Searches for New Physics at the Large Hadron Collider

Riley Patrick

Department of Physical Sciences
The University of Adelaide

This dissertation is submitted for the degree of
Doctor of Philosophy

February 2022

For Rose

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Riley Patrick
February 2022

Acknowledgements

Publications by the Author

A. Arhrib, D. Azevedo, R. Benbrik, H. Harouiz, S. Moretti, R. Patrick and R. Santos, *Scattering Interference effects on $H^+ \to t\bar{b}$ Signals in MSSM Benchmark Scenarios*, in *24th International Conference on Computing in High Energy and Nuclear Physics*.

A. Arhrib, D. Azevedo, R. Benbrik, H. Harouiz, S. Moretti, R. Patrick and R. Santos, *Signal versus background interference in $H^+ \to t\bar{b}$ signals for MSSM benchmark scenarios*, JHEP **10**, 209 (2020), 1905.02635.

R. Patrick, A. Scaffidi and P. Sharma, *Top polarisation as a probe of CP-mixing top-Higgs coupling in tjh signals*, Phys. Rev. D **101**(9), 093005 (2020), 1909.12772.

R. Patrick, P. Sharma and A. G. Williams, *Charged Higgs Search in Bosonic Decays Using Jet Substructure at the LHC*, Springer Proc. Phys. **203**, 297–300 (2018).

R. Patrick, P. Sharma and A. G. Williams, *Triple top signal as a probe of charged Higgs in a 2HDM*, Phys. Lett. B **780**, 603–607 (2018), 1710.08086.

M. van Beekveld, S. Caron, L. Hendriks, P. Jackson, A. Leinweber, S. Otten, R. Patrick, R. Ruiz de Austri, M. Santoni and M. White, *Combining outlier analysis algorithms to identify new physics at the LHC*, 2010.07940.

Abstract

The modern particle phenomenologist must be knowledgeable not only in both experimental and theoretical physics, but also in advanced machine learning techniques that have proven successful in recent years. This thesis provides a pedagogical overview of particle physics analyses at the Large Hadron Collider, including both tried-and-true supervised machine learning methods and also cutting edge novel unsupervised machine learning methods that are in development. It also contains three published studies, and another study that is in progress, in which these techniques are applied to improve current understanding of particle physics theory and experimental approaches.

In the first study the classic methodology of a manual cutflow is used to assess the discovery/exclusion potential of a charged Higgs boson in the Two Higgs Doublet Model using the process $pp \to tH^{\pm}$ followed by the $H^{\pm} \to W^{\pm}A$ and $A \to t\bar{t}$. This provides the signal $tt\bar{t}W^{\pm}$ and we study the final states with three leptons and with two leptons of the same charge. It is found that with minimal data in the early runs of the 14 TeV LHC the charged Higgs can be excluded at 95% confidence for masses up to 1 TeV if the mass splitting of the charged Higgs and pseudo-scalar Higgs is within 100 to 300 GeV.

In the second study the possibility of a beyond the Standard Model CP violating top-Higgs coupling is explored using the process $pp \to thj$. The angular variables of the decay products of the top quark are non-trivially effected by the level of CP violation of the top-Higgs coupling and can be used as a powerful probe into this coupling. A boosted decision tree analysis is performed to fully optimize the extraction of the thj signal. It is found that the combined effect of introducing the angular variable to the analysis as well as the usage of the boosted decision tree algorithm leads to a large increase in exclusion potential of CP violation in the top-Higgs coupling compared to previous literature.

In the third study the level of quantum scattering interference between signal and background of the process $pp \to t\bar{b}H^{-}$ followed by $H^{-} \to b\bar{t}$ (and conjugate process) and the irreducible backgrounds with final state $pp \to t\bar{t}b\bar{b}$ is investigated. It is found that in some areas of the parameter space - when

charged Higgs width to mass ratio is large - that this interference, which is traditionally assumed to be negligible in many analyses, is extremely large. In some instances it can be as large as the signal cross section itself.

