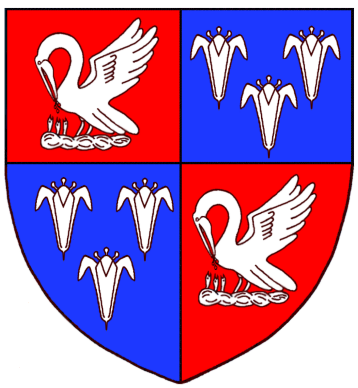
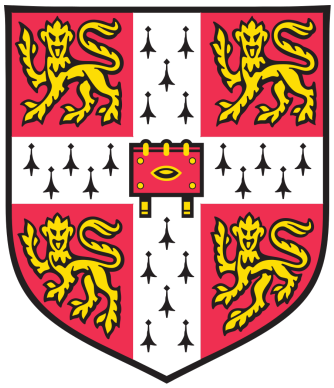
Machine learning based searches for new physics at the ATLAS experiment

Daniel Noel

Corpus Chrisi College
University of Cambridge





CERN-THESIS-2022-213
12/09/2022

This thesis is submitted for the degree of

Doctor of Philosophy

July 2022

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee

Daniel Noel

Abstract

A primary goal of the ATLAS experiment at the LHC is to discover new physics. In recent years, however, such discoveries have been scarce, creating a need for more sophisticated analysis techniques to probe for new physics in more intricate ways. This thesis focuses on two searches for new physics using machine learning techniques.

Firstly, a search for electroweak supersymmetry at ATLAS, using the full Run-2 dataset of 139 fb^{-1}at $\sqrt{s} = 13\,\text{TeV}$, is presented. This investigated a particularly challenging region of parameter space where there is a low mass-splitting between the supersymmetric particles considered – the lightest chargino and the lightest neutralino. Using a boosted decision tree to perform multiclass classification, separate regions in phase space can be defined that are enriched in either signal or a certain background. These regions are used to search for the supersymmetric signals and for improving the background modelling, respectively. Exclusion limits are set on the masses of the charginos and neutralinos, which cover an important gap in sensitivity between previous searches.

Secondly, a novel method for performing model-independent searches for parity-violating new physics is presented. This analysis method uses convolutional neural networks which are parity-odd by construction. Asymmetries in the parity-odd output of the convolutional neural network indicates that there is parity-violation

in the dataset. The efficacy of this method is demonstrated using a simplified model of parity-violating physics. Future searches using this method on data at the LHC allows for the investigation of previously unexplored forms of parity-violating physics at the energy scales of the LHC.

Acknowledgements

Firstly, I would like to thank my supervisor Tina Potter for the constant support and enthusiasm over the past years. Your guidance has made the whole PhD a very enjoyable experience! I would also, like to thank the Science and Technology Facilities council for funding me for the PhD, allowing me to carry out this research.

Thank you to everyone I collaborated with during the PhD. I would like to thank the entire 2L0J team, in particular Sonia Carrà, Margherita Primavera, Jordi Sabater, Eric Ballabene, Matteo Greco, Francesco Gravili and Gabriel Gallardo, and the parity team Chris Lester, Rupert Tombs and Radha Mastandrea. I would like to thank the entire Cambridge ATLAS group, in particular Sarah Williams and Holly Pacey for your advice when getting to grip with the two lepton analysis, John Chapman for your constant support during my qualification task, and Rupert Tombs, my perennial office-mate, thank you for all the useful discussions over the years! Also, I would like to thank Alexander Held, for extremely useful email correspondence related to the `cabinetry` package.

I would like to thank all my friends over the years, in particular those who I have lived with during the PhD: Harry Jones, Francesca Dakin, Sophia Junker, Kat Mitchell, Pim Verschuuren, Jonas Wuerzinger, James Shinner, Tanmay Sukthankar, Sophie Oldroyd, Becky Rickman, Silas Hope, Ben Farrar and Will Moody. It was always great knowing I didn't have to leave the house to have a laugh (especially when we couldn't leave the house!). Especially, I would like to thank Isabella Tarizzo for always supporting me throughout this PhD journey.

Finally, I want to thank my family for always supporting me throughout my life, and always being there for me. I dedicate this thesis to my mum, I know you're looking down on me smiling.

Preface

This thesis details my contributions to the search for new physics. To this end, I performed analyses both within and externally to the ATLAS collaboration. The thesis begins with introductory chapters: Chapter 1 presents a summary of the theoretical background relating to the Standard Model, issues with the Standard Model and supersymmetry as a theory which can resolve many of these issues. Chapter 2 then outlines relevant machine learning concepts used for the later analyses. Finally, Chapter 3 describes the ATLAS detector, and the methods used to reconstruct signals in the ATLAS detector to the physics objects ultimately used in the analyses.

Chapter 4 describes a search for electroweak supersymmetry, which the author performed as part of the ATLAS 'Two Lepton Zero Jet' supersymmetry team. This is a search for charginos decaying to neutralinos via W-bosons in a particularly challenging region of parameter space where the mass difference between the lightest chargino and the lightest neutralino is low. A boosted decision tree performing multiclass classification is used to separate the signal and different Standard Model backgrounds. This allows sensitivity to be attained for the supersymmetric signals, as well facilitating the construction of regions to better control background modelling. The author played a leading role in the analysis, designing and implementing the analysis. Contributions performed by collaborators are explicitly noted in the text. This analysis was initially made public as a conference note in Ref. [1], and subsequently as a paper in Ref. [2].

Chapter 5 describes a novel analysis technique for performing model-independent searches for parity-violating new physics. By training convolutional neural networks that are constructed to be parity-odd, parity-violating effects can be determined in

a dataset by the presence of asymmetries in the parity-odd output. The efficacy of this technique is demonstrated on a simplified model of parity-violating physics. This analysis was performed as part of a small team at Cambridge, where the author was solely responsible for the convolutional neural network development, the showering in PYTHIA and detector reconstruction in DELPHES, as detailed in the text. Contributions from collaborators are explicitly noted in the text. This analysis was made public in Ref. [3].

During the PhD, the author completed additional projects. The author was part of ATLAS 'Four lepton' supersymmetry team, which searched for supersymmetry in events with four or more leptons. The author designed control regions for improving the modelling of the dominant irreducible backgrounds, ZZ and $t\bar{t}Z$, and the author aided with the evaluation of the theoretical systematic uncertainties. This analysis was published in Ref. [4]. In addition, the author performed a qualification task and service work on the ATLAS simulation code. The author was involved with the re-writing of the Athena configuration code, to a new 'ComponentAccumulator' approach. This work focused on the configuration code that runs the simulation of the ATLAS detector. The simulation runs in the new style code developed, and is confirmed to replicate the output of the currently used configuration. All ATLAS simulation jobs plan to use this new configuration approach, starting during Run 3.

Contents

Chapter 1.

Theoretical background

"In this briefcase, I carry actual scientific facts. A briefcase of facts, if you will. Some of these fossils are over 200 million years old."

— Ross Geller

The Standard Model of particle physics is a theory developed over the 20th century which describes three of the four fundamental forces – the electromagnetic force, the weak force and the strong force – as well as all known elementary particles. Particle colliders were introduced in the 20th century, allowing the theory to be investigated at higher energies. Experimentally, the Standard Model agrees remarkably well with data over many orders of magnitudes [5]. Moreover, as of 2012, all particles predicted by the Standard Model have been observed in nature [6, 7].

This chapter begins with a brief review of the Standard Model, introducing the discrete parity symmetry, which was surprisingly found to be violated in the Standard Model [8]. Then, shortcomings of the Standard Model are discussed, with supersymmetry introduced as a theory that can explain many of the current gaps in the Standard Model.

1.1. The Standard Model

Symmetries are a fundamental concept for the Standard Model. Physics is symmetric when it is invariant under a certain transformation, and importantly, by Noether's theorem [9], continuous symmetries lead to conserved quantities. For example, spatial translational symmetry leads to conservation of momentum.

The Standard Model is a quantum field theory. Fields are the fundamental objects, and the physically observable elementary particles are excitations of such fields. The Standard Model is built upon two types of symmetries. First, *space-time symmetries* correspond to symmetries that act on the space-time coordinates. These are described by the Poincaré symmetry group of special relativity, which categorises space-time translations, rotations and boosts. Second, *internal* symmetries correspond to transformations of fields. The Standard Model is symmetric under local gauge transformation of the $SU(3)_C \times SU(2)_L \times U(1)_Y$. Quantum Chromodynamics (QCD), which describes the strong interactions, is defined by the $SU(3)_C$ group. The corresponding conserved charge is the colour charge, indicated by the subscript letter 'c'. The electroweak forces – weak and quantum electrodynamics (QED) – are defined by $SU(2)_L \times U(1)_Y$, with corresponding charges the third component of weak isospin I_3 and weak hypercharge Y[1]. Moreover, these gauge symmetries give rise to interacting gauge bosons, which mediate forces between particles with the corresponding charge.

Particles in the Standard Model come in two categories based on their spin: fermions and bosons. Fermions have half integer spin (in units of $\hbar$), whereas bosons have integer spin. Fundamental fermions are spin-1/2 and are the particles that make up matter. There are two types of fundamental fermions; *quarks* which feel the strong force, and *leptons* which do not. The quarks come in pairs with electric charge (+2/3, -1/3), and there are three generations of quarks with increasing mass. The leptons also

[1] Note that the $SU(2)_L \times U(1)_Y$ group is spontaneously broken by the Higgs mechanism [10, 11] to $U(1)_Q$. This results in the electric charge $Q = I_3 + \frac{1}{2}Y$ being the overall conserved charge.

come in pairs with electric charge (-1, 0), where the uncharged leptons are known as neutrinos. There are similarly three generations of leptons. Table 1.1 shows the Standard Model fermions, including each of the three generations.

The fundamental bosons consist of the spin-1 gauge bosons and the spin-0 Higgs boson. The gauge bosons that correspond to the $SU(3)_c$ symmetry are the *gluons*. There are eight gluons which are massless and couple to colour charge, which typically is labelled red, blue and green. The gluons also carry colour charge, so can self-couple to undergo self-interactions. QCD has the interesting property that the coupling constant α_S decreases as the momentum scale increases (corresponding to smaller distances). This gives rise to the property of *asymptotic freedom*, where at short distances quarks and gluons are essentially free particles. In addition, coloured particles have the property of *confinement*, since the coupling is very large at high distances coloured particles are never seen to be isolated. Only colourless particles can be observed, these are hadrons consisting of two or more quarks.

The electroweak interactions are based on the $SU(2)_L \times U(1)_Y$ symmetry group. The gauge bosons corresponding to this are W^1, W^2, W^3 from $SU(2)_L$ and B from $U(1)_Y$. These are necessarily massless to ensure gauge invariance, so cannot be the physical bosons that have been observed. The Higgs mechanism causes electroweak symmetry breaking, where these symmetries are spontaneously broken $SU(2)_L \times U(1)_Y \rightarrow U(1)_Q$. This results in three massive gauge bosons ($W^\pm$, Z) and one massless gauge boson (the photon γ). The Z-boson and photon are mixtures of the W^3 and B, and the $W^\pm$ are mixtures of W^1, W^2. The mass of the $W^\pm$, Z-bosons causes a short range effect, giving the appearance of a *weak* force. On the other hand, the massless photon has an infinite range carrying the electromagnetic force. Fermions in the Standard Model are in two chiral representations, left or right-handed. The $SU(2)_L$ bosons only couple to the left-handed fermions, indicated by the letter 'L' subscript. So, the $W^\pm$ only couple to left-handed particles and the Z-boson couples preferentially to left-handed fermions, however still couples to right-handed fermions

Fermion	Generation 1	2	3	Y	I_3	Q
Leptons	$\begin{pmatrix}\nu_e\\ e^-\end{pmatrix}_L$	$\begin{pmatrix}\nu_\mu\\ \mu^-\end{pmatrix}_L$	$\begin{pmatrix}\nu_\tau\\ \tau^-\end{pmatrix}_L$	-1	$\frac{1}{2}$ $-\frac{1}{2}$	0 -1
	e_R^-	μ_R^-	τ_R^-	-2	0	-1
Quarks	$\begin{pmatrix}u\\ d\end{pmatrix}_L$	$\begin{pmatrix}c\\ s\end{pmatrix}_L$	$\begin{pmatrix}t\\ b\end{pmatrix}_L$	$\frac{1}{3}$	$\frac{1}{2}$ $-\frac{1}{2}$	$\frac{2}{3}$ $-\frac{1}{3}$
	u_R	c_R	t_R	$\frac{4}{3}$	0	$\frac{2}{3}$
	d_R	s_R	b_R	$-\frac{2}{3}$	0	$-\frac{1}{3}$

Table 1.1.: The fermions of the Standard Model, and their electroweak quantum numbers. Y is the weak hypercharge, and I_3 is the third component of weak isospin. Q is the electric charge and is calculated as $Q = I_3 + Y/2$.

since it also contains a contribution from B. Table 1.1 includes the electroweak quantum numbers for the Standard Model particles.

The mass eigenstates of the quarks do not correspond to the weak interaction eigenstates. Interactions involving the $W^\pm$ bosons can involve terms which go across quark generations, for example $W^+ \to u\bar{s}$. This is because the quark states can mix, as described by the CKM matrix [12, 13]. The CKM matrix is parameterised by three mixing angles and a complex phase. The values of these four parameters are not predicted theoretically, so must be experimentally determined.

The Higgs mechanism causes electroweak symmetry breaking, giving masses to the $W^\pm$ and Z-bosons, as previously noted. The Higgs mechanism introduces the Higgs field: a complex SU(2) doublet, which has four degrees of freedom. When electroweak symmetry is broken, the Higgs field collapses to a vacuum state, and one degree of freedom gains a non-zero vacuum expectation value. The remaining

three of these degrees of freedom are "eaten" by the massive gauge bosons, to give the $W^{\pm}$ and Z-bosons longitudinal polarisation. The degree of freedom with a non-zero vacuum expectation value is a spin-0 particle, the Higgs boson. The mass of the Higgs boson is not predicted by the Standard Model, so must be experimentally determined.

Fermion mass terms are also forbidden by gauge invariance, so without the Higgs field all fermions would be massless. Fermions can couple to the Higgs field by Yukawa interactions, which gives masses to the fermions. These couplings are proportional to the mass of the fermion, and so are greatest for the top-quark, the most massive fermion. These masses are also not predicted by the Standard Model.

The parity transformation is defined by a spatial inversion through the origin,

$$P : \vec{x} \to -\vec{x}, \tag{1.1}$$

and it follows similarly that $\vec{p} \to -\vec{p}$. Parity is a symmetry for both the QED and QCD interactions. It was expected for parity to be a symmetry of the Standard model, however, Wu [8] observed that parity is violated by the weak interaction. Her experiment considered the β-decay of cobalt-60, which is driven by the weak interaction. The cobalt was polarised, with the spins aligned by a magnetic field. By measuring the rate of electron emission in different directions, parity was observed to be violated. This experiment can be formulated [14] in terms of a parity-odd measurement function $\vec{n} \cdot \vec{B}$, where $\vec{n}$ is the direction of the emitted electrons and $\vec{B}$ is the applied magnetic field. An asymmetry in this parity-odd function indicated that parity is violated. Overall, this experiment led to the formulation of the weak interaction as only interacting with the left-handed chiral fermions.

1.2. Issues with the Standard Model

Despite the numerous successes of the Standard Model, there are several issues with it from theoretical and experimental perspectives. These require new physics *beyond*

the Standard Model to be rectified. A selection of these issues will be reviewed in this section, and supersymmetry is introduced as a solution to many of the gaps.

Experimentally, neutrinos have been shown to have a non-zero mass, by observing neutrino oscillation [15–17]. There is currently no explanation for the origin of this mass (whether it be a Dirac or Majorana mass term), of the consequential existence of a right-handed neutrino, or of the very low mass scale of neutrinos in the Standard Model. In addition, there is a $4.2\,\sigma$ tension in the measurement of the muon anomalous magnetic moment [18], commonly referred to as $(g-2)_\mu$. Contributions from physics beyond the Standard Model could resolve this tension. Moreover, tests of lepton flavour universality in the b-sector have shown numerous discrepancies with the Standard Model [19–22]. These tests consider the transition $b \to s$ (or $\bar{b} \to \bar{s}$), in addition to the production of either a pair of electrons or muons. Interestingly, each result shows the "same deficit of muonic decays relative to electronic decays" [22], indicating that there could be a common cause of the discrepancies.

Theoretically, unifying all the Standard Model forces into one single force, with one coupling constant, is very desirable. Grand unified theories (GUTs) propose having a larger gauge group, such as SU(5) [23], in which the $\mathrm{SU(3)_C} \times \mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$ group structure of the Standard Model would be embedded. Looking at how the gauge couplings vary with energy in Figure 1.1, in the Standard Model (dashed lines), the couplings almost converge, but not exactly. With supersymmetry, the couplings can converge at high energies, which could indicate that supersymmetry is involved in a GUT.

Dark matter

A multitude of astronomical observations indicate that in the universe there is an abundance of massive matter which is only gravitationally interacting (non-luminous), known as *dark matter* [25, 26]. Dark matter accounts for $\sim$ 83% of the matter content

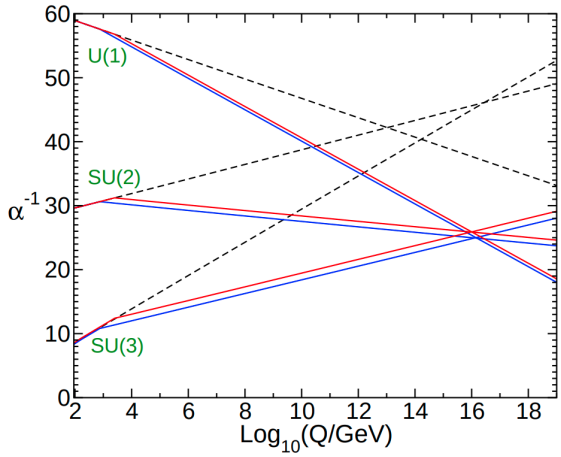


Figure 1.1.: The evolution of the U(1), SU(2) and SU(3) couplings with energy, the dashed lines show the Standard Model, and the solid lines the MSSM. Reproduced from Ref. [24].

in the universe, however it is not accounted for by the Standard Model since there is no suitable candidate particle.

Numerous experimental observations independently point to the existence of dark matter. Studies of the velocity distribution of stars (*rotation curves*), gravitational lensing, the kinematics of galactic collisions (*bullet cluster*) and numerical simulations indicate that the mass in galaxies is not concentrated in the centre. Galaxies are instead surrounded by a *halo* of dark matter, which account for a large fraction of the mass of the galaxy. Moreover, fluctuations in the cosmic microwave background (CMB) indicate that the matter density of dark matter is ~ 5 times as large as the Standard Model matter density. All these experimental observations necessitate the same conclusion; there is a new form of matter unexplained by the Standard Model. This dark matter only interacts gravitationally, is stable and is non-relativistic (cold).

In the early universe, Standard Model particles and dark matter particles were in thermal equilibrium – dark matter particles would have been created by annihilation of Standard Model particles, and Standard Model particles would have been created by annihilation of dark matter particles. As the universe expanded and the temperature cooled, the rate of dark matter particle annihilation reduced, and the amount of dark matter in the universe approached a constant amount, the *relic density*. This process

is commonly referred to as *freeze-out* of dark matter particles. A weakly interacting massive particle (WIMP), with mass around the electroweak scale, can explain the observed relic density. These energy scales are accessible at the LHC, motivating searches for the production of WIMPs as the LHC.

The hierarchy problem

A theoretical issue with the Standard Model is the hierarchy problem, relating to the dynamical stability of the Higgs boson mass and naturalness. The Higgs boson, as a fundamental scalar, receives quadratic corrections to its mass which are related to the highest scale Λ in the theory. New physics is expected at the Planck Scale, $M_P = \sqrt{\frac{\hbar c}{G}} \approx 10^{18}\,\text{GeV}$, where quantum gravity effects become important. New physics could provide very large corrections to the Higgs boson mass, resulting in it being far greater than the observed value of 125 GeV. It is technically possible to have a very precise cancellation of contributions to the Higgs boson mass over the many orders of magnitude between electroweak scale M_W and M_P, however this level of *fine-tuning* is undesirable. Moreover, the quadratic sensitivity of m_H to a high energy scale of new physics is "rather disturbing" [24].

The corrections to the Higgs boson mass can come from fermions and bosons, with example Feynman diagrams shown in Figure 1.2. These contributions to the Higgs boson mass are both quadratic, but fermions and bosons contribute with opposite signs, due to a factor of –1 picked up by a fermion loop. If there was a symmetry relating fermions and bosons, such that every fermion has an equivalent boson (and vice versa), then there would be an exact cancellation of the quadratic contributions over all energy scales. This symmetry is known as *supersymmetry*, and is discussed in the following section.

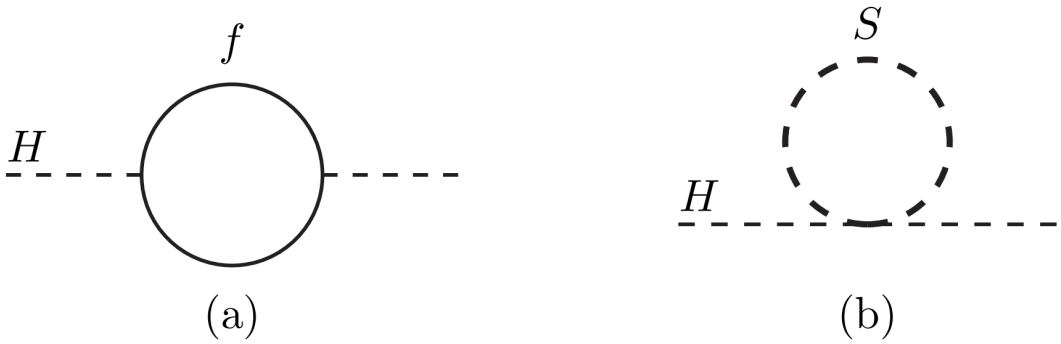


Figure 1.2.: Feynman diagrams of corrections to the Higgs boson mass. Figure taken from Ref. [24].

1.3. Supersymmetry

Symmetries are fundamental to the Standard Model, so a logical way of extending the Standard Model is by introducing new symmetries. Supersymmetry, commonly known as SUSY, is a proposed new symmetry of nature based on spin, which relates fermions and bosons. For each Standard Model fermion (boson), there is a corresponding boson (fermion) which has identical quantum numbers, differing only in spin. These particles are known as superpartner particles, or sparticles. If this were an exact symmetry, the quadratic sensitivity of the Higgs boson mass to Λ would be exactly cancelled, resolving the aforementioned hierarchy problem.

The minimal supersymmetric Standard Model (MSSM) introduces the fewest extra particles while being phenomenologically viable. The particles and their superpartners reside in supermultiplets, which consist of two components; a bosonic and fermionic component. There are two types of supermultiplets: chiral supermultiplets, which consist of two spin-0 bosons and a spin-1/2 fermion; and gauge supermultiplets, which consist of a massless spin-1 boson and a spin-1/2 fermion. The Standard Model fermions, quarks and leptons, reside in chiral supermultiplets. They have spin-0 superpartners, the squarks and sleptons, where the prepended 's' stands for scalar. The Standard Model bosons reside in gauge supermultiplets, and so have spin-1/2 superpartners. The electroweak gauge bosons before electroweak symmetry breaking, W^+, W^-, W^0 and B, have superpartner particles the *winos* and *bino*, $\widetilde{W}^+$, $\widetilde{W}^-$, $\widetilde{W}^0$

Chiral supermultiplets	spin 0	spin 1/2
squarks, quarks	$(\widetilde{u}, \widetilde{d})_L$	$(u, d)_L$
	$\widetilde{u}_R$	(u_R)
	$\widetilde{d}_R$	(d_R)
sleptons, leptons	$(\widetilde{\nu}_e, \widetilde{e})_L$	$(\nu_e, e)_L$
	$\widetilde{e}_R$	(e_R)
Higgs, higgsinos	(H_u^+, H_u^0)	$(\widetilde{H}_u^+, \widetilde{H}_u^0)$
	(H_d^0, H_d^-)	$(\widetilde{H}_d^0, \widetilde{H}_d^-)$
Gauge supermultiplets	**spin-1/2**	**spin-1**
gluino, gluons	$\widetilde{g}$	g
winos, W bosons	$\widetilde{W}^\pm, \widetilde{W}^0$	$W^\pm, W^0$
bino, B boson	$\widetilde{B}$	B

Table 1.2.: The Standard Model particles and corresponding supersymmetric particles in the MSSM. Only the first generation of quark and leptons are shown.

and $\widetilde{B}$. The gluon has the superpartner particle, the *gluino*, $\tilde{g}$. In the MSSM, there is an additional Higgs doublet, and the Higgs fields reside in chiral supermultiplets. These have spin-1/2 superpartners, the *higgsinos*, $\widetilde{H}_d^0$, $\widetilde{H}_d^-$, $\widetilde{H}_u^+$, and $\widetilde{H}_u^0$. The two Higgs doublets, after electroweak symmetry breaking, result in five predicted Higgs boson mass eigenstates. The lightest of these corresponds with the Standard Model Higgs boson. The particle content of the MSSM is summarised in Table 1.2.

The winos, bino and higgsinos mix to form the mass eigenstates, which would be ultimately detected. The charged winos and charged higgsinos mix to form *charginos*, $\tilde{\chi}_i^\pm$ $(i = 1, 2)$, and mixing of the neutral wino, bino and neutral higgsinos forms *neutralinos*, $\tilde{\chi}_i^0$ $(i = 1, 2, 3, 4)$. The labelling is in ascending mass order, such that $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_1^0$ are the lowest mass chargino and neutralino, respectively. The charginos and the neutralinos are electroweak sparticles that are searched for at the LHC, for example as in the search described in Chapter 4.

If supersymmetry was an exact symmetry, the sparticles would have identical masses to the Standard Model particles, and the low mass sparticles would have been detected long ago. Therefore, it must be a broken symmetry, which would result in the supersymmetric particles having a greater mass than their Standard Model counterparts. This breaking for the MSSM can be parameterised by introducing terms to the supersymmetric Lagrangian that explicitly break supersymmetry, whilst retaining the quadratic cancellation of the Higgs boson mass. This *soft* supersymmetry breaking introduces 105 new parameters [24].

Allowing terms in the supersymmetric Lagrangian that violate both baryon number B or lepton number L results in rapid proton decay by $p \to e^+ + \pi^0$. In order to prevent this, the multiplicatively conserved R-parity is introduced, defined as

$$R \equiv (-1)^{3(B-L)+2s} = \begin{cases} +1 & \text{Standard Model particles} \\ -1 & \text{Superpartner particles} \end{cases} \tag{1.2}$$

Models that conserve this are known as R-parity conserving (RPC) supersymmetry, which have characteristic features. Firstly, the lightest supersymmetric particle (LSP) is stable, since R-parity means it cannot decay into only Standard Model particles. Moreover, it should be electrically neutral, since no stable sparticle has been detected. In addition, vertices in Feynman diagrams consist of a pair of sparticles, so sparticles are produced in pairs in particle colliders and then each undergo a cascade of decays into lighter supersymmetric particles, until the LSP is formed. The final state consists of an even number of LSPs, typically two, and Standard Model decay products. These LSPs are assumed to be weakly interacting and consequently go undetected by the detector, giving the characteristic signature of RPC supersymmetry: large amounts of missing momentum in the final state. A very attractive feature of RPC theories is that, if the LSP is weakly interacting such as the $\tilde{\chi}_1^0$, it could be a viable WIMP dark matter candidate explaining the observed relic density.

Models with R-parity violation (RPV) are also viable if L or B are individually conserved. These models do not produce a dark matter candidate, but can solve the hierarchy problem, and provide a further way to search for supersymmetry, as in Ref. [4].

The parameter space in which to search for supersymmetry is vast, and so in order to guide our searches, we use *simplified models* [27]. These focus on a particular production and decay process, and make simplifying assumptions. These can help guide analysis strategies, providing a final state to be targeted for optimisation. Typically, only two sparticles have a low enough mass to be detectable, with all other sparticles having a much larger mass. Moreover, only certain decay modes are considered typically with a 100% branching fraction, with the other couplings set to zero. The neutralinos and charginos are taken to be unmixed, for example in Chapter 4 we consider a pure-wino $\tilde{\chi}_1^{\pm}$ and a pure-bino $\tilde{\chi}_1^0$. The results for simplified models can be visualised in the two-dimensional plane spanned by the two free mass parameters, where gaps in sensitivity can guide future searches.

Overall, the benefits from the extra particles introduced with supersymmetry are numerous, both from a theoretical and experimental perspective. Theoretically, supersymmetry can naturally resolve the hierarchy problem with cancellations arising from sparticles and can provide convergence of the gauge couplings at high energies, indicating it could be a part of a GUT for unifying the forces. Experimentally, supersymmetry with R-parity conservation can provide a dark matter candidate, in particular a $\tilde{\chi}_1^0$ could be the WIMP that provides the observed dark matter relic density. Currently, no supersymmetric particles have been observed at the LHC [28, 29]. Supersymmetry has been excluded in the most experimentally accessible areas of parameter space, so further searches should be made to probe regions of parameter space which are more experimentally challenging: we could have already produced many sparticles at the LHC, but they have as yet gone undetected since they are produced in an experimentally challenging region of parameter space.

Chapter 2.

Machine learning

"Hasta la vista, baby."

— The Terminator

Machine learning is concerned with creating computer programs that automatically improve with experience [30]. Supervised learning is where we know what the output should look like, since we have prelabelled examples. For example, when performing classification, we assign a discrete category: for example "cat" vs. "no cat" for a photo of a household pet; "0", "1", up to "9" for handwritten images of the numbers from 0 to 9; or "signal" vs. "background" when trying to distinguish a supersymmetric signal from the Standard Model background. In each of these cases, we have prelabelled data indicating the desired output which is to be learned.

For supervised learning, we have input variables $\vec{x}$ and a target output variable y. For *binary* classification there are two outputs classes, and y takes one of two values, typically 1 and 0 or — in the case of distinguishing supersymmetric signal from SM background — *signal* and *background*. For *multiclass* classification, y can take one of many values, for example 0 to 9 in the above example of classifying handwritten images of numbers. For a binary classification problem, logistic regression is a simple

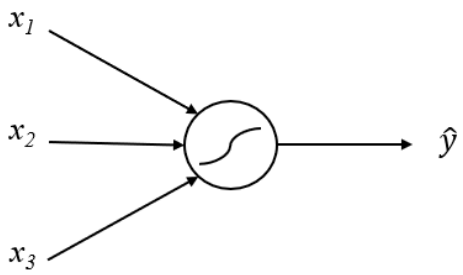


Figure 2.1.: An example logistic unit, used for logistic classification. It has three input variables and outputs a value between 0 and 1 using the sigmoid function.

classifier that can be used with the functional form

$$\hat{y} = \sigma(\vec{w} \cdot \vec{x} + b), \tag{2.1}$$

where $\sigma(z) = 1/(1 + e^{-z})$ is the sigmoid function, outputting values between 0 and 1. The output value is $\hat{y}$, and $\vec{w}$ and b are the tunable parameters of the model, conventionally known as the weights and biases. These can be conveniently combined to a single parameter $\vec{\theta} = (\vec{w}, b)$. Equation 2.1 is illustrated in Figure 2.1.

The aim is to learn a function $\hat{y} = f(x)$ such that it is a good predictor for y, meaning that the distribution of $\hat{y}$ closely matches the distribution of the truth values y. This is done by calculating a *loss*-function, which measures the error between a given prediction $f(x_i)$ and the true value y_i. By tuning the parameters $\vec{\theta}$ to minimise the loss, the model can learn a relationship between x and y.

The logistic loss function is typically used. For binary classification, this is known as binary cross entropy and is calculated as

$$\mathcal{L}(\hat{y}, y) = -\frac{1}{N} \sum_{i=1}^{N} \Big(y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \Big), \tag{2.2}$$

where the sum is over each training example from a total of N instances. This has the effect of maximising the likelihood of the predictions matching the true labels, ensuring the minimum distance between the output $\hat{y}$ and target underlying distribution y. For

multiclass classification, the logistic loss formula with multiple classes is,

$$\mathcal{L}(\hat{y}, y) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c} y_c \log \hat{y}_c, \tag{2.3}$$

where the second sum is over each output category c. Note that Equation 2.2 is a special case of this logistic loss with two classes. Additional terms can be added to the loss, for example the sum of the squares of the model weights $\lambda \sum_{i=1}^{N} w_i^2$, where λ controls the effect of this term. This term provides L2-regularisation by reducing the magnitude of the w_i and so helps prevent overfitting.

In order to find the minimum of the loss, gradient descent methods are used. For example, $\vec{\theta}$ is updated multiple times in the direction of the negative gradient,

$$\vec{\theta}_{i+1} = \vec{\theta}_i - \alpha \frac{\partial \mathcal{L}}{\partial \vec{\theta}_i}, \tag{2.4}$$

where α is the learning rate, which is set before training. The derivatives can be calculated by evaluating how the loss changes as a function of the parameters. An alternative method of gradient descent is Adam [31], which is used in Chapter 5.

2.1. Neural networks

By combining many logistic units, we can form a neural network, illustrated in Figure 2.2. These typically have numerous layers including, an input layer, multiple hidden layers, and an output layer. The output layer uses the sigmoid function to output a value between 0 and 1. In the hidden layers, the leaky ReLU activation function

$$\text{LReLU}(x) = \max(0, x) + c \times \min(0, x) = \begin{cases} x & \text{if } x > 0 \\ c \times x & \text{otherwise} \end{cases} \tag{2.5}$$

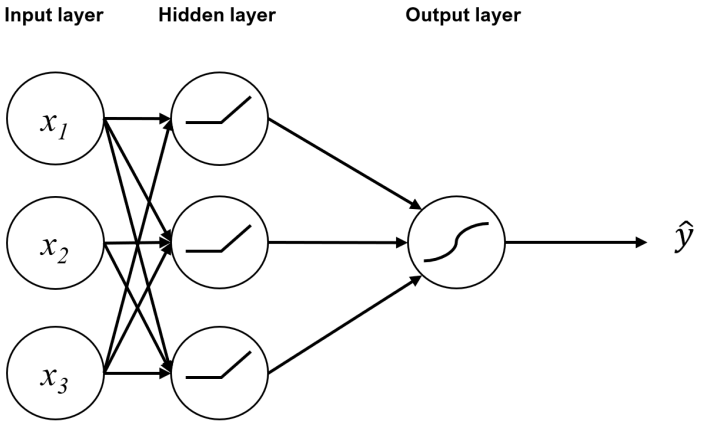


Figure 2.2.: An example simple neural network with one hidden layer and three input variables. The hidden layers are using the ReLU activation function, and the output layer the sigmoid activation to give an output between 0 and 1.

can be used instead of a sigmoid as an alternative way of introducing non-linearity into the network. It is common for c = 0.01 to be used, and when c = 0 we recover the ReLU activation function. Both ReLU and leaky ReLU can avoid issues with vanishing gradients when having lots of sigmoid functions in the hidden layers. These non-linear *activation functions* allow for complicated non-linear relationships to be modelled between the input x and the outputs y. Backpropagation [32] is used to calculate the gradients in each layer and so to update the weights and biases in each layer.

2.2. Convolutional neural networks

Convolutional neural networks (CNNs) are a type of neural networks that are particularly popular for image recognition tasks [33]. These use 2-dimensional (2D) convolutional layers, which operate on 2D matrices with 2D *kernel* functions with learnable weights. These kernels have a shape ($m \times n$), which is smaller than the input 2D matrix shape. At each point, the output is the sum of the input values multiplied

by the kernel weights. This is known as the convolution operator [34]

$$\text{Conv}(i,j) = \sum_m \sum_n M(i-m, j-n) \times K(m,n), \tag{2.6}$$

where K is the 2D kernel function, and M is a 2D input matrix. This operation is illustrated in Figure 2.3. The kernel weights are learned by gradient descent, in order to extract useful features from the input image. By applying many sequential convolutional layers of kernels, more complicated features can be extracted. For example, the earlier convolutional layers can be used to detect edges, and subsequent layers can build up these edges into more complicated overall structures.

Convolutions cause the output matrix to be smaller than the input, for example, in Figure 2.3 we go from a 3×4 matrix to a 2×3 matrix. *Padding* can be applied to the matrix before the convolutions are applied to maintain a constant shape. For example, *zero-padding* can be applied where rows and columns of zeros are added around the edge. *Cyclical-padding*, on the other hand, adds rows and columns that replicate those on the opposite edge.

The kernels operate over the entire image, so the kernel weights are effectively shared over the entire image. Since the kernels are smaller than the overall image, there are fewer weights to be learned than for a fully connected neural networks. Moreover, the convolution operator is equivariant to translation: if an object in the image moves, the output changes in the same way. This is useful for image recognition since the location of the image in the frame is not of interest, rather what the image is.

The pooling operator is often used in conjunction with the convolution. The max-pooling operator returns the maximum value contained within a prespecified area. Firstly, it reduces the dimensionality of the output, increasing the speed of training. In addition, it provides translational invariance, since the output will be unchanged regardless of the object location within the pooling range. In particular, if the pooling operator is performed over an entire axis, the network will be invariant to translations in that direction.

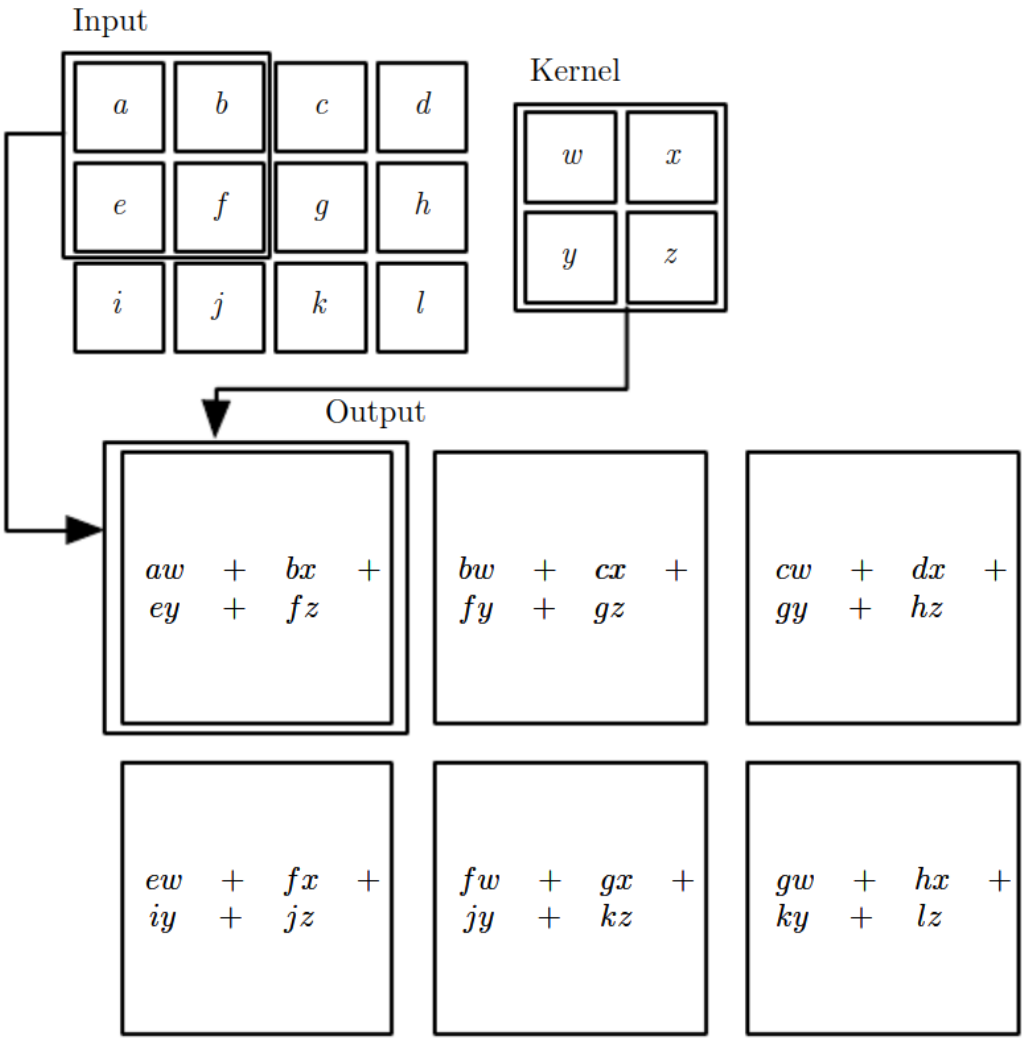


Figure 2.3.: An example of the 2D convolution. The arrows indicate how the upper left output value is calculated from the convolution of the kernel and the upper left part of the input. Figure taken from Ref. [34].

