University of Sheffield

School of Mathematical and Physical Sciences, University of Sheffield

Ph.D. thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the subject of Physics

A measurement of top quark pair production in association with a Z boson decaying to neutrinos

CANDIDATE: MICHAEL POSTILL

Supervisor: Davide Costanzo

November 26, 2025

Abstract

This thesis presents results for two ATLAS analyses measuring the cross section of $\mathrm{t\bar{t}Z}$, analyzing events where the Z-boson decays into a neutrinos pair ($Z \rightarrow \nu\bar{\nu}$), and one piece of technical research into improving the pile-up corrections for ATLAS's forward calorimeter using topological towers.

Both $\mathrm{t\bar{t}Z}$ analyses use the ATLAS full Run 2 dataset with centre-of-mass energy $\sqrt{s} = 13\ \mathrm{TeV}$ and an integrated luminosity of $140\ \mathrm{fb}^{-1}$. These analyses are presented in part II, the first analysis, a standalone zero-lepton channel analysis, while the second analyses is a combined zero, one and two-lepton analyses. The standalone analysis obtained a cross-section of $\sigma_{t\bar{t}Z\rightarrow\nu\bar{\nu}} = 0.11^{+0.08}_{-0.07}\ pb$ in good agreement with standard model predictions and a signal significance of $Z = 2.47\ \sigma$, this region is then used in the combined analysis with the one and two lepton channels, showing an expected signal significance of $Z = 4.9\ \sigma$. This result will allow dark matter and SUSY searches to further constrain their main background, improving their sensitivity.

The technical research is presented in part III and uses ATLAS Run 2 simulations to exploit topological towers in the forward region of the ATLAS calorimeter to model the impact of pile-up on jet reconstruction. This reconstruction method gives better modeling of forward jet pile-up corrections, improving over previous attempts using calorimeter topological clusters. We used this improved modeling of the pile-up to create an equation that would give a value for the pile-up correction needed based on the in and out-of-time pile-up conditions at ATLAS, which showed improved performance over previous methods. The implementation of this algorithm for Run 3 data will result in better precision of forward jets in measurements and will be especially important to rare standard model searches.

Author's contributions

This thesis presents the work from one ATLAS analysis and another piece of technical work. The analysis is the main thesis topic, with the technical work being the continuation of work that was conducted to become an ATLAS author, known as an ATLAS qualification task. The author's contributions to these works are listed below.:

- **Measurement of the inclusive $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$ cross section**: Lead analyzer in the 0L channel. This involved training, optimizing and configuring the neural network used to define all the signal, control and validation regions. Coded most of the variables in the data and Monte Carlo simulations used in both the training of the neural network and the region definitions. Defined and performed the fit for the 0L region.

- **Topological Towers and their use in the forward region of the ATLAS Calorimeter**: Lead developer in jet clustering studies investigating the use of topological towers in the forward region of the ATLAS calorimeter. This includes the implementation and update of various legacy tools that have been used previously with topological clusters that have been adapted to work with topological towers. This also includes the modeling of $\rho(\eta)$ and $\rho(\eta, \mathrm{N_{pv}}, \mu)$, as well as all of the testing of these models.

Acknowledgments

The research presented in this thesis has taken me three and a half years of my life, and has presented me with some of the toughest challenges of my academic career. It would not have been possible for me to complete this thesis without the encouragement and support of my supervisor, colleagues, friends and family.

First, I would like to thank my supervisor, Davide, for his patience in helping me to write this thesis with my dyslexia, which was no easy task. I would like to thank him for guiding me through my PhD and always treating me with respect. I would also like to thank my technical supervisor, Mark, for helping me with all my technical issues, of which there were many. I would like to thank Matt for teaching me how to code properly and training me in my first few years of my PhD. You always went above and beyond in preparing me to be a researcher at ATLAS. I would like to thank my colleagues in the $t\bar{t}Z \to \nu\bar{\nu}$ group for their help and patience during this long analysis. I would like to thank members of the Jet definitions group for helping me through my technical research on topo-towers, especially Peter and Mark.

I would like to thank my family for their support that helped me get onto this PhD and there continued belief in me throughout my PhD, in particular I would like to thank my mother for getting me the help I needed in school to get through my dyslexia and always believing that my mind was stronger than my penmanship. I would like to thank my grandparents, Ron and Silvia, for always supporting me and being like a second set of parents. I would like to thank my dad for always pushing me to achieve my goals.

I would like to thank my beautiful girlfriend, Alexandria, for always believing in me and listening to me ”yap” about my work even though it wasn't her thing. I always want to thank her for helping me through the stressful times of this PhD otherwise, I doubt I would have completed it. I would like to thank my friends Faye, Jothi, Valentina and Jemima, all of whom visited me in Geneva to see my work: Faye, whom I met on a physics course when I was 18 and first told I was going to be a doctor of physics, who never doubted me for a second. Faye, you are my best friend, and I am honored to have been the usher at your wedding. Jothi, who was one of the first friends I made when I first came to Sheffield 9 years ago, and who has stuck with me and gave me encouragement throughout my PhD. Val, who was always there for me and made the effort to see and support me through this endeavour. Having you take an interest in my work meant the world to me. A PhD student could never ask for a better friend than you. Jemima, who has always kept my spirits up and been a source of joy throughout my entire journey of my PhD, you are a kindred spirit and I am truly proud of to be called your friend.

Contents

Part I

Introduction to Particle Physics at the LHC and ATLAS

In part I we review the various theoretical, statistical and technical basis used when conducting research at ATLAS. Chapter 1 reviews the theory used in most of ATLAS research projects with a focus on topics related to this thesis; In chapter 2 the technical foundation of the ATLAS experiment and the LHC are will be covered. In Chapter 3, we review the statistical and analytical methods used by ATLAS. This will include details on how particle physics simulations using Monte Carlo samples are produced for ATLAS, statistical methods used to evaluate results at ATLAS, and how events are reconstructed to process the data at ATLAS. In chapter 4 we cover how neural networks are used in particle physics, introducing the relevant terminology that is used in Part II. Finally, in chapter 5 we review the theoretical basis for the production of $\mathrm{t\bar{t}Z}$ and discuss the experimental landscape to give context to the work presented in partII

Chapter 1

Theoretical introduction

This chapter covers the particle physics theory that is needed to comprehend the research presented in Part II and Part III. We review the standard model as a quantum gauge field theory, explaining which particles are predicted as well as the physics behind each of the fundamental forces. This section ends by explaining the introduction of the Higgs mechanism to the standard model. We also briefly explore the failings of the standard model to highlight that standard model physics is still an active field of research.

1.1 The Standard Model

The goal of any fundamental physics model is to produce a set of laws and theories that can make predictions that are testable through experiments. Any fundamental physics model at the subatomic level will need to cover 3 classes of interactions: strong, weak and electromagnetic. The only interaction that is not covered is gravity, which is negligible on the subatomic level but dominates at the astrophysical level. All these interactions other are described by quantum mechanics (quantum field theory) and special relativity. All these particle interactions have something in common: they all operate under the basic principle of gauge invariance [1]. Gauge theories are used as the framework for many modern physics theories. A gauge theory states that the Lagrangian density of the system and the system's internal degrees of freedom are invariant over local transformations performed by a specific group of operators[1].

To produce a gauge theory, certain parameters need to be set. First a scalar field ϕ is used to construct a Lagrangian density $\mathcal{L}[\phi, \partial_\phi \phi]$ which needs to be designed to be invariant under a D-dimensional continuous group of transformations (Γ). The transformations are of the form [1]

$$\phi'(x) = U(\theta^A)\phi(x), \tag{1.1}$$

where A is dimension of the field being operated on where $A = (1, 2..., D)$. The $U(\theta^{\mathrm{A}})$

term is defined as

$$U(\theta^A) = \exp(ig \sum_A \theta^A T^A), \tag{1.2}$$

where the generators of the Γ group of transformations for this gauge theory's field ϕ are the $\mathrm{T^A}$ terms. θ^{A} are numerical parameters, such as distance from the origin in the case of a group translation in an internal space. The constant g is the coupling constant that controls the strength of the transformation. In general, the generators for the gauge theory follow the commutation relation

$$[T^A, T^B] = iC_{ABC}T^C \tag{1.3}$$

Where C_{ABC} are the structure constants. If all the generators commute then the gauge theory is said to be abelian and the structure constants vanish[1]. If they do not commute then the theory is non-abelian.

In cases where the parameters of the Lagrangian depend on the spacetime coordinates, covariant derivatives are required to maintain gauge invariance. These covariant derivatives take the form

$$D_\mu = \partial_\mu + igV_\mu^A, \tag{1.4}$$

where $\mathrm{V_\mu^A}$ are gauge vector fields associated with the group's generators. For example, in a theory such as quantum electrodynamics this would be the photon. The covariant derivatives can be used to create a gauge-invariant kinetic energy term for the gauge fields by performing the following calculation

$$[D_\mu, D_\nu]\phi = ig(\partial_\mu V_\nu^A - \partial_\nu V_\mu^A - gC_{ABC}V_\nu^B V_\nu^C)\phi = igF_{\mu\nu}^A\phi, \tag{1.5}$$

where we define

$$F_{\mu\nu}^A = \partial_\mu V_\nu^A - \partial_\nu V_\mu^A - gC_{ABC}V_\nu^B V_\nu^C. \tag{1.6}$$

The $\partial_\mu V_\nu^A - \partial_\nu V_\mu^A$ term is the kinetic energy part of the equation, while the $gC_{ABC}V_\nu^B V_\nu^C$ shows the self-interactions of the gauge theory [2].

The $\mathcal{L}[\phi, D_\mu\phi]$, which is the interaction term for the Lagrangian, can also be represented in the Hermitian format as

$$\mathcal{L}[\phi, D_\mu\phi] = \frac{1}{2}[(D_\mu\phi)^2 - m^2\phi^2] - V(\phi). \tag{1.7}$$

where m is the mass term of the system and $V(\phi)$ is the interaction potential. Equation 1.7 can take different forms depending on the theory to which we apply it to, such as quantum chromo-dynamics or quantum electro-dynamics.

We now come to the Yang–Mills Lagrangian (which is by design invariant under Lorentz transformations), which takes the form [1]

$$\mathcal{L}_{YM} = -\frac{1}{4}\sum_{A} F^{A}_{\mu\nu}F^{\mu\nu A} + \mathcal{L}[\phi, D_\mu\phi], \tag{1.8}$$

Where the $-\frac{1}{4}\sum_A F^A_{\mu\nu}F^{\mu\nu A}$ is the kinetic part of the Lagrangian while the $\mathcal{L}[\phi, D_\mu\phi]$ is the interaction part. This equation can be applied to various aspects of the standard model to calculate their individual Lagrangian. These can then be combined to give an overall Lagrangian for the standard model.

The standard model(SM) is designed with 3 principles [3][4]:

1. There are 3 types of fundamental particles, quarks, leptons and bosons, with the first two being used to make most other particles.
2. The fundamental forces and their dynamics can be explained by gauge theories with the forces being mediated by bosons.
3. The Higgs field allows for spontaneous symmetry breaking to occur anywhere in the universe and therefore gives particles mass.

Principle 1 is expressed in Table 1.1 showing the 3 types of particles that form the SM.

Principle 2 builds off what was already discussed with gauge theories and uses them to describe three of the fundamental forces: strong, weak and electromagnetic. The SM is therefore defined as a 4-dimensional quantum non-Abelian gauge field theory based on the symmetry group[1]

$$G_{SM} = SU(3) \otimes SU(2) \otimes U(1). \tag{1.9}$$

The group contains 8 generators from the $SU(3)$ term, 3 generators from the $SU(2)$ term and 1 generator from the $U(1)$ term. This gives 12 generators in total. The $U(1)$ terms describe the electromagnetic interaction in quantum electro-dynamics (QED), the $SU(2)$ terms describe the weak interaction, and the $SU(3)$ term describes the strong interactions, which are also known Quantum Chromo-Dynamics(QCD). The $SU(2)$ and $U(1)$ terms are later unified in section 1.1.2[1].

Particle		Type	gen	Q	Colour	spin	mass
electron	e	lepton	1	-1	0	1/2	511 keV
muon	μ	lepton	2	-1	0	1/2	106 MeV
tau	τ	lepton	3	-1	0	1/2	1.78 GeV
electron neutrino	$\mathrm{v_e}$	lepton	1	0	0	1/2	$<$ 0.28 eV
muon neutrino	v_μ	lepton	2	0	0	1/2	$<$ 0.28 eV
tau neutrino	v_τ	lepton	3	0	0	1/2	$<$ 0.28 eV
up	u	quark	1	+2/3	3	1/2	2.2 MeV
down	d	quark	1	-1/3	3	1/2	4.7 MeV
charm	c	quark	2	+2/3	3	1/2	1.3 GeV
strange	s	quark	2	-1/3	3	1/2	93 MeV
top	t	quark	3	+2/3	3	1/2	172.9 GeV
bottom	b	quark	3	-1/3	3	1/2	4.2 GeV
photon	γ	vector boson		0	0	1	0
W boson	$\mathrm{W}^\pm$	vector boson		± 1	0	1	80.4 GeV
Z boson	Z	vector boson		0	0	1	91.2 GeV
gluon	g	vector boson		0	8	1	0
Higgs	h	scalar boson		0	1	0	125.1 GeV

Table 1.1: Summary table of all standard model particles including the particles, name, symbol, type, generation (gen), charge Q, Number of colour states, spin and mass [5].

The Standard Model Lagrangian can be broken up into

$$\mathcal{L}_{SM} = \mathcal{L}_{EW} + \mathcal{L}_{QCD} + \mathcal{L}_{other}. \tag{1.10}$$

The standard model also has specific types of transformations, such as the CPT transformations. We will quickly define these transformations here and mention in the following sections which of these theories are symmetric under these transformations.

- The C stands for a charge conjugation transformation. This transformation changes a particle into its corresponding antiparticle, flipping its internal charge but leaving all other properties such as mass and spin unchanged. For example, an electron under a C transformation would give a positron [2].
- The P stands for parity transformations. Party inverts the spatial coordinate $\overrightarrow{x} \to -\overrightarrow{x}$. This also affects the momenutm like so $\overrightarrow{p} \to -\overrightarrow{p}$, however angular momentum J remains unchanged by P transformations [2].
- The T stands for time reversal. Time reversal inverts the t axis $\mathrm{t} \to -\mathrm{t}$, This affects both the momentum $\overrightarrow{p} \to -\overrightarrow{p}$ and angular momentum $\overrightarrow{J} \to -\overrightarrow{J}$ of the system [2].

CPT all applied together implies that the mass, magentic moment and lifetime of any particle must be exactly equal to that of its antiparticle. The combined CPT transformation is an effective test for our gauge theroies, as any lorentz-invariant local quantum field theory with a hermitian Hamiltonian must be invariant under all 3 of these transformations [2].

1.1.1 Quantum Chromodynamics

SU(3) contains 8 colour generators($\mathrm{T^A}$ where A=1-8) which correspond to the 8 gluons which mediate the strong interaction. The gluon is a massless particle, and therefore, the strong force should have limitless range like the photon. However, the strong interaction range is constrained by a process known as confinement, where an isolated colour charge cannot exist. Only colour singlets that have no overall colour charge are allowed to exist, such as baryons and mesons. The potential of the strong force enforces this by increasing with distance rather than decreasing. This means that as two quarks move away from one another, the potential from the strong interaction increases until there is enough energy to produce more quarks. This leads to energetic groups of quarks producing additional final-state particles. This is known as hadronization and results in the formation of jets as will be discussed in section 3.1.3[1].

The covariant derivative for QCD takes the form:

$$D_\mu = \partial_\mu + ie_s g_\mu, \tag{1.11}$$

where $\mathrm{e_s}$ is the gauge coupling constant, g_μ is the sum of the gluon fields $\mathrm{g_\mu^A}$ and the QCD generators as shown:

$$g_\mu = \sum_{A=1}^{8} T^A g_\mu^A, \tag{1.12}$$

We can now use both equations 1.11 and 1.8 to produce the Lagrangian density for QCD.

$$\mathcal{L}_{QCD} = -\frac{1}{4}\sum_{A=1}^{8} F^{A\mu\nu}F^A_{\mu\nu} + \sum_{j=1}^{n_f} \bar{q}, (iD_\mu\gamma^\mu - m_j)q. \tag{1.13}$$

where γ^μ are the Dirac matrices, $\mathrm{q_j}$ are the quark fields with mass $\mathrm{m_j}$ and number of flavors $\mathrm{n_f}$ [1].

QCD has been observed to follow P and T symmetry and consequentially conserve CP symmetry with its Lagrangian defined in equation 1.13 is symmetric under CPT transformations. The CP symmetry of this group is more complicated than disscussed here however we do go into this issue further in section 1.2 on the strong CP problem.

1.1.2 Electroweak unification

For the electroweak interaction (EW), we have the unification of two theories and groups. The $U(1)$ group which controls electromagnetic interactions through the emission or abortion of a photon with its quantum field theory quantum electrodynamics (QED) and the $SU(2)$ groups which controls the weak interaction, mediated by the $\mathrm{W}^{\pm}$ and Z^{0} bosons. We will first cover QED.

QED is an abelian gauge theory with its associated gauge field being the photon. B_{μ} is the photon field. The photon mediates the electromagnetic interaction which controls the charged interactions of charged particles. The $U(1)$ group has a single charge generator Q. We can now define the covariant derivative from equation's 1.4 format for QED as [1]:

$$D_{\mu} = \partial_{\mu} + ieB_{\mu}Q, \tag{1.14}$$

where Q can be $+1$ or -1 depending on the charge of the particle. The Lagrangian density for QED can now be defined using equation 1.14 and equation 1.8 to produce:

$$\mathcal{L}_{\mathcal{QED}} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \sum_{i}\bar{\psi}_{i}(iD_{\mu}\gamma^{\mu} - m_{\psi})\psi_{i}, \tag{1.15}$$

where ψ represents the wave functions of each quantum state involved.

We combine QED with the weak interaction into one unified electroweak theory mediated by a new boson, which could allow for neutral decay interactions mediated by the Z boson. The Lagrangian density for this electroweak interaction can be divided into two sections [1]

$$\mathcal{L}_{EW} = \mathcal{L}_{gauge} + \mathcal{L}_{Higgs}. \tag{1.16}$$

We first start with the $\mathcal{L}_{gauge}$, with $\mathcal{L}_{Higgs}$ being presented in section 1.1.

We also need to define the covariant derivative D_{μ} for EW theory, with a Lagrangian, using the formatting from equation 1.4

$$D_{\mu} = \left[\partial_{\mu} + ig\sum_{A=1}^{3}t^{A}W_{\mu}^{A} + ig'\frac{Y}{2}B_{\mu}\right]\phi, \tag{1.17}$$

where t^{A} and $Y/2$ are the $SU(2)\otimes U(1)$ generators with their respective coupling constants g and g'; W_{μ}^{A} are the gauge fields corresponding to the three $SU(2)$ group generators and where $B_{\mu\nu}$ is the gauge tensor constructed from the gauge field B associated with the $U(1)$ group defined as

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu,$$

We will now define the $\mathcal{L}_{gauge}$ using equation 1.8 and equation 1.15 for gauge bosons and fermions

$$\mathcal{L}_{gauge} = -\frac{1}{4}\sum_{A=1}^{3} F^{A\mu\nu}F^{A}_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} + \bar{\psi_L}(iD_\mu\gamma^\mu)\psi_L + \bar{\psi_R}(iD_\mu\gamma^\mu)\psi_R, \tag{1.19}$$

where $F^{A}_{\mu\nu}$ is defined as

$$F^{A}_{\mu\nu} = \partial_\mu W^{A}_{\nu} - \partial_\nu W^{A}_{\mu} - g\epsilon_{ABC}W^{B}_{\mu}W^{C}_{\nu}, \tag{1.20}$$

and ϵ_{ABC} are the groups fine structure constants, and g is the gauge coupling for the weak interaction.

In equation 1.19 we also see left and right-handed fermion fields ψ_L and ψ_R, which for gauge field theories such as QED and QCD transform in the same way. This is because ψ_L and ψ_R have the same electric charge and the same colour. On the other hand, if ψ_L and ψ_R transform differentially under the gauge group transformations, this means we have a chiral gauge theory such as EW theory. The theory has to be chiral to allow for charge parity (CP) violations seen in experiments such as from kaons[6]. QED is also known to to violate C P and T symmetries individually, however, the combined CPT symmetry is maintained.

Typically, for fermions to have mass, the cross term $\bar{\psi_L}\psi_R + h.c$ needs to be allowed, however, this is only possible for gauge theories and not for chiral gauge theories due to the asymmetric way in which left and right-handed fermion fields transform. The $SU(2)\otimes U(1)$ symmetry for electroweak theory, therefore, does not allow for its fermions or its bosons to have mass. This is solved later in section 1.1.3 by introducing the Higgs field.

The charged W_μ fields are described by $W^{1,2}_{\mu}$ while the photon A_μ and the new weak neutral gauge boson Z_μ can be described by combinations of W^{3}_{μ} and B_μ. We defined both A_μ and Z_μ by using the weak mixing angle θ_W [1]

$$A_\mu = B_\mu\cos(\theta_W) + W^{3}_{\mu}\sin(\theta_W), \tag{1.21}$$

$$Z_\mu = -B_\mu\sin(\theta_W) + W^{3}_{\mu}\cos(\theta_W), \tag{1.22}$$

where the weak mixing angle determines the amount of mixing between weak and electromagnetic terms.

In total the $\mathrm{SU}(2) \otimes \mathrm{U}(1)$ group has 4 generators and therefore 4 gauge bosons: The previously mentioned γ, the W^+ and W^- for the weak interaction, and the Z-boson that mediates the united electroweak interactions. The range of these forces is controlled by their mass, making the range of the electromagnetic force effectively limitless as the photon is massless. The generators for $\mathrm{W}^\pm$ bosons and Z-boson have their symmetry spontaneously broken (discussed later in section 1.1.3) and therefore have large masses as shown in Table 1.1. This limits the range of both the weak and electroweak forces.

1.1.3 Spontaneous symmetry breaking

Principle 3 about the Higgs field fixes the mass issue for the Z and W-boson. If the $\mathrm{SU}(2) \otimes \mathrm{U}(1)$ symmetry held, then the Z and W-boson should be massless. However, this conflicted with the range of the weak interaction, which was known since the 1930s to have a very short range [1].

In electroweak theory, breaking the system's symmetry is known as spontaneous symmetry breaking (SSB). This symmetry is intrinsically tied to the systems minimum energy also known as the vacuum energy. This vacuum state is effectively the start of our universe for the SM. If this vacuum energy can be found in more than one location then it is not unique and there exists a continuum of degenerate states. This would break the symmetry by having the system, of infinite volume and degrees of freedom, be able to exist in any individual vacuum state. This choice has to be made especially for the SM, where there can only be one vacuum state and making this choice breaks the symmetry[1].

The Higgs mechanism adds a scalar field in the form of the Higgs field that creates a system with multiple degenerate vacuum states shown in figure 1.1[4].

The Higgs mechanism allows for the breaking of the $\mathrm{SU}(2) \otimes \mathrm{U}(1)$ symmetry, letting the Z boson and the W bosons have mass. In fact, the Higgs mechanism is coupled to all massive particles, allowing quarks and leptons to have mass. The Higgs boson mediates this interaction, its properties are shown in table 1.1.

To derive the Higgs Lagrangian we first consider a standard Lagrangian with a kinetic and potential term [1]

$$\mathcal{L} = \frac{1}{2}|D_\mu \phi|^2 - V(\phi), \tag{1.23}$$

where ϕ is the scalar field for the Higgs, $V(\phi)$ is the Higgs potential, which is expressed in the form

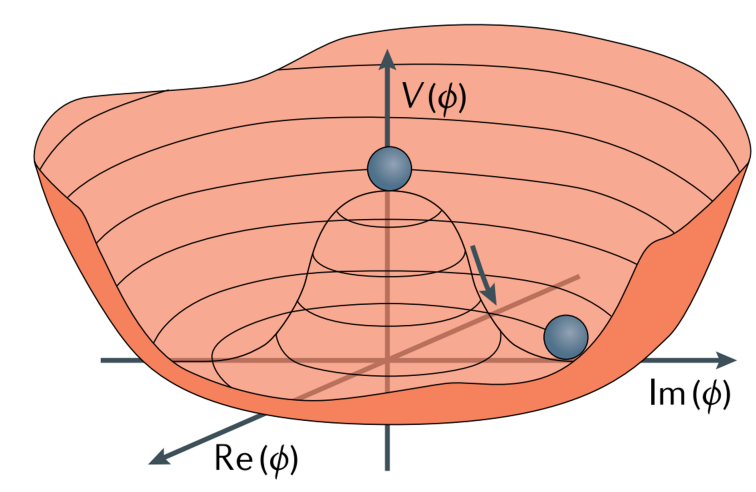


Figure 1.1: The Higgs potential $\mathrm{V}(\phi)$ for the scalar field ϕ and mass parameter $\mu^2 < 0$. [7]

$$V(\phi) = -\frac{1}{2}\mu^2\phi^2 + \frac{1}{2}\lambda(\phi^*\phi)^2. \tag{1.24}$$

Where μ is the mass term for the system, and λ is a constant. These two terms can be used to define the degenerate ground state $v = \phi^0$ that minimizes $V(\phi)$ as $v^2 = \mu^2/\lambda$. This mass term had to be negative to give the potential its unique shape shown in figure 1.1. This negative mass term is therefore also responsible for the SSB, as the minimum of the potential is no longer a single state but an orbit of states. Substituting the potential defined in equation 1.24 into equation 1.23 we get a simplified form for the Higgs Lagrangian [1]

$$\mathcal{L}_{Higgs} = \frac{1}{2}|D_\mu\phi|^2 + \mu^2\phi^*\phi - \frac{\lambda}{2}(\phi^*\phi)^2. \tag{1.25}$$

Now, we add the chiral terms

$$\mathcal{L}_{Higgs} = \frac{1}{2}|D_\mu\phi|^2 + \mu^2\phi^*\phi - \frac{\lambda}{2}(\phi^*\phi)^2 - \bar{\psi}_L\Gamma\psi_R\phi - \bar{\psi}_R\Gamma^*\psi_L\phi^*. \tag{1.26}$$

where Γ are the matrices which contain all the coupling constants, and are designed to make Yukawa couplings invariant under the Lorentz and gauge groups.

Now that we have the Higgs Lagrangian we want to show how this gives particles mass. We start with the vacuum expectation value

$$\langle 0|\phi(x)|0\rangle = v = \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{1.27}$$

Since we require that the photon remain massless, we have to set

$$Q|v\rangle = (t^3 + \frac{Y}{2})|v\rangle = 0. \tag{1.28}$$

We now move on to how the Higgs potential gives mass to the W boson. The W gains its mass from the quadratic terms in the W field. We substitute v into W field to get

$$m_W^2 = \frac{1}{2} g^2 v^2, \tag{1.29}$$

and by a similar process we also get the mass for the Z boson

$$m_Z^2 = \frac{g^2 v^2}{2\cos^2\theta_W}. \tag{1.30}$$

This gives the mass terms for the W and Z boson made of non-zero constants fixing the boson mass issue by breaking the $SU(2) \otimes U(1)$ symmetry with the Higgs potential. The values which these constants give for the masses of the two bosons and the mass of all fermions can be seen in table 1.1.

1.2 Beyond the Standard Model

The standard model's predictions discussed in the previous section have been tested to high levels of precision as shown in figure 1.2, summarizing the measurements made by the ATLAS collaboration in the past decade. Any statistically significant deviation between the SM prediction of a cross-section (XS) and the data measured for the XS would suggest physics beyond the standard model. So far, the SM has been confirmed with resounding success. The standard model, regardless of the XS measurements, has several limitations, which allow for further research into beyond standard model physics. Below is a list of the various limitations with the standard model:

The Fermion Problem: The entire universe could be formed out of particles from the 1st generation shown previously in table 1.1. However, we know there are 3 generations with no explanation from the SM as to why these particles should exist. The SM also doesn't explain the fermion masses that appear in a hierarchical pattern. On top of this, there is no explanation for why there is matter anti-matter asymmetry, with the universe predominantly made of matter. The CP violation from observed SM processes, such as kaon decays, natural D meson's decays, and strange B mesons [6][9][10], do not account for the large discrepancy between matter and anti-matter. Almost any kind of new Beyond SM physics or mechanisms could lead to new sources of CP violation, which can bridge the gap between the current measured CP violation and the CP violation needed for the universe we have today [2].

The Strong CP Problem: The weak interaction has been observed to have CP symmetry violation. An additional term can be added to the QCD Lagrangian density for it to allow the violation of P,T and CP symmetry. The addition of this term however would create an electric dipole moment for the neutron $\mathrm{d_n}$. An experimentally deduced limit is placed

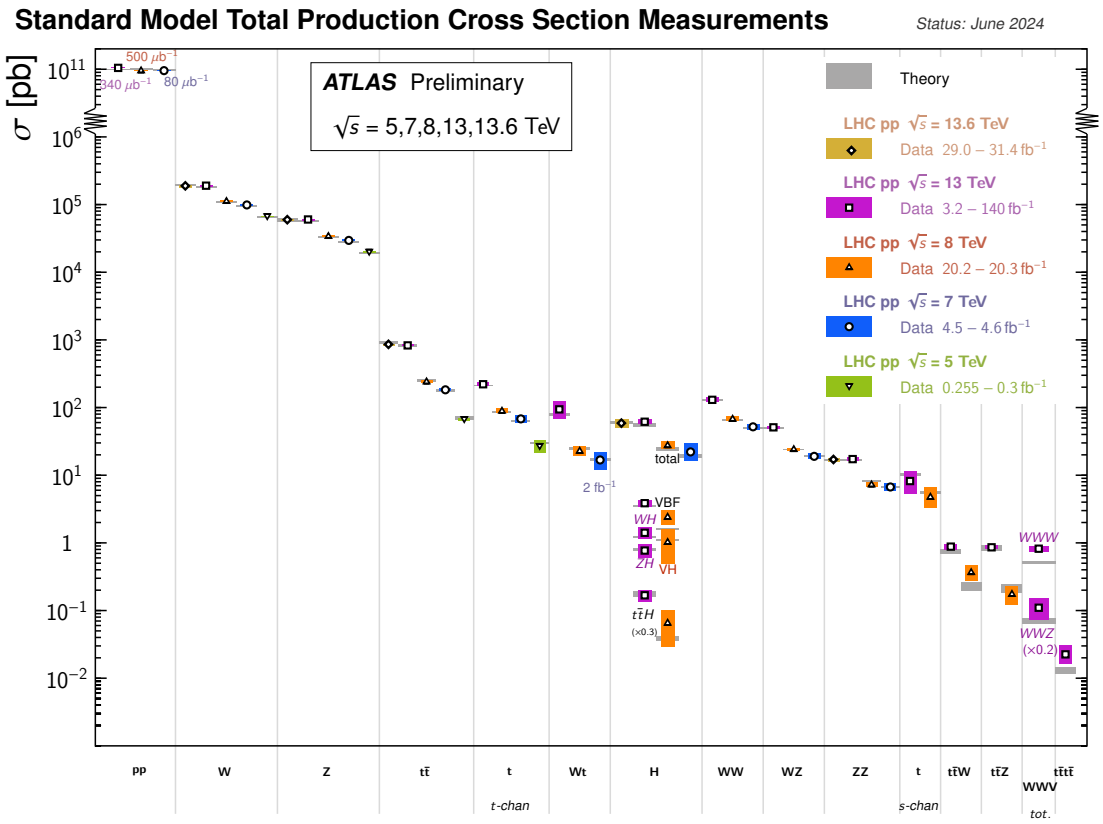


Figure 1.2: Summary of several Standard Model cross-section measurements. The measurements are corrected for branching ratios and are compared to the corresponding theoretical expectations. [8]

$$|d_n| < 3.0 \times 10^{-26} e \cdot cm. \tag{1.31}$$

This stringent limit sets an upper bound on the strength of the QCD CP-violation that is much higher than the measured strength of QCD. The fine-tuning required to correct for this is quite large and unnatural. Many solutions to this problem have been attempted but end up creating as many problems as they solve[2].

The Gauge Symmetry Problem: The SM is made up of three subgroups shown previously in equation 1.9. There isn't anything within the SM that should suggest that only the electroweak part of this symmetry is chiral and parity-violating. In a similar vein the SM does not infer why there is charge quantization or why the charges are in mutiples of e/3. The charge quantization is crucial to allowing charge neutrality which infers some kind of underling mechanism, such as a grand unification theory (GUT) to facilitate this. This GUT would not treat measured constants as axioms and instead provide mechanisms or generators for these properties, such as the quantitation of charge [2].

The Hierarchy Problem: The Higgs mechanism discussed in section 1.1 introduces the Higgs field which in turn provides mass to elementary particles. The Higgs field also self interacts to produce a mass for the Higgs boson. The mass of the Higgs boson is around $125\ \mathrm{GeV}$, which is similar to the W and Z mass at around $80\ \mathrm{GeV}$ and $91\ \mathrm{GeV}$ respectively. This observed mass is somewhat expected as the mass dictates the strength of an interaction and the Higgs-self interaction should not be much stronger than the weak interaction[2]. This is where the problems start arising. The equation for the mass of the Higgs boson is shown below[2].

$$M_H^2 = (M_H^2)_{bare} + O(\lambda, g^2, h^2)\Lambda^2, \tag{1.32}$$

Λ is the next highest scale in the theory. The $\mathrm{O}(\lambda, \mathrm{g}^2, \mathrm{h}^2)\Lambda^2$ term represents the Radiative corrections to the mass from both the Higgs self-interaction and its interaction with all other massive particles. The $(\mathrm{M_H})_{\mathrm{bare}}$ term represents the mass of the Higgs boson without any corrections due to virtual particles. Since these corrections are quadratically divergent, the correction becomes very large, which makes the value for $\mathrm{M_H}$ much larger than the observed value. This requires a large degree of fine-tunning of the $(\mathrm{M_H^2})_{\mathrm{bare}}$ term to the order of 10^{34}, making the theory unnatural. Theories trying to correct this issue have arisen like Supersymmetry (SUSY), which suggests new physics allowing for particles that are supersymmetric partners of already existing particles. These supersymetric partners would cancel out the quadratically divergent contributions from their SM physics partner, avoiding the issue of fine-tuning.

Gravity and General relativity: Gravity was not included as an interaction in the standard model. Gravity is not quantized and it is not unified with the other forces. General relativity is also not a quantum theory and it is not obvious how to include it in the Standard model. The cosmological constant predicted by General relativity (GR) can be thought of as the vacuum energy of the universe. However, the Higgs mechanism in the SM creates a vacuum expectation value and therefore would affect this constant. Attempts to merge these two mechanisms use two values for the cosmological constant: Λ_{bare} is the cosmological constant produced by GR and Λ_{H} the cosmological constant produced by the Higgs mechanism. The total cosmological constant is expected to be [2],

$$\Lambda_{Cosm} = \Lambda_{bare} + \Lambda_H. \tag{1.33}$$

The value for Λ_{H} is computed as $10^{56}\Lambda_{\mathrm{obs}}$, where Λ_{obs} is the observed cosmological constant measured in experiments. This is not physical, and the level of fine-tuning to make this value reasonable is unacceptable [2]. There is no accepted solution to this issue and therefore it remains an active field of study.

Dark matter: There is plenty of evidence from galactic rotation curves, gravitational lensing and the Cosmic microwave background that there is substantial undetected or dark matter in the universe. There are many dark matter candidates being suggested. Some of these candidates are already proven SM objects that are just difficult to observe. However, these types of candidates have been for the most part disproved. The most likely candidate is a particle outside of the SM. There is a small chance that dark matter or at least a portion of it could be primordial black holes(PBH). These PBH would have been formed in the very early universe and would have various limits placed on their mass. The narrow mass ranges at $10^{-13} - 10^{-8}\mathrm{M}_{\mathrm{sol}}$ and $10\mathrm{M}_{\mathrm{sol}}$, where $\mathrm{M}_{\mathrm{sol}}$ is the mass of the sun, would give black holes long-lived enough to have been formed close to the big bang and in a large enough density to be the cause of dark matter [2]. There is some experimental evidence to suggest that PBH could constitute small amounts of Dark Matter but not the entirety of it. It is therefore quite likely that Dark matter is some form of new particle, such as the Axion or WIMPs (weakly interacting massive particles). None of these new suggested particles have been discovered, leaving the field open for further study.

SM as an effective field theory: The standard model can also be used as an effective field theory (SMEFT)[11]. The SMEFT is a generalization of the standard model constructed out of a series of higher-dimensional operators. These operators operate in fields provided by the SM, including the Higgs field. The Lagrangian for this SMEFT is shown in equation 1.34 [11].

$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \mathcal{L}^{(5)} + \mathcal{L}^{(6)} + ... \tag{1.34}$$

Where $\mathcal{L}^{(5)}$, $\mathcal{L}^{(6)}$ and so on can be defined in the general sense as [11]:

$$\mathcal{L}^{(d)} = \sum_{i}^{N_d} \frac{C_i^{(d)}}{\Lambda^{d-4}} Q_i^{(d)} \tag{1.35}$$

The equations 1.34 and 1.35 are limited to above 4 dimensions by the cutoff scale

$\Lambda^{\mathrm{d}-4}$, and have $\mathrm{N_d}$ as the number of dimensions [11]. The strength of each of these new interactions represented by each of these higher-dimensional operators $Q_i^{(d)}$ is given by the Wilson coefficients $\mathrm{C_i^{(d)}}$.

We currently know the number of meaningful operators which cover the Lagrangian corrections shown in equation 1.35 up to dimension-8. Some of these SMEFT operators, in particular, are useful in top physics (discussed in section 5.1) using dimension-6 operators. These dimension-6 operators couple the 3rd generation quarks, such as the top, to electroweak bosons, such as the Z boson. These dimension-6 operators are useful as they have additional Higgs field terms that usually cancel out at lower dimensions. These Higgs terms cause a new minimum potential ν and create a new pre-factor $\frac{\nu^2}{\Lambda^2}$ which is not dependent on the energy of the interaction like the operators below dimension-6 $\frac{\mathrm{E}^2}{\Lambda^2}$ [12]. The Higgs new pre-factor allows these operators to be measured at lower energies as seen at the LHC [12].

The SMEFT is a useful framework as it provides characterizations of the low-energy limit of beyond-standard-model physics. These limits can be used to compare the Monte Carlo simulations that are produced with only the SM model assumed against data, which can show BSM behaviour[11].

Chapter 2

The LHC and ATLAS

In this chapter, we set the experimental basis for both the LHC and ATLAS. In section 2.1 we briefly explain the layout of the LHC and how it provides the beam to ATLAS. In section 2.2 and its subsections, we explain how the ATLAS detector functions. This includes the definition of the ATLAS coordinate system and where each sub-detector fits in that system, as well as how each sub-detector fulfills its role.

2.1 Proton creation and acceleration at the LHC

The Large Hadron Collider (LHC) is a two-ring superconducting hadron accelerator and collider as shown in figure 2.1. The LHC was built inside the LEP 26.7 km tunnel [13]. The process of creating a beam starts with hydrogen atoms being fired at a solid target that strips their electrons. The resultant protons are then accelerated using a series of accelerators, they are first sent to a series of linear accelerators to ramp up the protons to an initial energy of 50 MeV. The 50 MeV protons are then injected into the proton synchrotron, which increases the proton energy up to 1.4 GeV. These 1.4 GeV protons are injected into the super proton synchrotron to be accelerated to 450 GeV. Finally, these protons are sent to the main LHC ring, which completes the final acceleration of up to an energy of 6.5 TeV per beam [14].

The layout of the LHC follows the LEP tunnel with 8 arcs and 8 straight sections [13]. The straight sections are designed for experimental use, and particles are accelerated in a straight line using an electric field produced by radio frequency (RF) cavities, which oscillate at 400 MHz radio frequency. This oscillation is timed to change the electric field so that protons with the desired beam energy receive no acceleration. Protons with slightly different energies that arrive early or later than the ideally timed protons will accelerate or decelerate until they are at the correct energy. This will group the beam into discrete packets of protons separated by short time intervals (25ns) known as bunches [16].

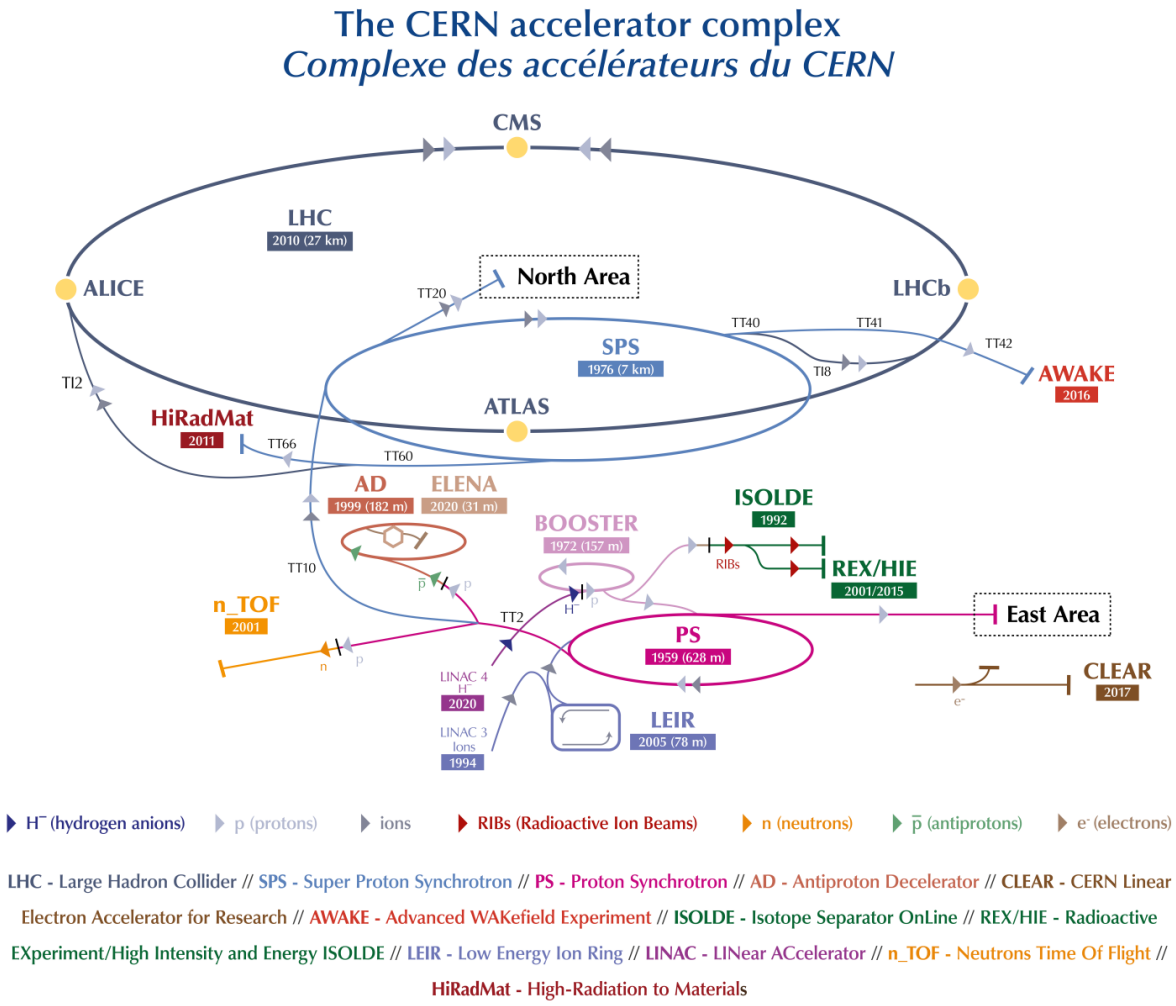


Figure 2.1: A diagram of the CERN accelerator complex [15].

The LHC ring also uses several magnet systems, such as the dipole and quadrupole magnet systems. The dipole magnets bend the beam around at the curved sections. The dipole magnet system consists of 1232 superconducting dipole magnets, each 15m long. The quadrupole magnets focus the beam. There are 474 quadrupole magnets, with each quadrupole being around 3m long. Both types of magnets use superconductive materials and are kept in cryostats at temperatures of 1.9K [13].

The two beams of 6.5 TeV collide at the 8 interaction points (Only 4 of which are used for major experiments), giving a center of mass energy of 13 TeV[13]. Figure 2.1 shows the four main detectors placed at where the beams collide: A Toroidal LHC ApparatuS (ATLAS), Compact Muon Solenoid (CMS), A Large Ion Collider Experiment (ALICE) and LHC beauty(LHCb) are placed at of the 4 interaction points along 4 of the straight sections.

ATLAS will be discussed further in section 2.2.

Multiple collisions happen for every bunch crossing at the interaction point (IP), hence, there are other collisions of interest occurring. Moreover, the detector in question might not be finished processing the previous collision before the next one starts. These are both forms of pile-up in the detector and need to be dealt with. Part III discusses this topic in detail.

2.2 The ATLAS detector

The ATLAS detector was created to measure collisions at the LHC in hopes of discovering evidence for the Higgs boson. The detector succeed in this search and was repurposed to probe rare and beyond standard model physics by detecting particles that are produced in proton-proton collisions by the LHC.

The ATLAS detector is situated at one of the 4 interaction points described in section 2.1[17]. The detector is 44 m in length, 25 m in height and weighs around 7000 tons. A drawing of the full detector is shown in figure 2.2.

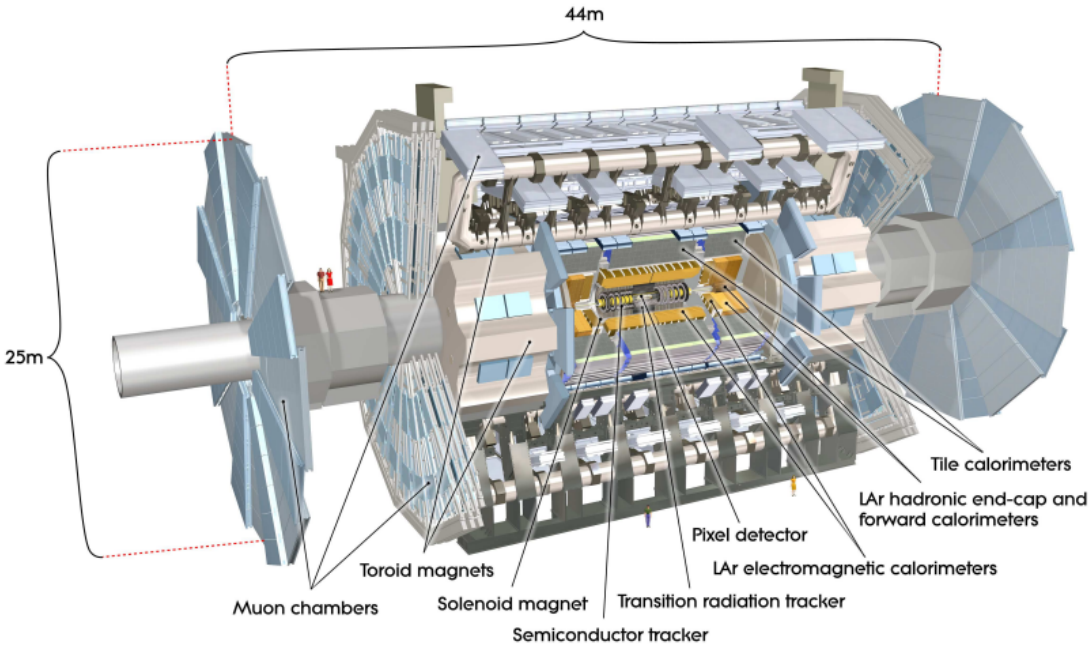


Figure 2.2: A cut-away view of the the ATLAS detector [17]

The detector, as shown in figure 2.2, has many interconnecting specialized sub-systems that work together to detect all the different types of particles and their properties. The highly accelerated protons in the LHC beams discussed in section 2.1 collide at the centre

of the detector, known as the interaction point (IP).