Finally, in the fourth study we show that a cutting edge technique known as a variational autoencoder, can be used to effectively parametrize composite images of events detected at the planned XENONnT dark matter detector. The variational autoencoder model is trained exclusively on electron recoil background images and builds a profile of the properties of these images. When the model is then exposed to a new dataset which includes a mixture of both electron recoil and simulated dark matter nuclear recoil events the two signals differ at 95% confidence. This acts as a proof of concept that the anomaly detection methodologies rapidly growing in popularity in many areas can find powerful applications in dark matter direct detection.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

The discovery of the Higgs boson in 2012 at the Large Hadron Collider (LHC) [1][2] marked the end of the search for the fundamental particles of the Standard Model and brought to a close a decades-long effort to understand and verify the building blocks and forces of the subatomic world as predicted by the theory. However, it is known that this theory is incomplete as many phenomena remain outside the realm of understanding.

As we have seen an enormous amount of success using the Standard Model it is likely that whatever theory emerges as its successor contains the symmetries and physics of the Standard Model within it, at least in the low-energy limit. As a result, modern theoretical physics largely boils down to the addition of fields and symmetries to the Standard Model Lagrangian in an attempt to describe experimental data. While modern experimental physics attempts to test the theories put forward by the community, it also revolves around the minimization of experimental error and development of new analytical techniques. This thesis will cover aspects of both these areas of study.

An excellent example of the theoretical perspective, and a major component of this thesis, are "supersymmetric models". These are models birthed from the Standard Model by extending the Poincaré algebra to a super-Poincaré algebra which in turn generates a symmetry between bosons and fermions. Supersymmetric models include an extended Higgs sector, dubbed the Two Higgs Doublet Model (2HDM) which lends itself very well to phenemonological study. The 2HDM comes in many forms which will be explored in this thesis. As a whole, supersymmetric models would naturally solve many open problems and lead to a rich phenemonology, which in many cases lies within the reach of current and near-future detectors, including the LHC and its planned upgrades.

On the experimental side of this thesis lies the development and application of machine learning (ML) techniques to study experimental data. ML algorithms

are algorithms which seek to perform tasks without an explicit set of instructions and rely on the learning of patterns in data. These algorithms mostly fall into two broad categories called "supervised" and "unsupervised" learning. Generally, supervised learning involves modelling of data in which the algorithm is trained by minimizing the difference between a true value and a prediction given by the algorithm. In contrast, unsupervised learning does not utilize a true value and is used for clustering and anomaly detection.

Since the explosion of computational power in recent years ML techniques are applied to almost every particle physics analysis in one way or another and development of improved or novel techniques are an enormous area of research. This thesis both employs ML algorithms for signal significance optimization and presents the development of novel techniques for anomaly detection.

This thesis is structured as follows: Chapter 2 will present an overview of the Standard Model and the gauge theories therein, whilst chapter 3 will present three extensions of the Standard Model of increasing complexity - a charge-parity violating top-Higgs coupling, the 2HDM and the minimal supersymmetric model (MSSM). Following this, chapter 4 will present an overview of collider technologies and terminologies and chapter 5 will present an example of an end-to-end pipeline of a collider physics study. Chapters 6, 7 and 8 then present three published studies which explore the three Standard Model extensions presented previously in chapter 3. Chapter 9 gives an overview of modern machine learning techniques for signal agnostic anomaly detection. These techniques are then applied to MSSM benchmarks in chapter 10 which consists of an exploration of an in peer review study. Chapter 11 presents an in-progress study in which anomaly detection techniques are applied to composite detector images of the XENON1T dark matter detector. Finally we conclude in chapter 12.

Chapter 2

The Standard Model

The "Standard Model" (SM) of particle physics is arguably one of humankind's greatest scientific achievements. It explains three of the four fundamental forces of nature - electromagnetism, the weak force and the strong force - and has led to extreme experimental success. It represents a century of development of quantum physics marked by the first developments of a quantum field theory (QFT) of electromagnetism by various physicists such as Werner Heisenberg, Max Born, Pascual Jordan and Paul Dirac in the 1920's. The Abelian electromagnetic quantum field theory was successfully extended to non-Abelian gauge groups by Chen Ning Yang and Robert Mills in the 1950's allowing for the development of a QFT for the strong force and shortly after the unification of the electromagnetic and weak forces by Sheldon Glashow in 1961. The model culminated in 1967 with addition of the Higgs mechanism to the electro-weak model by Steven Weinberg and Abdus Salam.

