eta| < 2.5$ currently covered by the ID. This will improve the reconstruction of forward jets found in many vector boson scattering and fusion processes and the acceptance for electrons. In total, the ITk will have 178 m^2 of active silicon sensors and a total of five billion readout channels. The pixel pitch is significantly reduced from the pitch used in the current pixel detector, and the 3D sensor technology employed for the IBL is again used for the innermost layer to withstand radiation doses of over 1.7 GRad. This work will focus on the strip detector [292], which uses double layers of silicon microstrip detectors with a small stereo angle of 26 mrad as in the SCT. In the strips detector, the front end electronics and sensors must survive total radiation doses of 50 MRad and displacement doses equivalent to 12×10^{14} neutrons per cm^2 at 100 MeV. The detector must support an average readout rate of 1 MHz and a latency of less than 25 μs.

ITk Strips Modules

The sensors used in the barrel have a pitch of 75.5 μm with 1280 strips in a row. Some sensors have two rows of short 24.10 mm strips, and others have two rows of long 42.80 mm

strips. In the endcap, the circular geometry requires more variation in the strip shape, but is generally chosen to closely match the strip pitch and length of the barrel modules. The front-end electronics responsible for powering and reading out each sensor are glued onto the sensor with an electrically resistive glue, and the strips are connected to the front-end readout electronics with wirebonds. This combination of sensor and readout electronics is known as a module and comprises the fundamental unit of the detector construction. The electronics are divided into two PCBs, the first of which is the powerboard responsible for monitoring and switching the high and low voltage connections to the sensor and front end electronics, and for converting the 12V low voltage input to the 1.2V needed by the front end electronics. The second is the hybrid, which hosts the front-end electronics to collect, digitize, and read out the signals from the silicon strips. In the case of short strip modules in the barrel, which have 2,560 rather than 1,280 strips, two hybrids are needed to read out each channel.

The main functions of the PCBs are performed by three flavors of application-specific integrated circuits (ASICs), which are custom-designed for the needs of the ITk strips detector. The Autonomous Monitoring and Control chip (AMAC) provides the logical functionality of the powerboard with monitoring and interlocking of the low voltage connection to the hybrid and the high voltage connection to the sensor. The ATLAS Binary Chip (ABC) is a front-end chip that processes signals from up to 256 strips into a single binary readout. The binary readout for each channel marks whether the signal has crossed a per-strip adjustable threshold. The threshold may be tuned using an on-chip charge injection circuit. Events with hits are organized into a cluster-based data format and stored in a buffer to await trigger commands, which cause the ABC to release the data. Each hybrid PCB has 10 ABC chips, each responsible for a 128×2 strip region of the sensor. The Hybrid Control Chip (HCC) controls the digital communication between the hybrid and the larger detector structures. A single HCC on each hybrid distributes clock and command signals from the larger detector system to the ABC chips and combines the ABC output, including triggered data and responses to commands, into a single data stream.

The remainder of this chapter details the author's contributions to an early version of the data acquisition (DAQ) software that will be used to operate the ITk strips detector and test beams to study the impact of single event effects from radiation on ITk strips modules.

DAQ Development for the ITk Strips Detector

The ITk strips data acquisition software must be used in several scenarios with a range of flexibility and performance requirements. Module research and development requires maximum flexibility to enable all characteristics and functionality of the ASICs and sensors to be tested. Quality control procedures performed during the construction of modules and larger detector structures require a fast, reliable implementation of basic communications tests and calibration scans, which can be operated at many sites. Operating in detector collision conditions requires integration with the wider detector control systems and ATLAS data acquisition system while meeting strict requirements on latency, rate, and robustness.

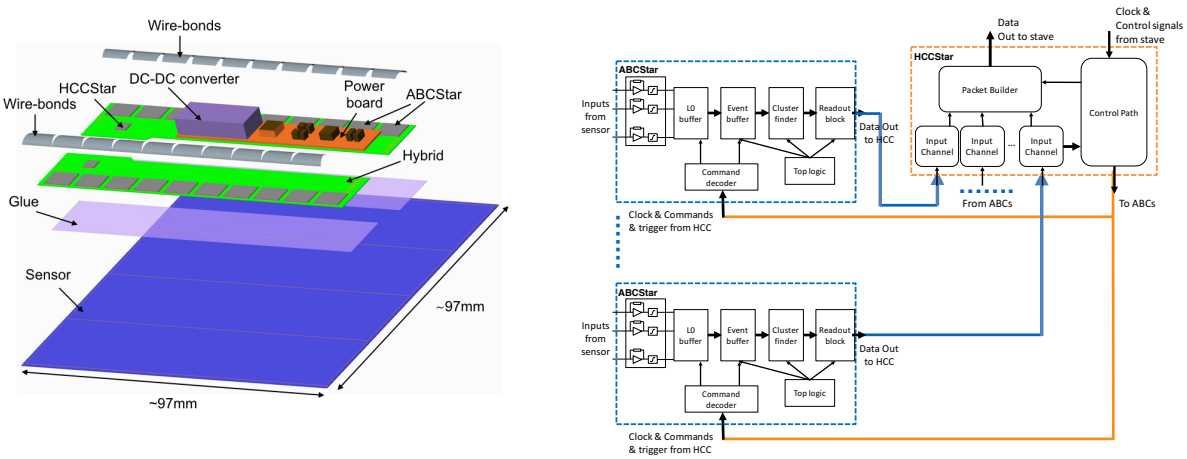


Figure C.4: Diagrams showing the construction of a short strip barrel module (left) and the flow of digital signals in a hybrid (right).

The same software should perform monitoring, calibration, and read-out operations. The author contributed to a migration from the `ITSDAQ` software used in the early R&D stage to the `YARR` software, which has since been integrated with the FELIX readout system [293] and can be used to operate large scale structures during integration activities or the entire ITk strips detector during operations. The development was performed using a partial module constructed with a single ABC chip connected to a "mini-sensor" consisting of 128 strips roughly 1 cm in length.

After initial development to make the existing `YARR` strips prototype compatible with the partial module, functionality was added to perform digital and analog quality control tests, read out the ABC and HCC registers, and perform the per-channel threshold calibration. The per-channel noise measurement, which requires sensitivity to the expected noise occupancy of $< 10^{-4}$, requires the readout of the module for a relatively large number of triggers. Operating by sending a single trigger and then waiting for the readout before sending the next causes this scan to take a substantial amount of time. Instead, an alternate data-taking procedure has been implemented based on the hit counters in the ABC chips. The FPGA that generates commands for the module is instructed to generate a batch of triggers, and then the hit counter registers are read out at the end of the loop. This change is illustrated in fig. C.6.

This procedure is fastest when sending more triggers per cycle before checking and refreshing the hit counters. However, the scan will fail if the hit counters overflow before being read out. The optimal point changes as a function of occupancy and must be adjusted throughout the scans. This is addressed with a feedback mechanism that throttles the trigger rate to keep the product of the per-trigger occupancy and the number of triggers per bunch safely below the maximum size of the hit counters. The flow of this mechanism is shown in fig. C.7.

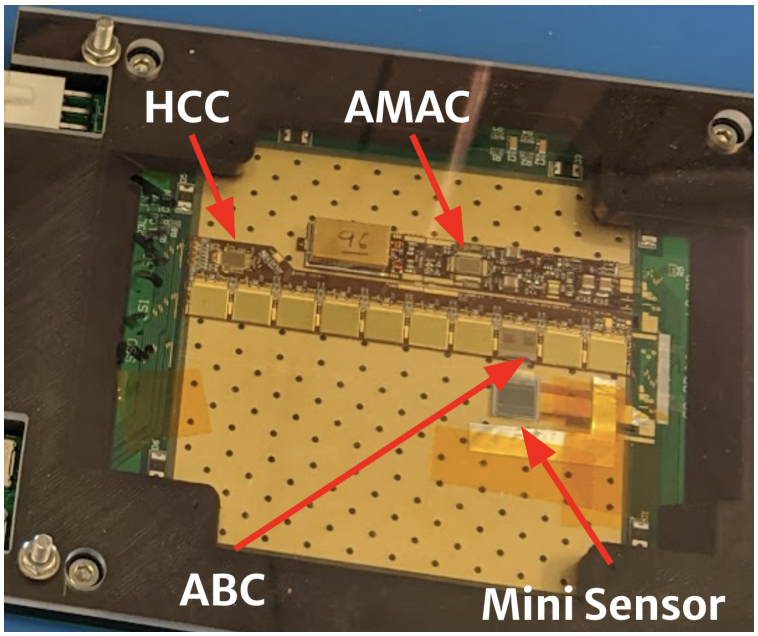


Figure C.5: Photograph of the partial module used for DAQ development.

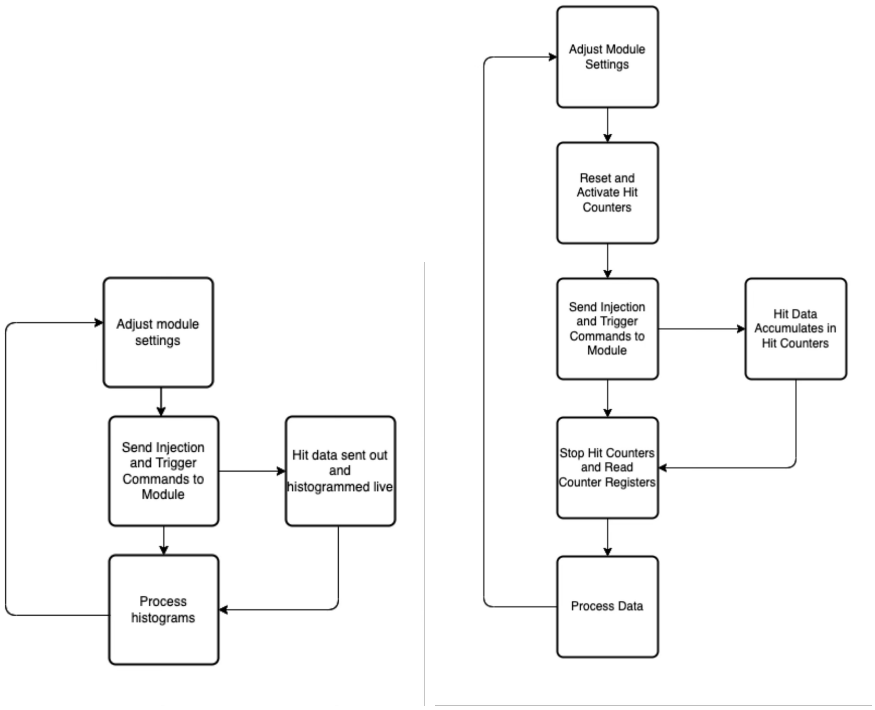


Figure C.6: (Left) Standard data taking flow for scans in `YARR`. (Right) Hit-counter-based data-taking flow.

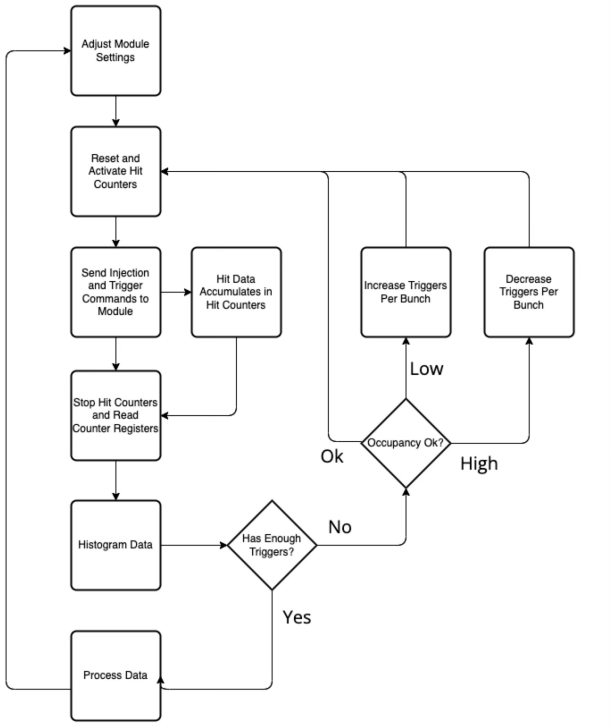


Figure C.7: Trigger throttling data flow.

The threshold calibration procedure adjusts the trim value of the set threshold for each channel in the ABC front end to achieve the desired occupancy for a fixed quantity of injected charge. In `ITSDAQ`, this is implemented as a scan over all possible values of the trim setting, followed by a fit to an error function to determine the trim value that produces the desired threshold. In `YARR`, the implementation is adapted from an existing setup used for the ITk pixel project that iteratively pushes the trim up or down if the occupancy is above or below the target value within some tolerance. The results of a trim of the partial module based on this approach are shown in fig. C.8.