The main subsystems of the detector are: The inner detector (ID), hadronic and electromagnetic calorimeter, forward detector, magnet systems, muon chambers, as well as various shielding for background radiation.

2.2.1 Coordinate System

Before we move on to the workings of the subsystems, we first need to define a coordinate system by which we can define both the range of the ATLAS detector that the subsystems cover and the positions of particles within the detector. The interaction point (IP) is the origin of our coordinate system for the ATLAS detector. In Cartesian coordinates, the beam axis direction defines the Z-axis (with the positive Z-axis direction being towards the beamline that leads to LHCb in figure 2.1) while the x-y plane is defined as being transverse to the beam direction. The x-axis points towards the centre of the LHC ring, and the y-axis is defined pointing upwards. The detector itself is cylindrical so using a radial distance (R) from the IP is a sensible choice. The azimuthal angle ϕ is measured from the beam in the x-y plane known as the beam axis, while θ is measured from the beam axis. The transverse energy, momentum and transverse missing energy are all defined in this plane. These variables are defined later in section 6.3. The particles coming out of the ATLAS IP are relativistic therefore, a Lorentz invariant coordinate system arises. For massive objects such as jets, this is the rapidity defined as [17]

$$y_r = \frac{1}{2}\ln(\frac{E + p_Z}{E - p_Z}), \tag{2.1}$$

Where E is the energy of the particle involved and $\mathrm{p_z}$ is the momentum in the Z axis. In cases where the objects can be approximated as massless, we use the pseudorapidity defined in equation 2.2.

$$\eta = -\ln(tan(\frac{\theta}{2})) \tag{2.2}$$

The η and $\mathrm{y_r}$ coordinate systems are useful since differences in rapidity are Lorentz invariant under boosts along the longitudinal axis, while θ is not. The distance in pseudorapidity-azimuthal angle space is $\Delta\mathrm{R} = \sqrt{\phi^2 + \eta^2}$. The ATLAS detector was designed around the sub-detectors being segmented in η. Therefore, the following sections will define sub-detector positions in terms of η.

2.2.2 The ATLAS magnet systems

The magnet systems used in the ATLAS detector are a hybrid system of four large superconducting magnets. These four magnets are the solenoid, the barrel toroid, and the two

end-cap toroids. Figure 2.3 shows the layout of the four magnets.

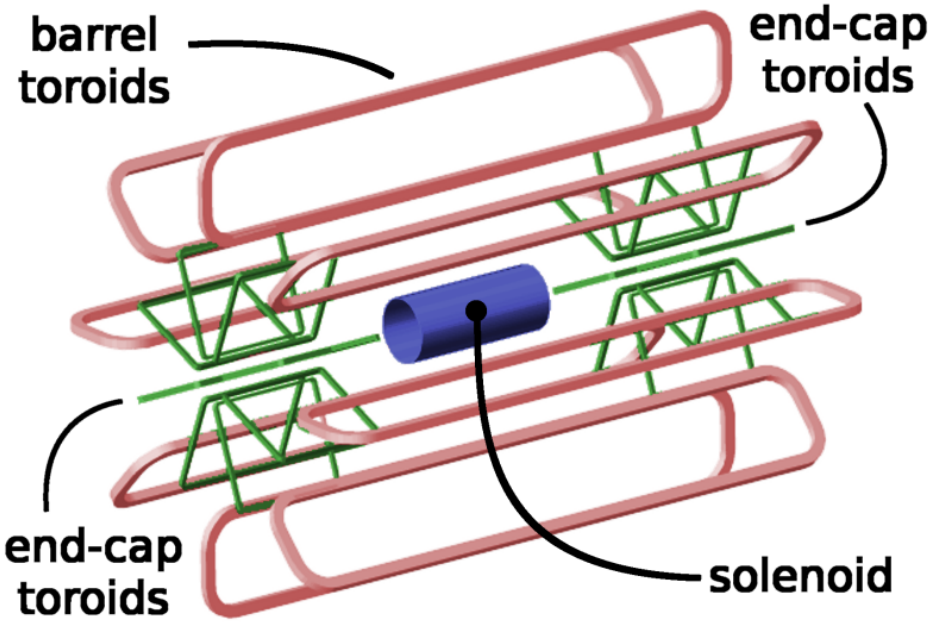


Figure 2.3: The ATLAS magnet systems [18].

The solenoid is hollow with an inner and outer diameter of 2.46 m and 2.56 m, respectively, with its axial length being 5.8 m. The solenoid surrounds the Inner Detector (ID), shown in figure 2.4, is aligned along the beam axis and generates a 2 T magnetic field. The barrel toroid is 25.3 m long with an inner and outer diameter of 9.4 m and 20.1 m, respectively. The two end-cap magnets have a length of 5 m and an inner and outer diameter of 1.65 m and 10.7 m, respectively. The barrel toroid and the two-end cap toroid produce a toroidal magnetic field of approximately 1 T for the muon detectors in the endcap regions and 0.5T for the muon detectors in the central region. Muon detectors are discussed later in 2.2.5.

2.2.3 The inner detector

The inner detector is the closest part of the detector to the IP, made up of 3 sub-components: the pixel, the silicon microstrip tracker (SCT) and the Transition Radiation Tracker (TRT). These sub-components work together to track and identify particles with high levels of precision from high detector granularity. The layout of the inner detector is shown in figure 2.4.

The tracking components of the ID cover $|\eta| < 2.5$ with an extremely high tracking resolution 2.3[17]

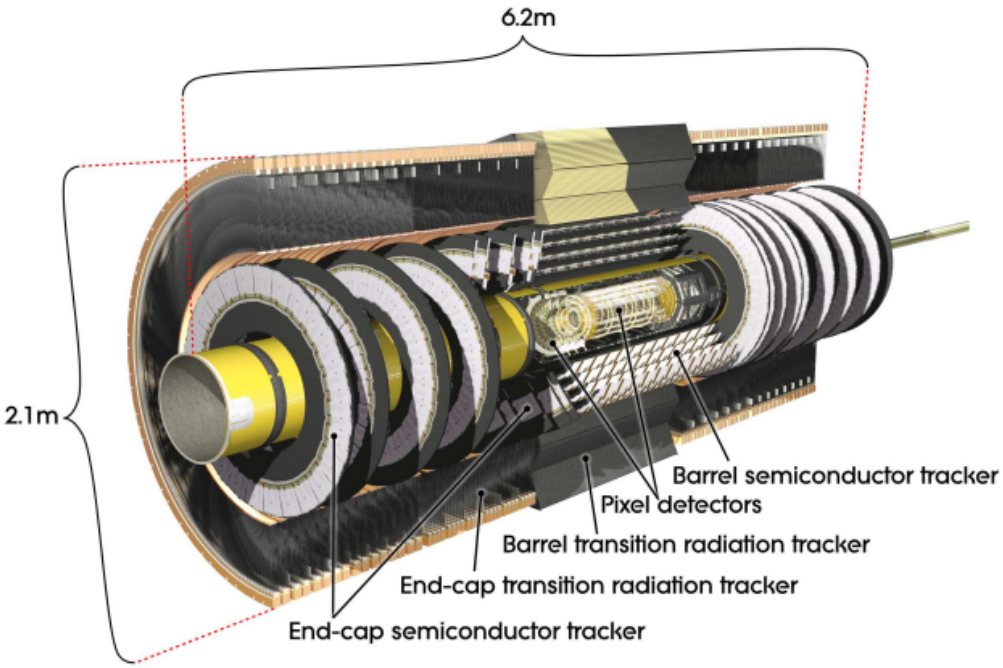


Figure 2.4: A cut-away view of the ATLAS Inner detector(ID) [17]

$$\frac{\sigma_{p_T}}{p_T} = 0.05\% p_T \oplus 1\%. \tag{2.3}$$

The ID is immersed in the 2 T magnetic field generated by the solenoid. In the barrel region, the SCT and pixels are arranged on concentric cylinders around the beam axis, while in the end-cap regions, they are located on disks perpendicular to the beam axis. The pixel detectors themselves are situated closest to the beam line (see figure 2.5), while the SCT is located just outside the pixel detector. The TRT is the outside layer of the ID surrounding the SCT and pixel detector. This is all shown in figure 2.5 and 2.6.

The pixel detectors are formed of multiple pixel layers and disks for the barrel and end-cap, respectively. The pixel layers and disks have a combined total of 87.2 million pixel channels. The pixel layers are segmented in the $R-\phi$ plane and Z with three pixel layers in the barrel and 3 disks in each end-cap, as shown in figures 2.5 and 2.6. All pixel channels are identical and have a minimum pixel size in $R-\phi$ of 50 $\times$ 400 μm. The pixel detector achieves high tracking accuracy in the barrel of 10 μm in the $R-\phi$ plane and 115 μm in the z axis. In the end-cap disks the accuracies are 10 μm in the $R-\phi$ plane and 115 μm in the R axis.

The Insertable B-Layer(IBL) is an additional 4th layer added to pixel in Run-2 [19]. A figure of how its position within the pixel is shown in figure 2.7. The IBL was added to

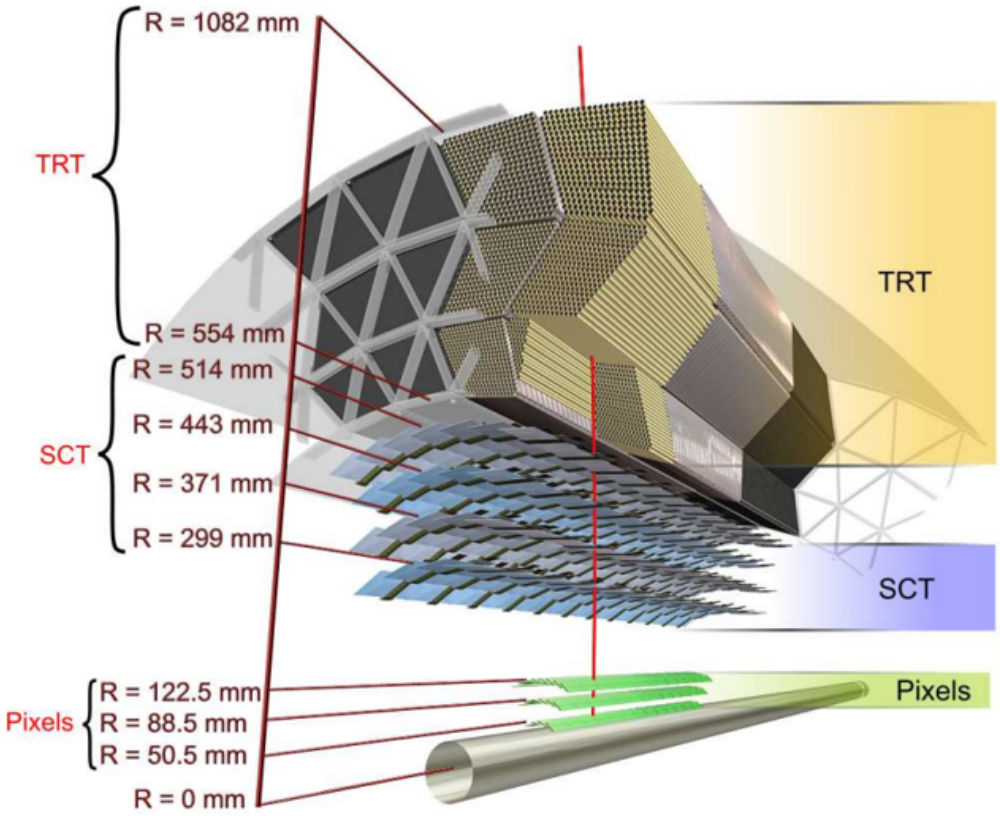


Figure 2.5: A drawing of the ATLAS inner detector showing the radial distribution of the Pixel, SCT and TRT along the barrel of the inner detector [17].

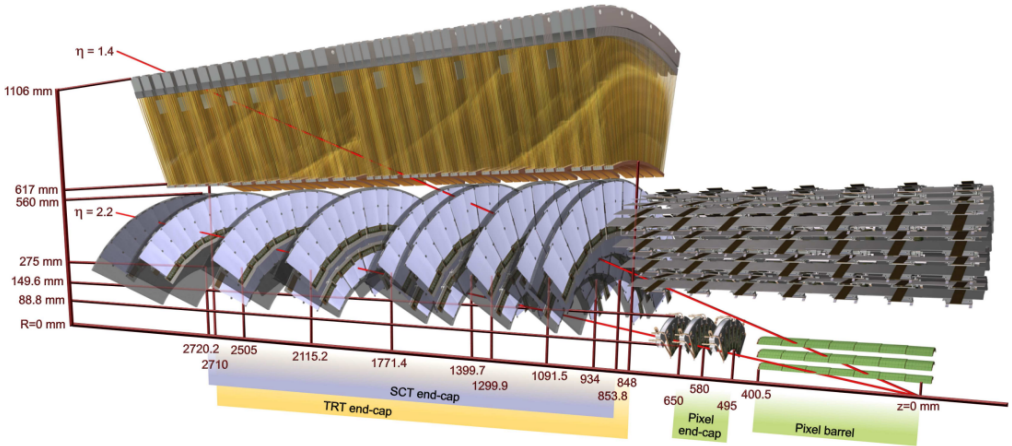


Figure 2.6: A drawing of the ATLAS inner detector showing the z distribution of the Pixel, SCT and TRT along the disks of the inner detector and its end-caps [17].

improve the tracking performance of the inner detector in particular for high luminosity scenarios as well as increase b-tagging efficiency. The IBL itself is the closest part of the ID

to the beam having an inner radius of 33 mm from the beam. The resolution of the IBL is 10 μm in the $R-\phi$ and 66.5 μm in the z-axis which is an improvement on the standard pixel resolution in the z-axis.

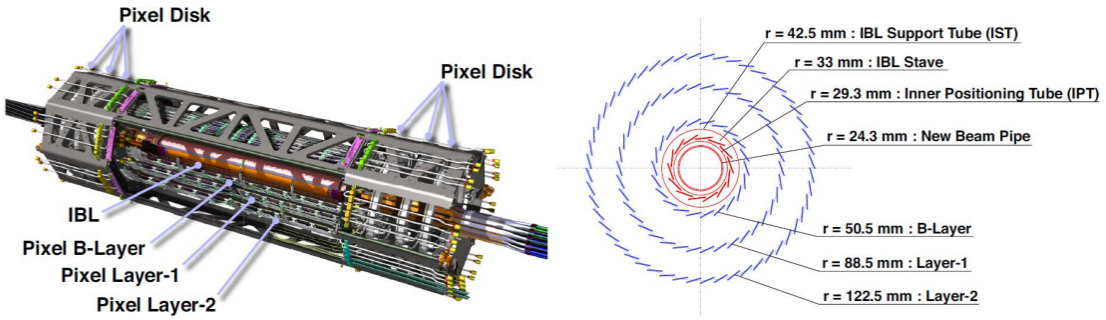


Figure 2.7: A diagram of the Pixel with the IBL inserted for Run-2. The IBL subsystems are highlighted in red. [19]

The SCT has 4 cylindrical layers in the barrel and 9 disks per end-cap as shown in figures 2.5 and 2.6. Each layer and disk consists of long silicon strips with 4088 modules spread between the barrel and the end-cap. The accuracy of the SCT is lower than that of the pixel at 17 μm in $R-\phi$ and 580 μm in Z in the barrel. The disks also have a lower accuracy of 17 μm in $R-\phi$ and 580 μm in R.

The TRT consists of 4 mm diameter straw tubes. The TRT allows for track-following for $\eta < |2|$. The straw tubes are filled with a mixture of gases such as CO_2 and xenon. These gases emit transition radiation when charged particles pass through them. The intensity of this radiation is proportional to the relativistic γ factor [20]. Therefore, the TRT is very effective at identifying electrons. The TRT provide $R-\phi$ information from its 73 straws in the barrel and 160 straws in the disk. The accuracy of the straws in $\mathrm{R}-\phi$ is 130 μm per straw. In the barrel, the straws are arranged parallel to the beam axis and are 144 cm long, while in the disk the straws are 37 cm long, arranged radially, as shown in figure 2.5 and 2.6.

2.2.4 The Calorimeter

The calorimeters used at ATLAS are sampling calorimeters, which use both a detecting medium and a denser absorbing material in an alternating layout. This ensures that the particles detected by the calorimeters are fully absorbed by it[17]. The calorimeters are used primarily for measuring the energy of particles that exit the inner detector. The calorimeters are split into two main parts, the hadronic (HCAL) and electromagnetic calorimeters (ECAL). The layout of the various calorimeter parts is shown in figure 2.8.

The ECAL is located directly outside and surrounding the inner detector and solenoid, and it is ideally suited to reconstructing electrons and photons. The HCAL is located outside and around the electromagnetic calorimeter; Photons and electrons are fully absorbed before reaching the HCAL. The HCAL also has coarser granularity that would not be suitable for electron or photon reconstruction; however, it is suitable enough for the reconstruction of jets and $\mathrm{E_T^{miss}}$. The calorimeters needed to be thick enough to contain the electronic and hadronic showers to stop the punch-through of these showers in the muon chambers. To calculate variables like $\mathrm{E_T^{miss}}$, described later in section 6.3, the calorimeters cover the entire η range of the detector up to $|\eta| < 4.9$.

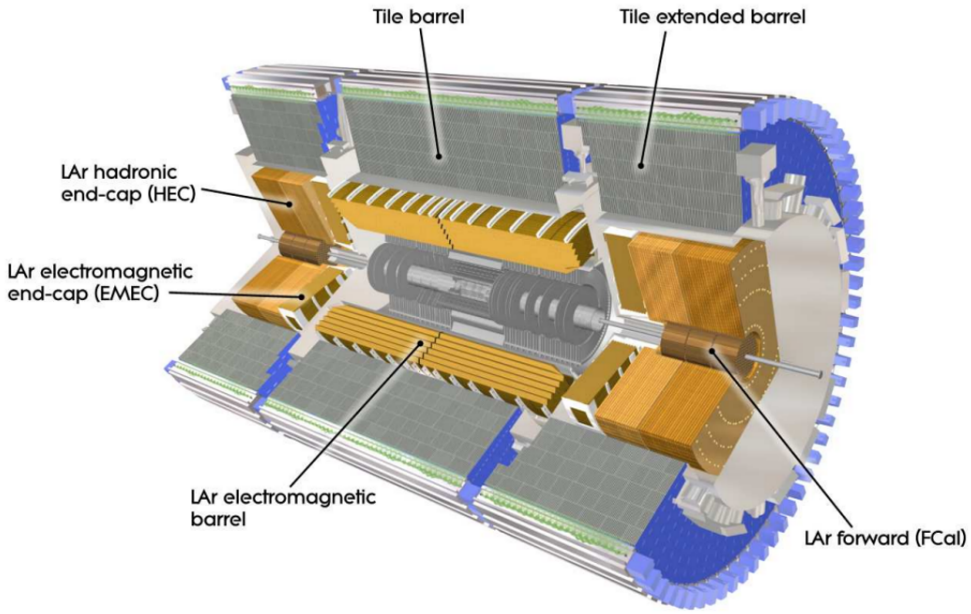


Figure 2.8: A cut-away view of the ATLAS Calorimeter[17]

The Liquid Argon Electromagnetic Calorimeter (LArEMCal) is broken up into a barrel portion and two end-cap portions, shown in figure 2.8. Both components consist of lead liquid argon detectors with accordion-shaped krypton electrodes and lead absorber plates. The accordion shape allows for complete ϕ symmetry. The barrel part of the LArEMCal covers the $\eta < |1.475|$ while the end-cap components cover the $1.375 < |\eta| < 3.2$ range. Both of these components are stored in a cryostat to keep the argon liquid. The LArEMcal in the barrel consists of two identical half barrels separated by a small gap of 4 mm at $Z = 0$. Each end-cap is divided into two coaxial wheels. The inner wheel, covering $2.5 < |\eta| < 3.2$, and an outer shell covering $1.375 < |\eta| < 2.5$. The detector is also segmented in layers: in the central region $|\eta| < 2.5$, the LArEMcal has 3 layers, in the end-cap, the LArEMcal has 2

layers; In the $|\eta| < 1.7$ region a single LAr-layer is placed on the surface of the innermost layer of the ECAL and used as a presampler detector. This is used to correct for energy lost from electrons and photons that did not reach the calorimeter. This has a 1.1cm thickness in the barrel and 0.5cm thickness in the end-cap.

The Hadronic Calorimeters are comprised of 3 main components: The Tile, the LAr hadronic end-cap calorimeter and the LAr forward calorimeter. The Hadronic calorimeter tile and end caps components are outside of and surround the LArEMCal.

The Tile Calorimeters is the part of the hadronic calorimeter which directly surrounds the LArEMCal. The tile consists of a barrel that covers the $|\eta| < 1$ range. It also has two extended barrels that cover the $0.8 < |\eta| < 1.7$. The barrels extend radially between $R = 2.28$ and $R = 4.25$ in 3 layers. The barrels and each end-cap have 64 modules, which consist of sampling calorimeters using steel as the absorber and scintillating tiles as the active material. The scintillating tiles are read out by wavelength-shifting fibres into two separate photomultiplier tubes to create the final signal.

The LAr hadronic end-cap calorimeter (HEC) is located directly behind the LArEMCal end-cap and consists of two independent wheels per end-cap. The end-cap components for both the hadronic and electronic calorimeters share the same LAr cryostat. The wheels are made of copper plates of varying sizes. There is a transition between the end-cap and the forward calorimeter at $|\eta| = 3.1$ to reduce a drop-off in material density, the HEC is extended beyond this transition point $|\eta| = 3.2$ to overlap with the forward calorimeter. The HEC also overlaps with the Tile calorimeter reaching $|\eta| = 1.5$. The Wheels are layered in two layers with 4 layers per end-cap. The wheels consist of 32 identical wedge-shaped modules.

The forward calorimeter (FCal) covers the range $3.1 < |\eta| < 4.9$. The FCal is integrated into the end-cap cryostats and consists of 3 modules in each end-cap. Each module consists of a metal matrix with regularly spaced longitudinal electrode structures. These electrode structures consist of concentric rods and tubes aligned parallel to the beam axis. The 3 modules are not identical with the first module being made of copper and optimized for EM measurements. The second and third modules are made of tungsten and mostly measure hadronic jets.

2.2.5 The Muon chambers

The Muon chambers, shown in figure 2.9, are the outermost part of the ATLAS detector and are designed to detect charged particles exiting the barrel and end-cap of the ATLAS

calorimeter. These charged particles should be muons, as they are not stopped by the calorimeter.

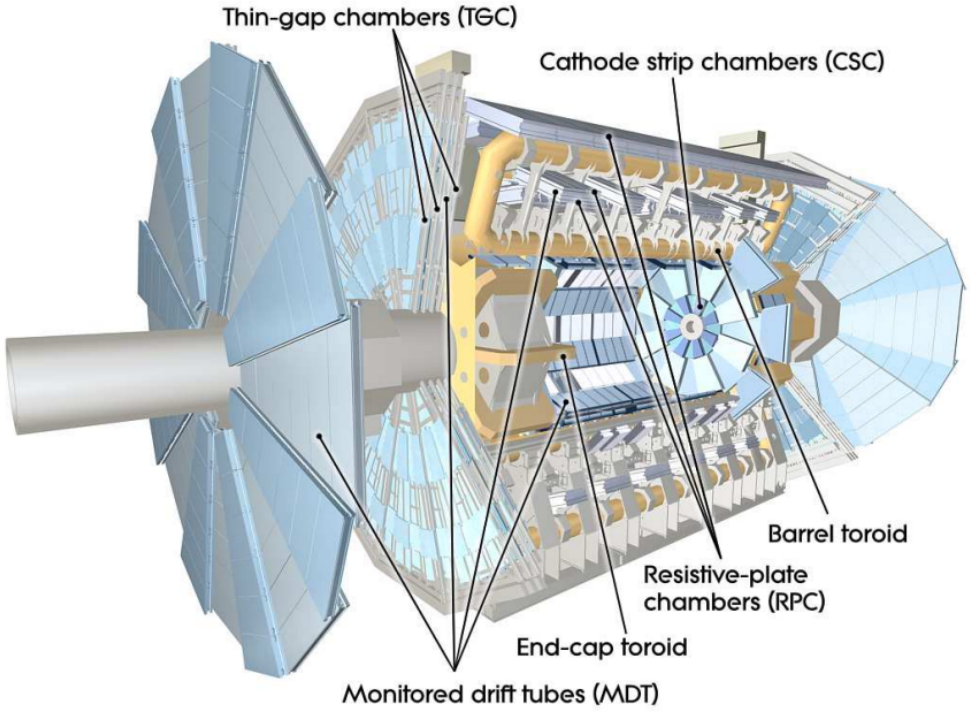


Figure 2.9: The ATLAS muon system [18].

The muon detection works by the deflection of muon tracks in a magnetic field. The magnetic field is provided by the toroid magnets $|\eta| < 1.4$ and two small end-cap magnets $1.6 < |\eta| < 2.7$ as shown in figure 2.3. This magnetic field bends the muons as they travel through the muon chambers.

The components that make up the muon spectrometer are Monitored Drift Tubes (MDT), Cathode Strip Chambers(CSC), Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC). The MDT cover the range $|\eta| < 2.7$ except in the innermost end-cap layer, where they reach $|\eta| < 2.4$. The MDT consists of 1088 chambers, each made of drift tubes. The CSC covers the range $2.0 < |\eta| < 2.7$, and consists of 32 chambers made of multiwire proportional chambers, which have cathode planes segmented in strips in the orthogonal directions. Both the CSC and MDT are dedicated towards the precision measurement of muon tracks. The RPC is a gaseous parallel electrode-plate detector and covers $|\eta| < 1.05$ with 606 chambers. This functions by having two resistive plates parallel to one another with an electric field between them while immersed in a gas. The TGC covers $1.05 < |\eta| < 2.7$ with 2588 chambers. The TGC are multi-wire proportional chamber with

the characteristic that the wire-to-wire distance is larger than the wire-to-cathode distance. The RPC and TGC are used to trigger muon tracking as well as measuring the muon's coordinate in the direction orthogonal to that determined by the precision-tracking chambers (the MDT).

The detection of charged particles occurs within the $|\eta| < 2.7$ while triggering of events on these particles occurs only within $|\eta| < 2.4$. The muon chambers in the barrel are arranged in three cylindrical layers around the beam axis. The muon chambers in the transition and end-cap region are also arranged in three layers although installed in planes perpendicular to the beam axis.

2.2.6 The forward detector

The forward detector consists of 3 subsystems: LUminosity measurement using Cerenkov Integrating Detector (LUCID), Absolute Luminosity For ATLAS (ALFA) and Zero-Degree Calorimeter (ZDC). LUCID and ALFA both measure the luminosity delivered to ATLAS, while ZDC measures inelastic proton-proton scattering in the forward region. LUCID consists of twenty aluminum tubes which encircle the beam-pipe. The tubes are oriented to point toward the interaction point. LUCID is the main online measure of the luminosity for ATLAS. ALFA consists of scintillating fibre trackers located inside Roman pots. ALFA acts under very specific beam conditions to measure luminosity and helps calibrate LUCID. The ZDC consists of alternating quartz rods and tungsten plates, which help to determine the centrality of heavy-ion collisions at $|\eta| \geq 8.2$

2.2.7 Radiation Shielding

There is large amounts of background radiation that need to be reduced through the use of shielding in ATLAS. The primary source of radiation at the ATLAS detector at full luminosity is the proton-proton collisions at the IP. In order to shield detector components from damage and interference from this radiation, the ATLAS detector uses 3000 tons of shielding arranged in 3 layers. The innermost layer to the beam is designed to stop high-energy hadrons using iron and copper plates. The second middle layer, made from doped polyethylene, stops neutron radiation that passes through the first layer. The outermost layer needs to stop photons, which are the most penetrating, therefore, the layer is made from steel and lead.

2.2.8 The trigger systems

The Trigger system at ATLAS use to consist of the L1 and L2 triggers as well as the event filter. The full ATLAS trigger system is shown in figure 2.10. The purpose of these triggers is to select the events which ATLAS will read and store, as only certain events may contain useful physics.

In run-2 the L2 trigger was merged with the event filter into the High-level Trigger (HLT) which allowed the trigger to better cope with the increase in luminosity between Run-1 to Run-2 [21].

The L1 trigger system searches for a wide array of high $\mathrm{p_T}$ particles such as: electrons, muons, hadrons (originating from tau decays), photons and jets. The L1 trigger selections are set based on calorimeter readout and trigger chambers in the barrel and end-cap for high $\mathrm{p_T}$ muons (discussed previously in section 2.2.5) as shown in figure 2.10. Once a particle is found matching the correct conditions for the L1 trigger they are then passed onto the next stage of the process. The L1 trigger collects data at a rate of up to a maximum of 100 kHz in run 2 [22]. The L1 trigger defines regions of interest (ROI) which are the (η, ϕ) coordinates of the interesting features. The ROIs are also defined with information about the interesting feature, such as the type of feature and trigger threshold passed.

The HLT then uses all the data within the ROI to see if it passes its selections. These initial selections are fast and reject events that won't yield good physics. The HLT then uses more CPU intensive algorithms which are similar to offline reconstruction to further select events. The HLT converts the 100 kHz of information from the L1 trigger to around a few hundred Hz that is then stored for offline analysis [21]. The HLT selection criteria also separates the events into streams based on their physics properties. Each of these physics streams is classified based on the physics objects that it believes caused the event. These tags allow for the offline processing of ATLAS data into the data streams. The controling of these streams is done through the trigger menu which contains thousands of trigger chains. These trigger chains is a series of selections which the event has passed or failed. Each trigger chain is tied to one or more physics streams. Figure 2.11 shows the streaming rates for the various physics streams as well as the streams overlap.

The data recorded by the triggers is then processed further, where objects such as tracks, jets, electrons, muons and photons are reconstructed using the information passed to them by the triggers. This is discussed in section 3.1.

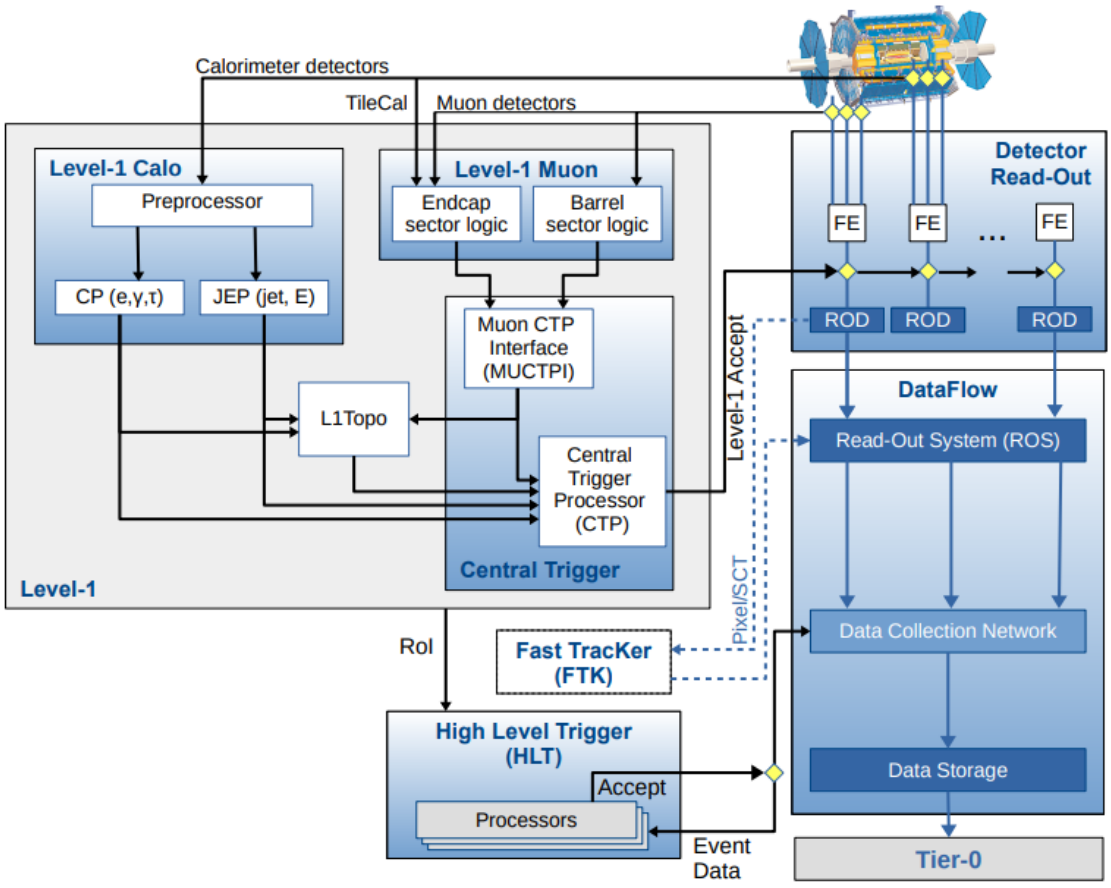


Figure 2.10: The ATLAS TDAQ system in Run 2 showing the components relevant for triggering as well as the detector read-out and data flow [21]

Stream	e	μ	Jet	γ	E_T^{miss} & τ	B-physics
e	31 ± 7.9	0.0056 ± 0.00058	$0.00053 \pm 6.2 \times 10^{-5}$	1.2 ± 0.4	1.4 ± 0.035	$(1.3 \pm 1.3) \times 10^{-5}$
μ	–	34 ± 8.7	0.021 ± 0.015	0.0028 ± 0.002	0.22 ± 0.022	0.076 ± 0.0043
Jet	–	–	38 ± 5.9	0.48 ± 0.4	0.71 ± 0.4	0 ± 0
γ	–	–	–	22 ± 5.7	0.22 ± 0.073	0 ± 0
E_T^{miss} & τ	–	–	–	–	32 ± 7.9	$(15 \pm 6.4) \times 10^{-6}$
B-physics	–	–	–	–	–	9.5 ± 5.5

Figure 2.11: Physics streams in Hz between the data streams at a luminosity of $10^{33}\text{cm}^2\text{s}^{-1}$. Diagonal elements in the table show overlap between streams [17].

Chapter 3

Analytical and statistical methods

This chapter covers the analytical and statistical methods used by ATLAS. In section 3.1 we cover the reconstruction of various physics objects used in both Part II and III. Section 3.2 describes the production methods for the Monte Carlo (MC) samples, which are used with the reconstructed data from section 3.1. Section 3.3 reviews the statistical methods used to evaluate data and MC. Methods such as the likelihood fit are used throughout this paper and are defined here. The Hypothesis testing is defined here to give context to the results of Part II.

3.1 Reconstructing data events at ATLAS

We have previously discussed in section 2.2 how the ATLAS detector and its subsystems work. This section will discuss how detector hits and readouts are translated into usable physics data. There are many physics objects that need to be reconstructed such as jets and their flavors, electrons, muons, tracks, missing energy and photons. Pile-up will be discussed in this section, as various objects need to be able to separate their signal from pile-up to correctly reconstruct the object. Pile-up is one type of background produced from two main sources: additional proton-proton interactions at the same time as the primary event being reconstructed, or from the residual impacts from the previous bunch crossing being reconstructed. Section 8.1 goes through pile-up and its suppression in more detail.

Figures 3.1 and 3.2 show a demonstration of how both a $t\bar{t}$ and a Higgs event at ATLAS will appear during reconstruction. The leptonic decays shown in both these figures shows how the ATLAS detector reconstructs both electron and muon events which will be further covered in sections 3.1.5 and 3.1.6. The $t\bar{t}$ decay in figure 3.2 also has examples of jet decays with two b-jets, these jet and jet flavors are discussed in sections 3.1.3 and 3.1.4.

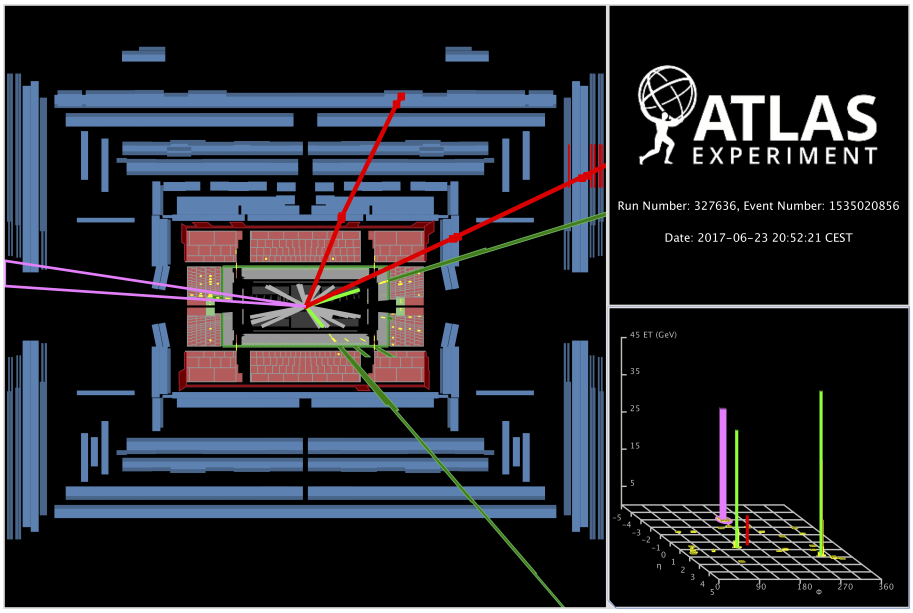


Figure 3.1: Display of a candidate Higgs boson event from proton-proton collisions recorded by ATLAS with LHC stable beams at a collision energy of 13 TeV. The candidate event is reconstructed in the $2e2\mu$ final state. In the left display, the red lines show the path of the two muons including the hits in the muon spectrometer, the green lines show the paths of the two electrons together with the energy deposit in the electromagnetic calorimeter, and the violet cone is a jet. As shown in the lego plot on the right [23].

3.1.1 Charged particle tracking

When a charged particle makes its way through a sub-detector with a magnetic field applied to it, it leaves a trail which can be reconstructed into a track. The change in the particle's direction as it moves through the ID can tell us the charge sign and mass(if using TRT) of the particle with a resolution defined in equation 2.3. These charged particle tracks are reconstructed in the inner detector using all three of the subsystems: the pixel, SCT and TRT.

5 quantities are used to define tracks:

- the p_T of the track
- its position in η
- its position in ϕ
- d_0, which is the minimum distance between the track and the primary vertex (discussed later in section 3.1.2) in the $\mathrm{R} - \phi$ plane.

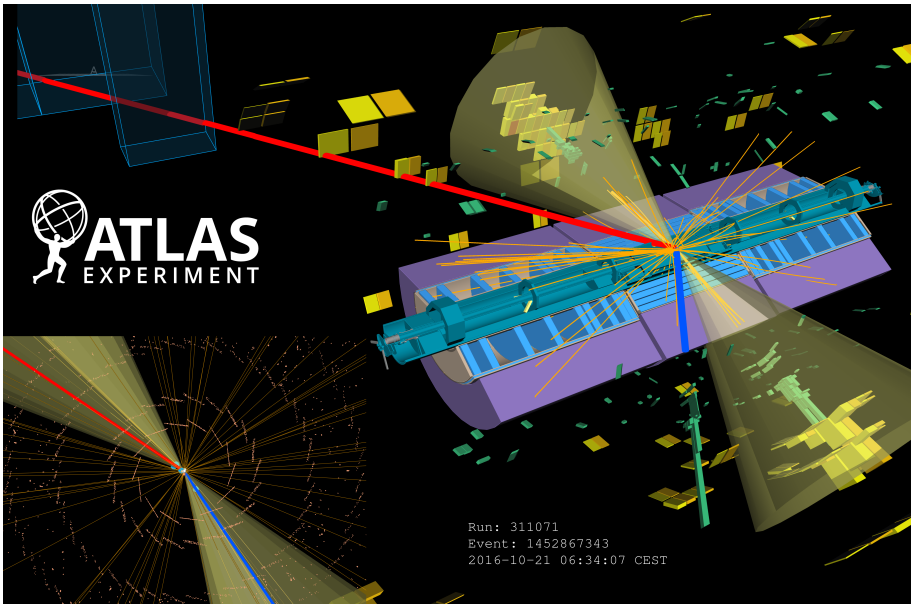


Figure 3.2: Display of a candidate top-antitop event from proton-proton collisions recorded by ATLAS with LHC stable beams at a collision energy of 13 TeV. The candidate event is reconstructed in the electron, muon and 2 b-tag jets final state. The path of the muon is indicated by the red line, the electron by the blue line. The energy deposits in the electromagnetic calorimeter are shown in green, the deposits in the hadronic calorimeter in yellow. The jets are represented by the two yellow cones. The bottom left insert presents a zoom on the proton-proton interaction region. The golden dots represent hits in the four layers of the pixel detector [23].

- $Z_0 \sin\theta$ is the distance of the track from the primary vertex measured in the longitudinal plane.

Pile-up and noise can also produce tracks, which are reduced by dedicated algorithms. After these pileup tracks have been removed, tracks are then required to have both $|\eta| < 2.5$ and $p_T > 400$ MeV [24]. Tracks are reconstructed with a higher efficiency $\approx 90\%$ towards lower η, however, this efficiency drops to $\approx 60\%$ around $|\eta| = 2.5$, and there is no tracking coverage for $|\eta| > 2.5$. The tracking efficiency across η is shown in figure 3.3 [24].

3.1.2 Vertices

A vertex in ATLAS is a point where 2 or more tracks meet. The primary vertex is defined as the vertex with the largest sum of the transverse momentum of tracks associated with it. This vertex originates from the main hard scatter event in the initial proton-proton collisions in the IP. The secondary vertex, which is a vertex which does not originate from the IP and instead comes from the decay of long-lived particles $(10^{-12} - 10^{-10}\ s)$ that did originate from the IP[24].

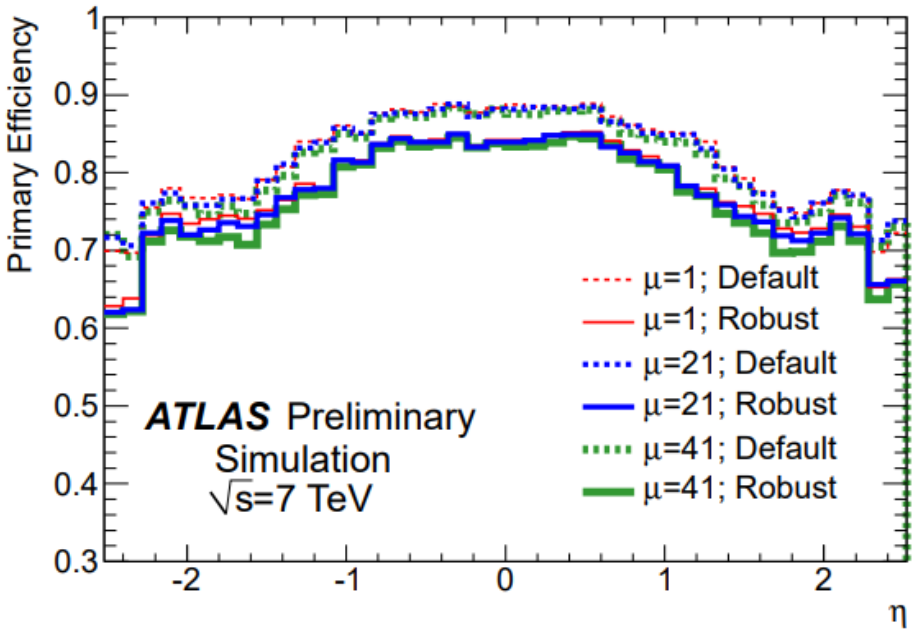


Figure 3.3: Track reconstruction efficiency of primary particles as a function of η [24]

Vertex reconstruction efficiency is determined by taking the ratio between events with reconstructed vertices and events with at least two reconstructed tracks. Beam-induced background events, such as the beam interacting with residual gas molecules in the beam pipe, are also removed[25]. The efficiency of the vertex reconstruction is $\approx 85\%$ when there are 2 tracks in the vertex reconstruction. This number increases to $\approx 97\%$ for 3 tracks, $\approx 99\%$ for 4 tracks and near 100% efficiency for 5 or more tracks [26].

In the event of multiple proton-proton collisions per bunch crossing, the number of primary vertices can be used to measure the in-time pileup. This variable is known as $\mathrm{N_{pv}}$ and will be discussed in more detail in part III.

3.1.3 Jets

Jets occur in the ATLAS calorimeter due to the color confinement of QCD discussed in section 1.1.1. These jets spray out in a cone from the IP into the detector, hitting multiple sub-detectors. Various algorithms have been developed to reconstruct jets. The main algorithm used is the anti-$\mathrm{k_T}$, a sequential recombination algorithm that works as follows [27]: First, identify a proto-jet. These proto-jets can be tracks or clusters of calorimeter cells (as will be discussed further in section 8.1.). The anti-$\mathrm{k_T}$ algorithm uses the distance between two proto-jets i and j to cluster proto-jets together. This distance is calculated as [27]

$$d_{ij} = \min(p_{Ti}^{-2}, p_{Tj}^{-2}) \frac{\Delta R_{ij}^2}{R^2}, \tag{3.1}$$

where $\Delta \mathrm{R}_{\mathrm{ij}}^2$ is the radial distance in the $\phi - \eta$ plane between the two proto-jets. The power of -2 in the momenta comes from the type of jet clustering algorithm. The $\mathrm{k_T}$ algorithm uses a positive power in the momenta modifier while the anti-$\mathrm{k_T}$ uses a negative power in the momenta, which is seen in equation 3.1. The absolute beam distance between proto-jet i and the beam is also calculated using the anti-$\mathrm{k_T}$ and the factor $\mathrm{d_{iB}} = \mathrm{p_T^{-2}}$. If $\mathrm{d_{iB}}$ is smaller than $\mathrm{d_{ij}}$ for any of the possible combinations of proto-jet i with any other proto-jet j, then proto-jet i is considered a jet. Otherwise, the proto-jet is merged with the other proto-jet j that gave the minimum $\mathrm{d_{ij}}$. This process is repeated until no proto-jets are left. These jets then collect softer jet contributions in a R=0.4 cone radius around them.

These jets can be affected by large amounts of background pile-up events in the detector. To reduce this impact, a multi-variant algorithm known as the jet vertex tagger (JVT) is used. This JVT uses a likelihood function with inputs from the constituents of candidate jets [28]. The jets are also calibrated using a Jet Energy Scale (JES) corrections to account for the miss-measurement of the reconstructed calorimeter energy [29]. Finally, the Jet Energy Resolution (JER) is then calculated from simulations using the jet response function [30]

$$R_E = \frac{E_{\mathrm{reco}}}{E_{\mathrm{truth}}}, \tag{3.2}$$

where E_{reco} is the reconstructed energy of a calorimeter jet, $\mathrm{E_{truth}}$ is the energy of a matched truth jet. The JER is calculated using the standard deviation of the Gaussian fitted to this response $\mathrm{R_E}$. This JER is used to measure how responsive the detector is to jets[30].

A jet area is calculated using ghost particles, which are injected into the algorithm as infinitely soft particles[1] so they uniformly cover the (η, ϕ) plane. Each ghost is massless and covers a fixed area $\mathrm{A_g}$. To give a jet an area, the algorithm defines a region where the ghost particles are clustered with a jet[31]. The sum of the area of the ghosts in this region is the area of the jet. This jet area is used in part III to calculate the transverse momentum density for pile-up jets.