Fig 2.1 presents the fermionic and bosonic particle content of the SM. The fermions are the fundamental matter of the theory, while the bosons are the force carriers of the fundamental forces of the theory. The electromagnetic force couples all particles which carry non-zero electromagnetic charge and is mediated by the photon. The weak force couples all particles which carry non-zero weak isospin and is mediated by the W^+, W^- and Z bosons. The strong force couples all particles that carry colour charge and is mediated by the gluon. Finally, mass is generated in the SM by the Higgs mechanism and all particles with mass couple to the Higgs boson.

The model contains a total of twelve fermions which comprise all the observed matter in the universe. This is comprised of six quarks which have mass, spin-$\frac{1}{2}$, electromagnetic charge of magnitude $\frac{1}{3}$ or $\frac{2}{3}$, and colour charge, meaning they interact with all three fundamental forces. The model also contains three leptons which all have mass, spin-$\frac{1}{2}$, carry electromagnetic charge of magnitude

Standard Model of Elementary Particles

	three generations of matter (fermions)			interactions / force carriers (bosons)	
	I	II	III		
mass charge spin	≃2.2 MeV/c² ⅔ ½ **u** up	≃1.28 GeV/c² ⅔ ½ **c** charm	≃173.1 GeV/c² ⅔ ½ **t** top	0 0 1 **g** gluon	≃124.97 GeV/c² 0 0 **H** higgs
QUARKS	≃4.7 MeV/c² −⅓ ½ **d** down	≃96 MeV/c² −⅓ ½ **s** strange	≃4.18 GeV/c² −⅓ ½ **b** bottom	0 0 1 **γ** photon	
	≃0.511 MeV/c² −1 ½ **e** electron	≃105.66 MeV/c² −1 ½ **μ** muon	≃1.7768 GeV/c² −1 ½ **τ** tau	≃91.19 GeV/c² 0 1 **Z** Z boson	
LEPTONS	<1.0 eV/c² 0 ½ **νe** electron neutrino	<0.17 MeV/c² 0 ½ **νμ** muon neutrino	<18.2 MeV/c² 0 ½ **ντ** tau neutrino	≃80.39 GeV/c² ±1 1 **W** W boson	

SCALAR BOSONS / GAUGE BOSONS VECTOR BOSONS

Fig. 2.1 A table of the particle content of the Standard Model [3].

1 and no colour charge and thus only interact electromagnetically and with the weak force. Along with each of these three leptons comes an associated neutrino which has mass, though very little, spin-$\frac{1}{2}$ and no electromagnetic charge or colour charge and so interact only with the weak force.

To date all particles in the current form of the SM have been observed at satisfactory statistical significance.

2.1 Mathematical Formulation

The Standard Model is a gauge quantum field theory with underlying internal symmetry under transformations which belong to the Lie group product $SU(3) \times SU(2) \times U(1)$. Its Lagrangian takes the form:

$$\mathcal{L} = -\frac{1}{4} Tr\left\{G_{\mu\nu}G^{\mu\nu}\right\} - \frac{1}{4}Tr\left\{W_{\mu\nu}W^{\mu\nu}\right\} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{2.1}$$

$$+ i\bar{\psi}\gamma^{\mu}D_{\mu}\psi + \{\text{hermitian conjugate}\} \tag{2.2}$$

$$+ \bar{\psi}\hat{Y}\psi\phi + \{\text{hermitian conjugate}\} \tag{2.3}$$

$$+ D_{\mu}\phi D^{\mu}\phi - V(\phi) \tag{2.4}$$

This presentation of the Lagrangian is written with brevity in mind and does not clearly display the immense detail that is contained within many of these terms. In the coming sections we will take a closer look at these terms and how they connect with each of the parts of the SM but for now let us look at it as a broad overview.

Line 2.1 of the above Lagrangian displays the self-coupling of all the gauge bosons of the SM. Each of the terms contains field strength tensors of their respective symmetry group.

Line 2.2 displays the kinetic terms of all the fermions of the model. ψ is a matrix of the fermion fields and D_{μ} is the gauge covariant derivative of each Lie group. These terms include all the interactions of the fermions with the force carriers.

Line 2.3 includes all of the Yukawa terms of the SM; the coupling of the fermions to the Higgs field. ϕ is the Higgs field and $\hat{Y}$ is a matrix of coupling strengths of the fermions to the Higgs field.

Line 2.4 contains all of the kinetic terms for the Higgs and the Higgs self-couplings. Finally, $V(\phi)$ is the Higgs potential, which takes the form $V(\phi) = \mu^2\phi^{\dagger}\phi - \frac{1}{2}\lambda(\phi^{\dagger}\phi)^2$, often called the "Mexican Hat potential".