The occupancies cluster around the desired value of 50%, but have a spread much larger than the target tolerance of 10% because of jitter in the individual steps of the iterative procedure. The efficacy of the threshold trim is evaluated by measuring the threshold value at which 50% occupancy is achieved for three different quantities of injected charge. This measurement is performed by scanning over the coarse threshold values while keeping the trims constant and measuring the occupancy at each step. The result is an S-curve like the one shown in fig. C.9, showing a steep drop off in the occupancy as the threshold passes above the size of the injected signal.

The gain values are extracted from a fit of the S-curve in each channel. The results of this fit are shown in fig. C.10. The values are compared between `YARR` and `ITSDAQ` to

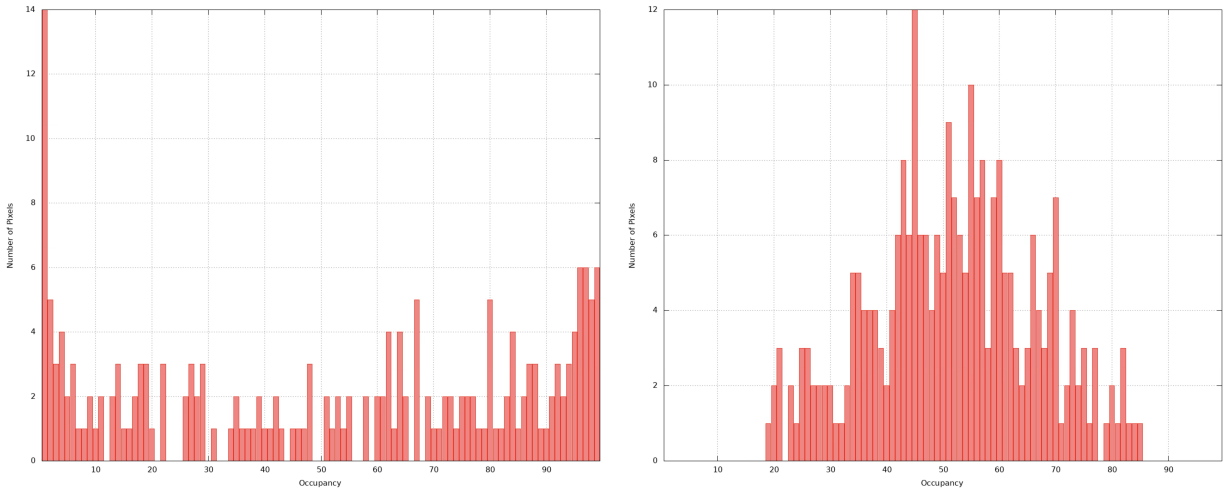


Figure C.8: Distribution of the channel occupancy in `YARR` before (left) and after (right) trimming the channel thresholds.

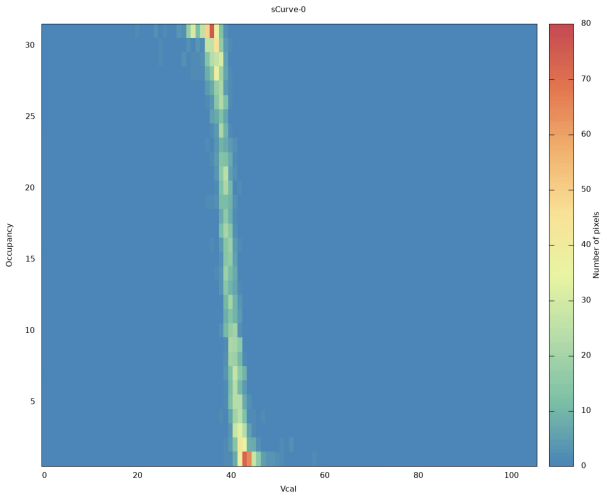


Figure C.9: Distribution of strip occupancies for different values of the charge threshold for a constant injected charge. The vertical axis shows the occupancy, and the horizontal axis shows the threshold in arbitrary units.

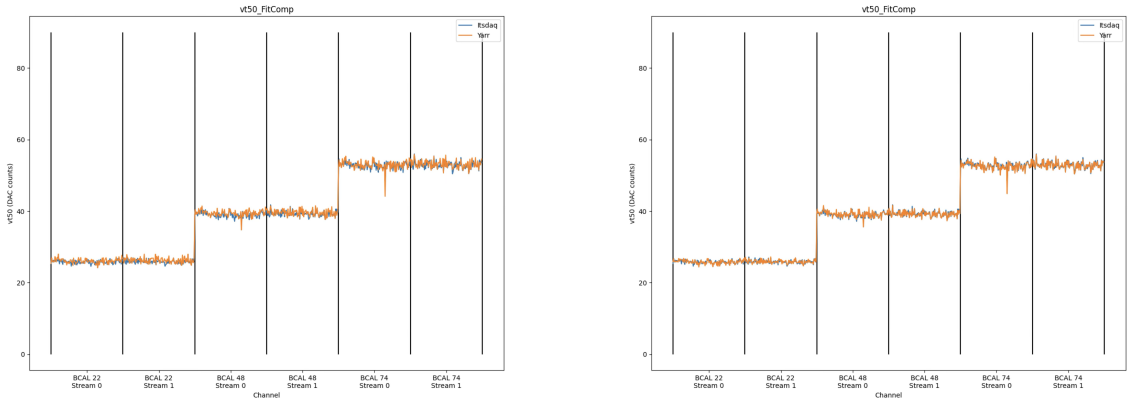


Figure C.10: Fitted thresholds compared between `YARR` and `ITSDAQ` using the trim values from each, respectively (left) and both using the trim values from `ITSDAQ` (right).

validate the procedure in `YARR` and are found to be generally consistent, with the same gain being measured and a similar level of fluctuation present between channels. The `YARR` scan shows slightly different gain values, which can be attributed to the differences between the trimming procedures. This is verified by repeating the gain measurement in `YARR` with the trim values determined from `ITSDAQ` injected, which results in near-perfect agreement in the measured threshold and noise values.

The basic functionality to operate ITk strips modules with `YARR` has since served as a foundation for the development of more advanced features and progress towards a fully functional DAQ system to operate the ITk strips detector.

C.2 SEE Test Beam Campaigns for the ITk Strips ASICs

In addition to the large integrated radiation doses, the ITk electronics must contend with high instantaneous rates, which can induce single event effects (SEEs) in the digital electronics. Single event effects may be divided into single event upsets (SEUs) in which one or more bits of an internal register are flipped, single event transients (SETs) in which a signal is induced in a digital logic pathway, and single event latch-ups (SELs) in which a transistor enters a permanently open state. The ITk strips ASICs have been designed to mitigate the effects of SEEs on the detector operations and data taking, primarily with a triplication scheme in which the chip clock, core logic, and critical registers are all triplicated. The results of logical operations use a majority voter method so that an SEE in one of the three pathways does not affect the chip's operation. Large buffers and data paths are not triplicated and are therefore susceptible to SEEs. Additionally, SEEs may occur in regions

of the chips responsible for communication with the other chips and therefore result in the transmission of invalid commands and data. These effects are mitigated by using the 6b8b and 8b10b encoding schemes, which allow error detection in addition to ensuring DC balance.

SEU effects have been studied in simulations of the ITk strips ASICs [294] but require verification *in situ* at test beam sites. Test beam campaigns for the AMAC, ABC, and HCC chips have been performed with both protons and heavy ions [295] to provide a precise measurement of the SEE rates as a function of the linear energy transfer (LET) imparted in an interaction and estimate the impact on the ITk strip detector in HL-LHC running conditions. The author has contributed to a set of test beam studies conducted in parallel with the ASIC tests to study module-level effects from SEEs. The remainder of this section describes that work, which is also presented in Ref. [296].

Module Level SEE Studies with BETSEE

The Board for the Evaluation of Single Event Effects (BETSEE) has been developed to allow the study of SEEs in the ITk strip system in a module-like environment with all three ASICs simultaneously in the beamspot as shown in fig. C.11. This parallel test provides additional redundancy to the SEE test campaign as the tests are performed with an independent hardware setup and monitoring software. Additionally, the module-like setup allows SEEs to be studied in a running configuration more similar to HL-LHC operations and affords sensitivity to possible interactions between ASICs, including the propagation of SEEs through the interchip communication lines.

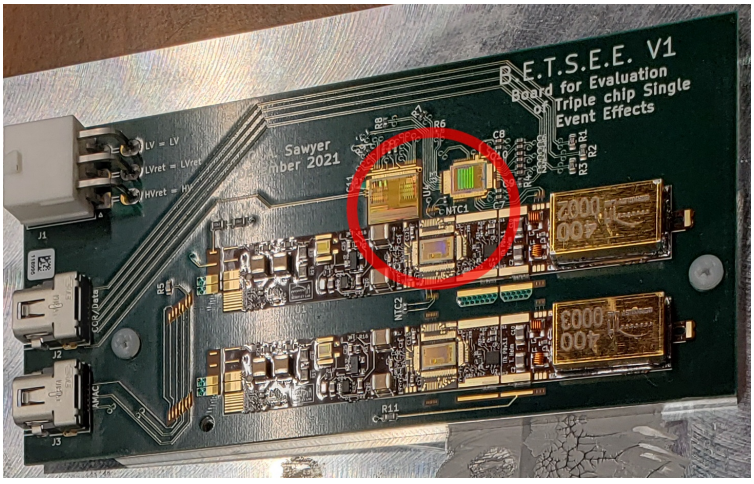


Figure C.11: Photograph of a BETSEE board mounted to the cooling plate used during heavy ion running. The red circle shows the rough location of the beam spot with the AMAC, ABC, and HCC inside.

The BETSEE layout consists of a custom PCB mounted with two powerboards. An

Ion	Energy [MeV]	LET [MeV/(mg/cm^2)]	Fluence [particles/cm^2]
$^{27}Al^{8+}$	250	5.7	3.00×10^7
$^{53}Cr^{16+}$	505	16.1	1.25×10^8
$^{84}Kr^{25+}$	769	32.4	1.50×10^8
p^+	480	N/A	6.35×10^{13}

Table C.1: Fluences, energies, and LET of different species used in test beam SEE runs with BETSEE. Proton collisions have a broad spectrum of LET rather than a narrow peak.

ABC and an HCC are mounted on the PCB near the AMAC of one powerboard, which is known as the primary powerboard. These three chips are connected as they would be on a module, but they must be arranged in this configuration rather than with the ABC and HCC on a hybrid to place all three chips in the 2.5 cm beam spot size available at the test beam facilities. The secondary powerboard is placed outside the beamspot and is used to test the OF functionality of the AMAC, which allows an AMAC to disable a neighboring module if communication with the neighbor is not possible. SEEs in the OF mechanism could potentially cause modules to be disabled during detector operations, so special care is taken to test that the secondary powerboard does not receive unexpected OF signals during operation.

BETSEE is operated with a portable setup controlled by a single small computer. The setup continually reads out the registers of all three ASICs under study to check for SEUs, monitors the status of the secondary powerboard, sends digital triggers to test the data readout, and monitors the chips' power consumption via interface with the power supply. Additional power usage and board temperature monitoring come from the AMAC. The ABC and HCC registers are read every two seconds, the digital scan is completed with the same frequency, and the module is reconfigured every minute to reset any lasting effects of SEEs. The raw data streams from the module are saved for later analysis, and summarized results are added to a database in real time. The module status and SEU rates are monitored in real time via a web interface using GrafAna. The setup has been run extensively in a lab environment to collect a baseline dataset and understand any non-SEE errors that occur when running the test loop used during test beam operations. Two test beam campaigns have been completed with BETSEE. The first, at the Heavy Ion Facility hosted at UC Louvain in February 2022, used ions of Al, Cr, and Kr and focused on the characterization of rare SEEs which may occur only at high LET. The second, at TRIUMF's Proton Irradiation Facility in summer 2022, used two BETSEE modules to double the effective luminosity and allowed the study of SEEs with a large fluence of particles with similar interactions to those that cause SEEs in LHC collisions. Table C.1 shows the collected fluence of each species during the two test beam campaigns.