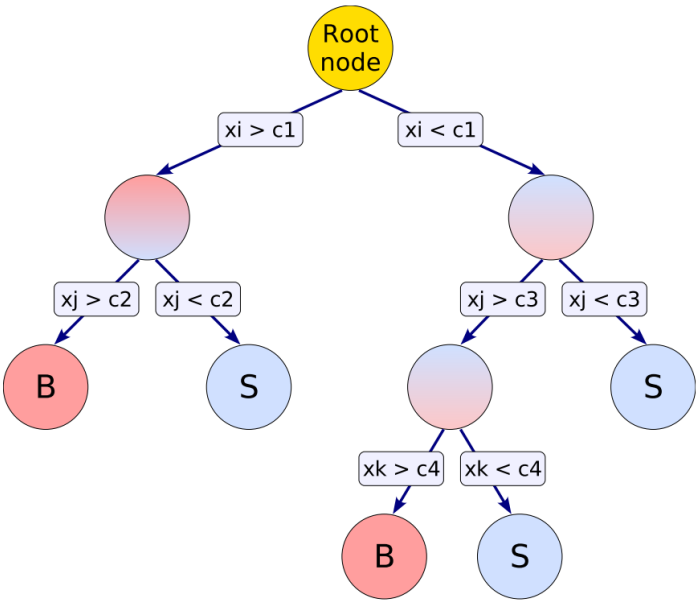


Figure 2.4.: Schematic of a Decision tree. Selections are made on the variables x_i, x_j, x_k to maximise the separation of background B and signal S. The leaf nodes are labelled B or S depending on the majority class in them. Figure taken from Ref. [36].

Frameworks such as PyTorch [35] allow for the creation of both neural networks and convolutional neural networks. It has efficient implementations of the convolutional operator and matrix multiplication, as well as automatically calculating the gradients for backpropagation. It allows the use of GPUs, which vastly speed up computation.

2.3. Boosted decision trees

Decision trees are simple models which separate classes based on simple higher or lower selections on the input variables. In a classification problem, where we seek to separate signal and background, sequential selections are made on the input features to separate into purer regions of signal and background. An example decision tree can be seen in Figure 2.4. For each input example, we can trace the path through the decision tree to determine which terminal node (leaf) it ends up in. Therefore, these models are very interpretable - we can exactly see why an example is labelled as it is.

Gradient boosting is used to combine many decision trees, in order to create a much more powerful overall model. An additive ensemble of decision trees is formed, with each subsequent decision tree trained to minimise the residual errors from the previous trees [37]. This model is known as a boosted decision tree (BDT). These residual errors correspond to negative gradients, so each subsequent tree added is performing a step of gradient descent on a loss function. Adding an additional tree gives a lower loss, and so the boosted model has a better performance.

Modern gradient boosting packages such as XGBoost [38] and LightGBM [39] have been designed to be highly scalable to large datasets, so can be quickly trained. Gradient descent is performed on the logistic losses in Equations 2.2 and 2.3 for binary and multiclass classification, respectively.

While adding a new tree in training, in order to calculate the best splits to be made, all possible splits are evaluated to find the one that minimises the loss. In LightGBM, the variables are binned into histograms before evaluating the splits, which is more memory efficient and faster since there are fewer potential points to make a selection.

By counting the number of times each feature is used in the trees for splitting, a feature importance metric can be derived.

2.4. Model training

Training a model involves finding the minimum of the loss function, while ensuring that we are not overtraining. Inputs are split into disjoint subsets for training and testing. The model is trained over the training set, then the performance of the model is evaluated on the previously unseen test set. This separation is necessary, since if we do not have a disjoint subset for testing, we do not know whether the model has learnt the underlying distribution or is just memorising each example of the training set.

Overtraining and undertraining are commonplace issues when training a machine learning model. Overtraining occurs when the model tunes to fluctuations in the

training set, and does not accurately describe the underlying distribution. Whereas, undertraining is where the model is not sufficiently complicated enough to model the underlying distribution. Training the optimal model involves finding the balance between these two regimes, also known as the bias variance trade-off [40].

When training a model, certain parameters are defined before the training procedure and do not change throughout. These are known as hyperparameters, and their choice is important in ensuring a good training and not being overtrained or undertrained. An important hyperparameter is the learning rate in Equation 2.4. For neural networks, hyperparameters that determine the structure of the network include: the number of layers, the number of nodes per layer, the kernel size, and the number of kernels. The hyperparameters that determine the structure of a BDT include: the number of trees, the maximum number of leaves per tree, and the minimum number of samples per each leaf.

There are no theoretically best set of hyperparameters, so to determine the optimal hyperparameters for the task at hand, we must try numerous combinations and use the best set. This is known as hyperparameter optimisation, where one option is to use a random grid search [41] amongst the hyperparameters we wish to optimise for. For this purpose a third, *validation*, set is used to evaluate the performance of numerous combinations of hyperparameters. The best performing set of hyperparameters are used for the training of the final model, then the ability of the model to generalise to a previously unseen set is calculated on the test set. In Section 5 this is done with a 60, 20, 20 split between train, validation, and test sets.

An alternative to having a validation set is to use K-folds cross validation [40], where the training set is split into K distinct subsets. One subset is used for validation, and the model is trained over the other K-1 subsets. The training is performed K times, such that the entire training dataset is used for validation. For example, when K = 5, we train over the first 4/5th of the data and validate over the final 1/5th. We repeat this process five times, averaging the performance metrics over the different 5 training iterations, as seen in Figure 2.5. This method is slower than the previous method, but

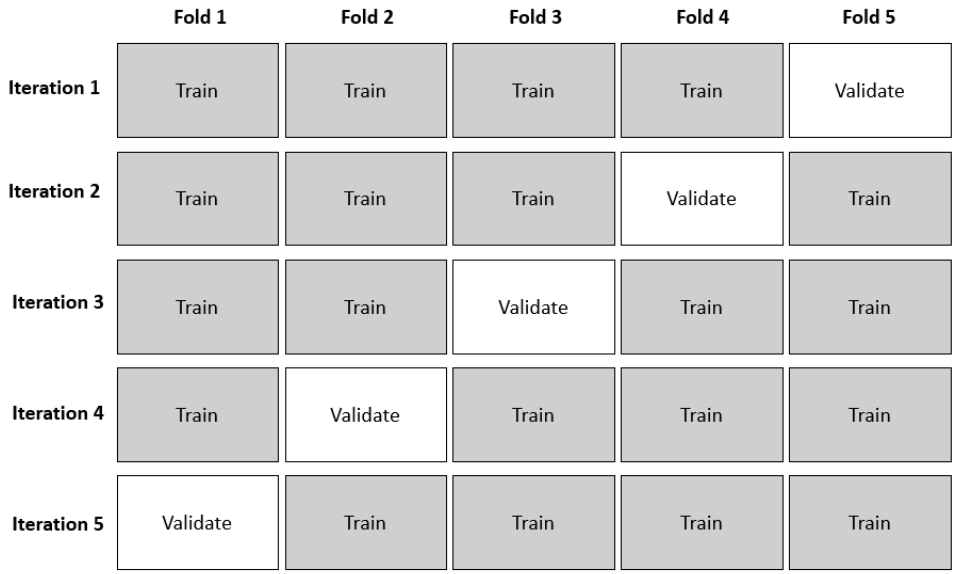


Figure 2.5.: An illustration of cross validation with 5 folds. The dataset is split into 5 folds, with each being used for validation in one of the 5 iterations. This allows for the entire dataset to be used for validation

for BDTs which are quick to train, allows for a more robust hyperparameter selection. Cross validation is used for hyperparameter selection in Section 4.

2.5. Model evaluation

ROC and PR curves

For a classification problem, we get an output distribution of *scores* for the two classes, such as the signal and background. We make a selection at a chosen value of the score and assign everything above the selection as predicted *signal*. For this predicted signal, we have either correctly classified it (*true positive*, TP) or incorrectly classified it (*false negative*, FN). For the predicted background, we have either correctly classified it (*true negative*, TN) or incorrectly classified it (*false positive*, FP). Figure 2.6a illustrates how these four categories are derived from a selection on the output score. Figure 2.6b illustrates how these four categories can be grouped into a *confusion matrix*. Further quantities can be calculated from the confusion matrix. The true positive rate or recall

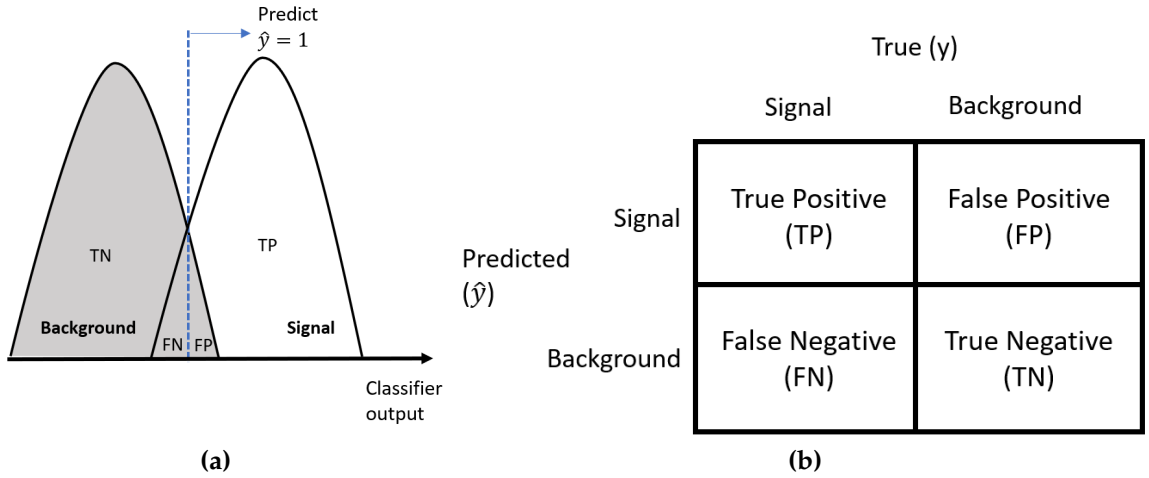


Figure 2.6.: (a) Plots of the classifier output for the background and signal. The dashed line indicates a selection made on the classifier output, where we assign signal to all values above the x-axis value. The regions corresponding to true positive (TP), true negative (TN), false positive (FP) and false negative (FN) values are indicated on the plot. (b) The selection made on the output distribution in left-hand plot, represented as a confusion matrix.

(TPR = recall = $\frac{TP}{TP+FN}$) is the fraction of the truly positive samples correctly classified as positive. The false positive rate (FPR = $\frac{FP}{FP+TN}$) is the fraction of the truly negative samples incorrectly classified as positive. Finally, the precision (precision = $\frac{TP}{TP+FP}$) is the fraction of the predicted positive samples that are truly positive.

Multiple selection values are considered at differing values of the score. For each of these selections, we obtain a different confusion matrix, from which we can calculate the TPR and the FPR. A better model would maximise the TPR whilst minimising the FPR, since this corresponds to higher signal selection and background rejection. Perfect separation corresponds to TPR = 1 and FPR = 0. In reality, there is a balance between the TPR and the FPR, which can be visualised by plotting these variables against each other as a 'receiver operating characteristic' (ROC) curve. The area under the curve, commonly referred to simply as AUC, is a numerical measure of the quality of the model. It corresponds to the probability that the model would rank a randomly chosen signal event higher than a randomly chosen background event. AUC = 1 represents perfect separation of signal and background, and AUC = 0.5 is what is expected from simply randomly guessing for each example.

One drawback of ROC curves is that the TPR and FPR only consider the true positive and true negative samples, regardless of the balance between the numbers of each category. That is, if we have far more negative samples than positive samples, then having a high precision is more important than a high FPR. This is typical when searching for a small supersymmetric signal in a region with a large Standard Model background. So, instead of plotting the FPR, the precision is plotted against the recall (TPR). This is known as a precision-recall (PR) curve, and does not take into account the background which can be easily removed, so can provide more information about the events that are classified as being positive.

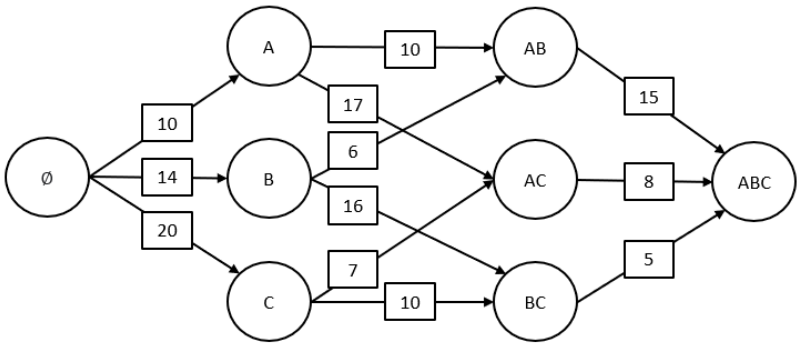
The discussion has been for a binary classification signal vs. background. If we wish to consider ROC and PR curves for a multiclass classification classifier, we can use the one-versus-all procedure. For example, if we are separating four classes $\{1, 2, 3, 4\}$, we can find the ROC curve for class 1 by setting signal as 1 and by grouping together classes $\{2, 3, 4\}$ as background. This procedure is then similarly repeated for each of the other classes.

SHAP values

Shapley values originate in game theory as a way of assigning credit in a cooperative game between many players [42]. For example, calculating how much each Alice, Bob, and Charlie should pay at a restaurant can be done using Shapley values. Figure 2.7 illustrates this, with Figure 2.7a indicating the cost when each subset of Alice, Bob, and Charlie is present. Figure 2.7b shows the marginal contribution when adding a person to form a larger subset. By averaging these marginal contributions over all paths to the complete set {A,B,C} we can calculate the Shapley value for Alice to be $\text{Shapley}(A) = \frac{1}{6}\left(2 \times 10 + 6 + 7 + 2 \times 5\right) = 7\frac{1}{6}$. Similarly, for Bob and Charlie the Shapley values are $10\ \frac{2}{3}$, $17\ \frac{2}{3}$. In this way, we have fairly allocated credit to unequally contributing agents.

Group	Cost
{A}	10
{B}	14
{C}	20
{A,B}	20
{A,C}	27
{B,C}	30
{A,B,C}	35

(a)



(b)

Figure 2.7.: (a) The costs with different combinations of Alice, Bob and Charlie when going to a reasonably priced restaurant, (b) an illustration of how different combinations can be formed starting from the empty set. The numbers attached to each line indicate the marginal cost of adding a certain person to the set, and can be calculated as the difference in the cost of that set and the cost of that set without the added person. By averaging these marginal contributions over all paths to the complete set {A,B,C} we can calculate the Shapley value for each person to be $7\frac{1}{6}$, $10\frac{2}{3}$, $17\frac{2}{3}$ respectively.

For a machine learning model, we can use Shapley values to calculate the contribution of each feature to the overall output score of the model. SHAP (SHapley Additive exPlanations) values [43, 44] provide an efficient implementation of these Shapley values, which is especially quick for tree-based models, such as BDTs. These help us improve in model explainability and transparency, which are important for understanding how machine learning models are working.

For each event, SHAP values are calculated for each feature to indicate the marginal contribution of each feature to the output score from the BDT. Starting with a base value, which is the mean BDT output over the training dataset. Looking at a specific collider event, the SHAP value for a particular variable is calculated by averaging over all permutations of adding the value of the variable to the event.

For example, the SHAP values for a supersymmetric event from Chapter 4 are shown in Figure 2.8. From this breakdown, we can see that for this event, the high E_T^{miss} significance value (METsig = 9) means that the event is more signal like and the

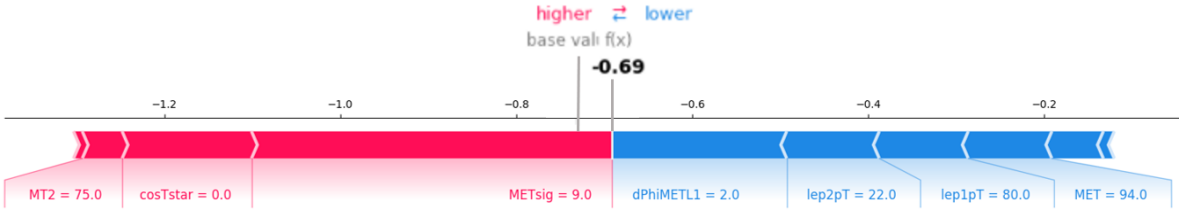


Figure 2.8.: The SHAP values for each variable in a particular collider event, with higher values corresponding to more *signal*-like. The value of the variables in red make the event more signal-like, and the values of the variables in blue make the event more background-like.

low subleading lepton p_{T} value (lep2pT = 22 GeV) means the event is more background-like.

Chapter 3.

The LHC and the ATLAS experiment

"Ogres have layers."

— Shrek

A necessary test of any theory is to verify whether its predictions agrees with data. The Standard Model is no different, and in particular, we require data from high energy particle collisions to both further scrutinise the Standard Model and to provide hints at physics beyond the Standard Model. Collecting such data is no trivial task; ATLAS [45] is an approximately cylindrical detector spanning 25 m $\times$ 44 m in height and length, and weighing around 7000 tonnes. ATLAS is designed to detect decay particles from proton-proton (pp) collisions at extremely high energies. This chapter introduces the Large Hadron Collider (LHC), which accelerates protons in an approximately circular 27 km tunnel, and then describes each layer of subdetectors that constitute the ATLAS detector. Finally, an introduction to the methods used to process the signals in the detector to the physics objects used in the analysis is presented.

3.1. The LHC

The LHC [46] at CERN is the highest energy particle accelerator ever built. It accelerates protons in two counter rotating beams in a 27 km in circumference tunnel that lies on average 100 m underground and spans the Swiss-French border near Geneva. Protons from ionised hydrogen are accelerated up to 450 GeV by a series of accelerators, as shown in Figure 3.1. These protons are then injected into the LHC, where they are subsequently accelerated to 6.5 TeV per beam[1]. The beams contain bunches of $\mathcal{O}(10^{11})$ protons with a spacing of 25 ns between bunches.

The beams are collided at four collision points along the ring, and the produced particles are measured by the four detectors ATLAS [45], CMS [47], LHCb [48] and ALICE [49]. ATLAS and CMS are both general purpose detectors, designed to detect a wide range of particle signatures and have a very broad physics programme. LHCb and ALICE are more specialised detectors, focusing on b-physics and heavy ion physics, respectively. In 2010-2012, the pp collisions took place at $\sqrt{s}$ = 7 TeV and 8 TeV, collectively referred to as Run 1. The pp collisions between 2015 and 2018 took place at $\sqrt{s}$ = 13 TeV. These collisions are referred to as Run 2, and the data used in this thesis were collected during Run 2.

A key parameter when describing the LHC is the luminosity $\mathcal{L}$. This relates the rate of events $\frac{dN}{dt}$ to the event cross-section σ,

$$\frac{dN}{dt} = \sigma \cdot \mathcal{L}. \tag{3.1}$$

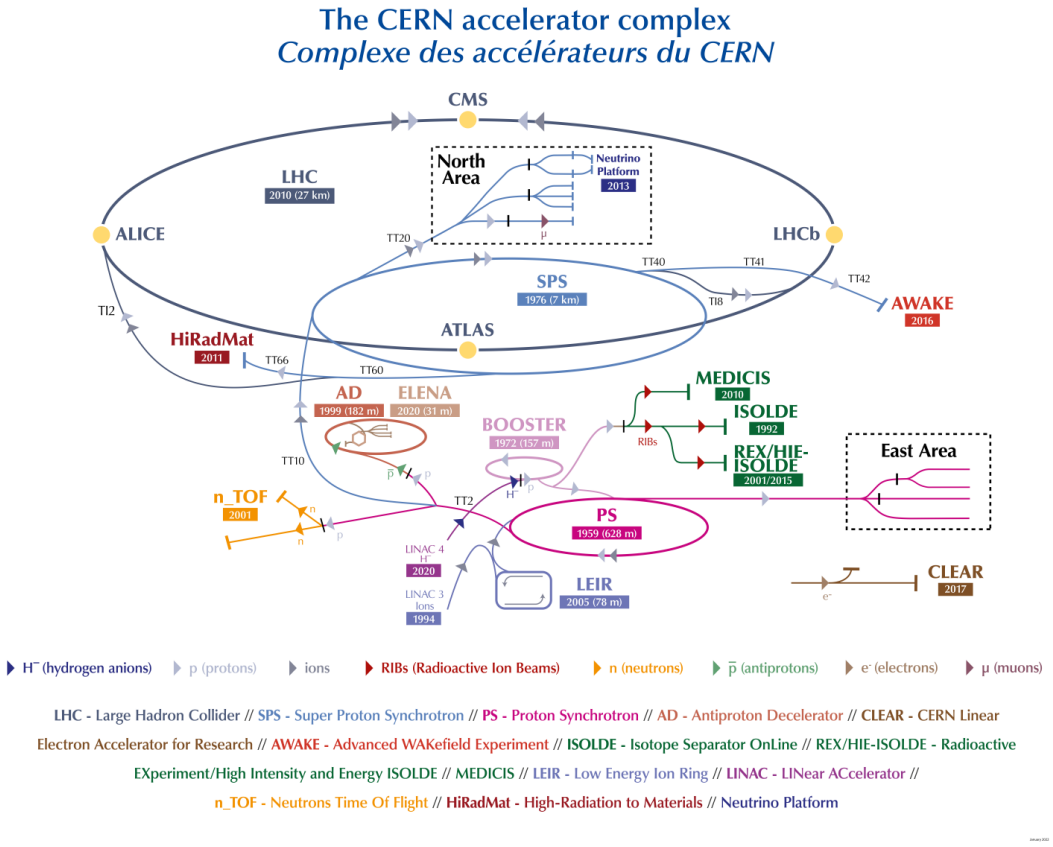
The integrated luminosity

$$L = \int \mathcal{L} dt \tag{3.2}$$

[1] With the start of Run 3, the protons are accelerated to 6.8 TeV.

gives a measure of the total amount of data collected, since it is related to the total number of events $N = \sigma \cdot L$. Higher luminosity is required for producing lower cross-section events in sufficient number to be both detected and studied. Undetected new physics, such as supersymmetry, will have a low cross-section, motivating higher luminosity runs of the LHC. A total of 139 $\text{fb}^{-1}$of data were collected in Run 2 of the LHC [50]. This amount has a 1.7% uncertainty [51], as determined by the LUCID-2 detector [52].

As luminosity increases, there are more proton-proton collisions which occur in addition to the collision of interest. These additional collisions are known as *pile-up*. *In-time* pile-up is from proton-proton collisions in the same bunch crossing, whereas *out-of-time* pile-up is from proton-proton collisions in different bunch crossings. Pile-up can be characterised by the mean number of interactions per bunch crossing $\langle \mu \rangle$. Pileup events tend to have lower momentum (softer) particles, and can be a background to interesting events as well as making the detector reconstruction worse.



(a)

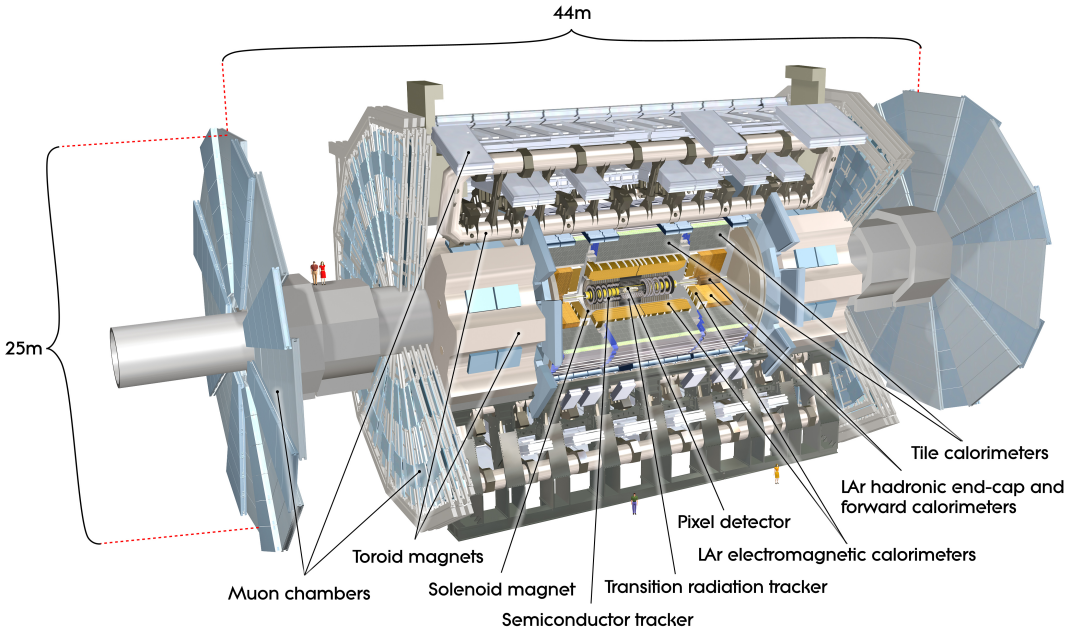
Figure 3.1.: The CERN accelerator complex. In particular, the accelerator chain Linac4, Proton Synchrotron booster, Proton Synchrotron (PS), and Super Proton Synchrotron (SPS) are used to accelerate protons to 450 GeV before injection into the LHC [53].

3.2. The ATLAS detector

The ATLAS detector [45] is designed to detect the signatures of a wide range of particles, and accurately measure their energy and momentum. When commissioned, the main physics goals of the ATLAS detector were the discovery of the Higgs boson, and further investigating new physics scenarios such as supersymmetry. The former was achieved in 2012 [6].

The detector consists of specialised detectors that concentrically surround the beam line. Closest to the beam line is the Inner Detector, providing momentum information derived from the trajectories of charged particles in addition to locating where the collision of interest took place (the primary vertex). Then lie the calorimeters, which measure the energies of particles by stopping them. There is the electromagnetic calorimeter, for measuring particles such as electrons and photons, and the hadronic calorimeter for measuring hadronic particles such as pions. Finally, surrounding all the other detectors is the Muon Spectrometer, which both measures the momenta of and triggers on muons. These are located the furthest from the beam line, since muons can traverse the entire detector. The detectors are immersed in magnetic fields that are produced by a superconducting solenoid magnet for the inner detector, and by toroid magnets for the Muon Spectrometer. The magnetic fields allow for measurement of the momentum of charged particles from the curvature of their trajectories. The layout of the ATLAS detector is shown in Figure 3.2.

Since the detector is approximately cylindrical, we use a cylindrical coordinate system about the beam line at the interaction point. The distance from this axis is measured by r and ϕ measures the azimuthal distance about the axis. The positions in the transverse plane are measured by x and y, with the x-axis pointing towards the centre of the LHC ring and y-axis pointing upwards. In terms of r and ϕ, these are $x = r\cos\phi$ and $y = r\sin\phi$. The z-axis lies along the beamline, completing the right-handed coordinate system. There is an unknown longitudinal boost as a result of partons in the initial state. Therefore, coordinates are used that are invariant to



(a)

Figure 3.2.: A schematic diagram of the ATLAS detector, including all its subdetectors. It spans 25 m $\times$ 44 m in height and length, and weighs around 7000 tonnes. Image taken from [45, 54].

longitudinal boosts. For example, differences in the rapidity $y = \frac{1}{2} \log \frac{E+p_Z}{E-p_Z}$ are Lorentz invariant. The pseudorapidity $\eta = -\ln \tan(\theta/2)$, which is the massless limit of the rapidity, is commonly used as an event variable. Moreover, the momentum in the transverse plane $p_T = \sqrt{p_x^2 + p_y^2}$ is used.

3.2.1. Inner Detector

The Inner Detector [55, 56] is closest to the beam line, extends radially to 1.05 metres and covers a pseudorapidity range of $|\eta| < 2.5$, as shown in Figure 3.3. It consists of three subdetectors; the high-resolution pixel detector, the strip-based semiconductor tracker (SCT), and the transition radiation tracker (TRT). These are immersed in a 2 T magnetic field, produced by a superconducting solenoid [57] to curve the trajectories of traversing electrically charged particles.

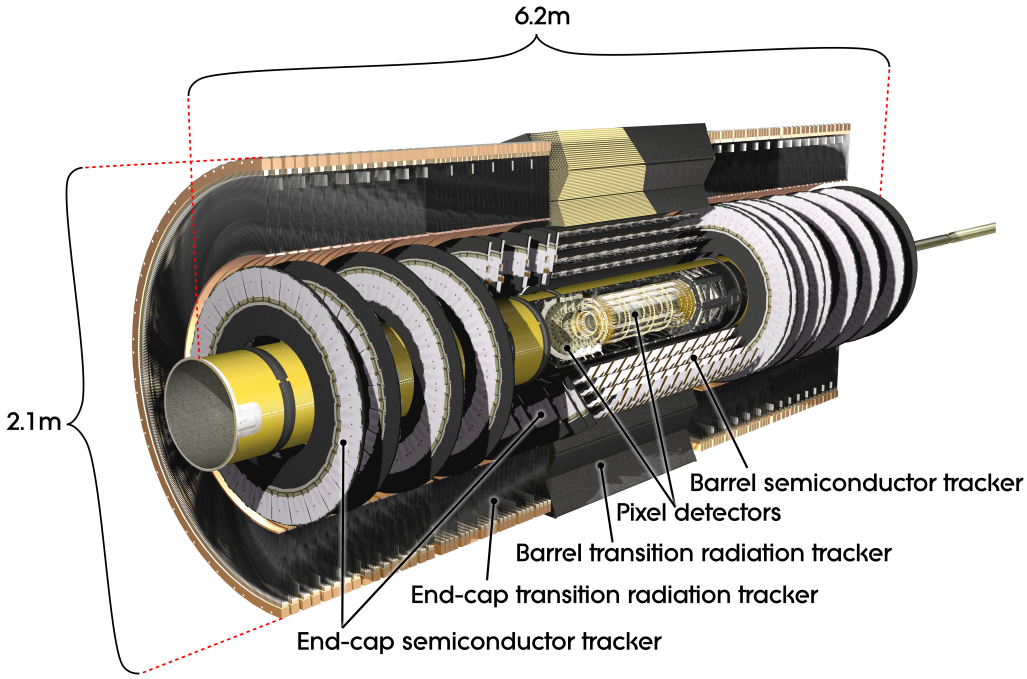
Electrically charged particles leave signals in the detectors, which are reconstructed to infer the trajectory of the particle. Using the curvature of these trajectories, measurements of the momentum of the traversing particle can be made. In addition, by extrapolating back the track, which primary vertex the particle originated from can be determined. Accurate determination of this primary vertex can indicate if there are any displaced vertices. These displaced vertices are where the charged particle is produced at a measurable distance from the interaction point.

The pixel detector [58] lies closest to the beam line, and consists of a total of 80 million silicon pixels of dimension 50×400 μm^2 in $R - \phi \times z$, covering a total active area of approximately 1.7 m^2. These are arranged into three concentric layers, such that at least three hits can be recorded per traversing charged particle. In the end-cap region, they consist of three disks. The pixels are p-n doped silicon, in which electron-hole pairs form in the silicon with the passage of a charged particle. The detection of these electrons and holes indicates where the charged particle travelled.

Between Run 1 and Run 2, a fourth layer was added between the beam-pipe and the original three pixel layers. This Insertable B-Layer (IBL) [59], provides greater tracking resolution at small radii, as well as better primary vertex detection. This improved precision of vertex location helps for identification of different particles, in particular b-jets, as described in Section 3.2.5.

The SCT [60] is located outside of the pixel detectors and also provides precision tracking information. It consists of 4088 silicon strip detectors arranged in four layers, and two endcaps each consisting of nine disks. These are designed such that a minimum of four strip layers will be traversed by a particle. The strips have a lower resolution than the pixels, but are less expensive, so the SCT covers a larger total volume.

Finally, the TRT [61] lies outside of the SCT and is made from straw tubes 4 mm in diameter. These are arranged parallel to the beam axis in the barrel region, and perpendicular to the beam axis in the end-cap region. The straws are filled with a



(a)

Figure 3.3.: A schematic diagram of the Inner detector. Image taken from [45, 62].

mixture of xenon, carbon dioxide and oxygen gasses. A traversing charged particle ionises the gas mixture, and the produced charge is detected and used to track the particle. Typically, there are a large number (~ 36 per track) of hits. This contributes to the momentum measurement, despite having a lower spatial precision.

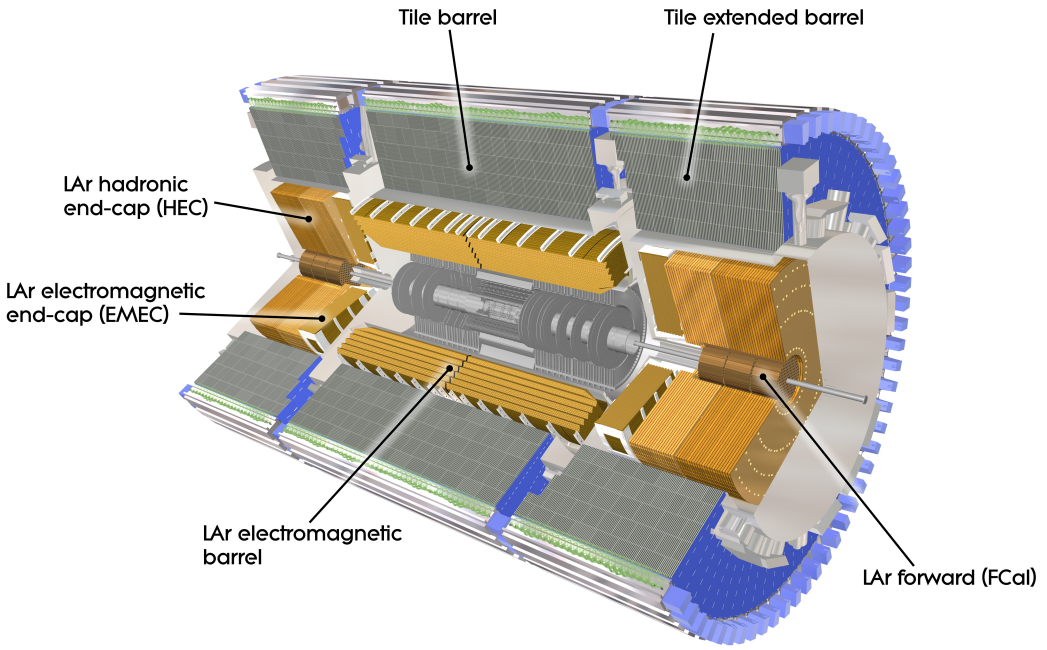
Between the straws are polymer fibres with a high refractive index, which instigate transition-radiation X-ray photons. This radiation is emitted due to the passage of relativistic charged particles through the transition between different materials. The production of this radiation depends on the Lorentz factor, $\gamma = E/m$, and is strongest for electrons - allowing for further particle identification. This is used, for example, to distinguish an electron from a pion.

3.2.2. Calorimeters

Moving outwards from the Inner Detector, we have the calorimeters [63]. These trigger on, and provide precision measurements of the energy of, both charged and neutral particles, such as electrons, photons and hadrons. The calorimeters cover a range $|\eta| < 4.9$, and radially extend to 4.25 metres from the beam-pipe. These are sampling calorimeters, where there are alternating layers of scintillating active media and absorbers. The absorbers are made of dense material which interacts with the particles and instigates a cascade of secondary particles to form. The energy of these particles is measured by the active medium and is proportional to the number of particles in the shower.

The electromagnetic (EM) calorimeter is designed to measure and distinguish electromagnetically interacting particles such as electrons and photons. This consists of two subdetectors; the EM barrel and the EM endcap, covering $|\eta| < 1.5$ and $1.4 < |\eta| < 3.2$ respectively. In the EM calorimeter, liquid argon [64] is used as the active medium and lead as the absorber. It has a maximum granularity of $\Delta\eta \times \Delta\phi$ of 0.1×0.1 and an accordion-shaped geometry to ensure complete coverage in ϕ.

The hadronic calorimeter is designed to detect particles that interact by the strong force, such as pions and neutrons. This consists of four subdetectors; the tile barrel, the tile extended barrel, the hadronic end-cap, and the forward detector. These cover $|\eta| < 1.0$, $0.8 < |\eta| < 1.7$, $1.5 < |\eta| < 3.2$ and $3.1 < |\eta| < 4.9$ respectively. In the hadronic calorimeter, the different subdetectors use different technologies. The tile calorimeter [65] consists of three layers, each using plastic scintillating tiles and steel as the absorber. The hadronic end-cap (HEC) consists of two disks per end cap. Each use liquid argon as the scintillator and copper as the absorber. The forward calorimeters [66] consist of three modules, each using liquid argon as the scintillator. For the absorber, copper is used for the first layer primarily for sensitivity for electromagnetic measurements, and tungsten for the outer two layers for sensitivity to hadronic measurements.