3.1.4 Jet Flavour

Jets also come in different flavours. These flavours represent the types of quarks the jets originate from, which are separated into: The light-flavour quarks like (u, d, s) and heavy-

[1]Particles of very low momentum $\mathrm{p_T} \approx 10^{-100}$ GeV so that they do not deposit detestably significant amounts of energy into the detector.

flavour quarks (c, b). The top quark is not part of these jet calibrations as it decays before it can form jets (which is discussed in section 5.1). The ability to reconstruct which quark caused the jet is vital to reconstructing events at ATLAS. The analyses in part II only use b-jets and therefore that will be the focus of this section.

b-hadrons are particularly useful as they have long lifetimes and consequently produce a secondary vertex alongside the primary vertex. This can be shown in figure 3.4.

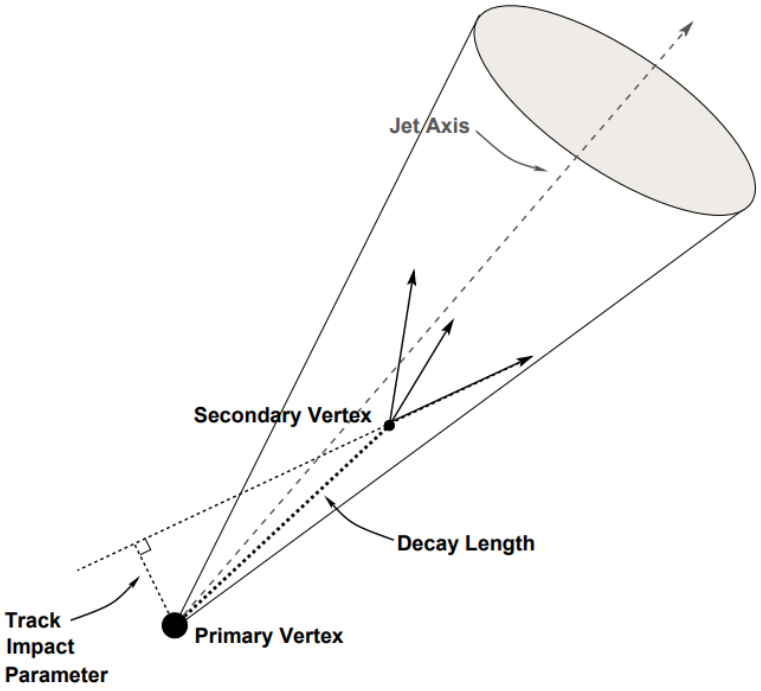


Figure 3.4: A diagram showing a secondary vertex produced by long-lived particles in the jet [32].

This secondary vertex is used in b-jet identification as the b-hadron travels an average of 8-9mm inside the ID, leaving a track that can be matched to this secondary vertex. This makes b-jets easier to identify compared to other flavours of jets.

The b-jets topology, secondary vertex reconstruction, and vertex impacts are combined using a multivariate algorithm called MV2c10[33]. MV2c10 is a Boosted Decision Tree (BDT), which separates b-jets from c-jets and light-quark jets.[33].

The BDT has scores for each event between 0 and 100% based on how likely it thinks an event is to be a b-jet. Working points in this context refer to where the cut on the BDT score is placed. Higher cuts ensure that the events kept are more likely to be b-tagged, but

reduce statistics. The analyses in part II use 77% working points for a fair balance between accuracy and statistics.

Jets may also form from gluons. These jets tend to have a larger hadron multiplicity with the jets themselves being larger in width and with softer contributions. ATLAS reconstruction measure the parameters with machine learning techniques to distinguish these events from quark jets. Being able to identify these events helps analyzes reduce the QCD background.

3.1.5 Electrons

Electrons are detected by ATLAS using tracks in the ID (section 2.2.3) and hits in the ECAL (section 2.2.4). In the ECAL, electrons are reconstructed using a sliding-window method. A similar sliding window algorithm is also discussed in more detail in section 8.4. The sliding-window algorithm used for the electrons works like so: Clusters of cells in the ECAL are selected as a seed if they pass the selection $\mathrm{p_T} > 2.5\ \mathrm{GeV}$. Tracks are seeded as coming from an electron if the track has $\mathrm{p_T} > 1\ \mathrm{GeV}$ within $|\eta| < 2.47$. If the track is found outside this acceptance, then the electrons are classified as forward electrons, which have a different classification system. The cluster seeds are then matched with the central track seeds. The sliding window finds clusters of 3×5 cells with each cell spanning an area $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$. The matched tracks to the cluster are then used to correct the (η, ϕ) coordinates of the reconstructed electron. The cluster seed is also corrected for the type of calorimeter used by re-clustering to the endcap 5×5 and barrel 3×7. The charge of the electron can be reconstructed from the curvature of the associated track, while the energy of the electron can be calculated from the cluster deposits [34].

Reconstructed electron signals can be swamped with fake electrons that can occur from jet and electrons produced from photon conversions that happen in the detector. These are all combined into one background group and need to be separated from real electrons. Electron identification is achieved through a likelihood-based identification.

Electrons also need to be isolated from the decay of heavy quarks, hadrons misidentified as leptons and photons and photons down converting into electron positron pairs. Electrons have two main types of isolation, track-based and calorimeter-based. Both methods use a variable

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}. \tag{3.3}$$

The track based isolation uses this ΔR and sums all tracks p_T within the cone other than the electron track itself. The isolation variable is then the ratio of this summed p_T compared to the electron's own p_T. The calorimeter based isolation take in the E_T of clus-

ter signals in a cone of ΔR around the identified electron cluster summing the energies within the cone. A correction factor is applied to the electrons energy representing the amount of leakage the electron shower has into the surrounding clusters. The isolation process also has working points similar to identification.

The analysis in this thesis in part II uses loose lepton working points, which corresponds to a 85% identification efficiency [34].

3.1.6 Muons

Muons are reconstructed using signals from both the muon spectrometer (MS) and the inner detector(ID), which were discussed in sections 2.2.5 and 2.2.3, respectively. The muons undergo a similar process to the electrons discussed in the previous section 3.1.5 in the ID. The reconstruction starts in the muon chambers. The reconstruction algorithm searches for hits in the muon chambers and fits them to tracks. These tracks are then aligned with charged tracks seen in the ID. Another track in the calorimeter is also used in Muon identification and are known as CT tracks. The muons are then subjected to Identification and isolation [35].

Identification : To identify muon signals, a few physics parameters are considered: The χ^2 produced from fitting the muon tracks between the MS and ID. The absolute difference in the ratio of measured charge to the measured momentum between the MS and ID muon tracks. The absolute $\mathrm{p_T}$ difference between the CT track and the MS muon track divided by the $\mathrm{p_T}$ of the combined (ID and MS) muon track [35].

Isolation: Isolating muons from other muon-like events is based on two variables. The first variable $(\mathrm{p_T^{varcone30}})$ uses the track information. $\mathrm{p_T^{varcone30}}$ is the scalar sum of the $\mathrm{p_T}$ of the tracks with $\mathrm{p_T} > 1\ \mathrm{GeV}$ that are within a cone of size $\Delta \mathrm{R} = \min(10\ \mathrm{Gev}/\mathrm{p_T}, 0.3)$ [35]. The second variable $(\mathrm{E_T^{topocone20}})$ uses the calorimeter information. $\mathrm{E_T^{topocone20}}$ is the sum of the $\mathrm{E_T}$ of each calorimeter energy cluster in a cone size of $\Delta R = 0.2$. These calorimeter energy clusters have already been corrected for pile-up effects and had the muon energy subtracted from them [35].

Both $\mathrm{p_T^{varcone30}}$ and $\mathrm{E_T^{topocone20}}$ are used in defining the muon isolation. Different working points have different selections for both of these variables [35]. There are three main working points for muons, which are called ”loose”, ”medium” and ”tight”. Loose working points represent the highest efficiency of muon reconstruction, but are likely to be swamped with fake muons, while tight working points have very pure muons, meaning it is very likely that the reconstructed muon is a real muon; however, this comes at a cost

to statistics. The medium working point is a middle-ground between the other two. The analysis discussed in this paper uses loose muon working points as the analysis struggled for statistics.

3.1.7 Photons

The analyses in this paper do not use photons and, therefore, only cover their reconstruction briefly. Photons can be reconstructed in the ECAL in a similar way to electrons in section 3.1.5. A cluster in the ECAL can be attributed to a photon if the matched track originates from the secondary vertex. This is deemed to come from an electron that was created by photon conversion, which are reconstructed and differentiated from regular electrons to reconstruct a photon [36].

A major background to photon reconstruction is non-prompt photons from hadron decays in jets. Similarly to leptons, photons have different working points. Photons can have both a loose and a tight working point. Tight photons are used for physics analyses while loose photons are used to trigger events [36].

3.1.8 Missing Transverse Energy

Momentum in the transverse plane p_T should be conserved in the ATLAS detector, therefore, the vector sum of all detected momenta from the same event should be zero. If there are any imbalances, this is known as missing transverse momentum $\overrightarrow{p}^{\mathrm{miss}}_{\mathrm{T}}$ and, by extension, missing transverse energy $\mathrm{E}^{\mathrm{miss}}_{\mathrm{T}}$.

The missing energy is reconstructed using both hard scatter objects and soft signals. This is because the hard scatter events are expected to be produced by the proton-proton collision and, therefore, have net zero momentum in the transverse plane. These hard objects consist of all the objects reconstructed in the previous sections, such as electrons, muons, jets, photons and taus. The soft contributions are tracks that are not included in the definition of other objects and are also not considered pile-up. Equation 3.4 shows the missing momentum defined as the vectoral sum of all soft signal and hard objects in the transverse plane. [37]

$$\overrightarrow{p}^{miss}_{T} = -\sum_{i\in\{hard\}} \overrightarrow{p}_{Ti} - \sum_{j\in\{soft\}} \overrightarrow{p}_{Tj} \tag{3.4}$$

The $\overrightarrow{p}^{\mathrm{miss}}_{\mathrm{T}}$ is converted into energy using equation 3.5 as it is assumed that the particles responsible for the missing momentum are massless.

$$E_T^{miss} = |\overrightarrow{p}_T^{miss}| \tag{3.5}$$

This missing energy could be coming from neutrinos, which are too weakly interacting to be detected by ATLAS. There are also prospects for this missing energy to be beyond standard model weakly interacting particles, such as Dark Matter particles. The $\mathrm{E_T^{miss}}$ is, therefore, very important to searches for new physics. This assumes the ATLAS detector is 100% efficient at detecting more strongly interacting particles. In reality, some interacting particles escape the acceptance of the detector or are inaccurately reconstructed or fail completely in being reconstructed. This gives fake $\mathrm{E_T^{miss}}$, which is discussed further in section 6.3.1.

3.2 Event Simulation

Monte Carlo (MC) Simulations are the backbone of ATLAS analysis techniques, allowing to compare data with MC Simulations, which are based on the principle that the behaviour of a statistic in random samples can be evaluated empirically from creating random arrays of samples and comparing this to known data [38]. The procedure for this has 5 steps shown below[38]:

1. Specify rules for the generation of a pseudo-population.
2. Take a sample from the pseudo-population with the sampling conditions being reflective of the statistical situation of interest.
3. Calculate a parameter θ in the sample and store it in a vector $\hat{\theta}$.
4. Repeat steps 2 and 3 T times. Where T is the number of trials.
5. Create a frequency distribution using $\hat{\theta}$ with T entries.

Now we have the basic outline of how to create MC simulations, we want to apply this to ATLAS. The production of these MC events is divided into 3 main steps. The first step simulates the generation of the particles in the event produced at the IP. It also covers these particles' immediate decay. The second step is to simulate how the particles interact with the detector and the physics interactions that take place as these particles propagate. The final step is to simulate the detector digitization of signal samples, which corresponds to the data readout in the ATLAS detector [39]. The simulated events are then processed using the same reconstruction algorithm described in section 3.1. These simulation steps can be slow; therefore, a fast method is also introduced.

3.2.1 Generation and decay

Event generation begins with the production of a set of particles, which are passed to a full or fast simulation. The generators compute matrix elements (ME), the calculation of SM interactions using Feyman diagrams, at varying orders of precision. The level of precision for these ME's corresponds to the number of loop corrections made to whichever process is generated in terms of the QCD coupling constant α_{s}. A few commonly used levels of precision are: The leading order (LO) loop corrections, also known as tree level, the next-to-leading order (NLO) loop corrections and next-to-next-to-leading order (NNLO) loop corrections.

There are many types of generators, general-purpose generators are used to produce complete events starting from proton-proton, proton-nucleus and nucleus–nucleus initial states. These generators can be used by themselves or in conjunction with more specialized generators. The general purpose generators, which produce events at leading order, used in ATLAS analyses are:

PYTHIA and **PythiaB**, which simulate events starting with a hard scatter processes [40]. The decay is modeled using both QCD and QED. The generator is very accurate when simulating radiation emitted at low angles. Its accuracy decreases at larger angles.

HERWIG which is used for complete event generation and is most used with ISAJET for modeling of beyond standard model supersymmetric processes [41] [42].

PYTHIA and HERWIG produce most of the events we use at ATLAS at LO. For higher precision, we often use NLO full event generators such as:

Sherpa which matches QCD matrix elements to QCD showers[43]. It also interfaces with the PYTHIA hadronization model. The generator processes complete events with better approximation than PYTHIA or HERWIG for final states with large numbers of isolated jets.

Specialized generators, unlike general-purpose generators, do not produce complete events and are instead attached onto other generators like PYTHIA or HERWIG. This can allow for extra precision for specialized measurements. Some examples are:

ISAJET, which is used to generate consistent sets of decay modes and associated masses for supersymmetric events [42]. This generator is not used in full-scale production but is instead used in conjunction with HERWIG for the generation of super-symmetric events.

EvtGen provides more of an in-depth description of B mesons and other hadron decays than PYTHIA or HERWIG [44].

MC@NLO simulates hard scatter processes at next-to-leading order in QCD perturbation theory [45]. MC@NLO is generally interfaced to other MC generators such as PYTHIA to produce these extra loop corrections in PYTHIA which otherwise can only produce events at LO. One particularly good use for MC@NLO is its ability to generate top quarks with more accurate distributions of $\mathrm{p_T}$ than PYTHIA or HERWIG alone.

Single particle generators are used for calibrating the ATLAS detector as well as testing reconstruction efficiencies. This is achieved through simulating the production of an event from a single primary particle. This is not physically possible, but it allows for testing as it doesn't have to simulate underlying events or proton remnants. The primary particle can be selected to be several different particles, like a muon or an electron. The particle's energy and direction can also be set to be in a specified range. An example of a use of the single particle generators is to estimate the cosmic ray background.

The HepMC event record is used to store all the status codes and history provided by the individual generators [46]. Only stable particles are selected to propagate through the simulation, and each particle is labelled with one of two status codes: 1 (stable) or 2 (unstable). The HEPMC information is used to model the underlying physics before any detector-related smearing effects have been added. The truth stores the entire history of all particles generated, whether they propagated through ATLAS or not. This truth information is useful in developing various physics analyses. Truth information is typically quite bulky and takes up around 30$\%$ of all the disk storage space for simulated events.

3.2.2 Simulating particle interactions with ATLAS

Each particle that is produced at the generation stage and is tagged as stable is then propagated through the detector using GEANT4 [47]. GEANT4 provides the models and infrastructure for particle travel through a geometry and the physics interactions with it. GEANT4 is interfaced with Athena, which handles the scoring (or hit) of each particle in the detector subsystems [48]. The score recording can be different for each subsystem. A calorimeter hit is measured as the energy deposited by the simulated particle into the detector. This scoring also records the time and position of the energy deposit. Each subdetector's scoring is optimized to provide the most realistic detector response to each type of particle.

The truth record created at the generation stage is updated by the information produced with the GEANT4 simulation of the particles interacting with the ATLAS detector. The number of interactions produced at this stage is too much to store, therefore, only the interactions which is the most relevant to physics analyses are saved.

3.2.3 Digitization

After particles have been generated and propagated through the detector, the simulation takes into account how responsive the detector electronics are to interactions with simulated particles. This process is fine-tuned to replicate the same readout for simulated particles from MC events as it would for real particles from data events.

This stage converts the simulated hits containing the energy and physics information deposited by each particle as it propagates through the detector into detector responses. These detector responses are known as digits. In the detector, a digit is produced when the voltage or current on some readout channel rises above some pre-determined value within a time window. The details of each subdetectors' noise, cross talk and sensitivities are modeled using subdetector-specific digitization software. The subdetector digitization packages are coordinated using a top-level PYTHON digitization package, which ensures consistent configuration across the subdetectors in the simulation. The digitization algorithms were tuned to reproduce the detector response recorded from real data.

3.2.4 Fast ATLAS simulation

A fast MC simulation was developed to allow large simulated samples to be produced. The ATLAS GEANT4 simulation deals with quite complex detector geometry and physics interactions, which make processing the "full simulation" slow. Many physics studies require high statistics, which would be impossible without a way to increase the speed of the MC production. Various simulation programs are made to complement GEANT4 simulation:

Fast G4 Simulation aims to increase the slowest part of the full simulation, which is accounted by particle propagation through the calorimeter and simulating electromagnetic particles. The way Fast G4 handles this is to remove low-energy electromagnetic particles from the calorimeter and replace them with pre-simulated showers stored in memory. CPU time is reduced by a factor of three in $\mathrm{t\bar{t}}$ using this technique with minimal physics loss ($\approx 1-5\%$ energy scale difference between Fast and full simulation).

ATLFAST-I is another fast MC algorithm it was designed for physics parameter space studies that require very large statistics but smaller physics detail than full simulation

[49]. The technique uses truth objects that are then smeared with detector resolutions to create physics objects similar to ones produced by the reconstruction. The ATLFAST-I has reduced details of efficiencies and fakes. This increases production speeds by 1000 times with enough details for many analyses [39].

ATLFAST-II is used to simulate events as fast as possible while still being able to run the full simulation. ATLFAST-II can be broken down into two parts: Fast ATLAS Tracking Simulation (Fatras) for the inner detector and muon system simulation and Fast calorimeter Simulation (FastCaloSim) for the calorimeter simulation [50] [51]. ATLFAST-II can also be set to run with just one sub-detector of interest with GEANT4, which speeds up the simulation with a high level of accuracy for that particular sub-detector. A decrease in processing time by a factor of 100 is achieved with Fatras and FastCaloSim [39].

3.2.5 Usable datasets for analyses

Once MC events have been fully simulated or a data event fully reconstructed (section 3.1), it is then stored in the form of an AOD (large Analysis Object Data). These AODs store all the ATLAS data. These AODs are, however, too large to be practical to use. These AODs are instead skimmed down to derived AODs (DAOD). These DAOD are produced from AODs with specific selections, reduction in objects and skimming applied based on what different groups of analysis might need. These DAOD are much faster to process for analysis users than the AODs. The DAODs are tagged based on what type of analysis they are produced for. The analysis in part II uses the tag $\mathrm{DAOD\_PHYS}$, which is used often in general physics analyses. The analysis in part II uses the$\mathrm{DAOD_JETM2}$ derivation, which is often used in jet studies. JETM2 was used as it contained both the tower and cluster objects information.

3.3 Statstical methods

3.3.1 Profile Likelihood fit

Particle physics experiments deal with large amounts of data with large degrees of randomness thrown in by both the nature of the physics and the detector itself. At ATLAS and other particle detectors, we have to make predictions based on this data and therefore have to take into account that the outcome of a study could potentially be from a statistical fluctuation rather than real physics. Consequently, we have to use a method of evaluating the statistics and randomness in the data to qualify the physics result.

To start, we define Bayes Theorem based on two events A and B [52]

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \tag{3.6}$$

where $\mathrm{P(A|B)}$ is the probability that event A occurs on the condition B occurs, $\mathrm{P(A)}$ is the probability of A occurring regardless of any conditions, $\mathrm{P(B)}$ is the probability of B occurring and $\mathrm{P(B|A)}$ that event B occurs on the condition A occurs. This is an example of conditional probability, and we can manipulate these conditions to be anything, including potential randomness from the detector or theory involved in the analysis.

Another useful tool is the probability distribution function. A discrete random variable n can have its' probability modeled for every possible value it can take as

$$P(n) = P(\{n\}), \tag{3.7}$$

where $\mathrm{\{n\}}$ is a discrete set of numbers containing n. This equation is fine for discrete cases; however, most of the variables we use are continuous; therefore, a probability distribution function (PDF) is better to use

$$P([x_1, x_2]) = \int_{x_1}^{x_2} f(x)dx. \tag{3.8}$$

The measurement from an event from a particle physics experiment can be modeled as a set of observed variables

$$\overrightarrow{x} = x_1, x_2 ... x_n. \tag{3.9}$$

The variable distributions are affected by both the theoretical and detector uncertainties. The theoretical and detector uncertainties themselves can be modeled using a set of nuisance variables

$$\overrightarrow{\theta} = \theta_1, \theta_2 ... \theta_m. \tag{3.10}$$

The values of these variables are typically unknown [52]. The variables defined in equation 3.9 and 3.10 can be used with the conditional probability $\mathrm{P(x|\theta)}$ to produce a PDF, which we call a Likelihood function

$$L = f(x_1, x_2 \ldots x_n; \theta_1, \theta_2 \ldots \theta_m). \tag{3.11}$$

Since most particle experiments measure more than one event and since each set of variables and nuisances is associated with one event, we need to amend equation 3.11 into

$$L = \prod_{i=1}^{N} f(x_1^i, x_2^i \dots x_n^i; \theta_1, \theta_2 \dots \theta_m), \tag{3.12}$$

which is a product over all events in a dataset. Equation 3.12 can be exploited to reduce theoretical and detector uncertainties in measured results by finding the maximum likelihood value that can be achieved with the nuisance parameters θ_{i} present in

$$\frac{\partial ln(L)}{\partial \theta_i} = 0, \tag{3.13}$$

and solving for L. This, in most cases, needs to be solved numerically. We now construct an extended likelihood function [52]

$$L = P(n; \theta_1, ..., \theta_m) \prod_{i=1}^{N} f(x_1^i, x_2^i \dots x_n^i; \theta_1, \theta_2 \dots \theta_m), \tag{3.14}$$

where $\mathrm{P(n;}\theta_1, ..., \theta_{\mathrm{m}})$ is the distribution of the various events n and in practice is almost always a Poisson distribution. This distribution is also a function of the unknown parameters 3.10. The equation for $\mathrm{P(n;}\theta_1, ..., \theta_{\mathrm{m}})$ is [52]

$$P(n; \theta_1, ..., \theta_m) = \frac{v(\theta_1, ..., \theta_m)^n e^{-v(\theta_1, ..., \theta_m)}}{n!}. \tag{3.15}$$

The $v(\theta_1, ..., \theta_{\mathrm{m}})$ function for a Poisson distribution of the number of fixed data observations m, which in practice for particle physics is always the case, is equal to the sum of the expected background (b) and signal (s) expected yields. This Possion process, as well as the assumption that each n variable will follow the same pdf, also allows for the extended likelihood function in equation 3.14 to be written as [52]

$$L = \frac{k^n e^{-k}}{n!} \prod_{i=1}^{n} (\frac{s}{k} P_s(x_i; \overrightarrow{\theta}) + \frac{b}{k} P_b(x_i; \overrightarrow{\theta})) \tag{3.16}$$

where k=s+b, $\mathrm{P_s}$ and $\mathrm{P_b}$ are the PDFs for the variables $\overrightarrow{x}$ for the signal and background respectively. Equation 3.16 can be simplified into

$$L = \frac{e^{-k}}{n!} \prod_{i=1}^{n} (sP_s(x_i; \overrightarrow{\theta}) + bP_b(x_i; \overrightarrow{\theta})). \tag{3.17}$$

The extended likelihood function here can provide additional information about θ. using this function will, therefore, improve the statistical uncertainties within the analysis.

3.3.2 Hypothesis testing

Hypothesis testing builds on the techniques discussed in section 3.3.1 to evaluate whether observed data are more compatible with one or more specific theories [52]. It is common practice to have two hypotheses H_0, the null hypothesis and H_1, the alternative Hypothesis [52]. H_0 is typically used for processes assuming background only with no signal or for BSM searches where only the SM process is present. The H_1 conversely would represent signal+background present, or for a BSM search, both the SM particles with the addition of BSM particles. To determine whether a set of data meets either hypothesis, a test statistic is constructed λ. The λ distribution is known under the two hypotheses and can be designed to be large for H_1 being true and low for H_0 being true. λ can be defined as the ratio of the likelihood of the H_1 and H_0 hypotheses under conditions x

$$\lambda = \frac{L(x|H_1)}{L(x|H_0)}. \tag{3.18}$$

This is one possible test statistic; we will use this value in defining other test statistics. This λ can be written as a function of k and defined using equation 3.17 to produce

$$\lambda(k) = \frac{L(\overrightarrow{x}, k, \hat{\hat{\theta}}(k))}{L(\overrightarrow{x}, \hat{k}, \hat{\theta})}. \tag{3.19}$$

In the denominator of equation 3.19 θ and k are fitted to $\hat{\theta}$ and $\hat{\mathrm{k}}$. In the numerator of equation 3.19 k is fixed, with $\hat{\hat{\theta}}(\mathrm{k})$ being the value of θ that is the best fit for the fixed value of k [52].

One test statistic that is used for the discovery of a process is

$$q_0 = -2\ln(\lambda(0)), \tag{3.20}$$

for $\hat{\mathrm{k}} \geq 0$. Equation 3.20 is equal to zero for $\hat{\mathrm{k}} \leq 0$ as we only consider positive signal and background strength as evidence against the background-only hypothesis.

Another test statistic that is often used is

$$q_k = -2\ln(\lambda(k)), \tag{3.21}$$

which tests the upper limit of a Hypothesis. Equation 3.21 only holds for $\hat{\mathrm{k}} \leq \mathrm{k}$ otherwise $\mathrm{q_k}$ is zero. This is because the equation represents upward fluctuations away from the estimated value $\hat{\mathrm{k}}$ that is fitted using the hypothesis being tested. Any value that does not at least exceed this estimate is not counted as a fluctuation away from the hypothesis [52].

Both equations 3.20 and 3.21 represent variables that disagree with the null hypothesis for the larger values of q_0 or q_k. $\mathrm{q_k}$ shows increasing incompatibility with the assumption of $\mathrm{H_0}$ as it increases in value, while $\mathrm{q_0}$ increases proportionally to the evidence against a background-only hypothesis.

Z is known as the significance level and represents how significant the results of an analysis are. To evaluate Z in particle physics, we first have to assume the conditions under which the search will be taking place. Most of the time in physics, when searching for a new particle, we know its mass and therefore have a mass peak to search for. In other cases, such as the Higgs boson, we did not know what the mass is. In this case, we have to search various parameter spaces with the assumption that there are only the background processes present $(k = 0)$. We then compute probability (assuming $\mathrm{H_0}$ to be true) of getting a value of the test static $\mathrm{q_0}$ that is at least as extreme as the observed value of the test static $\mathrm{q^{obs}}$ in the data. This probability can be expressed as

$$p_0 = \int_{q^{obs}}^{+\infty} f(q_0) dq_0. \tag{3.22}$$

This $\mathrm{p_0}$ can then be used to define

$$Z = \Phi^{-1}(1 - p_0), \tag{3.23}$$

which is the significance level (aka the signal significance), and Φ is the cumulative Gaussian distribution function. Z in equation 3.23 represents the number of standard deviations required for the integral of the Φ tail to evaluate $\mathrm{p_0}$. Z can be said to be the significance of rejecting $\mathrm{H_0}$ given that $\mathrm{H_0}$ is assumed to be true, where this probability is measured in standard deviations [52].

This significance level in literature is typically required to be at least 3σ for the measurements that confirm already well-known theories, such as evidence of a process predicted by the SM. In the case where a measured process would contradict or overturn a well-established theory, a higher constraint is typically placed, requiring Z to be at least 5σ [52].

Chapter 4

Neural Networks

In this chapter, we explain how Neural Networks work as they have a large contribution to the work presented in Part II, as in that part, a neural network is used to separate signal events from background events. The basics of a Neural network will be covered in section 4.1. The activation function and loss function are very important to the performance of a neural networks; their methodology and examples of these functions that are used in the analysis in part II will be covered in sections 4.2 and 4.3, respectively. followed by an overview of more specific features of neural networks in section 4.4 that are used in Part II.

4.1 Neural Network basics

A diagram of an artificial neural network in its most basic form is shown in figure 4.1. This figure shows the nodes arranged in 3 layers. The input layer represents the data that needs to be classified, being transformed into a compatible format and entering the network. The input layer has as many nodes as there are features of the dataset. The hidden layer is called as such as it is not observed directly by either the input or output of the network. The hidden layer serves as an intermediate between the input and output in simple neural networks. Figure 4.1 shows an example of the hidden layer; however, in neural networks, they may have multiple hidden layers. The output layer is fed by the hidden layer. The output can take a few forms; for example, a binary classifier will provide a number between 0 and 1 for the output nodes.

Each layer consists of nodes. These nodes take inputs from the previous layer and produce outputs for the next layer. The behavior of each node is determined by a set of weights. The behaviour of each node is shown in figure 4.2 with the weights W (the number multiplied to each variable from the previous layer that is connected to the node) and the variables x (the output of a node from a previous layer that is connected to the node in this layer or a number representing the input data). Each variable from the input

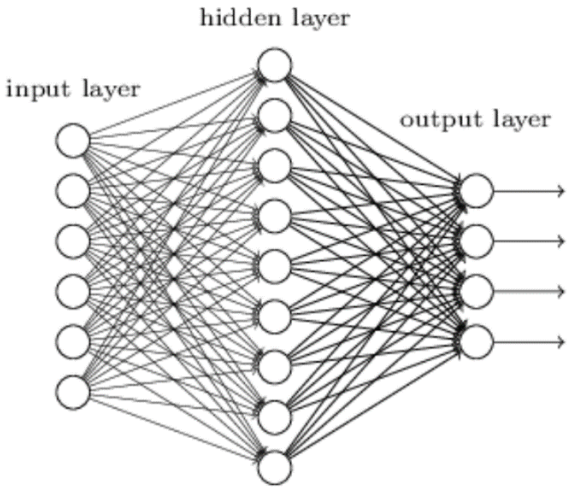


Figure 4.1: Diagram showing a 3-layered fully connected neural network. The layers consist of an input, hidden and output layer. Every node in each layer is fully connected to the nodes in the next layer. The nodes each linearly combine all their inputs from the previous layer, which are influenced by weights between each node.[53]

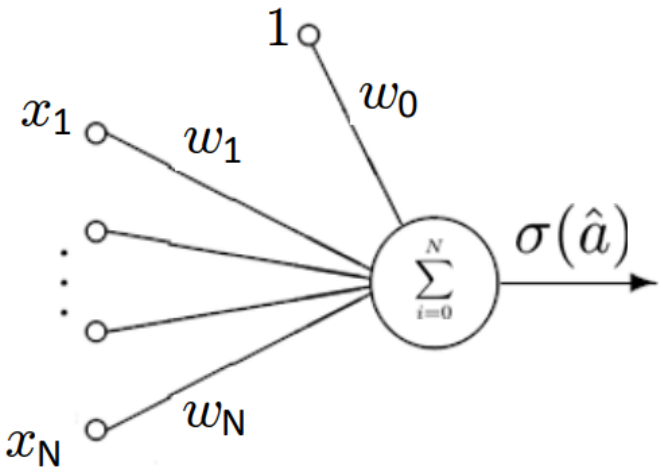


Figure 4.2: A diagram showing the visualisation of a node. The inputs x have a weighted summation before being nonlinearly (sigmoid) transformed at the output [54].

layer is multiplied by its corresponding weight before adding on a bias b; this is shown in equation 4.1[55].

$$a_j^\ell = \sum_{i=0}^{N} w_i^l x_i^{\ell-1} + b^\ell \tag{4.1}$$

Where a_j^l is the output of node j, which is then operated on by the activation function,

and ℓ is the layer of the network for multilayered neural networks. This a_j^l is calculated per node in the layer. Then, this number is put through an activation function, such as a sigmoid transformation, to produce the node's output [54].

4.2 Activation functions

Activation functions significantly improve network performance by introducing nonlinearity into the network output. These activation functions also keep the results bounded and make them more interpretable by the network. There are many activation functions neural networks can use. One main type which is used across a range of neural networks is the fixed shape activation function. Examples of this type of function are the Rectified Linear Unit(ReLU), tanh, and sigmoid. The sigmoid function produces outputs between 0 and 1 and takes in the a_j^l values from equation 4.1. The sigmoid function acts on the output of equation 4.1 as

$$\sigma(a_j^l) = \frac{1}{1 + e^{-a_j^l}}. \tag{4.2}$$

The tanh uses the hyperbolic tan function and takes in a^{l} from 4.1 as the variables to produce an output between -1 and 1. The ReLU function, instead of trying to bind the results in two directions to fixed numbers, binds only one side (the ReLU function can go between 0, which is a fixed number and a^ℓ, which is not fixed). The function is defined as [55]:

$$ReLU = \max(0, a^\ell) \tag{4.3}$$

This unbounded approach by the ReLU function helps to solve the vanishing gradient problem, which is an issue that occurs when the loss gradient (will be discussed in section 4.3), which is used to drive training, becomes very small. ReLU has improved performance over activation functions like sigmold and tanh, mostly for this reason [55]. The analysis in part II uses ReLU activation functions for its hidden layers and a sigmoid activation function for its output layer.

4.3 Training and loss functions

We now have the framework for the neural network and what it consists of, and we want to explore how the network learns from the inputs for the case of a binary classifier (as this is the classifier used in part II). The data is injected into the network at the input layer, which then is put through equation 4.1 with $\ell = 1$ and then has an activation function applied to a_j^ℓ. The data we use in training is tagged, and in the analysis in part II we use signal and background tags for each event determined by the Monte Carlo.

The network sends the MC event information through the hidden layers, with the weights at this stage set to random values. The output layer receives the information from the hidden layer, giving scores between 0 and 1 when using the sigmoid activation function defined in equation 4.2. The network now evaluates how well the weights at their current values have replicated the tag. If an output node is closer to 0, then it is classifying the MC event as being background, and if the tag is closer to 1, then it is classified as signal. The closer the value is to 1 or 0 then the more sure the neural network is in its classification. After this first run, the network is most likely going to give poor results. The network will now use the information from the first run to alter its weights to optimize the output to be as close to the tag as possible. The method most often used to train neural networks is called backpropagation. This method has nodes remembering their output values and pathways to the output nodes. If the output node is correct, more weight is added to the pathway. If the output node is incorrect, then it tries again with new weights and bias.

At the heart of the training process is the loss function. This loss function in classification networks measures the difference between the neural network's prediction and the tag with a value called loss (L). This loss represents how far the network's prediction is from perfect classification.

In neural networks, we measure this loss as a function of its weight. This weight space consists of a combination of all possible weights that the neural network can have and the loss that these weights will produce [54]. This loss function for these weights can take many forms, but whichever form it takes, we use the loss function to reduce the loss of the neural network by altering the weights. The exact weight space of a specific neural network is not mapped out. Instead, the network has a stepping method which steps along the weight space in small steps. Each time the neural network takes a step, it moves from one epoch to another. An epoch is the phase space the neural network is located during training. The neural network starts at epoch 0 and increases incrementally in integer steps until the training finishes. The steps the neural network takes in the weight space need to be taken in directions that reduce the overall loss of the system in as few epochs as possible. Therefore, the loss function's first derivative is calculated at each step, both in the weight space (w) and the bias (b) space calculating $(\frac{\partial \mathrm{L}}{\partial \mathrm{w}}, \frac{\partial \mathrm{L}}{\partial \mathrm{b}})$. This derivative is used to calculate which direction will give the optimal step. The network will keep this cycle of stepping (changing the weights), calculating the loss and the first derivative of the loss function at this new weight for each epoch until it finds a minimum loss. This method is not infallible, and issues such as local minima can affect the network's performance, or the loss function can converge to local minima. The step size can affect this issue; if the step size is too small, the minimization is more likely to get stuck in local minima, but if

the step size is too large, it might not be able to get close enough to the global minimum. Different types of loss functions provide different shapes and, therefore, allow the amount and size of local minima to be reduced. The choice of loss function in a neural network can have a large impact on the training.

The binary cross-entropy function is often used as a loss function in neural networks and is used in the neural network in part II. This function is useful for reducing the learning slowdown that other loss functions face towards the end of the training. We define the cross entropy loss function as [53][35]

$$L = -\frac{1}{n}\sum_{x}[y\ln(\sigma(a)) + (1-y)\ln(1-\sigma(a))], \tag{4.4}$$

where a is defined in equation 4.1, x is the input, y is the desired output and n is the total number of training variables. The equation increases the speed of training at the extremes of the sigmoid function $\sigma(\mathrm{a})$ (defined in equation 4.2). To express this, we show the first derivative of equation 4.4 as [53].

$$\frac{\partial L}{\partial w_j} = \frac{1}{n}\sum_{x} x_j(\sigma(a) - y) \tag{4.5}$$

For a binary classifier, the desired value of y is 1. The sigmoid function flattens as it approaches 1. By having the $(\sigma(\mathrm{a}) - \mathrm{y})$ term setup in this way in equation 4.5 it negates the slowest changing part of the function and instead makes it a function of x. This avoids the slower training seen in other loss functions at the flat end of the σ distribution.

4.4 Dense Neural Network features

Dense neural networks (DNN) are a type of neural network where each neuron in each layer is fully connected to every neuron in the previous layer. We will now cover a range of terms and features used in DNNs that are relevant to the DNN used in part II.

4.4.1 Batch Normalization Layers

A batch normalization layer is a normalization technique used for DNNs and aims to improve the stability and speed of training. These work by normalizing the input of each layer by calculating the mean μ and standard deviation σ of a mini-batch of inputs. This is then used in equation 4.6 to produce a normalized output $\hat{x}$ from the input layer x [56]:

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} \tag{4.6}$$

Where ϵ is a small constant.

4.4.2 Overfitting

Overfitting is an issue in DNNs where the network is trained so much on a training set of data that it starts to learn the noise of a feature. This is why Neural networks have both testing and validation datasets to go alongside the training dataset. The testing and validation datasets typically contain the same features but will have different noise parameters. The validation set is typically tested at each epoch as the network is trained, which can allow for methods like early stopping, which is discussed around figure 4.3. The testing set, however, is usually measured at the end and has no part in the training. If the neural network is fed the validation or testing set and has a large enough sample size to train on it should be able to classify it to the same degree of accuracy as the training set. Deviations between the accuracies of the DNN between the training and testing or validation sets can be a sign of overtraining/overfitting[1]. There are a few methods for reducing overfitting:

- Dropout layers in a DNN are nodes which can be dropped from the training to stop overfitting. This method works by randomly dropping out a small percentage of nodes during training. This helps the DNN generalize features and helps prevent it from becoming too specialized [56].

- Early stopping is a useful tool for DNNs in combating overfitting. Early stopping is a tunable parameter that aims to reduce overfitting in training by monitoring the validation loss over each epoch. An example of the early stopping is shown in figure 4.3.

The early stopping model attempts to stop the training before overtraining/overfitting occurs but while trying to underfit as little as possible. The user can use two parameters which control the early stopping process. The *patience* and *MinDelta*. The *Mindelta* is the minimum reduction in loss (that occurs from the current epoch compared to the previous epoch), before the network starts recording the number of epochs that fall below this threshold. The network will count the number of epochs which consecutively fail to pass the MinDelta until the reduction in loss increases above the MinDelta or the counter reaches the value defined by the patience. If during the training the reduction again falls below the MinDelta value, the counter will start again from zero. Once the network counter reaches the value of the patience, the training is stopped.

[1]Overtraining is where the neural network trains for too many epochs and learns the noise parameter, which is what causes the model to overfit.

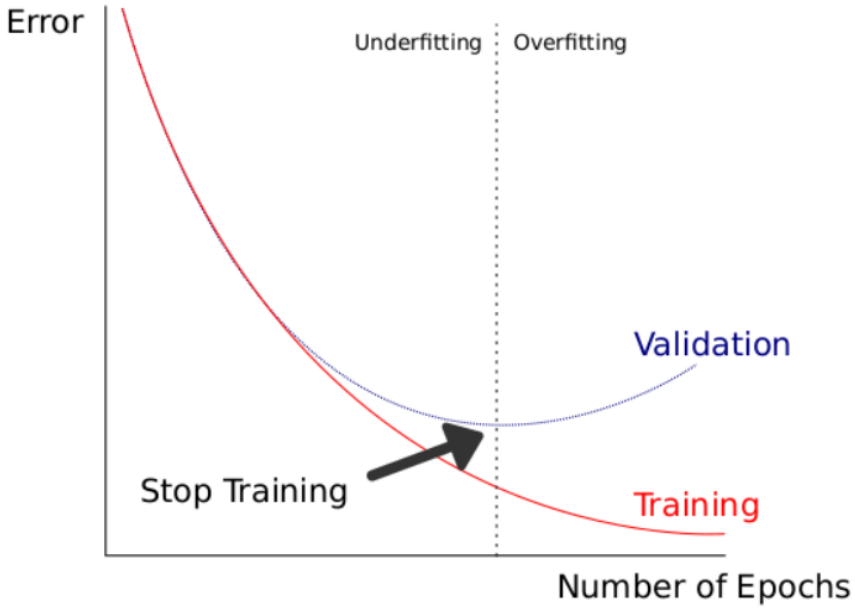


Figure 4.3: A sketch showing the principle of overfitting and underfitting for a DNN. The Validation set's error on a DNN model decreases with the number of epochs up until the model begins overfitting. The validation is the measure of this overtraining. When the validation error starts increasing with epochs, this is classed as overfitting, while before this point is classed as underfitting[56]

4.4.3 K-folding

Another method DNNs use to increase their performance is *K-folding*: This technique is needed as testing the results of a DNN on data that was not included in the training is vital. k-folding solves this issue by separating the data into k folds. Each fold is an equal subset of the data. One fold is used for validation while the others are used for training. Next, there are k models produced, each using a different combination of k folds for training, testing and validation. The models are trained on multiple folds and tested using the fold that corresponds to the k number of the model. For example, in a $k = 4$ situation: Model 1 (of 4) could train using folds 2,3 and 4 as the training set and use fold 1 for test. Model 2 would use fold 2 for testing, and the training would be done on folds 1,3, and 4 etc... Figure 4.4 shows a visualization of a $k = 4$ fold DNN.

4.4.4 Area under the operating curve

When evaluating DNNs, we need ways to evaluate how well the DNN is learning the information it is given so we can rank how well certain variables or sets of data help towards the training of a network. The receiver operating characteristic curve (ROC) was developed as

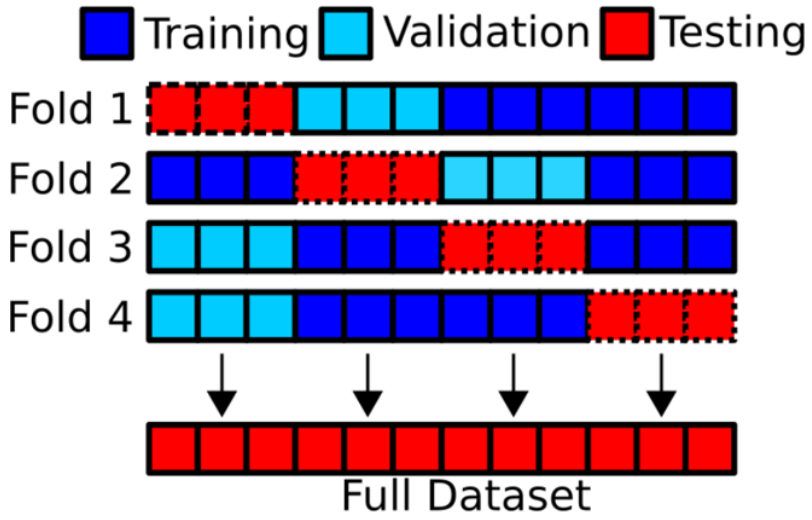


Figure 4.4: A diagram showing a $k = 4$ k-fold model for a DNN[56]

a way of evaluating the performance of a binary diagnostic classification method. The ROC is plotted using two parameters. True positive result and a false positive result, which in effect is the signal and noise. The false positive result and the true positive result are often referred to as the specificity and the sensitivity, respectively. The sensitivity and specificity both range between 0 and 1. Criteria can then be set for what determines a positive and negative result. The stricter the criteria for determining a positive result, the more points on the curve shift downwards and to the left. If the criteria is loser, the more points on the graph will shift up and to the right. An example of this graph is shown in figure 4.5 [57]

This technique is a good way for various studies in machine learning to test the validity of a result. Interpreting this graph, however, can be difficult; therefore, the Area Under the Curve (AUC) is used to quantify this ROC curve. The AUC is higher the more the curve is closer to the top left corner, as a perfect score for the AUC is 1. The minimum score for an AUC would be where the false positive score equals the true positive score. This would correspond to a 45-degree line on the ROC curve, giving an AUC of 0.5. The reason argued for the minimum here is that at this point, the chance of correctly identifying between two binary choices is akin to a coin toss. Any values lower than this would give less accuracy than just randomly guessing between the two. We now have a range between 0.5 and 1 for AUC, which is logical to use. Anything between 0.7 and 0.5 is considered a poor or a failed result. Between 0.7 and 0.8 is considered a fair result. A result above 0.8 is considered a good result, with results above 0.9 considered excellent. In general, the machine learning model with the higher AUC is considered to be the better model with better performance[57].

This AUC can now be used to determine feature importance in DNN studies. We create

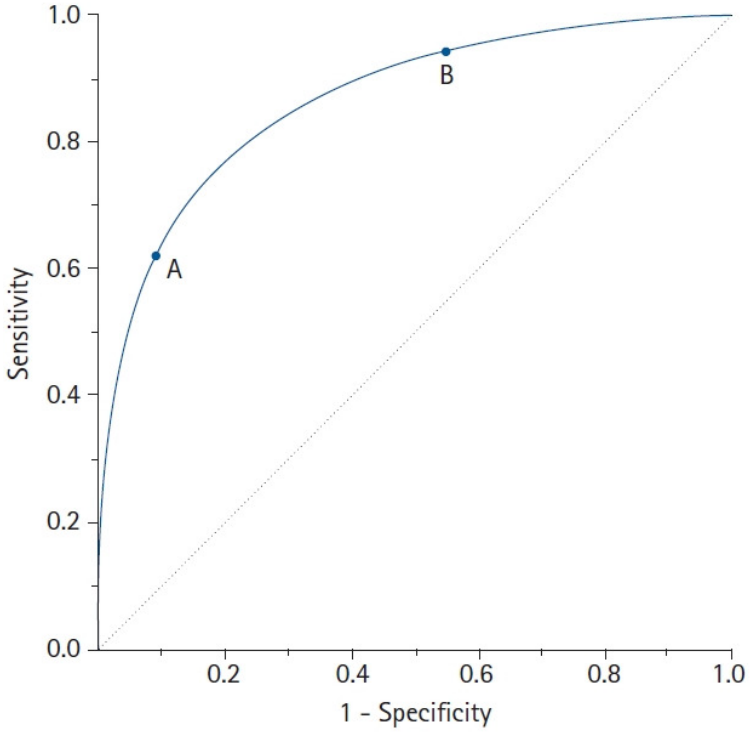


Figure 4.5: A receiver operating characteristic (ROC) curve connects coordinate points with 1 - specificity (false positive rate) as the x-axis and sensitivity(true positive rate) as the y-axis at all cut-off values measured from the test results. When a strict cut-off point (reference) value is applied, the point on the curve moves downward and to the left (Point A). When a loose cut-off point value is applied, the point moves upward and to the right (Point B). The 45° diagonal line serves as the reference line, since it is the ROC curve of random classification.[57]

a variable $\frac{\mathrm{AUC_{nom}}-\mathrm{AUC}}{\mathrm{AUC_{nom}}}$. The larger this variable the more impact the feature has on the training. This variable is used later in figure 6.6, the variable is created as follows:

1. Initial AUC scores are calculated during the training.

2. A feature or variable of the DNN should now be randomly permuted (randomly change the input values of the feature so it no longer is well fitted by the network). A new AUC score is now calculated after the feature is permuted.