BETSEE operated successfully throughout both test beam campaigns with no resets, power cycles, or permanent failure modes attributable to SEEs. SEUs are counted in the live

Chip	Observed Corrected SEUs (Kr)	σ_{SEU}(Kr) [cm^2/bit]	Observed Corrected SEUs (p^+)	$\sigma_{\text{SEU}}(p^+)$ [cm^2/bit]
ABC	31,225	2.5×10^{-7}	19,642	2.2×10^{-13}
HCC	5,749	3.4×10^{-7}	4,324	7.5×10^{-13}

Table C.2: Counts and cross-sections for corrected SEUs with protons and Krypton for ABC and HCC.

monitoring and the more detailed offline analysis, both by using the on-chip counters for SEU corrections in triplicated registers and by looking for changes in the values of non-triplicated registers. The observed SEU cross-sections for Krypton ions and protons are reported in table C.2 based on corrections in triplicated registers and are found to be consistent with the more precise measurements from tests of individual chips. Across the whole dataset in all three chips, no uncorrected SEUs in triplicated registers occurred, indicating that the protection of sensitive logic works as designed. The AMAC temperature measurements shown in fig. C.12 are stable after an initial warming period, indicating no sign of heat spikes from SELs.

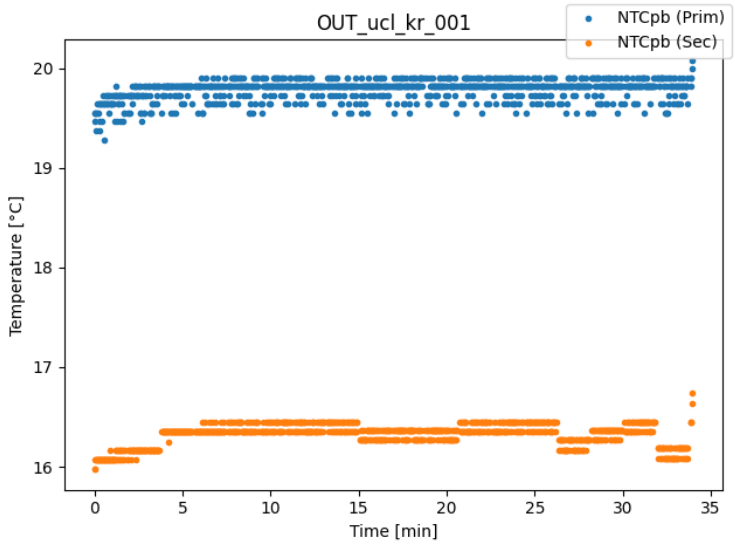


Figure C.12: Temperatures measured on two powerboards showing warm-up on start and stability during running.

While the online analysis and monitoring uses only data that can be successfully parsed, a more detailed offline inspection of the raw data output stream allows for a more complete understanding of any issues in BETSEE operation, including those potentially resulting in

lost or garbled data. The most common SEE after SEUs in the chip registers are bit flips in interchip communications, resulting in unparseable commands. The commands are 6b8b encoded and thus errors from a single bit flip are detectable but not correctable. The accumulation of these errors during running with Krypton is shown in fig. C.13. The errors sometimes occur in large bunches because an accumulation of errors can cause the synchronization of the communication to be lost. The ABC records many more communication errors than the HCC because all errors in commands passed to the HCC are forwarded to the ABC, and the ABC communications are sensitive to SEEs in the HCC output, the data path between the HCC and ABC, and the ABC input, whereas the HCC is only sensitive in the HCC input. The rate of these communication errors in HL-LHC conditions is expected to be minimal and have a negligible impact on detector operations and performance.

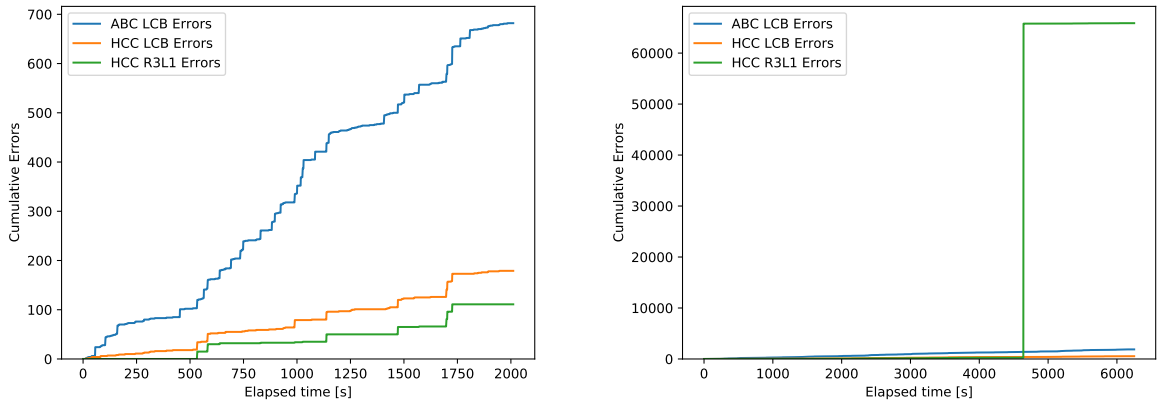


Figure C.13: Cumulative communications errors on ABC (blue) and HCC (orange and green) inputs during two runs with Kr. The run shown on the left shows normal behavior of the error counters. The run pictured on the right shows the sudden spike of errors associated with the R3L1 interruption.

The most significant disruption in BETSEE operations attributed to the beam was a five minute interruption in communications associated with the R3L1 input to the HCC, which is responsible for receiving some trigger signals. Detailed analysis of the data stream collected in this period reveals a three bit shift in the received R3L1 signal, which lasted for the duration of the interruption. The counter for R3L1 decoder errors filled rapidly after the error occurred, as shown in fig. C.13. During this time, no physics data was received, and the module did not respond to several regularly scheduled soft resets. After approximately five minutes, the module resumed normal operations asynchronously from any reset signal and without any direct intervention. This behavior has not been replicated in any other test beam campaign, in the lab without radiation, or in simulation. If it is a radiation effect, it must be exceedingly rare even at the LET deposited by Krypton ions and is therefore not

expected to affect HL-LHC operations. It is also possible that a module reset or power cycle could be used to recover from the issue quickly.

In total, the BETSEE test beam campaigns together with the single chip test beam campaigns show that the SEE mitigation in the design of the ITk strips ASICs works as intended and that the electronics can operate successfully in the difficult instantaneous radiation conditions of the HL-LHC. The excellent performance of the chips in these tests confirmed that the chip design was ready for mass production, which is now nearly complete.

Bibliography

[1] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Reading, USA: Addison-Wesley, 1995. ISBN: 978-0-429-49417-8. DOI: 10.1201/9780429503559.

[2] C. N. Yang and R. L. Mills. "Conservation of Isotopic Spin and Isotopic Gauge Invariance". In: *Phys. Rev.* 96 (1 Oct. 1954), pp. 191–195. DOI: 10.1103/PhysRev.96.191. URL: https://link.aps.org/doi/10.1103/PhysRev.96.191.

[3] Roberto D. Peccei. "The Strong CP Problem and Axions". In: *Axions*. Springer Berlin Heidelberg, 2008, pp. 3–17. ISBN: 9783540735182. DOI: 10.1007/978-3-540-73518-2_1. URL: http://dx.doi.org/10.1007/978-3-540-73518-2_1.

[4] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[5] Peter W. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Phys. Lett.* 12 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9.

[6] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[7] Peter W. Higgs. "Spontaneous Symmetry Breakdown without Massless Bosons". In: *Phys. Rev.* 145 (1966), pp. 1156–1163. DOI: 10.1103/PhysRev.145.1156.

[8] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.

[9] T. W. B. Kibble. "Symmetry Breaking in Non-Abelian Gauge Theories". In: *Phys. Rev.* 155 (1967), pp. 1554–1561. DOI: 10.1103/PhysRev.155.1554.

[10] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[11] S. Navas et al. "Review of particle physics". In: *Phys. Rev. D* 110.3 (2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001.

[12] Nathaniel Craig. *Naturalness: A Snowmass White Paper*. 2022. arXiv: 2205.05708 [hep-ph]. URL: https://arxiv.org/abs/2205.05708.

[13] Michael J. Ramsey-Musolf. "The electroweak phase transition: a collider target". In: *Journal of High Energy Physics* 2020.9 (Sept. 2020). ISSN: 1029-8479. DOI: 10.1007/jhep09(2020)179. URL: http://dx.doi.org/10.1007/JHEP09(2020)179.

[14] A. D. Sakharov. "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe". In: *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35. DOI: 10.1070/PU1991v034n05ABEH002497.

[15] F. Bezrukov and M. Shaposhnikov. "Why should we care about the top quark Yukawa coupling?" In: *Journal of Experimental and Theoretical Physics* 120.3 (Mar. 2015), pp. 335–343. ISSN: 1090-6509. DOI: 10.1134/s1063776115030152. URL: http://dx.doi.org/10.1134/S1063776115030152.

[16] Steven Weinberg. "Chapter 1: Effective field theory, past and future". In: *Int. J. Mod. Phys. A* 31.06 (2016), p. 1630007. DOI: 10.1142/9789813108332_0001.

[17] Zi-Yu Dong et al. "The new formulation of higgs effective field Theory". In: *Journal of High Energy Physics* 2023.9 (Sept. 2023). ISSN: 1029-8479. DOI: 10.1007/jhep09(2023)101. URL: http://dx.doi.org/10.1007/JHEP09(2023)101.

[18] B. Grzadkowski et al. "Dimension-six terms in the Standard Model Lagrangian". In: *Journal of High Energy Physics* 2010.10 (Oct. 2010). ISSN: 1029-8479. DOI: 10.1007/jhep10(2010)085. URL: http://dx.doi.org/10.1007/JHEP10(2010)085.

[19] Gino Isidori, Felix Wilsch, and Daniel Wyler. *The Standard Model effective field theory at work*. 2023. arXiv: 2303.16922 [hep-ph]. URL: https://arxiv.org/abs/2303.16922.

[20] Christopher W. Murphy. "Dimension-8 operators in the Standard Model Effective Field Theory". In: *Journal of High Energy Physics* 2020.10 (Oct. 2020). ISSN: 1029-8479. DOI: 10.1007/jhep10(2020)174. URL: http://dx.doi.org/10.1007/JHEP10(2020)174.

[21] D. de Florian et al. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector". In: (2016). DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].

[22] Federico Demartin et al. "Higgs characterisation at NLO in QCD: CP properties of the top-quark Yukawa interaction". In: *The European Physical Journal C* 74.9 (Sept. 2014). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-014-3065-2. URL: http://dx.doi.org/10.1140/epjc/s10052-014-3065-2.

[23] L. A. Harland-Lang et al. "Parton distributions in the LHC era: MMHT 2014 PDFs". In: *Eur. Phys. J. C* 75.5 (2015), p. 204. DOI: 10.1140/epjc/s10052-015-3397-6. arXiv: 1412.3989 [hep-ph].

[24] H. L. Lai et al. "Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions". In: *Eur. Phys. J. C* 12 (2000), p. 375. DOI: 10.1007/s100529900196. arXiv: hep-ph/9903282.

[25] NNPDF Collaboration, Richard D. Ball, et al. "Parton distributions for the LHC run II". In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[26] Toichiro Kinoshita. "Mass Singularities of Feynman Amplitudes". In: *Journal of Mathematical Physics* 3.4 (July 1962), pp. 650–677. ISSN: 0022-2488. DOI: 10.1063/1.1724268. eprint: https://pubs.aip.org/aip/jmp/article-pdf/3/4/650/19167464/650_1_online.pdf. URL: https://doi.org/10.1063/1.1724268.

[27] T. D. Lee and M. Nauenberg. "Degenerate Systems and Mass Singularities". In: *Phys. Rev.* 133 (6B Mar. 1964), B1549–B1562. DOI: 10.1103/PhysRev.133.B1549. URL: https://link.aps.org/doi/10.1103/PhysRev.133.B1549.

[28] R. Atkin. "Review of jet reconstruction algorithms". In: *J. Phys. Conf. Ser.* 645.1 (2015). Ed. by Alan S. Cornell and Bruce Mellado, p. 012008. DOI: 10.1088/1742-6596/645/1/012008.

[29] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[30] Rikkert Frederix and Stefano Frixione. "Merging meets matching in MC@NLO". In: *JHEP* 12 (2012), p. 061. DOI: 10.1007/JHEP12(2012)061. arXiv: 1209.6215 [hep-ph].

[31] Leif Lönnblad and Stefan Prestel. "Matching tree-level matrix elements with interleaved showers". In: *JHEP* 03 (2012), p. 019. DOI: 10.1007/JHEP03(2012)019. arXiv: 1109.4829 [hep-ph].