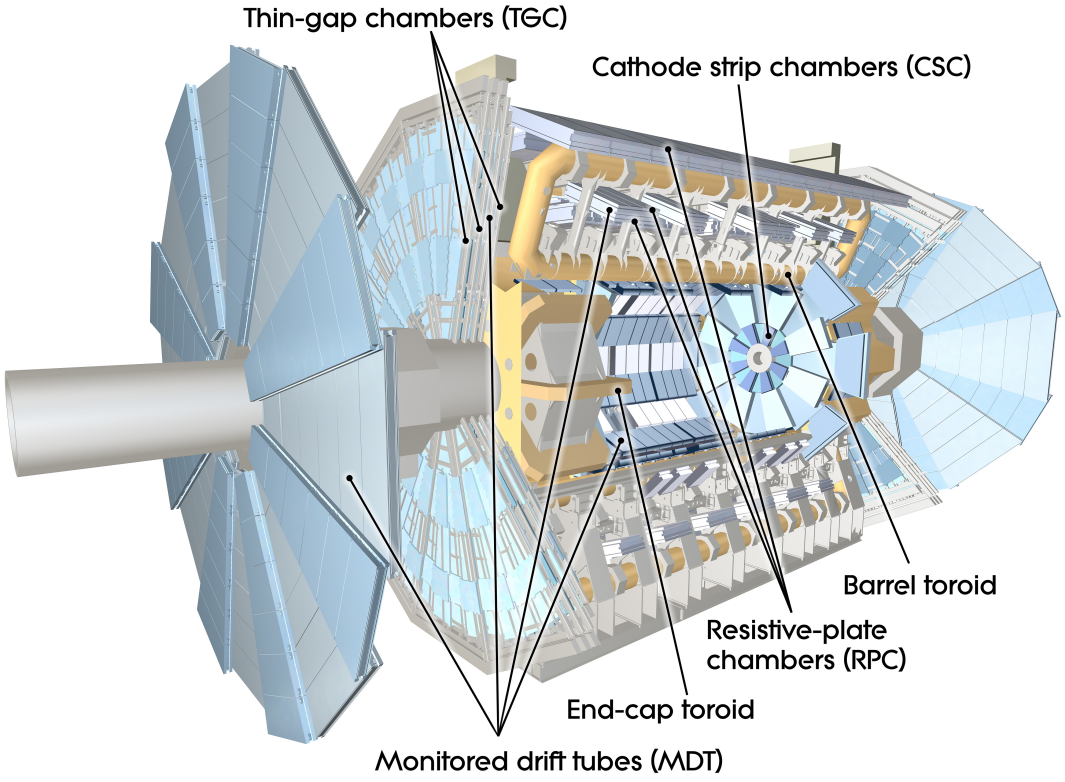


(a)

Figure 3.4.: A schematic diagram of the calorimeters. Image taken from [45, 67].

3.2.3. Muon Spectrometer

The outermost detector is the Muon Spectrometer [68], which covers a range of $|\eta| < 2.7$. It tracks the trajectories of muons in order to provide a momentum measurement, in addition to triggering on them. There are three superconducting toroidal magnet systems around the Muon Spectrometer, which each consist of eight coils. These are used to curve the tracks of the muons. The barrel toroid bends muons with $|\eta| < 1.4$ and the two end-cap toroids bends muons with $1.6 < |\eta| < 2.7$. The region $1.4 < |\eta| < 1.6$ is known as transition region, where deflection is provided by a combination of both magnet types. The structure of the Muon Spectrometer varies depending on these locations about the beam pipe. In the barrel region there are three cylindrical layers of detectors around the beam-pipe, whereas in the transition and end-cap region there are planes of detectors, forming large wheels.



(a)

Figure 3.5.: A schematic diagram of the Muon Spectrometer. Image taken from [45, 69]

The precision tracking is performed by monitored drift tubes (MDTs) for $|\eta| < 2.7$, as well as cathode strip chambers (CSCs) for the end-cap region, $2.0 < |\eta| < 2.7$. MDTs consist of 30 mm diameter tubes filled with argon and carbon dioxide. The CSCs are multiwire proportional chambers, which provide a higher timing resolution than the MDTs for the end-cap region.

Triggering is performed on muons in the region $|\eta| < 2.4$. The triggering is provided by the resistive plate chambers (RPCs) in the barrel region $|\eta| < 1.05$, and thin gap chambers (TGCs) in the higher pseudorapidity regions $1.05 < |\eta| < 2.4$. RPCs consist of two electrode-plates filled with gas in-between, the gas is ionised by the passing muon and the signal is read out. TGCs are multi-wire proportional chambers. These provide a fast response for triggering on the muons of 1.5 ns and 4 ns for the RPC and TGCs, respectively [45].

3.2.4. Triggering

With a 25 ns gap between bunch crossings, there is an initial rate of bunch crossings of 40 MHz. If we were to save every single event, it would take up over 50 TB per second, using an average event size of $\sim$1.3 MB [70]. Since most of these events are uninteresting, a triggering system [71, 72] is used to only select interesting events for further study. The triggers classify interesting objects, such as electrons, muons, or jets such that events containing these objects can be saved for further study. A two-stage triggering process is used, with a step-wise increase in thresholds.

Firstly, the level 1 (L1) hardware trigger reduces the rate from 40 MHz to 100 kHz. This uses information from the calorimeter and the Muon Spectrometer. The level-1 calorimeter (L1Calo) trigger selects clusters in the calorimeter with energies above a predefined threshold. The level-1 muon trigger is provided by the RPCs in the barrel region and TGCs in the end-cap regions. Coincident hits in these detectors that point towards the interaction point are selected. Secondly, the high level trigger (HLT) further reduces the rate from 100 kHz to 1 kHz. It uses information from other subsystems and algorithms similar to those used in offline reconstruction. This reduces memory requirements to a more manageable $\sim$1 GB per second, again considering an average event size of 1.3 MB.

For the analysis in chapter 4 a single-lepton trigger is used. An electron [73] or muon [74] is required to pass the triggering selection.

3.2.5. Object reconstruction

The signals read out from the ATLAS detector are reconstructed into objects that are used for further analysis. Each type of particle is reconstructed independently, with specific calibrations applied. This section outlines the reconstruction techniques used for the objects used in the analyses in this thesis.

Tracks

Charged particles with $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ and $|\eta| < 2.5$ are reconstructed as tracks in the Inner Detector. These tracks will point towards an origin, or vertex. The primary vertex for a bunch crossing is defined as the vertex with the highest scalar sum of squared p_{T}. One primary vertex is defined per bunch crossing, with an *event* being defined as all the objects associated with the given primary vertex.

Electrons

Electrons are reconstructed using information in the Inner Detector and the EM calorimeter. Clusters are formed in the EM calorimeter [75] and are matched to an Inner Detector track.

Identification (ID) requirements are applied to increase electron purity with respect to energy deposits from photons and hadronic jets. These requirements are based on shower shape, the associated tracks, and the amount of leakage into the hadronic calorimeter. In addition, selections are made on $|z_0 \sin\theta|$, where z_0 is the longitudinal impact parameter, and on $|d_0|/\sigma(d_0)$, where d_0 is the transverse impact parameter and $\sigma(d_0)$ is its uncertainty. Moreover, additional requirements on the number of hits in the IBL are applied. Three criteria *loose*, *medium*, and *tight* are defined [75].

Isolation requirements are applied [75] to suppress wrongly identified electrons, as well as those from, for example, decays of heavy-flavour hadrons. These consider the amount of activity in the detector around an electron candidate. Different working points are defined, which strike a balance between a pure sample of electrons and having a higher number of reconstructed electrons. These consider the relative p_{T} of the tracks and E_T inside a cone around the electron.

Muons

The muons in this thesis are reconstructed from tracks in the Muon Spectrometer, which are matched to tracks in the Inner Detector. By combining information from the Muon Spectrometer and the Inner Detector, we get a more robust identification and precise momentum measurement.

Identification requirements are applied to muons to reduce backgrounds, mainly from pion and kaon decays. As with electron identification, three criteria *loose*, *medium*, and *tight* are defined [76]. Isolation requirements [76] are applied in the same way as described for the electron.

Jets

Quarks and gluons produced from proton-proton collisions go on to produce a narrow cone of collimated hadrons from the hadronisation process. This cone of collimated hadrons is commonly known as a *shower*. These particle showers are reconstructed as a single object, commonly known as a *jet*. These are approximations to the parton which produced the shower.

The particle flow algorithm is used to reconstruct jets [77], which uses tracking information from the Inner Detector, in addition to calorimeter information. Tracks are used when the tracking performance is expected to be better than the calorimeter performance. The calorimeter deposits which match these tracks are then removed to avoid double counting. The remaining clusters and tracks matching the primary vertex are used to construct the jet using the anti-k_T algorithm [78], with $R = 0.4$.

Calibration is applied to the energy and resolution of the jets. For example, the entire energy of the jet may not be measured in the calorimeters, especially for hadrons due to, for example, energy deposited in inactive material, or leakage of the shower outside of the calorimeter. Jet energies are calibrated using the jet energy scale (JES) [79], derived from 13 TeV data and simulation.

Jets containing b-hadrons, commonly known as b-jets, can be distinguished from other jets. The lifetime of b-hadrons is relatively long, so they travel a measurable distance $\sim$1 mm before decaying. Therefore, these b-jets originate a measurable distance from the primary vertex. The DL1r [80] algorithm uses a neural network to identify b-jets, using information from the event including the primary vertex location, track positions and reconstructed hadronic jets. The analysis in Chapter 4 uses b-tagging to select b-jets from top decays.

Missing transverse momentum

The detector covers the entire 2π range in ϕ, so the momentum sum in the transverse plane can be used to infer the presence of undetected particles. The initial momentum in the z-direction is unknown, but is assumed to be zero in the transverse plane. So, a transverse momentum imbalance in the final state could be due to undetected neutrinos in the Standard Model, or escaping particles from new-physics processes.

Using the fact that the initial state has zero transverse momentum, the missing transverse momentum $\vec{p}_T^{\text{miss}}$ [81] can be written as

$$\vec{p}_T^{\text{miss}} = \sum_{invisible} \vec{p}_T = -\sum_{visible} \vec{p}_T. \tag{3.3}$$

For calculation of this term, each of the physics objects in the event are considered individually as

$$\vec{p}_T^{\text{miss}} = -\sum \vec{p}_T^{e} - \sum \vec{p}_T^{\gamma} - \sum \vec{p}_T^{\mu} - \sum \vec{p}_T^{\text{jets}} - \vec{p}_T^{\text{soft}}. \tag{3.4}$$

To avoid double-counting effects, the terms are added in the order they are written above. Once a term is added, any tracks and clusters associated with the term are removed. The sums over e, γ, μ and jets are collectively known as the *hard*-term. The final *soft*-term $\vec{p}_T^{\text{soft}}$ is calculated from any remaining tracks, and $p_{\text{T}} < 20\,\text{GeV}$ jets

associated to the primary vertex. This track-based soft term is more robust to pile-up than the previous calorimeter based soft term [81], which used calorimeter deposits for the calculation.

The magnitude of the missing transverse momentum is denoted $E_{\mathrm{T}}^{\mathrm{miss}}$. It is a commonly used variable for R-parity conserving supersymmetry searches, due to invisible LSPs escaping detection. The analysis in Chapter 4 uses $E_{\mathrm{T}}^{\mathrm{miss}}$ as an important discriminating variable.

Chapter 4.

Searching for electroweak supersymmetry

"Never ever ever give up."

— Michael Scott

4.1. Introduction

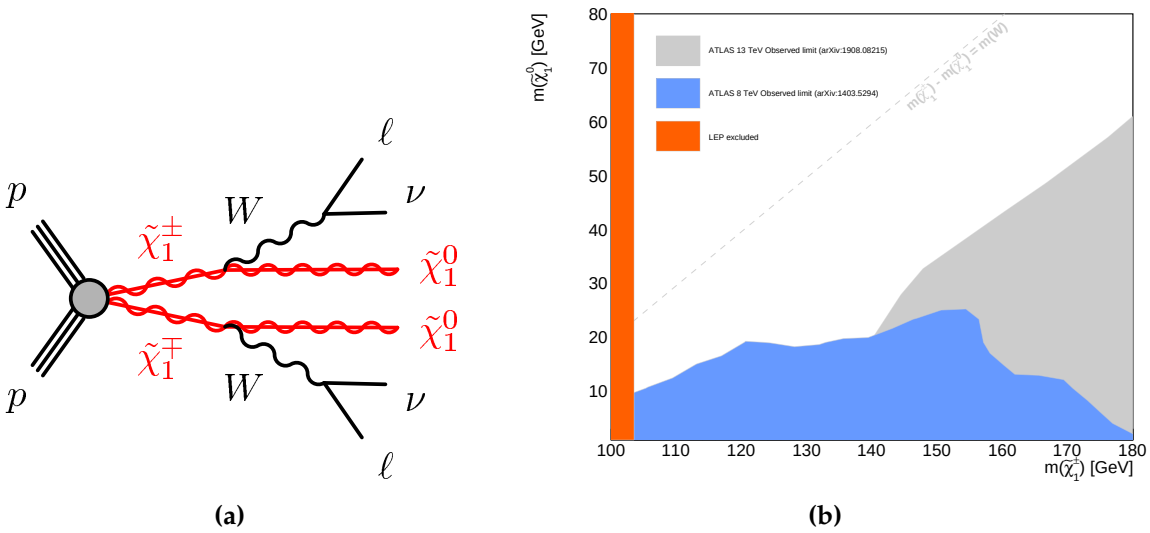
One of the main design goals for the ATLAS detector [45] is searching for new physics, such as supersymmetry. Strongly produced supersymmetric particles have the largest production cross-sections for a given mass at the LHC, so were the primary focus for early searches. However, these searches have shown no statistically significant excesses, with the current limits on the masses of the strongly produced supersymmetric particles being above the TeV scale [28]. Weakly produced supersymmetric particles – charginos, neutralinos and sleptons – have smaller production cross-sections, and so have far lower exclusion limits. If supersymmetry exists in nature with very heavy coloured sparticles, but with low mass charginos,

neutralinos and sleptons, then these electroweak supersymmetric particles could be the primary ones produced at the LHC.

This chapter presents a search for electroweak supersymmetry in events with two light leptons (e or μ) and $E_{\mathrm{T}}^{\mathrm{miss}}$ in the final state, using 139fb^{-1}of data collected by ATLAS during Run 2 of the LHC. This was made public in Refs. [1, 2]. The focus is the direct production of a pair of lightest charginos, $\tilde{\chi}_1^+ \tilde{\chi}_1^-$, each decaying to the lightest neutralino $\tilde{\chi}_1^0$ and a W-boson, which subsequently decays leptonically. An R-parity conserving model is considered, with the $\tilde{\chi}_1^0$ being the LSP, so there is a large amount of $E_{\mathrm{T}}^{\mathrm{miss}}$ in the final state from the escaping $\tilde{\chi}_1^0$ and neutrinos. A diagram for this process can be seen in Figure 4.1a.

The mass difference between the $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_1^0$, $\Delta m = m(\tilde{\chi}_1^\pm) - m(\tilde{\chi}_1^0)$, is the most important factor in determining the kinematics of the final state particles. The previous ATLAS analysis [82] of this process focused on the high Δm region, with a lack of sensitivity to the region with a lower Δm, as seen in Figure 4.1b. The latest CMS analysis [83], and future projections [84] of this process also focus on the high Δm region.

The analysis presented in this chapter focuses on the unexplored area of parameter space at low Δm, in particular the region around the W-boson mass. This is a particularly difficult region of parameter space, with all previous ATLAS analyses having no expected sensitivity in this region [82, 85]. There is a significant background from the diboson VV process, which is very kinematically similar to the signals, so machine learning techniques are utilised to probe this region. A BDT performing multiclass classification is developed using LightGBM [39], in order to identify regions of phase space where we can distinguish the supersymmetric signal from the background using kinematic features of the events. This is used to define regions of phase space with high supersymmetric signal and low background contribution, commonly referred to as *signal regions*. Moreover, multiclass classification has the added benefit over a traditional binary classification of being able to differentiate between different background types. This is leveraged in order to improve the



(a) **(b)**

Figure 4.1.: (a) A diagram of the supersymmetric model being considered; pair production of charginos, decaying via W-bosons. Diagram taken from Ref. [1]. (b) The exclusion limits on this model prior to the analysis described in this chapter. There was no expected exclusion from the 8 TeV analysis [85], with the plotted observed limit due to a downwards fluctuation in data.

background modelling, by designing regions enriched in certain backgrounds, commonly referred to as *control regions*. These are used to derive a data-driven background estimate by fitting the Monte Carlo simulations to data.

Firstly, Sections 4.2 and Section 4.3 introduce the signal and background models, including details about their reconstruction to the physics objects used in the analysis. Section 4.4 details the analysis methods, focusing on the kinematic variables used, the training of the BDT and the validation of the BDT output. Then, in Sections 4.5 and 4.6, the use of the BDT outputs to define the regions used in the analysis is discussed. Sections 4.7 and 4.8 outline the systematic uncertainties and describe the fitting procedure, before results are presented and discussed. Finally, concluding statements are made in Section 4.8.3.

The analysis was performed with collaborators from the 'Two Lepton Zero Jet' supersymmetry team in ATLAS, in which the author took an active role. The author performed the development, validation and implementation of the BDT classifier. In

addition, the author designed the signal regions, and specified preliminary definitions for the control and validation regions. The control and validation regions definitions were finalised in conjunction with a collaborator. All the results were obtained by the author and the plots shown have been produced by the author, unless otherwise noted.

4.2. Data and Monte Carlo samples

The full Run 2 ATLAS dataset, consisting of 139 fb^{-1} of data from 2015-2018, is used. This was collected using single-lepton triggers [73, 74] on electrons and muons. Different p_{T} thresholds are used depending on the data-taking period. Monte Carlo simulations are used to produce samples for the supersymmetric signal and Standard Model background with a two lepton and $E_{\text{T}}^{\text{miss}}$ final state. For the simulation of the detector response, GEANT 4 [86] is used for the background samples, whereas fast simulation [87] used for signal samples. The effect of pile-up is modelled by overlaying inelastic pp events simulated in PYTHIA [88] to the hard scatter event. These data and Monte Carlo samples were produced by a collaborator.

Supersymmetric signals

For the signal model simulation, simplified models [27] are used, as introduced in Section 1.3. Figure 4.1a illustrates the supersymmetric model considered; direct production of a pure wino $\tilde{\chi}_1^{\pm}$ pair, each of which decays to a pure bino $\tilde{\chi}_1^0$ and a W-boson. This is the only decay of the $\tilde{\chi}_1^{\pm}$ considered, with a 100% branching ratio, and the $\tilde{\chi}_1^0$ is the LSP. The masses of all other sparticles are set to very high values, so are not kinematically accessible by the LHC.

We detect the light leptons ($\ell = e, \mu$) produced by the W-boson decay, either directly or via the leptonic decay of the τ-lepton. For this simplified model, the free parameters

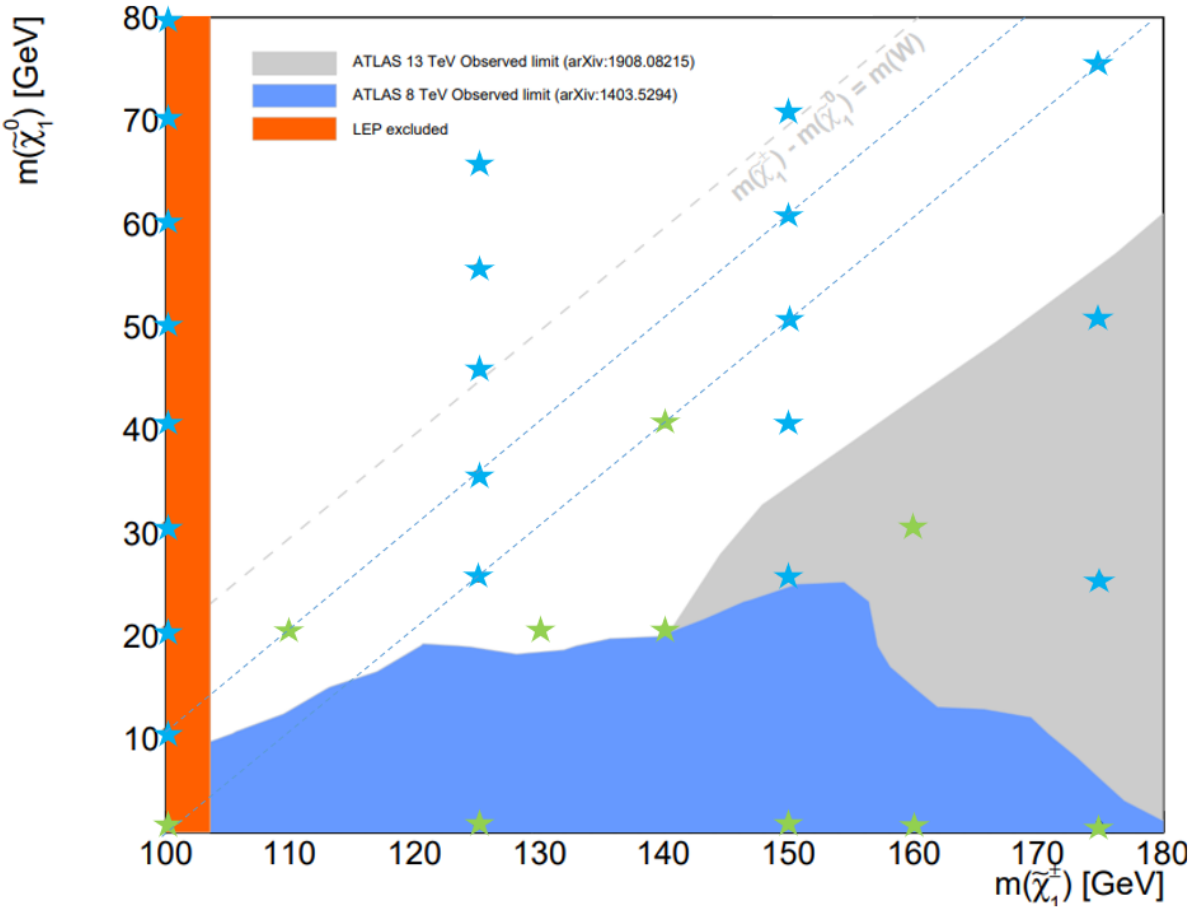


Figure 4.2.: A graph of the generated signal points, indicated by the coloured stars, in the region with $\Delta m \sim 80\,\text{GeV}$. The blue dotted lines indicate $\Delta m = 90$ and $100\,\text{GeV}$. The green points were generated just prior to the publication of Ref. [1] to increase the grid density in spare areas.

are the masses of the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_1^0$, $m(\tilde{\chi}_1^{\pm})$ and $m(\tilde{\chi}_1^0)$, respectively. The production cross-sections decrease as the mass of the $\tilde{\chi}_1^{\pm}$ increases.

A number of signal points were generated with varying $m(\tilde{\chi}_1^{\pm})$ and $m(\tilde{\chi}_1^0)$, as shown in Figure 4.2. The blue stars indicate some of the originally generated signals, which populate diagonals uniform in Δm from 100 GeV to 10 GeV in steps of 10 GeV. Just prior to the publication of Ref. [1], further signal points shown in green were produced to increase the grid density in sparse areas. The details of the generator settings used to generate these samples are in Table 4.1.

Backgrounds

The backgrounds can be separated into two types: irreducible and reducible. The irreducible backgrounds produce two real leptons, whereas the reducible backgrounds have at least one fake or non-prompt (FNP) lepton. *Real* leptons are reconstructed from promptly decaying particles. On the other hand, *fake* leptons can originate from incorrect object reconstruction. For example, a jet leaving a narrow deposit in the calorimeter may be identified as an electron. *Non-prompt* leptons originate from a non-prompt decay, for example, from a leptonic decay of a hadron within a jet.

The dominant irreducible backgrounds producing two real leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$ are diboson (VV) and top ($t\bar{t}$, Wt). The normalisation of these backgrounds is constrained using a simultaneous likelihood fit to data in control regions, as defined in Section 4.6. This lessens the reliance on the Monte Carlo generation and reconstruction processes, which could have a mismodelled normalisation in the region of phase space we are investigating. The 'other' minor Standard Model background processes, for example VVV, $Z/\gamma^*(\rightarrow \ell\ell)$+jets and $Z/\gamma^*(\rightarrow \tau\tau)$+jets, are taken from Monte Carlo directly. The settings used for the generation of these samples are summarised in Table 4.1, and for further details refer to Ref. [1]. The contribution from the reducible backgrounds, such as $W \rightarrow \ell\nu +$ jets is estimated from data using the Matrix method [89], as described in Section 4.6.2.

4.3. Physics object definitions

To define the objects, we use two categories, *baseline* and *signal*. Firstly, the looser *baseline* requirements are applied, then an overlap removal procedure is carried out to resolve ambiguities in double-counting the physics objects, and finally the tighter *signal* requirements are applied. The *signal* objects are the ones that are used for the

Physics process	Generator	Parton shower	Normalisation	Tune	PDF
Diboson VV	SHERPA 2.2.1 or 2.2.2 [90]	SHERPA 2.2.1 or 2.2.2 [91,92]	NLO [93–96]	SHERPA default [97]	NNPDF3.0NNLO [98]
$t\bar{t}$	POWHEG-BOX v2 [99–102]	PYTHIA 8.230 [103]	NNLO+NNLL [104]	A14 [105]	NNPDF3.0NLO [98]
Single top (Wt)	POWHEG-BOX v2 [100–102,106]	PYTHIA 8.230	NLO+NNLL [107,108]	A14	NNPDF3.0NLO
Triboson VVV	SHERPA 2.2.2	SHERPA 2.2.2	NLO	SHERPA default	NNPDF3.0NNLO
$t\bar{t}+V$	MADGRAPH5_aMC@NLO 2.3.3 [109]	PYTHIA 8.210 [103]	NLO [109,110]	A14	NNPDF3.0NLO
$t\bar{t}+H$	POWHEG-BOX v2 [99–102,111]	PYTHIA 8.230	NLO	A14	NNPDF3.0NLO
$t\bar{t}+WW$	MADGRAPH5_aMC@NLO 2.2.2	PYTHIA 8.186 [88]	NLO [109]	A14	NNPDF2.3LO
$t\bar{t}+WZ$	MADGRAPH5_aMC@NLO 2.3.3	PYTHIA 8.212 [88]	NLO [109]	A14	NNPDF2.3LO
$tZ, t\bar{t}t\bar{t}, t\bar{t}t$	MADGRAPH5_aMC@NLO 2.3.3	PYTHIA 8.230	NLO [109]	A14	NNPDF3.0NLO
$Z/\gamma^*(\rightarrow ll)$	SHERPA 2.2.1 [90]	SHERPA 2.2.1 [92]	NNLO [112]	SHERPA default	NNPDF3.0NNLO
$\tilde{\chi}_1^{\pm}\,\tilde{\chi}_1^{\mp}$	MADGRAPH5_aMC@NLO 2.6.2 [113]	PYTHIA 8.2 [114]	NLO+NLL	A14	NNPDF2.3LO [115]

Table 4.1.: The settings for the Monte Carlo simulation of the background and signal event samples, parton showering, cross-section order in α_s, underlying-event tune, and the PDF set used. The main irreducible backgrounds for this analysis are included at the top of the table, with the other smaller backgrounds in the following rows. The signal samples are in the final row. Table reproduced from Ref. [1].

final analysis regions. The *baseline* objects are also used in the FNP estimate and are an input for calculating $E_{\mathrm{T}}^{\mathrm{miss}}$.

Baseline electrons have $|\eta| < 2.47$, $p_{\mathrm{T}} > 9\,\mathrm{GeV}$, *LooseAndBLayerLLH* particle identification [116], and longitudinal impact parameter $|z_0 \sin\theta| < 0.5$ mm. For *signal* electrons, the identification is tightened to the *TightLLH* criteria [116], an additional *FCLoose* isolation requirement is applied, and the transverse impact parameter must satisfy $|d_0/\sigma(d_0)| < 5$.

Baseline muons have $|\eta| < 2.6$, $p_{\mathrm{T}} > 9\,\mathrm{GeV}$, *Medium* particle identification [117], and longitudinal impact parameter $|z_0 \sin\theta| < 0.5$ mm. For *signal* muons, an additional *FCLoose* isolation requirement is applied [117], and the transverse impact parameter must satisfy $|d_0/\sigma(d_0)| < 3$.

Particle flow [77] jets are reconstructed with the anti-kt algorithm [78], with $R = 0.4$. The jets have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, with *baseline* jets having $|\eta| < 2.8$ and *signal* jets $|\eta| < 2.4$. A further requirement for *signal* jets, with $p_{\mathrm{T}} < 60\,\mathrm{GeV}$, is the *Tight* JVT working point [118]. This aids in suppression of pileup jets.

The reconstruction of b-jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ is described in Section 3.2.5. For b-tagging, an 85% efficiency point is used for tagging b-jets, as measured in simulated $t\bar{t}$ events. For

the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation, *baseline* objects are used, where jets with $|\eta| > 2.4$ are required to have $p_{\mathrm{T}} > 30\,\mathrm{GeV}$.

Overlap removal is applied to resolve double-counting ambiguities between reconstructed objects. This is performed between the light leptons and jets, as they can often fake one other. Firstly, jet candidates that are close to electron and muon candidates are removed: jet candidates within $\Delta R = \sqrt{\Delta y^2 + \Delta\phi^2} = 0.2$ of an electron candidate are removed, as are jet candidates with fewer than three tracks within $\Delta R = 0.4$ of a muon candidate. These jet candidates will likely originate from the electron or muon, for example from the electron shower or muon bremsstrahlung. Secondly, electrons and muons within $\Delta R = \min(0.4, 0.04 + 10/p_{\mathrm{T}}[\mathrm{GeV}])$ of the remaining jet candidates are removed. This helps reject electrons and muons which could have been produced by decays of b- and c-jets. Finally, for electron and muons that share a track in the Inner Detector, if the muon is tagged in the calorimeter it is removed, otherwise, the electron is removed.

4.4. Analysis strategy

In order to distinguish the signals from the large irreducible background, machine learning techniques are used to exploit kinematic differences between events. Moreover, rather than a simple binary-classification, where we classify the signal against all the backgrounds, we use multiclass classification, where we classify four classes: signal, diboson, top ($t\bar{t}$ and single-top) and others (all remaining backgrounds). A BDT is used from the LightGBM [39] framework, which outputs four scores denoted as BDT-signal, BDT-VV, BDT-top and BDT-others. The sum of the four scores is equal to one.

Optimisation of the classifier includes the choice of hyperparameters, which variables to use, and which regions and signal points to train over. This optimisation is performed without looking at the data, with the data only finally looked at when

the development is complete and the background modelling is suitably validated. A metric is required for optimisation, constructed by comparing the signal and background yields. The metric used for optimisation is the Z_N value [119],

$$Z_N = \sqrt{2\left[n\log\left(\frac{n(b+\sigma^2)}{b^2+n\sigma^2}\right) - \frac{b^2}{\sigma^2}\log\left(1+\frac{\sigma^2(n-b)}{b(b+\sigma^2)}\right)\right]}, \tag{4.1}$$

where b is the expected background count, and σ is the uncertainty on this background. n is the observed number of events, which for optimisation we take to equal the sum of the expected background and signal counts (s), $n = s + b$. Z_N estimates the significance of observing n events when having an expected number of $b \pm \sigma$ background events [119], where significance refers to the number of standard deviations of a Gaussian distribution. For the potential exclusion of a signal point, we require $Z_N > 1.64$, whereas for a potential discovery we require $Z_N > 5$. The previous ATLAS analysis looking for this supersymmetric process had $\sim$15% systematic uncertainty [82], so a conservative 20% was used for the calculation of Z_N throughout the optimisation process.

4.4.1. Kinematic variables

The signals and backgrounds have different event kinematics and particle content. Numerous variables are considered to discriminate the supersymmetric signal from the background. Firstly, the signals produce oppositely charged leptons, hence we require *oppositely-signed* (OS) charged leptons for the final state in the analysis. *Same-signed* (SS) charged lepton final states are used for the validation of the fake estimates in Section 4.6.2. Events are categorised as *same-flavour* (SF) and *different-flavour* (DF), based on the flavour composition of the two leptons; ee and $\mu\mu$ compared to $e\mu$ and μe. This is particularly relevant since Z-bosons decay solely to SF leptons, hence there is a vastly different $Z/\gamma^*(\to \ell\ell)$+jets background contribution in the regions defined with SF and DF leptons. These are referred to as the SF and DF channels.

The number of jets (including b-jets) is denoted by n_{jet}. This is the sum of the number of non-b-jets, $n_{\text{non-}b\text{-tagged jets}}$, and the number of b-tagged jets, $n_{b\text{-tagged jets}}$. The signals produce zero jets in the final state, so for the training and signal regions we veto central jets; $n_{\text{jet}} = 0$.

The transverse momentum of the leading and sub-leading lepton ($p_T^{\ell_1}$, $p_T^{\ell_2}$) are used. From the four-momenta of the two leptons, we can calculate the invariant mass of these two leptons $m_{\ell\ell}$, which is particularly useful in identifying the $Z/\gamma^*(\rightarrow \ell\ell)$+jets background in the SF channel, since the $m_{\ell\ell}$ distribution has a peak around the Z-boson mass. $E_{\text{T}}^{\text{miss}}$, defined in Section 3.2.5, is a powerful discriminating variable due to the $\tilde{\chi}_1^0$ and neutrinos from the supersymmetric signal escaping undetected by ATLAS. This gives the signals a typically larger $E_{\text{T}}^{\text{miss}}$ than for the Standard Model background events.

E_T^{miss} significance

An extension to $E_{\text{T}}^{\text{miss}}$ is the object-based E_T^{miss} significance [120], which helps discriminate events where $E_{\text{T}}^{\text{miss}}$ arises from undetected particles as opposed to mismeasurement. It is defined as the log-likelihood ratio comparing the hypothesis that $p_{\text{T}}^{\text{inv}}$ is equal to zero to the hypothesis that it differs from zero, and can be expressed as

$$E_{\text{T}}^{\text{miss}}\text{significance} = \frac{|\vec{p}_{\text{T}}^{\text{miss}}|}{\sqrt{\sigma_L^2(1-\rho_{LT}^2)}} \tag{4.2}$$

where σ_L is the momentum resolution of the longitudinal component of $\vec{p}_{\text{T}}^{\text{miss}}$, and ρ_{LT} is the correlation factor between the resolutions in the parallel and transverse directions. This provides discrimination for events where the $E_{\text{T}}^{\text{miss}}$ in the background is due to object mismeasurement or finite detector acceptance or resolution (fake $E_{\text{T}}^{\text{miss}}$), rather than from invisible final state particles (real $E_{\text{T}}^{\text{miss}}$). The $Z/\gamma^*(\rightarrow \ell\ell)$+jets background

in particular tends to have low values of E_T^{miss} significance, since it produces no real $E_{\text{T}}^{\text{miss}}$.

m_{T2}

The transverse mass, m_T, can be used to calculate a bound on the mass of a semi-invisibly decaying object. For a two particle final state, it is defined as

$$m_T{}^2 = m_1{}^2 + m_2{}^2 + 2\left(E_{T,1}E_{T,2} - \vec{p}_{T,1}\cdot\vec{p}_{T,2}\right) \approx 2p_T^{\ell}p_T^{\text{miss}}\left(1-\cos\Delta\phi\right) \tag{4.3}$$

where $\Delta\phi$ is the angle between the lepton and the $E_{\text{T}}^{\text{miss}}$. For the latter expression, we have treated the electron, muon and neutrino masses to be zero, since they are much smaller than the momentum scales involved. The distribution of m_T has an endpoint at the mass of the parent particle, for example at the W-boson mass when considering the W-boson decay.

Extending this idea for events with a pair of semi-invisibly decaying particles, we can construct m_{T2} [121, 122]. Assuming massless particles are produced, it is defined as

$$m_{T2}(\vec{p}_T^{\,\ell 1}, \vec{p}_T^{\,\ell 2}, \vec{p}_T^{\,miss}) = \min_{\vec{q}_1+\vec{q}_2=\vec{p}_T^{\,miss}} \Big[\max\big(m_T(\vec{p}_T^{\,\ell 1}, \vec{q}_1), m_T(\vec{p}_T^{\,\ell 2}, \vec{q}_2)\big)\Big]. \tag{4.4}$$

As input, we have the observed lepton transverse momenta and missing transverse momentum. This is designed to provide a lower bound on the masses of the pair of semi-invisibly decaying particles. The maximum of the two m_T provides the better lower bound on the parent particle's mass. But, since the split of $\vec{p}_T^{\,miss}$ is unknown between the two invisible particles, all splits of $\vec{q}_1$ and $\vec{q}_2$ that to sum to $\vec{p}_T^{\,miss}$ are considered. If we choose the wrong split, the feature that m_T will have an endpoint at the parent particle mass is no longer necessarily correct, so the minimum value is taken across all splits to ensure that a valid lower bound is obtained. For example, for

Standard Model events, the distribution is expected to tail off around the W-boson mass.

When calculating m_{T2}, a further input can be the mass of the final state particles (m_χ). We take these as massless, since when investigating other values, there was no difference in sensitivity. When the correct m_χ is used, the m_{T2} distribution has the desired endpoint at the parent particle mass. In the case that we have the incorrect m_χ, the endpoint of the distribution becomes shifted to be $\sim m_\chi + \Delta m$. In the previous ATLAS analysis at 139 fb^{-1} [82], m_{T2} was used to gain sensitivity to high mass $\tilde{\chi}_1^\pm$ with a large Δm: the shape of the m_{T2} distribution for the supersymmetric signals extended out past the Standard Model background, which ended at around m_W. However, for the current analysis, we are considering low Δm events, so the endpoint of the m_{T2} distribution for the supersymmetric signals will not have the feature that it extends out past the Standard Model background. This is the primary reason for the different analysis strategy presented in this chapter compared to the previous analysis.

Angular variables

Angular variables are considered, which can distinguish between the geometry of supersymmetric and Standard Model events. Differences in the azimuthal angle $\Delta\phi$ are used, since the periodicity of ϕ for the cylindrical ATLAS detector means that absolute values of ϕ are defined with respect to an arbitrary origin. The azimuthal angle between the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the leading and sub-leading lepton $\Delta\phi_{E_{\mathrm{T}}^{\mathrm{miss}},\ell_1}$ and $\Delta\phi_{E_{\mathrm{T}}^{\mathrm{miss}},\ell_2}$ have discriminating power.

Two further angular variables are considered. Firstly, $\Delta\phi_{\mathrm{boost}}$ is defined as the azimuthal separation between the vectors $\vec{p}_T^{\,\mathrm{miss}}$ and $\vec{p}_T^{\,\mathrm{boost}}$, where $\vec{p}_T^{\,\mathrm{boost}}$ is the vector sum of the $\vec{p}_T^{\,\ell 1}$, $\vec{p}_T^{\,\ell 2}$ and $\vec{p}_T^{\,\mathrm{miss}}$. In addition, $\cos\theta^*_{\ell\ell}$ [123] is defined as $\cos\theta^*_{\ell\ell} = \tanh(\Delta\eta_{\ell\ell}/2)$. It was introduced in relation to spin-0 sleptons and is used in the slepton analysis in Ref. [1], where the sleptons tend to produce events with low $\cos\theta^*_{\ell\ell}$. The angular distribution of a pair of decaying of spin-1/2 $\tilde{\chi}_1^\pm$ via spin-1

W-bosons differs from the distribution of the VV background, due to the additional $\tilde{\chi}_1^{\pm}$–W-boson coupling: when comparing the distribution of $\cos\theta^*_{\ell\ell}$ for sleptons and the $\tilde{\chi}_1^{\pm}$ signals, we see a similar peak at low values, indicating that it can be similarly useful for discrimination.

During the development phase of the analysis, various other variables were considered, but ultimately did not improve the sensitivity to the supersymmetric signals. Overall, the following 10 variables were trained over: $p_T^{\ell_1}$, $p_T^{\ell_2}$, $m_{\ell\ell}$, $E_{\mathrm{T}}^{\mathrm{miss}}$, E_T^{miss} significance, m_{T2}, $\Delta\phi_{E_{\mathrm{T}}^{\mathrm{miss}},\ell_1}$, $\Delta\phi_{E_{\mathrm{T}}^{\mathrm{miss}},\ell_2}$, and $\cos\theta^*_{\ell\ell}$. Following the training of the BDTs, various metrics can be used to assign importances to the variables. Section 4.4.4 describes a selection of these, which can help us further understand the phase space selected by the BDT for identifying the signals from the backgrounds.