3. The difference between the original and new AUC score should now be calculated.

This difference is effectively the importance of the feature. The larger the difference the larger the importance of the feature to the network and the bigger the impact on training. This process is repeated for all variables in a DNN, giving scores and ranking for all

variables in the network.

This technique is good for determining which variables need to be dropped from the training to allow for better overall output of the network. This variable also allows the identification of rogue variables, which shouldn't be contributing as much towards the training of the network but are being given high importance due to some non-physics-based reason. The importance of plots can be misleading if certain variables display collinearity.

4.4.5 Input Scaling

Scaling for a DNN is an important step that normalizes the input data before being fed into the NN. This is important to configure correctly, as scaling helps the NN to converge faster and perform better. There are various types of scaling. Minmax scaling normalizes input features to numbers between 0 and 1. The input feature value has the minimum feature value subtracted from it and then divided by the range (the difference between the minimum and the maximum feature value)

$$X_{minmax-scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}. \tag{4.7}$$

where X is the input variable or feature, $\mathrm{X_{min}}$ and $\mathrm{X_{max}}$ are the minimum and maximum value of the variable respectively [56].

Standard scaling is also used in NN. This type of scaling, often referred to as Z-score normalisation, scales input features so that they have a mean of 0 and a standard deviation of 1. The scaling works by subtracting the mean of the feature from the feature value and then dividing by the standard deviation (σ) of the feature

$$X_{standard-scaled} = \frac{X - \mu}{\sigma}. \tag{4.8}$$

where X is the input variable and μ is the mean of the feature [56]. Standard scaling is better for normally distributed input data, while minmax scaling is more suited to variables within a fixed range.

We can also scale by using a group of objects of a similar type, as its expected that, for example, multiple momentum parameters would be on a similar scale and therefore be better scaled using their value ranges across these parameters. The same would not be true, for example, when comparing the η position (-5 to 5) of a jet to its momentum (MeV-GeV). In part II, we use particle physics variables such as the transverse momentum $(\mathrm{p_T})$; scaling is also used for $\mathrm{p_T}$ variables. $\mathrm{p_T}$ scaling works by taking the $\mathrm{p_T}$ value of the variable

and dividing it by the sum of all $\mathrm{p_T}$ objects of the same type. After this, the logarithm is taken

$$X_{p_T-scaled} = \log(\frac{p_{T,i}}{\sum_j p_{Tj}}). \tag{4.9}$$

where i is the index of the object you are scaling and j is the index of the other objects of the same type [56].

4.4.6 Vector Padding

Often, vector variables do not have a consistent size due to data features. Thus, this can be accounted for by adding padding vectors. These vectors will be filled into the otherwise empty positions of the vectors. This allows for variables that are not used in every event to be used with variables that are used in every event. An example of this is shown in figure 4.6 below [56].

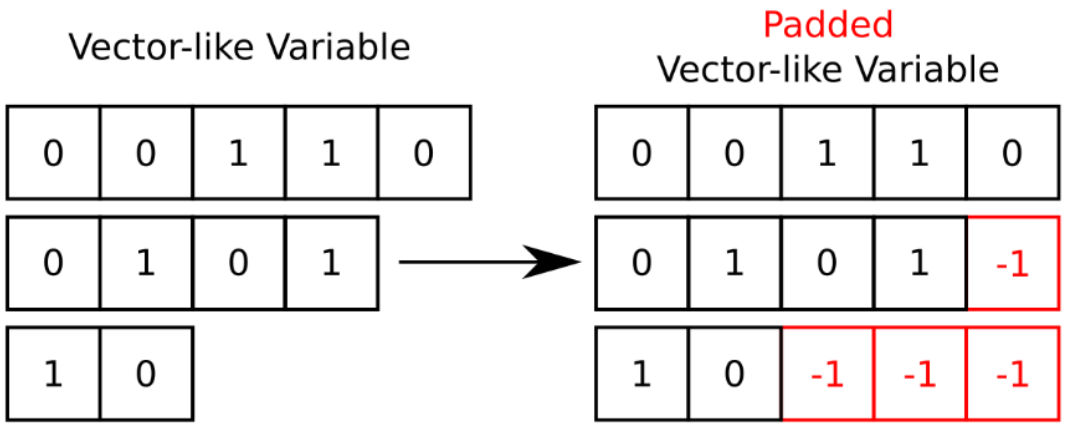


Figure 4.6: Shows an example of a padding vector with value -1 being applied to a variable example [56].

Similar to how the weights of the Monte Carlo need to be applied when using raw events in analysis scripts, the weights also need to be accounted for in the training of a DNN. The loss function is modified to consider the reduced or increased number of events from the weighting. This weighted loss is calculated as the weights multiplied by the corresponding average of the per-sample losses. This system makes the samples with higher weights contribute more towards the training. This average weighted loss

$$L(w) = \frac{1}{N}\sum_{i=1}^{N} w_i \cdot L_i, \tag{4.10}$$

where w_i is the weight of the sample at index i, L_i is the sample loss at index i. N is the total number of samples in the training set [56].

Chapter 5

Introduction to top quark production in association with a Z boson

The process $\mathrm{t\bar{t}Z}$ is covered in this chapter with both its theoretical background and experimental background with special focus on the process $\mathrm{t\bar{t}}(\mathrm{Z} \to \nu\bar{\nu})$ which is the research carried out in this thesis. In section 5.1 we cover the theory of both the top quark and the Z boson and their production together to give more context to the work presented in this thesis. In section 5.2 we cover the previous measurements of $\mathrm{t\bar{t}Z}$ to give the motivation for the work carried out in part II.

5.1 The theory of top-quark pair production with associated production of a Z boson

The top quark is the heaviest known elementary particle, as shown previously in table 1.1 at 172.9 GeV. This high mass makes the top quark very unstable with a decay time of around 10^{-25} s [5]. This makes direct top quark detection very difficult as the time it takes for quarks to hadronise into detectable jets is around 10^{-24} s. The top quark, therefore, can only be studied from its decay products.

The production of $\mathrm{t\bar{t}}$ pairs is the dominant process at ATLAS, with the production mediated primarily by the strong force via gluon-gluon or quark-antiquark introduction as shown in figure 5.1.(a), 5.1.(b) and 5.1.(c).

At $\sqrt{\mathrm{s}} = 13\mathrm{TeV}$ the gluon-fusion for s and t channels in figures 5.1 (a) and (b) receptively, makes up to $\approx 90\%$ (with t-channel dominating) of the $\mathrm{t\bar{t}}$ interactions from ATLAS proton-proton collisions, while the annihilation process in figure 5.1 (c) makes up the remaining 10% [5].

The top quark produced decays into the b-quark and W-boson around 99.8% of the

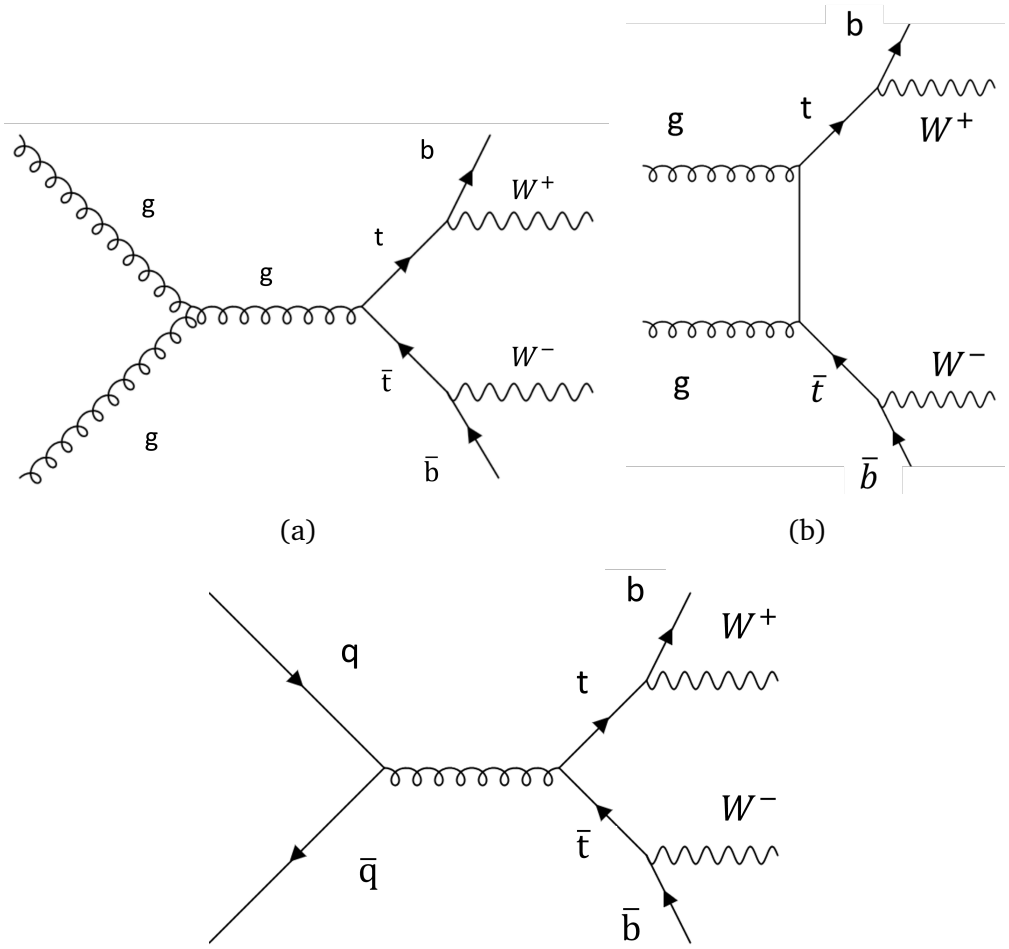


Figure 5.1: Feynman diagram of gluon fusion by s-channel in diagram (a), t-channel in diagram (b) and quark anti-quark annihilation in diagram (c); with all three producing $\mathrm{t\bar{t}}$ with the top and anti-top decaying into a b and anti-b quark and W bosons.

time, as shown in figure 5.1.(b). The W boson goes on to decay by a multitude of channels. Around 67.4% of the time, this is hadronic, while the remaining 32.6% of the time, it is leptonic with roughly equal splitting between the three types of leptons. The final state signature of the $\mathrm{t\bar{t}}$ process is classified as hadronic, semi-leptonic and leptonic channels based on the vertex combinations of the top and the anti-top, each decaying hadronically and leptonically [5].

The Z boson discussed in section 1.1, is the force carrier for the electroweak interaction and couples to all other SM particles except the gluon. The Z boson's mean lifetime is around 10^{-25} s, therefore, we must determine its properties from its decay products. The Z boson decays into hadrons around 70% of the time, while it decays into neutrinos (20%), and charged leptons 10% of the time. The Z boson has an approximately equal chance to

decay into any of the 3 generations of leptons [5].

In $\mathrm{t\bar{t}Z}$, the Z boson is produced in association with the $\mathrm{t\bar{t}}$ as is shown in figure **??**, its depends on the strength of the coupling between the top quark and the Z boson, and measuring its cross-section directly measures the strength of this coupling. The coupling is predicted to be weak, and the production of $\mathrm{t\bar{t}Z}$ is predicted to be 1000 times smaller than the production of $\mathrm{t\bar{t}}$ alone. This can occur in various channels of decay, denoted by the number of charged leptons produced.

The $\mathrm{t\bar{t}Z}$ vertex is predicted by the standard model and is used to produce MC events for $\mathrm{t\bar{t}Z}$. If there is a discrepancy between the SM prediction and the data observed, this would be a sign of new physics. However, previous $\mathrm{t\bar{t}Z}$ measurements confirmed that no deviation occurs, discussed further in section 5.2.

Previous measurements used the decay of Z into e^+, e^- and μ^+, μ^-, which corresponds to 6% of the of total decays, therefore, these measurements had limited statistics in this channel of phase space for the momentum of the Z boson past $p_T^Z > 200\ \mathrm{GeV}$. By using the decay of Z into neutrinos, which corresponds to 20% of total decays, we increase the statistics we have past $p_T^Z > 200$, allowing for a cross section measurement in that phase space. The tail end of this distribution towards higher $\mathrm{p_T^Z}$ is of particular interest to SMEFT as discussed in section 1.2

The final state of $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$ is studied in three different lepton channels and labeled 0L, 1L and 2L based on how many charged leptons (ℓ) there are in the final state, which are shown in figure 5.2. The 3L and 4L final states for $t\bar{t}Z$ mentioned later in 5.2 are shown in figure 5.3, where the Z boson decays into pairs of oppositely charged leptons

The Z boson in the 0L channel will produce all the $\mathrm{E_T^{miss}}$, meaning that the measurement of $\mathrm{E_T^{miss}}$ in the channel can directly be related to the $\mathrm{p_T}$ of the Z boson. This makes the zero lepton channel, in particular, interesting to studies targeting pure $\mathrm{p_T^Z}$.

The cross-section of $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$ was first calculated using new soft gluon corrections at next-to-next-to-leading-logarithm(NNLO), which was then matched to existing next-to-leading order(NLO) to give a (NNLO+NLO) prediction[58]. This calculation is then rescaled using the branching ratio of the Z boson to neutrinos. The cross-section for $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$ is

$$\sigma^{Theory}_{ttZ\to\nu\bar{\nu}} = 0.173^{+0.052}_{-0.052}\ \mathrm{pb}. \tag{5.1}$$

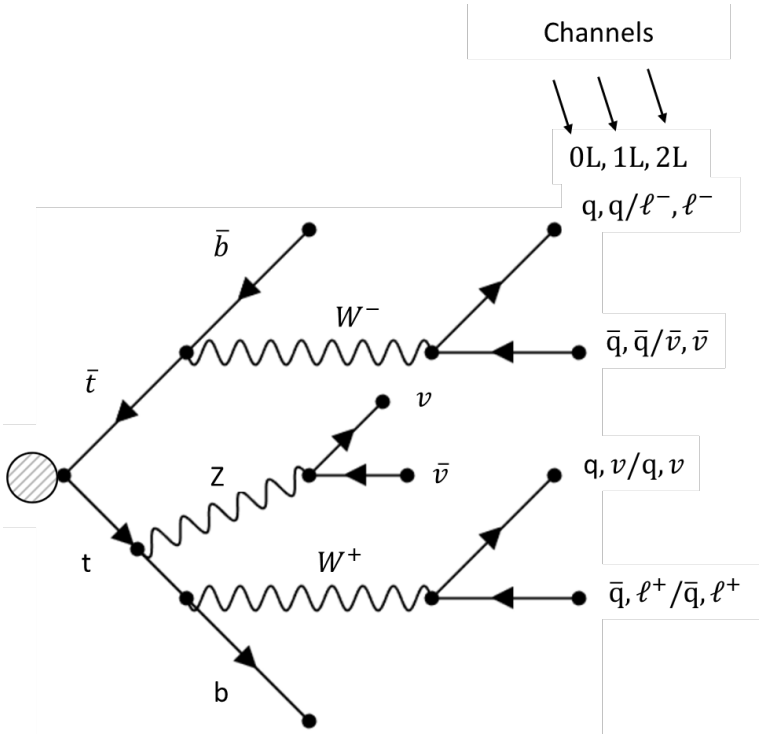


Figure 5.2: The decay of $t\bar{t}Z \rightarrow \nu\bar{\nu}$ into 0,1 and 2 charged lepton final state channels.

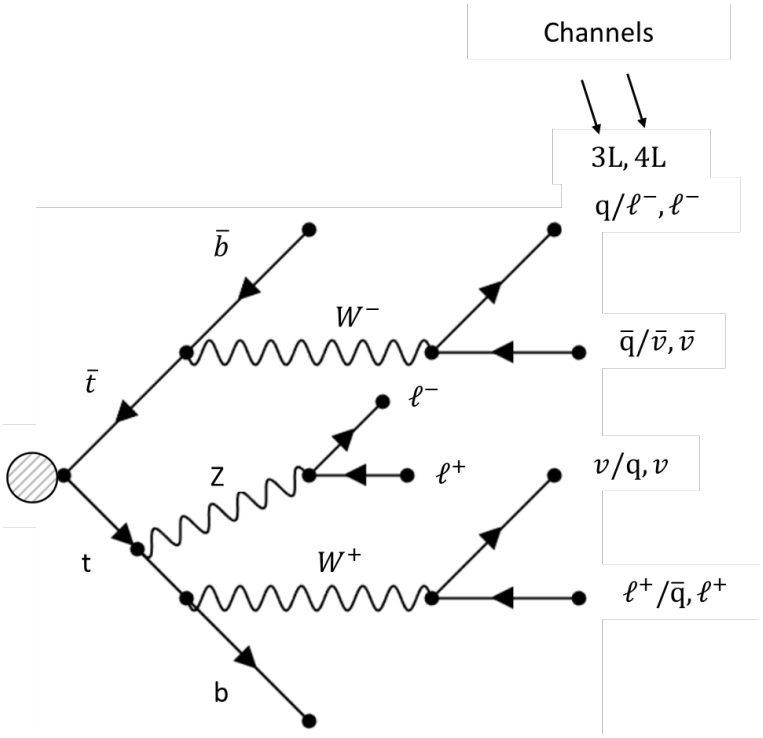


Figure 5.3: The decay of $t\bar{t}Z \rightarrow \nu\bar{\nu}$ into 3 and 4 charged lepton final state channels.

5.2 Previous measurements

A summary of all the cross-section measurements discussed in this section is shown in table 5.1

The cross-section of $\mathrm{t\bar{t}Z}$ was first measured in 2013 by CMS in the trilepton channel[59]. The same-sign dilepton channel was also measured to give a value for the ttV (V=Z,W) and $t\bar{t}Z$ cross section. These measurements were conducted with a luminosity of $5.0\ \mathrm{fb}^{-1}$ at $\sqrt{\mathrm{s}} = 7$. Measuring a cross-section of:

$$\sigma_{t\bar{t}Z} = 0.28^{+0.20}_{-0.14}\ pb \tag{5.2}$$

This measurement was classed as statistically significant as it achieved a significance of $3.3\ \sigma$. The measurement was also consistent with the theoretical value of the cross-section predicted by the standard model:

$$\sigma^{pred}_{t\bar{t}Z} = 0.137^{+0.012}_{-0.016}\ pb \tag{5.3}$$

ATLAS first cross-section measurement of $\mathrm{t\bar{t}Z}$ in 2015 expanded upon the previous measurement made by CMS. This measurement used 2, 3 and 4 lepton channels with both the opposite and same-sign charged leptons used in the 2 lepton region[60]. The paper measures both the cross-section of $\mathrm{t\bar{t}Z}$ and $\mathrm{t\bar{t}W}$. The data corresponds to a integrated luminosity of $20.3\mathrm{pb}^{-1}$ and $\sqrt{\mathrm{s}}$ = 8. This produced a cross section for $\mathrm{t\bar{t}Z}$ of

$$\sigma_{t\bar{t}Z} = 0.176^{+0.058}_{-0.052} pb. \tag{5.4}$$

The NLO QCD theoretical calculation agreed with the measured value

$$\sigma^{pred}_{t\bar{t}Z} = 0.218^{+0.026}_{-0.034}\ pb, \tag{5.5}$$

with had an improved significance of $4.2\ \sigma$ compared to CMS.

CMS in 2016 preformed a new measurement of $\mathrm{t\bar{t}V}$ using a luminosity of $19.5\ \mathrm{fb}^{-1}$ at $\sqrt{\mathrm{s}}$ = 8 TeV [61]. This included a new cross-section measurement of $\mathrm{t\bar{t}Z}$ where the Z boson decays into 2 charged leptons. They used 2, 3 and 4 lepton channels with at least 2 of the same flavour opposite sign leptons coming from the Z boson. The cross-section measured is:

$$\sigma_{t\bar{t}Z} = 0.242^{+0.065}_{-0.055}\ pb \tag{5.6}$$

The measured value has the largest signal significance of any $\mathrm{t\bar{t}Z}$ measurement so far of $6.4\ \sigma$. The value of the cross-section is also consistent with the theoretical prediction given by the standard model:

$$\sigma_{t\bar{t}Z}^{pred} = 0.206^{+0.062}_{-0.052}\ pb \tag{5.7}$$

where the change in predicted cross section between equations 5.7 and 5.5 is that CMS analysis only generated its MC samples for $\mathrm{t\bar{t}Z}$ at tree level, while ATLAS produced it at NLO.

So far, we have covered inclusive cross-section measurements. The first differential cross-section measurement of $\mathrm{t\bar{t}Z}$ was conducted in 2019-2020 by ATLAS alongside an inclusive cross-section measurement across the 3 and 4 lepton channels [62]. The full run 2 dataset was used and a luminosity of $139\ \mathrm{fb}^{-1}$ and $\sqrt{\mathrm{s}}$=13 TeV. The inclusive cross section measured is

$$\sigma_{t\bar{t}Z} = 1.09 \pm 0.10\ pb. \tag{5.8}$$

This cross-section agrees with the theoretical value predicted by the standard model.

$$\sigma_{t\bar{t}Z}^{pred} = 0.863^{+0.12}_{-0.13}\ pb \tag{5.9}$$

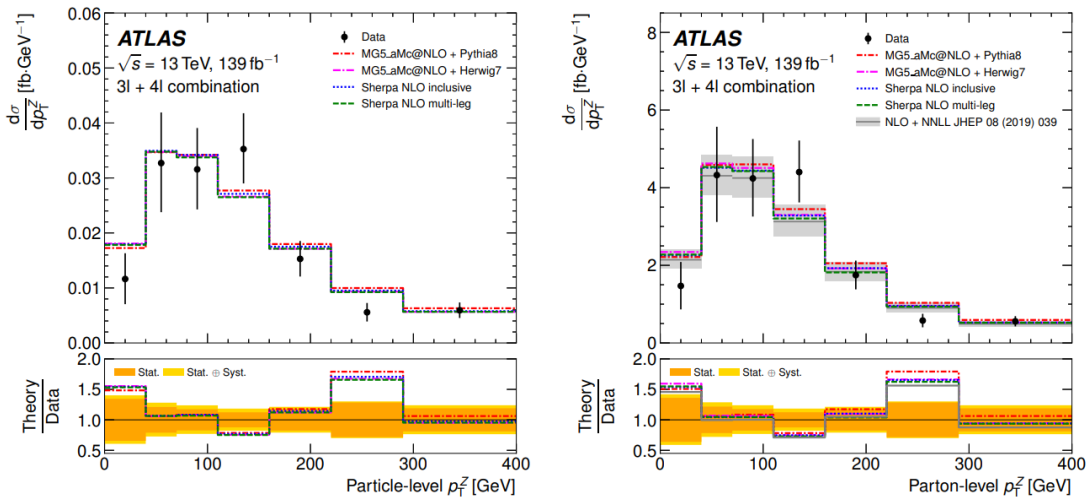
The differential cross-section used the same data and MC setup as the inclusive measurement. The differential measurement was performed at the reconstruction level in the 3L and 4L channels. A combination of both channels was also used. In general, a good level of agreement between the unfolded data and the various predictions is observed. An example plot of the differential cross for the transverse momentum of the reconstructed Z boson is shown in figure 5.4:

A more refined differential cross-section measurement was made in 2023 using the 3L and 4L channels [63]. The full run 2 dataset was used at a luminosity of $140\ \mathrm{fb}^{-1}$ and $\sqrt{\mathrm{s}}$=13 TeV. They, too, conducted an inclusive and differential cross-section measurement. The main improvement of this analysis over previous measurements is the use of Dense Neural Networks to improve the discrimination between signal and background events. The refined measurement of the inclusive cross section gave a value of:

$$\sigma_{t\bar{t}Z} = 0.86 \pm 0.08\ pb \tag{5.10}$$

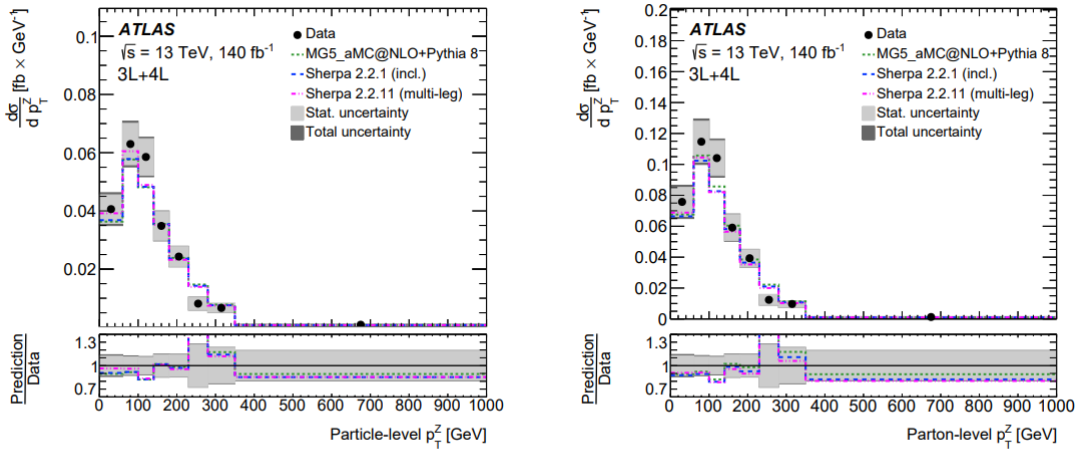
This value is centred nicely on the theoretical value given in equation 5.9.

The differential measurement used various observables that probed the kinematics of the $\mathrm{t\bar{t}Z}$ system. Figure 5.5 shows the differential cross section as a function of the transverse momentum of the reconstructed Z boson:



(a) (b)

Figure 5.4: Absolute differential $t\bar{t}Z$ cross sections measured at (a) particle and (b) parton level as a function of the transverse momentum of the reconstructed Z boson. The large difference in the y-axis scales in (a) and (b) results from different efficiency and acceptance corrections between the particle- and parton-level measurements. [62].



(a) (b)

Figure 5.5: Absolute Differential cross-section measurements as a function of p_T^Z observable in the combined 3ℓ and 4ℓ channels, at particle (a) and parton level (b)[63].

The 2023 refined differential cross-section measurement evaluated the compatibility between the unfolded distributions and the predictions by using calculated χ^2 values produced using the uncertainties of the unfolded distribution and their correlations. The unfolded data showed good agreement with the predictions for most variables[63].

This is where the $t\bar{t}Z \to \nu\bar{\nu}$ analysis presented in this thesis is relevant (part II). This

(Collaboration)	channel	$\sqrt{s}$	luminosity	$\sigma_{t\bar{t}Z}^{measured}$	$\sigma_{t\bar{t}Z}^{predicted}$	$\frac{d\sigma}{dx}$
CMS 2013 [59]	3L	7TeV	$5\ fb^{-1}$	$0.28^{+0.2}_{-0.14}pb$	$0.137^{+0.012}_{-0.016}pb$	No
ATLAS 2015 [60]	2L, 3L and 4L	8TeV	$20.3\ fb^{-1}$	$0.176^{+0.058}_{-0.052}pb$	$0.218^{+0.026}_{-0.034}pb$	No
CMS 2016 [61]	2L, 3L and 4L	8TeV	$19.5\ fb^{-1}$	$0.242^{+0.065}_{-0.055}pb$	$0.206^{+0.062}_{-0.052}pb$	No
ATLAS 2019 [62]	3L and 4L	13TeV	$139\ fb^{-1}$	$1.09 \pm 0.10pb$	$0.863^{+0.12}_{-0.13}pb$	Yes
ATLAS 2023 [63]	3L and 4L	13TeV	$140\ ^{-1}$	$0.86 \pm 0.08 pb$	$0.863^{+0.12}_{-0.13}pb$	Yes

Table 5.1: Summary of $t\bar{t}Z$ cross-section measurements. The $\frac{d\sigma}{dx}$ represents whether the analysis conducted a differential cross-section measurement, where the x represents any possible variable the cross section was measured over.

analysis aims to fill in the gap left by previous analyses. Table 5.1 shows a summary of the previous analyses. We have seen focus on the 2L, 3L and 4L but no measurements in the 0L and 1L. The 2L has not yet been fully covered with measurements of Z to 2 oppositely charged leptons, final state negating Z to neutrinos. This measurement has been avoided in the past due to the difficulty in finding a signal region where the signal isn't washed out by the overwhelming $t\bar{t}$ and Z+jets background.

At the same time, a series of SUSY searches, which contained final states with top-quarks and missing transverse momentum, were conducted in 2018 and found promising signal regions for a 0L, 1L and 2L $t\bar{t}Z \rightarrow \nu\bar{\nu}$ cross section measurement [64][65],[66]. The paper searching for the stop pair production in the 0L channel (stop-0L) paper which was the SUSY search in the 0L channel looking for a SUSY partner for the top quark had a final state that produced 2 top quarks plus missing momentum (the neutralino) as shown in figure 5.6[64]. The signal regions they found for this measurement had a large amount

of $t\bar{t}Z \to \nu\bar{\nu}$ quoted as being a large and ineradicable background. Figure 5.7 shows an example of the signal regions composition. The regions SRA-TT, SRA-TW and SRA-TO show healthy amounts of $t\bar{t}Z \to \nu\bar{\nu}$ with the main backgrounds still being $t\bar{t}$ and Z+jets but significantly reduced. This gave our analysis a starting point where we were able to use variables that discriminated well against $t\bar{t}$ and Z+jets without rejecting significant $t\bar{t}Z \to \nu\bar{\nu}$ events.

The analysis discussed in part II builds upon the stop-0L analysis by using many of the variables it used with a DNN. This DNN provided a low background signal region, which was then used to perform the first inclusive cross-section measurement of the 0L channel for $t\bar{t}Z$, as well as the first ever inclusive cross-section measurement of $t\bar{t}Z \to \nu\bar{\nu}$ across all 3 channels 0L, 1L and 2L.

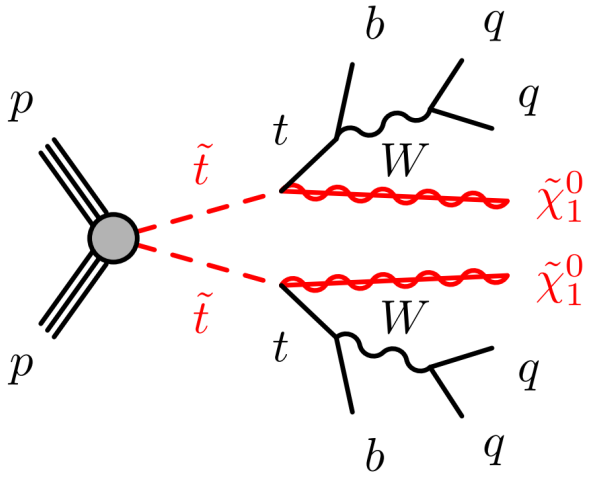


Figure 5.6: Decay topologies of a theoretical two-body decay of $t\bar{t}$ with a neutralino, where the top quark decays hadronically [64].

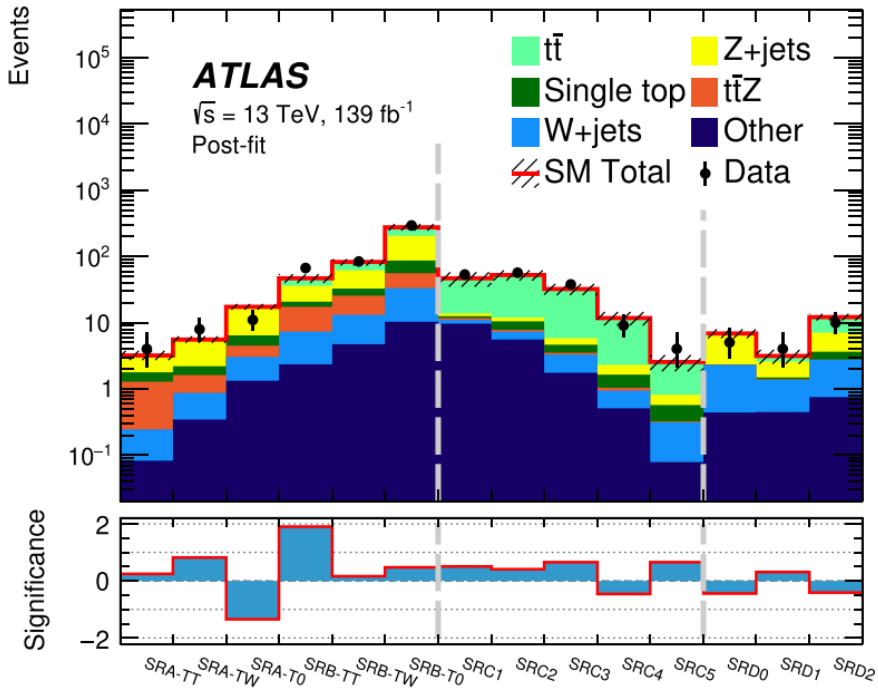


Figure 5.7: Event yields comparing data (points) to the SM prediction (stacked histograms) in all signal regions after the SRA, SRB, SRC and SRD background-only fits. The hatched uncertainty band around the SM prediction shows the total uncertainty, which includes the MC statistical uncertainties, detector-related systematic uncertainties, and theoretical uncertainties. The signal regions included in the SRA–B, SRC and SRD fits are separated by vertical dashed lines [64].

Part II

Top quark pair production in association with a Z boson decaying exclusively into neutrinos

In part 2, we present the analysis of $t\bar{t}Z \to \nu\bar{\nu}$, for both a standalone 0L fit and a combined 0L, 1L and 2L combined fit, presented in two chapters. Chapter 6 reviews the strategies the analysis used, including the samples selected, the systematics used, the objects/variables used, the neural network application and finally the selection of the analysis regions. In chapter 7, the results of this analysis are shown. This includes the fits setups for both the standalone 0L and combined fit, the results of the 0L only fit and the results of the 0L,1L and 2L combined fit.

Chapter 6

Analysis strategy

In this chapter, we cover the analytical setup and methodology for the analyses of $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$, especially covering the 0L channel as this is where the candidate contributed to the analyses. In section 6.2 we overview the samples used for data and MC. In section 6.3 we defined the objects used in the analyses, in section 7.2 we cover the systematics used in this analysis, in section 6.4 we define the pre-selection for the 0L channel while outlining challenges the region faces, in section 6.5 we also explain the choice of training variables and the choice of DNN features used. In section 6.6 we define the various signal regions used for the 0L only fit. Finally, in section 6.7 we overview the combined channel definitions for the combined fit using all three lepton channels.

It is important to note that there are two distinct analyses being presented in this chapter. The standalone 0L analysis being conducted in the 0L channel of $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$. The second analysis being presented in this chapter is a combined 0L, 1L and 2L channel analysis for $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$. Sections 6.2, 6.3, 7.2, 6.4, 6.5 and 6.6 describe the setup for the standalone 0L analysis, then section 6.7 describes the setup for the combined analysis.

6.1 Background modeling

In order to measure $t\bar{t}Z \to \nu\bar{\nu}$ we need to model and contain the main backgrounds within the region. In this section we will cover what these backgrounds and how much of an impact they have to the measurement.

The main backgrounds in the 0L region are $\mathrm{t\bar{t}}$, Z+jets and single-top. The minor contributing backgrounds are W+jets, and $\mathrm{t\bar{t}W}$. We will now overview each of these backgrounds and their final states.

$\mathrm{t\bar{t}}$ is very similar to figure 5.2 and 5.3 but lacking the Z boson. $\mathrm{t\bar{t}}$. The final states for $\mathrm{t\bar{t}}$ are not contributing equally, with $\mathrm{t\bar{t}}$ 0L final state being fully hadronic they have

no way of producing missing energy in the detector. Since later we select on the missing energy this veto the all hadronic $\mathrm{t\bar{t}}$ decays from the background. Instead the vast majority of the background comes from $\mathrm{t\bar{t}}$ with 1L in the final state with the lepton failing to be reconstructed allowing it to pass any lepton cuts requiring exactly 0L leptons. There is also smaller amounts of $\mathrm{t\bar{t}}$ 2L which have both leptons fail to reconstruct but this is a rarer concurrence.

Z+jets is normally the Z-boson decaying into hadronic jets which occurs for 70% of Z boson decays. Since hadronic Z boson decays would also have the same issue as the hadronic $\mathrm{t\bar{t}}$ only $Z \to \nu\bar{\nu}$ would make it through the missing energy cuts and make up the background in the 0L region.

single-top similarly to Z+jets and $\mathrm{t\bar{t}}$ has its hadronic decays vetoed due to not producing enough missing energy. The single top itself can decay in two channels both t and s, most of the background comes from the t channel single-top decay as its cross section is much higher than s channel production. The single-top produced goes through the same decay final states that one top quark of the $\mathrm{t\bar{t}}$ goes through in figure 5.2. The leptonic decays which fail to reconstruct will pass the selection and be the main background. The single-top can also produce 2 b-jets from gluon splitting in the final state which can minic the $\mathrm{t\bar{t}}Z$ signature making it a difficult background to separate.

$\mathrm{t\bar{t}}W$ is also a background though due to it having an extra charged lepton from the decay of its W boson it can pass the missing energy cut when both the top and anti top decay hadronically. The lepton cut however requiring zero charged leptons would still require this lepton to fail to be reconstructed. It is much less likely for 2L and 3L final states to have all their leptons fail to reconstruct so so most of the $\mathrm{t\bar{t}}W$ will come from the 1L final state with the charged lepton coming from either of the top quarks or associated W boson. Since $\mathrm{t\bar{t}}W$ has a similar cross section to $\mathrm{t\bar{t}}Z$ but requires failing to reconstruct a 1L final state the amount of $\mathrm{t\bar{t}}W$ is very small.

W+jets similarly to other backgrounds cannot pass the missing energy cuts when it decays into hadronic jets, instead it needs to decay leptonically and then have that lepton fail to be reconstructed.

6.2 Sample selection

For these analyses we use the full ATLAS Run 2 datasets with integrated luminosity of 140 fb^{-1} at centre-of-mass energy $\sqrt{s} = 13\ \mathrm{TeV}$. For all samples, we use $\mathrm{DAOD\_PHYS}$ data format and the event simulation described in section 3.2. The MC simulation generators

used are described below:

- Simulation samples were generated for $\mathrm{t\bar{t}}$ and single-top we use the Powheg+Pythia8 generators.
- For Z+jets and W+jets simulations, we use Sherpa 2.2.11.
- $\mathrm{t\bar{t}W}$ is produced using sherpa 2.2.10.
- For simulating our signal ($\mathrm{t\bar{t}Z}$) we used $\mathrm{aMC@NLO}$ with Pyhtia8.

6.3 Variables used for the 0L analysis

We will show the various variables which are used to help reject background and define the signal, control, and validation regions. These variables are also used to train the neural networks presented in this paper. A full summary of all the variables used is shown in table 6.1.

6.3.1 Object-based missing transverse energy significance.

Missing transverse energy significance (S), while is derived from $\mathrm{E_T^{miss}}$ defined in 3.1.8. We use $\mathrm{E_T^{miss}}$ mainly as a measure of how many weakly interacting particles there are in the detector. However, this assumes that the ATLAS detector is 100% efficient at detecting more strongly interacting particles. In reality, some interacting particles escape the acceptance of the detector, are inaccurately reconstructed, or fail to be reconstructed. This is known as fake $\mathrm{E_T^{miss}}$.

The significance S evaluates the degree to which $\mathrm{E_T^{miss}}$ is consistent with momentum resolution and particle identification efficiencies. S is evaluated on an event-by-event basis from the log-likelihood ratio that the reconstructed $\mathrm{E_T^{miss}}$ is consistent with the null hypothesis of having zero real $\mathrm{E_T^{miss}}$. A high value of S indicates that $\mathrm{E_T^{miss}}$ likely came from a neutrino or another non-interacting particle; a low value indicates that it is more likely to be a fake $\mathrm{E_T^{miss}}$. The Object-based $\mathrm{E_T^{miss}}$ significance is defined as [64]

$$S = \frac{E_T^{miss}}{\sqrt{\sigma_L^2(1-\rho_{TL}^2)}} \tag{6.1}$$

Where ρ_{TL} is the correlation factor between the transverse and longitudinal momentum resolutions for all objects, σ_{L} is the is the expected resolution of the total longitudinal

momentum of all objects in the event as a function of the $\mathrm{p_T}$ of each object.

6.3.2 HT

HT is the scalar sum of all visible transverse momenta in an event

$$HT = \sum_i |p_{T,i}|. \tag{6.2}$$

HT, therefore, represents the overall energy scale of the detected part of the event.

6.3.3 Mass variables

The transverse mass ($\mathrm{m_T}$) is designed to give an event-by-event measure of the mass of a system where particles are pair-produced and one of them is undetected and massless. $\mathrm{m_T}$ is defined as[67]:

$$m_T^2 = 2(E_T^l E_T^\nu - \overrightarrow{p}_T^{\,l} \cdot \overrightarrow{p}_T^{\,\nu}) \tag{6.3}$$

Where $\mathrm{E_T^\nu}$ and $\overrightarrow{p}_{\mathrm{T}}^{\,\nu}$ are the transverse energy and momentum of the neutrino or invisible part of the process. $(\mathrm{p_T^\nu})^2 = (\mathrm{E_T^\nu})^2$ as we assumed the invisible part, the neutrino, is massless. $\mathrm{E_T^l}$ and $\overrightarrow{p}_{\mathrm{T}}^{\,\mathrm{l}}$ represent the lepton or the visible part of the process.

The stransverse mass ($\mathrm{m_{T2}}$) is defined similarly to $\mathrm{m_T}$ but for the case where the undetected particle is not massless. This measurement requires the decay of two massive particles into two pairs of detected and undetected particles. $\mathrm{m_{T2}}$ is defined as.

$$m_{T2}^2 = \min[\max(m_T^2(\overrightarrow{p}_T^{\,l1}, \overrightarrow{p}_1^{\,x}), m_T^2(\overrightarrow{p}_T^{\,l2}, \overrightarrow{p}_2^{\,x}))] \tag{6.4}$$

This minimization is performed across all possible 2 momenta for the two undetected particles, $\overrightarrow{p}_1^{\,\mathrm{x}}$ and $\overrightarrow{p}_2^{\,\mathrm{x}}$, such that they satisfy $\overrightarrow{p}_1^{\,\mathrm{x}} + \overrightarrow{p}_2^{\,\mathrm{x}} = \overrightarrow{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. This $\mathrm{m_{T2}}$ can have various forms depending on what process it is applied to. We will go through all the forms of $\mathrm{m_{T2}}$ that are used in this analysis:

First we are going to cover $\mathrm{m_{T2}}(\chi^2)$. This variable uses the previously mentioned $\mathrm{m_{T2}}$ method and uses the direction and magnitude of the missing momentum as well as the momentum of each of the top candidates reconstructed using a χ^2 method. The χ^2 method

for reconstructing the top candidate is used in a form of χ^2-like penalty function defined as [64].

$$\chi^2 = \frac{(m_{cand} - m_{true})^2}{m_{true}} \tag{6.5}$$

Where $\mathrm{m_{true}}$ is set to 173.2 GeV for the top quark candidate. $\mathrm{m_{cand}}$ is the mass of the top candidate. Selections are done to select top candidates, and start with pairs or single R=0.4 jets. The choice of whether to use a single or a pair is determined by whichever configuration gives the closest mass to the mass of the W-boson $\mathrm{m_W}$. The selections chosen on the $\mathrm{m_W}$ mass are then required to have an additional b jet in the same event to create a top-mass candidate. After the top quark candidates have been selected, their momentum can be used in the $\mathrm{m_{T2}}(\chi^2)$, which uses the momentum of the reconstructed top candidates previously mentioned and the missing momentum; however, it does not use the reconstructed values for the mass of the top candidate and instead uses the theoretical value 173.2 GeV. The missing momentum particles are assumed to be massless.

The transverse mass m_T and the stransverse mass m_{T2} combinations used are:

- $\mathrm{m_{T2}(bb)}$ which uses the $\mathrm{m_{T2}}$ reconstructed from the leading and sub-leading b-jets and the missing $\mathrm{p_T}$.
- $\mathrm{m_T(b, E_T^{miss})^{min}}$ is defined as being the transverse mass calculated from the $\mathrm{E_T^{miss}}$ and the b-tagged jet closest in ϕ to the direction of the missing momentum [64]

$$m_T(b, E_T^{miss})^{min} = \sqrt{2{p_T}^B E_T^{miss}[1 - cos\Delta\phi({p_T}^B, {p_T}^{miss})]}. \tag{6.6}$$

- $\mathrm{m_T(b, E_T^{miss})^{max}}$ is the same as $\mathrm{m_T(b, E_T^{miss})^{min}}$ but the b-tagged jet in this case is chosen as the one which is furthest away in ϕ from the missing momentum.

6.3.4 Lepton veto variables

Tau veto: Tau lepton's are difficult to directly detect in the ATLAS detector as they have a very short lifetime $(10^{-13}s)$ and are usually reconstructed from their decay products. In the zero lepton region by definition, we want to measure zero tau decay events in the final state and therefore veto these events. The tau veto is performed using 2 variables. The first variable is $nTau$, the number of reconstructed taus. This is the most effective of the 2 vetos and removes the majority of the tau event. The second part of the veto is called jet aligned $\mathrm{E_T^{miss}}$ cleaning and this is done in three steps:

1. The first check is that the jet is not a b-tag jet.

2. Next the algorithm checks the $\Delta\phi$ of the jet and the missing $\mathrm{p_T}$. If this value is less than $\pi/5$, then the jet is deemed to be originating from a tau event that came from a $\mathrm{W} \to \tau\nu$ decay where the tau decayed hadronically. The missing energy from the neutrino is therefore aligned with the hadronic jet.

3. tau events are assigned a tag so they can be vetoed.

The electron and muon vetos use variables called nEl and nMu which characterise the number of electrons and muons in the event. These are both set to zero to veto events with electrons and muons in the final state.

6.3.5 Jet variables

We have a number of variables defined using the antikt jet algorithm with radius parameter (R=0.4):

- The mass of the 1st, and 4th leading jet $\mathrm{m(j_1)}$,$\mathrm{m(j_4)}$.
- The number of jets $\mathrm{n_{jets}}$
- The momentum of the 1st, 2nd and 4th leading jet $\mathrm{p_T(j_1)}$, $\mathrm{p_T(j_2)}$ and $\mathrm{p_T(j_4)}$.
- $\mathrm{n_{b-jet}^{77\%}}$ the number of b-tag jets using 77% b-tag efficiency,
- b-tag working points (described in section 3.1.4) of the leading and subleading jet btagWP 1jet, btagWP 2jet
- $\Delta\mathrm{R(b1, b2)}$, the angular difference ($\Delta\mathrm{R}$) between the leading and subleading b jet
- $\Delta\phi[1..4]^{\mathrm{min}}$ the minimum difference in the azimuthal angle(ϕ) between the missing $\mathrm{p_T}$ and the leading four jets [64].