[32] Stefan Höche et al. "A critical appraisal of NLO+PS matching methods". In: *JHEP* 09 (2012), p. 049. DOI: 10.1007/JHEP09(2012)049. arXiv: 1111.1220 [hep-ph].

[33] Bo Andersson et al. "Parton fragmentation and string dynamics". In: *Phys. Rept.* 97 (1983), pp. 31–145. DOI: 10.1016/0370-1573(83)90080-7.

[34] Jan-Christopher Winter, Frank Krauss, and Gerhard Soff. "A modified cluster - hadronisation model". In: *Eur. Phys. J. C* 36 (2004), pp. 381–395. DOI: 10.1140/epjc/s2004-01960-8. arXiv: hep-ph/0311085.

[35] Christian Bierlich et al. *A comprehensive guide to the physics and usage of PYTHIA 8.3*. 2022. arXiv: 2203.11601 [hep-ph]. URL: https://arxiv.org/abs/2203.11601.

[36] P. Ferreira da Silva. "Physics of the Top Quark at the LHC: An Appraisal and Outlook of the Road Ahead". In: *Annual Review of Nuclear and Particle Science* 73.Volume 73, 2023 (2023), pp. 255–284. ISSN: 1545-4134. DOI: https://doi.org/10.1146/annurev-nucl-102419-052854. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-nucl-102419-052854.

[37] ATLAS Collaboration. *Standard Model Summary Plots June 2024*. ATL-PHYS-PUB-2024-011. 2024. URL: https://cds.cern.ch/record/2903866.

[38] Vernon Barger, Wai-Yee Keung, and Brian Yencho. "Triple-top signal of new physics at the LHC". In: *Phys. Lett. B* 687 (2010), p. 70. DOI: 10.1016/j.physletb.2010.03.001. arXiv: 1001.0221 [hep-ph].

[39] Chuan-Ren Chen. "Searching for new physics with triple-top signal at the LHC". In: *Phys. Lett. B* 736 (2014), p. 321. DOI: 10.1016/j.physletb.2014.07.041.

[40] Eduard Boos and Lev Dudko. "Triple top quark production in standard model". In: *Int. J. Mod. Phys. A* 37.05 (2022), p. 2250023. DOI: 10.1142/S0217751X22500233. arXiv: 2107.07629 [hep-ph].

[41] Rikkert Frederix, Davide Pagani, and Marco Zaro. "Large NLO corrections in $t\bar{t}W^{\pm}$ and $t\bar{t}t\bar{t}$ hadroproduction from supposedly subleading EW contributions". In: *JHEP* 02 (2018), p. 031. DOI: 10.1007/JHEP02(2018)031. arXiv: 1711.02116 [hep-ph].

[42] G. Bevilacqua and M. Worek. "Constraining BSM Physics at the LHC: Four top final states with NLO accuracy in perturbative QCD". In: *JHEP* 07 (2012), p. 111. DOI: 10.1007/JHEP07(2012)111. arXiv: 1206.3064 [hep-ph].

[43] Tomáš Ježo and Manfred Kraus. "Hadroproduction of four top quarks in the powheg box". In: *Phys. Rev. D* 105.11 (2022), p. 114024. DOI: 10.1103/PhysRevD.105.114024. arXiv: 2110.15159 [hep-ph].

[44] Melissa van Beekveld, Anna Kulesza, and Laura Moreno Valero. "Threshold resummation for the production of four top quarks at the LHC". In: (Dec. 2022). arXiv: 2212.03259 [hep-ph].

[45] Qing-Hong Cao et al. "Limiting top quark-Higgs boson interaction and Higgs-boson width from multitop productions". In: *Phys. Rev. D* 99.11 (2019), p. 113003. DOI: 10.1103/PhysRevD.99.113003. arXiv: 1901.04567 [hep-ph].

[46] Qing-Hong Cao, Shao-Long Chen, and Yandong Liu. "Probing Higgs width and top quark Yukawa coupling from $t\bar{t}H$ and $t\bar{t}t\bar{t}$ productions". In: *Phys. Rev. D* 95.5 (2017), p. 053004. DOI: 10.1103/PhysRevD.95.053004. arXiv: 1602.01934 [hep-ph].

[47] Hans Peter Nilles. "Supersymmetry, supergravity and particle physics". In: *Phys. Rept.* 110 (1984), pp. 1–162. DOI: 10.1016/0370-1573(84)90008-5.

[48] Glennys R. Farrar and Pierre Fayet. "Phenomenology of the production, decay, and detection of new hadronic states associated with supersymmetry". In: *Phys. Lett. B* 76 (1978), pp. 575–579. DOI: 10.1016/0370-2693(78)90858-4.

[49] Tilman Plehn and Tim M. P. Tait. "Seeking sgluons". In: *J. Phys. G* 36 (2009), p. 075001. DOI: 10.1088/0954-3899/36/7/075001. arXiv: 0810.3919 [hep-ph].

[50] Samuel Calvet et al. "Searching for sgluons in multitop events at a center-of-mass energy of 8 TeV". In: *JHEP* 04 (2013), p. 043. DOI: 10.1007/JHEP04(2013)043. arXiv: 1212.3360 [hep-ph].

[51] D. Dicus, A. Stange, and S. Willenbrock. "Higgs decay to top quarks at hadron colliders". In: *Phys. Lett. B* 333 (1994), pp. 126–131. DOI: 10.1016/0370-2693(94)91017-0. arXiv: hep-ph/9404359 [hep-ph].

[52] Nathaniel Craig et al. "The hunt for the rest of the Higgs bosons". In: *JHEP* 06 (2015), p. 137. DOI: 10.1007/JHEP06(2015)137. arXiv: 1504.04630 [hep-ph].

[53] Nathaniel Craig et al. "Heavy Higgs bosons at low $\tan\beta$: from the LHC to 100 TeV". In: *JHEP* 01 (2017), p. 018. DOI: 10.1007/JHEP01(2017)018. arXiv: 1605.08744 [hep-ph].

[54] Alex Pomarol and Javi Serra. "Top quark compositeness: Feasibility and implications". In: *Phys. Rev. D* 78 (7 Oct. 2008), p. 074026. DOI: 10.1103/PhysRevD.78.074026. arXiv: 0806.3247 [hep-ph].

[55] Christoph Englert et al. "The $\hat{H}$-parameter: an oblique Higgs view". In: *JHEP* 09 (2019), p. 041. DOI: 10.1007/JHEP09(2019)041. arXiv: 1903.07725 [hep-ph].

[56] Celine Degrande et al. "Non-resonant new physics in top pair production at hadron colliders". In: *JHEP* 03 (2011), p. 125. DOI: 10.1007/JHEP03(2011)125. arXiv: 1010.6304 [hep-ph].

[57] Cen Zhang. "Constraining *qqtt* operators from four-top production: a case for enhanced EFT sensitivity". In: *Chin. Phys. C* 42.2 (2018), p. 023104. DOI: 10.1088/1674-1137/42/2/023104. arXiv: 1708.05928 [hep-ph].

[58] Giovanni Banelli et al. "The present and future of four top operators". In: *JHEP* 02 (2021), p. 043. DOI: 10.1007/JHEP02(2021)043. arXiv: 2010.05915 [hep-ph].

[59] Rafael Aoude et al. "Complete SMEFT predictions for four top quark production at hadron colliders". In: *JHEP* 10 (2022), p. 163. DOI: 10.1007/JHEP10(2022)163. arXiv: 2208.04962 [hep-ph].

[60] ATLAS Collaboration. "Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton–proton collisions at $\sqrt{s} = 13$TeV with the ATLAS detector". In: *Eur. Phys. J. C* 80 (2020), p. 1085. DOI: 10.1140/epjc/s10052-020-08509-3. arXiv: 2007.14858 [hep-ex].

[61] ATLAS Collaboration. "Measurement of the $t\bar{t}t\bar{t}$ production cross section in *pp* collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *JHEP* 11 (2021), p. 118. DOI: 10.1007/JHEP11(2021)118. arXiv: 2106.11683 [hep-ex].

[62] CMS Collaboration. "Evidence for four-top quark production in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: (Mar. 2023). arXiv: 2303.03864 [hep-ex].

[63] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716 (2012), p. 1. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[64] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Phys. Lett. B* 716 (2012), p. 30. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[65] ATLAS Collaboration. "Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector". In: *Eur. Phys. J. C* 75 (2015), p. 476. DOI: 10.1140/epjc/s10052-015-3685-1. arXiv: 1506.05669 [hep-ex]. Erratum: in: *Eur. Phys. J. C* 76 (2016), p. 152. DOI: 10.1140/epjc/s10052-016-3934-y.

[66] CMS Collaboration. "Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV". In: *Phys. Rev. D* 92 (2015), p. 012004. DOI: 10.1103/PhysRevD.92.012004. arXiv: 1411.3441 [hep-ex].

[67] ATLAS Collaboration. "Combined Measurement of the Higgs Boson Mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ Decay Channels with the ATLAS Detector Using $\sqrt{s} = 7, 8,$ and 13 TeV pp Collision Data". In: *Phys. Rev. Lett.* 131 (2023), p. 251802. DOI: 10.1103/PhysRevLett.131.251802. arXiv: 2308.04775 [hep-ex].

[68] CMS Collaboration. "A measurement of the Higgs boson mass in the diphoton decay channel". In: *Phys. Lett. B* 805 (2020), p. 135425. DOI: 10.1016/j.physletb.2020.135425. arXiv: 2002.06398 [hep-ex].

[69] ATLAS Collaboration. "A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery". In: *Nature* 607 (2022), pp. 52–59. DOI: 10.1038/s41586-022-04893-w. arXiv: 2207.00092 [hep-ex]. Erratum: in: *Nature* 612 (2022), E24. DOI: 10.1038/s41586-022-05581-5.

[70] ATLAS Collaboration. In: *Nature* 612 (2022), E24. DOI: 10.1038/s41586-022-05581-5.

[71] CMS Collaboration. "A portrait of the Higgs boson by the CMS experiment ten years after the discovery". In: *Nature* 607 (2022), pp. 60–68. DOI: 10.1038/s41586-022-04892-x. arXiv: 2207.00043 [hep-ex]. Erratum: in: *Nature* 623 (2023), E4. DOI: 10.1038/s41586-023-06164-8.

[72] CMS Collaboration. In: *Nature* 623 (2023), E4. DOI: 10.1038/s41586-023-06164-8.

[73] ATLAS and CMS Collaborations. "Evidence for the Higgs Boson Decay to a Z Boson and a Photon at the LHC". In: *Phys. Rev. Lett.* 132 (2024), p. 021803. DOI: 10.1103/PhysRevLett.132.021803. arXiv: 2309.03501 [hep-ex].

[74] ATLAS Collaboration. "A search for the dimuon decay of the Standard Model Higgs boson with the ATLAS detector". In: *Phys. Lett. B* 812 (2021), p. 135980. DOI: 10.1016/j.physletb.2020.135980. arXiv: 2007.07830 [hep-ex].

[75] CMS Collaboration. "Search for associated production of a Higgs boson and a single top quark in proton–proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. D* 99 (2019), p. 092005. DOI: 10.1103/PhysRevD.99.092005. arXiv: 1811.09696 [hep-ex].

[76] ATLAS Collaboration. "Measurements of WH and ZH production with Higgs boson decays into bottom quarks and direct constraints on the charm Yukawa coupling in 13 TeV pp collisions with the ATLAS detector". In: (2024). arXiv: 2410.19611 [hep-ex].

[77] ATLAS Collaboration. "Test of CP-invariance of the Higgs boson in vector-boson fusion production and its decay into four leptons". In: *JHEP* 05 (2024), p. 105. DOI: 10.1007/JHEP05(2024)105. arXiv: 2304.09612 [hep-ex].

[78] ATLAS Collaboration. "Test of CP Invariance in Higgs Boson Vector-Boson-Fusion Production Using the $H \to \gamma\gamma$ Channel with the ATLAS Detector". In: *Phys. Rev. Lett.* 131 (2023), p. 061802. DOI: 10.1103/PhysRevLett.131.061802. arXiv: 2208.02338 [hep-ex].

[79] CMS Collaboration. "Constraints on anomalous Higgs boson couplings to vector bosons and fermions from the production of Higgs bosons using the $\tau\tau$ final state". In: *Phys. Rev. D* 108 (2023), p. 032013. DOI: 10.1103/PhysRevD.108.032013. arXiv: 2205.05120 [hep-ex].

[80] CMS Collaboration. "Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state". In: *Phys. Rev. D* 104 (2021), p. 052004. DOI: 10.1103/PhysRevD.104.052004. arXiv: 2104.12152 [hep-ex].