Let us now take a closer look at each of the theories that make up the SM.

2.1.1 Electroweak Theory

Electroweak theory governs the interaction of particles which carry electromagnetic charge and weak isospin. It is a gauge field theory of the group product $SU(2)_L \times U(1)_Y$ Lie group. Here the subscript L indicates this symmetry is held for the left handed chirality particles of the model and the subscript Y indicates the symmetry is held for particles carrying non-zero hyper-charge. At an energy scale of approximately 220 GeV, known as the electroweak symmetry

breaking scale, the theory separates into two separate forces via the Higgs mechanism - the electromagnetic force and the weak force - and has symmetry group $U(1)_Q$. Here the subscript Q indicates the symmetry is held for particles carrying non-zero electric charge.

The electromagnetic force and the weak force are best understood via a full review of the electroweak model and the Higgs mechanism which will be covered in the coming sections.

Higgs Mechanism

The Higgs mechanism is the result of the spontaneously broken symmetry of the electroweak sector and the imposition of local gauge invariance.

As we wish to utilize perturbation theory when doing quantum field theoretic calculations of the electroweak sector we require a unique vacuum to perturb around. However a general potential such as the Higgs potential does not necessarily have to have a unique vacuum, in fact as we will soon see the famous quartic Higgs potential has infinite vacuum states. As a result we are free to choose one of these vacuum states and there is no inherent reason that the $SU(2)_L \times U(1)_Y$ symmetry of the electroweak sector must be preserved at this point of the potential. When the symmetry is broken by this choice of vacuum it leads to the appearance of Goldstone boson terms in the Lagrangian [4, 5].

However, remarkably by imposition of local gauge invariance these Goldstone boson terms of the Lagrangian are removed and terms which generate non-zero mass for other particles appear. This is the essence of the Higgs mechanism, however an illustrative model with a more basic Lagrangian is helpful and we will review this before moving onto the SM implementation.

An illustrative model

As a non-physical but informative exercise we will follow the lead of Ref. [6] and explore a toy model. Imagine the Higgs mechanism occurring within a spinless, charged field. We use two complex scalar fields ϕ and ϕ^* with Lagrangian:

$$\mathcal{L} = \frac{1}{2}|D^\mu \phi|^2 - V(\phi^*\phi) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{2.5}$$

where we define the potential as such:

$$V(\phi^*\phi) = -\frac{1}{2}\mu^2 \phi^*\phi + \frac{1}{4}\lambda^2 (\phi^*\phi)^2 \tag{2.6}$$

and,

- μ and λ are real parameters,

- $\phi = \phi_1 + i\phi_2$,

- $D_\mu = \partial_\mu + iqA_\mu$ with q a real scalar,

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

and the Lagrangian is invariant under local phase transformations such as $\phi \to e^{-iq\alpha(x)}\phi$. Now clearly, if the first coefficient of the potential is less than zero then the potential would have a unique minimum (vacuum) at $\phi = 0$, and this minimum has the same symmetry as the Lagrangian. However, if this coefficient is greater than zero the potential has an infinite number of minima lying on a circle of radius $\frac{\mu}{\lambda}$, but none of them have the same symmetry as the Lagrangian (the local phase rotation breaks down). Thus, regardless of our choice of minimum our symmetry is spontaneously broken. Our choice of minimum is mathematically arbitrary, but sensible choices do exist for the sake of simplicity, one such is the purely real vacuum with:

$$\langle\phi_1\rangle = \frac{\mu}{\lambda} \equiv \nu \qquad (2.7)$$

$$\langle\phi_2\rangle = 0 \qquad (2.8)$$

We say the field ϕ (and hence ϕ^* as we chose a real value) acquires the vacuum expectation value (VEV): $\langle\phi\rangle = \langle\phi_1\rangle^2 + \langle\phi_2\rangle^2 = \nu^2$.