Only the most useful (Had high impact on DNN training in section 6.5 or was used in defining a region) jet parameters were kept,

6.3.6 Re-clustered jet variables

Re-clustered jets (rcjets) are produced using calibrated antikt(discussed in section 3.1.3) jets with radius parameter (R=0.4) as the inputs to antikt algorithm but with a larger clustering radius parameter (R=1.2) [68]. The rcjet variables used are:

- $\mathrm{m(j_1)^{R=1.2}}$, $\mathrm{m(j_2)^{R=1.2}}$ which is the mass for the leading and subleading rcjet

- $\mathrm{n}_{\mathrm{jets}}^{\mathrm{R=1.2}}$ the number of rcjets
- $\mathrm{n}_{\mathrm{constjets}}^{\mathrm{R=1.2}}$ the number of constituent jets (R=0.4 jets that have been clustered together into an rcjet) in the leading rcjet.
- The $\Delta\mathrm{R}$ between the leading and subleading constituent jets that contribute towards the leading rcjet $\Delta\mathrm{R}(\mathrm{j}_{\mathrm{1,const1}}^{\mathrm{R=1.2}}, \mathrm{j}_{\mathrm{1,const2}}^{\mathrm{R=1.2}})$
- The $\Delta\mathrm{R}$ between the leading and subleading RC jet $\Delta\mathrm{R}(\mathrm{j}_{1}^{\mathrm{R=1.2}}, \mathrm{j}_{2}^{\mathrm{R=1.2}})$.

Variable	Rough Description
$\mathrm{E_T^{miss}}$	missing transverse energy
$\mathrm{n_{b-jets}^{77\%}}$	b jet multiplicity.
$\mathrm{n_{jets}}$	R=0.4 jet multiplicity
$n\mathrm{EL}$	electron multiplicity
$n\mathrm{MU}$	muon multiplicity
$n\mathrm{Tau}$	tau multiplicity
τ jet cleaning	Selects jet events that aligned too closely with the $\mathrm{p_T^{miss}}$ in ϕ
$\mathrm{p_T(j4)}$	The transverse momentum of the 4th leading jet
$\mathrm{p_T(j2)}$	The transverse momentum of the 2nd leading jet
$\mathrm{p_T(j1)}$	The transverse momentum of the 1st leading jet
$\mathrm{m_T(b, E_T^{miss})^{min}}$	transverse mass calculated from the $\mathrm{E_T^{miss}}$ and a b-tag jet closest in ϕ to the direction of the missing momentum.
$\mathrm{m_T(b, E_T^{miss})^{max}}$	transverse mass calculated from the $\mathrm{E_T^{miss}}$ and a b-tag jet furthest away in ϕ to the direction of the missing momentum.
$\Delta\phi(1..4)^{\mathrm{min}}$	The minimum difference in azimuthal angle(ϕ) between the missing $\mathrm{p_T}$ and the leading four jets.
$\mathrm{E_T^{miss}significance}(S)$	object-based missing transverse energy significance, which calculates how related to real physics the missing energy is. Higher values represent more physical neutrino-like events, and lower values represent detector faults.
$\mathrm{m_{T2}}\chi^2$	The stransverse mass calculated from 2 top candidates selected using a χ^2 method.
$\mathrm{m_{T2}bb}$	stransverse mass calculated from the first two leading b jets.
$\Delta\mathrm{R(b1, b2)}$	The angular difference $\Delta\mathrm{R}$ between the leading and subleading b jet.
$\mathrm{m(j_1)^{R=1.2}}$	mass of the leading reclustered jet.
$\mathrm{m(j_2)^{R=1.2}}$	mass of the sub-leading reclustered jet.
$\mathrm{m(j_1)}$	mass of the leading jet.
$\mathrm{m(j_4)}$	mass of the 4th-leading jet.
$\mathrm{n_{constjets}^{R=1.2}}$	number of constituents in the leading RC jet.
$\Delta\mathrm{R(j_1^{R=1.2}, j_2^{R=1.2})}$	The $\Delta\mathrm{R}$ between the leading and subleading RC jet.
$\Delta\mathrm{R(j_{1,const1}^{R=1.2}, j_{1,const2}^{R=1.2})}$	The $\Delta\mathrm{R}$ between the leading RC jets and sub-leading constituent jets.

Table 6.1: list of variable definitions

6.4 Pre-selection

We start by defining a pre-selection of the events from which we refine the analyses to maximise the signal significance of $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$. The 0L region of $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$ pre-selection-mtb

is defined as:

- We require 4 jets, 2 of which are b-tagged with b-tag efficiency of 77%.
- We require $E_T^{miss} > 250$ GeV as this is the requirement to pass the missing transverse momentum trigger [69].
- We require $S > 5$ GeV in order to reject fake E_T^{miss}.
- The number of leptons, including electrons, muons and taus, is required to be 0.
- The tau suppression makes use of the tau jet cleaning variable defined in section 6.3 to remove hadronically decaying taus from $\mathrm{W} \to \tau\nu$ decays.
- The $\mathrm{p_T}$ of the 2nd leading and 4th leading jets are required to be greater than 80 and 40 GeV, respectively.
- $\mathrm{m_T(b, E_T^{miss})^{min}} > 200$ GeV and $\mathrm{m_T(b, E_T^{miss})^{max}} > 200$ GeV

The E_T^{miss} distribution is shown for the pre-selection-mtb in figure 6.1.

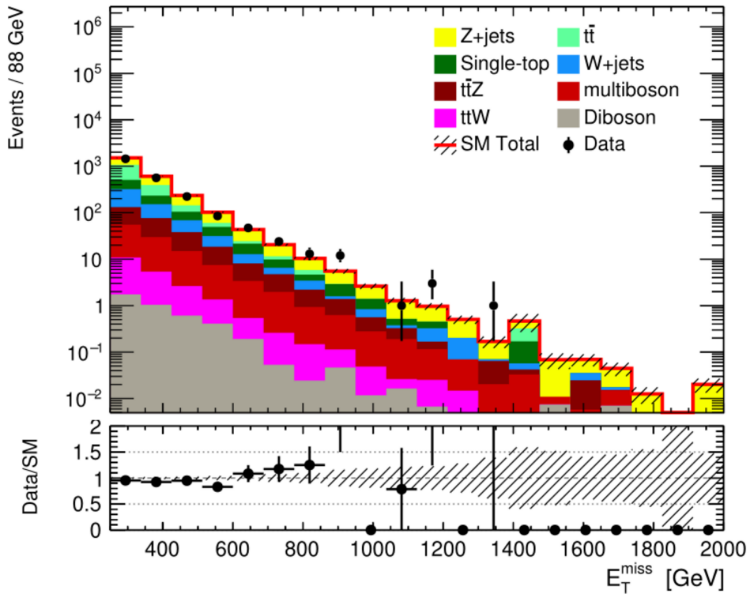


Figure 6.1: pre-selection-mtb data Monte Carlo plot. The MC is stacked so that the red line represents the total MC in that bin. The data used in this plot if from the full Run 2 data and luminosity 140 fb^{-1}. The MC and data are plotted across missing transverse energy in 20 bins between 250 and 2000 GeV.

The 0L region has an extremely large $\mathrm{t\bar{t}}$ background, which dominates the region, with most of it coming from the 1L channel when a lepton fails to reconstruct, as shown in

figure 6.2. A minority of the $t\bar{t}$ come from 2L channel, which has failed to reconstruct both leptons. None of the $t\bar{t}$ comes from the 0L channel as the 0L $t\bar{t}$ has no way of producing missing energy and would therefore not pass the pre-selection cut. The two cuts $m_T(b, E_T^{miss})^{min} > 200$ GeV and $m_T(b, E_T^{miss})^{max} > 200$ GeV, defined in equation 6.6, massively discriminate against $t\bar{t}$ while barely affecting $t\bar{t}Z$. This can be seen in figures 6.2 and 6.3 which show the distributions of these two variables in a region without the 200 GeV cuts on these variables made but with a 50 GeV cut made on $m_T(b, E_T^{miss})^{min}$. With these selections applied the largest contributor to the background events is now Z+jets as shown in figure 6.1.

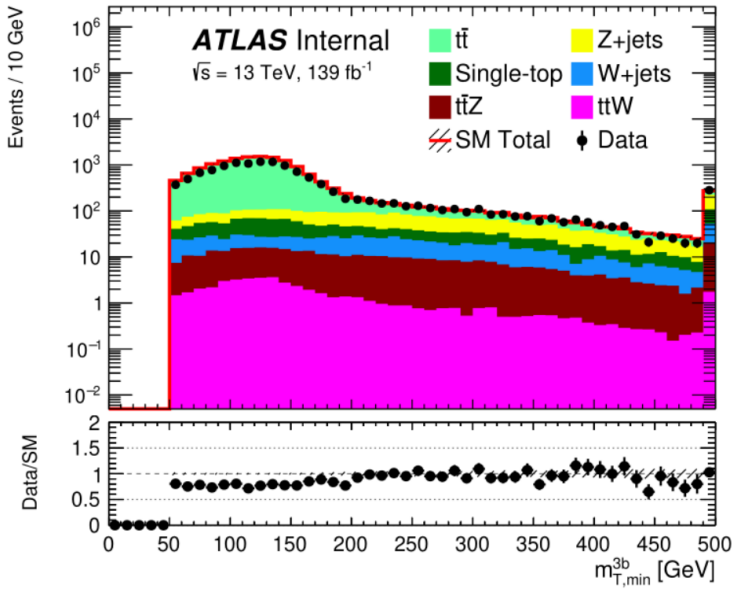


Figure 6.2: pre-selection-mtb data Monte Carlo plot. The MC is stacked so that the red line represents the total MC in that bin. The data used in this plot if from the full Run 2 data and luminosity 140 fb^{-1}. The MC and data are plotted $m_T(b, E_T^{miss})^{min}$ in 50 bins between 0 and 500 GeV.

This pre-selection-mtb region is used in the definition of all other regions used in the 0L analysis. The region has large amounts of background events that need to be filtered to allow for a meaningful signal region as can be seen in table 6.2 which shows the yields for all types of events in the pre-selection-mtb region. We therefore utilized a dense neural network that could utilize the most out of all the available variables. This is discussed in the next section.

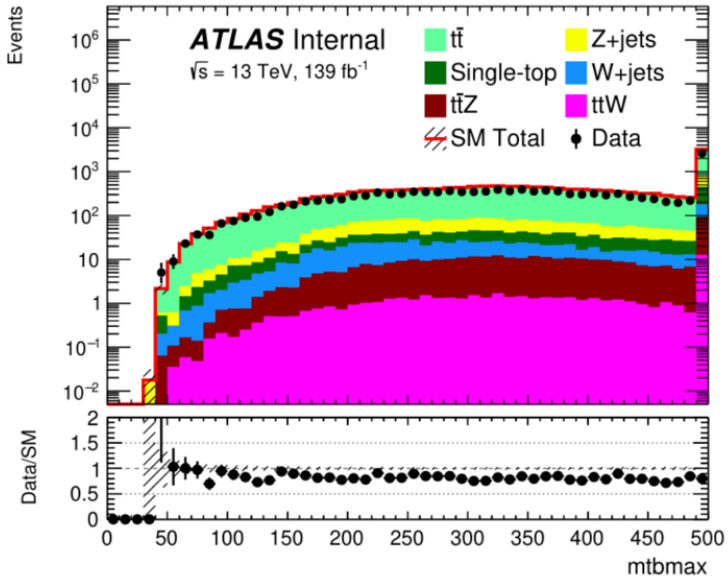


Figure 6.3: pre-selection-mtb data Monte Carlo plot. The MC is stacked so that the red line represents the total MC in that bin. The data used in this plot if from the full Run 2 data and luminosity 140 fb^{-1}. The MC and data are plotted $\mathrm{m_T(b, E_T^{miss})^{max}}$ in 50 bins between 0 and 500 GeV.

Process	Events
Data	2414.00 ± 49.13
Z+jets	825.35 ± 9.24
$\mathrm{t\bar{t}}$	781.58 ± 10.47
Single-top	337.06 ± 7.23
W+jets	317.27 ± 9.33
$\mathrm{t\bar{t}}Z$	167.62 ± 2.04
multiboson	93.73 ± 1.29
$\mathrm{t\bar{t}}W$	17.31 ± 0.38
Diboson	4.17 ± 0.19
SM	2544.09 ± 18.44

Table 6.2: Pre-selection-mtb yields

6.5 The Dense neural network for the standalone 0L channel

The MVA framework (version v2.0.9 [56]) was used for all training and evaluating of the 0L DNN, section 4.4 describes the definitions of the various DNN settings. A binary classifier was used to separate the signal $\mathrm{t\bar{t}Z}$ from the main backgrounds Z+jets, $\mathrm{t\bar{t}}$, W+jets and single-top, within the pre-selection-mtb. A BDT was also tested to separate out these variables, but the binary classifier was found to perform better than the BDT.

5 k-folds were used in the final DNN. Increasing number of k-folds increases the number of MC events that were being trained on. This was useful to use in the pre-selection-mtb region as one of its main issues was low statistics. We tested k-folds from 2 to 5, with 5 giving the largest separation. Increasing the k-folds above 5 would have risked affecting the shape of the output distribution for the k models. This is because the more models that are introduced, the fewer events there are in each model. Eventually the number of models is going to become so large that the DNN is no longer analysing physically equal subsets of the data. The conclusions and patterns that can be drawn from the variables will be different for each model leading to diverging models. This level of k-fold model divergence can be measured by the models shape on the ROC curve as well as their total AUC converging or not.

The ROC (defined in section 4.4.4) curve for the 5 k-models is shown in figure 6.4. The area under the operating curve (AUC, which is defined in 4.4.4) for the 5 k-models goes between 0.77-0.8. The AUC for our fits is a measure of how well the DNN has learned the signal events from the background events. 0.77-0.8 is a reasonable score, especially considering how few statistics we had to train on. These k-models' coverage, which means we aren't overspliting our training data over too many k-folds.

The MC have associated weights to simulate the reconstruction effects seen from data. The weights that are used per event for training the DNN are: the MC weight, the cross-section weight, the b-tag weight, the jet vertex tagging weight, the pileup weight, the year weight[1] and the lepton weight. Combinations of different weights and having no weights were tested to see if they yielded better results for the separation or signal significance. The MVA trainer was able to differentiate signal from background and keep the separation large in most configurations of the weights tested. Therefore, the main reason we included all the weights was due to this giving the largest value for the AUC total.

[1]ATLAS datasets for each run are broken up by the year in which they were taken and therefore require weighting based on how much data was taken that year.

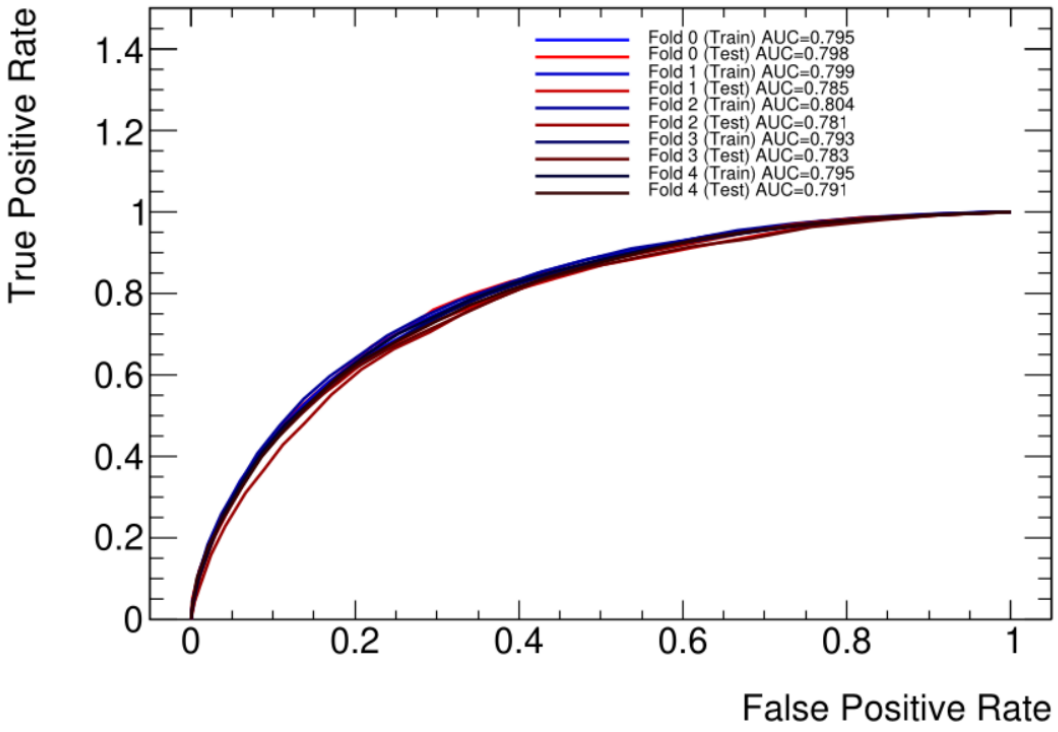


Figure 6.4: ROC curve for a k-5 fold model with a Binary classifier trained on pre-selection-mtb

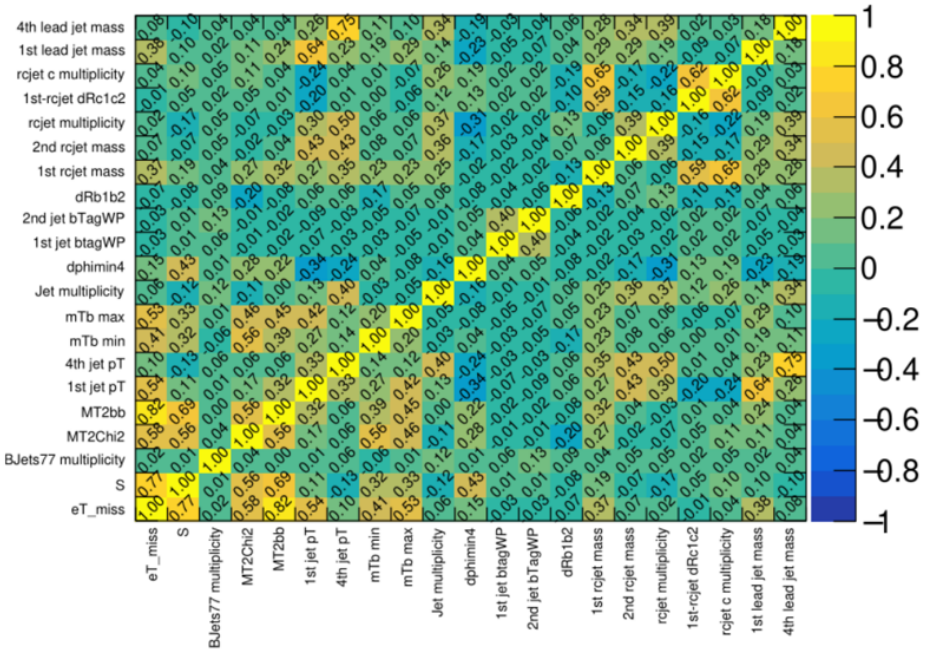


Figure 6.5: A correlation plot showing the degree to which two variables are related. 1 signifies that the variables are fully correlated. 0 represents completely unrelated variables. This plot corresponds to the variables used in the training of the binary DNN.

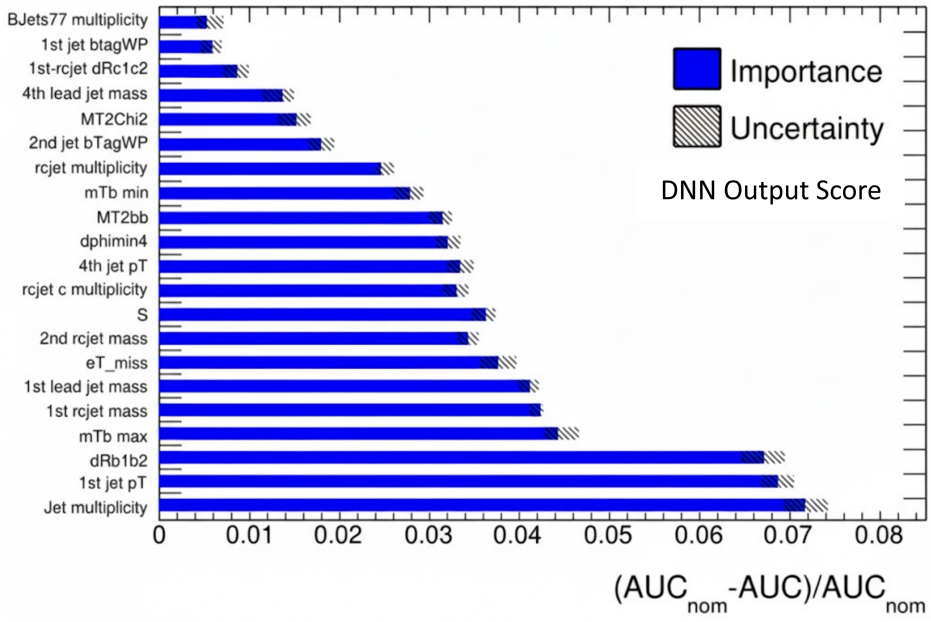


Figure 6.6: A plot showing the ranking of variables used in the training of a binary DNN. The ranking is based off the difference in AUC. The initial AUC, $\mathrm{AUC}_{\mathrm{nom}}$, is taken from the evaluation stage after the training of the DNN has finished. This value is then contrasted with a randomly permeated AUC. This randomly permeated AUC is taken from randomly changing each variable's values and then re-evaluating the AUC for each variable.

The input scaling model used was minmax. There were attempts to use $\mathrm{p_T}$ scaling and standard scaling for specific variables that would have been more suited to these scaling methods in the MVA's mixed scaling option. These attempts at mixed scaling resulted in overall lower separation, and therefore, only minmax scaling was used going forward.

The DNN itself was built using 2 hidden layers, each with 30 nodes. The loss function used in the training is the binary cross-entropy with a Nadam optimizer with a learning rate of 0.0005. The (ReLU) is used in the hidden layers; the output layer, instead, uses the sigmoid activation function. The batch normalization index was set to 1, which means batch normalization is applied at the first layer. The validation size per model was set to 25%.

The selection of input variables started from a large number of various kinematic variables that were known or expected as being good discriminants from previous SUSY searches, such as the variables listed at the beginning of this section [64]. We removed variables from our initial list from training for a few reasons:

- Incompatibilities between variables, where we kept the best-performing of the two variables.

- Certain variables' phase spaces were already covered by other variables highlighted in figure 6.5, showing the correlation between variables.
- Poor performance by variables in the importance plots.

The final list of variables used is shown in the final importance plot (After all poor preforming variables had been removed) in figure 6.6.

The MC samples used for background training were $\mathrm{t\bar{t}}$, Z+jets, W+jets and single top. Diboson, multiboson and $\mathrm{t\bar{t}W}$ samples were not significant enough backgrounds and were removed from the training, thus allowing the DNN to focus on the high-impact backgrounds.

6.6 Signal, Control and Validation region selection for the standalone 0L channel

The DNN described in section 6.5 is used to define our signal, control, and validation regions for the analysis. The DNN's performance is shown in figure 6.7, which is broken up into individual backgrounds in figure 6.8. When we plot the Data/MC distribution of this DNN variable in figure 6.9, we can also see what this performance means in terms of events.

A summary of the signal, control and validation region deviations is shown in table 6.3. There is also a visualization of this summary in figure 6.10.

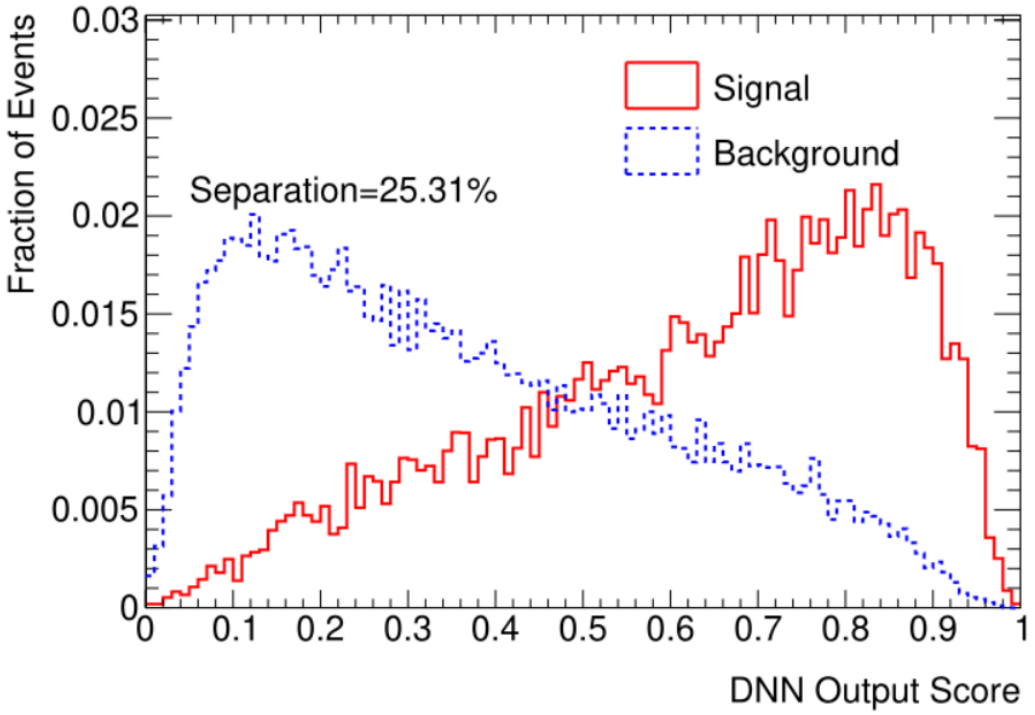


Figure 6.7: 1D DNN output score for a DNN Binary classifier trained to target $t\bar{t}Z$ against backgrounds of $t\bar{t}$, Z+jets, single top and W+jets with 100 bins

We can see from figure 6.7 that in the DNN output score range 0.75-0.9 we get the peak in the number of signal events. In figure 6.8 we can see most backgrounds are significantly reduced for DNN output scores above 0.4 with W+jets, Z+jets and $t\bar{t}$ all peaking around 0.2, however, single top is not as suppressed as the other backgrounds, and a wide peak is shown in figure 6.9.

St is incredibly difficult to separate from $t\bar{t}$, making selecting a control region for it

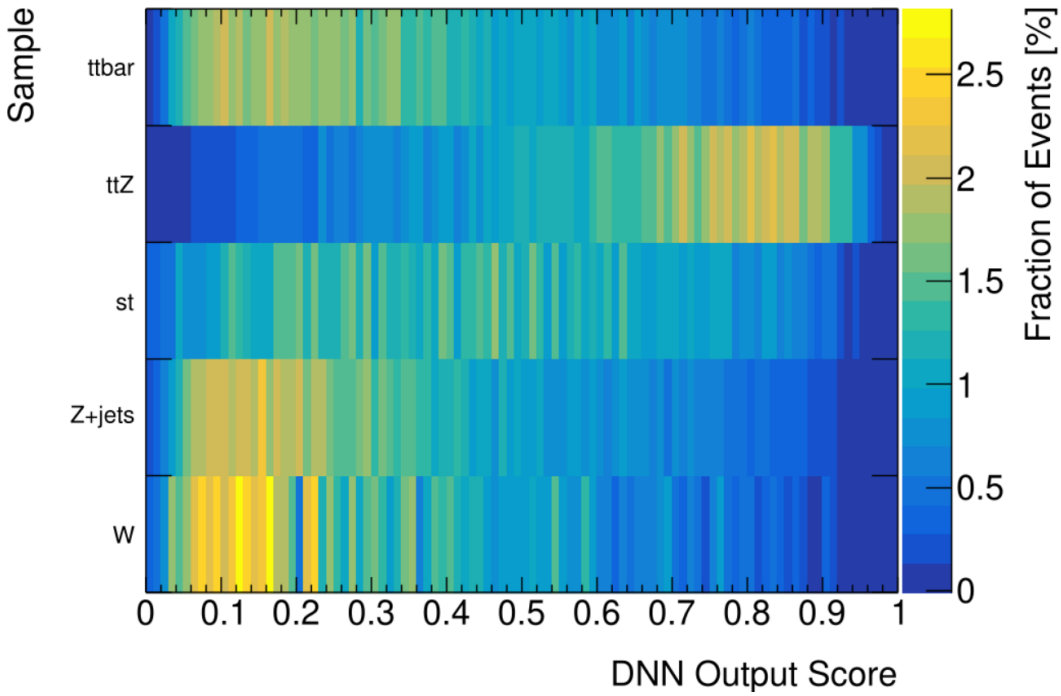


Figure 6.8: 2D DNN output score for a DNN Binary classifier trained to target $t\bar{t}Z$ against backgrounds of $t\bar{t}$, Z+jets, st and W+jets with 100 bins.

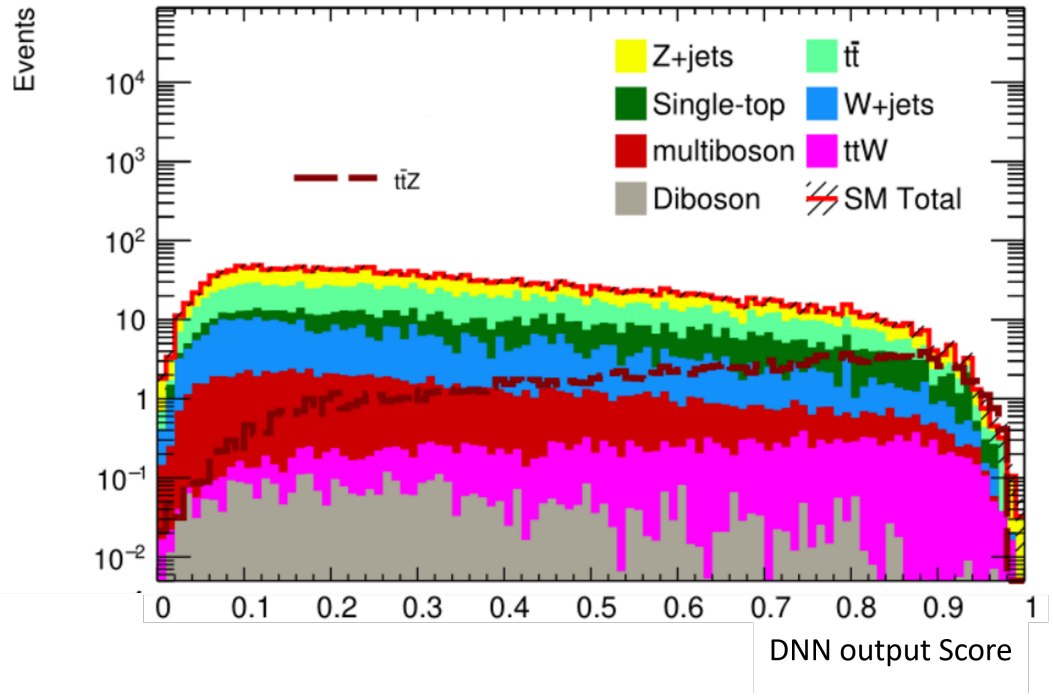


Figure 6.9: DNN output score for a DNN Binary classifier trained to target $t\bar{t}Z$ against backgrounds $t\bar{t}$, Z+jets, single top and W+jets with 100 bins displaying number of events for each bin.

complex. A large single-top contamination in the signal region, combined with a poor st control region, would impact the overall significance of the measurement. The signal region selection, therefore, needed to reduce the single-top where possible. Comparing these figures with the signal significance from the Data/MC figure 6.9, we selected for DNN output score > 0.75 and planned to bin across the DNN output score. These cuts are applied on top of the pre-selection-mtb.

The validation region had to be orthogonal to our signal region but also close enough in phase space to be kinematically similar. The validation region was therefore chosen to be between 0.75 and 0.63 in the DNN output binary score, with the phase space below 0.63 being used exclusively for control regions.

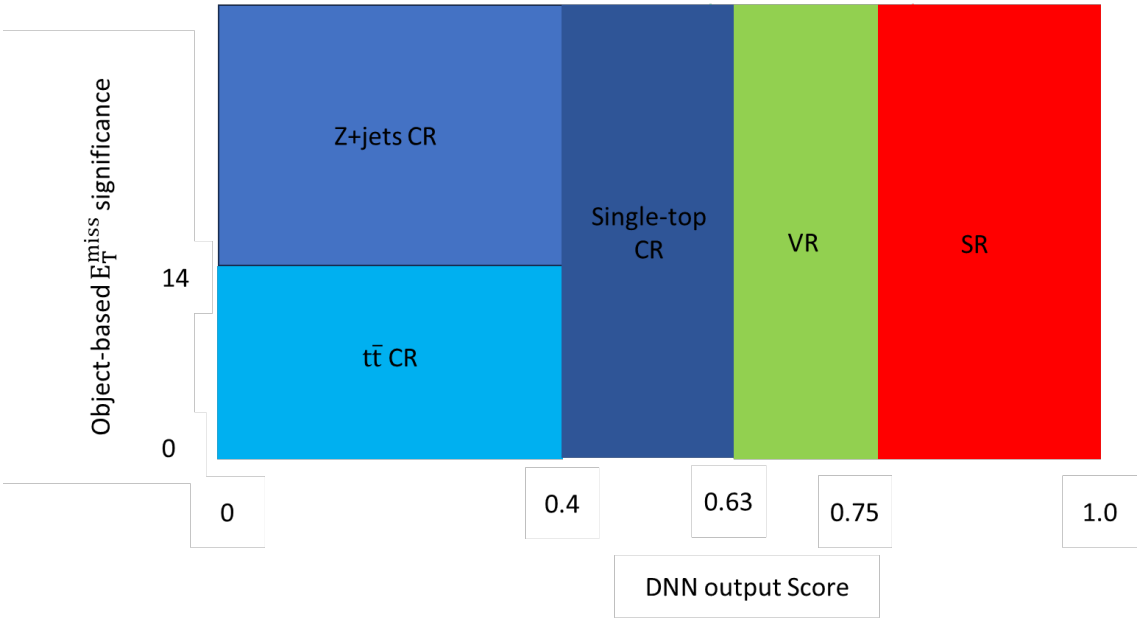


Figure 6.10: Visualisation of signal validation, and control region phase space.

The three biggest contributions to background in the signal region are Z+jets, single top and $t\bar{t}$ as shown in figure 7.4. These require control regions to constrain them. For the Control regions we can see in figure 6.8 that Z+jets and $t\bar{t}$ are peaking between 0 and 0.4 DNN output score. The single-top, however, is seen to have a more spread-out distribution over 0 to 0.6 DNN output score. A selection can therefore be placed at 0.4 DNN output score with the Z+jets and $t\bar{t}$ control regions existing below this cut (< 0.4) with additional cuts needed to provide orthogonality between Z+jets and $t\bar{t}$ while the single top control region is being placed above the selection (< 0.4) but below the 0.63 DNN output score to both ensure orthogonality with the VR and allow for as little signal contamination($t\bar{t}Z$) as possible.

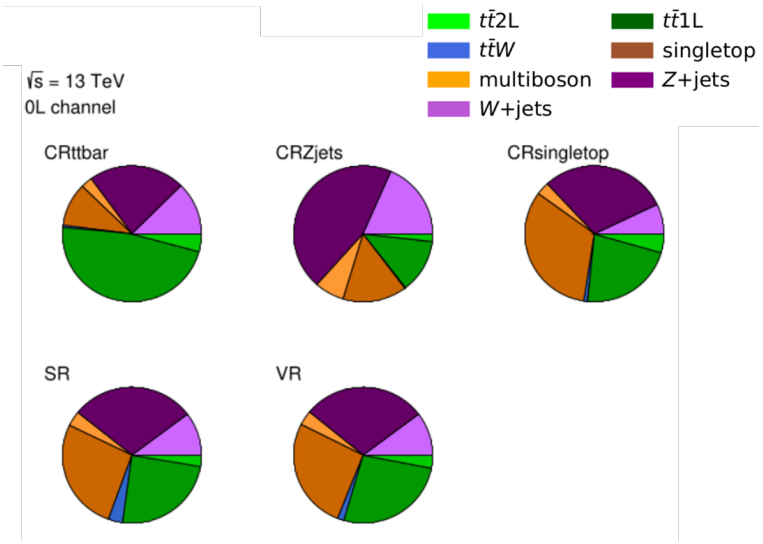


Figure 6.11: A pie chart of the background composition in each analysis region, created by TrexFitter.

Additional selections are applied to the single-top CR to improve the purity of the single-top for the region. The further selections made were: $\mathrm{n_{jets}} < 7$, $\Delta\phi(1..2)^{\mathrm{min}} > 0.8$, $S > 14$, $\Delta\mathrm{R(b1, b2)} > 2$, $\mathrm{m(j_1)^{R=1.2}} > 70$ GeV and $\mathrm{m_T(b, E_T^{missT})} > 220$ GeV. These cuts all reduced the substantial Z+jets and $\mathrm{t\bar{t}}$ backgrounds in the single top control region with the selection $\Delta\mathrm{R(b1, b2)} > 2$ removing most of the W+jets contribution. These cuts were made to improve the signal significance of st in its control region

The Z+jets control region builds from the < 0.4 DNN output score region and adds additional selections $\mathrm{m(j_1)^{R=1.2}} < 100$ GeV and $S > 14$ to discriminate against the $\mathrm{t\bar{t}}$ background.

The $\mathrm{t\bar{t}}$ control region(for $\mathrm{t\bar{t}}$ 1L) also builds from the <0.4 DNN output score cut and adds inverted versions of the selections from the Z+jets control region to ensure orthogonality. The inverse cuts are $\mathrm{m(j_1)^{R=1.2}} > 100$ GeV and $S < 14$, these cuts also ensure that the control region has plenty of $\mathrm{t\bar{t}}$ and little Z+jets.

Variable	$t\bar{t}$ CR	Z+jets CR	single top CR	VR	SR
pre-selection-mtb	applied	applied	applied	applied	applied
DNN output score	< 0.4	< 0.4	$> 0.4 \& < 0.63$	$> 0.63 \& < 0.75$	> 0.75
S	< 14	> 14			
$\mathrm{m(j_1)^{R=1.2}}$	> 100 GeV	< 100 GeV	> 70 GeV		
$\mathrm{n_{jets}}$			< 7		
$\Delta\phi(1..2)^{\mathrm{min}}$			> 0.8		
$\Delta\phi(1..4)^{\mathrm{min}}$					
$\Delta \mathrm{R(b1, b2)}$			> 2		
$\mathrm{m_T(b, E_T^{missT})}$			> 220		

Table 6.3: Signal, validation and control region definitions

6.7 Signal, Control and Validation region definitions for the combined channels

We now cover the combined 0L, 1L and 2L channels region definitions. The regions and variables defined for all three channels were designed using the same framework. This allowed for all three channels to be orthogonal to each other during the fitting stage of the analysis.

The work on the 1L and 2L channels was conducted by collaborators in the analysis group with the intention of combining the channels for combined results. We will now define the pre-selections for the 1L and 2L channels.

The pre-selection for the 1L used the stop-1L paper's pre-selection as inspiration [65]. It required that $nEl + nMu = 1$ with the transverse momentum of the single lepton required to be $\mathrm{p_T}(\ell) > 27$ GeV. The number of jets was required to be at least 4 and the number of b-jets to be at least 2. These cuts were made to best recreate the 1L final state.

The rest of the pre-selection cuts for the 1L remove specific sources of backgrounds.

$\Delta\phi(1,2)^{\mathrm{min}}$, which is the minimum difference in azimuthal angle(ϕ) between the missing $\mathrm{p_T}$ and the leading two jets, was required to be greater than 0.4 in order to remove fake sources of $\mathrm{E_T^{miss}}$. To remove semileptonic $\mathrm{t\bar{t}}$ we use $\mathrm{m_T}(\ell)$, which is the transverse mass reconstructed using the single lepton as the visible component, was required to be greater than 190 GeV. This is because this variable has an endpoint for the semi-leptonic $\mathrm{t\bar{t}}$ at the W mass due to its missing energy coming exclusively from the $W \to \ell\nu$, while the signal for this variable typically has a much larger value due to the extra missing energy from the $\mathrm{Z} \to \nu\bar{\nu}$. The di-leptonic $\mathrm{t\bar{t}}$ is also reduced by placing a selection on two variables the Topness and $\mathrm{m_{T2}^{\tau}}$. Topness is a variable which measures how likely an event is to come from a top quark, was required to be greater than 0. $\mathrm{m_{T2}^{\tau}}$, which is the stranverse mass using an available tau and the missing energy in its calculation, was required to be greater than 100 GeV. Events where no $\mathrm{m_{T2}^{\tau}}$ could be constructed were also accepted.

The two charged lepton pre-selection was built using the stop-2L pre-selection [66]. The following set of cuts isolates the $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$ 2L final state. The 2L pre-selection has a strict tau veto and required exactly two leptons other than a tau. The $\mathrm{p_T}$ for the first lepton was required to be $\mathrm{p_T}(\ell_1) > 27\ \mathrm{GeV}$ while the second leptons transverse momentum was required to be $\mathrm{p_T}(\ell_2) > 20\ \mathrm{GeV}$. The number of jets for the 2L was required to be at least 2.

The following pre-selection cuts remove specific background events. The stanverse mass for same-flavour leptons was required to be $81 < \mathrm{m_{T2}^{SF}}(\ell\ell) < 101\ \mathrm{GeV}$, in order to move them outside of the Z mass peak, to remove $\mathrm{Z} + \mathrm{jets}$ events. To reduce $\mathrm{t\bar{t}}$ events, we require the stansverse mass with the two leptons as input to be $\mathrm{m_T^{\ell\ell}} > 50\ \mathrm{GeV}$ and $S > 8$. The change in azimuthal angle between $\mathrm{p_T^{miss}}$ and $\mathrm{p_{T,boost}^{\ell\ell}}$ is required to be $\Delta\phi_{\mathrm{boost}} < 1.0$ to remove none collimated none-hadronic background events. $\mathrm{p_{T,boost}^{\ell\ell}}$ is the vectorial sum of the p_T^{miss}, $\mathrm{p_T}(\ell_1)$ and $\mathrm{p_T}(\ell_2)$.

The 0L region setup is nearly identical to the one in section 6.6, other than a few changes. The validation region in the 0L was made into a secondary signal region in the combined fit, with some slight adjustments. Some minor adjustments were made to the single-top control region as well. These changes were made as the combined measurements had validation regions in their 1L and 2L channels, meaning there was some validation of the samples occurring. This allowed the combined measurement to utilize the 0L validation region's phase space as an additional signal region, as the 0L struggled for significance compared with the other channels. A separate DNN was trained for the 0L in the combined measurement using the same configuration and produced almost identical results, however, there were a few phase space changes needed to fully optimize this and hence the small changes to the regions. A full breakdown of the 0L regions used for the combined measurement are shown in table 6.4.

Variable	$t\bar{t}$ CR	Z+jets CR	single-top CR	SR2	SR
pre-selection-mtb	applied	applied	applied	applied	applied
DNN output score	< 0.4	< 0.4	> 0.4 & < 0.6	> 0.6 & < 0.75	> 0.75
S	< 14	> 14			
$m(j_1)^{R=1.2}$	> 100 GeV	< 100 GeV			
n_{jets}			< 6		

Table 6.4: Signal, validation and control region definitions for the 0L in the combined fit

The 1L region is defined using a multi-class DNN which trains to identify multiple types of processes. The 1L channel set the DNN to classify its two main backgrounds $t\bar{t}$ and single-top, as well as the signal. This produced 3 different DNN scores. The 1L channel used its DNN scores to define its signal, control and validation regions. This is summarized in table 6.5.

Variable	$t\bar{t}$ 1L CR	single-top 1L CR	VR	SR1	SR2
pre-selection-1L	applied	applied	applied	applied	applied
DNN $t\bar{t}$ score	≥ 0.4	< 0.16	$(-)$ (≥ 0.4)	< 0.16	$(\geq 0.16$ & $< 0.4)$ (< 0.16)
DNN single-top score	< 0.21	≥ 0.84	$(\geq 0.21$& $< 0.4)$ $(\geq 0.4$ & $< 0.84)$	< 0.21	$(\geq 0.21$ & $< 0.4)$ (< 0.4)

Table 6.5: Signal, validation and control region definitions for the 1L in the combined fit. The cuts in the brackets are applied exclusively together, while cuts not in brackets are applied always.

The 2L channel had already well-defined regions from previous SUSY studies [66] and therefore did not require a DNN. The region also had very low statistics and therefore would struggle with training of its DNN. The region instead made use of the E_T^{miss} and

object-based $\mathrm{E_T^{miss}}$ significance. It also made use of a 2L only variable $\mathrm{m_{T2}}(\ell\ell)$, which can be defined as the $\mathrm{m_{T2}}$ calculation(from equation 6.4) using 2 leptons as the inputs. The regions the 2L used are defined in table 6.6

Variable	$\mathrm{t\bar{t}}$ 2L CR	0b 2L CR	SR
pre-selection-2L	applied	applied	applied
$\mathrm{n_{b-jets}^{77\%}}$	≥ 1	$= 0$	≥ 1
$\mathrm{M_{T2}ll}$	> 80 GeV	> 100 GeV	> 100 GeV
S	< 12	> 12	> 12
$\mathrm{E_T^{Miss}}$	> 100 GeV	> 200 GeV	> 200 GeV
$\mathrm{m_T}(\ell_1)$			$< 160GeV$
$\mathrm{m_T}(\ell_2)$			$< 150GeV$

Table 6.6: Signal, validation and control region definitions for the 2L in the combined fit

Chapter 7

Top-quark pair production in association with a Z boson fit and results

This chapter carries on from the previous chapter 6 and uses the regions defined in section 6.6 to fit the MC to the data samples for the 0L fit. We make a cross-section measurement of $\mathrm{t\bar{t}Z} \rightarrow \nu\bar{\nu}$ and compare it to the value predicted by the SM and test for new physics. In section 7.1 we cover the fitting methodology used to fit the 0L regions and measure the cross-section. In section 7.3 we cover the results of the 0L channel fit. In section 7.4 we briefly cover the setup for the combined fit, which uses a slightly modified version of the 0L fit from section 7.1. This fit is still blinded and therefore only the systematic ranking and the region setup is covered. In section 7.5 we covered the results of the combined fit. Since the results are still blinded, only the predicted accuracies are relevant in this section.

It is important to note in this chapter, we present the fitting strategy and fitting results for two distinct analyses and The standalone 0L channel fit for $\mathrm{t\bar{t}Z} \rightarrow \nu\bar{\nu}$, which is covered in sections 7.1 and 7.3 and the combined 0L, 1L and 2L channel fit for $\mathrm{t\bar{t}Z} \rightarrow \nu\bar{\nu}$ which is covered in sections 7.4 and 7.5.

7.1 Fitting strategy 0L

The fit was performed using the TrexFitter (version 4.28[70]) package and the fit used the regions defined in table 6.3 to fit its signal, control and validation regions. The signal region was binned in 4 bins using the DNN output score between 0.75 and 1. Two of these bins are very pure in $\mathrm{t\bar{t}Z}$ but with low statistics, while the bins are the lower end of the DNN output score have more statistics but less signal purity. In figure 7.1.(e) we show the pre-fit plot for the signal region.

The validation region. $\mathrm{t\bar{t}}$ control region and the Z+jets control region were all binned in HT using 4 bins between 250 to 1000 GeV. the st control region was binned in 2bins

between . The HT distribution in the pre-fit plots for these regions is shown in figure 7.1(a-d).