[81] ATLAS Collaboration. "Measurement of the CP properties of Higgs boson interactions with τ-leptons with the ATLAS detector". In: *Eur. Phys. J. C* 83 (2023), p. 563. DOI: 10.1140/epjc/s10052-023-11583-y. arXiv: 2212.05833 [hep-ex].

[82] CMS Collaboration. "Measurement of the Higgs boson mass and width using the four-lepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: (2024). arXiv: 2409.13663 [hep-ex].

[83] CMS Collaboration. "Limits on the Higgs boson lifetime and width from its decay to four charged leptons". In: *Phys. Rev. D* 92 (2015), p. 072010. DOI: 10.1103/PhysRevD.92.072010. arXiv: 1507.06656 [hep-ex].

[84] Nikolas Kauer and Giampiero Passarino. "Inadequacy of zero-width approximation for a light Higgs boson signal". In: *JHEP* 08 (2012), p. 116. DOI: 10.1007/JHEP08(2012)116. arXiv: 1206.4803 [hep-ph].

[85] Fabrizio Caola and Kirill Melnikov. "Constraining the Higgs boson width with ZZ production at the LHC". In: *Phys. Rev. D* 88 (2013), p. 054024. DOI: 10.1103/PhysRevD.88.054024. arXiv: 1307.4935 [hep-ph].

[86] John M. Campbell, R. Keith Ellis, and Ciaran Williams. "Bounding the Higgs width at the LHC using full analytic results for $gg \to e^-e^+\mu^-\mu^+$". In: *JHEP* 04 (2014), p. 060. DOI: 10.1007/JHEP04(2014)060. arXiv: 1311.3589 [hep-ph].

[87] ATLAS Collaboration. "Measurement of off-shell Higgs boson production in the $H^* \to ZZ \to 4\ell$ decay channel using a neural simulation-based inference technique in 13 TeV pp collisions with the ATLAS detector". In: (2024). arXiv: 2412.01548 [hep-ex].

[88] CMS Collaboration. "Measurement of the Higgs boson width and evidence of its off-shell contributions to ZZ production". In: *Nature Phys.* 18 (2022), pp. 1329–1334. DOI: 10.1038/s41567-022-01682-0. arXiv: 2202.06923 [hep-ex].

[89] ATLAS Collaboration. "Constraining off-shell Higgs boson production and the Higgs boson total width using $WW \to \ell\nu\ell\nu$ final states with the ATLAS detector". In: (2025). arXiv: 2504.07710 [hep-ex].

[90] Malte Buschmann et al. "Mass effects in the Higgs-gluon coupling: boosted vs off-shell production". In: *JHEP* 02 (2015), p. 038. DOI: 10.1007/JHEP02(2015)038. arXiv: 1410.5806 [hep-ph].

[91] Christoph Englert and Michael Spannowsky. "Limitations and Opportunities of Off-Shell Coupling Measurements". In: *Phys. Rev. D* 90 (2014), p. 053003. DOI: 10.1103/PhysRevD.90.053003. arXiv: 1405.0285 [hep-ph].

[92] Qing-Hong Cao et al. "What can we learn from the total width of the Higgs boson?" In: *Chin. Phys. C* 47.3 (2023), p. 033101. DOI: 10.1088/1674-1137/aca8f6. arXiv: 2107.08343 [hep-ph].

[93] ATLAS Collaboration. "Measurement of the associated production of a top-antitop-quark pair and a Higgs boson decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ using the ATLAS detector at the LHC". In: (2024). arXiv: 2407.10904 [hep-ex].

[94] CMS Collaboration. "Measurement of the $t\bar{t}H$ and tH production rates in the $H \to b\bar{b}$ decay channel using proton–proton collision data at $\sqrt{s} = 13\,\text{TeV}$". In: (2024). arXiv: 2407.10896 [hep-ex].

[95] ATLAS Collaboration. "Measurement of the properties of Higgs boson production at $\sqrt{s} = 13\,\text{TeV}$ in the $H \to \gamma\gamma$ channel using $139\,\text{fb}^{-1}$ of pp collision data with the ATLAS experiment". In: *JHEP* 07 (2023), p. 088. DOI: 10.1007/JHEP07(2023)088. arXiv: 2207.00348 [hep-ex].

[96] ATLAS Collaboration. "Differential cross-section measurements of Higgs boson production in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: (2024). arXiv: 2407.16320 [hep-ex].

[97] CMS Collaboration. "Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J. C* 81 (2021), p. 378. DOI: 10.1140/epjc/s10052-021-09014-x. arXiv: 2011.03652 [hep-ex].

[98] *Top Quarks + X Summary Plots April 2024*. ATL-PHYS-PUB-2024-005. 2024. URL: https://cds.cern.ch/record/2896021/.

[99] CMS Collaboration. "Measurement of the top quark Yukawa coupling from $t\bar{t}$ kinematic distributions in the lepton+jets final state in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Phys. Rev. D* 100 (2019), p. 072007. DOI: 10.1103/PhysRevD.100.072007. arXiv: 1907.01590 [hep-ex].

[100] CMS Collaboration. "Search for production of four top quarks in final states with same-sign or multiple leptons in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J. C* 80 (2020), p. 75. DOI: 10.1140/epjc/s10052-019-7593-7. arXiv: 1908.06463 [hep-ex].

[101] Joachim Brod, Ulrich Haisch, and Jure Zupan. "Constraints on CP-violating Higgs couplings to the third generation". In: *Journal of High Energy Physics* 2013.11 (Nov. 2013). ISSN: 1029-8479. DOI: 10.1007/jhep11(2013)180. URL: http://dx.doi.org/10.1007/JHEP11(2013)180.

[102] Henning Bahl et al. "Constraining the $\mathcal{CP}$ structure of Higgs-fermion couplings with a global LHC fit, the electron EDM and baryogenesis". In: *The European Physical Journal C* 82.7 (July 2022). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-022-10528-1. URL: http://dx.doi.org/10.1140/epjc/s10052-022-10528-1.

[103] J. Baron et al. "Order of Magnitude Smaller Limit on the Electric Dipole Moment of the Electron". In: *Science* 343.6168 (Jan. 2014), pp. 269–272. ISSN: 1095-9203. DOI: 10.1126/science.1248213. URL: http://dx.doi.org/10.1126/science.1248213.

[104] CMS Collaboration. "Measurements of $t\bar{t}H$ Production and the CP Structure of the Yukawa Interaction between the Higgs Boson and Top Quark in the Diphoton Decay Channel". In: *Phys. Rev. Lett.* 125 (2020), p. 061801. DOI: 10.1103/PhysRevLett.125.061801. arXiv: 2003.10866 [hep-ex].

[105] ATLAS Collaboration. "CP Properties of Higgs Boson Interactions with Top Quarks in the $t\bar{t}H$ and tH Processes Using $H \to \gamma\gamma$ with the ATLAS Detector". In: *Phys. Rev. Lett.* 125 (2020), p. 061802. DOI: 10.1103/PhysRevLett.125.061802. arXiv: 2004.04545 [hep-ex].

[106] ATLAS Collaboration. "Probing the CP nature of the top-Higgs Yukawa coupling in $t\bar{t}H$ and tH events with $H \to b\bar{b}$ decays using the ATLAS detector at the LHC". In: *Phys. Lett. B* 849 (2024), p. 138469. DOI: 10.1016/j.physletb.2024.138469. arXiv: 2303.05974 [hep-ex].

[107] CMS Collaboration. "Search for CP violation in $t\bar{t}H$ and tH production in multilepton channels in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *JHEP* 07 (2023), p. 092. DOI: 10.1007/JHEP07(2023)092. arXiv: 2208.02686 [hep-ex].

[108] Lyndon Evans and Philip Bryant. "LHC Machine". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001. URL: https://dx.doi.org/10.1088/1748-0221/3/08/S08001.

[109] Oliver Brüning et al. "70 years at the high-energy frontier with the CERN accelerator complex". In: *Nature Rev. Phys.* 6.10 (2024), pp. 628–637. DOI: 10.1038/s42254-024-00758-5.

[110] The ATLAS Collaboration. "Improved luminosity determination in pp collisions at sqrt(s) = 7 TeV using the ATLAS detector at the LHC". In: *Eur. Phys. J. C* 73.8 (2013), p. 2518. DOI: 10.1140/epjc/s10052-013-2518-3. arXiv: 1302.4393 [hep-ex].

[111] The ATLAS Collaboration. Public Plots for 2012 Luminosity. Jan. 2025. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun1.

[112] The ATLAS Collaboration. "Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC". In: *The European Physical Journal C* 83.10 (Nov. 2023). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-023-11747-w. URL: http://dx.doi.org/10.1140/epjc/s10052-023-11747-w.

[113] The ATLAS Collaboration. Public Plots for Run 3 Luminosity. Jan. 2025. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3.

[114] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[115] ATLAS Collaboration. "The ATLAS Inner Detector commissioning and calibration". In: *Eur. Phys. J. C* 70 (2010), p. 787. DOI: 10.1140/epjc/s10052-010-1366-7. arXiv: 1004.5293 [physics.ins-det].

[116] ATLAS Collaboration. In: *Journal of Instrumentation* 3.07 (July 2008), P07007. DOI: 10.1088/1748-0221/3/07/P07007. URL: https://dx.doi.org/10.1088/1748-0221/3/07/P07007.

[117] ATLAS Collaboration. *ATLAS Insertable B-Layer: Technical Design Report*. ATLAS-TDR-19; CERN-LHCC-2010-013. 2010. URL: https://cds.cern.ch/record/1291633. Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009. 2012. URL: https://cds.cern.ch/record/1451888.

[118] B. Abbott et al. "Production and integration of the ATLAS Insertable B-Layer". In: *JINST* 13 (2018), T05008. DOI: 10.1088/1748-0221/13/05/T05008. arXiv: 1803.00844 [physics.ins-det].

[119] A. Ahmad et al. "The Silicon microstrip sensors of the ATLAS semiconductor tracker". In: *Nucl. Instrum. Meth. A* 578 (2007), pp. 98–118. DOI: 10.1016/j.nima.2007.04.157.

[120] T. Akesson et al. "ATLAS Transition Radiation Tracker test-beam results". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 522.1 (2004). TRDs for the Third Millenium. Proceedings of the 2nd Workshop on Advanced Transition Radiation Detectors for Accelerator and Space Applications, pp. 50–55. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2004.01.017. URL: https://www.sciencedirect.com/science/article/pii/S0168900204000841.

[121] ATLAS Collaboration. "Readiness of the ATLAS liquid argon calorimeter for LHC collisions". In: *Eur. Phys. J. C* 70 (2010), p. 723. DOI: 10.1140/epjc/s10052-010-1354-y. arXiv: 0912.2642 [physics.ins-det].

[122] ATLAS Collaboration. "Readiness of the ATLAS Tile Calorimeter for LHC collisions". In: *Eur. Phys. J. C* 70 (2010), p. 1193. DOI: 10.1140/epjc/s10052-010-1508-y. arXiv: 1007.5423 [physics.ins-det].

[123] ATLAS Collaboration. "Commissioning of the ATLAS Muon Spectrometer with cosmic rays". In: *Eur. Phys. J. C* 70 (2010), p. 875. DOI: 10.1140/epjc/s10052-010-1415-2. arXiv: 1006.4384 [physics.ins-det].

[124] G. Avoni et al. "The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS". In: *JINST* 13 (2018), P07017. DOI: 10.1088/1748-0221/13/07/P07017.

[125] ATLAS Collaboration. "Luminosity determination in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ using the ATLAS detector at the LHC". In: *Eur. Phys. J. C* 83 (2023), p. 982. DOI: 10.1140/epjc/s10052-023-11747-w. arXiv: 2212.09379 [hep-ex].

[126] ATLAS Collaboration. "Performance of the ATLAS trigger system in 2015". In: *Eur. Phys. J. C* 77 (2017), p. 317. DOI: 10.1140/epjc/s10052-017-4852-3. arXiv: 1611.09661 [hep-ex].

[127] ATLAS Collaboration. *The ATLAS Collaboration Software and Firmware*. ATL-SOFT-PUB-2021-001. 2021. URL: https://cds.cern.ch/record/2767187.

[128] ATLAS Collaboration. "Software and computing for Run 3 of the ATLAS experiment at the LHC". In: (2024). arXiv: 2404.06335 [hep-ex].

[129] ATLAS Collaboration. "Observation of four-top-quark production in the multilepton final state with the ATLAS detector". In: *Eur. Phys. J. C* 83 (2023), p. 496. DOI: 10.1140/epjc/s10052-023-11573-0. arXiv: 2303.15061 [hep-ex]. Erratum: in: *Eur. Phys. J. C* 84 (2024), p. 156. DOI: 10.1140/epjc/s10052-024-12458-6.