4.4.2. Event selection

Preselection requirements are applied to select the final state we are interested in. These are loose requirements to remove the areas of kinematic phase space dominated by backgrounds, resulting in more manageable background yields. Two OS leptons are required, with $p_T^{\ell_1} > 27\,\mathrm{GeV}$ and $p_T^{\ell_2} > 9\,\mathrm{GeV}$, to ensure we are on the trigger efficiency plateau. A requirement of $m_{\ell\ell} > 11\,\mathrm{GeV}$ is applied to remove low mass resonances such as J/ψ and Y, and for SF events $m_{\ell\ell}$ being further than 15 GeV from the Z-boson mass in order to reduce the $Z/\gamma^*(\to \ell\ell)$+jets component. This requirement is commonly referred to as a 'Z-veto'. In addition, E_T^{miss} significance > 3 is applied to significantly reduce the background component, in particular backgrounds without real $E_{\mathrm{T}}^{\mathrm{miss}}$ such as $Z/\gamma^*(\to \ell\ell)$+jets . These preselection selections are summarised in Table 4.2.

The targeted supersymmetric signals have a final state with zero jets, so the classifier is trained over a region with a zero jet requirement, $n_{\mathrm{jet}} = 0$. This reduces the top backgrounds in particular. During development, training over final states with one jet

was also considered. However, adding these regions provided no additional sensitivity, so these were not further considered.

The background composition is different when considering the DF and SF channels, predominantly due to Z-boson decays producing only SF leptons. Hence, training is performed separately over the DF and SF channels. Figures 4.3 and 4.4 illustrate the DF and SF events with zero jets and the preselection requirements. From these plots, it is clear that the $Z/\gamma^*(\rightarrow \ell\ell)$+jets contribution in the SF channel is far greater. Table 4.3 indicates the number of events for each background after these preselection, jet and DF or SF requirements.

Variable	Requirement
$N_{\text{OS leptons}}$	$= 2$
$p_T^{\ell_1}$	> 27 GeV
$p_T^{\ell_2}$	> 9 GeV
$m_{\ell\ell}$	> 11 GeV
n_{jet}	≤ 1
E_T^{miss} significance	> 3
$\lvert m_{\ell\ell} - m_Z \rvert$	> 15 GeV(SF events only)

Table 4.2.: The preselection requirements applied in general. $n_{\text{jet}} = 0$ is required for training of the BDT, whereas $n_{b\text{-tagged jets}} = 1$ is required for top control regions.

4.4.3. Classifier training

As described in Section 4.2, the supersymmetric signal points are generated with different Δm, and the kinematics of the signal models are closely related to the mass difference Δm. For example, the leptons produced have lower p_{T} as Δm gets smaller, so there is a lower signal yield when using leptonic triggers. This can be seen in Table 4.3, with a reduced yield for the lower Δm scenario with an equal mass $\tilde{\chi}_1^{\pm}$.

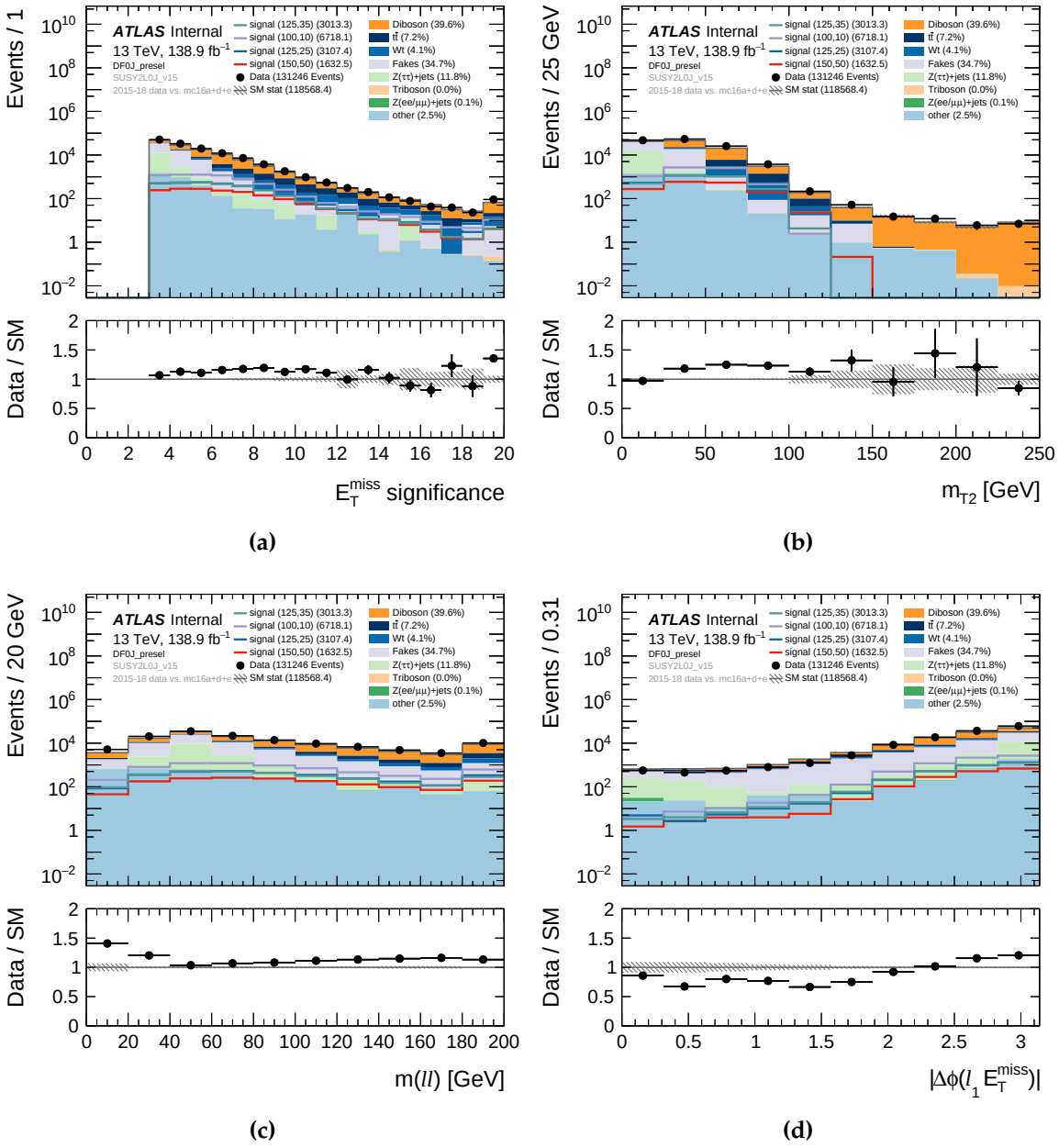


Figure 4.3.: Kinematic distributions of data and Monte Carlo simulated signals and backgrounds with DF leptons, zero jets and the preselection requirements in Table 4.2 applied. The FNP backgrounds calculated as described in Section 4.6.2, and 'other' includes all other rare backgrounds.

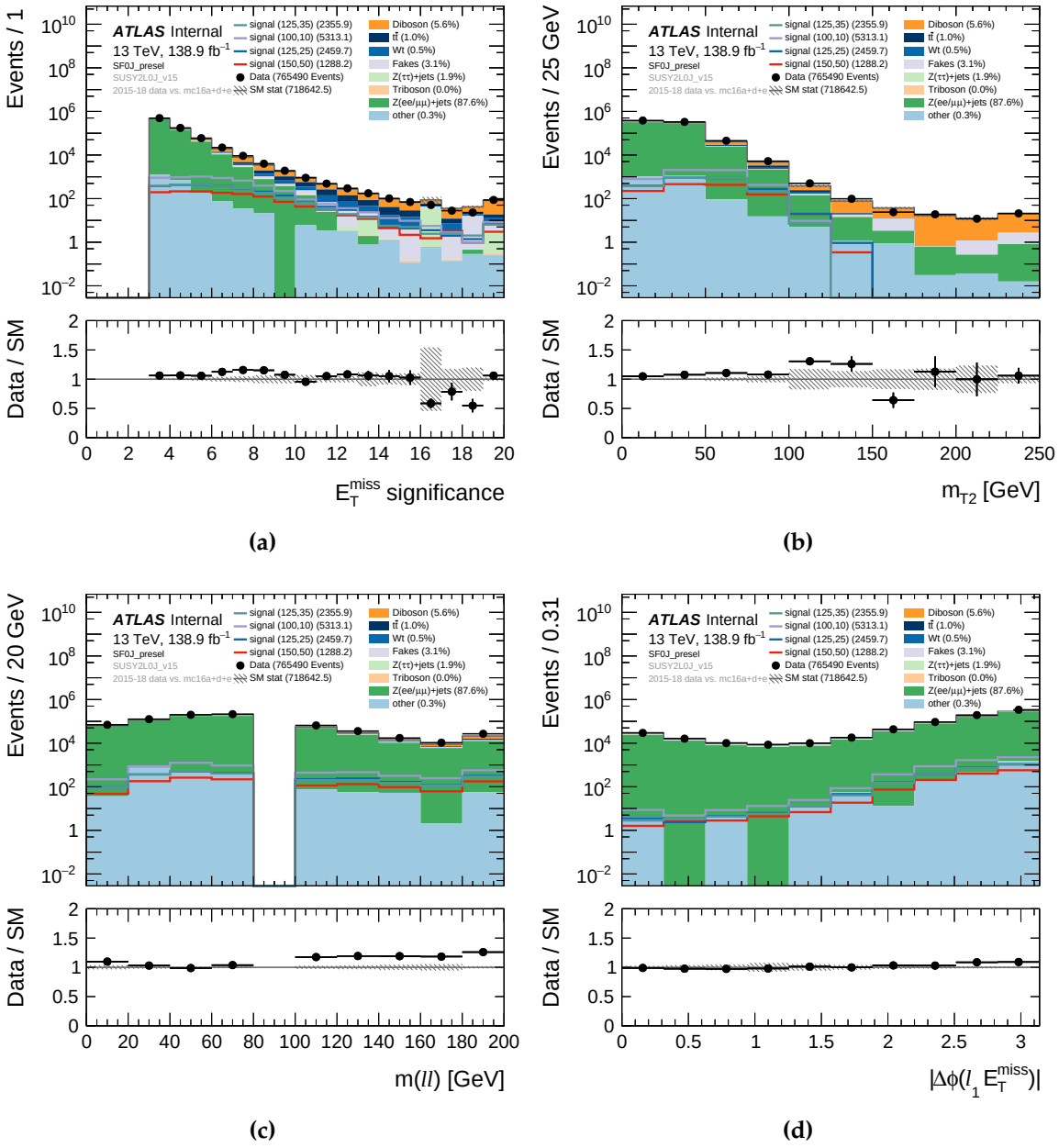


Figure 4.4.: Kinematic distributions of data and Monte Carlo simulated signals and backgrounds with SF leptons, zero jets and the preselection requirements in Table 4.2 applied. The FNP backgrounds calculated as described in Section 4.6.2, and 'other' includes all other rare backgrounds.

Sample	DF0J raw events	DF0J yield at 139fb^{-1}	SF0J raw events	SF0J yield at 139fb^{-1}
VV	1328177	46941.4 ± 76.5	1262416	40142.0 ± 72.0
$t\bar{t}$	242826	8573.9 ± 18.3	190406	6899.4 ± 16.6
Wt	38969	4865.7 ± 25.6	30280	3881.4 ± 23.1
$Z/\gamma^*(\to \ell\ell)$+jets	48	65.1 ± 28.2	330092	631188.3 ± 3868.2
$Z/\gamma^*(\to \tau\tau)$+jets	8822	13996.4 ± 420.7	8696	13564.6 ± 418.4
VVV	2147	4.3 ± 0.1	1592	3.3 ± 0.1
others	60517	3020.3 ± 269.3	53639	2245.4 ± 238.1
m($\tilde{\chi}_1^\pm$, $\tilde{\chi}_1^0$) = (125, 25) GeV	17119	3107.4 ± 24.9	13206	2461.7 ± 22.3
m($\tilde{\chi}_1^\pm$, $\tilde{\chi}_1^0$) = (150, 50) GeV	9927	1632.5 ± 17.1	7650	1290.3 ± 15.4
m($\tilde{\chi}_1^\pm$, $\tilde{\chi}_1^0$) = (125, 85) GeV	5617	589.1 ± 9.0	4831	505.8 ± 8.3

Table 4.3.: The number of raw events and event yields at 139fb^{-1}, following the preselection requirements in Table 4.2.

During the development stage, various training strategies were considered. These strategies consisted of training over points with the same value of Δm, either individually or combined with other diagonals of uniform Δm. Overall, the best sensitivity was obtained by training over the signals with Δm = 100 GeV and 90 GeV combined. These signals are the blue stars that lie on the blue dashed lines in Figure 4.2. The more compressed points with $\Delta m \leq 80$ GeV were considered, but even with trained classifiers no sensitivity to these could be obtained, so they were not considered further. Appendix A details further why it is particularly difficult to have sensitivity to the more compressed points with leptonic triggers, since the leptons from the W-boson decay are very soft. Future searches for this supersymmetric signal in the compressed region could proceed as in Refs. [124, 125], using a trigger on $E_{\mathrm{T}}^{\mathrm{miss}}$ and considering scenarios with an initial-state radiation (ISR) jet.

The overall simulated dataset is randomly split into two equally sized, statistically independent subsets. The BDT was trained over one half and tested over the other half, such that we can investigate how well the BDT generalises to unseen data. By doing this, the number of events used for further analysis is reduced by a factor of two, which can increase uncertainties. So, a second BDT was trained with the training and test sets swapped with respect to the previous BDT, as illustrated in Figure 4.5.

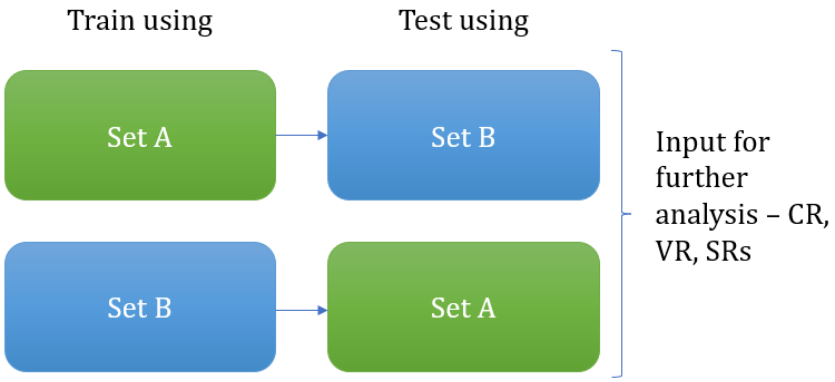


Figure 4.5.: The training and testing strategy used for the analysis. By training two BDTs, we can use the entire dataset for further analysis, ensuring that we are always testing over a statistically independent data sample to the one that the BDT is trained over.

This allows us to use the entire dataset for both training and testing, whilst ensuring we always test using a statistically independent data sample to the training sample.

Hyperparameter tuning was performed as described in Section 2.4. Using 5-fold cross validation, the following hyperparameters of the BDT were tuned: number of trees, number of leaves, the minimum number of samples per leaf and the learning rate. Ultimately, we observed that, in general, having a high number of trees with a small number of leaves gives good performance. With this setup, the sensitivity is comparable over a large range of hyperparameter selections, so a further more detailed hyperparameter optimisation was not necessary.

The classifiers for the DF and SF channels were independently optimised, with the chosen hyperparameters for the following results being as follows. For the DF BDT we have a learning rate of 0.06, 100 trees, each tree having 24 leaves and a minimum of 116 samples per tree. Whereas, for the SF BDT we have a learning rate of 0.05, 150 trees, each tree having 20 leaves and a minimum of 17 samples per tree. All other hyperparameters remain at their default values set by LightGBM [39].

Following the hyperparameter optimisation, the BDTs were trained over the entire dataset, as illustrated by Figure 4.5. The four BDT scores on the test-sets are combined

and can be seen in Figures 4.6 and 4.7. We can clearly see that signal, VV, top and the other backgrounds are preferentially selected by the BDT-signal, -VV, -top and -other scores, respectively. High BDT-signal score is used for defining the signal regions in Section 4.5. Selections on the BDT-scores for the backgrounds are used to define the validation and control regions in Section 4.6.1. For example, a requirement of low BDT-other is used in the SF regions to remove events from $Z/\gamma^*(\rightarrow \ell\ell)$+jets events.

4.4.4. Classifier evaluation

Following training, various checks are performed on the classifiers. Firstly, checks are made to ensure that they are not overtrained, and the quality of the classification is assessed using ROC curves and PR curves as described in Section 2.5. In addition, several variable importance metrics are calculated in order to get a fuller understanding of how each variable contributes to the classification performance. These include simpler metrics, such as simply enumerating the number of times each variable is used for a split in the BDT. Finally, more involved methods of evaluating variable importance are investigated; SHAP values are calculated as described in Section 2.5. This section focuses on the DF classifier, which is the more sensitive. Comparable results are seen when evaluating the SF classifier.

Train and test set plot

A first check is made by plotting the training and test set for both signal and background on the same plot in Figure 4.8. Visually, we can see a good agreement between the training set (bars) and the test set (points) for both the background and signal samples, indicating that the classifier is not overtrained. The background sample includes all the three classes VV, top and other.

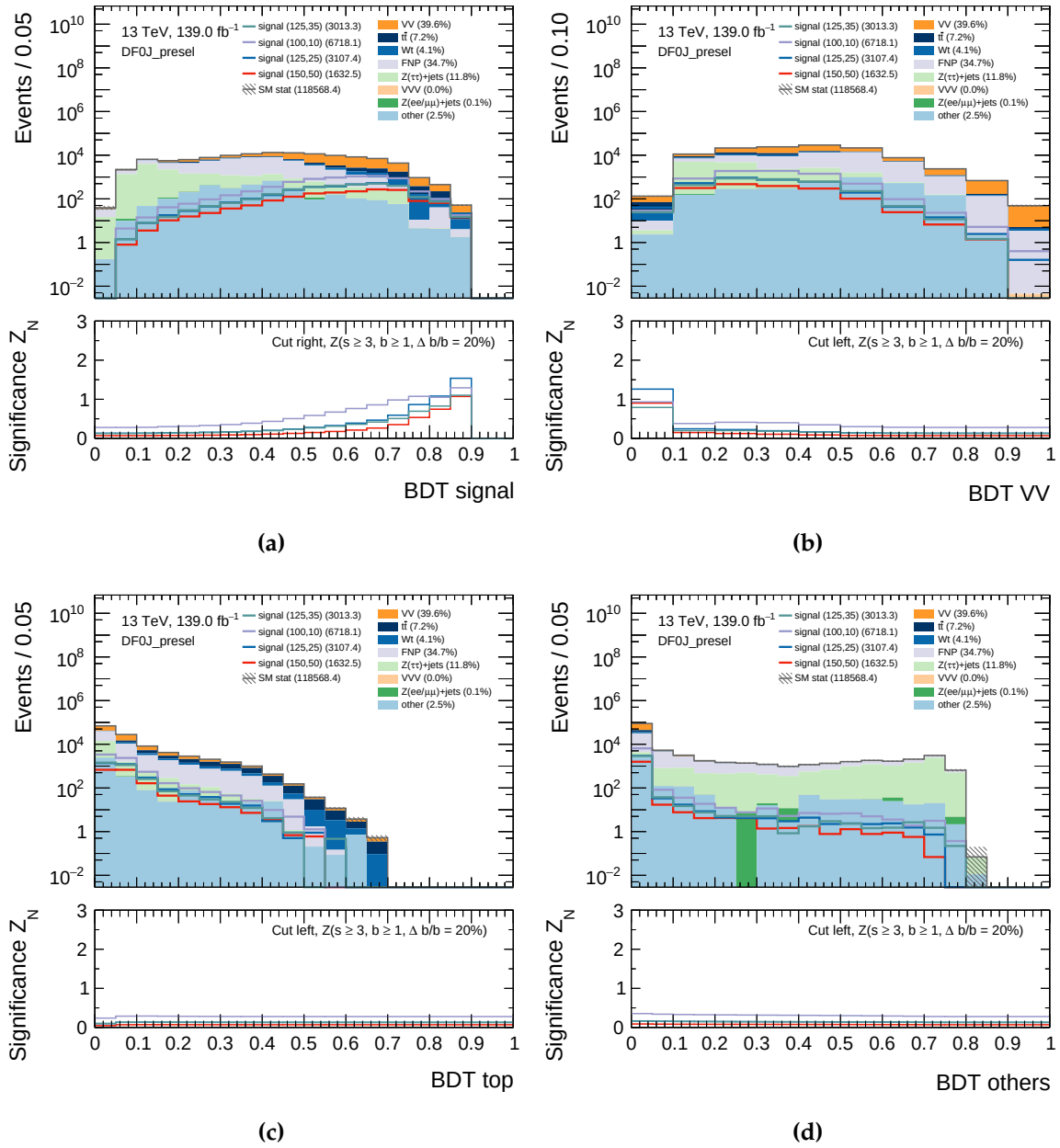


Figure 4.6.: Distributions of the four BDT scores for the classifier trained on DF, zero jet samples. The significance in the bottom panel is calculated using Equation 4.1, with a selection to the right for BDT-signal and to the left for the other BDT scores. The FNP backgrounds calculated as described in Section 4.6.2, and 'other' includes all other rare backgrounds.

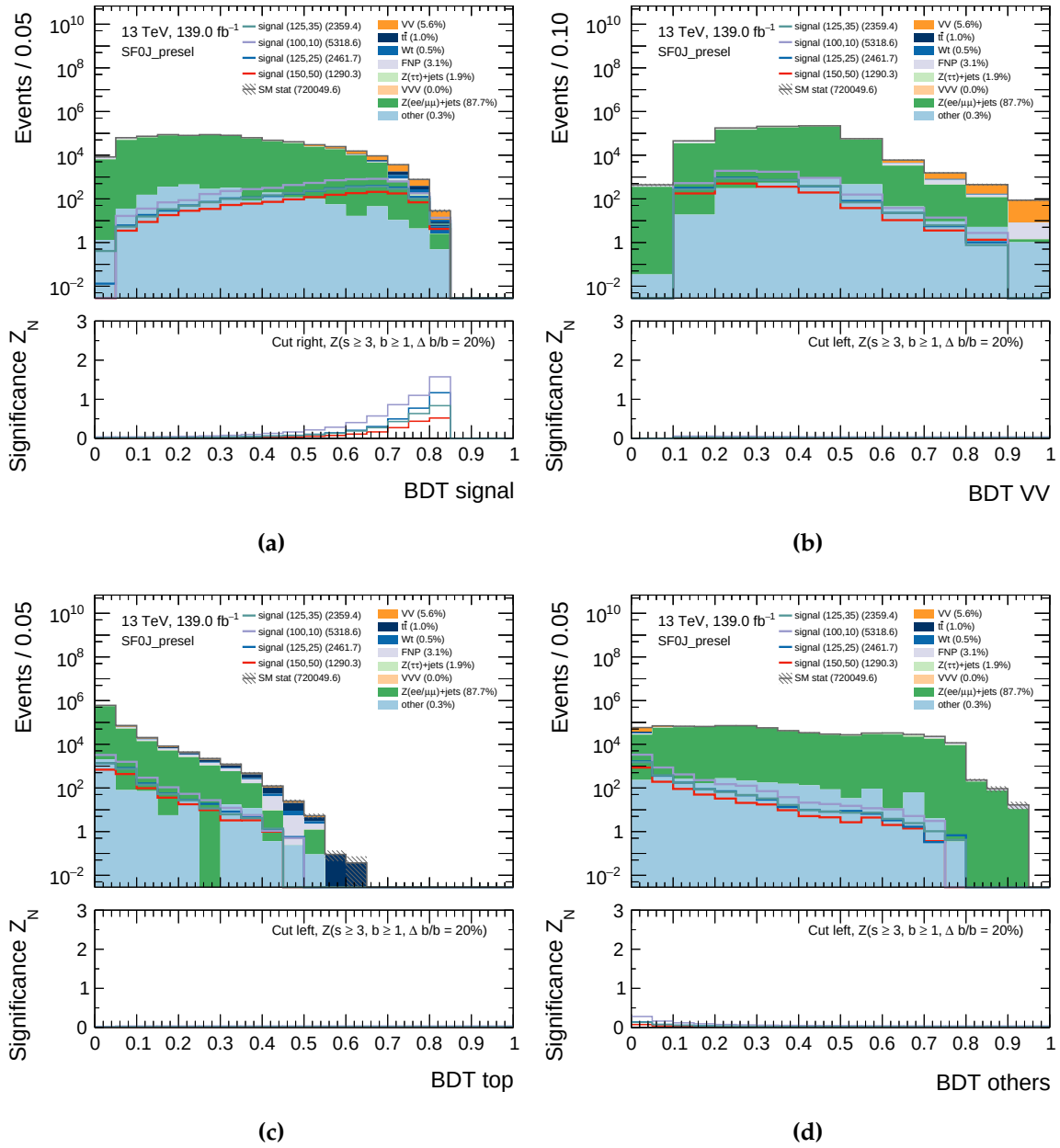


Figure 4.7.: Distributions of the four BDT scores for the classifier trained on SF, zero jet samples. The significance in the bottom panel is calculated using Equation 4.1, with a selection to the right for BDT-signal and to the left for the other BDT scores. The FNP backgrounds calculated as described in Section 4.6.2, and 'other' includes all other rare backgrounds.

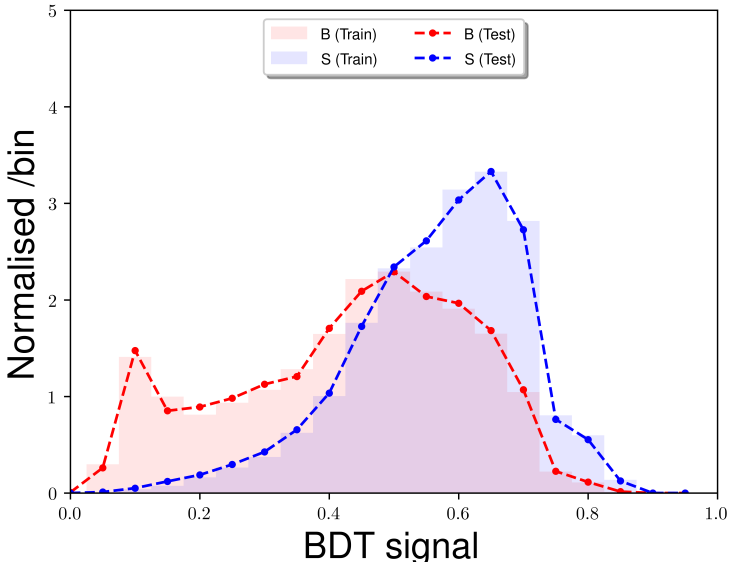
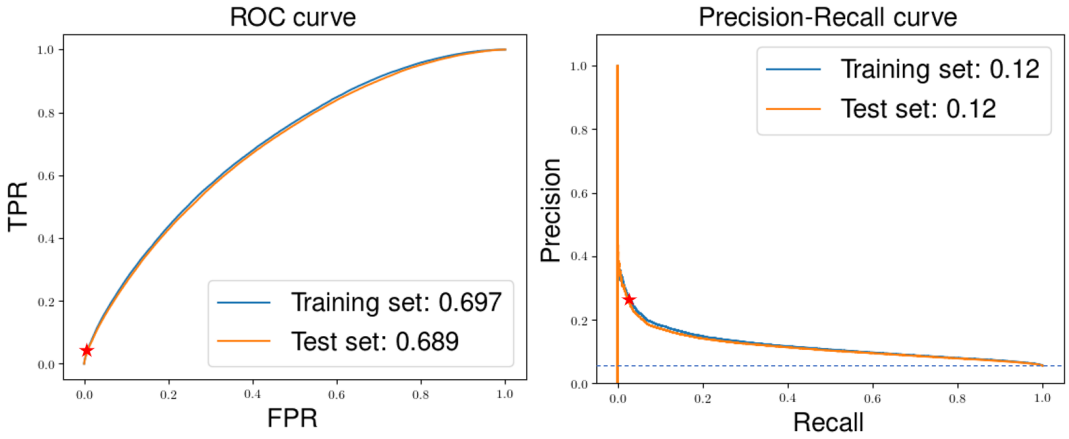


Figure 4.8.: Distribution of the BDT-signal score for the training and test set for the signal (S) and background (B) samples.

ROC and PR curves

The ROC and PR curves are calculated as described in Section 2.5. These are shown in Figure 4.9 when comparing the signal sample against all other backgrounds combined. The 'area under the curve' (AUC) values are calculated and included in the legend. The AUC is very similar for both the training and testing set, providing further indication that there is no overtraining.

In addition, we can calculate ROC curves for each of the four categories, by considering classifying them against all the other categories combined. Figure 4.10 shows this and indicates which of the categories are easiest to distinguish from the rest. With a high AUC value of 0.93, we can see that the 'other' background is the easiest to separate from the remaining classes (signal, VV and top), due to the 'other' background having very different event kinematics.



(a) **(b)**

Figure 4.9.: The Receiver operating characteristic (ROC) and precision-recall (PR) curves for the training and testing sets. These have been calculated considering the signal against all the other backgrounds. The ROC area under curve (AUC) and average-precision values are included in the legend. The star indicates where the SRs begin. TPR and FPR are the true and false positive rates, respectively, as described in Section 2.5.

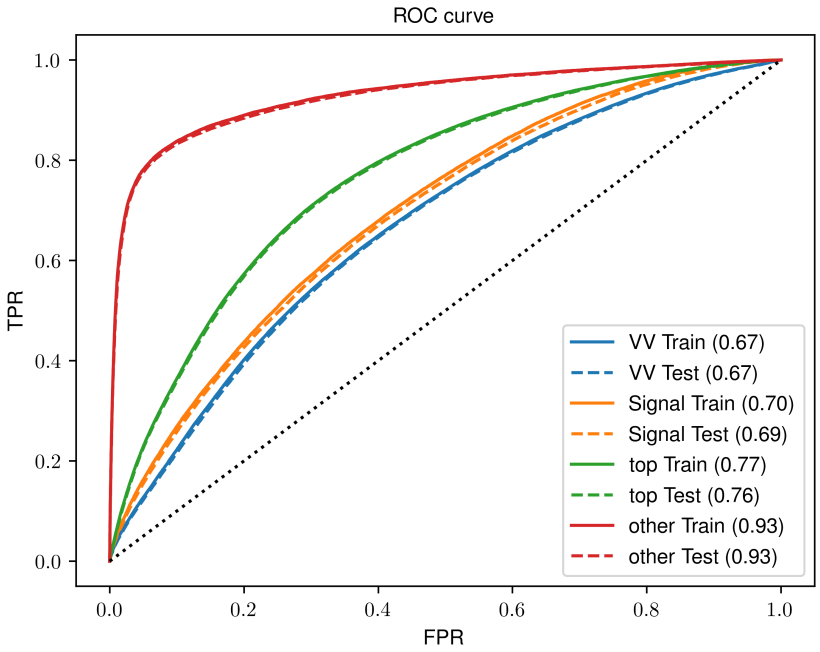


Figure 4.10.: The Receiver operating characteristic (ROC) for each of the four categories (VV, signal, top, other) in the multiclass classifier. The curves use the respective BDT score and measure the separation to the other categories. The area under curve (AUC) are included in the legend, for the training and test sets. TPR and FPR are the true and false positive rates, respectively, as described in Section 2.5

Variable importances

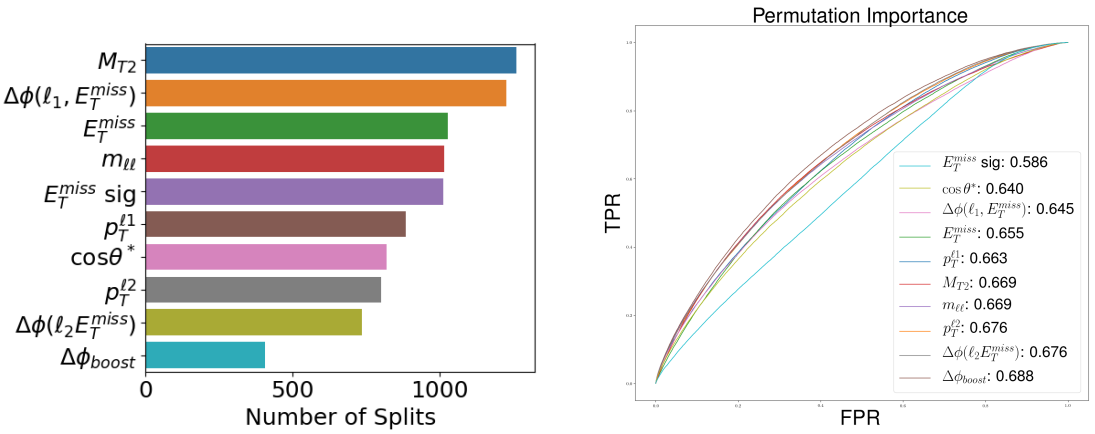
Understanding how important each variable is to the classifier gives a good insight into interpreting the classifier output. We evaluate two such methods in this section, and consider SHAP values in the following section.

A first simple metric can be derived by enumerating the number of times each variable is used by the BDT to make a split. Variables that are used more often are likely to be providing better discrimination. These counts are indicated by the x-axis in Figure 4.11a. This indicates that variables m_{T2}, $\Delta\phi_{E_{\mathrm{T}}^{\mathrm{miss}},\ell_1}$, and $E_{\mathrm{T}}^{\mathrm{miss}}$ are used the most when constructing the BDT. This, however, has limitations; it does not give tell us how much discrimination between different classes is achieved through each of these splits.

A further method to determine variable importances is through *permutation importance*. In essence, we wish to consider how a trained classifier performs on the test set when we remove any information about a given variable. If one variable is important for classification, then removing the information about this variable will result in worse classifier performance.

It is not possible to simply remove a variable and run the classifier, since there is no prescription of what to do when evaluating a split on a removed variable. However, we can replace the values of the variable with random noise drawn from the same distribution. In this case, no useful information can be gained from that variable, and it will have the desired effect of removing the variable. This replacement by random noise is achieved by randomly permuting the values for a given variable.

Figure 4.11b reports the test-set signal ROC curves, where one variable is permuted. Each line corresponds to the specific permuted variable, and the ROC-AUC values are indicated in the legend. Interestingly, we see here that E_T^{miss} significance is the most important variable, since its removal leads to the worst performance. We see that E_T^{miss} significance is used fewer times for splitting the trees, but is nevertheless important for separation at each split.



(a) **(b)**

Figure 4.11.: (a) The number of times each variable is used to split in the BDT. (b) The ROC curves for the test set with one variable randomly shuffled across all events, as indicated by the legend. Lower values of the AUC indicate a more important variable.

These two variable importances metrics are inherently different – the number of times each variable is used in contrast to AUC. It is not unexpected that there are some differences in ordering. Both methods provide complementary information and help us understand the BDT, just like the SHAP values described in the following section.

SHAP values

SHAP values [43] as introduced in Section 2.5 are also considered for the trained BDT. These provide even further information about how the BDT is using each variable.

Figure 4.12 illustrates the SHAP values for the BDT-signal score. An event corresponds to a point in each row of the plot, with the colour of the point indicating the value of the variable. For example, a red point in the top row corresponds to an event with a high E_T^{miss} significance value. The position of the point along the x-axis indicates the contribution to the BDT-signal score from the value of the variable. For example, a red point in the top row which is far to the right along the x-axis means

that, for this event, the high E_T^{miss} significance value results in a higher BDT-signal score.

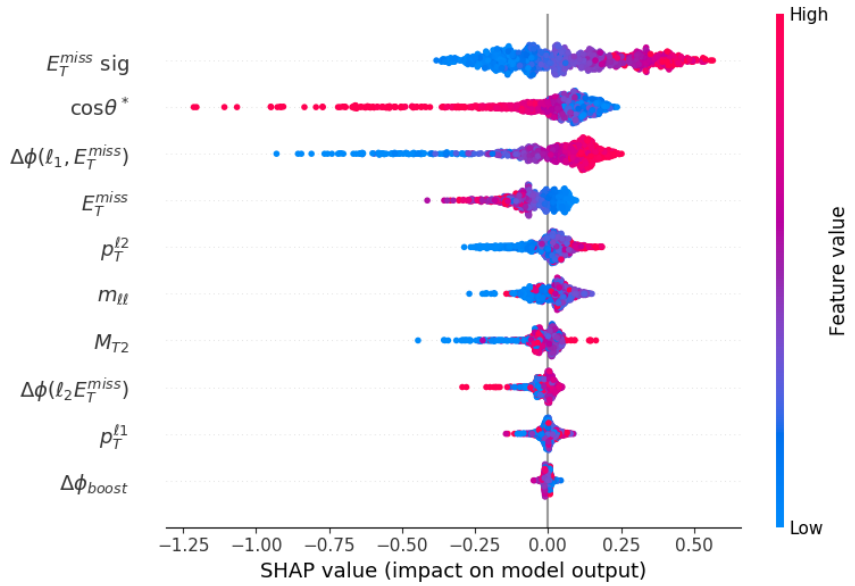


Figure 4.12.: The SHAP values for the BDT-signal score. An event corresponds to a point in each row of the plot, with the colour of the point corresponding to the value of the variable. The x-axis indicates the SHAP value, where points to the right are more signal-like and to the left are more background-like.

SHAP values can be calculated for each of the other BDT scores as well, giving us four sets of SHAP values. By taking the magnitude of the SHAP values, we can get an indication of the importance of a variable; a variable that does not have a large impact on the value of the BDT score will have a low SHAP value. These are plotted as a 'SHAP bar chart' in Figure 4.13, from which can assess the importance of each variable for each background. For example, it is evident that m_{T2} is particularly important for classifying the other backgrounds (mainly $Z/\gamma^*(\to \tau\tau)$+jets), since these tend to have low m_{T2}. This provides further insight into how the classifier is performing the separation into the four classes in a way not achieved by the previously considered variable importances.