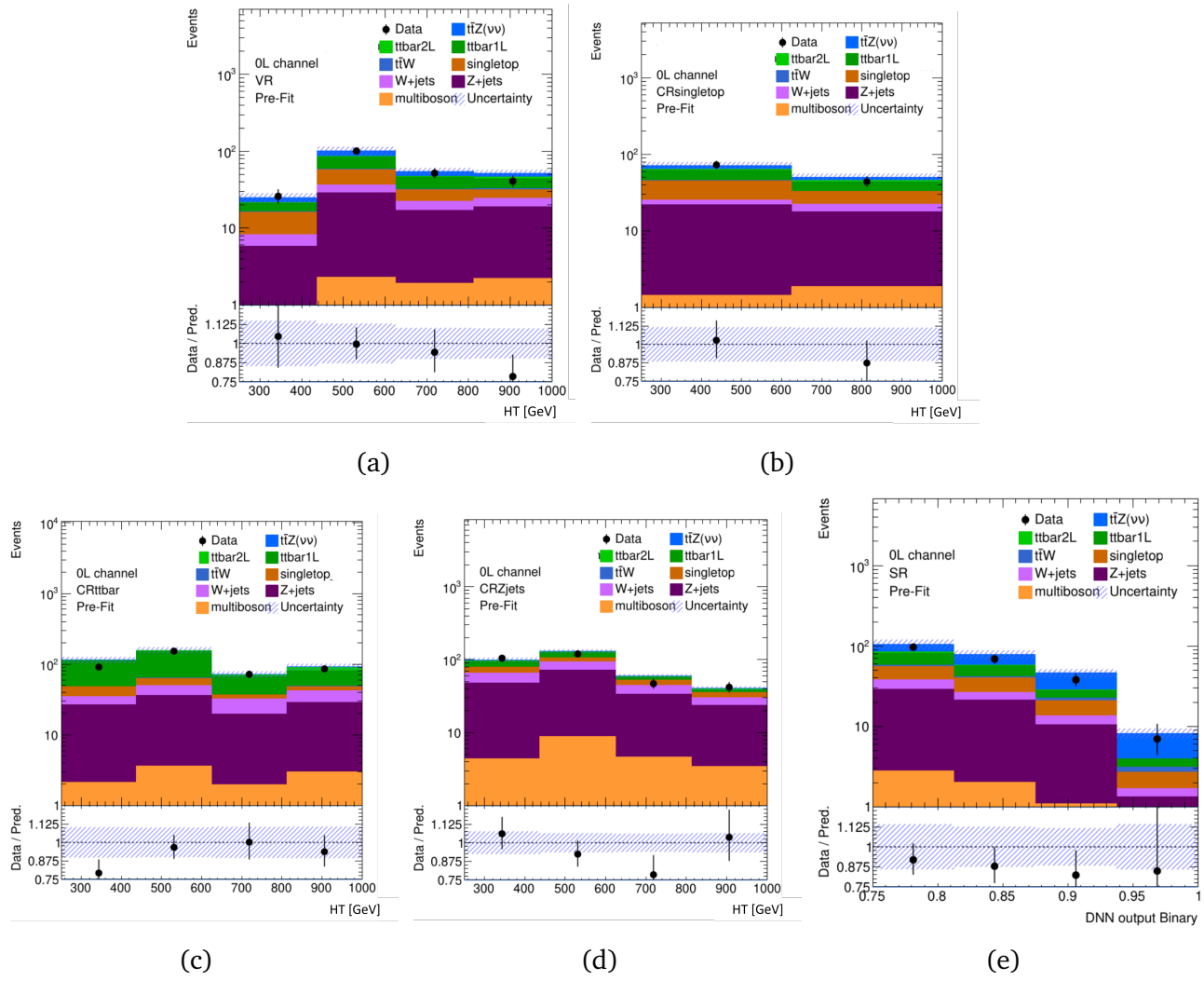


Figure 7.1: Histograms for the 0L channel in 4 bins in HT for: (a) the validation region, (c) the $t\bar{t}$ control region and (d) the Z+jets control region. (b) shows the single-top control region in 2 bins in HT, and (e) shows the DNN output score in the signal region in 4 bins. These are pre-fit plots, which are produced before the fit is applied. The histograms are stacked so the vertical widths represent the number of events in each bin. The subplot at the bottom of each plot shows the MC agreement with data.

The plots shown in figure 7.1 are used as inputs to the 0L fit. The fit will use these pre-fit plots to scale the backgrounds: Z+jets, singletop and $t\bar{t}$ MC and the signal $t\bar{t}Z$ to the data in their respective control and signal regions while keeping the MC in the validation region in line with the data. These produce scale factors for each major background and the signal. The scale factors for the 3 main backgrounds are floated as nuisance parameters without constraints, while the systematic uncertainties covered in section 7.2 are

floated as nuisance parameters with gaussian constraints. This fit is preformed simultaneously across the signal and control regions. The parameter of interest for this fit is the scale factor for the predicted cross-section (XS) for the process $\mathrm{t\bar{t}Z} \rightarrow \nu\bar{\nu}$ also known as the signal strength. The signal strength is the relative ratio of observed events vs the predicted events from the MC simulations. The cross section is the effective area, measured in barns ($10^{-28}\mathrm{m}^2$), that a incident particle must hit for an interaction to occur. It is related to the probability that a process will occur given some conditions. Measuring the signal strength will allow us to infer information about the cross-section.

A summary of these pre-selection plots and data/MC composition in each region is shown in figure 7.2, and the background compositions are also represented in figure 7.4. The signal in each of these regions, other than the signal region, should be minimal to avoid biasing the fit modelling of the scale factors on the backgrounds. This signal contamination in the control regions is shown in figure 7.3 in terms of approximate signal significance ($S/\sqrt{B}$) to determine how much of an impact the signal contamination will make in each region.

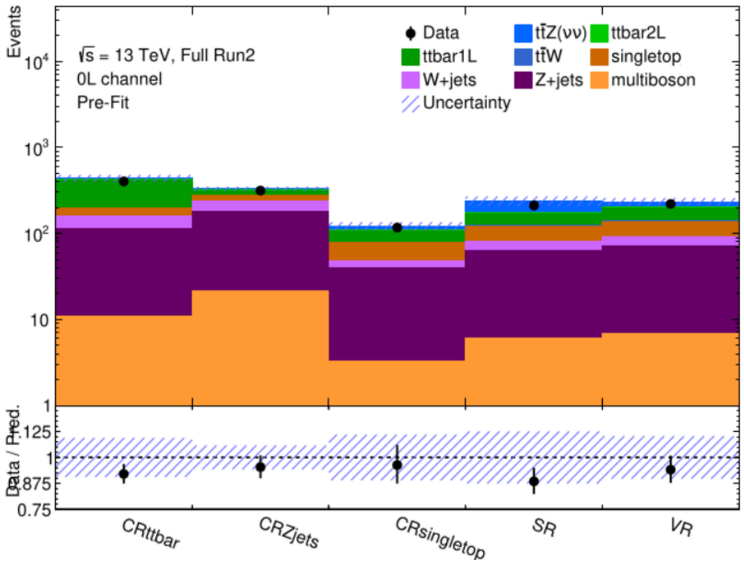


Figure 7.2: A summary of all the regions pre-fit, for the 0L channel fit, fitted by Trexfitter.

The largest backgrounds in the signal region, shown in figure 7.4, in descending order are Z+jets, $\mathrm{t\bar{t}}$ and single-top, since these are the largest backgrounds we define control regions to constrain them. We also see across figure 7.2 that there is a general MC excess across all the regions, suggesting that most of the scale factors will be less than 1.

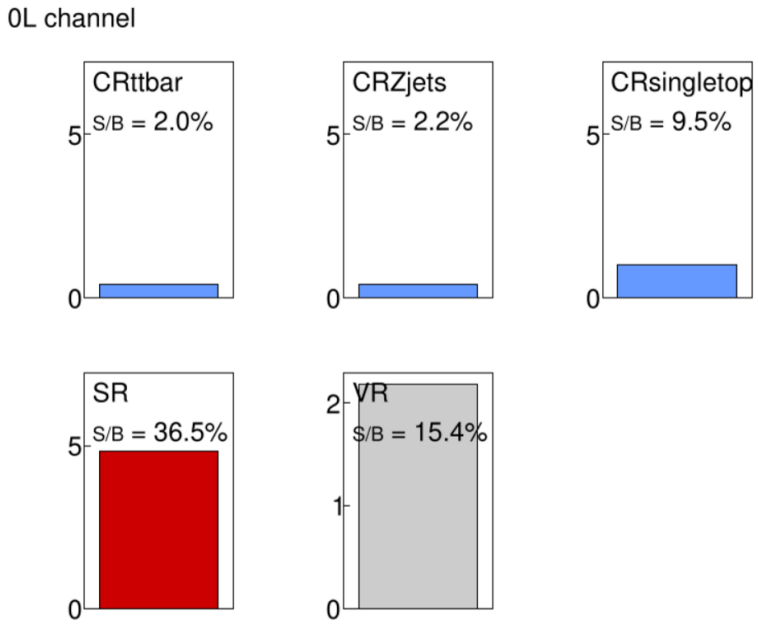


Figure 7.3: A diagram showing the signal purity in each region for the 0L channel fit, where S and B are the signal and background counts in each region. The y-axis shows $S/\sqrt{B}$

.

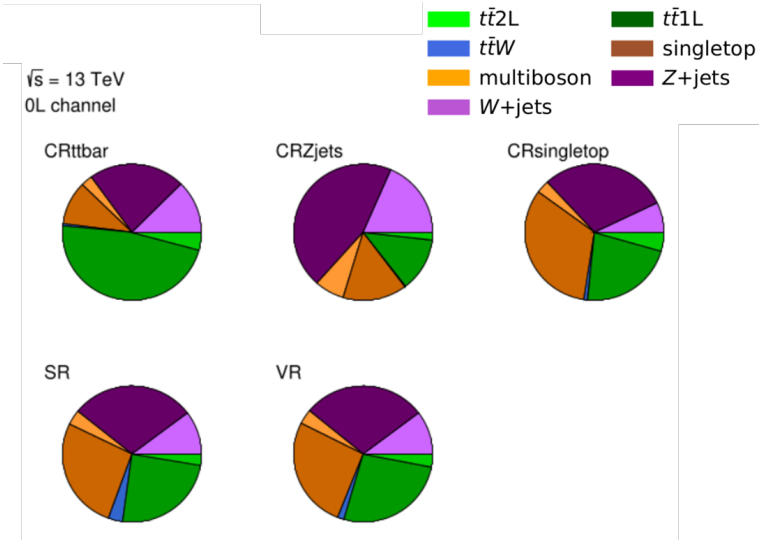


Figure 7.4: A pie chart of the background composition in each pre-fit region for the 0L channel fit, created by TrexFitter.

7.2 Systematics uncertainties

The systematic uncertainties included when evaluating the 0L fit are split into two categories: detector and theory systematics.

7.2.1 Detector systematics

The ATLAS detector has various calibrations and objects it reconstructs from data, which carry with them uncertainties that often need to be transferred onto the MC simulations. The detector systematics we consider in this analysis are:

- e/γ systematics, which are the electron/photon efficiencies for reconstruction, identification, and isolation. These efficiencies need to be passed on to the MC through scale factors. The scale factors have uncertainties associated with these efficiencies e/γ, which are represented as systematics in our MC samples.

- The muon reconstruction, isolation and identification efficiencies in the reconstruction of data events are passed along to the MC in order for it to accurately represent the data its modeling. This is done through a scale factor with associated uncertainties.

- The jet energy scale (JES) corrects jets for mismeasured calorimeter energy in event reconstruction; these corrections are not perfect and have associated uncertainties that must be passed onto the MC events.

- The jet energy resolution (JER), which measures how responsive the jets are to the detector. The variations calculated from the JER need to be propagated for the calculation of $\mathrm{E_T^{miss}}$. The MC events need to follow this reconstruction process and take the associated uncertainties with the JER into account.

- The jet vertex tagging (JVT) has an efficiency associated with it. In simulation, we try to account for this with a scale factor based on the JVT efficiency. The uncertainties related to this scale factor have to be applied to jet MC events.

- The b-tagging of data events has a b-tagging efficiency (as discussed in section 3.1.4) and related uncertainties that depend on the p_T and η. These uncertainties need to be passed on to MC events containing jets depending on their flavour, p_T and η.

- The LHC luminosity value used to scale all produced MC events has associated uncertainties that need to be estimated. The calibrations are done by using LUCID (more details in section 2.2.6) running at lower luminosity and then scaled to high luminosity. The uncertainties associated with this process is passed on to the MC events.

- Pile-up in the detector needs to be accounted for in MC events. MC events are rescaled based on pile-up conditions. The uncertainties associated with this rescaling are passed on to the MC events. This systematics is named pile-up reweighting (PRW).

- The $\mathrm{E_T^{miss}}$ calculation requires information from the whole event to reconstruct, including both hard and soft contributions discussed in section 3.1.8. Uncertainties in measuring the soft contributions and the resolution of $\mathrm{E_T^{miss}}$ need to be passed onto the MC events.

7.2.2 Theory systematics

The theoretical systematics come from the MC generators' ability, or lack thereof, to accurately produce events. The MC events can have fluctuations from systematics affecting their generation. These fluctuations can affect the kinematic distribution of MC events and the normalization applied to the background in the signal region. Therefore, we have to classify how much of an impact these fluctuations may have. These systematics are quantified empirically through comparing different MC generators output and using the spread in values as the $\pm\sigma$. The theoretical systematics we consider in this analysis are:

- The renormalization scale factor μ_R, which defines the α_s used in the generators matrix elements, and the factorization scale factor μ_F, which regulates the generator's PDF, need to have their related uncertainties accounted for in the $\mathrm{t\bar{t}Z}$, $\mathrm{t\bar{t}}$ and single-top MC samples.

- The parton distribution function for protons at the LHC(PDF4LHC) is critical for event generation as this helps to simulate the initial collision. The uncertainties associated with the PDF4LHC are needed for the $\mathrm{t\bar{t}Z}$, $\mathrm{t\bar{t}}$ and single-top events.

- The simulation of the top quark showering has uncertainties associated with it that need to taken in account when generating $\mathrm{t\bar{t}Z}$, $\mathrm{t\bar{t}}$ and single-top events.

- In POWHEG generation there is a parameter called h_{damp} which regulates high p_T radiation and since both $\mathrm{t\bar{t}}$ and single-top samples are produced using POWHEG we need to take into account the uncertainties associated with h_{damp}.

- p_T(hard1) is a systematic that describes the uncertainty on the choice of an internal parameter of POWHEG which is used to veto phase space in parton showering. This uncertainty effects $\mathrm{t\bar{t}}$ and single-top events.

7.3 Results for 0L only fit

In this section we present the results for the 0L only fit for $t\bar{t}Z \to \nu\bar{\nu}$. The scale factors of the post-fit regions for the signal, Validation and control regions are shown in figures 7.5.

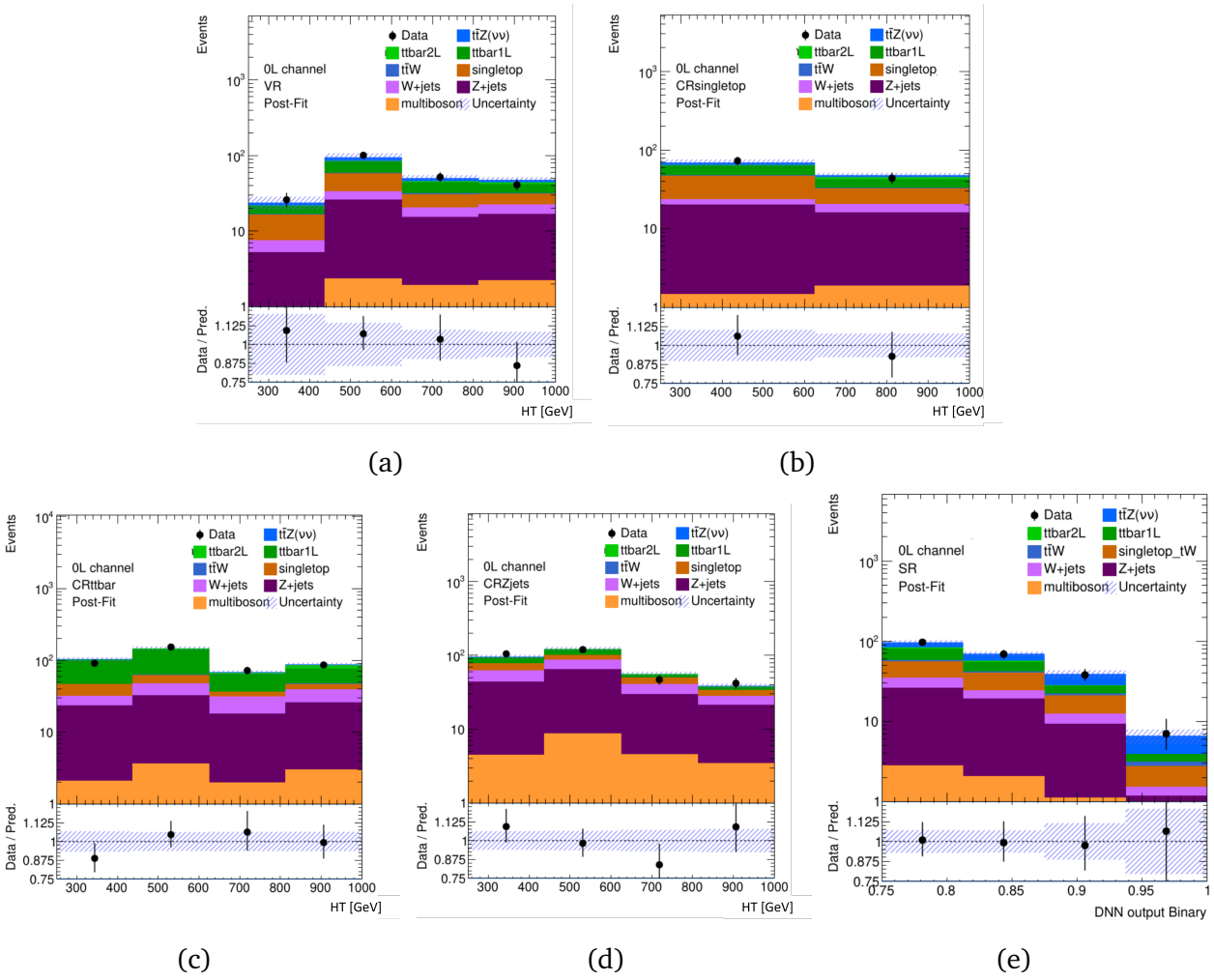


Figure 7.5: Histograms plotted for the 0L channel in 4 bins in HT for: (a) the validation region, (b) the single-top control region, (c) the $t\bar{t}$ control region, (d) the Z+jets control region and (e) the signal region which is binned in 4 bins in the DNN output score. This is created in TrexFitter. These are post-fit plots produced after the fit is applied. The histograms are stacked so the vertical widths represent the number of events in each bin. The subplot at the bottom of each plot shows the MC agreement with data.

The plots in figure 7.5 have had the scale factors produced from the fit applied to them, changing them from the pre-fit plots shown in figure 7.1. These plots can be summarized in figure 7.6, where we can see the scale factors have matched the MC and data together in each region. We also show figure 7.7 to show how the background distributions have

changed with the scale factor application.

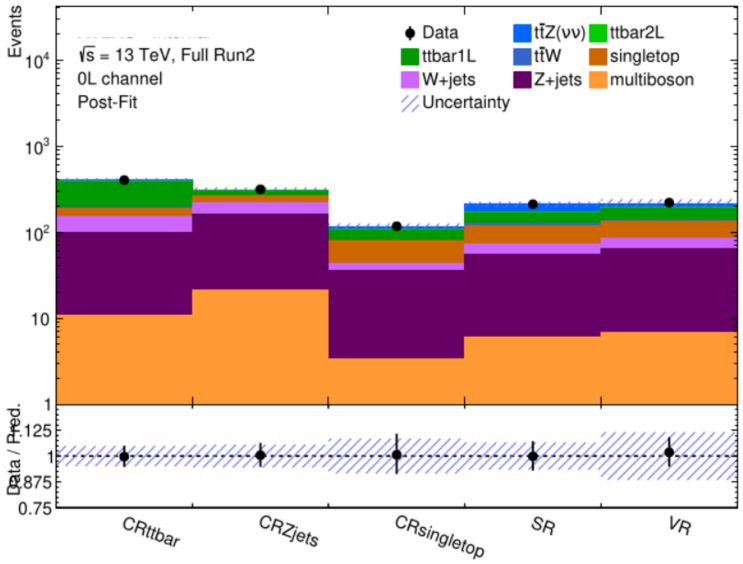
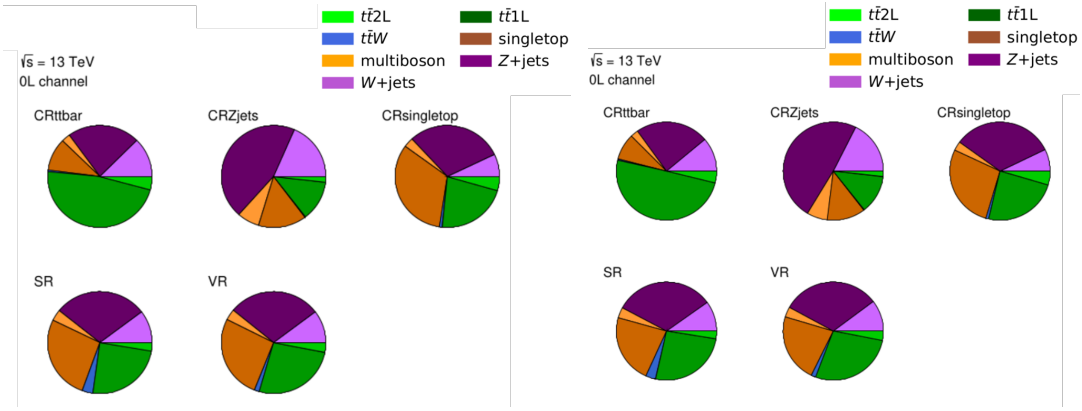


Figure 7.6: A summary of all the regions post-fit, for the 0L channel fit, fitted by TrexFitter.



(a) (b)

Figure 7.7: A pie chart of the background in each pre-fit (a) and post-fit (b) region for the 0L channel fit, created in TrexFitter.

As can be seen in figure 7.6, the MC has been scaled down to match the data. The relative distribution of the backgrounds has not changed significantly between figures 7.4 and 7.7. The scale factors for the signal and control regions are shown in figure 7.8. As can be seen, the errors on the single top scale factors are the largest. This is due to the poor single-top purity and statistics in the region.

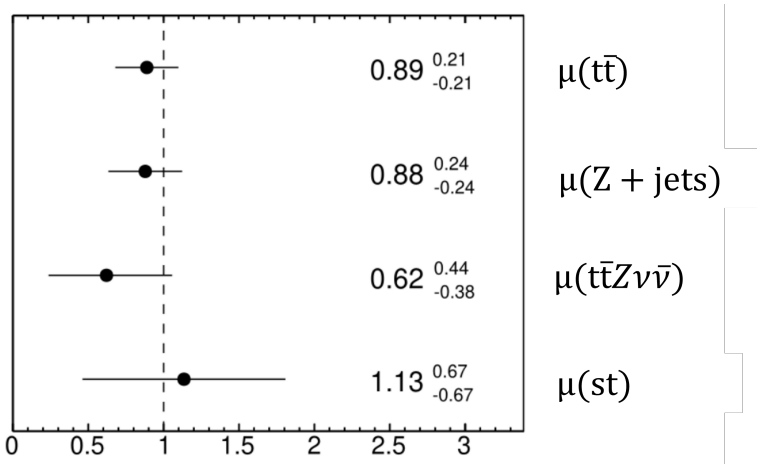


Figure 7.8: A scale factor plot showing the scale factors applied to each control region and to the cross section of $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$ in the signal region, for the 0L channel fit, by TrexFitter.

The ranking of the highest impacting systematics and the highest impacting scale factor uncertainties on the fit results are shown in figure 7.9.

As can be seen from figure 7.9 the scale factor (μ) on the single-top control region, the $\mathrm{t\bar{t}}$ 1L showering uncertainty, the μ on the Z+jets control region and the $\mathrm{t\bar{t}Z}\nu\bar{\nu}$ showering uncertainties have the biggest impact on the cross section measurement. This is reasonable as the single-top is not very well separated from other processes in its control region (seen in figure 7.4) and roughly makes up a third of the background in the signal region. This lack of separation is due to the single-top being very closely correlated with $\mathrm{t\bar{t}}$. This is made worse by the dominance of Z+jets in the pre-selection-mtb, which was also difficult to separate.

To calculate the cross-section measured in this analysis, we take a best guess for what the cross-section is. We guess that the MC scale factor for the signal region is 1, which means the MC and data match in the region. If they don't match then the scale factor we apply in the signal region can be smaller or greater than 1 to compensate. We take the scale factor and apply it to the theoretical cross section for $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$ calculated in equation 5.1 to gain our measured cross section.

The $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$ scale factor, shown in figure 7.8, was extracted and multiplied by the cross-section theoretical value used in producing the samples. The inclusive cross-section measurement for $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$ is:

$$\sigma_{t\bar{t}Z\to\nu\bar{\nu}} = 0.11^{+0.08}_{-0.07}\ \mathrm{pb}, \tag{7.1}$$

with a signal significance of Z = $2.47\ \sigma$ under a $\mathrm{H_0}$ hypothesis that we have agreement

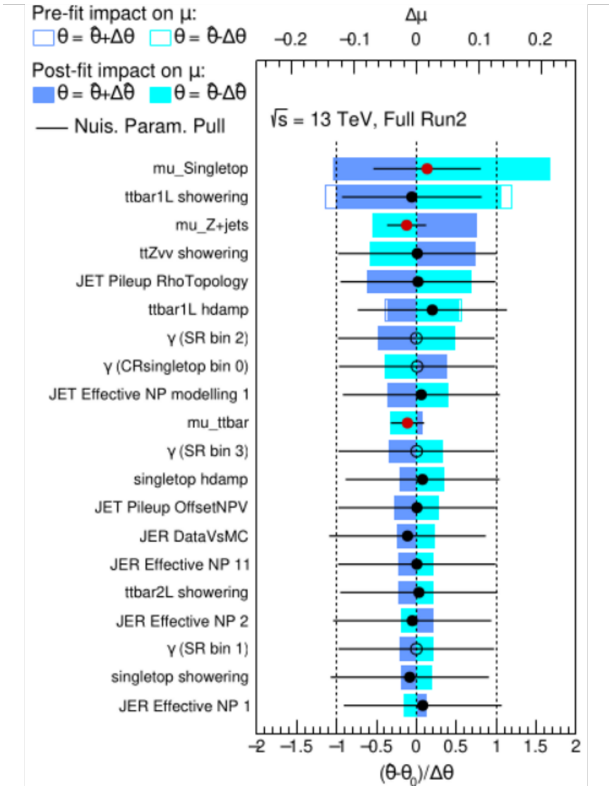


Figure 7.9: The ranking of the top-rated systematics impact on the scale factor for the cross section of $t\bar{t}Z \rightarrow \nu\bar{\nu}$, including the impacts of each of the control regions' scale factors. This is shown with the largest contributing parameter at the top. The dots for each parameter represent the direction in which the parameter is pulled during the fitting process. The color of the dot represents where the background is constrained or not. Normalization factors on specific backgrounds are not constrained. The dots with holes in them represent the gamma pulls of individual bins for the regions

with the standard model. This significance is not higher than the normally required $3\ \sigma$ for a measurement, but this result can be used in combination with 1L and 2L channels to achieve over $3\ \sigma$. This result can be compared with the theoretical value presented in section 5.2

$$\sigma^{Theory}_{t\bar{t}Z\rightarrow\nu\bar{\nu}} = 0.173^{+0.052}_{-0.052}\ \text{pb}. \tag{7.2}$$

The measured cross section in equation 7.1 shows agreement with the standard model prediction shown in equation 7.2. This means there is no significant disagreement with the standard model in this 0L measurement, but it is the first measurement of the $t\bar{t}Z \rightarrow \nu\bar{\nu}$ cross section in the 0L channel. The combined fit may also find more confirmation of the standard model, as the 0L did.

7.4 Fitting strategy combined channels

The combined fit was run across the three lepton channels, discussed in 6. With each channel having its own set of signal, control and validation regions, defined in 6.7.

The combined 0L signal regions are also split into SR1 and SR2. SR1 was binned in 7 bins in the 0L DNN output Score and SR2 was binned in 5 bins in the 0L DNN output Score. The 0L control region for $t\bar{t}$ and Z+jets are binned in S in 4 bins. The single-top control region for the 0L is binned in $\Delta R(b1, b2)$ in 5 bins.

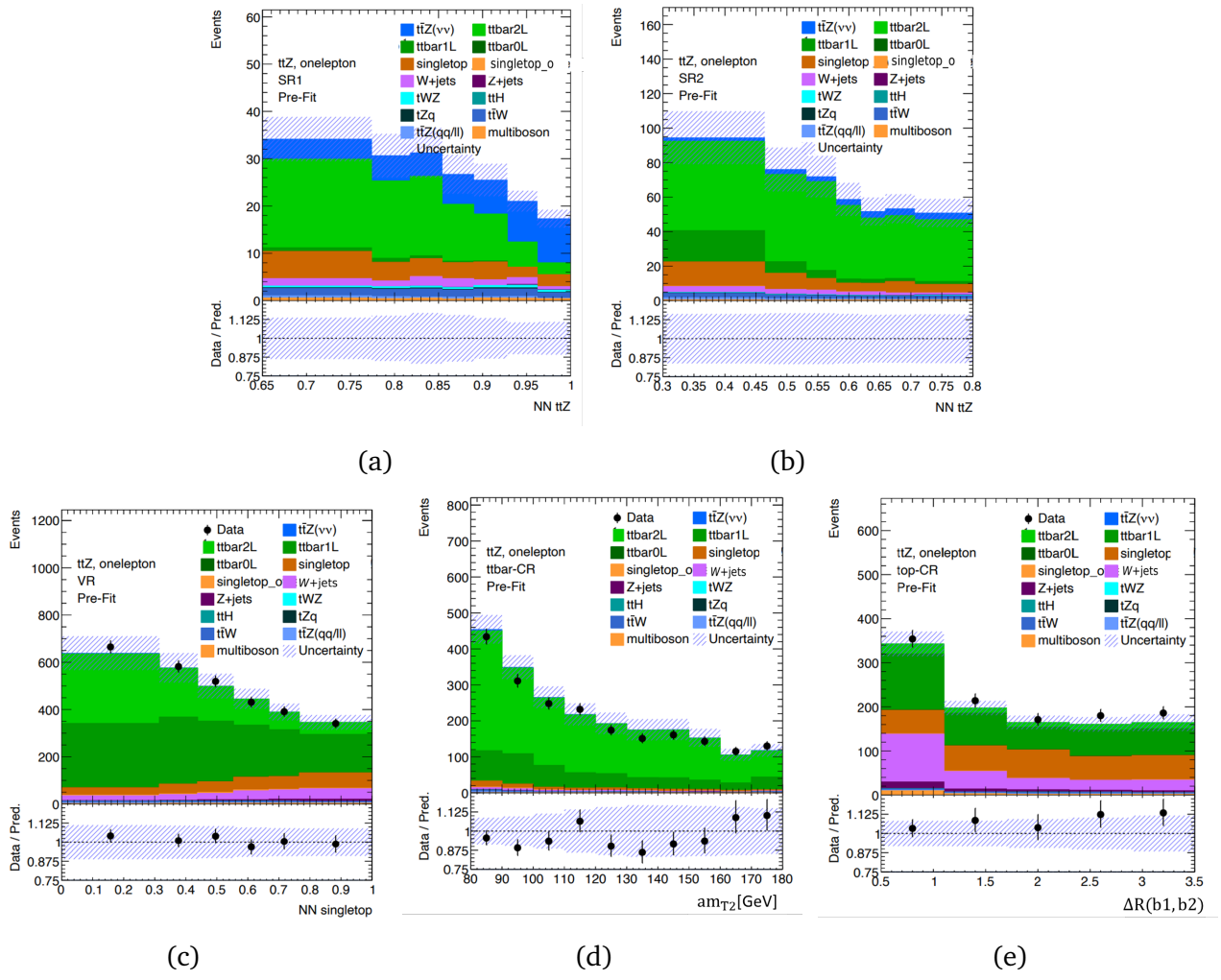


Figure 7.10: Histograms plotted for the 1L channel: (a) the blinded signal region SR1, (b) the blinded signal region SR2, (c) the validation region, (d) $t\bar{t}$ control region and (e) the single-top control region. This is created in TrexFitter. These are post-fit plots produced after the fit is applied. The histograms are stacked so the vertical widths represent the number of events in each bin. The subplot at the bottom of each plot shows the MC agreement with data.

The combined 1L signal regions are split into SR1 and SR2, these are then binned in the 1L DNN $t\bar{t}Z$ score, with SR1 binned in 7 bins between 0.65 and 1, while SR2 is binned between 0 and 0.65. The 1L control region for $t\bar{t}$ is binned in the asymmetric transverse mass $\mathrm{am_{T2}}$, which is the stanverse mass calculated by grouping together the one lepton to each of the two b-jets into objects v_1 and v_2, these two objects are then used in the m_{T2} calculation. The variable was split into 10 bins from 80 to 180 GeV. The 1L control region for single-top was binned in $\Delta\mathrm{R(b1, b2)}$ from 0.5 to 3.5 in 5 bins. The 1L validation region was binned in the DNN single-top score from 0 to 1 in 6 bins. The pre-fit plots for this region are shown in figure 7.11

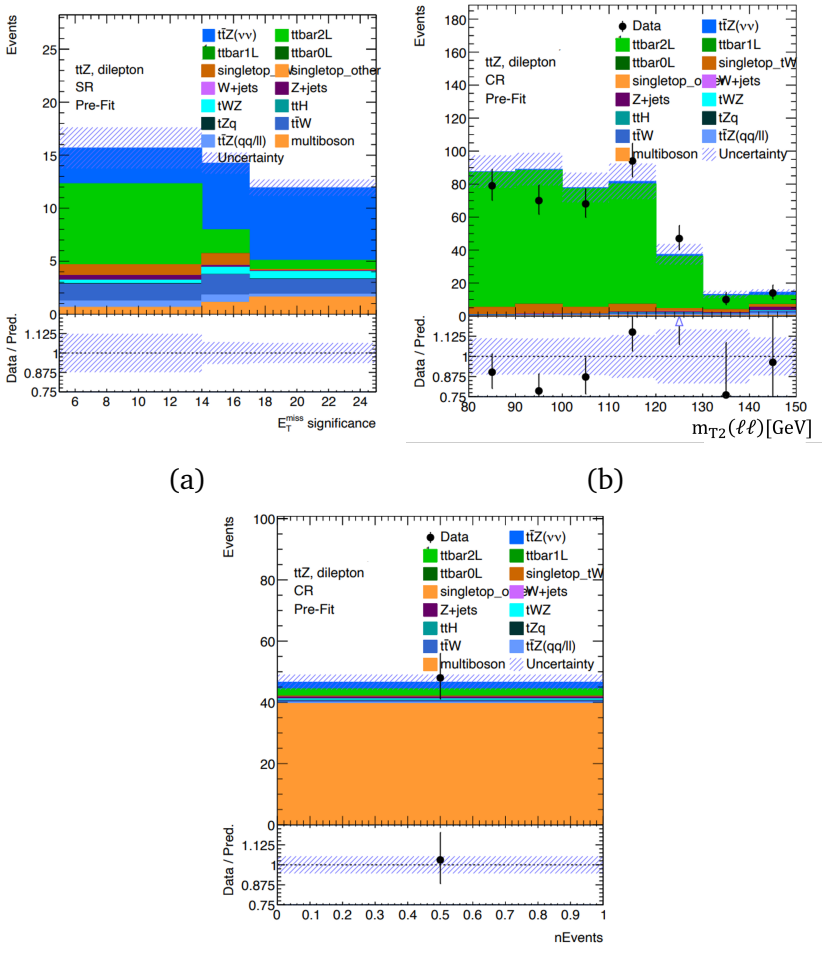


Figure 7.11: Histograms plotted for the 2L channel: (a) the blinded signal region, (b) the $\mathrm{t\bar{t}}$ control region, (c) the multi-boson control region. This is created in TrexFitter. These are post-fit plots produced after the fit is applied. The histograms are stacked so the vertical widths represent the number of events in each bin. The subplot at the bottom of each plot shows the MC agreement with data.

The combined 2L signal region is binned in E_T^{miss} significance in 3 bins between 5 to 25 in S. The 2L control region for $t\bar{t}$ is binned in $m_{T2}(\ell\ell)$ between 80 to 150 GeV in 7 bins. The 2L control region for multi-boson is binned in a single bin, measuring the number of events in the region.

7.5 Results of the combined fit

The results of the combined fit are currently blinded. The blinded fits here show the expected significance that the fit is estimated to achieve.

The scale factors for the combined fit for all three channels are shown in figure 7.12. The expected signal significance from the combined fit is 4.9 σ.

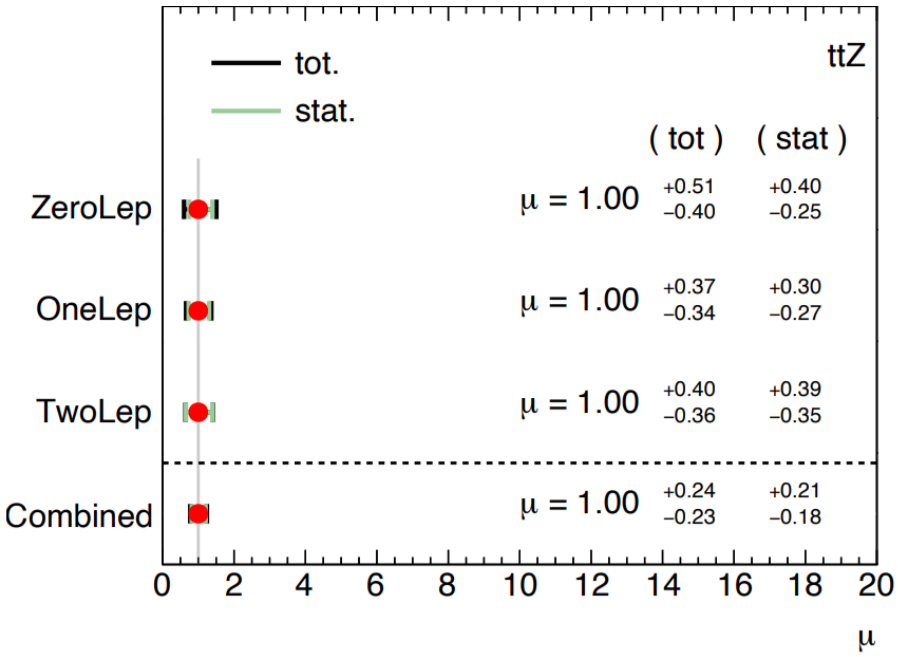


Figure 7.12: A scale factor plot showing the scale factors applied to each control region and to the cross section of $t\bar{t}Z \to \nu\bar{\nu}$ in the signal region, for the combined channels fit, by TrexFitter.

As can be seen from figure 7.12, the scale factor for $t\bar{t}Z \to \nu\bar{\nu}$ is mainly statistically dominated, which was expected from the low statistics in each of the channels. The 1L channel had the smallest contribution to the uncertainty of the combined scale-factor, while the 0L gave the largest contribution. The higher accuracy of the combined scale-factors on the $t\bar{t}Z \to \nu\bar{\nu}$ cross section, if its unblinded value is the same as in equation 7.1, could suggest the SM prediction is too large when compared with experimental data. If this is true, then the SM cross section for $t\bar{t}Z \to \nu\bar{\nu}$ would need to be reduced in simulations and could po-

tentially reduce the background to SUSY and dark matter searches. This high significance $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$ will also be very useful to EFT interpretations at high p_T^Z.

Part III

Topological towers of calorimeter clusters

Part 3 covers the technical work on topological towers and their use in improving the $\mathrm{p_T}$ correction of forward jets in the ATLAS calorimeter. This will be covered in two chapters. Chapter 8 covers the core concepts of what topological towers are, why they are needed and how they are implemented. Chapter 9 shows the results of implementing these towers.

Chapter 8

Calorimeter energy reconstruction in the forward region

This chapter aims to cover the methodology of how calorimeter energy deposited in cells clustered into topological towers is used in improving the transverse momentum density readout in the forward region of the ATLAS Calorimeter. Section 8.1 covers the theory and methodology of topological clusters. In section 8.3, we cover the replacement for the clusters, the topological towers. In section 8.2, we cover the failings of the current method in the central region as well as other attempts at fixing the issue with topological clusters. This will lead us onto section 8.4, which covers how we use topological towers and a sliding window algorithm to model p_T density as a function of η, we then model this density as a function of both the in-time and out-of-time pileup as well.

8.1 Topological clusters

As mentioned in section 2.2.4 the ATLAS detector uses calorimeters to measure the energy of various particles. The calorimeter cells are often clustered into groups of cells in order to boost the ratio of signal to noise as much as possible. One way in which the signal to noise is reduced is by clustering using topological clusters. These 3D topological clusters(topo-clusters) are topologically connected cells with the formation of the clusters driven by the cell signal significance($\varsigma_{\mathrm{cell}}^{\mathrm{EM}}$) defined as [71]

$$\varsigma_{cell}^{EM} = \frac{E_{cell}^{EM}}{\sigma_{noise,cell}^{EM}}, \tag{8.1}$$

where $\mathrm{E}_{\mathrm{cell}}^{\mathrm{EM}}$ and $\sigma_{\mathrm{noise,cell}}^{\mathrm{EM}}$ are the cell signal and average expected noise, both measured at the electromagnetic (EM) scale. The EM scale is the reconstructed energy deposited by particles, but without any corrections for the loss of signal for hadrons in the calorimeters' inactive material. The cell signal significance is used as a starting point to the grow clusters. The growing of cluster cells starts with a cell with a high cell signal significance. This is called the seeding phase. The cluster then grows from the seeded cell, adding more

cells into the cluster. This is the growing phase. The cluster will continue growing until it reaches cells for which ς_{cell}^{EM} is below the threshold to be added to the cluster. The seeding and growth stages are controlled by 3 main parameters. Seeding(S), growth(N) and a filter(P)[71]. The cell signal significance from equation 8.1 is then used to define the selection parameters

$$\varsigma_{cell}^{EM} > S, \tag{8.2}$$

$$\varsigma_{cell}^{EM} > N, \tag{8.3}$$

$$\varsigma_{cell}^{EM} > P. \tag{8.4}$$

With the rules of the algorithm defined, the topo-clusters are built from a combination of seed and growth steps. The steps are repeated until all cells satisfying equations 8.2 and 8.3 and their direct neighbors passing equation 8.4 are collected.

The seed and grow steps function as follows: A proto-cluster is formed from each seed cell (cells which have passed the S selection in equation 8.2) and its neighbouring cells which satisfy equation 8.3 and 8.4. The neighbouring cells which pass the 8.3 selection have their neighbours collected also. Proto-clusters merge if the neighbour is the seed cell of another proto-cluster. If one of the cells in the growing phase is attached to two different proto-clusters and passes equation 8.3, then the two proto-clusters merge. The procedure is then repeatedly applied to further and further neighbours until all the neighbours considered have met the threshold P but have not passed threshold N. At this point, the formation stops. The proto-cluster now consists of a core of highly significant cells surrounded by less significant cells. The next stage in the cluster formation collects any of these directly neighbouring less significant cells which pass 8.4 into the cluster.

For the 4-2-0 algorithm, which is used in the formation of topo-clusters, the thresholds are set to 4,2,0 for S,N and P, respectively. With P set to zero in this case, the entire set of cells that surround the proto-cluster, after the growing phase has finished, are added to the proto-cluster. Using this clustering 4-2-0 algorithm leads to significant improvements in various aspects of the calorimeter performance[71]. A 2D visual example of the 4-2-0 algorithm is shown in figures 8.1, 8.2 and 8.3 (note the algorithm is actually performed in 3D). Figure 8.1 shows the seeding stage with cells with $\varsigma_{cell}^{EM} \geq 4$ highlighted in red. Figure 8.2 shows the cells added to the cluster during the growing stage in blue that pass ς_{cell}^{EM} and are next to a red or a blue cell. In figure 8.3 we see the final stage of cluster building were we add any cells that are next to any red or blue cells and add them to the cluster.

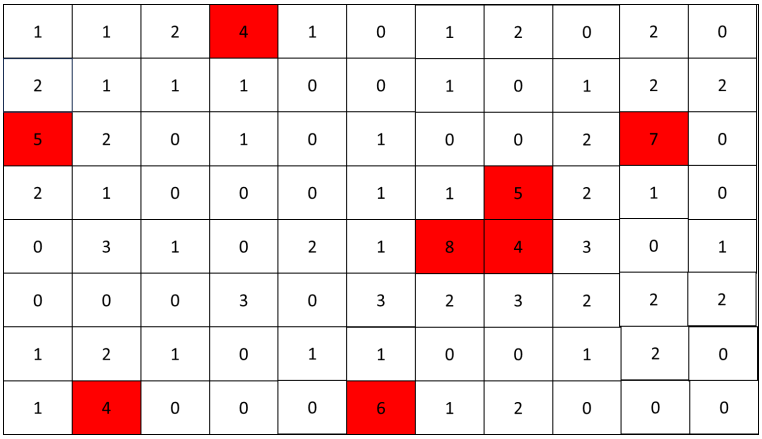


Figure 8.1: Diagram showing the seeding step of a 4-2-0 topo-cluster. The boxes represent cells. Numbers in the boxes show the cell signal significance. The boxes colored in red show the cells which have passed the 4-2-0 algorithm's seeding requirements shown in equation 8.2.

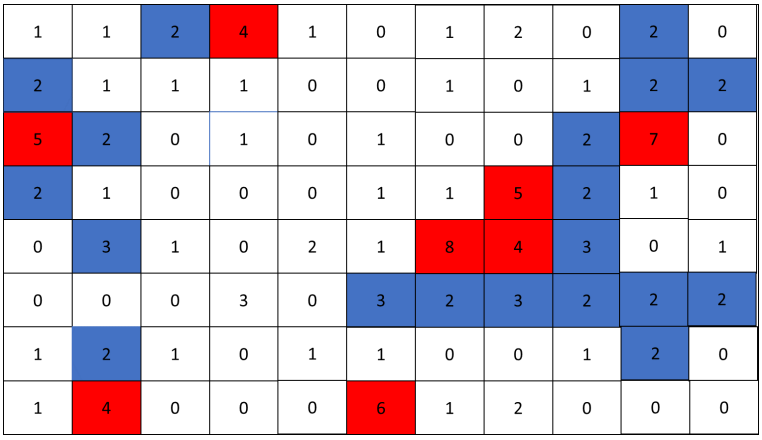


Figure 8.2: Diagram showing the growing step of a 4-2-0 topo-cluster. The boxes represent cells. Numbers in the boxes show the cell signal significance. The boxes colored in red show the cells which have passed the 4-2-0 algorithm's seeding requirements shown in equation 8.2. The boxes colored in blue show the cells which have passed the 4-2-0 algorithm for growth shown in equation 8.3.