[130] ATLAS Collaboration. In: *Eur. Phys. J. C* 84 (2024), p. 156. DOI: 10.1140/epjc/s10052-024-12458-6.

[131] ATLAS Collaboration. *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV*. ATL-PHYS-PUB-2015-026. 2015. URL: https://cds.cern.ch/record/2037717.

[132] ATLAS Collaboration. "Measurement of Higgs boson production in the diphoton decay channel in *pp* collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector". In: *Phys. Rev. D* 90 (2014), p. 112015. DOI: 10.1103/PhysRevD.90.112015. arXiv: 1408.7084 [hep-ex].

[133] ATLAS Collaboration. "Muon reconstruction and identification efficiency in ATLAS using the full Run 2 *pp* collision data set at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 81 (2021), p. 578. DOI: 10.1140/epjc/s10052-021-09233-2. arXiv: 2012.00578 [hep-ex].

[134] ATLAS Collaboration. "Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data". In: *JINST* 14 (2019), P12006. DOI: 10.1088/1748-0221/14/12/P12006. arXiv: 1908.00005 [hep-ex].

[135] ATLAS Collaboration. "Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector". In: *Phys. Rev. D* 97 (2018), p. 072003. DOI: 10.1103/PhysRevD.97.072003. arXiv: 1712.08891 [hep-ex].

[136] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet user manual". In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[137] ATLAS Collaboration. "Jet reconstruction and performance using particle flow with the ATLAS Detector". In: *Eur. Phys. J. C* 77 (2017), p. 466. DOI: 10.1140/epjc/s10052-017-5031-2. arXiv: 1703.10485 [hep-ex].

[138] ATLAS Collaboration. "Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Eur. Phys. J. C* 81 (2021), p. 689. DOI: 10.1140/epjc/s10052-021-09402-3. arXiv: 2007.02645 [hep-ex].

[139] ATLAS Collaboration. *Tagging and suppression of pileup jets with the ATLAS detector*. ATLAS-CONF-2014-018. 2014. URL: https://cds.cern.ch/record/1700870.

[140] ATLAS Collaboration. "ATLAS flavour-tagging algorithms for the LHC Run 2 *pp* collision dataset". In: *Eur. Phys. J. C* 83 (2023), p. 681. DOI: 10.1140/epjc/s10052-023-11699-1. arXiv: 2211.16345 [physics.data-an].

[141] ATLAS Collaboration. *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*. ATL-PHYS-PUB-2017-013. 2017. URL: https://cds.cern.ch/record/2273281.

[142] ATLAS Collaboration. "Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 78 (2018), p. 903. DOI: 10.1140/epjc/s10052-018-6288-9. arXiv: 1802.08168 [hep-ex].

[143] ATLAS Collaboration. "ATLAS data quality operations and performance for 2015–2018 data-taking". In: *JINST* 15 (2020), P04003. DOI: 10.1088/1748-0221/15/04/P04003. arXiv: 1911.04632 [physics.ins-det].

[144] ATLAS Collaboration. "Performance of electron and photon triggers in ATLAS during LHC Run 2". In: *Eur. Phys. J. C* 80 (2020), p. 47. DOI: 10.1140/epjc/s10052-019-7500-2. arXiv: 1909.00761 [hep-ex].

[145] ATLAS Collaboration. "Performance of the ATLAS muon triggers in Run 2". In: *JINST* 15 (2020), P09015. DOI: 10.1088/1748-0221/15/09/p09015. arXiv: 2004.13447 [physics.ins-det].

[146] ATLAS Collaboration. "Operation of the ATLAS trigger system in Run 2". In: *JINST* 15 (2020), P10004. DOI: 10.1088/1748-0221/15/10/P10004. arXiv: 2007.12539 [physics.ins-det].

[147] ATLAS Collaboration. *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model.* ATL-PHYS-PUB-2016-017. 2016. URL: https://cds.cern.ch/record/2206965.

[148] ATLAS Collaboration. *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim.* ATL-PHYS-PUB-2010-013. 2010. URL: https://cds.cern.ch/record/1300517.

[149] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[150] Paolo Nason. "A new method for combining NLO QCD with shower Monte Carlo algorithms". In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146.

[151] Stefano Frixione, Giovanni Ridolfi, and Paolo Nason. "A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction". In: *JHEP* 09 (2007), p. 126. DOI: 10.1088/1126-6708/2007/09/126. arXiv: 0707.3088 [hep-ph].

[152] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with parton shower simulations: the POWHEG method". In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[153] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[154] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[155] Enrico Bothmann et al. "Event generation with Sherpa 2.2". In: *SciPost Phys.* 7.3 (2019), p. 034. DOI: 10.21468/SciPostPhys.7.3.034. arXiv: 1905.09127 [hep-ph].

[156] T. Gleisberg et al. "Event generation with SHERPA 1.1". In: *JHEP* 02 (2009), p. 007. DOI: 10.1088/1126-6708/2009/02/007. arXiv: 0811.4622 [hep-ph].

[157] M. Bähr et al. "Herwig++ physics and manual". In: *Eur. Phys. J. C* 58 (2008), p. 639. DOI: 10.1140/epjc/s10052-008-0798-9. arXiv: 0803.0883 [hep-ph].

[158] Johannes Bellm et al. "Herwig 7.0/Herwig++ 3.0 release note". In: *Eur. Phys. J. C* 76.4 (2016), p. 196. DOI: 10.1140/epjc/s10052-016-4018-8. arXiv: 1512.01178 [hep-ph].

[159] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nucl. Instrum. Meth. A* 462 (2001), p. 152. DOI: 10.1016/S0168-9002(01)00089-4.

[160] Piotr Golonka and Zbigniew Was. "PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays". In: *Eur. Phys. J. C* 45 (2006), pp. 97–107. DOI: 10.1140/epjc/s2005-02396-4. arXiv: hep-ph/0506026.

[161] S. Agostinelli et al. "GEANT4 – a simulation toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.

[162] Adam Alloul et al. "FeynRules 2.0 - A complete toolbox for tree-level phenomenology". In: *Comput. Phys. Commun.* 185 (2014), pp. 2250–2300. DOI: 10.1016/j.cpc.2014.04.012. arXiv: 1310.1921 [hep-ph].

[163] Celine Degrande et al. "UFO - The Universal FeynRules Output". In: *Comput. Phys. Commun.* 183 (2012), pp. 1201–1214. DOI: 10.1016/j.cpc.2012.01.022. arXiv: 1108.2040 [hep-ph].

[164] P. Artoisenet et al. "A framework for Higgs characterisation". In: *JHEP* 11 (2013), p. 043. DOI: 10.1007/JHEP11(2013)043. arXiv: 1306.6464 [hep-ph].

[165] Céline Degrande et al. "Automated one-loop computations in the SMEFT". In: *Phys. Rev. D* 103.9 (2021), p. 096024. DOI: 10.1103/physrevd.103.096024. arXiv: 2008.11743 [hep-ph].

[166] Fabio Cascioli, Philipp Maierhöfer, and Stefano Pozzorini. "Scattering Amplitudes with Open Loops". In: *Phys. Rev. Lett.* 108 (2012), p. 111601. DOI: 10.1103/PhysRevLett.108.111601. arXiv: 1111.5206 [hep-ph].

[167] Tanju Gleisberg and Stefan Höche. "Comix, a new matrix element generator". In: *JHEP* 12 (2008), p. 039. DOI: 10.1088/1126-6708/2008/12/039. arXiv: 0808.3674 [hep-ph].

[168] Pierre Artoisenet et al. "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations". In: *JHEP* 03 (2013), p. 015. DOI: 10.1007/JHEP03(2013)015. arXiv: 1212.3460 [hep-ph].

[169] ATLAS Collaboration. *ATLAS Pythia 8 tunes to 7 TeV data*. ATL-PHYS-PUB-2014-021. 2014. URL: https://cds.cern.ch/record/1966419.

[170] Stefan Höche et al. "QCD matrix elements + parton showers. The NLO case". In: *JHEP* 04 (2013), p. 027. DOI: 10.1007/JHEP04(2013)027. arXiv: 1207.5030 [hep-ph].

[171] Stefan Kallweit et al. "NLO QCD+EW predictions for V + jets including off-shell vector-boson decays and multijet merging". In: *JHEP* 04 (2016), p. 021. DOI: 10.1007/JHEP04(2016)021. arXiv: 1511.08692 [hep-ph].

[172] Christian Gütschow, Jonas M. Lindert, and Marek Schönherr. "Multi-jet merged top-pair production including electroweak corrections". In: *Eur. Phys. J. C* 78.4 (2018), p. 317. DOI: 10.1140/epjc/s10052-018-5804-2. arXiv: 1803.00950 [hep-ph].

[173] Rikkert Frederix and Ioannis Tsinikos. "On improving NLO merging for $t\bar{t}W$ production". In: *JHEP* 11 (2021), p. 029. DOI: 10.1007/JHEP11(2021)029. arXiv: 2108.07826 [hep-ph].

[174] Michal Czakon and Alexander Mitov. "Top++: A program for the calculation of the top-pair cross-section at hadron colliders". In: *Comput. Phys. Commun.* 185 (2014), p. 2930. DOI: 10.1016/j.cpc.2014.06.021. arXiv: 1112.5675 [hep-ph].

[175] Nikolaos Kidonakis. "Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-". In: *Phys. Rev. D* 82 (2010), p. 054018. DOI: 10.1103/PhysRevD.82.054018. arXiv: 1005.4451 [hep-ph].

[176] Nikolaos Kidonakis. "Next-to-next-to-leading-logarithm resummation for s-channel single top quark production". In: *Phys. Rev. D* 81 (2010), p. 054028. DOI: 10.1103/PhysRevD.81.054028. arXiv: 1001.5034 [hep-ph].

[177] Nikolaos Kidonakis. "Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production". In: *Phys. Rev. D* 83 (2011), p. 091503. DOI: 10.1103/PhysRevD.83.091503. arXiv: 1103.2792 [hep-ph].

[178] Stefano Frixione et al. "Single-top hadroproduction in association with a W boson". In: *JHEP* 07 (2008), p. 029. DOI: 10.1088/1126-6708/2008/07/029. arXiv: 0805.3067 [hep-ph].

[179] Steffen Schumann and Frank Krauss. "A parton shower algorithm based on Catani–Seymour dipole factorisation". In: *JHEP* 03 (2008), p. 038. DOI: 10.1088/1126-6708/2008/03/038. arXiv: 0709.1027 [hep-ph].

[180] Glen Cowan. *Statistical data analysis*. 1998. ISBN: 978-0-19-850156-5.

[181] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71 (2011), p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an]. Erratum: in: *Eur. Phys. J. C* 73 (2013), p. 2501. DOI: 10.1140/epjc/s10052-013-2501-z.

[182] Glen Cowan et al. In: *Eur. Phys. J. C* 73 (2013), p. 2501. DOI: 10.1140/epjc/s10052-013-2501-z.

[183] Kyle Cranmer et al. "HistFactory: A tool for creating statistical models for use with RooFit and RooStats". In: (June 2012).

[184] Wouter Verkerke and David Kirkby. *The RooFit toolkit for data modeling*. 2003. arXiv: physics/0306116 [physics.data-an].

[185] S. S. Wilks. "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses". In: *Annals Math. Statist.* 9.1 (1938), pp. 60–62. DOI: 10.1214/aoms/1177732360.

[186] Alexander L. Read. "Presentation of search results: the CL_S technique". In: *J. Phys. G* 28 (2002), p. 2693. DOI: 10.1088/0954-3899/28/10/313.

[187] John Ellis et al. "Disentangling Higgs-top couplings in associated production". In: *Journal of High Energy Physics* 2014.4 (Apr. 2014). ISSN: 1029-8479. DOI: 10.1007/jhep04(2014)004. URL: http://dx.doi.org/10.1007/JHEP04(2014)004.

[188] Steve Baker and Robert D. Cousins. "Clarification of the Use of Chi Square and Likelihood Functions in Fits to Histograms". In: *Nucl. Instrum. Meth.* 221 (1984), pp. 437–442. DOI: 10.1016/0167-5087(84)90016-4.

[189] Robert D. Cousins. *Generalization of Chisquare Goodness-of Fit Test for Binned Data Using Saturated Models, with Application to Histograms*. 2013. URL: https://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf.