As is often done, it is useful to redefine our field in terms of the deviation from the vacuum:

$$\eta \equiv \phi_1 - \nu \qquad (2.9)$$

$$\zeta \equiv \phi_2 \qquad (2.10)$$

We can now expand our Lagrangian in terms of these fields:

$$
\begin{aligned}
\mathcal{L} = \ & \frac{1}{2}\left|(\partial_\mu + iqA)(\eta + \nu + i\zeta)\right|^2 \\
& - \frac{1}{2}\mu^2(\eta + \nu - i\zeta)(\eta + \nu + i\zeta) + \frac{1}{4}\lambda^2\left[(\eta + \nu - i\zeta)(\eta + \nu + i\zeta)\right]^2 \\
& - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\
= \ & \frac{1}{2}\left[(\partial_\mu\eta)^2 - \mu^2\eta^2\right] + \frac{1}{2}(\partial_\mu\zeta)^2 + \frac{1}{2}\left[q^2\nu^2A^2 - F_{\mu\nu}F^{\mu\nu}\right] && (2.11) \\
& - 2iq\nu(\partial_\mu\zeta)A^\mu + \Bigg\{ q\left[\eta(\partial_\mu\zeta) - \zeta(\partial_\mu\eta)\right]A^\mu + \nu q^2(\eta A^2) && (2.12) \\
& + \frac{q^2}{2}(\zeta^2 + \eta^2)A^2 - \lambda\mu(\eta^3 + \eta\zeta^2) - \frac{\lambda^2}{4}(\eta^4 + 2\eta^2\zeta^2 + \zeta^4) \Bigg\} + \frac{\mu^2\nu^2}{4} \\
& && (2.13)
\end{aligned}
$$

Now we can clearly see from the first term in line 2.11 we have a massive particle η of mass $\sqrt{2}\mu$. We also interpret the second term as a massless particle ζ (the Goldstone boson) and the third term as a massive gauge field, with mass $q\nu > 0$, A_μ (a massive photon). The remainder of the terms in lines 2.12 and 2.13, aside from the term involving $(\partial_\mu\zeta)A^\mu$, are various currents and couplings between fields. $(\partial_\mu\zeta)A^\mu$ in particular is troublesome as it does not have a physical interpretation in the Feynman formalism. We can remove this term, and the massless Goldstone boson, by exploiting the local gauge invariance. That is:

$$
\begin{aligned}
\phi \to \phi' &= e^{-iq\alpha(x)}\phi \\
&= \left[\phi_1\cos(-q\alpha(x)) - \phi_2\sin(-q\alpha(x))\right] + i[\phi_1\sin(-q\alpha(x)) + \phi_2\cos(-q\alpha(x))] \\
& \hspace{10cm} (2.14)
\end{aligned}
$$

and the gauge field transforms as usual $A_\mu \to A'_\mu = A_\mu + \partial_\mu\alpha(x)$.

Now importantly, note that we can choose $\alpha(x)$ to be any arbitrary function of spacetime to fix the gauge, so we may choose $\alpha(x) = \frac{1}{q}\arctan\left(\frac{\phi_2(x)}{\phi_1(x)}\right)$ and this will enforce our new field ϕ' to be real. In performing this transformation we of course have not changed our physics, and we get the Lagrangian [6]:

$$
\begin{aligned}
\mathcal{L} = \ & \left[\frac{1}{2}(\partial_\mu\eta)^2 - \mu^2\eta^2\right] + \left[\frac{q^2\nu^2}{2}A^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}\right] \\
& + \left\{ q^2\nu(\eta A^2) + \frac{q^2\eta^2}{2}A^2 - \lambda\mu\eta^3 - \frac{\lambda^2\eta^4}{4}\right\} + \frac{\mu^2\nu^2}{4} && (2.15)
\end{aligned}
$$

Through this imposition of local gauge invariance, our massless Goldstone boson field, ζ, has disappeared. We are left with a scalar field η of mass $\sqrt{2}\mu$ and a massive gauge field A_μ.