Finally, we can plot scatter plots of each variable against the respective SHAP value due to that variable. For example, Figure 4.14 illustrates the scatter plots for $\Delta\phi_{E_{\text{T}}^{\text{miss}},\ell_1}$ and $\cos\theta^*_{\ell\ell}$. Most of the events tend to have high $\Delta\phi_{E_{\text{T}}^{\text{miss}},\ell_1}$, near π, which indicates

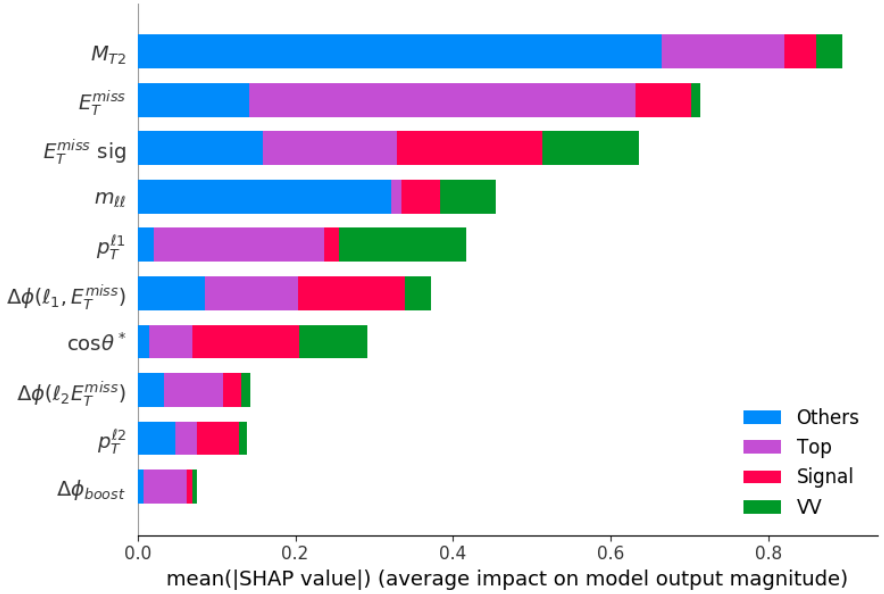
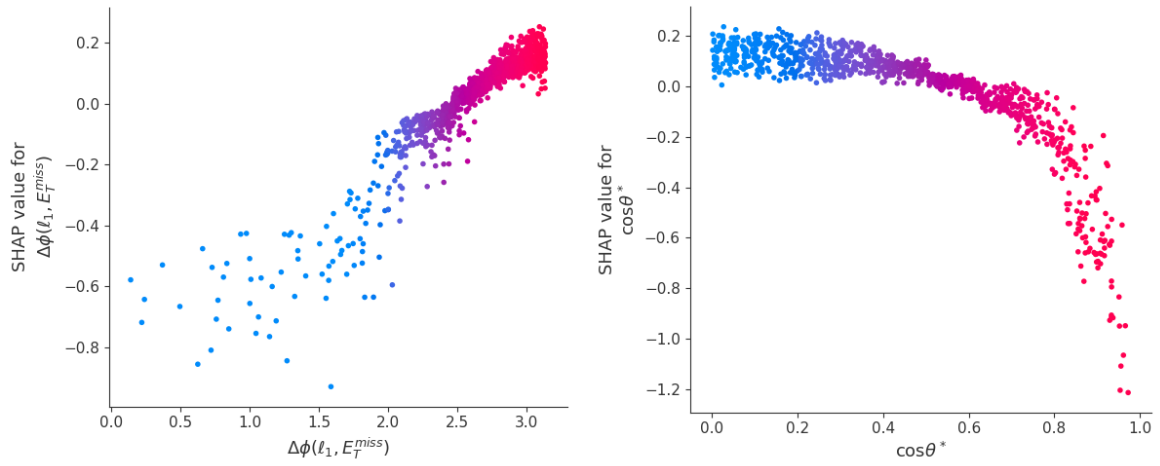


Figure 4.13.: A bar chart of the magnitude of the SHAP values for each BDT score. These indicate the average impact of each variable on each specific BDT score.



(a) **(b)**

Figure 4.14.: A scatter plot of the signal-SHAP values against the variable value for (a) $\Delta\phi_{E_T^{\text{miss}},\ell_1}$ (b) $\cos\theta^*_{\ell\ell}$. For $\Delta\phi_{E_T^{\text{miss}},\ell_1}$, high values tend to be more signal-like, whereas, for $\cos\theta^*_{\ell\ell}$, the high values tend to be more background-like. The colour of the points corresponds to the value of the variable on the x-axis.

a more signal-like event. In addition, we see that there are fewer events with high $\cos\theta^*_{\ell\ell}$, close to 1, and these tend to be more background-like.

4.5. Signal region definitions

The signal regions are defined using the BDT-signal score as the main discriminatory variable. Figures 4.6 and 4.7 illustrate this discrimination power; the signal events have higher values of BDT-signal. Two types of signal regions are defined. Firstly, the *binned* signal regions consist of multiple subregions defined by selections on the BDT-signal score in order to target specific signal models. These are fit simultaneously in order to calculate exclusion limits. By statistically combining many signal region bins, we include discrimination information from the shapes of the signal and background distributions and so can make stronger statements on the exclusion of the targeted models. In addition, multiple *inclusive* signal regions, which each consist of a single requirement on the BDT-signal score, are defined in order to make model-independent statements about sensitivity to new physics. Each inclusive region is fit independently to one another.

As when optimising the BDT, we remain blind to data when designing the signal regions. This is in order to avoid any biases when constructing the signal regions arising from knowledge of the data distribution. The data are considered only when the signal regions are fully defined, and we are satisfied with the background modelling.

Before going into the details about the SR definitions, we make some preliminary clean-up requirements. Looking at m_{T2} and E_T^{miss} significance distributions with high BDT-signal score for the DF and SF channels in Figure 4.15, we can manually make the selections $m_{\mathrm{T2}} > 50\,\mathrm{GeV}$ and E_T^{miss} significance > 8 without changing the events chosen by the BDTs, since the BDT identifies that more signal like events have high E_T^{miss} significance and m_{T2}. Applying these selections to other regions ensures that they are kinematically close to the signal regions.

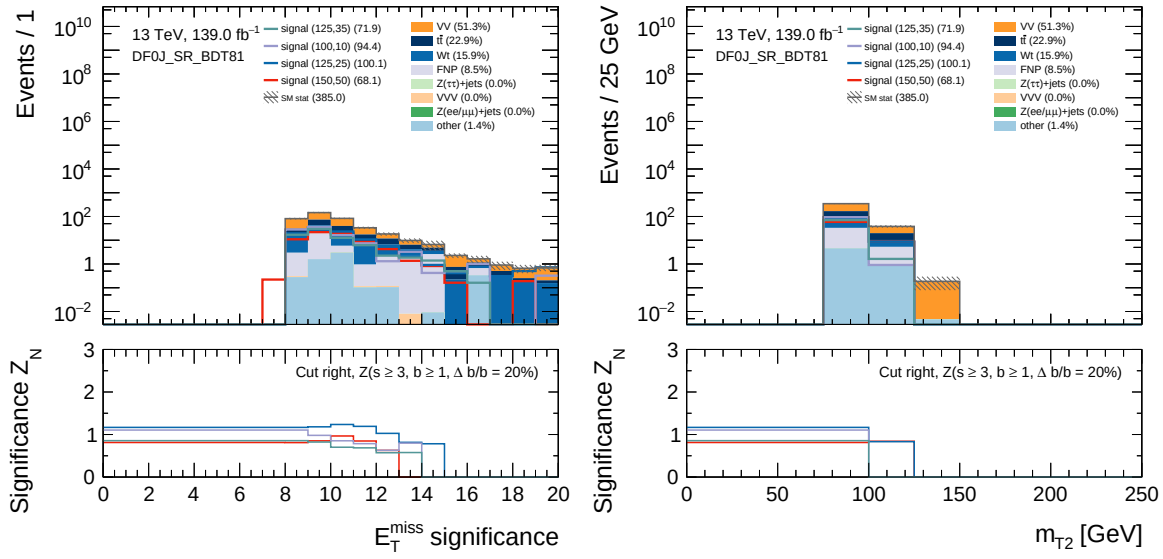


Figure 4.15.: The E_T^{miss} significance and m_{T2} distributions in a region with the high BDT-signal score requirements of the signal regions. The upper panels show DF events with the requirement of BDT-signal > 0.81 and the lower panels show SF events with the requirement of BDT-signal > 0.77. The significance in the bottom panel is calculated using Equation 4.1. The FNP backgrounds calculated as described in Section 4.6.2, and 'other' includes all other rare backgrounds.

Binned signal regions

In the region of high BDT-signal score, as the BDT-signal score increases, the backgrounds decrease, whilst the signals remain relatively constant from bin to bin. This shape difference can be exploited to perform a shape-fit on these bins to achieve a greater sensitivity. When defining these signal regions, it was observed that having $\sim$30 background events in each bin ensured that the exclusion fit, as described in Section 4.8.2, remained stable when using all the systematic uncertainties. This meant that wider bins were used at higher BDT-signal scores, where the signal and background contributions are lower. The representative points m($\tilde{\chi}_1^{\pm}$, $\tilde{\chi}_1^0$) = (125, 25) (150, 50) and (125, 35) GeV were considered for this optimisation.

In order to define the binned signal regions, since using Z_N is not appropriate, an exclusion fit in `HistFitter` [126] was performed over all the proposed bins. The exclusion fit used a 20% flat systematic uncertainty, and outputs CL_s values[1], where a lower value corresponds to a greater sensitivity; $CL_s < 0.05$ indicates the signal point can be excluded at 95% confidence level.

The general method for defining the signal regions was to make a single selection on the BDT-signal score and consider binning the region above this. For the DF channel, splitting this region of BDT-signal > 0.84 into four bins gave a greater sensitivity. Moreover, including more bins below down to 0.81 included more shape information and further increased the sensitivity. The CL_s values for these are reported in Table 4.4, with the sensitivity increasing for each successive case.

For the SF channel, a further selection on BDT-others < 0.01 is made to reduce the $Z/\gamma^*(\rightarrow \ell\ell)$+jets contribution. This ensures the signal region events to be very unlike $Z/\gamma^*(\rightarrow \ell\ell)$+jets events. The same procedure as the DF channel is then performed: splitting the region with BDT-signal > 0.785 into 5 bins is first considered. Then introducing three more bins further increases the sensitivity, as reported in Table 4.4.

[1] CL_s values are defined in Section 4.8.2.

Finally, combining the bins from both the DF and SF channels, we get the greatest sensitivity, as can be seen in the top row of Table 4.4. The values of the selections used to define these binned signal regions are reported in Table 4.5.

Region	(125,25) GeV CL_s	(150,50) GeV CL_s	(125,35) GeV CL_s
DF0J and SF0J (26 bins)	0.0028	0.038	0.030
DF0J $>$ 0.81 (16 bins)	0.0056	0.046	0.043
DF0J $>$ 0.84 (4 bins)	0.021	0.10	0.098
DF0J $>$ 0.84 (1 bin)	0.032	0.13	0.13
SF0J $>$ 0.77 (8 bins)	0.051	0.26	0.17
SF0J $>$ 0.785 (5 bins)	0.087	0.43	0.22
SF0J $>$ 0.785 (1 bin)	0.099	0.44	0.25

Table 4.4.: The expected CL_s values (as defined in Section 4.8.2) with statistical uncertainties and a flat 20% systematic uncertainty. These are calculated for the three signal points m($\tilde{\chi}_1^{\pm}$, $\tilde{\chi}_1^0$) = (125, 25), (150, 50) and (125, 35) GeV. Each row indicates a different setup of bins, with the number of bins for each row shown in brackets. The top row illustrates the full binning setup, which uses with both DF and SF combined. Lower CL_s values indicate better sensitivity.

Inclusive signal regions

For the model-independent signal regions, we define wider regions to be sensitive to a greater range of potential models of beyond the Standard Model physics. Multiple overlapping signal regions are defined, which are each considered in a separate fit.

The first, most general, region considers all the phase space considered by the binned signal regions. That is, including both DF and SF events, with requirements of BDT-signal $>$ 0.81 for DF events and $>$ 0.77 for SF events. Tighter regions are also considered for the DF and SF channels separately, as described in Table 4.5.

Distributions are displayed in Figure 4.16 for the inclusive DF region with BDT-signal > 0.81 and in Figure 4.17 for the inclusive SF region with SF0J BDT-signal > 0.77. Similar regions of phase space are selected in both the DF and SF channels. For

example, in the signal regions we have high values of $\Delta\phi_{E_{\mathrm{T}}^{\mathrm{miss}},\ell_1}$, indicating that the $E_{\mathrm{T}}^{\mathrm{miss}}$ and leading lepton tend to be back-to-back. In addition, the E_T^{miss} significance values are high since we have real $E_{\mathrm{T}}^{\mathrm{miss}}$ from the escaping $\tilde{\chi}_1^0$ and neutrinos.

Table 4.5.: The definitions of the signal regions. These selections are made in addition to the preselection requirements in Table 4.2. For the binned signal region, all the signal regions are fit together. For the inclusive signal regions, each signal region is fit independently to one another. Table reproduced from Ref. [1].

Signal region (SR)	SR-DF	SR-SF
$n_{b\text{-tagged jets}}$	$= 0$	
$n_{\text{non-}b\text{-tagged jets}}$	$= 0$	
E_T^{miss} significance	>8	
m_{T2} [GeV]	>50	
BDT-others		< 0.01
Binned SRs	Binned SRs	
	$\in$(0.81,0.8125]	$\in$(0.77,0.775]
	$\in$(0.8125,0.815]	$\in$(0.775,0.78]
	$\in$(0.815,0.8175]	$\in$(0.78,0.785]
	$\in$(0.8175,0.82]	$\in$(0.785,0.79]
	$\in$(0.82,0.8225]	$\in$(0.79,0.795]
	$\in$(0.8225,0.825]	$\in$(0.795,0.80]
	$\in$(0.825,0.8275]	$\in$(0.80,0.81]
	$\in$(0.8275,0.83]	$\in$(0.81,1]
	$\in$(0.83,0.8325]	
	$\in$(0.8325,0.835]	
	$\in$(0.835,0.8375]	
	$\in$(0.8375,0.84]	
	$\in$(0.84,0.845]	
	$\in$(0.845,0.85]	
	$\in$(0.85,0.86]	
	$\in$(0.86,1]	
	Inclusive SRs	
BDT-signal	$\in$(0.81,1]	$\in$(0.77,1]
	$\in$(0.81,1]	
	$\in$(0.82,1]	
	$\in$(0.83,1]	
	$\in$(0.84,1]	
	$\in$(0.85,1]	
		$\in$(0.77,1]
		$\in$(0.78,1]
		$\in$(0.79,1]
		$\in$(0.80,1]

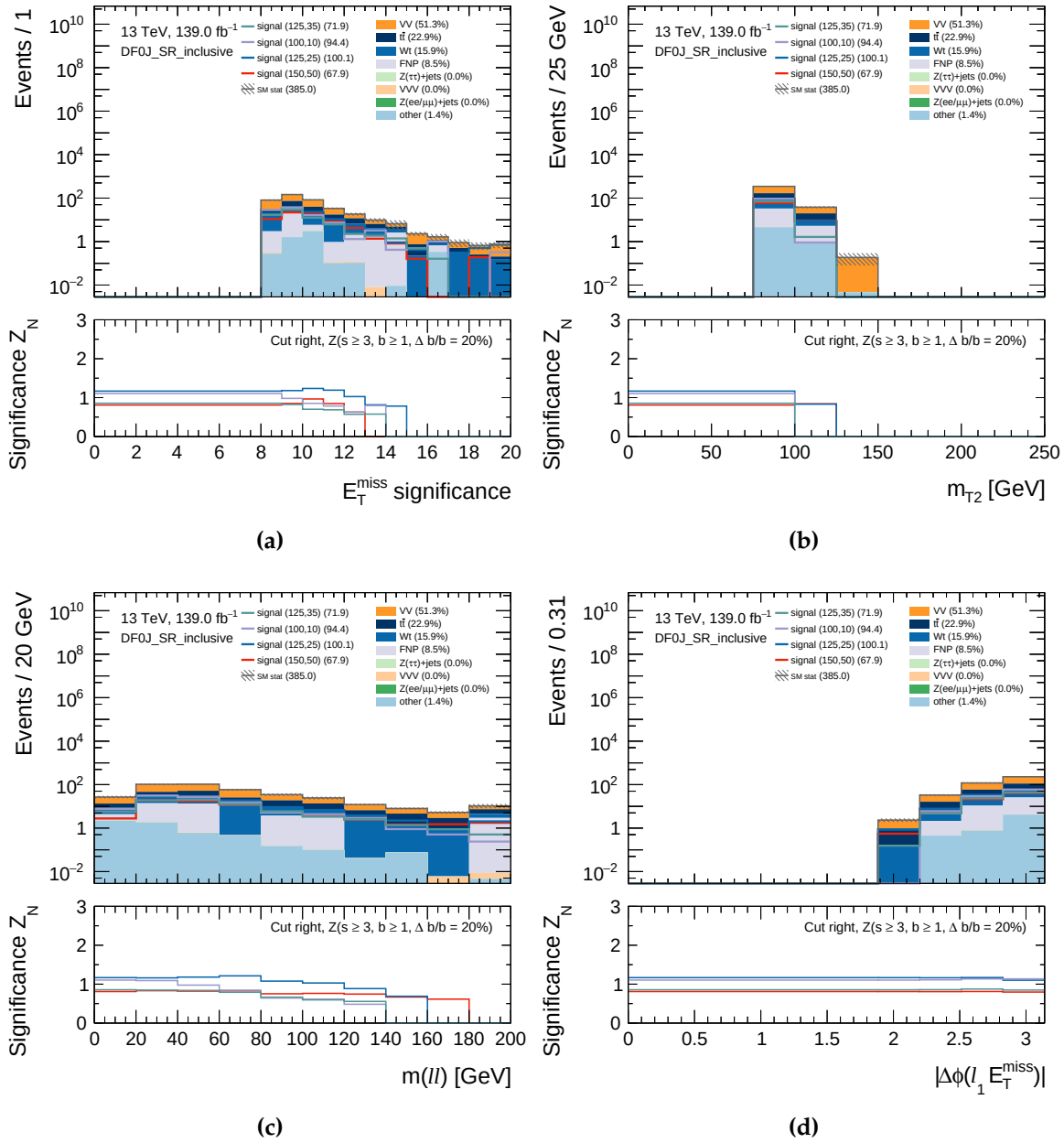


Figure 4.16.: Kinematic distributions of Monte Carlo simulated signals and backgrounds in the loosest DF inclusive signal region defined by BDT-signal > 0.81, E_T^{miss} significance > 8 and $m_{\text{T2}} > 50\,\text{GeV}$. The significance in the bottom panel is calculated using Equation 4.1, with a selection to the right for BDT-signal and to the left for the other BDT scores. The FNP backgrounds are calculated as described in Section 4.6.2, and 'other' includes all other rare backgrounds.

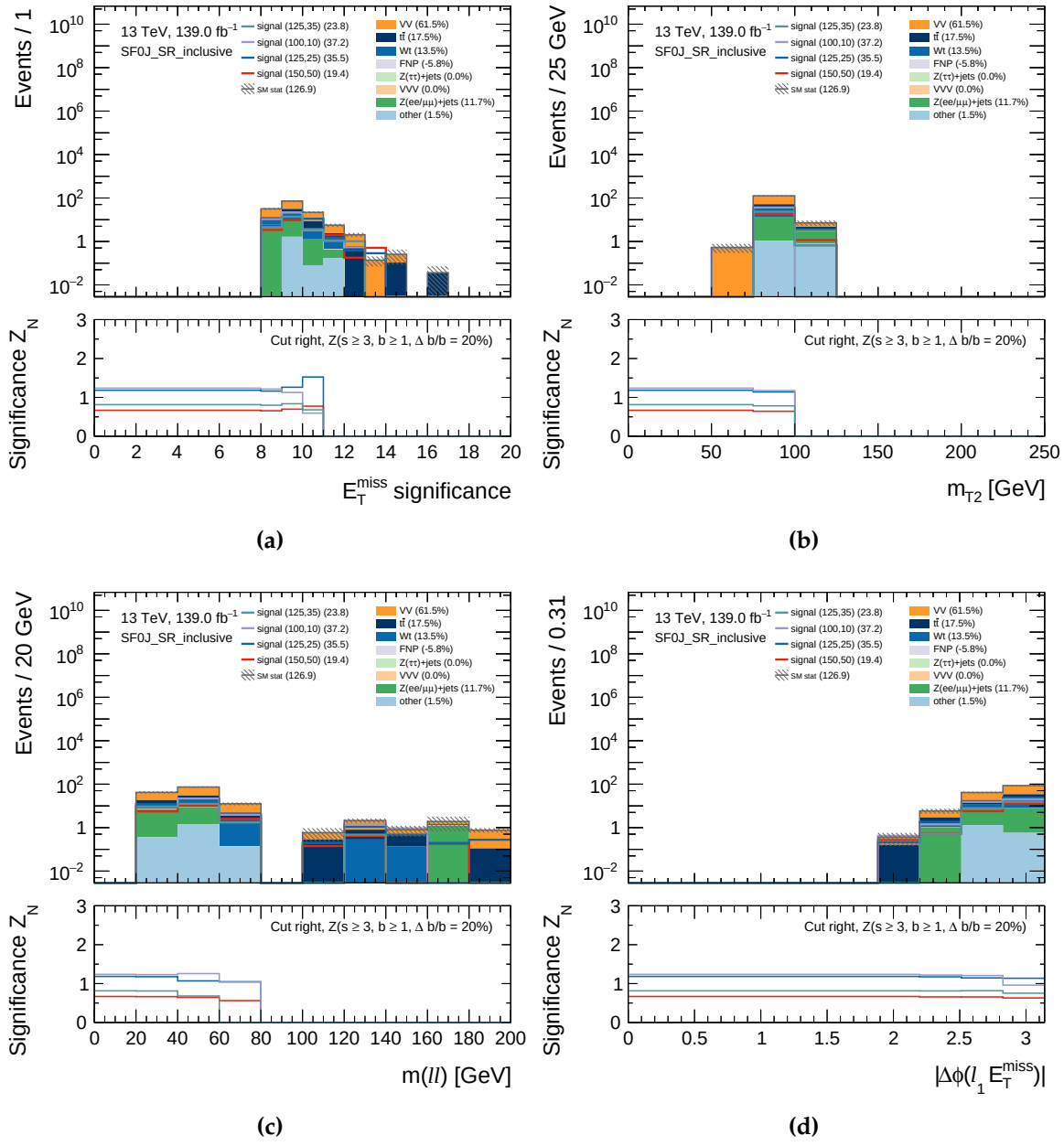


Figure 4.17.: Kinematic distributions of Monte Carlo simulated signals and backgrounds in the loosest SF inclusive signal region defined by BDT-signal > 0.77, E_T^{miss} significance > 8 and $m_{\text{T2}} > 50\,\text{GeV}$. The significance in the bottom panel is calculated using Equation 4.1, with a selection to the right for BDT-signal and to the left for the other BDT scores. The FNP backgrounds are calculated as described in Section 4.6.2, and 'other' includes all other rare backgrounds.

4.6. Background estimation

Having accurate background modelling is important to be able to search for new physics processes. This section outlines the background estimation techniques performed. The dominant irreducible backgrounds in the signal regions are the VV, $t\bar{t}$ and Wt backgrounds. These are normalised to data in dedicated control regions (CRs) using a simultaneous likelihood fit. These control regions are defined to be enriched in the desired background, while having a high yield to reduce statistical uncertainties. Moreover, they are designed to have similar kinematic properties to the signal regions, whilst having a low signal contamination. The normalisations from the control regions are extrapolated to the signal regions. Hence, validation regions (VRs) are defined in-between the control and signal regions in order to verify that the modelling is accurate before unblinding the signal regions. The validation regions do not constrain the fit, rather they allow us to qualitatively assess the agreement between the predicted background and the data. Each of these defined regions are orthogonal in the fit, as illustrated by the schematic in Figure 4.18, where we use requirements on BDT-signal, $n_{b\text{-tagged jets}}$ and BDT-VV to ensure this.

The reducible backgrounds are estimated using the Matrix Method [89], as described below. All the other backgrounds have a small effect in the signal regions and are estimated directly from Monte Carlo simulation.

4.6.1. Irreducible backgrounds

The signal regions are located in the region of phase space with high BDT-signal score with events containing zero jets. In order for all the defined regions to be mutually-orthogonal, the control and validation regions are defined in regions with lower BDT-signal score, or a requirement on having one b-tagged jet. Figure 4.18 illustrates this structure. For each region, further selections on BDT-VV, BDT-top and BDT-others are applied to increase the purity of the desired background, and

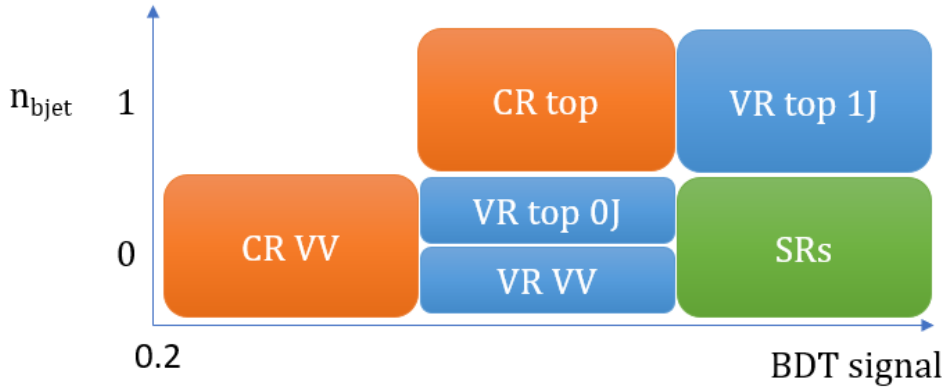


Figure 4.18.: A schematic outlining the locations of the signal regions (SRs), validation regions (VRs) and control regions (CRs) in the 2-dimensional phase space mapped by BDT-signal and $n_{b\text{-tagged jets}}$. Further selections are made on the BDT-VV, BDT-top and BDT-others scores to define regions enriched in VV and top backgrounds. Each region is orthogonal, using selections on the BDT-signal, $n_{b\text{-tagged jets}}$ (as illustrated), and BDT-VV.

in the case of VR-top-0J and VR-VV the selections on BDT-VV ensure orthogonality. The details of these definitions are described in the following sections, with the final definitions reported in Table 4.6. The construction of the VR-top-0J regions were performed by a collaborator.

VV control and validation regions

The control and validation regions are defined by looking individually at the DF and SF channels. The requirements E_T^{miss} significance > 8, $m_{\text{T2}} > 50\,\text{GeV}$ and BDT-others < 0.01 (for the SF channel) are applied to ensure we are kinematically close to the signal regions. For the control regions, the region 0.2 $<$ BDT-signal $\leq$ 0.65 is used. The validation regions lie between control and signal regions; 0.65 $<$ BDT-signal $\leq$ 0.81 and 0.65 $<$ BDT-signal $\leq$ 0.77 for DF and SF, respectively.

A combination of selections on the three scores BDT-VV, BDT-top and BDT-others are considered in order to increase the purity of VV (fraction of VV events), whilst ensuring a large VV event yield. There is typically a balance between these two

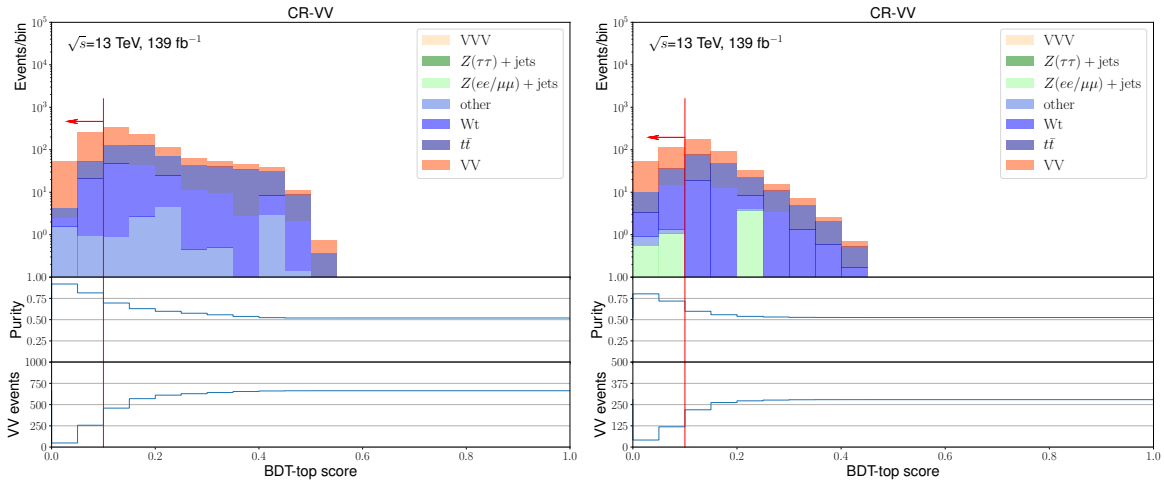


Figure 4.19.: Distributions of the BDT-top output of the Monte Carlo backgrounds in the (a) DF and (b) SF channel, used to define the VV control regions. The plotted region is defined by the preselection selections in Table 4.2, E_T^{miss} significance > 8, $m_{\text{T2}} >$ 50 GeV, 0.2 $<$ BDT-signal $\leq$ 0.65 and BDT-VV $>$0.2. The fraction of VV events (purity), and number of VV events when making a selection to the left are shown in the panels below. The red arrow indicates the selection value used to define the VV control regions.

requirements, since having tighter selections gives regions with fewer events, but also often increases the purity.

The best combination was found by combining a relatively loose selection of BDT-VV > 0.2 with a tighter selection of BDT-top < 0.1 for both the DF and SF channels. The plots illustrating the final BDT-top selection, and the resultant purity and VV event yield, can be seen in Figure 4.19 for both DF and SF. In this case, we obtain purities of 82% and 72% in the DF and SF channels, respectively. When initially performing the fit with the control regions separated, both channels were observed to have very similar normalisation factors. Hence, the DF and SF channels were combined to one control region 'CR-VV'.

The process of defining the validation regions was identical, and the same BDT-VV > 0.2 with a tighter selection on BDT-top < 0.1 were found to be optimal. These define the validation regions, VR-VV-DF and VR-VV-SF.

Top control and validation regions

The process of defining the control and validation regions for the top backgrounds is the same as previously described for the VV regions. Firstly, selections on E_T^{miss} significance > 8, $m_{\text{T2}} > 50\,\text{GeV}$ and BDT-others < 0.01 (for the SF channel) are applied. Then, we use the fact that the top backgrounds produce b-jets to isolate these by requiring a region with one b-tagged jet. This effectively selects the top backgrounds; Figure 4.20 indicates the purity improvement from a selection on BDT-top when considering the control regions. No further selection is necessary – a region with 98% purity for both the DF and SF channel is obtained without a requirement on BDT-top. Just as for the VV backgrounds, very similar normalisation factors were observed, so the DF and SF control regions top are combined in the fit as the 'CR-top'.

For the top-validation regions, we have two different types. These have either one b-tagged jet and a high BDT-signal score, or zero jets and an intermediate BDT-signal score. For the former region, we require no further selections on BDT-top to ensure good purity for VR-top-DF and VR-top-SF. For the latter region, a low requirement on BDT-$VV < 0.15$ is applied to both increase top purity, and ensure orthogonality to the VV validation regions to obtain VR-top0J-DF and VR-top0J-SF. This region validates the extrapolation of the top normalisation across the number of b-tagged jets.

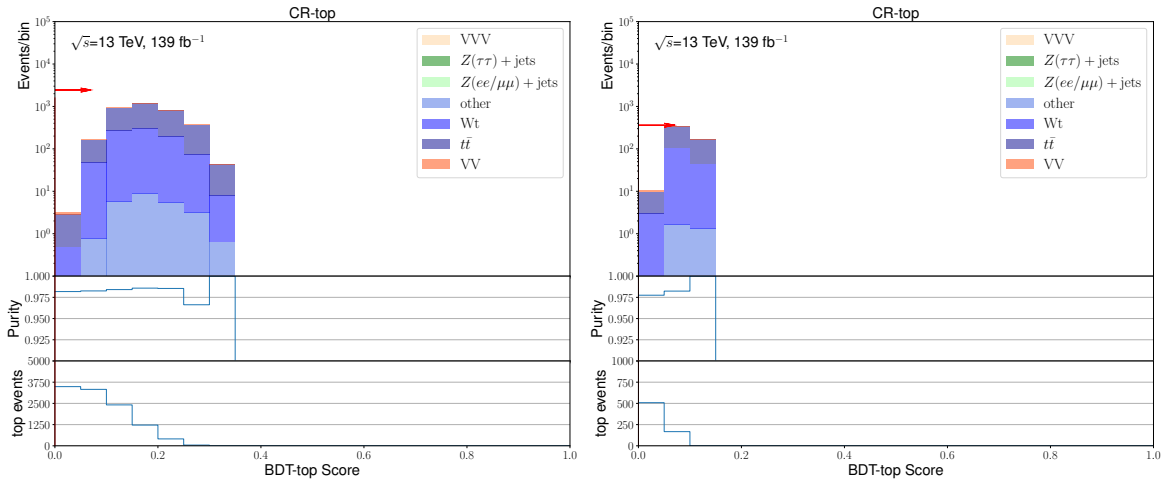


Figure 4.20.: Distributions of the BDT-top output of the Monte Carlo backgrounds in the (a) DF and (b) SF channel, used to define the top control regions. The plotted region is defined by the preselection requirements in Table 4.2, E_T^{miss} significance > 8, $m_{\text{T2}} > 50\,\text{GeV}$, $0.5 <$ BDT-signal ≤ 0.7 for DF and $0.7 <$ BDT-signal ≤ 0.75 for SF. The fraction of top events (purity), and number of top events when making a selection to the right are shown in the panels below. The red arrow indicates that no requirement on the BDT-top score was needed to define the top control regions.

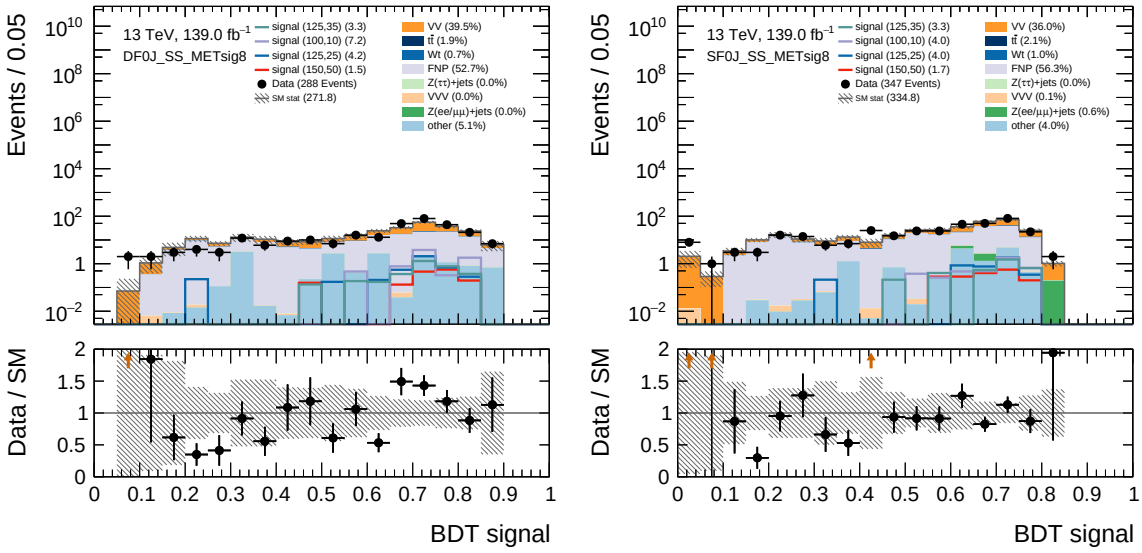
Control region (CR)	CR-VV		CR-top	
E_T^{miss} significance	> 8			
m_{T2} [GeV]	> 50			
$n_{\text{non-}b\text{-tagged jets}}$	$= 0$			
Leptons flavour	DF	SF	DF	SF
$n_{b\text{-tagged jets}}$	$= 0$	$= 0$	$= 1$	$= 1$
BDT-other	-	< 0.01	-	< 0.01
BDT-signal	$\in (0.2, 0.65]$	$\in (0.2, 0.65]$	$\in (0.5, 0.7]$	$\in (0.7, 0.75]$
BDT-VV	> 0.2	> 0.2	-	-
BDT-top	< 0.1	< 0.1	-	-

	VR-VV-DF	VR-VV-SF	VR-top-DF	VR-top-SF	VR-top0J-DF	VR-top0J-SF
E_T^{miss} significance	> 8					
m_{T2} [GeV]	> 50					
$n_{\text{non-}b\text{-tagged jets}}$	$= 0$					
$n_{b\text{-tagged jets}}$	$= 0$	$= 0$	$= 1$	$= 1$	$= 0$	$= 0$
BDT-other	-	< 0.01	-	< 0.01	-	< 0.01
BDT-signal	$\in (0.65, 0.81]$	$\in (0.65, 0.77]$	$\in (0.7, 1]$	$\in (0.75, 1]$	$\in (0.5, 0.81]$	$\in (0.5, 0.77]$
BDT-VV	> 0.2	> 0.2	-	-	< 0.15	< 0.15
BDT-top	< 0.1	< 0.1	-	-	-	-

Table 4.6.: The control and validation region definitions. The control regions are used to extract normalisation factors for the VV and top backgrounds, and the validation regions are used to assess the background modelling. Table reproduced from Ref. [1].

4.6.2. Reducible backgrounds

Reducible backgrounds have fewer than two prompt leptons, for example the W + jets process. This produces one real lepton from the W-boson and a FNP lepton, for example from the decays of a hadron within a jet, or the reconstruction of the jet as a lepton. These fake effects are related to the detector, so are not expected to be accurately modelled in simulation. The data-driven matrix method (MM) [89] is used to estimate FNP background. The FNP background estimate was performed by a collaborator, with a brief description provided below. Further details can be found in Refs. [1, 2].



(a) **(b)**

Figure 4.21.: The data and Monte Carlo background distribution of the BDT-signal score for (a) DF and (b) SF in the SS validation region defined by the preselection requirements in Table 4.2 with SS leptons, E_T^{miss} significance > 8 and $m_{\text{T2}} > 50\,\text{GeV}$. The uncertainties shown are statistical. The FNP backgrounds are calculated from data, and 'other' includes all other rare backgrounds.

The MM uses *signal* and *baseline* leptons, as introduced in Section 4.3. The probability of passing signal identification, given that the object has passed baseline object identification, is calculated for both real and fake leptons, giving a *real-efficiency* and a *fake-rate* respectively. With these probabilities, equations relating the observed number of events to the contributing real and fake components can be written as a matrix multiplication. By inverting this matrix, an estimate of the fake contribution to the signal leptons in the analysis regions can be calculated as a function of the observed data, real-efficiency and fake-rate. The real-efficiency and fake-rate are calculated in specific fake control regions as functions of p_{T} and η.

In order to validate the FNP modelling, a validation region with SS leptons is considered. This consists of the preselection requirements in Table 4.2 with SS leptons instead of OS leptons and E_T^{miss} significance > 8. Figure 4.21 shows the BDT-signal distribution in this region, which shows good agreement across the entire distribution.

4.7. Systematic uncertainties

There are several sources of uncertainty relating to the background estimates, and understanding these uncertainties is essential to being able to search for new physics. There are statistical uncertainties related to the expected yields in the analysis regions, and systematic uncertainties. The systematic uncertainties come in two categories: experimental uncertainties, for example relating to the calibration and reconstruction of physics objects; and theoretical uncertainties from the Monte Carlo modelling of the background. These are evaluated by varying a given parameter within its error, propagating this change through the analysis pipeline, and then evaluating the effect on the yields in the analysis regions by comparing the yields with the nominal yields obtained. The values for the dominant systematic uncertainties are reported in Table 4.7.