[190] Jerome H. Friedman. "Greedy Function Approximation: A Gradient Boosting Machine". In: *The Annals of Statistics* 29.5 (2001), pp. 1189–1232. ISSN: 00905364, 21688966. URL: http://www.jstor.org/stable/2699986 (visited on 01/27/2025).

[191] Anqi Mao, Mehryar Mohri, and Yutao Zhong. *Cross-Entropy Loss Functions: Theoretical Analysis and Applications*. 2023. arXiv: 2304.07288 [cs.LG]. URL: https://arxiv.org/abs/2304.07288.

[192] Xinyu Peng, Li Li, and Fei-Yue Wang. *Accelerating Minibatch Stochastic Gradient Descent using Typicality Sampling*. 2019. arXiv: 1903.04192 [cs.LG]. URL: https://arxiv.org/abs/1903.04192.

[193] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG]. URL: https://arxiv.org/abs/1412.6980.

[194] Vinod Nair and Geoffrey E. Hinton. "Rectified linear units improve restricted boltzmann machines". In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. ICML'10. Haifa, Israel: Omnipress, 2010, pp. 807–814. ISBN: 9781605589077.

[195] Geoffrey Hinton et al. "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups". In: *IEEE Signal Processing Magazine* 29.6 (2012), pp. 82–97. DOI: 10.1109/MSP.2012.2205597.

[196] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. "Multilayer feedforward networks are universal approximators". In: *Neural Networks* 2.5 (1989), pp. 359–366. ISSN: 0893-6080. DOI: https://doi.org/10.1016/0893-6080(89)90020-8. URL: https://www.sciencedirect.com/science/article/pii/0893608089900208.

[197] M. Gori, G. Monfardini, and F. Scarselli. "A new model for learning in graph domains". In: *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.* Vol. 2. 2005, 729–734 vol. 2. DOI: 10.1109/IJCNN.2005.1555942.

[198] Franco Scarselli et al. "Graph Neural Networks for Ranking Web Pages." In: Jan. 2005, pp. 666–672. DOI: 10.1109/WI.2005.67.

[199] Franco Scarselli et al. "The Graph Neural Network Model". In: *IEEE Transactions on Neural Networks* 20 (2009), pp. 61–80. URL: https://api.semanticscholar.org/CorpusID:206756462.

[200] Franco Scarselli et al. "Computational Capabilities of Graph Neural Networks". In: *IEEE Transactions on Neural Networks* 20 (2009), pp. 81–102. URL: https://api.semanticscholar.org/CorpusID:206756425.

[201] Yujia Li et al. "Gated Graph Sequence Neural Networks". In: *arXiv: Learning* (2015). URL: https://api.semanticscholar.org/CorpusID:8393918.

[202] Peter W. Battaglia et al. "Relational inductive biases, deep learning, and graph networks". In: (2018). DOI: 10.48550/ARXIV.1806.01261. arXiv: 1806.01261.

[203] Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant. "Graph neural networks in particle physics". In: *Machine Learning: Science and Technology* 2.2 (Jan. 2021), p. 021001. ISSN: 2632-2153. DOI: 10.1088/2632-2153/abbf9a. URL: http://dx.doi.org/10.1088/2632-2153/abbf9a.

[204] Savannah Thais et al. *Graph Neural Networks in Particle Physics: Implementations, Innovations, and Challenges*. 2022. arXiv: 2203.12852 [hep-ex]. URL: https://arxiv.org/abs/2203.12852.

[205] Xiangyang Ju et al. "Performance of a geometric deep learning pipeline for HL-LHC particle tracking". In: *Eur. Phys. J. C* 81.10 (2021), p. 876. DOI: 10.1140/epjc/s10052-021-09675-8. arXiv: 2103.06995 [physics.data-an].

[206] Daniel Murnane, Savannah Thais, and Ameya Thete. "Equivariant Graph Neural Networks for Charged Particle Tracking". In: *21th International Workshop on Advanced Computing and Analysis Techniques in Physics Research: AI meets Reality*. Apr. 2023. arXiv: 2304.05293 [physics.ins-det].

[207] Shah Rukh Qasim et al. "Learning representations of irregular particle-detector geometry with distance-weighted graph networks". In: *The European Physical Journal C* 79.7 (July 2019). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-7113-9. URL: http://dx.doi.org/10.1140/epjc/s10052-019-7113-9.

[208] Shah Rukh Qasim et al. *Multi-particle reconstruction in the High Granularity Calorimeter using object condensation and graph neural networks*. 2021. arXiv: 2106.01832 [physics.ins-det]. URL: https://arxiv.org/abs/2106.01832.

[209] Manzil Zaheer et al. "Deep Sets". In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017.

[210] Harrison Edwards and Amos Storkey. *Towards a Neural Statistician*. 2017. arXiv: 1606.02185 [stat.ML]. URL: https://arxiv.org/abs/1606.02185.

[211] Tomas Pevny and Petr Somol. *Using Neural Network Formalism to Solve Multiple-Instance Problems*. 2017. arXiv: 1609.07257 [cs.LG]. URL: https://arxiv.org/abs/1609.07257.

[212] Yue Gao et al. "3-D Object Retrieval and Recognition With Hypergraph Analysis". In: *IEEE Transactions on Image Processing* 21 (2012), pp. 4290–4303. URL: https://api.semanticscholar.org/CorpusID:17879431.

[213] Yifan Feng et al. *Hypergraph Neural Networks*. 2019. arXiv: 1809.09401 [cs.LG]. URL: https://arxiv.org/abs/1809.09401.

[214] Chuan Shi et al. "A survey of heterogeneous information network analysis". In: *IEEE Transactions on Knowledge and Data Engineering* 29.1 (2017), pp. 17–37. DOI: 10.1109/TKDE.2016.2598561.

[215] Xiaocheng Yang et al. *Simple and Efficient Heterogeneous Graph Neural Network*. 2023. arXiv: 2207.02547 [cs.LG]. URL: https://arxiv.org/abs/2207.02547.

[216] Ashish Vaswani et al. *Attention Is All You Need*. 2023. arXiv: 1706.03762 [cs.CL]. URL: https://arxiv.org/abs/1706.03762.

[217] Tom B. Brown et al. *Language Models are Few-Shot Learners*. 2020. arXiv: 2005.14165 [cs.CL]. URL: https://arxiv.org/abs/2005.14165.

[218] Alexey Dosovitskiy et al. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. 2021. arXiv: 2010.11929 [cs.CV]. URL: https://arxiv.org/abs/2010.11929.

[219] John M. Jumper et al. "Highly accurate protein structure prediction with AlphaFold". In: *Nature* 596 (2021), pp. 583–589. URL: https://api.semanticscholar.org/CorpusID:235959867.

[220] Huilin Qu and Loukas Gouskos. "Jet tagging via particle clouds". In: *Physical Review D* 101.5 (Mar. 2020). ISSN: 2470-0029. DOI: 10.1103/physrevd.101.056019. URL: http://dx.doi.org/10.1103/PhysRevD.101.056019.

[221] ATLAS Collaboration. *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*. ATL-PHYS-PUB-2022-027. 2022. URL: https://cds.cern.ch/record/2811135.

[222] Jonas Spinner et al. *Lorentz-Equivariant Geometric Algebra Transformers for High-Energy Physics*. 2024. arXiv: 2405.14806 [physics.data-an]. URL: https://arxiv.org/abs/2405.14806.

[223] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[224] Adam Paszke et al. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. 2019. arXiv: 1912.01703 [cs.LG]. URL: https://arxiv.org/abs/1912.01703.

[225] James Bradbury et al. *JAX: composable transformations of Python+NumPy programs*. Version 0.3.13. 2018. URL: http://github.com/jax-ml/jax.

[226] Matthias Fey and Jan Eric Lenssen. *Fast Graph Representation Learning with PyTorch Geometric*. 2019. arXiv: 1903.02428 [cs.LG]. URL: https://arxiv.org/abs/1903.02428.

[227] Oleksandr Ferludin et al. *TF-GNN: Graph Neural Networks in TensorFlow*. 2023. arXiv: 2207.03522 [cs.LG]. URL: https://arxiv.org/abs/2207.03522.

[228] Minjie Wang et al. *Deep Graph Library: A Graph-Centric, Highly-Performant Package for Graph Neural Networks*. 2020. arXiv: 1909.01315 [cs.LG]. URL: https://arxiv.org/abs/1909.01315.

[229] Tianqi Chen et al. *MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems*. 2015. arXiv: 1512.01274 [cs.DC]. URL: https://arxiv.org/abs/1512.01274.

[230] Michael James Fenton et al. "Permutationless many-jet event reconstruction with symmetry preserving attention networks". In: *Physical Review D* 105.11 (June 2022). ISSN: 2470-0029. DOI: 10.1103/physrevd.105.112008. URL: http://dx.doi.org/10.1103/PhysRevD.105.112008.

[231] Alexander Shmakov et al. "SPANet: Generalized permutationless set assignment for particle physics using symmetry preserving attention". In: *SciPost Physics* 12.5 (May 2022). ISSN: 2542-4653. DOI: 10.21468/scipostphys.12.5.178. URL: http://dx.doi.org/10.21468/SciPostPhys.12.5.178.

[232] Michael James Fenton et al. "Reconstruction of unstable heavy particles using deep symmetry-preserving attention networks". In: *Communications Physics* 7.1 (Apr. 2024). ISSN: 2399-3650. DOI: 10.1038/s42005-024-01627-4. URL: http://dx.doi.org/10.1038/s42005-024-01627-4.

[233] *SALT Framework*. 2025. URL: https://ftag-salt.docs.cern.ch/.

[234] Xiangyang Ju et al. "Graph Neural Networks for Particle Reconstruction in High Energy Physics detectors". In: *33rd Annual Conference on Neural Information Processing Systems*. Mar. 2020. arXiv: 2003.11603 [physics.ins-det].

[235] I. Antcheva et al. "ROOT — A C++ framework for petabyte data storage, statistical analysis and visualization". In: *Computer Physics Communications* 180.12 (Dec. 2009), pp. 2499–2512. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2009.08.005. URL: http://dx.doi.org/10.1016/j.cpc.2009.08.005.

[236] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. *Layer Normalization*. 2016. arXiv: 1607.06450 [stat.ML]. URL: https://arxiv.org/abs/1607.06450.

[237] Nitish Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

[238] CMS Collaboration. "Evidence for four-top quark production in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Phys. Lett. B* 844 (2023), p. 138076. DOI: 10.1016/j.physletb.2023.138076. arXiv: 2303.03864 [hep-ex].

[239] ATLAS Collaboration. "Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *Phys. Rev. D* 99 (2019), p. 072009. DOI: 10.1103/PhysRevD.99.072009. arXiv: 1901.03584 [hep-ex].

[240] CMS Collaboration. "Measurement of the cross section of top quark–antiquark pair production in association with a W boson in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *JHEP* 07 (2023), p. 219. DOI: 10.1007/JHEP07(2023)219. arXiv: 2208.06485 [hep-ex].

[241] ATLAS Collaboration. "Measurement of the total and differential cross-sections of $t\bar{t}W$ production in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *JHEP* 05 (2024), p. 131. DOI: 10.1007/JHEP05(2024)131. arXiv: 2401.05299 [hep-ex].

[242] ATLAS Collaboration. "Search for leptonic charge asymmetry in $t\bar{t}W$ production in final states with three leptons at $\sqrt{s} = 13\,\text{TeV}$". In: *JHEP* 07 (2023), p. 033. DOI: 10.1007/JHEP07(2023)033. arXiv: 2301.04245 [hep-ex].

[243] S. D. Ellis, R. Kleiss, and W. J. Stirling. "W's, Z's and jets". In: *Phys. Lett. B* 154.5 (1985), pp. 435–440. DOI: https://doi.org/10.1016/0370-2693(85)90425-3.

[244] F. A. Berends et al. "Multijet production in W, Z events at pp colliders". In: *Phys. Lett. B* 224.1 (1989), pp. 237–242. DOI: https://doi.org/10.1016/0370-2693(89)91081-2.

[245] W. T. Giele and W. J. Stirling. "Top search at fermilab: Multijet signals and backgrounds". In: *Nucl. Phys. B* 343.1 (1990), pp. 14–30. DOI: https://doi.org/10.1016/0550-3213(90)90592-2.

[246] Erik Gerwick et al. "Scaling patterns for QCD jets". In: *JHEP* 10 (2012), p. 162. DOI: 10.1007/JHEP10(2012)162. arXiv: 1208.3676 [hep-ph].

[247] ATLAS Collaboration. *Study of ttbb and ttW background modelling for ttH analyses*. ATL-PHYS-PUB-2022-026. 2022. URL: https://cds.cern.ch/record/2810864.