These clusters are treated as massless pseudo-particles in physics object reconstruction. In theory, a cluster mass could be calculated from cell signals and their spatial distribution; however, due to issues with the noise thresholds and the shower spreading, the cluster mass calculated in this way would not be physically meaningful. Hadronic showers also tend to be captured by two or more clusters for single particles, and in most cases, the ATLAS proton-proton conditions produce crowded environments where the clusters will contain information from several particles. This increases the difficulty in relating any cluster mass measurement to a real particle. The assumption that the cluster is a massless

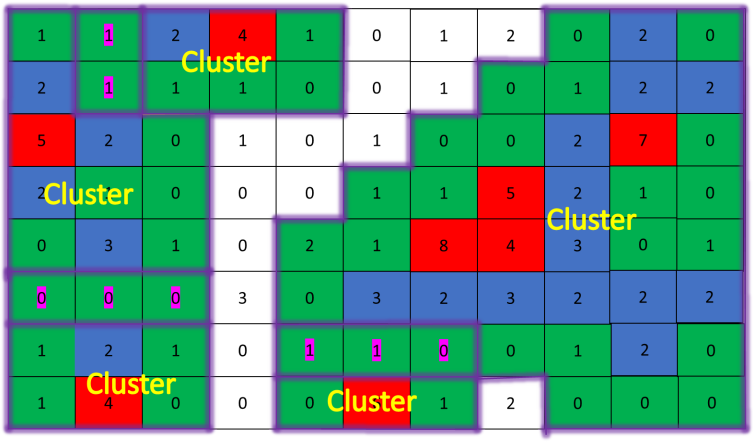


Figure 8.3: Diagram showing the filter step of a 4-2-0 topo-cluster. The boxes represent cells. Numbers in the boxes show the cell signal significance. The boxes colored in red show the cells which have passed the 4-2-0 algorithm's seeding requirements shown in equation 8.2. The boxes colored in blue show the cells which have passed the 4-2-0 algorithm for growth shown in equation 8.3. The boxes colored in green show the cells which have passed the 4-2-0 algorithm for filter step shown in equation 8.4. The purple lines define boundaries for the clusters. The cells with numbers highlighted in pink show cells which overlap between two clusters and need to have the cell shared.

pseudo-particle is, therefore, used.

Proto-clusters built using the 4-2-0 algorithm can be too large to provide good measurements of energy flow, therefore, a cluster splitting algorithm is used to split proto-clusters into smaller clusters. The cluster-splitting algorithm locate cells which are neighbors to two or more signal maxima, such as the two signal maxima in the top left-hand corner of figure 8.3. These cells, highlighted in purple, are then assigned to the two highest energy clusters after the splitting of the original cluster it was associated with. This assures that a cell is never shared with more than two topo clusters at a time. The sharing of these cells between the two cluster energies $\mathrm{E}^{\mathrm{EM}}_{\mathrm{clus},1}$ and $\mathrm{E}^{\mathrm{EM}}_{\mathrm{clus},2}$ is measured by the geometrical weights $\mathrm{w}^{\mathrm{geo}}_{\mathrm{cell},1}$ $\mathrm{w}^{\mathrm{geo}}_{\mathrm{cell},2}$ defined as

$$w^{geo}_{cell,1} = \frac{E^{EM}_{clus,1}}{E^{EM}_{clus,1} + rE^{EM}_{clus,2}}, \tag{8.5}$$

$$w^{geo}_{cell,2} = 1 - w^{geo}_{cell,1}, \tag{8.6}$$

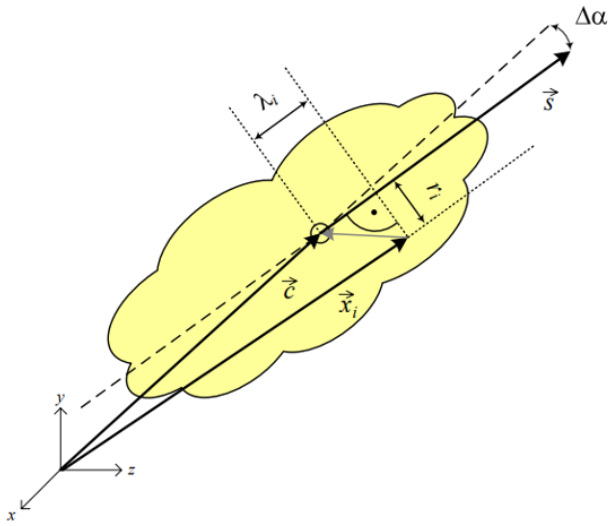
where $\mathrm{r} = \mathrm{e}^{\mathrm{d}_1 - \mathrm{d}_2}$, and d_1 and d_2 are the distances between the cell and the centre of gravity of the 2 clusters.

8.1.1 Toplogical cluster moments

So far, we have summarized how to form a cluster. We will now cover a range of reconstructed cluster observables known as cluster moments. Figure 8.4 shows a diagram and summary of the various geometrical moments for topo-clusters. Cluster moments can generally be defined as a weighted sum of its cell variables v_{cell} [71]

$$\langle v^n_{cell} \rangle = \frac{\sum_{(i|E^{EM}_{cell,i}>0)} w^{geo}_{cell,i} E^{EM}_{cell} v^n_{cell,i}}{\sum_{(i|E^{EM}_{cell,i}>0)} w^{geo}_{cell,i} E^{EM}_{cell}}, \tag{8.7}$$

where $\langle v^n_{cell} \rangle$ is the topo-cluster moment and v_{cell} is the cell variable for the type of moment you are calculating. The exponent n is the order of moments we are calculating; typically, this is set to 1 or 2, as higher-level moments are not used.



$\vec{c}$ *centre of gravity of cluster, measured from the nominal vertex* $(x = 0, y = 0, z = 0)$ *in ATLAS*

$\vec{x}_i$ *geometrical centre of a calorimeter cell in the cluster, measured from the nominal detector centre of ATLAS*

$\vec{s}$ *particle direction of flight (shower axis)*

$\Delta\alpha$ *angular distance* $\Delta\alpha = \angle(\vec{c}, \vec{s})$ *between cluster centre of gravity and shower axis* $\vec{s}$

λ_i *distance of cell at* $\vec{x}_i$ *from the cluster centre of gravity measured along shower axis* $\vec{s}$ $(\lambda_i < 0$ *is possible)*

r_i *radial (shortest) distance of cell at* $\vec{x}_i$ *from shower axis* $\vec{s}$ $(r_i \geq 0)$

Figure 8.4: Diagram of a topo-cluster's Geometrical moments [71].

The location of these clusters is defined by their centre of gravity in 3-dimensional space $\overrightarrow{c}$, which is shown in figure 8.4. $\overrightarrow{c}$ is calculated using the Cartesian coordinates first moments, which specify the calormeter cell centers. The locations are provided in the frame of reference of the interaction point(IP) defined in chapter 2.

The direction of the topo cluster is defined as a signal-weighted barycentre $(\eta_{\text{clus}}, \phi_{\text{clus}})$. Only the absolute value for $\text{E}^{\text{EM}}_{\text{cell}}$ is taken when calculating the direction. Using positive or negative $\text{E}^{\text{EM}}_{\text{cell}}$ leads to distortion or bias in the direction. Equation 8.7 is used to define the

two moments η_{cluster} and ϕ_{cluster} where the variable v_{cell} is η_{cell} and ϕ_{cell}, respectively [71].

$$\eta_{clus} = \frac{\sum_{i=1}^{N_{cell}} w_{cell,i}^{geo} \cdot |E_{cell}^{EM}| \cdot \eta_{cell,i}}{\sum_{i=1}^{N_{cell}} w_{cell,i}^{geo} \cdot |E_{cell}^{EM}|}, \tag{8.8}$$

$$\phi_{clus} = \frac{\sum_{i=1}^{N_{cell}} w_{cell,i}^{geo} \cdot |E_{cell}^{EM}| \cdot \phi_{cell,i}}{\sum_{i=1}^{N_{cell}} w_{cell,i}^{geo} \cdot |E_{cell}^{EM}|}. \tag{8.9}$$

Where the first directional moments are calculated with $n = 1$. These directions and their moments are defined in reference to the IP.

The size of a topo-cluster is represented with respect to the shower axis (the direction to which the jet is traveling through the calorimeter shown in figure 8.4) $\overrightarrow{s}$ and center of gravity $\overrightarrow{c}$ shown in figure 8.4. Therefore, a cell's distance from the axis, within the cluster, can be represented as [71]

$$r_i = |(\overrightarrow{x_i} - \overrightarrow{c}) \times \overrightarrow{s}|, \tag{8.10}$$

$$\lambda_i = (\overrightarrow{x_i} - \overrightarrow{c}) \cdot \overrightarrow{s}, \tag{8.11}$$

where $\overrightarrow{x}$ is the cell's geometrical center position, $\mathrm{r_i}$ and λ_i are the radial distance to the shower axis and the longitudinal distance from the shower center of gravity. These variables are then fed into 8.7 as the $\mathrm{v_{cell,i}}$ component with n=2 to derive the $\langle \mathrm{r}^2 \rangle$ and $\langle \lambda^2 \rangle$ variables.

Equation 8.7 is calculated using two different values for $\mathrm{r_i}$ and λ_i coordinates for its two most energetic cells depending on whether we are calculating $\langle \mathrm{r}^2 \rangle_{\mathrm{out}}$ or $\langle \mathrm{r}^2 \rangle_{\mathrm{core}}$, and $\langle \lambda^2 \rangle_{\mathrm{out}}$ or $\langle \lambda^2 \rangle_{\mathrm{core}}$. We set $v_{cell,i} = r_i$ and $v_{cell,i} = \lambda_i$ for all cells.

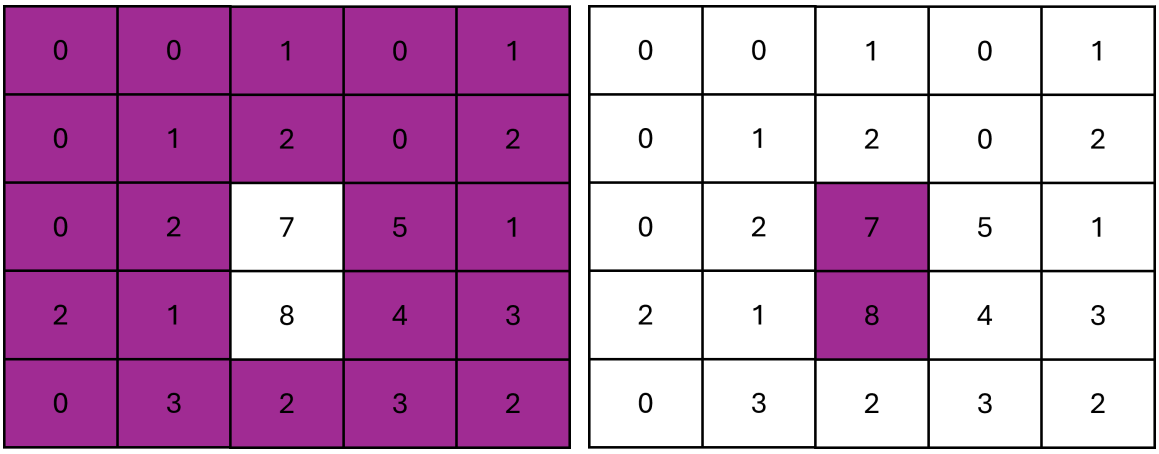
When calculating the out variables ($\langle \mathrm{r}^2 \rangle_{\mathrm{out}}$ and $\langle \lambda^2 \rangle_{\mathrm{out}}$), we set the two most energetic cells $\lambda_i = 0$ and $r_i = 0$ respectively. We then use the rest of the non-zero value cells in the calculation of $\langle \lambda^2 \rangle_{out}$ and $\langle r^2 \rangle_{out}$. The cells we use for this calculation are shown in figure 8.5.(a).

For $\langle \mathrm{r}^2 \rangle_{\mathrm{core}}$ and $\langle \lambda^2 \rangle_{\mathrm{core}}$ the two most energetic cells are set to $\mathrm{r_i} = \mathrm{r_{core}}$ and $\lambda_\mathrm{i} = \lambda_{\mathrm{core}}$ while all other cells have $r_i = 0$ and $\lambda_i = 0$. We then use the two central cells in calculating $\langle \mathrm{r}^2 \rangle_{\mathrm{core}}$ and $\langle \lambda^2 \rangle_{\mathrm{core}}$ as shown in figure 8.5.(b).

These core and out variables are then used like so[71]

$$m_{long}^2 = \frac{\langle \lambda^2 \rangle_{out}}{\langle \lambda^2 \rangle_{out} + \langle \lambda^2 \rangle_{core}}, \tag{8.12}$$

$$m_{lat}^2 = \frac{\langle r^2 \rangle_{out}}{\langle r^2 \rangle_{out} + \langle r^2 \rangle_{core}}, \tag{8.13}$$



(a) (b)

Figure 8.5: A 2D visual demonstration of the cells used in calculating $\langle \mathrm{r}^2 \rangle_{\mathrm{out}}$ and $\langle \lambda^2 \rangle_{\mathrm{out}}$ in (a). In (b) we show the cells used for the calculation of $\langle \mathrm{r}^2 \rangle_{\mathrm{core}}$ and $\langle \lambda^2 \rangle_{\mathrm{core}}$. The purple represents a cell that is turned on and has a non-zero value, while white represents cells that are turned off and have a zero value.

where $\mathrm{m}_{\mathrm{long}}^2$ and $\mathrm{m}_{\mathrm{lat}}^2$ are the normalized longitudinal and lateral energy dispersion. $\mathrm{m}_{\mathrm{long}}^2$ and $\mathrm{m}_{\mathrm{lat}}^2$ are effectively measurements of the cluster length and width, respectively. While they may not directly measure the spatial cluster dimensions, they provide a measure of the energy dispersion of cells within the cluster. Larger values of $\mathrm{m}_{\mathrm{long}}^2$ and $\mathrm{m}_{\mathrm{lat}}^2$ indicate a long and wide topo-cluster while lower values of $\mathrm{m}_{\mathrm{long}}^2$ and $\mathrm{m}_{\mathrm{lat}}^2$ indicate a short and narrow topo-cluster. The effective size of the topo-cluster in (η, ϕ) space is approximated as

$$\sigma_\eta \approx \sigma_\phi \approx \left(\frac{\sqrt{\langle r^2 \rangle}}{|\overrightarrow{C}|}\right) \times \cosh(\eta), \tag{8.14}$$

where σ_η and σ_ϕ are the full width at half maximum of the energy-weighted η and ϕ distributions of positive energy cells within the cluster. This approximation can be made due to the unique granularity of the ATLAS calorimeters [71].

8.1.2 Topological cluster weighting

The clusters can be set to either be weighted at the EM scale or the local cluster weighting (LCW) scale. LCW seeks to improve the cluster energy resolution by taking more than the cluster signal into account when applying the weighting, by also considering the non-compensating calorimeter response, signal losses due to clustering and signal losses due to energy lost in inactive material(see section 2.2.4 for more detail). Figure 8.6 shows an overview of these corrections [71].

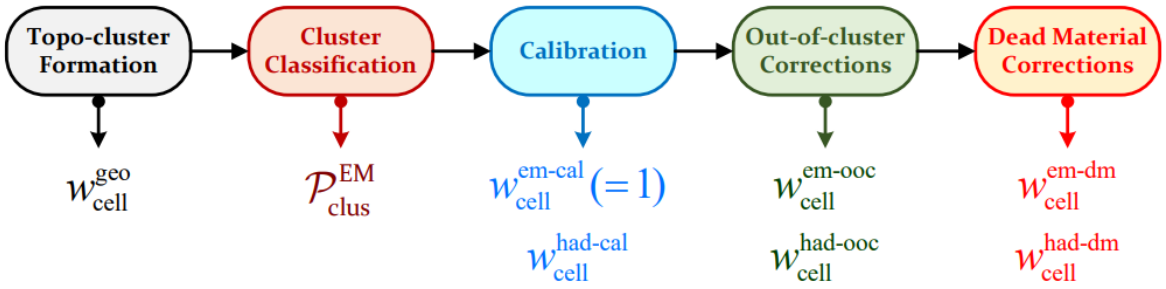


Figure 8.6: LCW calibration scheme for topo-clusters. It starts with the formation of a topo-cluster along with its geometrical signal weight $\mathrm{w}_{\text{cell}}^{\text{geo}}$. Following the formation of a topo-cluster, the cluster is assigned a probability $\mathcal{P}_{clus}^{EM}$ that it was created by electromagnetic energy deposit. Next, weights are produced for the two possible cases hadronic and electromagnetic, for calibration, out-of-cluster and dead material corrections. [71].

The LCW attempts to apply local corrections so that, on average, the energy deposited gives the same signal regardless of the type of particle that caused it. Applying a calibration has a myriad of benefits such as improving the resolution of jets that use both hadronic and electromagnetic signals and improving the construction of variables like $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ which require knowledge of the entire event to reconstruct including responses from electromagnetic and hadronic parts of the calorimeter which need to be combined[71].

Signal loss due to clustering is the intrinsic noise suppression that happens during the topo-cluster formation, which loses signal due to cells not being added to the cluster. This can be seen in figure 8.3 where the clusters have left out cells with a non-zero cell significance. The amount of signal lost depends on the noise thresholds, which are determined by the pile-up conditions. This signal loss can be significant and therefore needs to be adjusted by applying out-of-cluster corrections, which is a part of the LCW corrections[71].

A correction is needed for signal losses due to energy lost in dead material. This comes from energy lost to nearby inactive material in front, between, and inside the calorimeter

modules. The LCW weights adjust for this [71].

There is also the cluster classification step in figure 8.6 which assigns a probability $\mathcal{P}_{clus}^{EM}$ that the cluster was created by an electromagnetic energy deposit over a hadronic contribution $(1 - \mathcal{P}_{clus}^{EM})$. This is used to assign the specific correction factors based on whether the cluster was formed from a hadronic or electromagnetic source.

8.2 The problem in the forward region

Jets are reconstructed in the ATLAS detector as discussed in section 3.1.3, using particle-flow[1] (Pflow-jets) objects or using topo-clusters (topo-jets). These jets need to have their transverse momentum adjusted to account for pile-up. A good measure of these pile-up contributions is the $\mathrm{p_T}$ density, also known as ρ[73]. The ρ of an event can be estimated by the median of $\mathrm{k_T}$ jets' ρ distribution. These $\mathrm{k_T}$ jets are constructed with no minimum $\mathrm{p_T}$ threshold. This can be expressed as

$$\rho_{event} = \mathrm{median}\left(\frac{p_{T,i}^{jet}}{A_i^{jet}}\right), \tag{8.15}$$

where each $\mathrm{k_T}$ jet i has transverse momentum $\mathrm{p_{T,i}^{jet}}$ and area $\mathrm{A_i^{jet}}$[73]. The jet area $\mathrm{A_i^{jet}}$ was defined in section 3.1.3

The inputs into these $\mathrm{k_T}$ jets are topo-clusters(disscused in section 8.1) that are within $|\eta| \leq 2$. The clusters are kept within $|\eta| \leq 2$ as this region has high and relatively simple calorimeter occupancy. ρ_{event} can then be used to correct jets as

$$p_{T,i}^{corr} = p_{T,i}^{jet} - \rho_{event} \times A_i^{jet}. \tag{8.16}$$

This method for correcting jets in the central region works very well due to topo-cluster inputs only being taken from an area where ρ is large, and relatively flatly distributed. This is shown in figure 8.7. Equation 8.16 works well for jets within the central region, while jets that don't fall within the central region tend to be overcorrected by equation 8.16. The overcorrection comes from the ρ intrinsically decreasing in the forward region due to low occupancy of clusters, and to the calorimeter segmentation and response. For $|\eta| > 2.5$, the calorimeter cell size increases, which causes issues as topo-clusters are seeded (see equation 8.2 and 8.1) based on relative signal against noise rather than absolute values. Therefore, as cells increase in size and decrease in population, it decreases the probability that a cell fluctuates up due to pile-up or electronic noise. Fewer cells are therefore seeded

[1]Reconstructs jets based on information from both the ATLAS tracking and calorimeter, splitting the reconstruction into charged and neutral objects, with charged hadrons gaining improved performance over using separate track and cluster reconstruction, for further reading see [72]

due to pileup, leading to a drop in ρ. This is further made worse at higher η, as the probability of hadronic showers overlapping increases, making it even less likely for a cluster to be seeded by pile-up.

The intrinsically lower ρ in the forward region makes correcting the jets with the centrally calculated ρ_{event}, given in equation 8.15, lead to a larger and inaccurate correction. This will then lead to a smaller and incorrect value for $\mathrm{p_T^{corr}}$, which can be fixed with residual corrections. However, the systematics of these corrections increase with their magnitude; therefore, ρ as a function of η is preferred to correct jets in the forward region.

An attempt was made to model ρ as a function of η using topo-clusters. The results are shown in figure 8.7[73]. The distribution shown in figure 8.7 is created by using a sliding window that calculates the median ρ in an η window that steps across the full η range of the detector placing the median value of the transverse momentum density $\langle\rho\rangle$, calculated within the window, in the center of the window in η. (The sliding window algorithm will be discussed in more detail in the next section.) The value of $\langle\rho\rangle$ is calculated for various average numbers of proton-proton interactions per bunch crossing (μ).

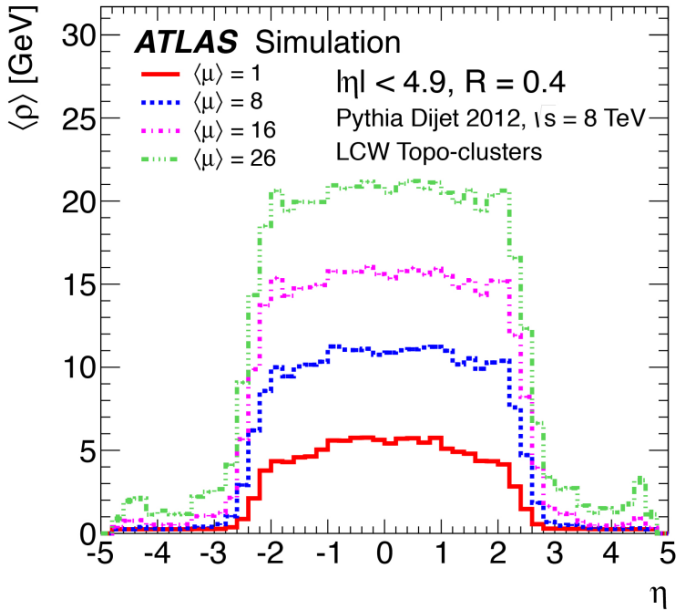


Figure 8.7: The median estimated $\rho(\eta)$, in simulated PYTHIA8 dijet events [73]

We can also see from figure 8.7 that $\rho(\eta)$ drops dramatically beyond $|\eta| > 2$ with $\rho(\eta)$ heading to near 0 at all values of $\langle\mu\rangle$ past $|\eta| > 2.5$. The drop-off shown towards 2.5 $|\eta|$ is too severe to be a description of the pile-up and would effectively underestimate the pile-up. The method shown in equation 8.16 was kept until a better method could be found.

The cause of the larger drop in measured median $\rho(\eta)$ was due to the cluster area being both poorly defined and its size increasing towards larger η. The clusters in the forward region have a much larger area than the clusters in the central region. This larger area dominates the ρ ($\mathrm{p_T}$ density) calculation in the forward region. Ideally, there would be a way to keep the noise suppression technique of the clusters while also reducing and better defining their areas in the forward region. This is where forward topo-towers become relevant.

8.3 Topological towers

Topo towers are another form of clustering that takes advantage of the noise suppression of topo-clusters. Topo towers are built from the topo-cluster signals with much smaller areas at 0.1×0.1 in (η, ϕ) space. These cluster signals can be represented as one massless four-momentum with $(\mathrm{E_{clus}}, \mathrm{p_{clus}})$ with $\mathrm{E_{clus}} = |\mathrm{p_{clus}}|$. The topo towers distribute this cluster four-momentum into many four-momenta populating the (η, ϕ) around the cluster direction $(\eta_{\mathrm{clus}}, \phi_{\mathrm{clus}})$. The number of towers that are built from a cluster is dependent on both the extension of the cluster in the (η, ϕ) plane(catchment area) and the possible catchment area overlap by clusters staggered along $\overrightarrow{s}$. The topo-tower's kinematics are available at both EM and LCW scales, in the same way as topo-clusters are. The kinematics at both scales follow the four-vector notation $(\mathrm{E}, \eta, \phi, \mathrm{m})$

$$p_{tower}^{EM} = (E_{tower}^{EM}, \eta_{tower}^{EM}, \phi_{tower}^{EM}, 0), \tag{8.17}$$

$$p_{tower}^{LCW} = (E_{tower}^{LCW}, \eta_{tower}^{LCW}, \phi_{tower}^{LCW}, 0). \tag{8.18}$$

The tower reconstructs the momentum from the energy collected by the tower with the assumption that the tower acts like a massless pseudo-particle similar to topo-clusters, hence the zero component for the mass (m) in the four-vector.

The tower direction is also not the nominal direction associated with the tower grid in (η, ϕ) space. The direction in which the tower faces is instead reconstructed from the direction of the cells and their energies, with the energies weighted by the area overlap ratio between the tower and contributing cells. The direction reconstruction of the towers follows the same rules as the topo-clusters previously discussed. The direction reconstruction of a tower at a grid location $(\eta_{\mathrm{grid}}, \phi_{\mathrm{grid}})$ is $(\eta_{\mathrm{tower}}, \phi_{\mathrm{tower}})$ with the following constraints

$$\eta_{grid} - \frac{\Delta\eta}{2} \leq \eta_{tower} \leq \eta_{grid} + \frac{\Delta\eta}{2}, \tag{8.19}$$

$$\phi_{grid} - \frac{\Delta\phi}{2} \leq \phi_{tower} \leq \phi_{grid} + \frac{\Delta\phi}{2}, \tag{8.20}$$

where $\Delta\eta$ $\Delta\phi$ are the size of the tower in (η, ϕ) space. A 2D representation of these towers is show in figure 8.8. These constraints can be violated in the case of cells bordering the edge of the tower where the cell has its η or ϕ coordinate of its barycenter outside of the tower, but have an area overlap. These cells are kept as they contribute towards the tower's signal. At EM scale, the cells that contribute to the tower's signal are weighted based on area overlap with the tower. At LCW scale, the weight of each cell's contribution to the tower is the product of the area overlap between the cell and the tower and the LCW calibration factor applied to each cell.

The topo towers described here are forward topo towers and are only applied in the forward region with the requirement that $|\eta_{\text{clus}}| > 2.5$.

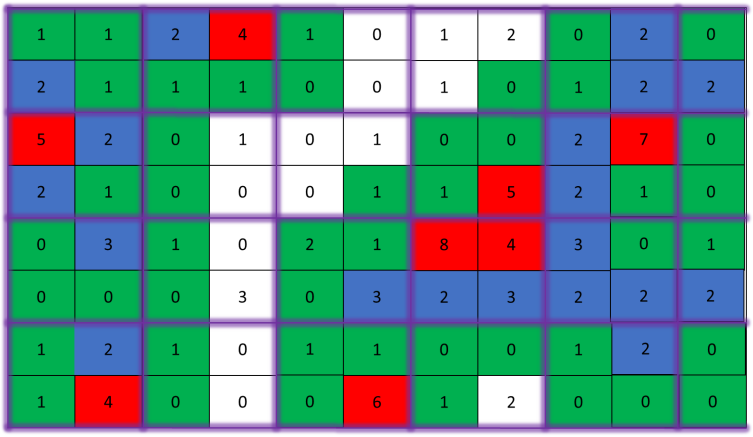


Figure 8.8: Diagram showing the filter step of a 4-2-0 topo-cluster. The boxes represent cells. Numbers in the boxes show the cell signal significance. The boxes colored in red show the cells which have passed the 4-2-0 algorithm's seeding requirements shown in equation 8.2. The boxes colored in blue show the cells which have passed the 4-2-0 algorithm for growth shown in equation 8.3. The boxes colored in green show the cells which have passed the 4-2-0 algorithm for filter step shown in equation 8.4. The purple lines define boundaries for the topological towers which split up the cluster signals.

8.4 Implementation

The problem with topo-clusters in the forward region can be solved by replacing them with topo-towers, as their area remains small and constant over the entire η range. Clusters still function well in the central region $(|\eta| < 2.5)$ as shown in figure 8.7 in the previous section; therefore, we are using a mix of clusters and towers, while the forward towers are being

used exclusively in the forward region ($|\eta| > 2.5$).

We use the sliding window algorithm mentioned in the previous section, this sliding window algorithm takes the median value of ρ found within a window of $\Delta\eta_{\text{window}} = 0.8$ and places the value it calculates at the centre of the window. The window is stepped across the entire η plane in small steps of $\Delta\eta_{step} = 0.1$. When the window's edge gets to the edge of the detector in η the windows are too large to get the centre any closer than $\eta = |4.6|$. The window, therefore, has to shrink with each step towards the boundary. The final bin will be at $\eta = |4.9|$ as the window cannot shrink any further. This sliding window algorithm is illustrated in figure 8.9.

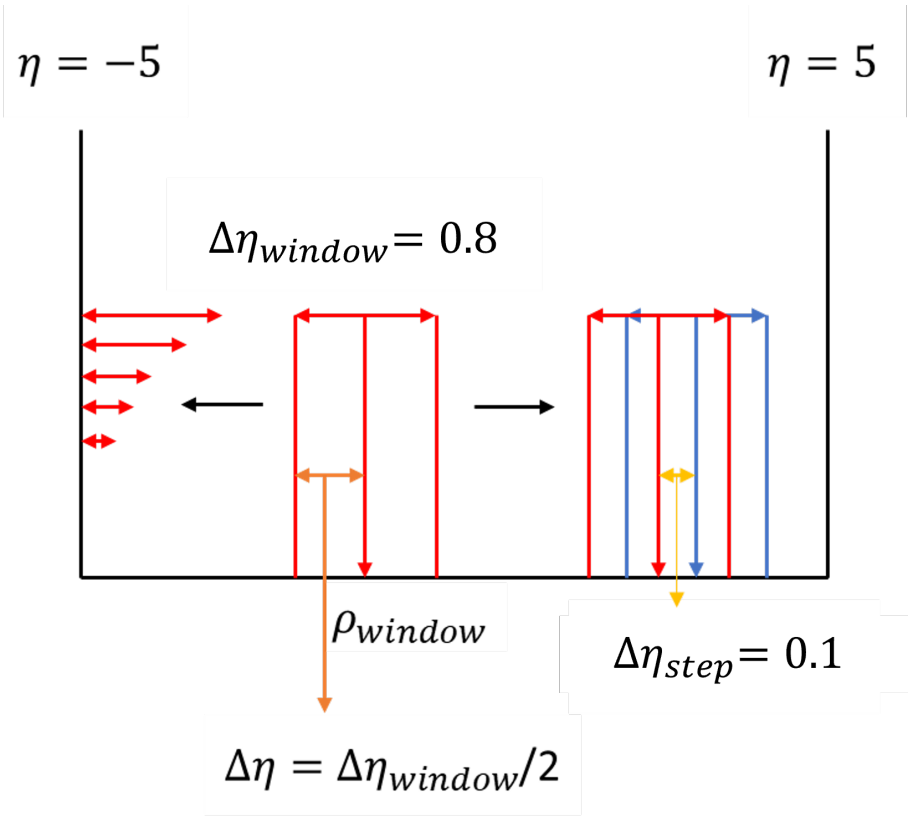


Figure 8.9: A diagram of the sliding window algorithm. The sliding window is $\Delta\eta = 0.8$ across. The sliding window is depicted as placing the ρ_{window} calculated within the window at the center of the window. The window is then depicted stepping in small steps of $\Delta\eta_{step} = 0.1$. The steps are taken from $\eta = -5$ to $\eta = 5$. At each step, the window recalculates its ρ_{window}, placing this value at the new center of the window. On the left side of the diagram, we see what happens when the window reaches an edge case. Upon reaching either limit of $\eta = -5$ or $\eta = 5$, the window shrinks with each step as it approaches the limit.

The ρ that the window averages over can be quite complex as in some windows we are

mixing two different types of objects: clusters and towers. The ρ for each tower can be calculated at both EM and LCW scale as

$$\rho^{EM|LCW} = \frac{p_T^{EM|LCW}}{\Delta\eta \times \Delta\phi}, \tag{8.21}$$

where the $\Delta\eta \times \Delta\phi$ is the area of the tower. The methodology to evaluate the topo-cluster ρ follows a similar methodology to the towers by using all the $\mathrm{p_T}$ collected in the cluster divided by the cluster area. The topo-cluster area is not well defined, and there are several ways in which the area for the topo-clusters can be represented. There is the elliptical cluster area defined as

$$A_{clus}^{ellipse} = \pi\sigma_\eta\sigma_\phi. \tag{8.22}$$

where σ_η and σ_ϕ are defined in equation 8.14. A different definition that can be used for clusters is the linear cluster area defined as:

$$A_{clus}^{linear} = 2\sigma_\eta\sigma_\phi. \tag{8.23}$$

Equations 8.22 and 8.23 both define areas that would be suitable for a ρ calculation; however, these definitions do not take into account the fact that clusters can overlap longitudinally. The method we used for the area of the clusters comes from the constituent subtraction (CS) algorithm [74].

We use a method that takes the constituent jets that are associated with topo-clusters and inclusive jets as the source of the constituent jets. We calculate the area of these constituents using a Fastjet Algorithm[75]. This algorithm models the area of the jet constituents using a Voronoi area[2] in the (η, ϕ) plane. The Voronoi areas (example shown in figure 8.10) of the constituents are calculated by considering the intersection between the Voronoi cell of each particle and a circle of radius R centred on the particle's position in the (η, ϕ) plane. The particle in this case is the jet constituent. This method of defining the jet areas avoids longitudinal overlap. Defining the areas in this way can lead to over-estimated areas in regions of the calomieter where the extension of the cluster is large. Since we only use the clusters in the central region, this is not an issue. The ρ for clusters is therefore just transverse momentum at the EM or LCW scale divided by the area of the clusters calculated using a Voronoi area of the constituent jet associated with that cluster.

The sliding window median ρ calculated from moving the window across both towers and clusters in the full η range is used to create a ρ distribution as a function of η. This is

[2]a mathematical way of dividing spaces into regions based on proximity to a set of points in a plane, for more information see [76]

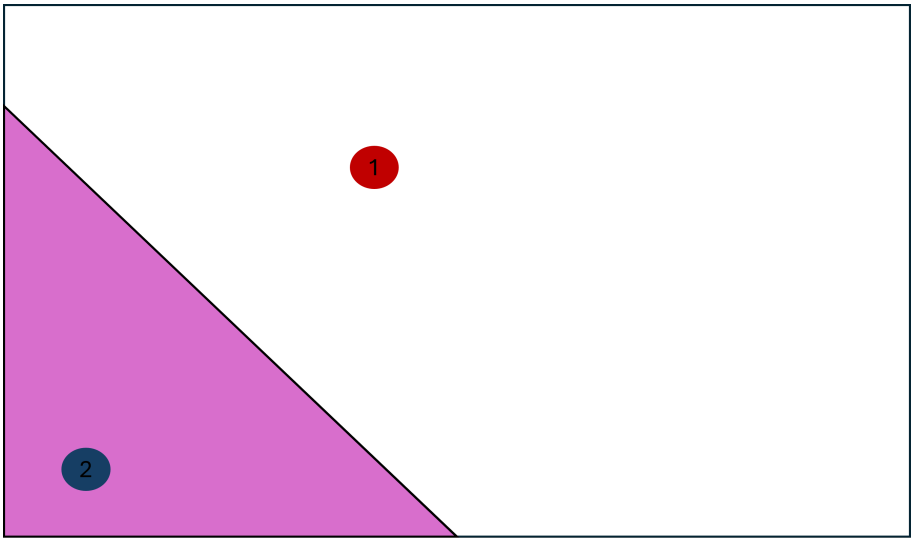


Figure 8.10: Simple diagram of a Voronoi area with 2 points. The area that is closer to point 1 than point 2 is white, while the areas that are closer to point 2 than point 1 are shaded in purple.

done on an event-by-event basis. The distribution for each event is stored until all events have been processed, and the distributions are then superimposed into one distribution.

The algorithm was run on dijet Monte Carlo(MC20) simulations simulating the run conditions for ATLAS run2. We used particle flow jets reconstructed using the anti-Kt algorithm (discussed in section 3.1.3) at EM scale.

The distribution created from these samples was then fitted, giving an equation for ρ as a function of η. This is not the only parameter that the ρ depends on. The ρ also depends on the number of primary vertices (N_{pv}) and the average number of proton-proton interactions per bunch crossing μ. The N_{pv} measures the in-time pileup that occurs when there are multiple proton-proton collisions in a single bunch crossing, and treats these excess collisions as pile-up that originated at the same time as our event. μ, on the other hand, measures the out-of-time pileup. This is when the previous event energy deposits in the calorimeter have not been dispersed fully by the time the next event occurs. For the ρ to be a function of both N_{pv} and μ we implemented a grid method. The N_{pv} and μ distribution for a MC20 sample is shown in figure 8.11 and 8.12.

The aim of the grid method was to use these distributions to produce a set of zones in which to calculate the $\rho(\eta)$ using the sliding window method. The zones are spaced along both μ and $\mathrm{N_{pv}}$ to try and give enough statistics to each zone while also covering a large enough range of both $\mathrm{N_{pv}}$ and μ to be impactful. Figure 8.13 shows the zones selected for both $\mathrm{N_{pv}}$ and μ with figure 8.14 showing the number of average μ and N_{pv} calculations

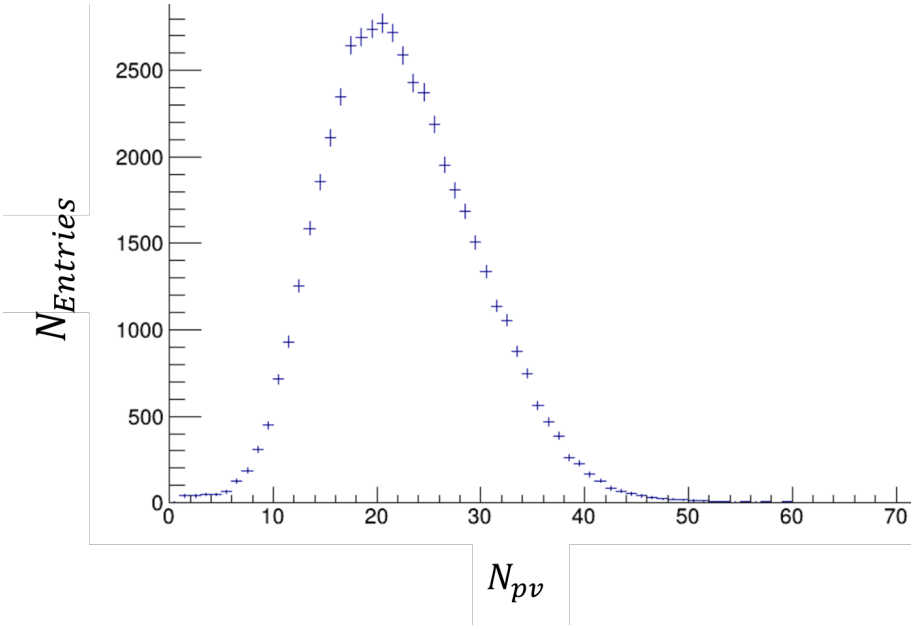


Figure 8.11: The distribution of the number of primary vertices $\mathrm{N_{pv}}$ in a MC20 sample with 50k events.

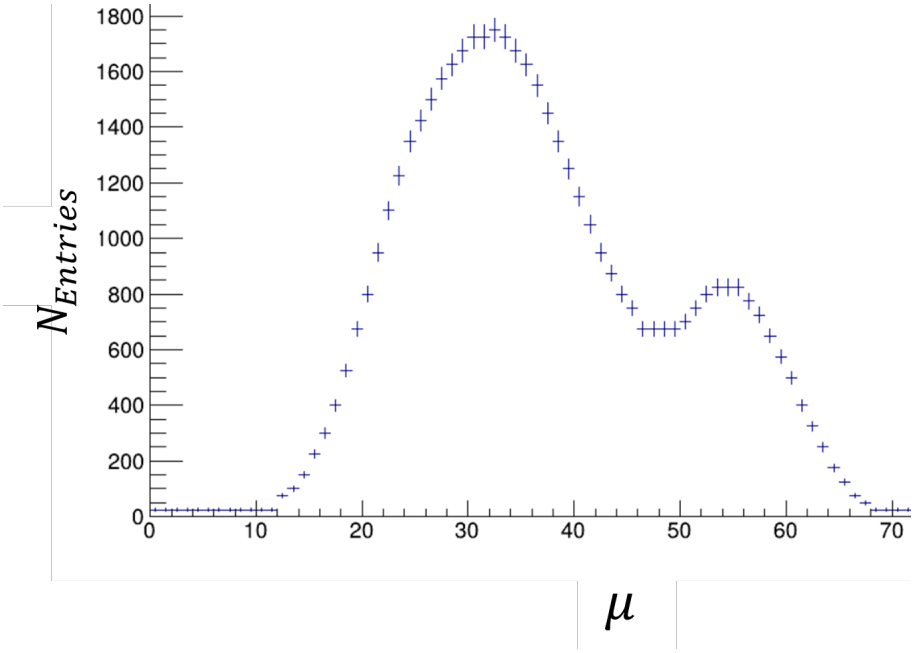


Figure 8.12: The distribution of average number of proton–proton interactions per bunch crossing (μ) in a MC20 sample with 50k events.

made in each zone. Since the figures 8.14 (a) and (b) show roughly the same N_{pv} or μ per zone we can say that the zoning didn't cause significant bias in N_{pv} or μ. This being said, grid zone 5 and 21 do seem to struggle for statistics as their calculation of the average μ and N_{pv} seems to have a lot more spread on them compared with other zones.

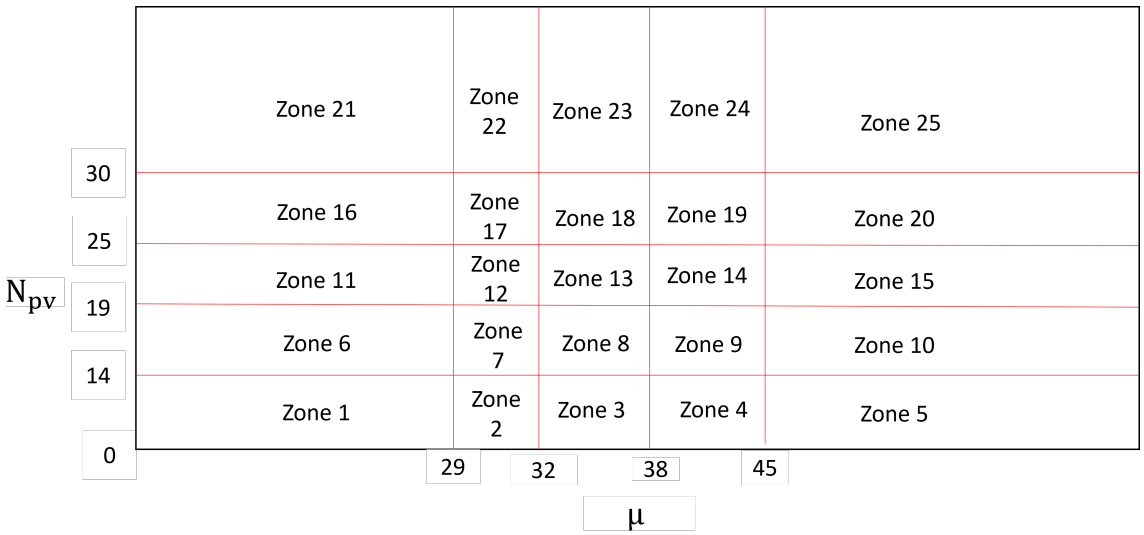
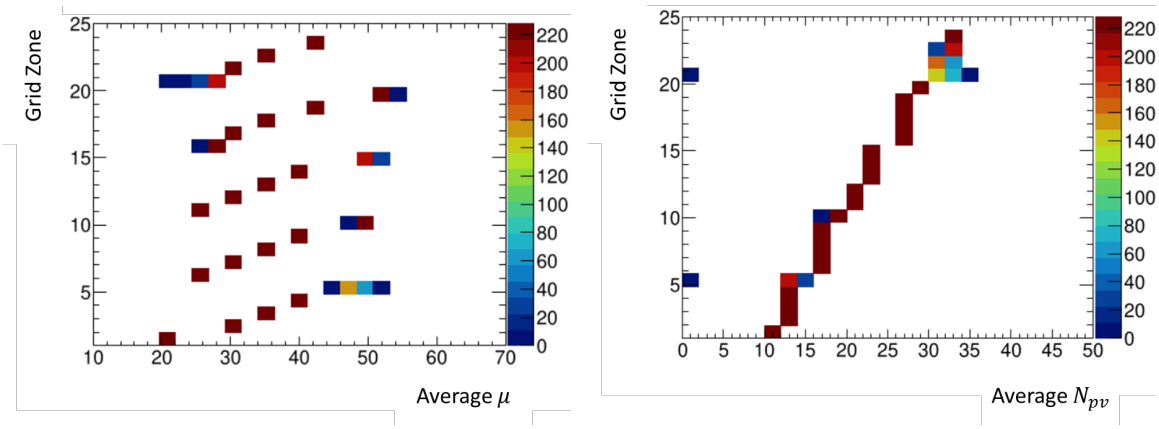


Figure 8.13: Grid representation of the various zones that the sliding window algorithm operates on. The red line represents the boundaries in N_{pv} and μ between the zones.



(a) (b)

Figure 8.14: Heatmaps showing the in (a) the number of average μ calculations that occur in each grid zone, and in (b) the number of average N_{pv} calculations that occur in each zone. The total number of calculations performed in both heat maps was 5700, with each calculation making up 2000 events.

The zones could then be used to produce an equation for the ρ dependence on $\mathrm{N_{pv}}$ and μ. This was done using the average values for $\mathrm{N_{pv}}$ and μ within these zones and the values of the parameters used to fit the $\rho(\eta)$ produced for each zone. This will be discussed further in the chapter 9.

Chapter 9

Topo tower performance

In this chapter, we evaluate the effectiveness of the new topo tower sliding window method described in section 8.4. In section 9.1 we show plots produced using the sliding window method with topological towers in the forward region, and also show segmented plots in μ and $\mathrm{N_{pv}}$. We then present the equation that evaluates ρ as a function of η, μ and $\mathrm{N_{pv}}$. In section 9.2, we define two metrics to evaluate the performance of our method in reducing the residual correction. We then show performance plots produced using truth-level information. These plots show how our new method has improved performance over the legacy method.

9.1 Transverse momentum density drop-off evaluation and equation construction

In this section, we evaluate how well the new sliding window method combined with topo-towers produces $\rho(\eta)$ for $|\eta| > 2.5$. We then use the results from this new modeling of $\rho(\eta)$ to derive a distribution function capable of reducing the residual correction.

9.1.1 Transverse momentum density drop-off evaluation

Previously in figure 8.7, we showed that previous studies had found the measured ρ to be almost zero for $|\eta| \geq 3$ at every value of μ tested. The first check for this tool is, therefore, to see if the $\rho(\eta)$ produced using our sliding window algorithm gives non-zero ρ values for $|\eta| \geq 3$ when using topo towers. The $\rho(\eta)$ we plot is then normalized using the average value of $\rho(\eta)$ calculated within $|\eta| < 2$ to produce figure 9.1. This normalization is chosen to allow for the slope of ρ vs η, in $2.5 < |\eta| < 5$ to be modeled in a way that can be normalized to any ρ_{event} (defined in equation 8.15). This normalization also allows us to see the percentage decrease of ρ across η.