The dominant experimental uncertainties relate to the jets and the $E_{\mathrm{T}}^{\mathrm{miss}}$. These have contributions to the background uncertainty between 5 and 15 % in the signal regions. The process of calibrating the four-momenta of reconstructed jets gives the jet energy scale uncertainty (JES). The spread of the energy of these calibrated jets around the true energy is characterised by the jet energy resolution (JER). The uncertainties on the JES and JER are derived using dijet, Z+jet and γ+jet data and simulated events and are calculated as a function of the p_{T} and η of the jet [79, 127, 128]. Additionally, uncertainties relating to the modelling of the soft term in $E_{\mathrm{T}}^{\mathrm{miss}}$ is calculated by considering the scale and the resolution of the soft term using $Z \to ee$ events [81]. Additional subdominant experimental uncertainties include those related to the FNP estimate, and related to leptons. These uncertainties have much smaller contributions of typically less than 1 % in the signal regions. The FNP uncertainties are calculated by evaluating the effect of experimental uncertainties on the real-efficiency and fake-rate. The uncertainties on the leptons include the electron and muon energy scale and resolution, the reconstruction efficiency and trigger efficiencies [75, 76].

Finally, uncertainties on the b-tagging scale factor, pile-up modelling and luminosity were considered but were negligible in all regions used in the analysis.

Theoretical uncertainties relate to the Monte Carlo simulation, and in particular the choices made for certain parameters which do not have theoretically predicted values. We calculate an uncertainty related to our choice of these parameters. These are considered for the VV, top and $Z/\gamma^*(\rightarrow \ell\ell)$+jets backgrounds, however are negligible for $Z/\gamma^*(\rightarrow \ell\ell)$+jets . For these we use parton level backgrounds, with a smearing procedure used to emulate detector reconstruction, following the recommended procedure from the ATLAS supersymmetry group. The inputs for these theoretical uncertainties were prepared by a collaborator. These have a smaller effect than the dominant experimental systematic uncertainties, of between 1 and 10% in the signal regions.

For the VV background, the effect of the parton distribution (PDF) uncertainties is considered. In addition, the QCD factorisation scale, QCD renormalisation scale, resummation scale, and CKKW scale are varied up and down in order to derive an uncertainty [129]. For the $t\bar{t}$ background, the QCD factorisation scale and QCD renormalisation scale are likewise considered. The effect of the choice of generator for the parton showering and the event generation, is assessed by comparing each to a different respective generator. The effect of initial-state radiation (ISR) and final-state radiation (FSR) are calculated by varying parameters in the generator [130]. Finally, for single-top there is an uncertainty calculated concerning the modelling of the Wt and $t\bar{t}$ interference [131, 132].

Region	SR-DF	SR-SF
Signal BDT score	$\in (0.81, 1]$	$\in (0.77, 1]$
Total background expectation	474	168
$E_{\mathrm{T}}^{\mathrm{miss}}$ modelling	6.5%	11.4%
Jet energy scale	4.3%	5.2%
VV normalisation	3.2%	3.5%
Diboson theoretical uncertainties	2.7%	8.3%
Top theoretical uncertainties	1.4%	3.0%
FNP leptons	1.3%	$< 1\%$
Top normalisation	1.1%	$< 1\%$
Jet energy resolution	$< 1\%$	5.3%
Lepton modelling	$< 1\%$	1.3%
MC statistical uncertainties	1.7%	4.5%
Total systematic uncertainty	9.7%	18.4%

Table 4.7.: The breakdown of the dominant systematic uncertainties in the two inclusive signal regions SR-DF BDT-signal $\in (0.81, 1]$ and SR-SF BDT-signal $\in (0.77, 1]$. The quoted values are relative to the total background expectation in each region.

4.8. Results and statistical interpretation

The results are obtained from a simultaneous profile likelihood fit [133] to data in the control regions and signal regions. A likelihood function is constructed for the observed data, and is a function of the parameters of the model. These parameters are the normalisation factors for the VV and top backgrounds, and the nuisance parameters corresponding to the systematic uncertainties. The contributions from the remaining 'Other' backgrounds are minimal, and are taken directly taken from Monte Carlo. Maximum likelihood fits are performed to calculate the model parameters that best fit the data.

A first maximum likelihood fit is performed just using the control regions, assuming no signal contamination. This is the *background-only* fit, from which we can derive

normalisation factors for the VV and top backgrounds. These normalisation factors are first assessed in the validation regions, before being used in the signal regions when looking at the data during *unblinding*.

Since no significant excesses are seen of data over the predicted backgrounds in the signal regions, we perform exclusion fits – both model-independent and model-dependent. The model-dependent exclusion fit uses the exclusive, binned, signal regions to set limits on the specific model of $\tilde{\chi}_1^\pm$ decaying via W-bosons to $\tilde{\chi}_1^0$. These limits are set for each model point, defined by the masses of the $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_1^0$, and are plotted in the m($\tilde{\chi}_1^\pm$)-m($\tilde{\chi}_1^0$) plane for visualisation. The model-independent exclusion fit quantifies how much of any new signal model can be present in the inclusive signal regions defined in Section 4.5.

The statistical analysis is performed in `pyhf` [134, 135], using inputs from `HistFitter` [126] produced by a collaborator. `Cabinetry` [136] is used for manipulation of the `pyhf` output for generation of yield tables, post-fit plots and evaluation of the contributions of systematic uncertainties. The author liaised with the `cabinetry` author in order to produce new `cabinetry` features in order to produce these yield tables and post-fit plots. Example scripts using these features can be found in Ref. [137].

4.8.1. Background-only fit

Firstly, we perform the background-only fit; a maximum likelihood fit in the control regions in order to calculate normalisation factors for the dominant backgrounds. This fit assumes no signal is present in the control regions. A simultaneous fit to the data in both CR-VV and CR-top is performed in order to derive the normalisation for the VV, Wt, $t\bar{t}$ backgrounds. Table 4.8 indicates the yields in the control regions after the background-only fit, and Figure 4.22 shows some example kinematic plots in the control regions.

The fitted parameters are then used in the validation regions in order to assess the background modelling in an area of phase space close to the signal regions. The yields in this region can be seen in Table 4.9. This is an important step before looking at the data in the signal regions, to ensure that we are satisfied with the background modelling. Figures 4.23 and 4.24 show example kinematic distributions in the validation regions. We see consistently good modelling in the validation regions, with the data and post-fit yields agreeing within uncertainties, so we have confidence in the background modelling. We are now ready for unblinding of the signal regions and searching for new physics.

Figure 4.25 summaries the yields in the signal regions, with the significance as calculated with Equation 4.1 shown in the panel below for each individual signal region bin. Unfortunately, there are no significant excesses in any of the signal region bins, indicating that the data are consistent with the Standard Model hypothesis within uncertainties. This allows us to set limits on rates of new physics processes using exclusion fits, as described in the following sections.

Region	CR-VV	CR-top
Observed events	634	4468
Fitted backgrounds	634 ± 25	4468 ± 67
Fitted VV	520 ± 27	68 ± 12
Fitted $t\bar{t}$	69 ± 7	3243 ± 105
Fitted single top	40 ± 6	1129 ± 89
$Z/\gamma^*(\rightarrow \ell\ell)$+jets	$1.6^{+5.4}_{-1.6}$	–
Other backgrounds	3.0 ± 0.5	28 ± 5
FNP leptons	$0.011^{+1.41}_{-0.011}$	$0.019^{+11.88}_{-0.019}$
Simulated VV	376	49
Simulated $t\bar{t}$	63	2974
Simulated single top	37	1040

Table 4.8.: The observed and predicted event yields in the control regions, after the background-only fit. The uncertainties include both systematic and statistical uncertainties. The FNP are calculated as described in Section 4.6.2, and 'other' includes all the other backgrounds.

Regions	VR-VV-DF	VR-VV-SF	VR-top-DF	VR-top-SF	VR-top0J-DF	VR-top0J-SF
Observed events	972	596	1910	95	810	17
Fitted backgrounds	941 ± 61	668 ± 80	1896 ± 40	101 ± 10	881 ± 42	18 ± 3
Fitted VV	727 ± 51	402 ± 38	32 ± 13	2.2 ± 2.1	427 ± 29	8.1 ± 2.6
Fitted $t\bar{t}$	116 ± 12	111 ± 9	1348 ± 48	67 ± 6	260 ± 21	5.8 ± 1.7
Fitted single top	94 ± 19	75 ± 10	503 ± 57	27 ± 7	168 ± 18	4 ± 1
Other backgrounds	3.0 ± 1.5	72 ± 69	13.6 ± 2.5	0.7 ± 0.4	5.2 ± 1.9	0.05 ± 0.05
FNP leptons	$0.01^{+2.3}_{-0.01}$	7 ± 4	$0.015^{+5}_{-0.015}$	4.2 ± 1.3	20 ± 8	$0.05^{+0.15}_{-0.05}$
Simulated VV	527	291	23	1.6	309	5.9
Simulated $t\bar{t}$	106	102	1240	61	239	5.3
Simulated single top	87	69	460	25	154	3.2

Table 4.9.: The observed and predicted event yields in the validation regions, after the background-only fit in the control regions. The uncertainties include both systematic and statistical uncertainties. The FNP backgrounds are calculated as described in Section 4.6.2, and 'Other backgrounds' includes all the other backgrounds.

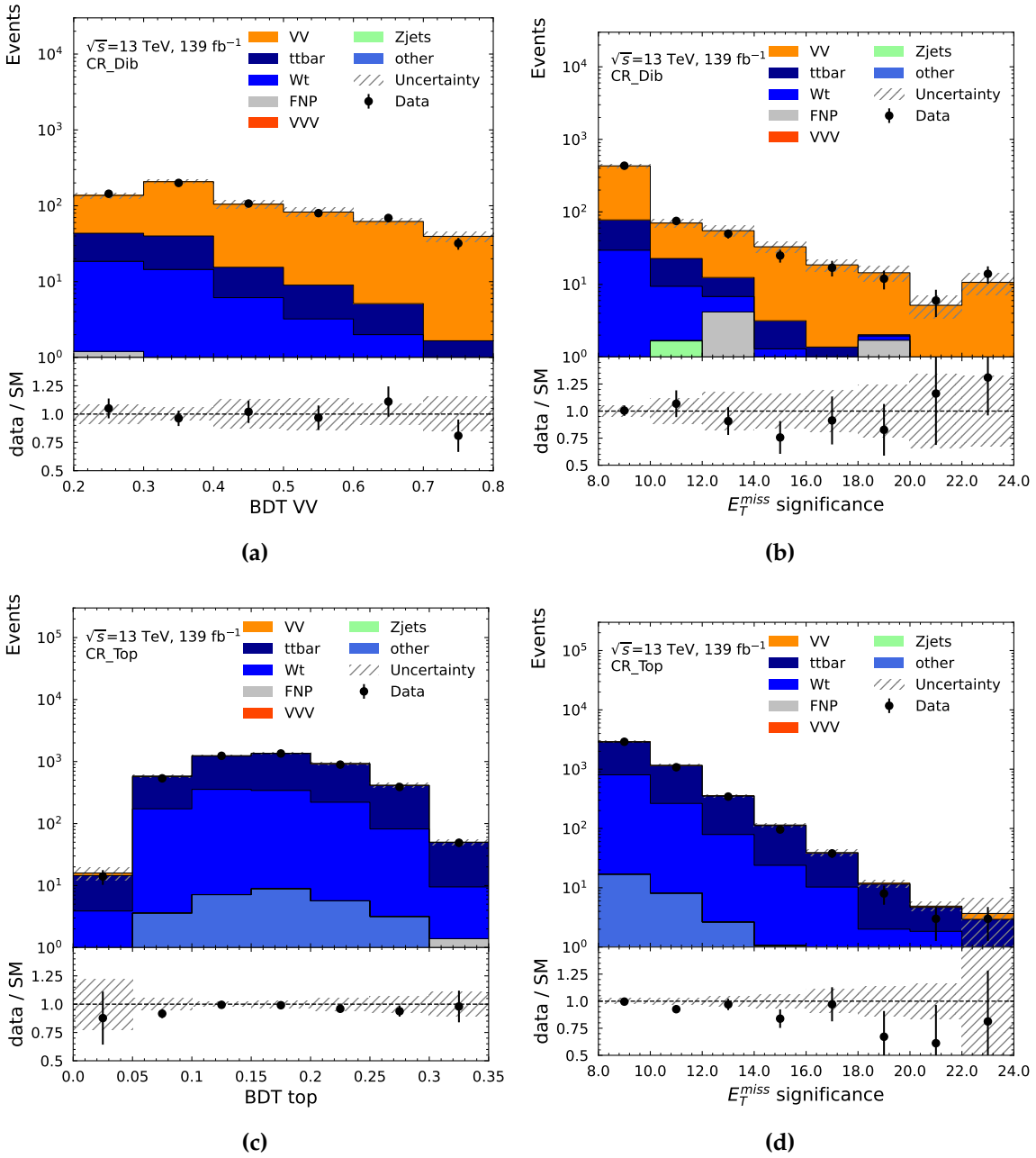


Figure 4.22.: Kinematic distributions of the data and background predictions in the control regions after the background-only fit. The uncertainties plotted include both the statistical and systematic uncertainties. The FNP backgrounds are calculated as described in Section 4.6.2, and 'other' includes all other rare backgrounds.

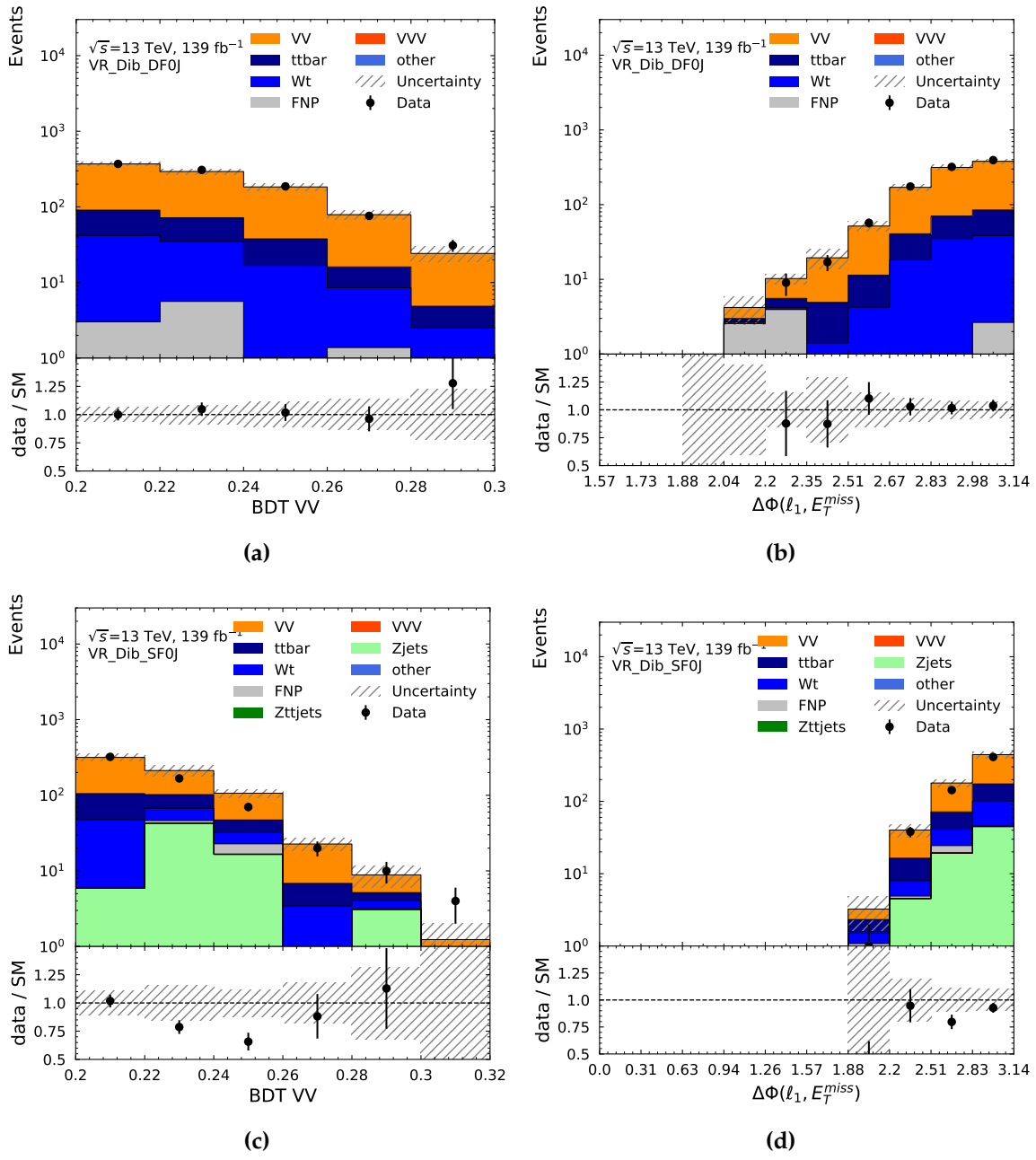


Figure 4.23.: Kinematic distributions of the data and background predictions in the VV validation regions after the background-only fit. The uncertainties plotted include both the statistical and systematic uncertainties. The FNP backgrounds are calculated as described in Section 4.6.2, and 'other' includes all other rare backgrounds.

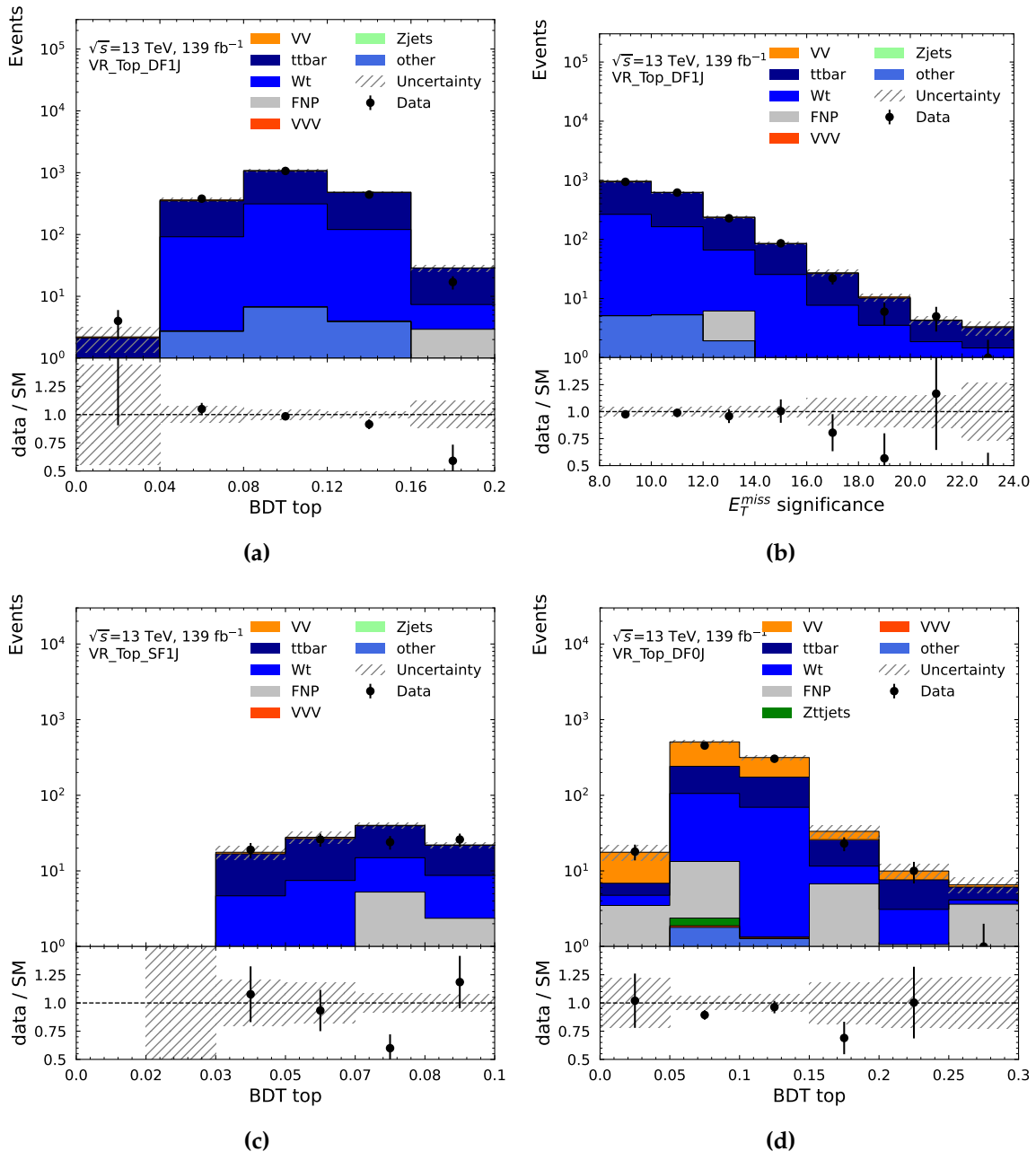


Figure 4.24.: Kinematic distributions of the data and background predictions in the top validation regions after the background-only fit. The uncertainties plotted include both the statistical and systematic uncertainties. The FNP backgrounds are calculated as described in Section 4.6.2, and 'other' includes all other rare backgrounds.

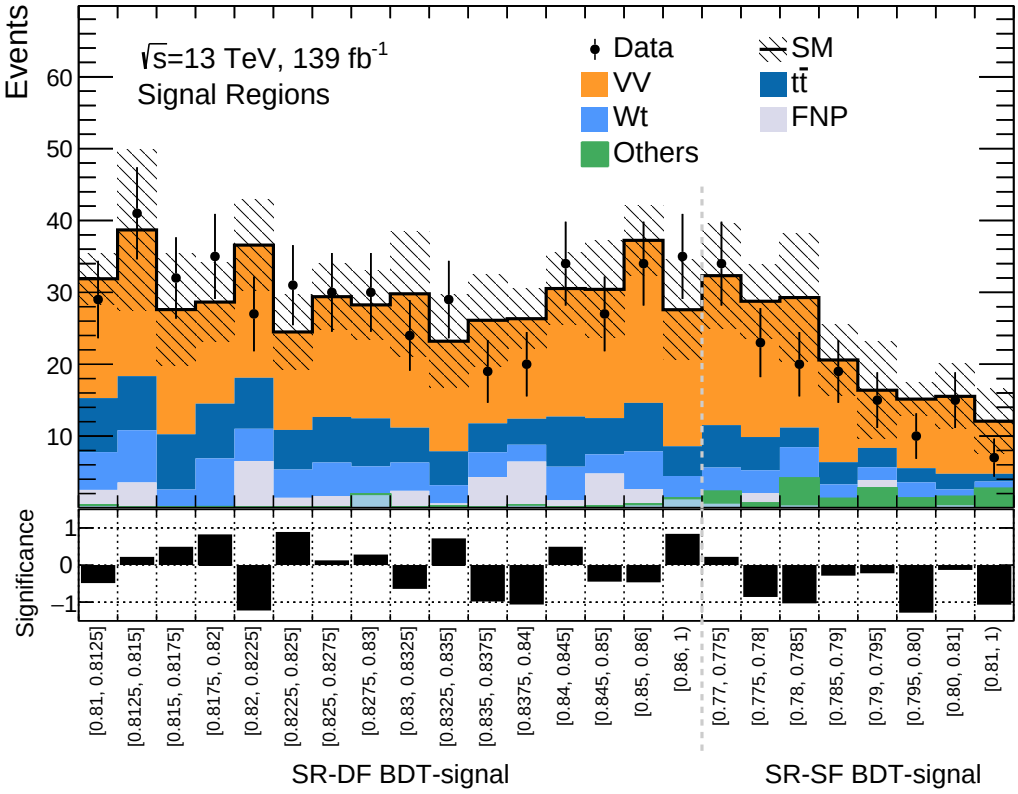


Figure 4.25.: The observed data and background predictions in the signal regions after the background-only fit. The uncertainty includes both systematic and statistical uncertainties. The lower panel shows the significance of the observed data in each bin as calculated with Equation 4.1. The FNP backgrounds calculated as described in Section 4.6.2 and 'Others' includes all other backgrounds.

4.8.2. Exclusion fit

In order to calculate limits on the number of new physics processes, we run an *exclusion* fit. We include signal into the fit, and correspondingly introduce the signal strength parameter μ_{SIG} which accordingly scales up and down the signal yield such that the expected number of events can be written as $N = B + \mu_{\text{SIG}}S$, for expected signal and background yields, S and B[2]. We wish to compare the background-only hypothesis that there is no signal ($\mu_{\text{SIG}} = 0$) to the signal plus background hypothesis ($\mu_{\text{SIG}} = 1$). We run a hypothesis test with the null hypothesis as the signal plus background hypothesis for each signal point, and wish to calculate whether we can reject this hypothesis at the 95% confidence level (CL).

In order to compare two hypotheses, we use a test statistic. This is essentially a single number which provides an ordering rule, where the order relates to increasing agreement between data and the hypothesis. We can calculate the probability of observing a given number or greater of data by integration of the test statistic. The Neyman-Pearson lemma [138] states that the most powerful test statistic for separating two hypotheses is given by the likelihood ratio. So we use a test statistic that is the negative logarithm of the profile likelihood ratio as described in Ref. [133]. Higher values of this test statistic indicate greater incompatibility of data with the hypothesised μ_{SIG} value.

We use p-values to quantify how dissimilar data are to predictions, with a small p-value indicating that the prediction is unlikely. More formally, a p-value indicates the probability, under a given hypothesis, of obtaining a value of the test-statistic equal to or greater than the value obtained from the observed data. It is common to quote the equivalent Gaussian significance of the p-values obtained with p-value = 0.05 corresponding to 1.64 σ, p-value = 0.001 corresponding to 3σ and p-value = 3×10^{-7} corresponding to 5 σ. If the p-value of the null hypothesis is less than 0.05 we can exclude the given model at 95% CL.

[2] We can express B = $\mu_{VV}B_{VV}$ + $\mu_{\text{top}}B_{\text{top}}$ + B′, where B_{VV}, B_{top} are the expected VV and top pre-fit yields and B′ is the expected yield of the remaining backgrounds.

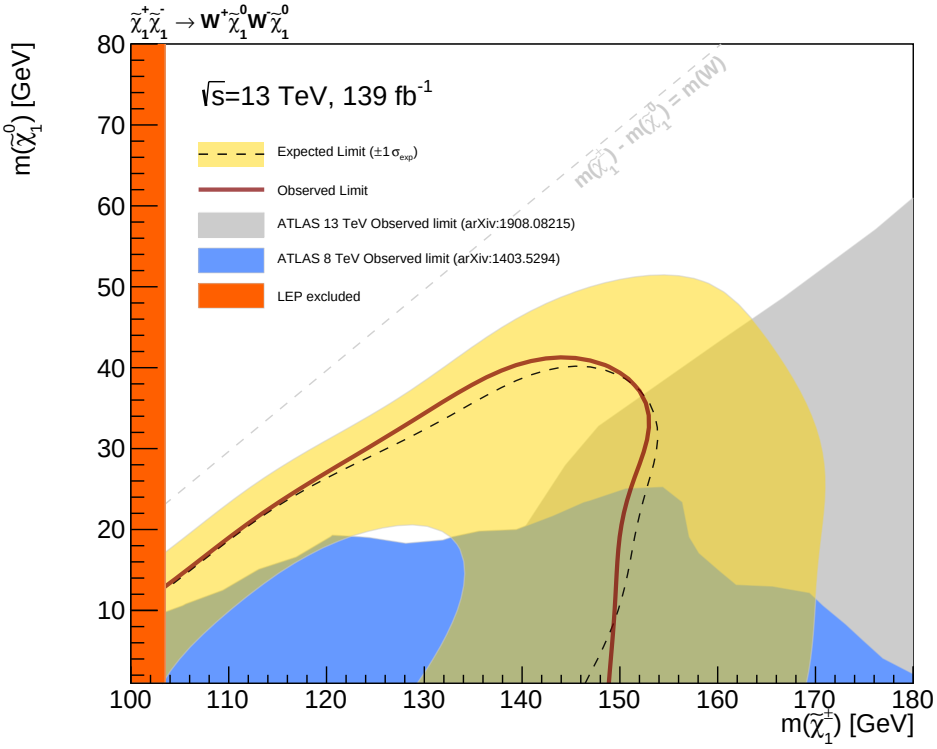
In order to calculate the p-value, we need to integrate above the observed test statistic value, to calculate this probability of observing the observed data or greater. However, the probability distribution function of the profile likelihood test statistic is not trivial to integrate. Fortunately, in the large-sample limit, the p-values of the test statistic can be calculated analytically [133]. This is the method used for calculation in this thesis. An alternative method exists, whereby the probability distribution of the test statistic can be estimated by performing multiple pseudo-experiments. However, this is very computationally intensive and is not necessary for these fits since we have sufficient yields in the signal regions for the asymptotic approximation to hold.

An important point to note is that if we have a large enough downward fluctuation of data in the signal region, we could exclude the background-only hypothesis at 95% CL. In this case, we could exclude any signal model, even when there is no sensitivity to the given model. This situation is rectified by normalising the p-value for the signal plus background hypothesis by the p-value for the background only hypothesis. This reduces the exclusion power when we have poor compatibility with the background only hypothesis, and is known as the CL_s prescription [139]. This normalised p-value, CL_s, is used to set limits: when $CL_s < 0.05$ we can claim exclusion for a signal point at 95% CL.

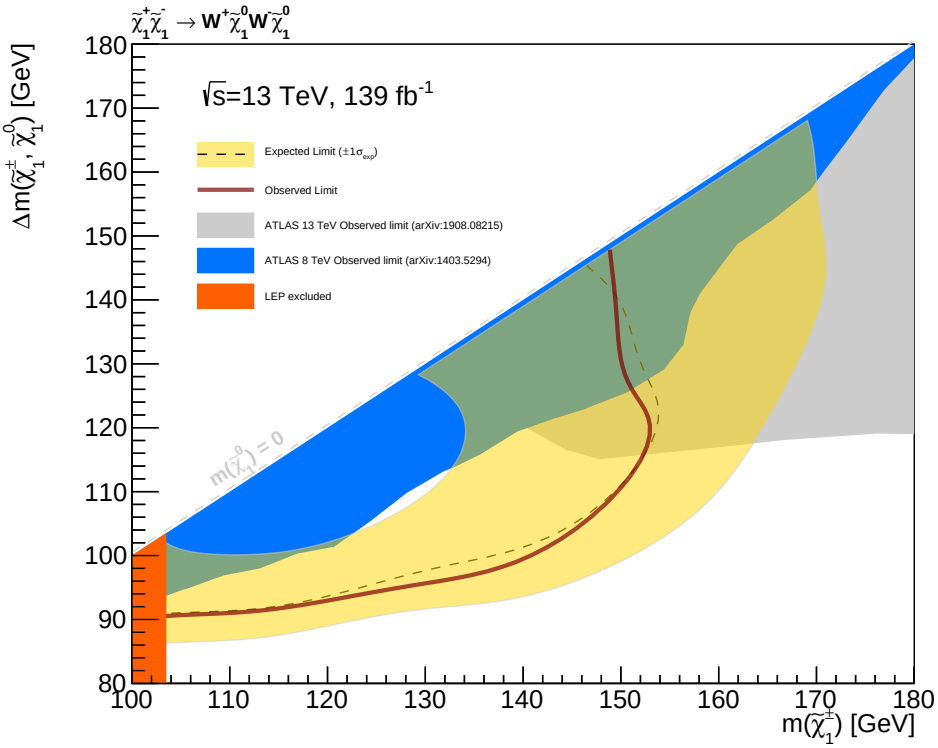
Exclusion fit (model dependent)

The model-dependent exclusion fit is run using the binned signal regions, in order to calculate CL_s values for signal points from the targeted supersymmetric model. Separate hypothesis tests are performed for each signal point, which are defined by different masses m($\tilde{\chi}_1^{\pm}$) and m($\tilde{\chi}_1^0$) as described in Section 4.2. A simultaneous likelihood fit is performed over the control regions and all the binned signal regions in order to calculate the CL_s values. Contamination from signal events is included in the control regions during this fit.

The CL_s values are plotted in the m($\tilde{\chi}_1^{\pm}$)–m($\tilde{\chi}_1^0$) and the Δm–m($\tilde{\chi}_1^0$) planes in Figure 4.26. The yellow band on the expected limit indicates the $\pm 1\sigma$ deviations from all the systematic uncertainties. Interpolated curves are drawn for the expected and observed limits between the CL_s values. The limit covers the gap between the LEP limit [140] and the previous 13 TeV ATLAS analysis [82]. The 8 TeV analysis [85] had no expected sensitivity, with the plotted observed sensitivity being solely due to an under-fluctuation in data compared to the expected yield. This analysis extends the observed limit from the 8 TeV analysis, and fills the previously missing gap in expected sensitivity between the LEP and 13 TeV analysis.



(a)



(b)

Figure 4.26.: The observed and expected exclusion limits for the supersymmetric signal model considered, shown in the (a) in m($\tilde{\chi}_1^\pm$)–m($\tilde{\chi}_1^0$) (b) m($\tilde{\chi}_1^\pm$) –Δm planes. The yellow band indicates the $\pm 1\sigma_{\text{exp}}$ from the systematic uncertainties. There was no expected sensitivity of the 8 TeV ATLAS analysis [85], so this gap in expected sensitivity has been covered by this analysis.

Exclusion fit (model-independent)

Finally, we perform the model-independent exclusion fit, looking for any new physics in an inclusive signal region. The fit consists of the two control regions, and one inclusive signal region.

As before, the signal strength parameter μ_{SIG} is used, but in this case is set to be equivalent to the number of events of any new physics process. By performing multiple hypothesis tests with varying μ_{SIG} values, we can evaluate the value where CL_s equals 0.05. From this, we can infer the number of events from any new physics process that can be excluded in the given signal region. The results from this model independent exclusion fit for each inclusive signal region are shown in Table 4.10, which include the fitted backgrounds yields in these signal regions. The observed and expected 95% CL limits on the number of beyond-the-SM events, $S^{0.95}_{\text{obs}}$ and $S^{0.95}_{\text{exp}}$, are shown. In addition, these tables include upper limits on the visible cross-section σ_{obs} = $S^{0.95}_{\text{obs}}$ / $\mathcal{L}$, which is the fraction of the total cross-section to which the signal regions have sensitivity.

A final *discovery* fit is performed for the model independent case. This is a hypothesis test where the null hypothesis is the background-only hypothesis. From this, we can calculate the p-value (p_0) of the background-only hypothesis ($\mu_{\text{SIG}} = 0$). If there were new physics events in the signal region, and the observed data were in excess of the expected data, then if we could exclude the background-only hypothesis at 5 σ we could claim a discovery. The p_0 values are reported in Table 4.10. These are very large, with most being > 0.5 due to a deficit of data compared to the expected yield. This indicates good compatibility of the data with the background-only hypothesis.

Table 4.10.: The observed event yields and predicted background yields from the background-only fit for the inclusive signal regions, as defined in Table 4.5. A '–' indicates a negligible contribution from the background. The lower section of each table indicates the model independent exclusion fit results; the observed and expected 95% CL limits on the number of new physics events, $S^{0.95}_{\text{obs}}$ and $S^{0.95}_{\text{exp}}$. $\sigma_{\text{obs}} = S^{0.95}_{\text{obs}} \,/\, \mathcal{L}$ is the upper limit on the visible cross-section. The final row indicates the p_0-value for the background-only hypothesis, as calculated with the discovery fit. This value is capped at 0.50. Note that the regions are overlapping, so the yields are not independent.

Region	SR-DF	SR-DF	SR-DF	SR-DF	SR-DF
Signal BDT score	$\in(0.81,1]$	$\in(0.82,1]$	$\in(0.83,1]$	$\in(0.84,1]$	$\in(0.85,1]$
Observed events	477	340	222	130	69
Fitted backgrounds	474 ± 46	348 ± 36	231 ± 26	126 ± 15	65 ± 10
Fitted VV	273 ± 28	204 ± 23	140 ± 17	77 ± 10	42 ± 7
Fitted $t\bar{t}$	96 ± 13	66 ± 10	40 ± 8	23 ± 6	10.9 ± 3.3
Fitted single top	67 ± 16	45 ± 12	28 ± 7	16 ± 5	8.3 ± 3.0
$Z/\gamma^*(\to \ell\ell)$+jets	–	–	–	–	–
Other backgrounds	5.4 ± 1.8	4.8 ± 1.2	2.8 ± 1.3	2.1 ± 0.7	1.8 ± 0.4
FNP leptons	33 ± 6	29 ± 6	21 ± 4	7.8 ± 1.2	2.2 ± 1.1
$S^{0.95}_{\text{obs}}$	106.30	75.95	52.19	39.91	30.86
$S^{0.95}_{\text{exp}}$	$104.3^{+41.9}_{-27.1}$	$81.4^{+32.8}_{-23.1}$	$57.8^{+22.7}_{-16.1}$	$37.4^{+14.9}_{-10.6}$	$27.9^{+12.0}_{-8.3}$
σ_{obs} [fb]	0.76	0.54	0.37	0.29	0.22
p_0	0.47	0.50	0.50	0.42	0.38

Region	SR-DF SF	SR-SF	SR-SF	SR-SF	SR-SF
Signal BDT score	-DF$\in(0.81,1]$ -SF$\in(0.77,1]$	$\in(0.77,1]$	$\in(0.78,1]$	$\in(0.79,1]$	$\in(0.80,1]$
Observed events	620	143	86	47	22
Fitted backgrounds	634 ± 74	168 ± 31	108 ± 22	58 ± 14	28 ± 7
Fitted VV	380 ± 46	108 ± 20	68 ± 13	36 ± 8	18 ± 5
Fitted $t\bar{t}$	120 ± 18	24 ± 6	14 ± 7	8 ± 8	3.3 ± 1.8
Fitted single top	85 ± 20	19 ± 7	12 ± 5	5.9 ± 2.7	1.9 ± 1.1
$Z/\gamma^*(\to \ell\ell)$+jets	15 ± 10	15 ± 10	13 ± 9	8 ± 6	4^{+5}_{-4}
Other backgrounds	7.3 ± 2.8	1.9 ± 1.4	1.1 ± 0.6	0.5 ± 0.4	0.49 ± 0.32
FNP leptons	26 ± 5	–	–	–	–
$S^{0.95}_{\text{obs}}$	165.93	65.51	43.21	30.23	16.88
$S^{0.95}_{\text{exp}}$	$175.2^{+69.2}_{-49.2}$	$80.7^{+34.2}_{-23.0}$	$55.9^{+21.8}_{-15.5}$	$37.1^{+14.4}_{-10.1}$	$20.3^{+10.2}_{-6.3}$
σ_{obs} [fb]	1.19	0.47	0.31	0.22	0.12
p_0	0.50	0.50	0.50	0.50	0.50

4.8.3. Conclusions

The analysis in this chapter searched for electroweak supersymmetry, focusing on the decay of $\tilde{\chi}_1^{\pm}$ to $\tilde{\chi}_1^0$ via leptonically decaying W-bosons. The region where there is small mass difference between the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_1^0$ had a lack of sensitivity from previous analyses. The analysis described in this chapter bridged this gap in sensitivity using machine learning methods. Multiclass classification was used, which enabled for very pure control and validation regions to be defined for the improvement of background modelling. Novel statistical tools, `pyhf` and `cabinetry`, have been utilised for fitting and presentation of results.