[248] Christian F. Nielsen et al. "Precision Measurement of Trident Production in Strong Electromagnetic Fields". In: *Phys. Rev. Lett.* 130 (7 Feb. 2023), p. 071601. DOI: 10.1103/PhysRevLett.130.071601. URL: https://link.aps.org/doi/10.1103/PhysRevLett.130.071601.

[249] ATLAS Collaboration. "ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J. C* 79 (2019), p. 970. DOI: 10.1140/epjc/s10052-019-7450-8. arXiv: 1907.05120 [hep-ex].

[250] ATLAS Collaboration. "Measurement of the c-jet mistagging efficiency in $t\bar{t}$ events using pp collision data at $\sqrt{s} = 13\,\text{TeV}$ collected with the ATLAS detector". In: *Eur. Phys. J. C* 82 (2022), p. 95. DOI: 10.1140/epjc/s10052-021-09843-w. arXiv: 2109.10627 [hep-ex].

[251] ATLAS Collaboration. "Calibration of the light-flavour jet mistagging efficiency of the b-tagging algorithms with Z+jets events using 139 fb^{-1} of ATLAS proton-proton collision data at $\sqrt{s} = 13$ TeV". In: (Jan. 2023). arXiv: 2301.06319 [hep-ex].

[252] ATLAS Collaboration. "Jet energy measurement with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 7\,\text{TeV}$". In: *Eur. Phys. J. C* 73 (2013), p. 2304. DOI: 10.1140/epjc/s10052-013-2304-2. arXiv: 1112.6426 [hep-ex].

[253] ATLAS Collaboration. "Jet energy resolution in proton–proton collisions at $\sqrt{s} = 7\,\text{TeV}$ recorded in 2010 with the ATLAS detector". In: *Eur. Phys. J. C* 73 (2013), p. 2306. DOI: 10.1140/epjc/s10052-013-2306-0. arXiv: 1210.6210 [hep-ex].

[254] ATLAS Collaboration. "Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ using the ATLAS detector". In: *Eur. Phys. J. C* 76 (2016), p. 581. DOI: 10.1140/epjc/s10052-016-4395-z. arXiv: 1510.03823 [hep-ex].

[255] ATLAS Collaboration. "Measurement of the inelastic proton–proton cross-section at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector". In: *Nature Commun.* 2 (2011), p. 463. DOI: 10.1038/ncomms1472. arXiv: 1104.0326 [hep-ex].

[256] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *J. Phys. G* 43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph].

[257] ATLAS Collaboration. "Measurement of the production cross-section of a single top quark in association with a Z boson in proton–proton collisions at 13 TeV with the ATLAS detector". In: *Phys. Lett. B* 780 (2018), p. 557. DOI: 10.1016/j.physletb.2018.03.023. arXiv: 1710.03659 [hep-ex].

[258] Hesham El Faham. "tWZ production at NLO in QCD in the SMEFT". In: *14th International Workshop on Top Quark Physics*. Dec. 2021. arXiv: 2112.13282 [hep-ph].

[259] Olga Bessidskaia Bylund. "Modelling Wt and tWZ production at NLO for ATLAS analyses". In: *9th International Workshop on Top Quark Physics*. Nov. 2016. arXiv: 1612.00440 [hep-ph].

[260] ATLAS Collaboration. "Measurement of $W^{\pm}Z$ production cross sections and gauge boson polarisation in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *Eur. Phys. J. C* 79 (2019), p. 535. DOI: 10.1140/epjc/s10052-019-7027-6. arXiv: 1902.05759 [hep-ex].

[261] ATLAS Collaboration. "Search for new phenomena in events with same-charge leptons and b-jets in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *JHEP* 12 (2018), p. 039. DOI: 10.1007/JHEP12(2018)039. arXiv: 1807.11883 [hep-ex].

[262] ATLAS Collaboration. "Measurements of inclusive and differential fiducial cross-sections of $t\bar{t}$ production with additional heavy-flavour jets in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *JHEP* 04 (2019), p. 046. DOI: 10.1007/JHEP04(2019)046. arXiv: 1811.12113 [hep-ex].

[263] Roger Barlow and Christine Beeston. "Fitting using finite Monte Carlo samples". In: *Comput. Phys. Commun.* 77 (1993), p. 219. DOI: 10.1016/0010-4655(93)90005-W.

[264] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. *Searching for Activation Functions*. 2017. arXiv: 1710.05941 [cs.NE]. URL: https://arxiv.org/abs/1710.05941.

[265] Qing-Hong Cao et al. "What can we learn from triple top-quark production?" In: *Phys. Rev. D* 100.5 (2019), p. 055035. DOI: 10.1103/PhysRevD.100.055035. arXiv: 1901.04643 [hep-ph].

[266] Hamzeh Khanpour. "Probing top quark FCNC couplings in the triple-top signal at the high energy LHC and future circular collider". In: *Nucl. Phys. B* 958 (2020), p. 115141. DOI: 10.1016/j.nuclphysb.2020.115141. arXiv: 1909.03998 [hep-ph].

[267] ATLAS Collaboration. "Combined measurements of Higgs boson production and decay using up to $80\,\text{fb}^{-1}$ of proton–proton collision data at $\sqrt{s} = 13\,\text{TeV}$ collected with the ATLAS experiment". In: *Phys. Rev. D* 101 (2020), p. 012002. DOI: 10.1103/PhysRevD.101.012002. arXiv: 1909.02845 [hep-ex].

[268] ATLAS Collaboration. "Constraints on the off-shell Higgs boson signal strength in the high-mass ZZ and WW final states with the ATLAS detector". In: *Eur. Phys. J. C* 75 (2015), p. 335. DOI: 10.1140/epjc/s10052-015-3542-2. arXiv: 1503.01060 [hep-ex].

[269] ATLAS Collaboration. "Measurement of Higgs boson decay into b-quarks in associated production with a top-quark pair in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *JHEP* 06 (2022), p. 097. DOI: 10.1007/JHEP06(2022)097. arXiv: 2111.06712 [hep-ex].

[270] ATLAS Collaboration. "Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector". In: *Eur. Phys. J. C* 81 (2021), p. 178. DOI: 10.1140/epjc/s10052-020-08677-2. arXiv: 2007.02873 [hep-ex].

[271] ATLAS Collaboration. "Measurement of the associated production of a Higgs boson decaying into b-quarks with a vector boson at high transverse momentum in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *Phys. Lett. B* 816 (2021), p. 136204. DOI: 10.1016/j.physletb.2021.136204. arXiv: 2008.02508 [hep-ex].

[272] ATLAS Collaboration. "Measurements of Higgs bosons decaying to bottom quarks from vector boson fusion production with the ATLAS experiment at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J. C* 81 (2021), p. 537. DOI: 10.1140/epjc/s10052-021-09192-8. arXiv: 2011.08280 [hep-ex].

[273] ATLAS Collaboration. "Higgs boson production cross-section measurements and their EFT interpretation in the 4ℓ decay channel at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *Eur. Phys. J. C* 80 (2020), p. 957. DOI: 10.1140/epjc/s10052-020-8227-9. arXiv: 2004.03447 [hep-ex]. Erratum: in: *Eur. Phys. J. C* 81 (2021), p. 29. DOI: 10.1140/epjc/s10052-020-08644-x. Erratum: in: *Eur. Phys. J. C* 81 (2021), p. 398. DOI: 10.1140/epjc/s10052-021-09116-6.

[274] ATLAS Collaboration. "Measurements of Higgs boson production by gluon–gluon fusion and vector-boson fusion using $H \to WW^* \to e\nu\mu\nu$ decays in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *Phys. Rev. D* 108 (2023), p. 032005. DOI: 10.1103/PhysRevD.108.032005. arXiv: 2207.00338 [hep-ex].

[275] ATLAS Collaboration. "Measurement of the production cross section for a Higgs boson in association with a vector boson in the $H \to WW^* \to \ell\nu\ell\nu$ channel in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *Phys. Lett. B* 798 (2019), p. 134949. DOI: 10.1016/j.physletb.2019.134949. arXiv: 1903.10052 [hep-ex].

[276] ATLAS Collaboration. "Measurements of Higgs boson production cross-sections in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *JHEP* 08 (2022), p. 175. DOI: 10.1007/JHEP08(2022)175. arXiv: 2201.08269 [hep-ex].

[277] ATLAS Collaboration. "A search for the $Z\gamma$ decay mode of the Higgs boson in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *Phys. Lett. B* 809 (2020), p. 135754. DOI: 10.1016/j.physletb.2020.135754. arXiv: 2005.05382 [hep-ex].

[278] ATLAS Collaboration. "Constraint on the total width of the Higgs boson from Higgs boson and four-top-quark measurements in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: (2024). arXiv: 2407.10631 [hep-ex].

[279] J. R. Andersen et al. "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties". In: (July 2013). DOI: 10.5170/CERN-2013-004. arXiv: 1307.1347 [hep-ph].

[280] ATLAS Collaboration. "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector". In: *Phys. Lett. B* 784 (2018), p. 173. DOI: 10.1016/j.physletb.2018.07.035. arXiv: 1806.00425 [hep-ex].

[281] ATLAS Collaboration. "Measurements of observables sensitive to colour reconnection in $t\bar{t}$ events with the ATLAS detector at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J. C* 83 (2023), p. 518. DOI: 10.1140/epjc/s10052-023-11479-x. arXiv: 2209.07874 [hep-ex].

[282] ATLAS Collaboration. "Electron and photon energy calibration with the ATLAS detector using LHC Run 2 data". In: *JINST* 19 (2024), P02009. DOI: 10.1088/1748-0221/19/02/P02009. arXiv: 2309.05471 [hep-ex].

[283] Arnaud Duperrin. "Flavour tagging with graph neural networks with the ATLAS detector". In: *30th International Workshop on Deep-Inelastic Scattering and Related Subjects*. June 2023. arXiv: 2306.04415 [hep-ex].

[284] Yujia Li et al. *Gated Graph Sequence Neural Networks*. 2017. arXiv: 1511.05493 [cs.LG]. URL: https://arxiv.org/abs/1511.05493.

[285] *Transformer for Energy Calibration in the ATLAS Electromagnetic Calorimeter*. ATL-EGAM-2023-01. 2024. URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/EGAM-2023-01/.

[286] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. *Explaining and Harnessing Adversarial Examples*. 2015. arXiv: 1412.6572 [stat.ML]. URL: https://arxiv.org/abs/1412.6572.

[287] Yaroslav Ganin et al. *Domain-Adversarial Training of Neural Networks*. 2016. arXiv: 1505.07818 [stat.ML]. URL: https://arxiv.org/abs/1505.07818.

[288] Oliver Brüning and M. Zerlauth. "Overall status of the HL-LHC project". In: *JACoW* IPAC2023 (2023), TUYG1. DOI: 10.18429/JACoW-IPAC2023-TUYG1.

[289] *Highlights of the HL-LHC physics projections by ATLAS and CMS*. ATL-PHYS-PUB-2025-01. 2025. URL: https://cds.cern.ch/record/2928907/.

[290] *The ATLAS Upgrade for the HL-LHC*. ATL-UPGRADE-PUB-2025-001. 2025. URL: https://cds.cern.ch/record/2928798/.

[291] *ATLAS Software and Computing for the Future*. ATL-SOFT-PUB-2025-002. 2025. URL: https://cds.cern.ch/record/2928841/.

[292] ATLAS Collaboration. *ATLAS Inner Tracker Strip Detector: Technical Design Report*. ATLAS-TDR-025; CERN-LHCC-2017-005. 2017. URL: https://cds.cern.ch/record/2257755.

[293] William Panduro Vazquez. "FELIX: the new detector interface for ATLAS". In: *EPJ Web Conf.* 245 (2020). Ed. by C. Doglioni et al., p. 01037. DOI: 10.1051/epjconf/202024501037.

[294] W. J. Ashmanskas et al. "Verification of simulated ASIC functionality and radiation tolerance for the HL-LHC ATLAS ITk Strip Detector". In: *JINST* 18.01 (2023), p. C01029. DOI: 10.1088/1748-0221/18/01/C01029.

[295] J. R. Dandoy et al. "Irradiation testing of ASICs for the HL-LHC ATLAS ITk Strip Detector". In: *JINST* 18.02 (2023), p. C02044. DOI: 10.1088/1748-0221/18/02/C02044.

[296] C. Belanger-Champagne et al. "BETSEE: testing for system-wide effects of single event effects on ITk strip modules". In: *JINST* 18.01 (2023), p. C01019. DOI: 10.1088/1748-0221/18/01/C01019.