Figure 9.1 shows the $\rho(\eta)$ distribution, which now extends into the $|\eta| \geq 3$ region, only

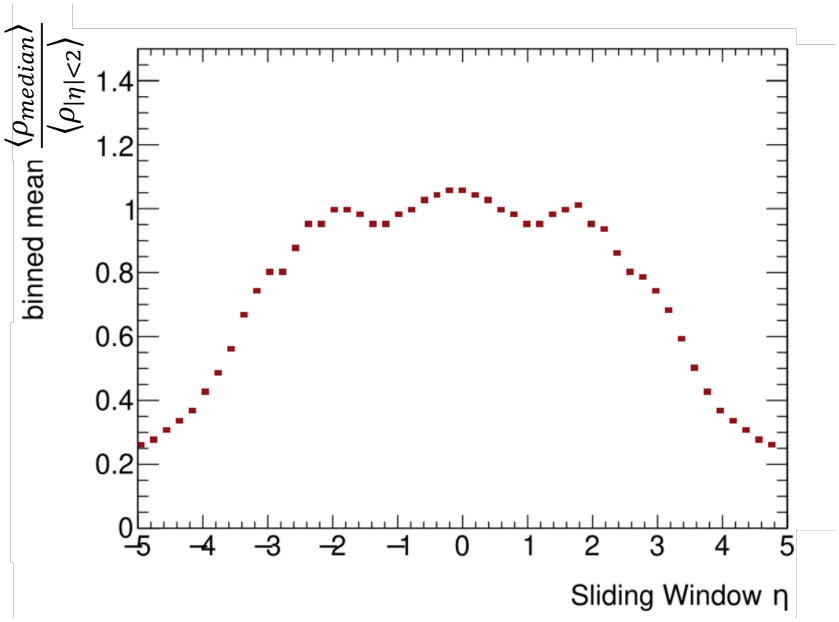


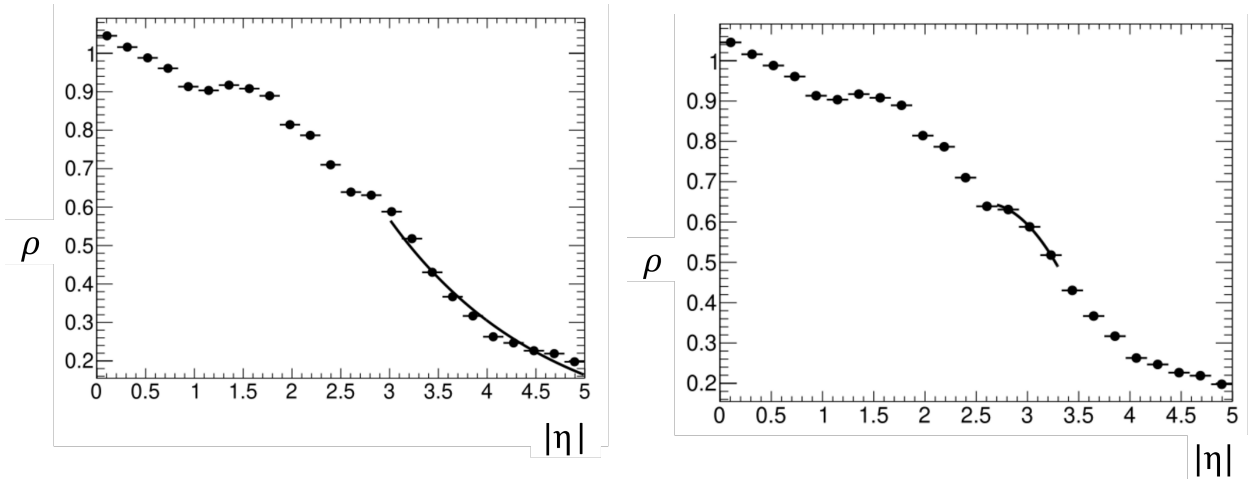
Figure 9.1: Ratio of the mean average ρ_{medain} in each η bin to the average value calculated within $|\eta| < 2$. The ρ_{median} is the ρ value calculated from the sliding window algorithm. This plot was produced on MC20 JETM2 dijet samples.

decreasing to around 20% of its maximum value towards the end of the detector. This is compared to figure 8.7, which loses 99% of its events for $|\eta| \geq 3$, and is a massive improvement in not just the fraction of events kept but also for the smoothness in the modeling of the forward region. Figure 9.1 also shows symmetry in the modeling of $\rho(\eta)$ for both positive and negative η. Therefore, we used the modulus of η in performance plots, using the mean ρ from the positive and negative η counterparts.

9.1.2 Equation construction

Figure 9.1 is plotted for all values of $\mathrm{N_{pv}}$ and μ, to produce an equation as a function of these variable we need to segment this figure into bins of $\mathrm{N_{pv}}$ and μ as specified in the zoning in figure 8.13. Figure **??** shows an example of how the $\rho(\eta)$ distribution in figure 9.1 changes when selecting events in grid zone 10.

Figure **??** shows two plots, each parameterized using a different function, an exponential and a polynomial of order 2 (pol2). These functions are designed to overlap with the exponential function parameterization between $3 \leq |\eta| \leq 4.9$, while the pol2 part of the parameterization is between $2.7 \leq |\eta| \leq 3.3$. This two-parameter method was designed to smooth off the transition between the slope and the shoulder that appears between $2.5 \leq |\eta| \leq 3$. The fits are combined using a weighted average that takes the average of the output produced by both the exponential and pol2 parameterization at the same value



(a) (b)

Figure 9.2: The ρ distribution produced using the sliding window algorithm and topo-towers in zone 10. The ρ shown on the y-axis is the ratio of the mean average ρ_{median} in each η bin to the average value calculated within $|\eta| < 2$. A parameterization is shown for an exponential function between $3 \leq |\eta| \leq 4.9$ in plot (a). A parameterization is shown for a second-degree polynomial for $2.7 \leq |\eta| \leq 3.3$ in plot (b). The ρ corresponding to the negative and positive values of η were averaged to give the values corresponding to $|\eta|$.

of η between $3 \leq |\eta| \leq 3.3$. The average is weighted with a linear function to give increased bias towards the exponential at $|\eta| = 3.3$ and increased bias towards pol2 at $|\eta| = 3$. The central region itself is not modeled as the ρ_{event} method works well in that region. The two parameterizations are modeled as

$$\rho(\eta) = Ae^{B\eta}, \tag{9.1}$$

$$\rho(\eta) = P_2\eta^2 + P_1\eta + P_0, \tag{9.2}$$

Where P_2, P_1, P_0, A and B are constants to be determined from the parameterization. The parameterizations from equations 9.1 and 9.2 were performed for every zone from 1 to 25 in figure 8.13. We want to determine how the constants change with N_{pv} and μ. The average N_{pv} and μ in each zone were calculated and used as the zone's N_{pv} and μ coordinates. As can be seen in figure 8.13, the zone setup allows for N_{pv} to be changed while μ is kept constant for 5 values of N_{pv}. This can be repeated for 5 constant μ values across each JETM2 sample(defined in section 3.2.5) used, giving 5 fits per sample. A slice of these files is shown for the exponential part in figure 9.3 and for the pol2 part in figure 9.4. These plots are fitted to get the constants' dependence on N_{pv}.

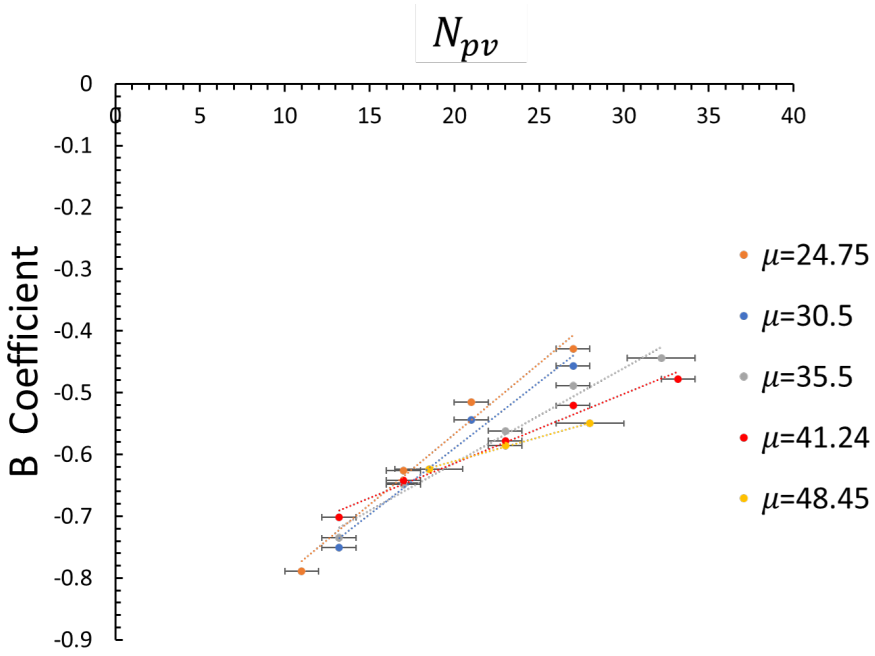


Figure 9.3: The B coefficient, for the exponential function controlling the ρ decrease with η, plotted against $\mathrm{N_{pv}}$. The constant value for μ over which $\mathrm{N_{pv}}$ is changed is listed on the key. The fits are colour-coded to the data points. This plot was fitted with exponential fits. The χ^2 evaluation for all the fits was over 0.999.

As can be seen in figure 9.4, the constant value of μ has a small impact on the fit results (as there is no visible trend for differences in μ and many points overlap within error bars); on the other hand, the $\mathrm{N_{pv}}$ dependence of these plots is large. The $\mathrm{P_0}$ and $\mathrm{P_2}$ constants are modeled using a logarithmic function while $\mathrm{P_1}$ uses an exponential. The multiple fits corresponding to the different contact values of μ are averaged and then used to give the $\mathrm{N_{pv}}$ dependence for the pol2 function in equation 9.2. The equations for the $\mathrm{P_0}$, $\mathrm{P_1}$ and $\mathrm{P_2}$ are

$$P_0 = A_{P_0} \log(N_{pv}) + B_{P_0}, \tag{9.3}$$

$$P_1 = A_{P_1} \exp(N_{pv} B_{P_1}), \tag{9.4}$$

$$P_2 = A_{P_2} \log(N_{pv}) + B_{P_2}, \tag{9.5}$$

where $\mathrm{A_{P_0}}$, $\mathrm{A_{P_1}}$, $\mathrm{A_{P_2}}$, $\mathrm{B_{P_0}}$ $\mathrm{B_{P_1}}$ and $\mathrm{B_{P_2}}$ are all constants to be determine from the fits in figure 9.4. Figure 9.3, however, does have an observable and consistent constant μ dependence, which will need to be modeled. The $\mathrm{N_{pv}}$ dependence of the B constant can be modeled as

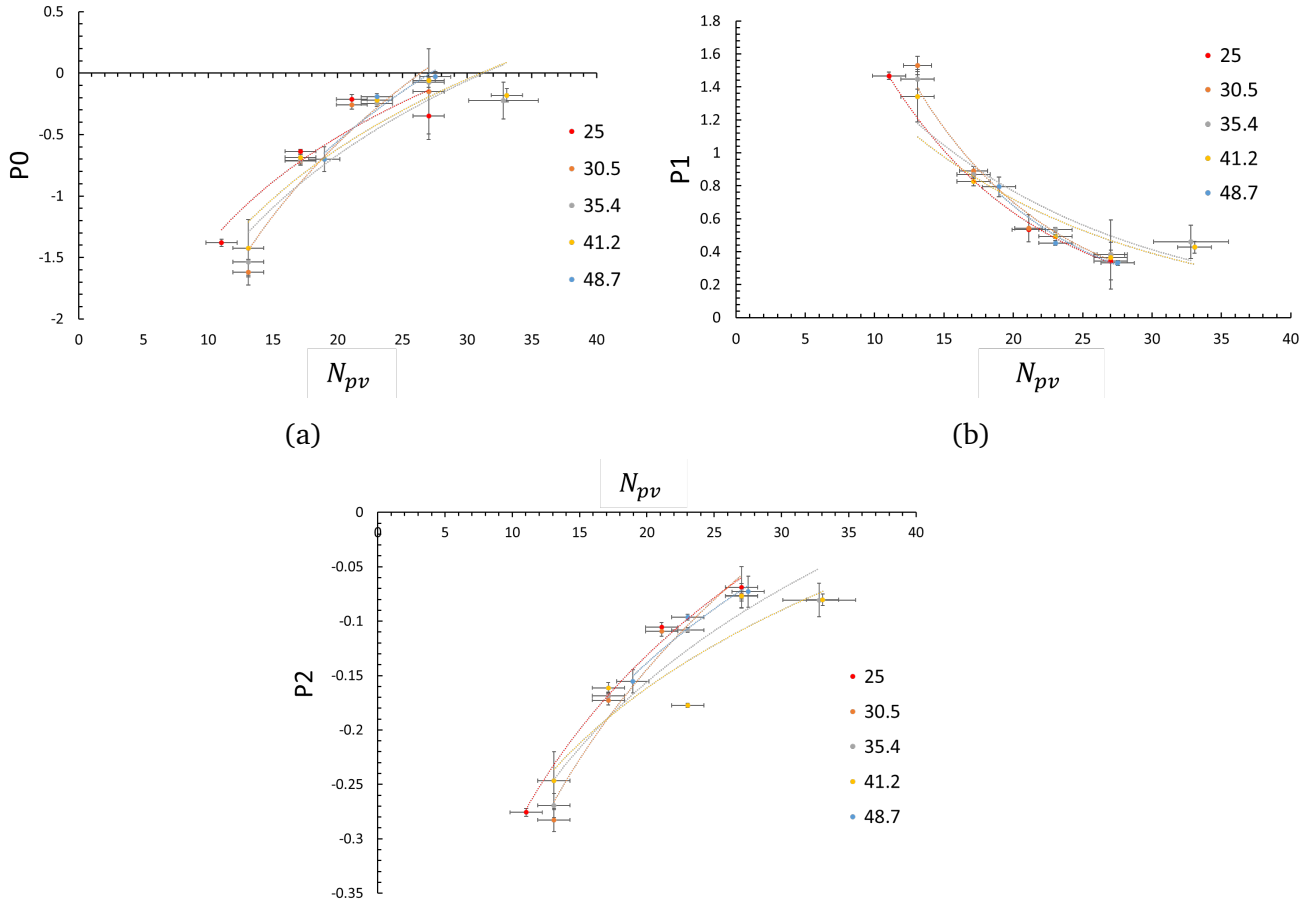


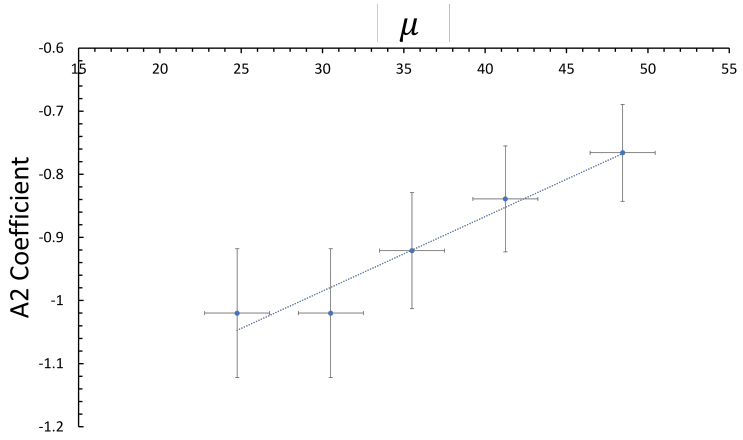
Figure 9.4: The P_0, P_1 and P_2 coefficients, for the pol2 function controlling the shoulder modeling in the ρ vs $|\eta|$ distribution between $2.7 < |\eta| < 3.3$, plotted against N_{pv}. The constant value for μ over which N_{pv} is changed is listed on the key. The fits are colour-coded to the data points. These plots were fitted with a logarithm, exponential and logarithm for plots a, b and c, respectively. For plot (a) the χ^2 for each fit are: $\chi^2_{25} = 0.926$, $\chi^2_{30.5} = 0.929$, $\chi^2_{35.4} = 0.965$, $\chi^2_{41.2} = 0.967$, $\chi^2_{48.7} = 0.84$. In plots (b) and (c) we had $\chi^2 > 0.995$ and $\chi^2 > 0.999$ for all fits in each plot, respectively.

$$B = B_2 N_{pv} + A_2, \tag{9.6}$$

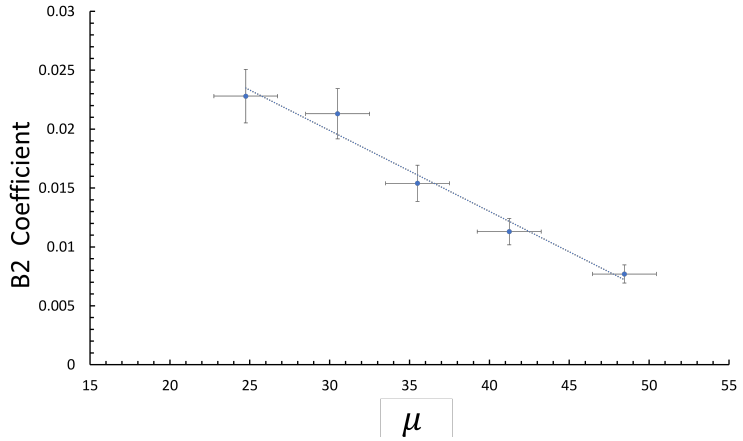
where A_2 and B_2 are both μ dependent constants found from fitting the plots in figure 9.5.

As can be seen in figure 9.5 the μ dependence of the B_2 and A_2 are linear. They are modeled like so

$$B_2 = m_1 \mu + c_1, \tag{9.7}$$



(a)



(b)

Figure 9.5: Plot (b) shows the B_2 coefficient, of the B coefficient, plotted against μ. Plot (a) shows the A_2 coefficient, of the B coefficient, plotted against μ. These plots were fitted with linear fits. For both fits in plots (a) and (b) we had $\chi^2 > 0.999$.

$$A_2 = m_2\mu + c_2, \tag{9.8}$$

where m_1, m_2, c_1 and c_2 are constants found empirically from the plots in figure 9.5.

This now gives us an equation that gives a value of ρ that is dependent on η, N_{pv} and μ. The steps to obtain this equation are summarized in figure 9.6. The constants for figure 9.6 are found from the fits and are shown in table 9.1

This equation needs to be tested to see if it matches well with the truth level information as well as how much it affects the already existing residual correction. This will be discussed in the next section.

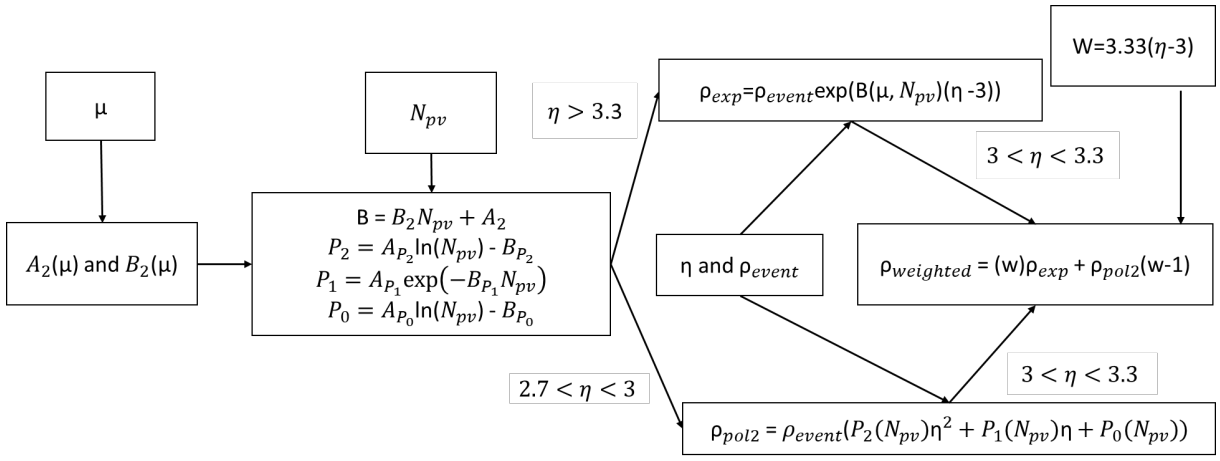


Figure 9.6: A diagram detailing how a new method of calculating the $\rho(\eta, \mu, \mathrm{N_{pv}})$ jet correction factor. This $\rho(\eta, \mu, \mathrm{N_{pv}})$ correction has an exponential part and a polynomial of the order 2 part. The exponential deals with $|\eta| > 3$ and the pol2 part deals with the range $2.7 < |\eta| < 3.3$. In the regions of overlap between the two functions, the two functions are combined by a weighted average.

Constant	value
A	ρ_{event}
$\mathrm{A_{P_0}}$	1.6 $\pm$ 0.5
$\mathrm{A_{P_1}}$	3.9 $\pm$ 0.4
$\mathrm{A_{P_2}}$	0.23 $\pm$ 0.02
$\mathrm{B_{P_0}}$	-4.2 $\pm$ 1.6
$\mathrm{B_{P_1}}$	-0.83 $\pm$ 0.01
$\mathrm{B_{P_2}}$	-0.83 $\pm$ 0.01
$\mathrm{m_1}$	-0.0007 $\pm$ 0.0002
$\mathrm{m_2}$	0.012 $\pm$ 0.003
$\mathrm{c_1}$	0.044 $\pm$ 0.009
$\mathrm{c_2}$	-1.43 $\pm$ 0.03

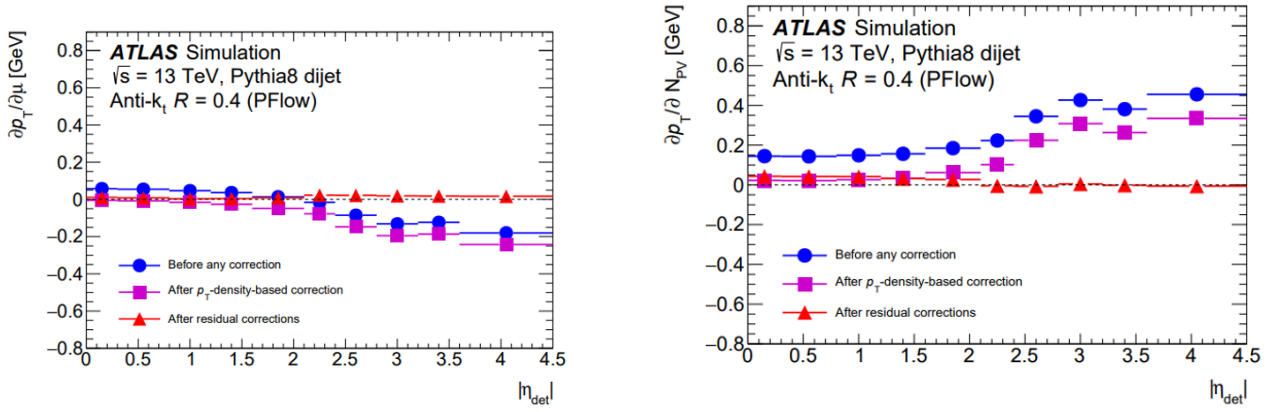
Table 9.1: A list of constants determined from a range of fits on data produced from MC20 dijet files. Constant A changes based on what event is being processed.

9.2 Truth level level performance plots

This section evaluates the performance, namely the impact of this equation on the residual correction, of the equation presented in figure 9.6. In ATLAS, the metric which is used as an evaluation of the residual corrections are $\frac{\partial \mathrm{p_T}}{\partial \mu}$ and $\frac{\partial \mathrm{p_T}}{\partial \mathrm{n_{pv}}}$ [77], which is a measure of the dependence of the p_T on N_{pv} and μ. These measurements are fitted over η to determine the correction factors applied to jets by the residual correction. In this thesis, corrections are calculated using particle flow(Pflow) jets, which were matched to truth jets and filled into $\mathrm{p_T}$ bins centered around truth jet $\mathrm{p_T} = 25\ \mathrm{GeV}$. The Pflow jets are then processed, with the processing being different depending on the level of correction. The processing for the $\mathrm{p_T}$ based correction applies the area-based correction described in equation 8.16. The corrected jet is then binned in η, where a fit is performed in each η bin. The $\mathrm{p_T}$ is plotted against both $\mathrm{N_{pv}}$ and μ in each bin. The fits in these η bins give the variables $\frac{\partial \mathrm{p_T}}{\partial \mathrm{N_{pv}}}$ and $\frac{\partial \mathrm{p_T}}{\partial \mu}$. These are plotted in each η bin and then fitted to provide the residual correction factors α and β for $\frac{\partial \mathrm{p_T}}{\partial \mathrm{N_{pv}}}$ and $\frac{\partial \mathrm{p_T}}{\partial \mu}$ respectively. The residual correction can be added to 8.16 in the following way [77]

$$p_{T,i}^{corr} = p_{T,i}^{jet} - \rho_{event} \times A_i^{jet} - \alpha(N_{pv} - 1) + \beta\mu. \tag{9.9}$$

In figure 9.7 we see both $\frac{\partial \mathrm{p_T}}{\partial \mu}$ and $\frac{\partial \mathrm{p_T}}{\partial \mathrm{N_{pv}}}$ plotted without, with $\rho_{\mathrm{event}}\mathrm{A_i^{jet}}$ and with $-\alpha(\mathrm{N_{pv}} - 1) + \beta\mu$ corrections [77].



(a) (b)

Figure 9.7: (a) is the η dependence of the residual correction measured in $\mathrm{N_{pv}}$. (b) is the η dependence of the residual correction measured in μ [77]

We see from figures 9.7 that after the residual correction is applied there is a roughly zero dependency of the p_T on the μ or N_{pv}. The difference between the various corrections dependency on these variables and zero is the size of the residual correction needed which we wish to be as small as possible. We see from figures 9.7 that the ρ_{event} based corrections

decreases the $\frac{\partial \mathrm{p_T}}{\partial \mathrm{npv}}$ correction needed, however, it has the opposite effect on the $\frac{\partial \mathrm{p_T}}{\partial \mu}$ making its correction larger. The equation we derived in section 9.1 was created to reduce these residual corrections. We follow the same method for producing these plots as the ones shown in figure 9.7, but instead use the new correction method defined in 9.6 for the jet $\mathrm{p_T}$ correction. In figure 9.8.(a) and 9.8.(b) the correction factors defined using our new method, defined in figure 9.6 are shown. We also produce control plots in figures 9.8.(c) and 9.8.(d) using the old correction to compare.

As can be seen in figure 9.8, when comparing the control plots produced using the old method to the new methods, the improvements of the new method are η dependent with both metrics $\frac{\partial p_T}{\partial \mu}$ and $\frac{\partial p_T}{\partial N_{pv}}$ showing varying performance over different η ranges.

In figure 9.8 for the μ dependency we see improved performance in $3.5 < |\eta| < 5$ with roughly a 20% smaller residual correction needed. The $3 < |\eta| < 3.5$ bin sees roughly the same performance while the $2.5 < |\eta| < 3$ bin sees worse performance. The worse performance in the $2.5 < |\eta| < 3$ bin could be due to the lack of μ modeling in the region. The improved performance is still lower than what were expecting in the region which could suggest issues with the μ modeling, thankfully for future studies the modeling seems to give consistent modeling in the forward when compared with the old method, which means the problem could likely be fixed with some scale factor applied to the μ modeling in the equation represented in figure 9.6.

In figure 9.8 for the dependency of N_{pv}, we see improved performance for $2.5 < |\eta| < 3.5$, but decreased performance for $3.5 < |\eta| < 4.5$. There is also the issue seen in the bins between $2.5 < |\eta| < 4.5$ where the modeling of N_{pv} gets worse as η increases. This suggests some failing in the modeling of N_{pv} and its relationship with η. The modeling between $2.5 < |\eta| < 3$ shows the largest improvement and is modeled using the pol2 function rather than the exponential function. In future studies, the N_{pv} modeling will need to be evaluated.

We can see that the new method has not eliminated the need for a residual correction in the forward region, as was hoped. The residual correction applied to the plots in figure 9.8 would give very similar corrections for α in equation 9.9 but would give a reduced value for β. Using this new method, therefore, would require smaller residual corrections in the forward region, and yield smaller systematic uncertainties. Further studies into the modeling of N_{pv} and μ using the studies provided in this thesis as a starting off point will reduce the residual correction further.

The tool created here is a proof of concept that modeling the forward regions $\rho(\eta)$ using

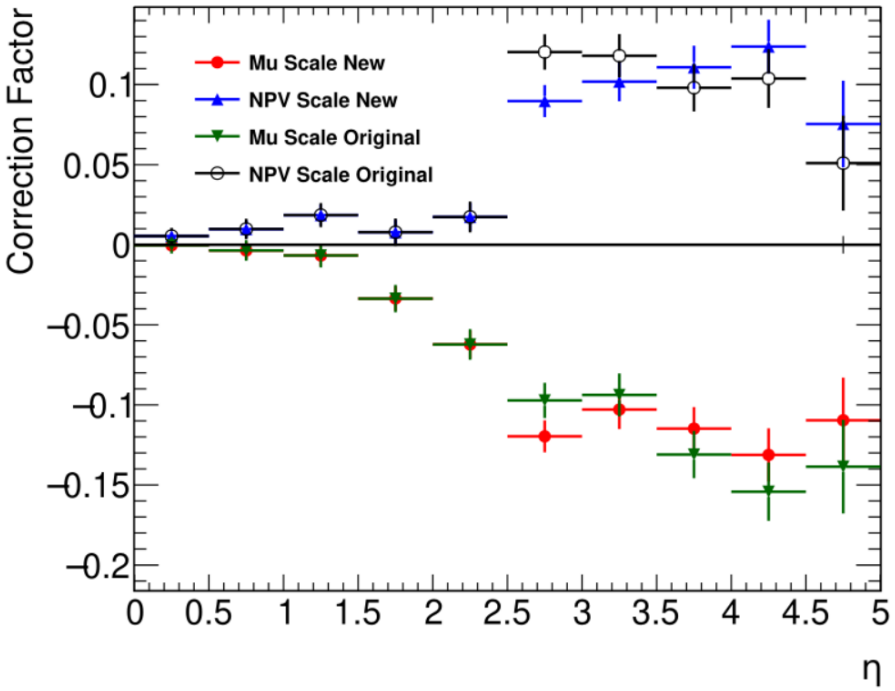


Figure 9.8: The η dependence of the residual correction measured in $\mathrm{N_{pv}}$ and μ. The new key highlights the new topological tower method in the forward in blue and red while the old methods using clusters is shown in white and green.

topo towers as a function of N_{pv} and μ could reduce the residual correction in the forward region. We have shown some promising reductions in residual corrections for specific η ranges, especially in $\frac{\partial p_T}{\partial \mu}$. We have successfully modeled $\rho(\eta)$ in the forward region using topo-towers where topo-clusters failed. Making the plots shown in this thesis the first novel measurements of $\rho(\eta)$ in the forward region. We have also shown the first $\rho(\eta, N_{pv}, \mu)$ however the modeling of this will require further study, as we would expect actuate modeling of $\rho(\eta, N_{pv}, \mu)$ to reduce the residual correction to zero which does not happen. Once this tool has been fully developed, we expect the reduction in the residual correction to allow forward jets to be used in measurements. At the moment, the uncertainties on the residual correction wash out any gain analysis that would get from using them. Reducing the residual correction and allowing these forward jets to be used in analyses would vastly boost the number of events available in measurements. This means analyses searching for rare events at ATLAS could significantly benefit from the increased number of events. The reconstruction of $\mathrm{E_T^{miss}}$ could also greatly benefit from this, as it requires full event information to reconstruct, including forward jets. Analyses that plan to use $\mathrm{E_T^{miss}}$ in their analysis, such as the one covered in this thesis in part II, would greatly benefit from the reduced systematic uncertainties on this object.

Chapter 10

Conclusion

In this thesis, we have presented two sets of research, targeting both the physics and technical aspects of ATLAS.

In chapter 7, we showed the results of measuring the cross section of $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$ in two different fits. The unblinded standalone 0L only fit, and the blinded combined 0L, 1L, and 2L fit. The 0L only fit produced a cross-section value of $\sigma_{t\bar{t}Z\to\nu\bar{\nu}} = 0.11^{+0.08}_{-0.07}\ pb$, this shows agreement with the standard model prediction with a signal significance $Z = 2.47\ \sigma$. The combined measurement of $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$ presented in 7 had an expected signal significance of $Z = 4.9\ \sigma$. This will be the best cross-section measurement of $\mathrm{t\bar{t}Z}$ in the $p_T^Z > 200\ \mathrm{GeV}$ range. This measurement will also constrain the background for both SUSY and dark matter searches, potentially improving their significance to $Z \geq 3\ \sigma$, which in turn could help in the discovery of new physics.

In chapter 9, we showed the results of using the new method of calculating pile-up in the forward region of the ATLAS calorimeter. The topological towers perform much better than their topological cluster counterparts at measuring the transverse momentum density in the forward region. This improved modeling in itself is a novel measurement and can be used for future studies into pile-up density in the forward region. We performed a study to see if this improved modeling in the forward region could also be made a function of N_{pv} and μ, which would allow us to reduce the residual correction by replacing the original method of correcting jets with one that changes the momentum correction based on the event's pile-up parameters. This method only partially succeeded with an improvement in performance for the $\frac{\partial p_T}{\partial \mu}$ part of the residual correction, and mixed results for the $\frac{\partial p_T}{\partial N_{pv}}$ part, which performed better at lower $|\eta|$ and worse towards higher $|\eta|$. Improvements of this tool will be implemented to process run-3 data and would allow forward jets to be used in measurements, which would allow studies which struggle for statistics to gain events, with improved forward jets performance and reduced systematic uncertainties on E_T^{miss} events;

which would in turn improve the significance of studies such as the $\mathrm{t\bar{t}Z} \to \nu\bar{\nu}$ conducted in this thesis.

Bibliography

[1] G. Altarelli, "Collider physics within the standard model: a primer," 2013.

[2] P. Langacker, *The standard model and beyond*. Series in high energy physics, cosmology, and gravitation, Boca Raton, [Florida] ; London, [England] ; New York: CRC Press, second edition. ed., 2017.

[3] Y. Nagashima, *Beyond the standard model of elementary particle physics*. Weinheim, Germany: Wiley-VCH Verlag, 1st ed. ed., 2014.

[4] P. W. Higgs, "Broken symmetries and the masses of gauge bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, Oct 1964.

[5] M. Tanabashi and et al, "Review of particle physics," *Phys. Rev. D*, vol. 98, p. 030001, Aug 2018.

[6] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, "Evidence for the 2π decay of the k_2^0 meson," *Phys. Rev. Lett.*, vol. 13, pp. 138–140, Jul 1964.

[7] S. D. Bass, A. De Roeck, and M. Kado, "The higgs boson implications and prospects for future discoveries," *Nature reviews physics*, vol. 3, no. 9, pp. 608–624, 2021.

[8] ATLAS collaboration, "Standard Model Summary Plots June 2024," tech. rep., CERN, Geneva, 2024. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2024-011.

[9] A. Carbone, "A search for time-integrated cp violation in $d^0 \to h^- h^+$ decays," 2012.

[10] LHCb Collaboration, "First observation of cp violation in the decays of B_s^0 mesons," *Phys. Rev. Lett.*, vol. 110, p. 221601, May 2013.

[11] I. Brivio and M. Trott, "The standard model as an effective field theory," *Physics Reports*, vol. 793, pp. 1–98, 2019. The standard model as an effective field theory.

[12] J. Adelman, B. Alvarez Gonzalez, Y. Bai, M. Baumgart, R. K. Ellis, A. Khanov, A. Loginov, and M. Vos, "Top Couplings: Pre-Snowmass Energy Frontier 2013 Overview," in *Snowmass 2013: Snowmass on the Mississippi*, 9 2013.

[13] L. Evans and P. Bryant, "LHC machine," *Journal of Instrumentation*, vol. 3, p. S08001, aug 2008.

[14] M. Anthony, "Search for the supersymmetric partner to the top quark in the all-hadronic final state with the atlas detector," 2020.

[15] E. Mobs, "The CERN accelerator complex in 2019. Complexe des accélérateurs du CERN en 2019," 2019. General Photo.

[16] "Radiofrequency cavities," 2012.

[17] ATLAS collaboration, "The atlas experiment at the cern large hadron collider," *Journal of instrumentation*, vol. 3, no. 8, pp. S08003–S08003, 2008.

[18] B. Ravina, "Measurement of the top quark coupling to the z boson, and applications to searches for physics beyond the standard model." April 2020.

[19] M. P. Giordani, "The Upgraded Pixel Detector of the ATLAS Experiment for Run-2 at the LHC," tech. rep., CERN, Geneva, 2016.

[20] A. Vogel, "ATLAS Transition Radiation Tracker (TRT): Straw Tube Gaseous Detectors at High Rates," tech. rep., CERN, Geneva, 2013.

[21] T. A. collaboration, "Operation of the atlas trigger system in run 2," *Journal of Instrumentation*, vol. 15, p. P10004–P10004, Oct. 2020.

[22] ATLAS collaboration, "Performance of the ATLAS Trigger System in 2015. Performance of the ATLAS Trigger System in 2015," *Eur. Phys. J. C*, vol. 77, no. 5, p. 317, 2017. 77 pages in total, author list starting page 61, 50 figures, 1 table. Published in Eur. Phys. J. C. All figures including auxiliary figures are available at http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/TRIG-2016-01/.

[23] ATLAS Collaboration, "Event Displays for Run 2 Physics," 2018. Accessed on 2025-09-09.

[24] C. Wasicki, "Track and vertex reconstruction of the atlas inner detector in the high multiplicity lhc environment," *Journal of Physics: Conference Series*, vol. 396, no. 2, pp. 22056–6, 2012.

[25] E. Nurse, T. Kuhl, and A. K. Morley, "Charged-particle multiplicities in $\sqrt{s}$=13 TeV pp interactions measured with the ATLAS detector at the LHC," tech. rep., CERN, Geneva, 2015. Draft for 13 TeV minimum bias paper.

[26] ATLAS Collaboration, "Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV,$''$ 2015.

[27] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-ktjet clustering algorithm," *Journal of High Energy Physics*, vol. 2008, p. 063–063, Apr. 2008.

[28] ATLAS Collaboration, "Tagging and suppression of pileup jets with the ATLAS detector," tech. rep., CERN, Geneva, 2014. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2014-018.

[29] ATLAS Collaboration, "Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the atlas detector," *Phys. Rev. D*, vol. 96, p. 072002, Oct 2017.

[30] A. Onofre, N. F. Castro, and A. Collaboration, "Determination of jet calibration and energy resolution in proton–proton collisions at $\sqrt{s} = 8$ TeV using the atlas detector," *The European physical journal. C, Particles and fields*, vol. 80, no. 12, 2020.

[31] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas," *Physics letters. B*, vol. 659, no. 1-2, pp. 119–126, 2008.

[32] ATLAS collaboration, "Performance of the ATLAS Secondary Vertex b-tagging Algorithm in 900 GeV Collision Data," tech. rep., CERN, Geneva, 2010. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2010-004.

[33] ATLAS collaboration, "Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run," tech. rep., CERN, Geneva, 2016. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2016-012.

[34] A. Onofre, N. F. Castro, and A. Collaboration, "Electron reconstruction and identification in the atlas experiment using the 2015 and 2016 lhc proton-proton collision data at $\sqrt{s} = 13$ TeV," *The European physical journal. C, Particles and fields*, vol. 79, no. 8, 2019.

[35] Y. Shen, "Loss functions for binary classification and class probability estimation," 2005.

[36] ATLAS Collaboration, "Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016," *Eur. Phys. J. C*, vol. 79, no. 3, p. 205, 2019. 55 pages in total, author list starting page 39, 18 figures, 2 tables, published in EPJC. All figures including auxiliary figures are available at http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/PERF-2017-02.

[37] A. Onofre, N. F. Castro, and A. Collaboration, "Performance of missing transverse momentum reconstruction with the atlas detector using proton–proton collisions at $\sqrt{s} = 13$ TeV," *The European physical journal. C, Particles and fields*, vol. 78, no. 11, 2018.

[38] C. Z. Mooney, *Monte Carlo simulation*. Quantitative applications in the social sciences ; no. 07-116, Thousand Oaks, Calif. ; London: SAGE, 1997.

[39] A. Abdesselam and et al, "The atlas simulation infrastructure," *The European physical journal. C, Particles and fields*, vol. 70, no. 3, pp. 823–874, 2010.

[40] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to pythia 8.2," *Computer Physics Communications*, vol. 191, pp. 159–177, 2015.

[41] M. Bähr, S. Gieseke, M. A. Gigg, D. Grellscheid, K. Hamilton, O. Latunde-Dada, S. Plätzer, P. Richardson, M. H. Seymour, A. Sherstnev, and B. R. Webber, "Herwig++ physics and manual," *The European physical journal. C, Particles and fields*, vol. 58, no. 4, pp. 639–707, 2008.

[42] H. Baer, F. E. Paige, S. D. Protopescu, and X. Tata, "Isajet 7.69: A monte carlo event generator for pp, $\bar{p}p$, and $e^{=}e^{-}$ reactions," 2003.

[43] E. Bothmann, G. S. Chahal, S. Höche, J. Krause, F. Krauss, S. Kuttimalai, S. Liebschner, D. Napoletano, M. Schönherr, H. Schulz, S. Schumann, and F. Siegert, "Event generation with Sherpa 2.2," *SciPost Phys.*, vol. 7, p. 034, 2019.

[44] F. Abudinén, J. Back, M. Kreps, and T. Latham, "Evtgen — recent developments and prospects," *EPJ Web of conferences*, vol. 295, p. 3012, 2024.

[45] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations," *Journal of High Energy Physics*, vol. 2002, p. 029, jul 2002.

[46] M. Dobbs and J. B. Hansen, "The hepmc c++ monte carlo event record for high energy physics," *Computer Physics Communications*, vol. 134, no. 1, pp. 41–46, 2001.

[47] S. Agostinelli and et al, "Geant4—a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, 2003.

[48] P. Calafiura, W. Lavrijsen, C. Leggett, M. Marino, and D. Quarrie, "The Athena Control Framework in Production, New Developments and Lessons Learned," 2005.

[49] E. Richter-Was, D. Froidevaux, and L. Poggioli, "ATLFAST 2.0 a fast simulation package for ATLAS," tech. rep., CERN, Geneva, 1998.

[50] K. Edmonds, S. Fleischmann, T. Lenz, C. Magass, J. Mechnich, and A. Salzburger, "The Fast ATLAS Track Simulation (FATRAS)," tech. rep., CERN, Geneva, 2008. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-SOFT-PUB-2008-001.

[51] T. Yamanaka and A. Collaboration, "The atlas calorimeter simulation fastcalosim," *Journal of Physics: Conference Series*, vol. 331, p. 032053, dec 2011.

[52] L. Lista, "Practical statistics for particle physicists," *arXiv: Data Analysis, Statistics and Probability*, 2016.

[53] M. Nielson, *Neural Networks and Deep Learning*. LibreTexts, 2023.

[54] J. Therhaag, A. Lucotte, and T. Delemontex, "Introduction to neural networks in high energy physics," *EPJ Web of Conferences*, vol. 55, p. 2003, 2013.

[55] A. Apicella, F. Donnarumma, F. Isgrò, and R. Prevete, "A survey on modern trainable activation functions," *Neural networks*, vol. 138, pp. 14–32, 2021.

[56] S. Korn, "Mva-trainer documentation," 2025. `https://mva-trainer-docs-site.docs.cern.ch/` [Accessed: (16/01/2025)].

[57] F. S. Nahm, "Receiver operating characteristic curve: overview and practical use for clinicians," *Korean journal of anesthesiology*, vol. 75, no. 1, pp. 25–36, 2022.

[58] A. Kulesza, L. Motyka, D. Schwartländer, T. Stebel, and V. Theeuwes, "Associated production of a top quark pair with a heavy electroweak gauge boson at nlo+nnll accuracy," *The European physical journal. C, Particles and fields*, vol. 79, no. 3, pp. 1–15, 2019.

[59] S. Chatrchyan and et al, "Measurement of associated production of vector bosons and top quark-antiquark pairs in pp collisions at $\sqrt{s}$=7 TeV,"

[60] G. Aad and et al., "Measurement of the $t\bar{t}w$ and $t\bar{t}z$ production cross sections in pp collisions at $\sqrt{s} = 8$ TeV with the atlas detector," *The journal of high energy physics*, vol. 2015, no. 11, pp. 1–48, 2015.

[61] V. Khachatryan and et al, "Observation of top quark pairs produced in association with a vector boson in pp collisions at $\sqrt{s} = 8$ TeV," *Journal of High Energy Physics*, vol. 2016, no. 1.

[62] ATLAS Collaboration, "Measurements of the inclusive and differential production cross sections of a top-quark–antiquark pair in association with a z boson at $\sqrt{s} = 13$ TeV with the atlas detector," 2021.

[63] ATLAS Collaboration, "Inclusive and differential cross-section measurements of $t\bar{t}z$ production in pp collisions at $\sqrt{s} = 13$ TeV with the atlas detector, including EFT and spin-correlation interpretations," 2024.

[64] E. Alunno Camelia and et al, "Search for a scalar partner of the top quark in the all-hadronic tt^{-} plus missing transverse momentum final state at $\sqrt{s} = 13$ TeV with the atlas detector," *The European physical journal. C, Particles and fields*, vol. 80, no. 8, 2020.

[65] ATLAS Collaboration, "Search for new phenomena with top quark pairs in final states with one lepton, jets, and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the atlas detector," *The journal of high energy physics*, vol. 2021, no. 4, 2021.

[66] ATLAS Collaboration, "Search for new phenomena in events with two opposite-charge leptons, jets and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the atlas detector," *The journal of high energy physics*, vol. 2021, no. 4, 2021.

[67] C. Lester and D. Summers, "Measuring masses of semi-invisibly decaying particle pairs produced at hadron colliders," *Physics Letters B*, vol. 463, no. 1, pp. 99–103, 1999.

[68] ATLAS Collaboration, "Jet reclustering and close-by effects in ATLAS run II," tech. rep., CERN, Geneva, 2017. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2017-062.

[69] ATLAS Collaboration, "2015 start-up trigger menu and initial performance assessment of the ATLAS trigger using Run-2 data," tech. rep., CERN, Geneva, 2016. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-DAQ-PUB-2016-001.

[70] M. Aly and et al, "Trexfitter documentation," 2025. https://trexfitter-docs.web.cern.ch/trexfitter-docs/.

[71] ATLAS Collaboration, "Topological cell clustering in the atlas calorimeters and its performance in lhc run 1," 2017.

[72] ATLAS Collaboration, "Jet reconstruction and performance using particle flow with the atlas detector," *The European Physical Journal C*, vol. 77, July 2017.

[73] A. Onofre, N. F. S. F. Castro, and A. Collaboration, "Performance of pile-up mitigation techniques for jets in pp collisions at s=8 TeV using the atlas detector," *The European physical journal. C, Particles and fields*, vol. 76, no. 11, 2016.

[74] P. Berta, L. Masetti, D. Miller, and M. Spousta, "Pileup and underlying event mitigation with iterative constituent subtraction," *The journal of high energy physics*, vol. 2019, no. 8, pp. 1–36, 2019.

[75] M. Cacciari, G. P. Salam, and G. Soyez, "Fastjet user manual: (for version 3.0.2)," *The European physical journal. C, Particles and fields*, vol. 72, no. 3, pp. 1–54, 2012.

[76] F. Aurenhammer, "Voronoi diagrams—a survey of a fundamental geometric data structure," *ACM computing surveys*, vol. 23, no. 3, pp. 345–405, 1991.

[77] A. Onofre, N. F. Castro, and A. Collaboration, "Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the atlas detector," *The European physical journal. C, Particles and fields*, vol. 81, no. 8, 2021.