No significant deviations from the Standard Model hypothesis were observed, so model-dependent and model-independent limits have been set on new physics. Charginos with masses up to 140 GeV have been excluded at 95% CL for a mass-splitting of 100 GeV. This area of parameter space is particularly challenging, with soft leptons produced from the W-boson decays, and a very large kinematically similar contribution from the VV background. In order to search for even more compressed scenarios, leptonic triggers are not suitable, due to the soft leptons from the W-boson decay. Future searches for the more compressed region of this process could consider scenarios with an initial-state radiation (ISR) jet and using a trigger based on $E_{\mathrm{T}}^{\mathrm{miss}}$, such that lower thresholds on lepton momenta can be used.

With the lack of discovery of supersymmetry, searching in more challenging areas of parameter space is imperative. Starting to probe this very challenging region of parameter space with low Δm is an important and necessary step in searching for supersymmetry.

Chapter 5.

Hunting for unlikely forms of parity violation at the LHC

"Will you look into the mirror?"
"What will I see?"
"Even the wisest cannot tell. For the mirror shows many things."
— Galadriel and Frodo

5.1. Introduction

Searches for parity violation at the LHC are considerably more challenging than the search performed by Wu [8], as described in Section 1.1. Firstly, the LHC beams are unpolarised, and secondly the detectors are not sensitive to the spin of the produced particles. Considering matrix elements[1], conclusions can be made about parity-violating new physics models, as described in Ref. [141]: the terms in matrix elements can be parity-even and parity-odd, where the parity-odd terms contain a γ_5

[1] Matrix elements are related to the cross-section of scattering processes.

matrix. Performing the spin-averaging on the matrix elements involves performing traces of the gamma matrices, where traces containing a γ_5 can result in the antisymmetric tensor ϵ [142]. This leaves us with one parity-odd term $\epsilon_{\mu\nu\rho\sigma}p_1^\mu p_2^\nu p_3^\rho p_4^\sigma$, where p_1, p_2, p_3, p_4 are four momenta. This term is both parity-odd and time-odd, since it contains an odd number of spacelike and timelike momenta, respectively. Assuming CPT-symmetry, being time-odd implies it is also CP-odd. Hence, parity violation cannot be generated by a CP-conserving locally Lorentz invariant QFT.

Searches for parity violation at the TeV scales probed by the LHC have so far been lacking[2], not least due to a lack of theoretical motivation. That does not mean unforeseen sources of parity violation could in fact be present at these scales; non-Standard-Model parity-violating physics may be occurring at the LHC currently, however this violation could be as yet not seen simply because it is not being searched for.

This chapter introduces a method for performing data-driven searches for new parity-violating physics, which can be performed with no reference to a particular model [3]. This builds on the search strategies developed in Refs. [143, 144], in order to search for parity violation within a simulated collider-like multijet dataset. The dataset is simulated from a simplified model with parity-violation, which is from the minimal Standard Model Extension (mSME) framework [145].

A general introduction to the search strategy for parity-violating physics is presented in Section 5.2. Then, the generation of the parity-violating physics dataset is described in Section 5.3. The analysis strategy is outlined in 5.4, including a discussion of the construction of a CNN to respect cylindrical symmetries and the training of the CNN on an unlabelled dataset. Finally, results are presented in Section 5.5.

The MADGRAPH implementation of the parity-violating model was performed by collaborators. The implementation was performed by Rupert Tombs, with the theory developed by Radha Mastandrea. The author was solely responsible for the showering

[2]One controversial search was performed in Ref. [141], which is discussed further in this chapter.

in PYTHIA and the reconstruction of the samples in DELPHES, as well as the search with the CNNs.

5.2. Searches for parity violation at the LHC

In order to motivate parity searches in general, we can consider drawing from a distribution $\mathrm{D}(x)$, obtaining a set of events $\{x_1, x_2, ..., x_n\}$. By analysing these events, we wish to investigate whether $\mathrm{D}(x)$ violates parity symmetry or not. That is, whether $\mathrm{D}(x)$ differs from $\mathrm{D}(\mathrm{P}x)$; whether events are produced at the same rate as their parity flipped counterparts.

By constructing a suitable variable $f(x)$ that is odd under parity symmetry, we can perform such an analysis [14]. If $\mathrm{D}(x)$ is parity symmetric, $f(x)$ would have a symmetric distribution since it is parity-odd. Whereas, if $\mathrm{D}(x)$ violates parity symmetry, we could see an asymmetry in the distribution of $f(x)$. For example, if $f(x)$ has a significantly non-zero mean, we can conclude that $\mathrm{D}(x)$ does violate parity.

When considering searches at the LHC, we can note that there are certain symmetries of a cylindrical detector like ATLAS. We have two identical proton beams being collided, so we wish to be invariant to any rotations that leave the beams unchanged: rotations about the beam axis (in ϕ) and rotations 180° about a perpendicular axis (corresponding to swapping the beams). By blinding ourselves to these symmetries, we can simplify the analysis, by not having to consider symmetrically equivalent events that are represented differently. This is done in standard analyses, for example using event variables that are invariant to rotations in ϕ such as p_T and $m_{\ell\ell}$. For the analysis that follows, events that are equivalent under these symmetries are represented the same way. This is done by modifying the structure of the CNNs used for this search, as described in Section 5.4.2.

There has only been one search for non-standard parity-violating physics at the energy scales probed by the LHC [141]. This is a model-independent data-driven

search using the scalar triple product of jet momenta

$$\alpha = \arcsin\left(\frac{\vec{p}^{\,j_1} \times \vec{p}^{\,j_2}}{|\vec{p}^{\,j_1} \times \vec{p}^{\,j_2}|} \cdot \frac{\vec{p}^{\,j_3}}{|\vec{p}^{\,j_3}|}\right), \tag{5.1}$$

as the parity-odd event variable in order to perform the analysis in the aforementioned way. This event variable α is a pseudoscalar so is rotationally invariant and parity-odd by construction; it is parity-odd since each momentum vector gains a factor of –1 under a parity flip. Hence, by looking for an asymmetry in α we can perform a search for parity-violating new physics.

The authors of Ref. [141] acknowledge that the variable is not fully general nor optimal, intending for this search to simply be a first search for parity-violating physics at the LHC. We aim to improve on this variable, by constructing event variables that are more sensitive to new parity-violating physics. Moreover, we aim to have a more general method which can be performed on any number of jets, or other physics objects as desired.

Finally, seeing parity violation in *data* does neither necessarily imply that there is parity violation in *nature*. The data are affected by the response of the detector, which could introduce parity asymmetries. Known parity-violating effects in the detector can be dealt with as described in Ref. [144], however *unknown unknowns* could introduce unexpected parity asymmetries. Hence, seeing parity violation in data indicates either that we have new physics in nature *or* we that have a previously unforeseen parity-violating effect in the detector. The former is the ultimate aim of this method, however the latter is not at all a negative result; it allows us for better calibration and understanding of our detector, improving other analyses using the detector.

5.3. Data generation

5.3.1. Minimal Standard Model extension

The minimal Standard Model Extension (mSME) [145] is a framework of beyond the Standard Model physics which is Lorentz violating. It has various couplings, some of which can provide parity violation. For this analysis, we construct a parity violating model from the mSME by considering a specific part of its QCD coupling. We denote this simplified model we use as 'PV-mSME'. This can produce measurable parity-violating signatures in the ATLAS detector. We can control the amount of parity violation using a single real parameter λ_{PV}.

We consider parity violation in the quark-gluon sector, noting that the Standard Model quark-quark-gluon vertex has the form

$$f_\nu^{\text{SM}} = -ig_s \frac{\lambda_{ij}^a}{2} \gamma_\nu. \tag{5.2}$$

For the PV-mSME model considered, an axial vector term is introduced, containing with coupling matrix $(c_A)_{\mu\nu}$, giving a modified vertex

$$f_\nu^{\text{PV-mSME}} = -ig_s \frac{\lambda_{ij}^a}{2} \left(\gamma_\nu + \gamma^\mu (c_A)_{\mu\nu} \gamma^5 \right). \tag{5.3}$$

The coupling matrix is given by

$$(c_A)_{\mu\nu} = \lambda_{\text{PV}} \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \tag{5.4}$$

and λ_{PV} is the aforementioned coupling strength parameter. The indices μ and ν take values from 0 to 3, corresponding to the usual energy and momentum components of the four-momentum. $\lambda_{\text{PV}} = 1$ gives terms of comparable magnitude to the Standard Model terms, and $\lambda_{\text{PV}} = 0$ recovers the Standard Model. Appendix C of Ref. [3] contains further details about this model.

5.3.2. MADGRAPH simulation

We define these $(c_A)_{\mu\nu}$ couplings in the lab-frame and simulate the matrix-elements in the lab-frame using MADGRAPH. Samples containing three or four gluons or light quarks (up, down, strange or charm) are produced at $\sqrt{s} = 13\,\text{TeV}$. We first generate the benchmark coupling $\lambda_{\text{PV}} = 1$, for initial investigations. More samples are then produced, reducing λ_{PV} down to 0 in steps of 0.1, to generate a total of 10 mSME models in addition to the Standard Model ($\lambda_{\text{PV}} = 0$).

It should be noted that, looking at the distributions of reconstructed jet momenta in Figure 5.1 there are different kinematics produced by the PV-mSME. We do not concern ourselves with these differences, since our interest lies in detecting the parity violation within the dataset, rather than the specifics of the kinematic shapes. Moreover, it should be emphasised that we are not concerned with searching for a specific model, such as this PV-mSME. Rather, the aim is to develop a general method for model-independent searches for any parity-violating physics, that could occur in a detectable way at the LHC.

5.3.3. PYTHIA and DELPHES reconstruction

PYTHIA [114] is used to perform the parton shower and hadronisation and DELPHES [146–148] is used to approximate the detector effects of the ATLAS detector. Since full reconstruction using GEANT 4 [86] is slow, DELPHES is used as a fast-simulation alternative. An approximation to the ATLAS detector is used in

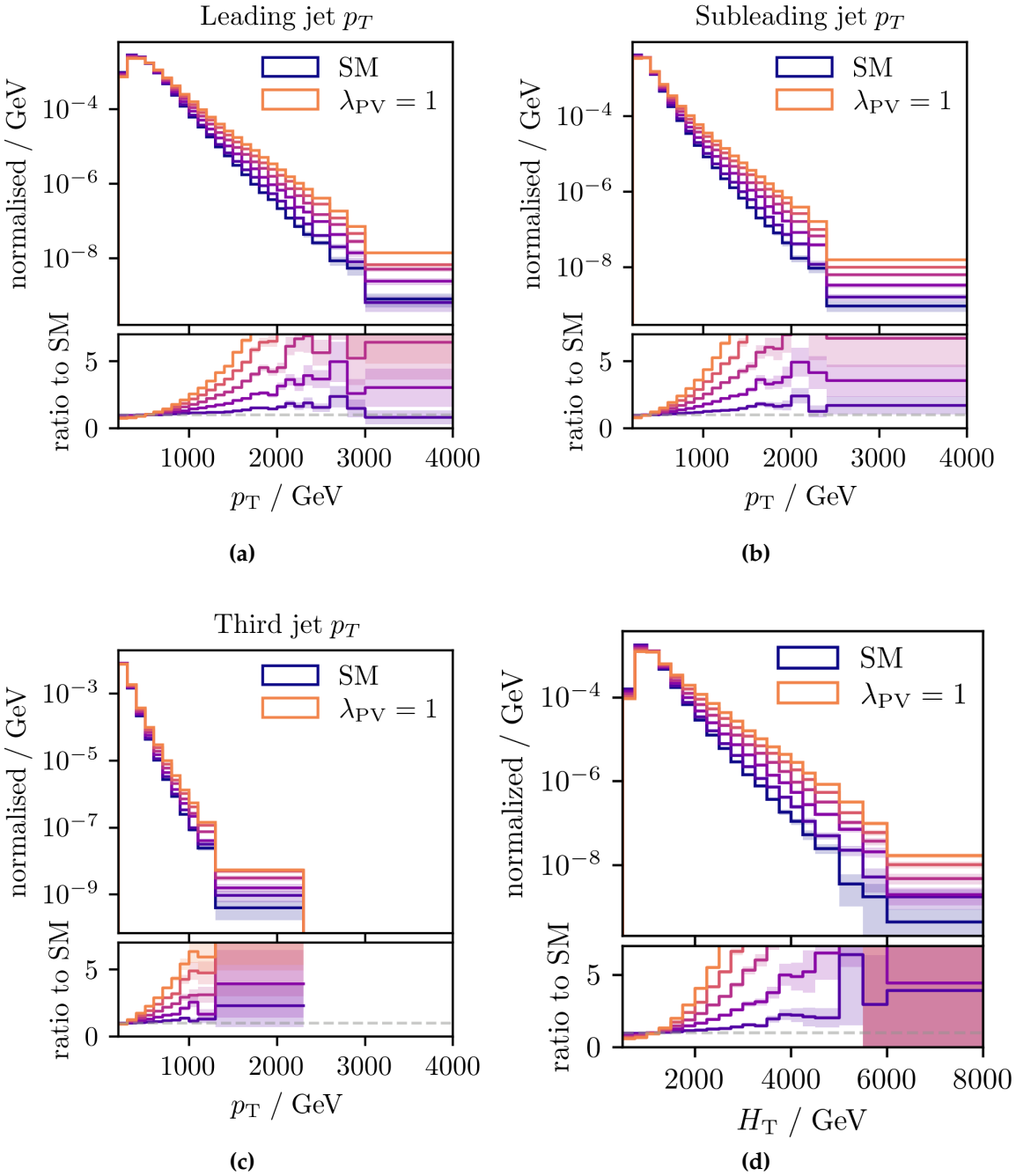


Figure 5.1.: The p_T distributions of the (a) leading (b) subleading (c) third jet and (d) the scalar sum of the leading three reconstructed jet p_T. The coloured lines linearly interpolate between the Standard Model ($\lambda_{PV} = 0$) and $\lambda_{PV} = 1$.

DELPHES, where stable particles are propagated within a uniform axial magnetic field, and the energy of these particles is deposited in the calorimeters.

In this approximation, the EM calorimeter and hadronic calorimeter are perfectly overlaid and segmented in η and ϕ. Electrons and photons deposit all their energy in the EM calorimeter, and hadrons deposit all their energy in the hadronic calorimeter. Then, the energies for each particle are smeared and summed to produce the total energy deposited in each calorimeter pixel. The energy deposits in the calorimeters are an output from DELPHES. Using these deposits, the jets are reconstructed using the Anti-kt algorithm [78] with $R = 0.4$.

Following reconstruction, kinematic selections are made on the leading three jets to approximate a three-jet trigger efficiency threshold. The leading three jets are required to have $p_{\mathrm{T}} > 220\,\mathrm{GeV}$ and $|\eta| < 2.8$, and an event is rejected which does not pass these criteria. The events passing the trigger criteria are used for further analysis, considering either the energy deposits or reconstructed jets as inputs for further analysis. When analysing the events using the reconstructed jets, any number of jets are used; all jets that have $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.8$ are included for analysis.

Pile-up simulation

The effect of pile-up is also considered, in order to more accurately represent the conditions at the ATLAS detector. Pile-up events are simulated using PYTHIA and overlaid along the z-axis of the hard scatter event. An average of 50 minimum bias events are overlaid per hard scatter event before reconstruction is performed by DELPHES. DELPHES subsamples from the simulated pile-up sample and overlays with a random ϕ-rotation. Since ϕ-invariant event variables are considered, we do not want the CNN to be able to individually learn each pile-up event that has been overlaid over the hard scatter event. Hence, for each batch of 200 000 hard scatter events generated, an individual sample of 200 000 pileup events are produced from which DELPHES can subsample.

The jet kinematics are similar when reconstruction is performed with and without pile-up, as can be seen in Figure 5.2. This is expected, since high p_{T} jets are considered and pile-up tends to result in softer jet kinematics. Overall, pile-up is included to more closely match the conditions seen in an actual experiment.

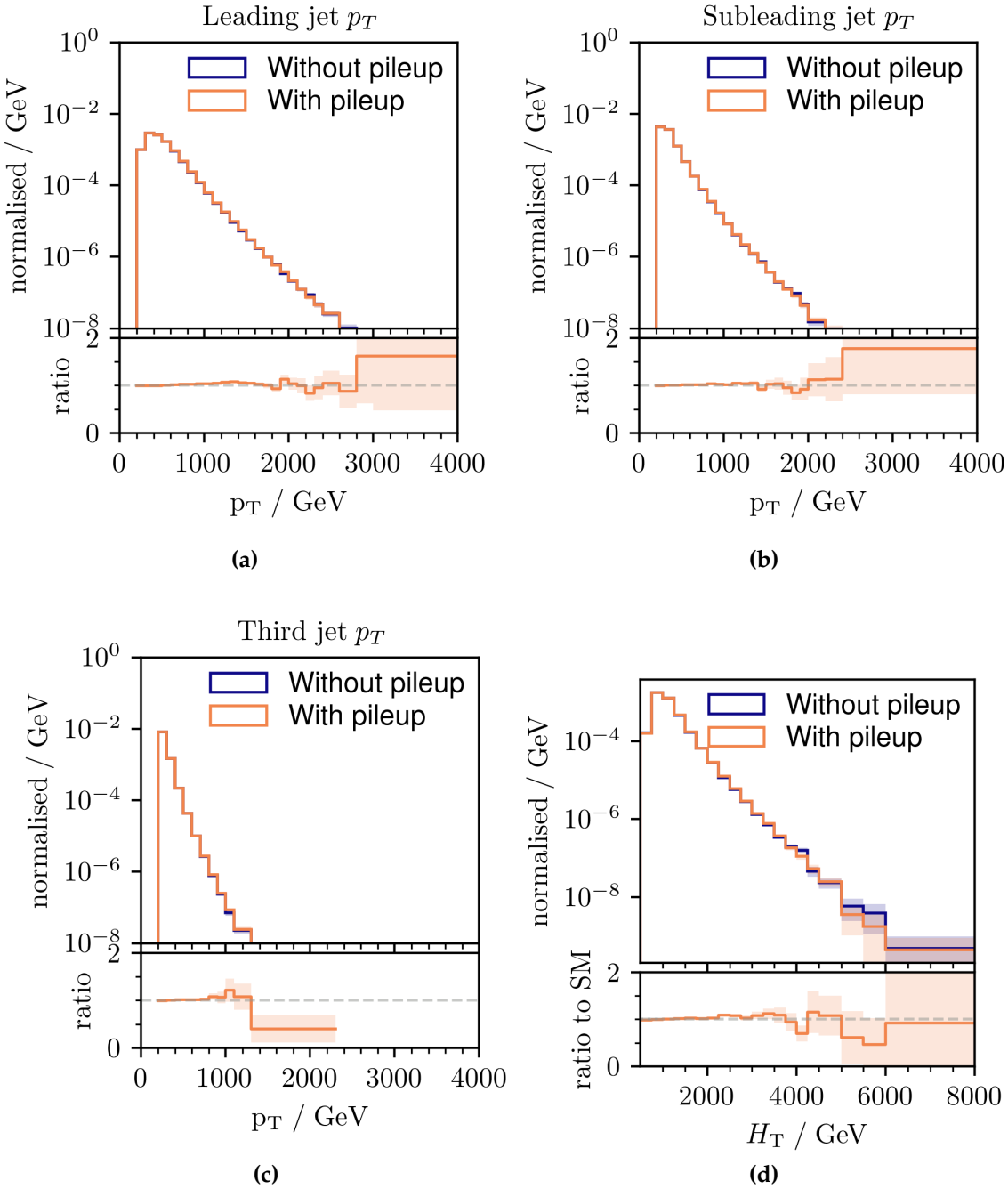


Figure 5.2.: The p_T distributions of the (a) leading (b) subleading (c) third jet and (d) the scalar sum of the leading three reconstructed jet p_T, for Standard Model events with and without pileup simulation.

5.4. Analysis strategy

Parity-violating physics is searched for using a parity-odd event variable $f(x)$ as described in Section 5.2. Such a variable can be simply constructed from an arbitrary function $g(x)$ as

$$f(x) = g(x) - g(\mathrm{P}x). \tag{5.5}$$

CNNs are used as a general way to assign the functional form for $g(x)$. As described in Section 2.2, these take images as input. In this case, images that are derived from either calorimeter energy deposits or jet momenta are used, as described in the following section.

5.4.1. Detector images

The CNN takes images as input, so images in the η-ϕ plane are considered. These images are formed of 32×32 pixels in $\eta \times \phi$, with ϕ covering the whole 2π-range and $|\eta| < 3.2$. This corresponds to the extent of the EM calorimeter and hadronic end cap in ATLAS. Note that the jet images have jets with $|\eta| < 2.8$, due to the jet selections described previously.

Two image formats are considered, either calorimeter energy deposits or reconstructed jet p_{T} (reco-jet p_{T}). Figure 5.3 illustrates images from an example event in each of these formats. Each pixel corresponds to either the energy deposited, or the sum p_{T} of the jets with a η and ϕ within the pixel. Both of these input formats are considered for the CNN, to investigate whether applying jet reconstruction helps the classifier or whether it is better to provide the classifier more information in the energy deposits. In addition, images of the parton p_{T} (truth-jet p_{T}) before any detector reconstruction is applied are considered, in order to assess the effect of the reconstruction on the sensitivity of the method.

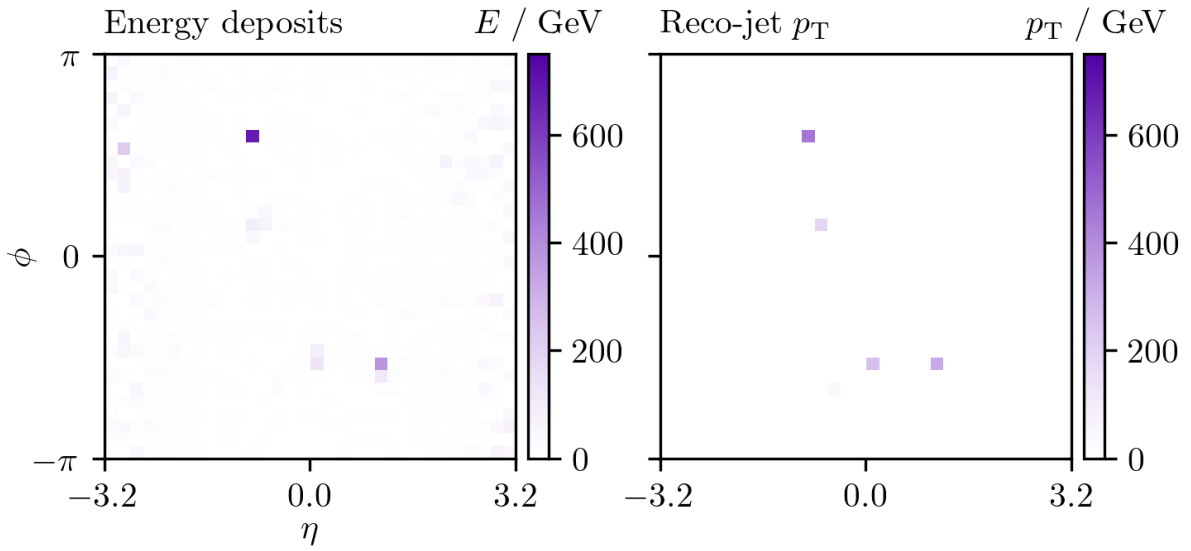


Figure 5.3.: The calorimeter energy deposits and p_{T} of reconstructed jets for an example Standard Model event in the calorimeter for an example event.

5.4.2. Building in symmetries and asymmetries into neural networks

The rotational symmetries of collisions in a cylindrical detector, described in Section 5.2 ,are in-built into the structure of the CNN. These symmetry-transforms can be conveniently visualised in Figure 5.4, which shows them on an example energy deposit that resembles a letter 'R'.

The ϕ-rotation symmetry is inbuilt into the network structure. Firstly, a translationally invariant cyclical padding is applied before each convolution. Then, a max-pooling layer is applied over the entire ϕ-axis. This selects the largest pixel value for each slice in ϕ, ensuring invariance to translations in ϕ.

For the invariance to rotations about the beam axis, R_{180}, we proceed similarly to how parity-oddness is achieved in Equation 5.5. The function $g(x) = h(x) + h(R_{180}x)$ is defined. This is invariant to R_{180}, since $R_{180}R_{180}x = x$. Then, the overall function

$$f(x) = [h(x) + h(R_{180}x)] - [h(\mathrm{P}x) + h(R_{180}\mathrm{P}x)] \tag{5.6}$$

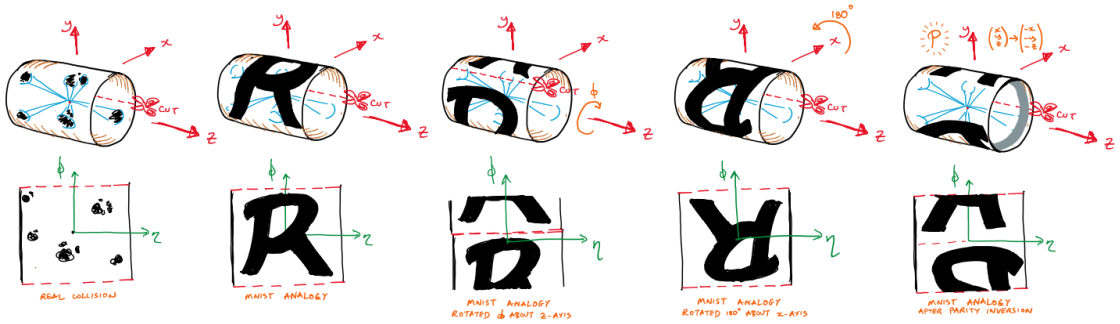


Figure 5.4.: Sketch of a detector, used to illustrate the rotational invariance in ϕ, symmetry to rotations 180-degrees about the x-axis, and parity flipping. The top row corresponds to a 'real' three-dimensional calorimeter energy deposit. The bottom row illustrates in an unrolled two-dimensional (η,ϕ)-space. Figure taken from Ref. [14].

maintains the parity-odd property $f(x) = -f(\mathrm{P}x)$ and introduces the beam-flip symmetry $f(x) = f(R_{180}x)$.

The parity flip operator transforms $\eta \to -\eta$ and $\phi \to \phi + \pi \mod 2\pi$. With a ϕ-rotationally invariant CNN, the parity flip operator is just a flip in η, or simply a mirror image through the line $\eta = 0$ when visualised in the η–ϕ plane.

Figure 5.5 shows the reco-jet p_{T} image of an example event, with the different symmetry transforms applied to it. The original event image is in the upper-left pane, and the transforms are written above each image. The outputs of an untrained CNN are reported in the figure caption; it can be seen that the network is indeed parity-odd whilst being symmetric to beam-flips. In addition, it can be seen that the network is invariant to translations in ϕ by comparing (c) and (d). Furthermore, comparing (a) to (d) verifies that a parity flip on an image in the η–ϕ plane is equivalent to a mirror flip in $\eta = 0$.

5.4.3. Training the CNN

For a typical supervised learning context, we have class labels. So when training a network, we can minimise the loss functions described in Section 2. However, in our

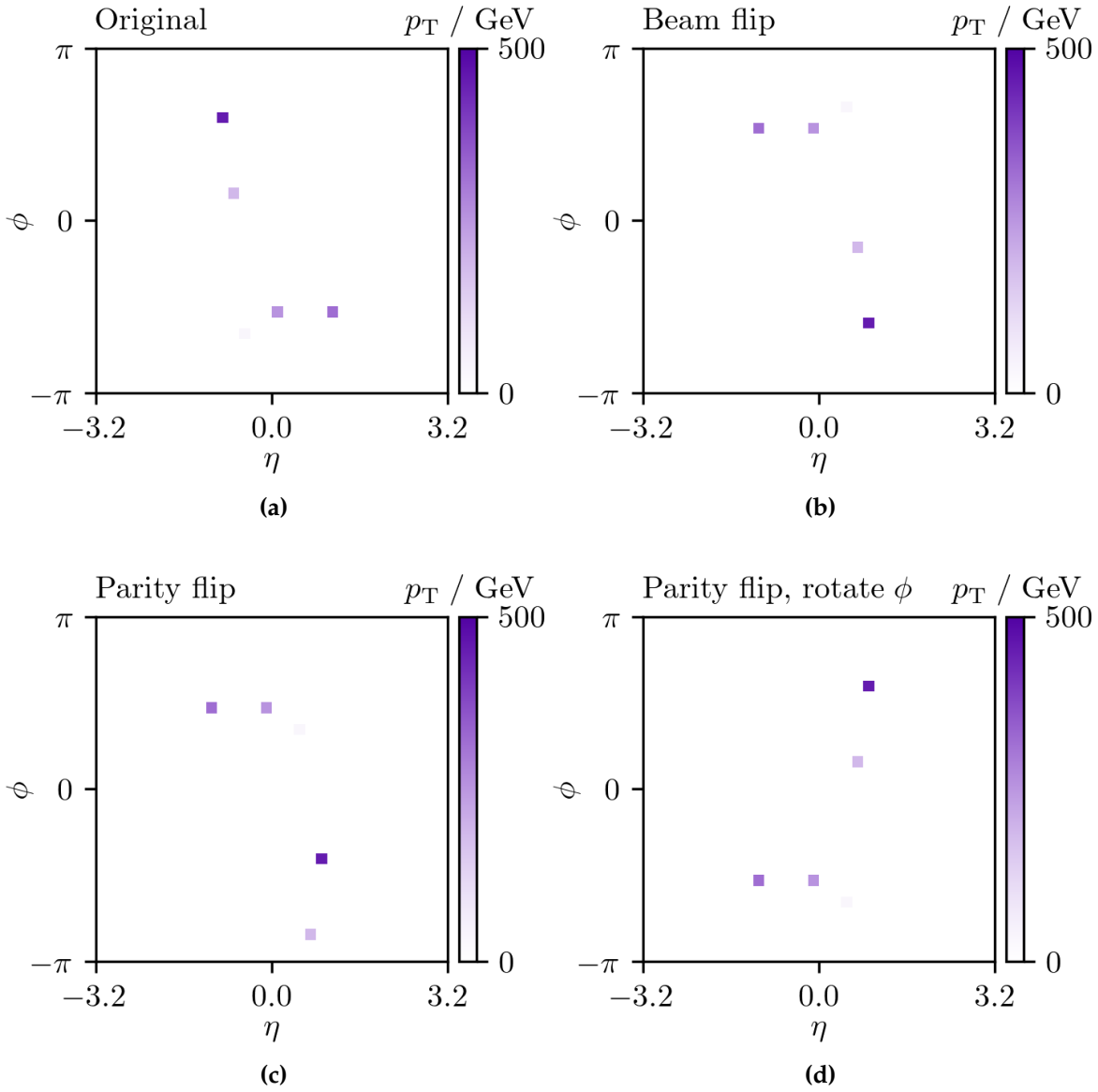


Figure 5.5.: The p_T of jets for an Standard Model event projected in the η–ϕ plane. Different transformations are applied: (a) the original event image x (b) x rotated 180° about the x-axis (c) x parity flipped (d) x parity flipped and translated by $\pi/2$ in ϕ. An untrained CNN output $f(x)$ is -0.132 for (a) and (b), and 0.132 for (c) and (d).

case, we do not have class labels; rather, there is a dataset about which we wish to say whether it is parity-violating or not. Hence, we use a loss function designed for an unlabelled dataset, and can be used to detect violations of a given symmetry [144].

Loss functions

As described in Section 5.2, we have events x from a dataset. The task can be cast as a classification problem of labelling the pair $\{x, \mathrm{P}x\}$ as 'real–fake' or 'fake–real'. Where x is an event, and $\mathrm{P}x$ is its parity flipped counterpart.

For this classification problem, the "Which is real?" loss function

$$L_{\text{Which is real?}} = -\frac{1}{N}\sum_{i=1}^{N} \log p(\text{real–fake} \mid \{x_i, \mathrm{P}x_i\}) \tag{5.7}$$

$$= \frac{1}{N}\sum_{i=1}^{N} \log(1 + e^{-f(x_i)}) \tag{5.8}$$

is used as introduced in Ref. [144]. Notice that the network output $f(x)$ in Equation 5.5 has been converted to a probability using the sigmoid function $\sigma(z) = 1/(1+e^{-z})$.

In order to derive this loss function, we can consider drawing points from a distribution $\mathrm{D}(x)$, and wish to label the pair $\{x, \mathrm{P}x\}$ as 'real–fake' or 'fake–real'. We can calculate the probability of 'real–fake' as

$$p(\text{real–fake}|\{x, \mathrm{P}x\}) = \frac{p(\text{real–fake}|\{x, \mathrm{P}x\})}{p(\text{real–fake}|\{x, \mathrm{P}x\})) + p(\text{fake–real}|\{x, \mathrm{P}x\})} \tag{5.9}$$

$$= \frac{1}{1 + \frac{p(\text{fake–real}|\{x, \mathrm{P}x\})}{p(\text{real–fake}|\{x, \mathrm{P}x\})}}. \tag{5.10}$$

Noting that the joint probability can be expressed as

$$p(\text{real–fake}|\{x, \mathrm{P}x\}) = p(\text{real–fake}, \{x, \mathrm{P}x\})p(\{x, \mathrm{P}x\}), \tag{5.11}$$

we can reexpress the ratio on the denominator as joint probabilities

$$\frac{p(\text{fake–real}|\{x, \text{P}x\})}{p(\text{real–fake}|\{x, \text{P}x\})} = \frac{p(\text{fake–real}, \{x, \text{P}x\})}{p(\text{real–fake}, \{x, \text{P}x\})}. \tag{5.12}$$

The joint probability $p(\text{real–fake}, \{x, \text{P}x\})$ can be written as

$$\begin{aligned} p(\text{real–fake}, \{x, \text{P}x\}) &= p(x \text{ real}, \text{P}x \text{ fake}) && (5.13)\\ &= p(\text{P}x \text{ fake}|x \text{ real})p(x \text{ real}) && (5.14)\\ &= p(x \text{ real}) && (5.15)\\ &= D(x), && (5.16) \end{aligned}$$

since $p(\text{P}x \text{ fake})|x \text{ real})$ is equal to 1 by construction, and $p(x \text{ real})$ is simply the probability that x is drawn from the underlying distribution $D(x)$.

Hence, the ratio in Equation 5.12 can be expressed as

$$\frac{p(\text{fake–real}, \{x, \text{P}x\})}{p(\text{real–fake}, \{x, \text{P}x\})} = \frac{D(\text{P}x)}{D(x)}. \tag{5.17}$$

The logarithm of the ratio 5.12 can be expressed as a difference between two terms,

$$\log \frac{p(\text{fake–real}|\{x, \text{P}x\})}{p(\text{real–fake}|\{x, \text{P}x\})} = \log D(\text{P}x) - \log D(x) = \text{g}(\text{P}x) - \text{g}(x) = -f(x), \tag{5.18}$$

where $f(x)$ is as defined in Equation 5.5, and $g(x)$ is a function that can be learned. Note that learning $g(x)$ is an approximation to the logarithm of the data distribution log D(x). But it will not necessarily be exact since there can be, for example, additive constants in $g(x)$.

Finally, returning to Equation 5.9 we can express the probability as

$$p(\text{real–fake}|\{x, \mathrm{P}x\}) = \frac{1}{1+e^{-f(x)}}. \tag{5.19}$$

Note that the derivation has resulted in $f(x)$ being converted to a probability using the sigmoid function $\sigma(z) = 1/(1+e^{-z})$. We obtain the previously quoted "Which is real?" loss by calculating the average negative log-probability,

$$L_{\text{Which is real?}} = -\frac{1}{N}\sum_{i=1}^{N} \log p(\text{real–fake} \mid \{x_i, \mathrm{P}x_i\}) \tag{5.20}$$

$$= \frac{1}{N}\sum_{i=1}^{N} \log(1+e^{-f(x_i)}), \tag{5.21}$$

such that when minimised we maximise the probability of real–fake labelling. By maximising this probability, we learn a parity-odd event variable $f(x)$ for separating an event from its parity-flipped image.

We are training a classifier to distinguish an event from its parity-flipped image. If the dataset $D(x)$ is parity symmetric, then events x and $\mathrm{P}x$ will be produced at the same rate. Hence, a well-trained classifier will not be able to output a greater probability of real–fake or fake–real labelling. However, if the dataset $D(x)$ violates parity symmetry, then x and $\mathrm{P}x$ can be produced at the different rates. Hence, a trained classifier can ascribe a probability greater than $1/2$ for either real–fake or fake–real labelling.

For illustration, we consider a specific event x_1. If we produce the event x_1 at twice the rate as $\mathrm{P}x_1$, when sampling over the entire dataset, we will be classifying $\{x_1, \mathrm{P}x_1\}$ twice as frequently as $\{\mathrm{P}x_1, x_1\}$. The classification of $\{x_1, \mathrm{P}x_1\}$ will have a probability $> 1/2$, and of $\{\mathrm{P}x_1, x_1\}$ will have a probability $< 1/2$. Since we see twice as many points $\{x_1, \mathrm{P}x_1\}$, the mean probability of real–fake will be greater than $1/2$ over the entire dataset in the presence of this parity violation.

The above discussion was in terms of probabilities. The parity-odd event variable we use is the network output $f(x)$ in Equation 5.5. As previously mentioned, the probability is simply the sigmoid function acting on $f(x)$. Probabilities of $1/2$ correspond to $f(x) = 0$, and a probability greater than $1/2$ is equivalent to $f(x) > 0$. Indeed, this is what we visualise when looking at the final results in Figure 5.6; we assess whether the positive half of $f(x)$ has a greater yield than the negative half of $f(x)$. If this is the case, we have an indication that parity violating physics is occurring.

Metrics for optimisation

For evaluating performance, a model comparison approach is used. Comparing to a symmetric model which would output $p_{\text{sym}} = 1/2$ for each event, the likelihood ratio for an event x_i is

$$\text{LR} = \frac{p(\text{real–fake} \mid \{x_i, \text{P}x_i\})}{p_{\text{sym}}} = \frac{p(\text{real–fake} \mid \{x_i, \text{P}x_i\})}{1/2}. \tag{5.22}$$

From this we can calculate the mean log-likelihood ratio

$$Q = \frac{1}{N} \sum_{i=1}^{N} \log p(\text{real–fake} \mid \{x_i, \text{P}x_i\}) - \log \frac{1}{2}. \tag{5.23}$$

For perfect classification, each event can be unambiguously distinguished from the parity-flipped counterpart. In this case, $p(\text{real–fake} \mid \{x_i, \text{P}x_i\})$ =1, and Q has a maximum of $\log(2) \approx 0.693$. If the asymmetric model does not predict better than a symmetric model would on the dataset, Q takes a value of ≈ 0.

By estimating the mean and standard deviation of Q, a significance metric can be calculated by evaluating how much Q is greater than 0. This metric is used for the analysis to indicate whether we can observe parity-violation.

Training details

The overall dataset of around 10 million samples is split into train, validation and test subsets in the ratio 60, 20, 20. The benchmark model with $\lambda_{\mathrm{PV}} = 1$ was used for an initial tuning and to check for sensitivity. The tuning of the hyperparameters was performed on the validation set, and the test set was not looked at until the models were finalised.

The CNN is implemented using PyTorch [35]. The network used for the final results consists of two convolutional layers with 5×5 kernels each outputting 6 channels. Before the convolutional layers, there is cyclical-padding in ϕ to ensure rotational invariance, and zero-padding in η. The convolutions are followed by two fully connected layers, with leaky ReLU in Equation 2.5 providing the non-linearity between layers. It is trained with Adam [31] using a learning rate of 0.001 and L2 regularisation penalty of 0.1.

The network is trained on batches of 512 images for a maximum of three epochs, or until the score Q on the validation set saturates. The validation score is evaluated every 1000 steps, and the training is terminated if it has not increased for 10 consecutive evaluations.

5.5. Results

Firstly, we can assess the asymmetry in the distribution of the parity-odd CNN output $f(x)$ by comparing the distribution of its positive and negative halves. The output distributions for the benchmark λ_{PV}=1 sample for truth jets, reconstructed jets and calorimeter energy deposits are presented in Figure 5.6. In all distributions, there is clearly an asymmetry, with the positive values of $f(x)$ having a greater yield. This indicates that the parity violation in the dataset is visible with these trained event variables.

The Q values from Equation 5.23 on the test set are 710 ± 30, 860 ± 30, and 2140 ± 60 ($\times 10^{-6}$) for the truth-jets, jet-p_{T} images and energy deposits, respectively. This is significantly above zero, and provides a further indication of the parity violation being visible. These scores are calculated with 2.3 million testing data. With the 60: 20: 20 split of training, validation and testing set, this corresponds to a luminosity of 53 fb^{-1} at the Standard Model cross-section of 0.22 nb.

Figure 5.6d shows the distributions of α from Equation 5.1. There is no separation between the positive and negative halves of this parity-odd event variable, indicating that it is blind to the parity violation in the dataset. This is evident from the ratio plot, which has a much reduced scale as compared to the other plots.

Investigating now smaller λ_{PV} couplings, Figure 5.7 shows the Q values for the truth jet-p_{T}, reconstructed jet-p_{T} and calorimeter energy deposit images. As λ_{PV} is decreased, Q decreases, as is expected.

Interestingly, the energy deposit images have a much better performance than the momenta based images. These have more information than the images just containing jet p_{T} information. For example, the energy deposit images preserve flavour information, for example the size of the energy cluster can be affected by the flavour of the quark that produced it. This cluster would subsequently be reconstructed to a jet, so this flavour information would be lost for the jet p_{T} image. Flavour information is seen to be very important for sensitivity; when using momentum information and flavour information in truth samples, there is a far greater performance than just momentum information (see Appendix A of Ref. [3]). In addition, subtle effects in the showering, such as colour connection between partons could be preserved in the energy deposit images providing further information for the classifier to utilise.

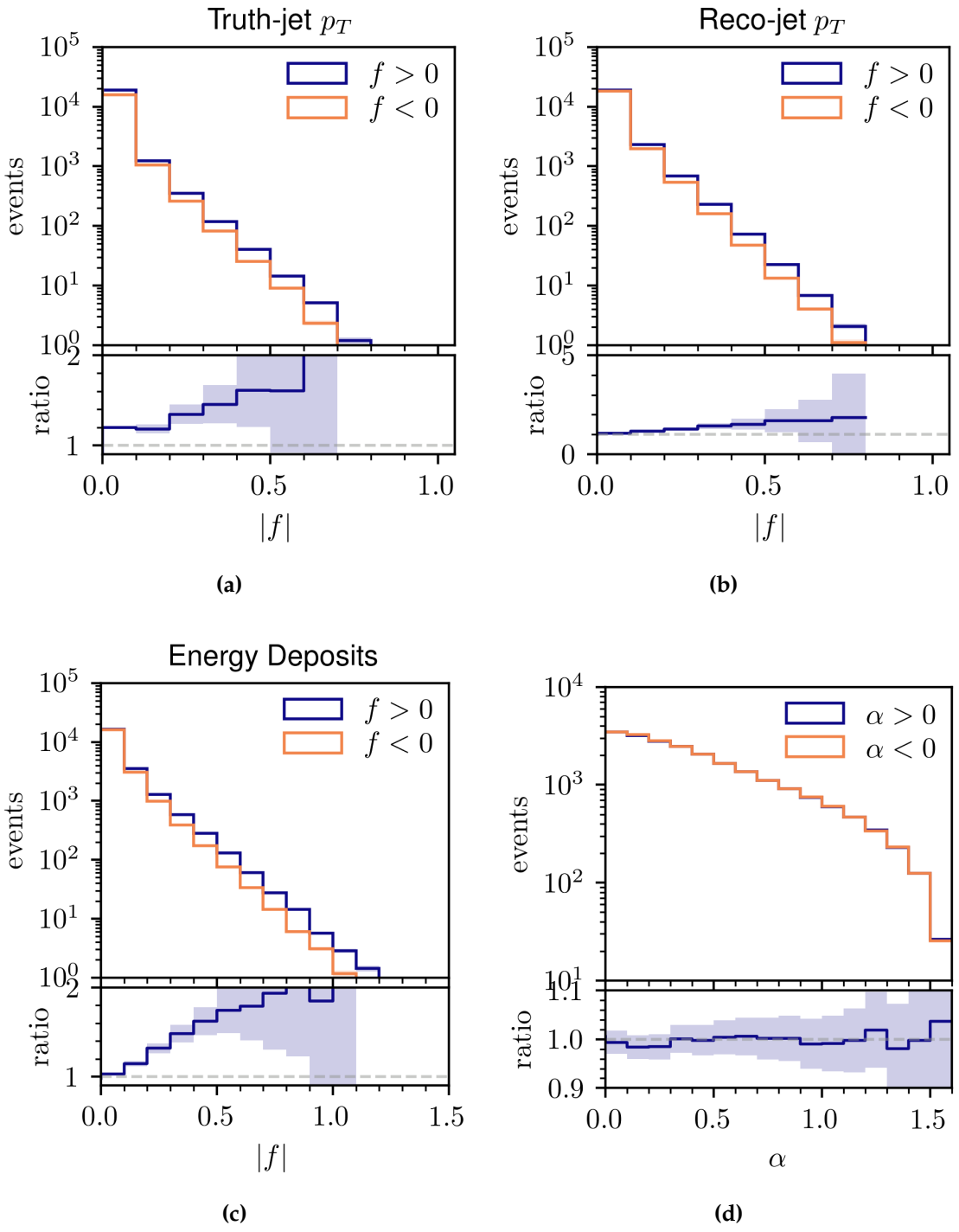


Figure 5.6.: The distribution of the network output on the test set for λ_{PV} =1 for (a) truth-jet p_{T} images (b) reconstructed jet images (c) calorimeter energy deposit images. The distribution of α from Equation 5.1 is shown in (d), with a reduced scale on the ratio plot due to greater similarity between positive and negative halves of the distribution. The positive and negative halves of the distribution are overlaid, with their ratio shown below. The error bars shown are statistical.

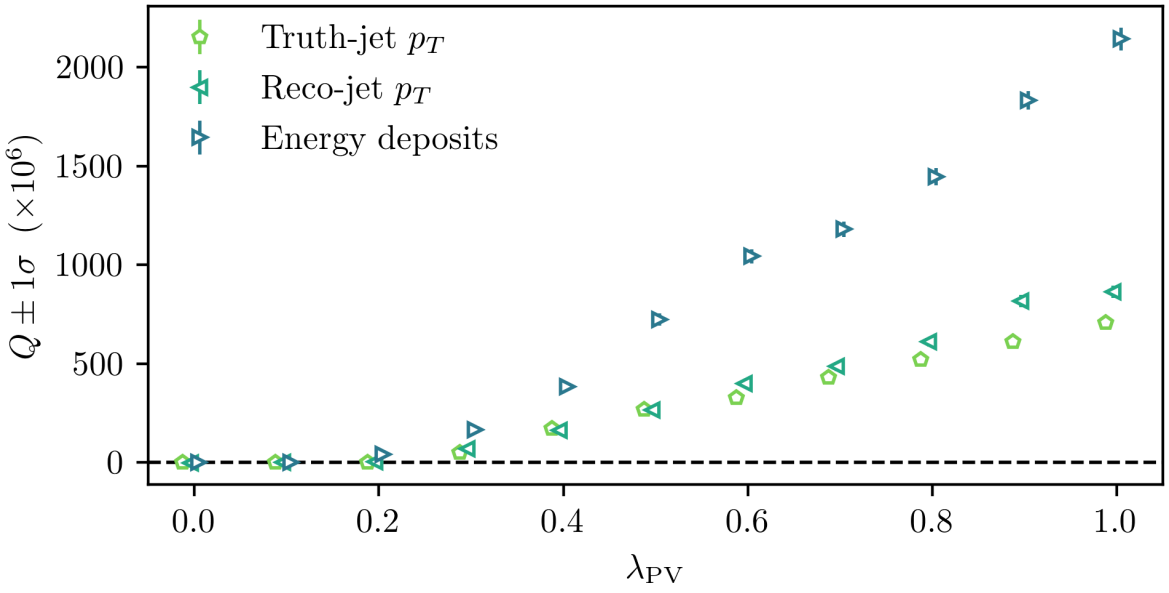


Figure 5.7.: The Q value evaluated on the test set as a function of the coupling λ_{PV}. This is displayed for different input formats; truth jet-p_T, reconstructed jet-p_T and calorimeter energy deposits. Each point uses an individually trained CNN. The markers are slightly offset at each point to aid visibility.

5.6. Discussion and conclusions

Parity-violating signals may already have been produced at the LHC, but simply have not yet been searched for. This chapter details a new model-independent data-driven method for searching for such parity-violating new physics at the LHC. Parity-odd event variables are used, which respect the rotational symmetry of a cylindrical detector at the LHC. These are generated using CNNs trained on images in the η–ϕ plane, and have sensitivity to parity-violating new physics. From Figure 5.6 it can clearly be seen that this method is far more sensitive than the only previous search for this parity-violating new physics [141].

The method is developed for a multijet final state, but this is not a requirement; the method is general and can be applied to other final states with other physics objects. This study is the first investigation into an effective way to perform a general, model-independent parity violation search, which has no reliance on Monte Carlo simulations. The next step would be to apply this method to real data at ATLAS or

CMS, and investigate previously unexplored forms of parity-violating physics at the energy scales of the LHC.

Chapter 6.

Concluding remarks

"Trust me, everything's going to be fine...
You met me at a very strange time in my life."
— The Narrator, Fight Club

This thesis describes searches for new physics using machine learning methods. It focuses on searching for new physics in intricate ways in order to probe particularly challenging new physics scenarios, which previous searches have not had sensitivity to – or in fact physics models that have never previously been searched for.

Chapter 4 presents a search for electroweak supersymmetry using the full Run 2 dataset of 139 fb^{-1}at $\sqrt{s} = 13\,\text{TeV}$. This is the second ATLAS search of the decay of $\tilde{\chi}_1^{\pm}$ decaying to $\tilde{\chi}_1^0$ via leptonically decaying W-bosons using the Run 2 dataset, where this analysis focuses on a very challenging region of parameter space with low mass-splitting between $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_1^0$. This analysis utilises a BDT performing multiclass classification, which is trained to separate the signal and different background classes from each other, in order to gain sensitivity in this region. Moreover, it allows for regions enriched in certain backgrounds to be developed for improving background modelling. No significant excesses were observed, so exclusion limits were set on the masses of the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_1^0$: for a mass-splitting of 100 GeV, $\tilde{\chi}_1^{\pm}$ masses up to 140 GeV

have been excluded at 95% CL. These limits cover an important gap in sensitivity between previous searches.

Chapter 5 presents a novel method for searching for parity-violating new physics, and allows us to test this symmetry at the energy scales of the LHC. This method is model-independent and can be applied directly to data with a variety of final states. Moreover, it allows for an analysis to be performed solely on data, with no reliance on Monte Carlo. This chapter uses convolutional neural networks which are by construction parity-odd in order to perform the search for parity-violating physics. Asymmetries in this parity-odd output indicate that there is parity-violation in the dataset. The efficacy of this method is demonstrated using a simplified model of parity-violating physics – when this method is applied to the simplified model, the parity-violation is visible.

With the lack of new physics discovered at the LHC, more sophisticated analysis techniques need to be developed. Machine learning provides an excellent framework in which to develop these sophisticated techniques, allowing us to probe for new physics in more intricate ways. Exploring further the Standard Model – with such novel analysis techniques, with higher luminosity and at higher energies – allows us to have a more fundamental understanding of the particles and forces that make up the universe.

Appendix A.

Compressed chargino signals

The analysis in Chapter 4 considers the chargino signal, focusing on the signals with Δm = 90 and 100 GeV. Further signal points were generated with lower Δm values, however during the development phase of the analysis no sensitivity could be attained to these points. This appendix summarises the reasons that this is not currently possible with the analysis set-up described in Chapter 4.

Firstly, as evident in Table 4.2, signal points that have lower Δm have a lower yield at preselection level, despite having the same production cross-section for the same mass $\tilde{\chi}_1^{\pm}$. This is because, as Δm is reduced, softer leptons are produced as can be seen in Figure A.1 (a) and (b). This plot is normalised, and, evidently, a large fraction of the events with low Δm are removed by the trigger cuts $p_T^{\ell_1} > 27$ GeV and $p_T^{\ell_2} > 9$ GeV. For example, the number of events at 139 fb^{-1}for m($\tilde{\chi}_1^{\pm}$, $\tilde{\chi}_1^0$) = (125, 85) GeV is 5.3 times lower than the number for the (125, 35) GeV signal point. With the current luminosity, too few events compressed events are produced to be able to perform a successful search.

Secondly, Figure A.1 also shows, when Δm= 80 GeV ($\sim M_W$), the signals are kinematically extremely similar to the VV-background. Even with trained BDTs, it was not possible to distinguish the signal points with Δm= 80 GeV from the

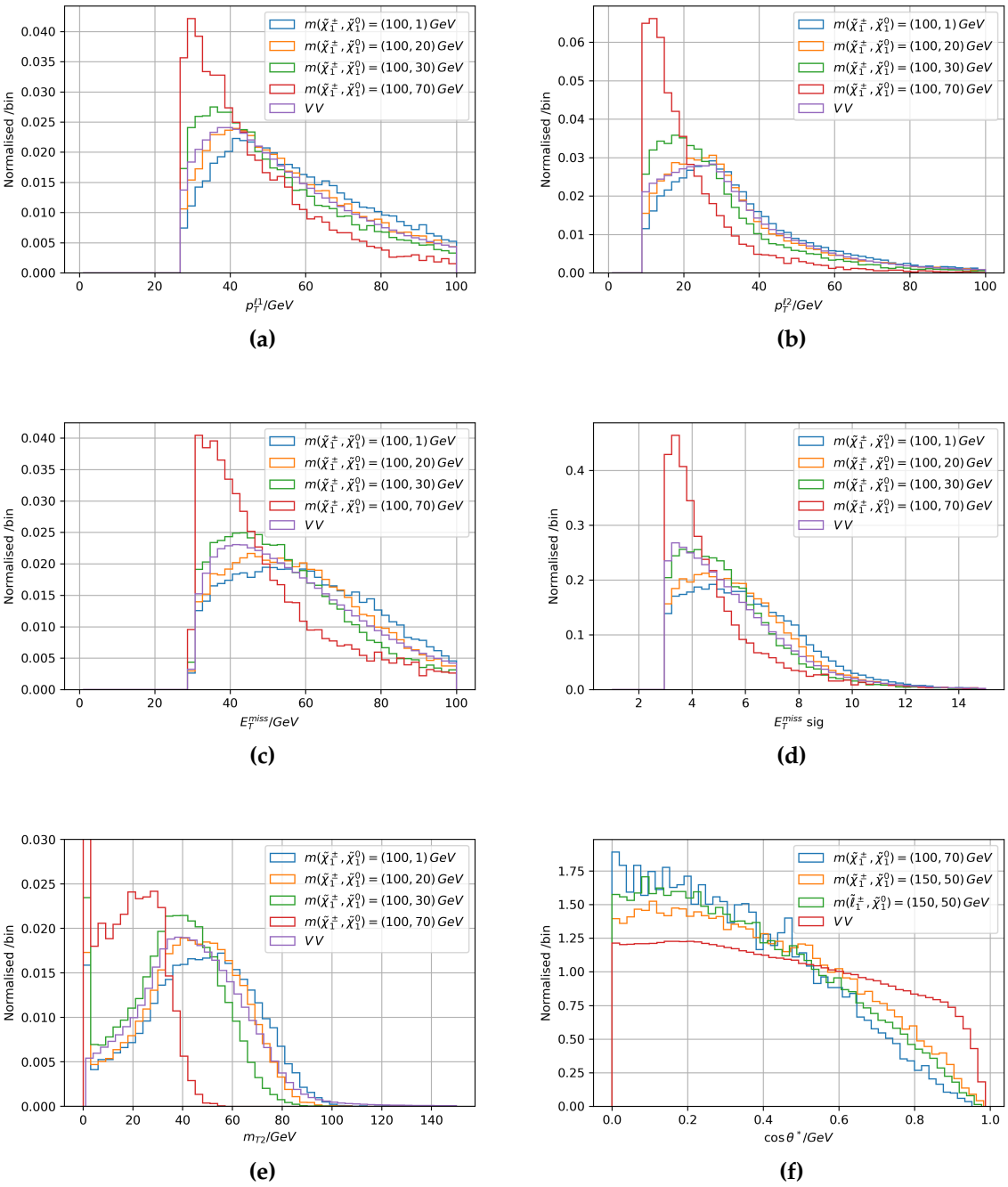


Figure A.1.: The leading and subleading lepton p_{T}, $E_{\mathrm{T}}^{\mathrm{miss}}$, E_T^{miss} significance, m_{T2} and $\cos\theta^*_{\ell\ell}$ distribution for chargino signal points with varying Δm, and the VV background. As Δm decreases, the distributions of p_{T}, $E_{\mathrm{T}}^{\mathrm{miss}}$, E_T^{miss} significance and m_{T2} shift towards lower values. In addition, the signal point m($\tilde{\chi}_1^\pm$,$\tilde{\chi}_1^0$) =(125, 85) GeV is very kinematically similar to the VV background. In Figure (f) the slepton signal from Ref. [1] is included for comparison with the chargino signals.

overwhelming VV background, since these signals are kinematically too similar to the VV background.

Colophon

This thesis was made in LaTeX 2_ε using the "hepthesis" class [149].

Bibliography

[1] ATLAS Collaboration, " Search for direct pair production of sleptons and charginos decaying to two leptons and neutralinos with mass splittings near the W boson mass in $\sqrt{s}$=13 TeV pp collisions with the ATLAS detector." ATLAS-CONF-2022-006, 2022. https://cds.cern.ch/record/2805051/.

[2] ATLAS Collaboration, *Search for direct pair production of sleptons and charginos decaying to two leptons and neutralinos with mass splittings near the W-boson mass in* $\sqrt{s} = 13\,TeV$ *pp collisions with the ATLAS detector*, arXiv:2209.13935 [hep-ex].

[3] C. G. Lester, R. Mastandrea, D. Noel, and R. Tombs, *Hunting for vampires and other unlikely forms of parity violation at the Large Hadron Collider*, JHEP **08** (2022) 231, arXiv:2205.09876 [hep-ph].

[4] ATLAS Collaboration, *Search for supersymmetry in events with four or more charged leptons in* $139\,fb^{-1}$ *of* $\sqrt{s} = 13\,TeV$ *pp collisions with the ATLAS detector*, JHEP **07** (2021) 167, arXiv:2103.11684 [hep-ex].

[5] ATLAS Collaboration, *Standard Model Summary Plots March 2021*,. https://cds.cern.ch/record/2758261.

[6] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv:1207.7214 [hep-ex].

[7] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv:1207.7235 [hep-ex].

[8] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, *Experimental Test of Parity Conservation in Beta Decay*, Phys. Rev. **105** (1957) 1413–1415. https://link.aps.org/doi/10.1103/PhysRev.105.1413.

[9] E. Noether, *Invariante Variationsprobleme*, Nachrichten von der Gesellschaft der

Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse **1918** (1918) 235–257. http://eudml.org/doc/59024.

[10] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Physical Review Letters **13** no. 16, (1964) 508.

[11] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Physical review letters **13** no. 9, (1964) 321.

[12] N. Cabibbo, *Unitary symmetry and leptonic decays*, Physical Review Letters **10** no. 12, (1963) 531.

[13] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652–657.

[14] C. G. Lester, *Chiral Measurements*, 2111.00623. http://arxiv.org/abs/2111.00623.

[15] Kamiokande-II Collaboration, K. S. Hirata et al., *Observation of B-8 Solar Neutrinos in the Kamiokande-II Detector*, Phys. Rev. Lett. **63** (1989) 16.

[16] Super-Kamiokande Collaboration, Y. Fukuda et al., *Measurements of the solar neutrino flux from Super-Kamiokande's first 300 days*, Phys. Rev. Lett. **81** (1998) 1158–1162, arXiv:hep-ex/9805021.

[17] SNO Collaboration, Q. R. Ahmad et al., *Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **89** (2002) 011301, arXiv:nucl-ex/0204008.

[18] B. Abi et al., *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm*, PHYS. REV. LETT. **126** (2021) 141801.

[19] LHCb Collaboration, R. Aaij et al., *Test of lepton universality in beauty-quark decays*, Nature Phys. **18** no. 3, (2022) 277–282, arXiv:2103.11769 [hep-ex].

[20] LHCb Collaboration, R. Aaij et al., *Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays*, JHEP **08** (2017) 055, arXiv:1705.05802 [hep-ex].

[21] LHCb Collaboration, R. Aaij et al., *Test of lepton universality with $\Lambda_b^0 \to pK^-\ell^+\ell^-$ decays*, JHEP **05** (2020) 040, arXiv:1912.08139 [hep-ex].

[22] LHCb Collaboration, R. Aaij et al., *Tests of lepton universality using $B^0 \to K_S^0\ell^+\ell^-$ and $B^+ \to K^{*+}\ell^+\ell^-$ decays*, Phys. Rev. Lett. **128** no. 19, (2022) 191802,

arXiv:2110.09501 [hep-ex].

[23] H. Georgi and S. L. Glashow, *Unity of All Elementary Particle Forces*, Phys. Rev. Lett. **32** (1974) 438–441.

[24] S. P. Martin, *A supersymmetry primer*, arXiv:hep-ph/9709356v7 [hep-ph].

[25] G. Bertone and D. Hooper, *History of dark matter*, Rev. Mod. Phys. **90** no. 4, (2018) 045002, arXiv:1605.04909 [astro-ph.CO].

[26] K. Garrett and G. Duda, *Dark Matter: A Primer*, Adv. Astron. **2011** (2011) 968283, arXiv:1006.2483 [hep-ph].

[27] J. Alwall, P. C. Schuster, and N. Toro, *Simplified models for a first characterization of new physics at the LHC*, Physical Review D **79** no. 7, (2009) 075020.

[28] ATLAS Collaboration, *SUSY March 2022 Summary Plot Update*,. https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-013/.

[29] CMS Collaboration, *CMS supersymmetry physics public results*, Online. https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS.

[30] T. M. Mitchell et al., *Machine learning. 1997*, Burr Ridge, IL: McGraw Hill **45** no. 37, (1997) 870–877.

[31] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, arXiv:1412.6980.

[32] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning representations by back-propagating errors*,.

[33] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, *Object recognition with gradient-based learning*,.

[34] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[35] A. Paszke et al., *PyTorch: An Imperative Style, High-Performance Deep Learning Library*,. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[36] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*,

physics/0703039. http://arxiv.org/abs/physics/0703039.

[37] J. H. Friedman, *Greedy function approximation: A gradient boosting machine.*, The Annals of Statistics **29** no. 5, (2001) 1189 – 1232. https://doi.org/10.1214/aos/1013203451.

[38] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, 1603.02754. http://arxiv.org/abs/1603.02754.

[39] G. Ke et al., *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*,. http://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf.

[40] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer, 2009.

[41] J. Bergstra and Y. Bengio, *Random Search for Hyper-Parameter Optimization*,.

[42] L. S. Shapley, *A value for n-person games*, Contributions to the Theory of Games **2.28** (1953) 307–317.

[43] S. M. Lundberg and S.-I. Lee, *A Unified Approach to Interpreting Model Predictions (SHAP)*,. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

[44] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, *From local explanations to global understanding with explainable AI for trees*,. https://www.nature.com/articles/s42256-019-0138-9. Number: 1 Publisher: Nature Publishing Group.

[45] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[46] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[47] CMS Collaboration, *The CMS experiment at the CERN LHC*,.

[48] LHCb Collaboration, *The LHCb detector at the LHC*, Journal of instrumentation **3** no. 08, (2008) S08005.

[49] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, Journal of

Instrumentation **3** no. 08, (2008) S08002.

[50] ATLAS Collaboration, *ATLAS luminosity public results Run 2*,. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[51] ATLAS Collaboration, *Luminosity determination in pp collisions at* $\sqrt{s} = 13$ *TeV using the ATLAS detector at the LHC*,. https://cds.cern.ch/record/2677054.

[52] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** no. 07, (2018) P07017.

[53] CERN, *The CERN accelerator complex, layout in 2022*,. https://cds.cern.ch/images/CERN-GRAPHICS-2022-001-1.

[54] J. Pequenao, "Computer generated image of the whole ATLAS detector." https://cds.cern.ch/record/1095924. Number: CERN-GE-0803012.

[55] ATLAS Collaboration, *ATLAS Inner Detector: Technical Design Report, Volume 1*, 1997. https://cds.cern.ch/record/331063.

[56] ATLAS Collaboration, *ATLAS Inner Detector: Technical Design Report, Volume 2*, 1997. https://cds.cern.ch/record/331064.

[57] ATLAS Collaboration, *ATLAS Barrel Toroid: Magnet Project Technical Design Report, Volume 2*, 1997. https://cds.cern.ch/record/331065.

[58] ATLAS Collaboration, *ATLAS Pixel Detector: Technical Design Report*, 1998. https://cds.cern.ch/record/381263.

[59] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, 2010. https://cds.cern.ch/record/1291633.

[60] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor tracker*, JINST **9** (2014) P08009, arXiv:1404.7473 [hep-ex].

[61] E. Abat, T. Addy, T. Åkesson, J. Alison, F. Anghinolfi, E. Arik, M. Arik, G. Atoian, B. Auerbach, O. Baker, et al., *The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance*, Journal of Instrumentation **3** no. 02, (2008) P02013.

[62] J. Pequenao, "Computer generated image of the ATLAS inner detector." https://cds.cern.ch/record/1095926. Number: CERN-GE-0803014.

[63] ATLAS Collaboration, *ATLAS Calorimeter Performance: Technical Design Report*, 1996. https://cds.cern.ch/record/331059.

[64] ATLAS Collaboration, *ATLAS Liquid Argon Calorimeter: Technical Design Report*, 1996. https://cds.cern.ch/record/331061.

[65] ATLAS Collaboration, *ATLAS Tile Calorimeter: Technical Design Report*, 1996. https://cds.cern.ch/record/331062.

[66] A. Artamonov, D. Bailey, G. Belanger, M. Cadabeschi, T.-Y. Chen, V. Epshteyn, P. Gorbounov, K. Joo, M. Khakzad, V. Khovanskiy, et al., *The ATLAS forward calorimeter*, Journal of Instrumentation **3** no. 02, (2008) P02010.

[67] J. Pequenao, "Computer generated image of the ATLAS calorimeter." https://cds.cern.ch/record/1095927. Number: CERN-GE-0803015.

[68] ATLAS Collaboration, *ATLAS Muon Spectrometer: Technical Design Report*, 1997. https://cds.cern.ch/record/331068.

[69] J. Pequenao, "Computer generated image of the ATLAS muons subsystem." https://cds.cern.ch/record/1095929. Number: CERN-GE-0803017.

[70] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2010*, Eur. Phys. J. C **72** (2012) 1849, arXiv:1110.1530 [hep-ex].

[71] ATLAS Collaboration, *Operation of the ATLAS trigger system in Run 2*, JINST **15** (2020) P10004, arXiv:2007.12539 [hep-ex].

[72] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. C **77** (2017) 317, arXiv:1611.09661 [hep-ex].

[73] ATLAS Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, Eur. Phys. J. C **80** (2020) 47, arXiv:1909.00761 [hep-ex].

[74] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*, JINST **15** (2020) P09015, arXiv:2004.13447 [hep-ex].

[75] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, JINST **14** (2019) P12006, arXiv:1908.00005 [hep-ex].

[76] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s} = 13\,TeV$, Eur. Phys. J. C **76** (2016) 292,

arXiv:1603.05598 [hep-ex].

[77] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv:1703.10485 [hep-ex].

[78] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv:0802.1189 [hep-ph].

[79] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2020) 689, arXiv:2007.02645 [hep-ex].

[80] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79** (2019) 970, arXiv:1907.05120 [hep-ex].

[81] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **78** (2018) 903, arXiv:1802.08168 [hep-ex].

[82] ATLAS Collaboration, *Search for electroweak production of charginos and sleptons decaying into final states with two leptons and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector*, Eur. Phys. J. C **80** (2020) 123, arXiv:1908.08215 [hep-ex].

[83] CMS Collaboration, *Searches for pair production of charginos and top squarks in final states with two oppositely charged leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV*, JHEP **11** (2018) 079, arXiv:1807.07799 [hep-ex].

[84] S. Carra, T. Lari, D. Noel, and T. Potter, "Beyond the Standard Model Physics at the HL-LHC and HE-LHC, Chargino pair production at HL-LHC." CERN-LPCC2018-05, 2018. arXiv Preprint:1812.07831 [hep-ph].

[85] ATLAS Collaboration, *Search for direct production of charginos, neutralinos and sleptons in final states with two leptons and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **05** (2014) 071, arXiv:1403.5294 [hep-ex].

[86] GEANT4 Collaboration, S. Agostinelli, et al., GEANT4 – *a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[87] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70**

(2010) 823, arXiv:1005.4568 [physics.ins-det].

[88] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820 [hep-ph].

[89] ATLAS collaboration, *Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton-proton collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, ATLAS-CONF-2014-058 (2014). https://cds.cern.ch/record/1951336.

[90] E. Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Phys. **7** no. 3, (2019) 034, arXiv:1905.09127 [hep-ph].

[91] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, arXiv:0808.3674 [hep-ph].

[92] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour dipole factorisation*, JHEP **03** (2008) 038, arXiv:0709.1027 [hep-ph].

[93] S. Höche, F. Krauss, M. Schönherr, and F. Siegert, *A critical appraisal of NLO+PS matching methods*, JHEP **09** (2012) 049, arXiv:1111.1220 [hep-ph].

[94] S. Höche, F. Krauss, M. Schönherr, and F. Siegert, *QCD matrix elements + parton showers. The NLO case*, JHEP **04** (2013) 027, arXiv:1207.5030 [hep-ph].

[95] S. Catani, F. Krauss, B. R. Webber, and R. Kuhn, *QCD Matrix Elements + Parton Showers*, JHEP **11** (2001) 063, arXiv:hep-ph/0109231.

[96] S. Höche, F. Krauss, S. Schumann, and F. Siegert, *QCD matrix elements and truncated showers*, JHEP **05** (2009) 053, arXiv:0903.1219 [hep-ph].

[97] ATLAS Collaboration, "Multi-Boson Simulation for 13 TeV ATLAS Analyses." ATL-PHYS-PUB-2017-005, 2017. https://cds.cern.ch/record/2261933.

[98] NNPDF Collaboration, R. D. Ball et al., *Parton distributions for the LHC run II*, JHEP **04** (2015) 040, arXiv:1410.8849 [hep-ph].

[99] S. Frixione, G. Ridolfi, and P. Nason, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09** (2007) 126, arXiv:0707.3088 [hep-ph].

[100] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv:hep-ph/0409146.

[101] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv:0709.2092 [hep-ph].

[102] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv:1002.2581 [hep-ph].

[103] 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*, Comput. Phys. Commun. **191** (2015) 159, arXiv:1410.3012 [hep-ph].

[104] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair cross-section at hadron colliders*, Comput. Phys. Commun. **185** (2014) 2930, arXiv:1112.5675 [hep-ph].

[105] ATLAS Collaboration, "ATLAS Pythia 8 tunes to 7 TeV data." ATL-PHYS-PUB-2014-021, 2014. https://cds.cern.ch/record/1966419.

[106] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. C **71** (2011) 1547, arXiv:1009.2450 [hep-ph].

[107] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-*, Phys. Rev. D **82** (2010) 054018, arXiv:1005.4451 [hep-ph].

[108] N. Kidonakis, *Top Quark Production*, arXiv:1311.0283 [hep-ph].

[109] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv:1405.0301 [hep-ph].

[110] ATLAS Collaboration, "Modelling of the $t\bar{t}H$ and $t\bar{t}V(V = W, Z)$ processes for $\sqrt{s} = 13$ TeV ATLAS analyses." ATL-PHYS-PUB-2016-005, 2016. https://cds.cern.ch/record/2120826.

[111] H. B. Hartanto, B. Jäger, L. Reina, and D. Wackeroth, *Higgs boson production in association with top quarks in the POWHEG BOX*, Phys. Rev. D **91** no. 9, (2015) 094003, arXiv:1501.04498 [hep-ph].

[112] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, *High precision QCD at*

hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order, Phys. Rev. D **69** (2004) 094008, arXiv:hep-ph/0312266.

[113] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, Journal of High Energy Physics **2014** no. 7, (2014) 1–157, arXiv:1405.0301.

[114] 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*,. https://linkinghub.elsevier.com/retrieve/pii/S0010465515000442.

[115] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, et al., *Parton distributions with LHC data*, Nuclear Physics B **867** no. 2, (2013) 244–289, arXiv:1207.1303.

[116] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at* $\sqrt{s} = 13\,TeV$, Eur. Phys. J. C **79** (2019) 639, arXiv:1902.04655 [hep-ex].

[117] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at* $\sqrt{s} = 13\,TeV$, Eur. Phys. J. C **81** (2021) 578, arXiv:2012.00578 [hep-ex].

[118] ATLAS Collaboration, "Tagging and suppression of pileup jets." ATL-PHYS-PUB-2014-001, 2014. https://cds.cern.ch/record/1643929.

[119] ATLAS collaboration, *Formulae for Estimating Significance*, ATL-PHYS-PUB-2020-025 (2020). https://cds.cern.ch/record/2736148.

[120] ATLAS collaboration, *Object-based missing transverse momentum significance in the ATLAS detector*, ATLAS-CONF-2018-038 (2018). https://cds.cern.ch/record/2630948.

[121] C. G. Lester and D. Summers, *Measuring masses of semi-invisibly decaying particle pairs produced at hadron colliders*, Phys. Lett. B **463** no. 1, (1999) 99–103, arXiv:hep-ph/9906349.

[122] A. Barr, C. Lester, and P. Stephens, *A variable for measuring masses at hadron colliders when missing energy is expected; mT2: the truth behind the glamour*, Journal

of Physics G: Nuclear and Particle Physics **29** no. 10, (2003) 2343, arXiv:hep-ph/0304226.

[123] A. J. Barr, *Measuring slepton spin at the LHC*, Journal of High Energy Physics **2006** no. 02, (2006) 042, arXiv:hep-ph/0511115.

[124] ATLAS Collaboration, *Search for electroweak production of supersymmetric states in scenarios with compressed mass spectra at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Rev. D **97** (2018) 052010, arXiv:1712.08119 [hep-ex].

[125] ATLAS Collaboration, *Searches for electroweak production of supersymmetric particles with compressed mass spectra in* $\sqrt{s} = 13$ *TeV* pp *collisions with the ATLAS detector*, Phys. Rev. D **101** (2020) 052005, arXiv:1911.12606 [hep-ex].

[126] M. Baak, G. J. Besjes, D. Cote, A. Koutsman, J. Lorenz, and D. Short, *HistFitter software framework for statistical data analysis*, Eur. Phys. J. C **75** (2015) 153, arXiv:1410.1280 [hep-ex].

[127] 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 **96** (2017) 072002, arXiv:1703.09665 [hep-ex].

[128] ATLAS Collaboration, *Determination of jet calibration and energy resolution in proton–proton collisions at* $\sqrt{s} = 8$ *TeV using the ATLAS detector*, Eur. Phys. J. C **80** (2020) 1104, arXiv:1910.04482 [hep-ex].

[129] E. Bothmann, M. Schönherr, and S. Schumann, *Reweighting QCD matrix-element and parton-shower calculations*, Eur. Phys. J. C **76** no. 11, (2016) 590, arXiv:1606.08753 [hep-ph].

[130] ATLAS Collaboration, "Studies on top-quark Monte Carlo modelling with Sherpa and MG5_aMC@NLO." ATL-PHYS-PUB-2017-007, 2017. https://cds.cern.ch/record/2261938.

[131] S. Frixione, E. Laenen, P. Motylinski, C. White, and B. R. Webber, *Single-top hadroproduction in association with a W boson*, JHEP **07** (2008) 029, arXiv:0805.3067 [hep-ph].

[132] ATLAS Collaboration, "Studies on top-quark Monte Carlo modelling for Top2016." ATL-PHYS-PUB-2016-020, 2016. https://cds.cern.ch/record/2216168.

[133] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*,. http://arxiv.org/abs/1007.1727.

[134] L. Heinrich, M. Feickert, and G. Stark, *pyhf: v0.6.3*,. https://doi.org/10.5281/zenodo.1169739. https://github.com/scikit-hep/pyhf/releases/tag/v0.6.3.

[135] L. Heinrich, M. Feickert, G. Stark, and K. Cranmer, *pyhf: pure-Python implementation of HistFactory statistical models*, Journal of Open Source Software **6** no. 58, (2021) 2823. https://doi.org/10.21105/joss.02823.

[136] K. Cranmer and A. Held, *Building and steering binned template fits with cabinetry*,. https://www.epj-conferences.org/10.1051/epjconf/202125103067.

[137] D. Noel, "hep-fitting-scripts." https://github.com/Daniel-Noel/hep-fitting-scripts.

[138] J. Neyman and E. S. Pearson, *IX. On the problem of the most efficient tests of statistical hypotheses*, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character **231** no. 694-706, (1933) 289–337.

[139] A. L. Read, *Modified Frequentist Analysis of Search Results (The CLs method)*,. https://cds.cern.ch/record/451614.

[140] LEPSUSYWG, ALEPH, DELPHI, L3, OPAL Collaborations, *Combined LEP Selectron/Smuon/Stau Results, 183-208 GeV*, LEPSUSYWG/01-03.1 (2001). http://lepsusy.web.cern.ch/lepsusy/www/inos_moriond01/charginos_pub.html.

[141] C. G. Lester and M. Schott, *Search for Non-Standard Sources of Parity Violation in Jets at $\sqrt{s}$=8 TeV with CMS Open Data*, 1904.11195. http://arxiv.org/abs/1904.11195.

[142] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[143] C. G. Lester and R. Tombs, *Stressed GANs snag desserts, a.k.a Spotting Symmetry Violation with Symmetric Functions*, 2111.00616. http://arxiv.org/abs/2111.00616.

[144] R. Tombs and C. G. Lester, *A method to challenge symmetries in data with self-supervised learning*, JINST **17** no. 08, (2022) P08024, arXiv:2111.05442

[hep-ph].

[145] D. Colladay and V. A. Kostelecky, *Lorentz violating extension of the standard model*, Phys. Rev. D **58** (1998) 116002, arXiv:hep-ph/9809521.

[146] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, 1307.6346. http://arxiv.org/abs/1307.6346.

[147] S. Ovyn, X. Rouby, and V. Lemaitre, *Delphes, a framework for fast simulation of a generic collider experiment*, 0903.2225. http://arxiv.org/abs/0903.2225.

[148] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896, arXiv:1111.6097 [hep-ph].

[149] A. Buckley, "The hepthesis LaTeX class.".