Higgs Mechanism of the Standard Model

The Standard Model expands the above ideas with a more complex symmetry. The local gauge symmetry is now $SU(2)_L \times U(1)_Y$, the direct product of the group of 2×2 unitary matrices with determinant equal to one, and the group of local phase rotations. The symmetry now breaks to $U(1)_Q$. The subscript L indicates the transformation is only acting on left handed doublets of iso-spin and the subscript Y is hypercharge ($Y = 2Q - 2I_3$ where I_3 is weak isospin and Q is electric charge) which acts as the generator of U(1). The subscript Q denotes the fact that after spontaneous symmetry breaking the generator corresponding to electric charge does not break. A summary of this process is presented in Fig 2.2.
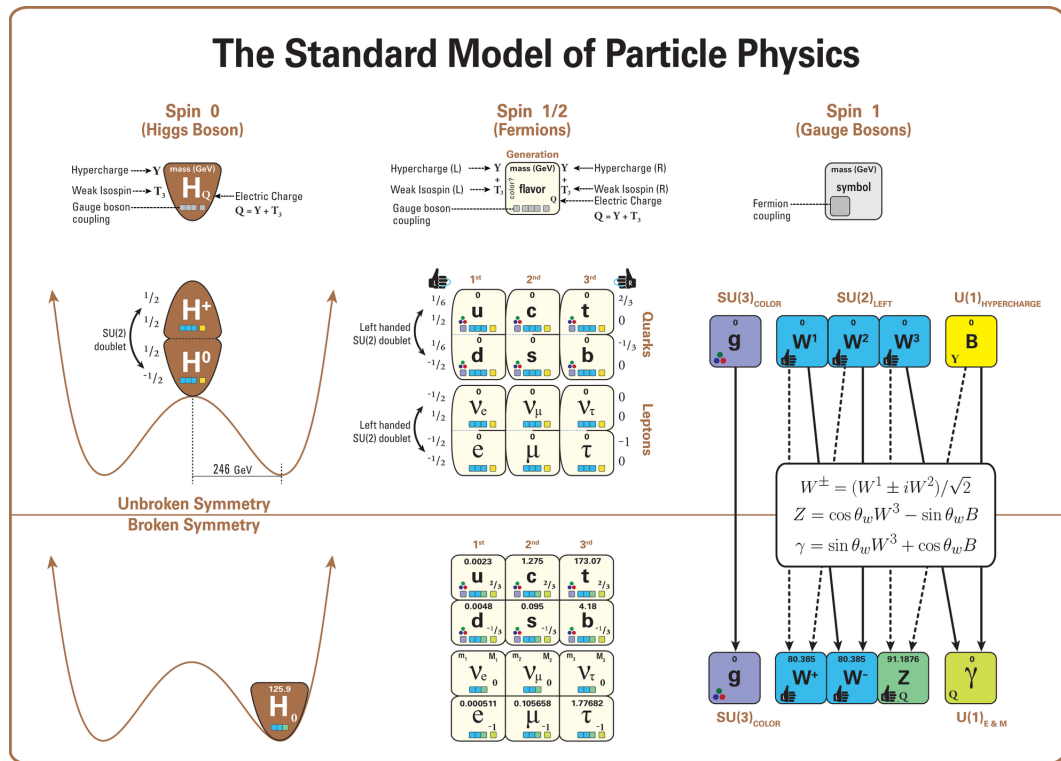
Fig. 2.2 The Standard Model before (above) and after (below) electroweak symmetry breaking [7].

We start with a scalar doublet,

$$\mathbf{\Phi} = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{2.16}$$

which transforms through elements of $SU(2)_L$. As described above the subscript L means the transformation is only acting on left-handed, negative chirality, fermions which implies the weak hypercharge $Y = 1$.

So as with our illustrative model, the doublet must acquire a VEV that is not zero, and we choose:

$$\langle \Phi \rangle = \begin{pmatrix} 0 \\ \frac{\nu}{\sqrt{2}} \end{pmatrix} \tag{2.17}$$

The generators of $SU(2)_L$ are the Pauli matrices $\sigma_i/2$, and the generator of $U(1)_Y$ is Y. Thus these generators all inherit the Lie algebra of the Pauli matrices with an extension:

$$[\sigma_i, \sigma_j] = i\epsilon_{ijk}\sigma_k \tag{2.18}$$

$$[\sigma_i, Y] = 0 \tag{2.19}$$

Now through elementary calculation we can see that every one of these generators has non-zero eigenvalues:

$$\frac{1}{2}\sigma_1\langle \Phi \rangle = \frac{1}{2}\begin{pmatrix} \frac{\nu}{\sqrt{2}} \\ 0 \end{pmatrix} \qquad \frac{1}{2}\sigma_3\langle \Phi \rangle = \frac{1}{2}\begin{pmatrix} 0 \\ -\frac{\nu}{\sqrt{2}} \end{pmatrix}$$
$$\frac{1}{2}\sigma_2\langle \Phi \rangle = \frac{1}{2}\begin{pmatrix} \frac{-i\nu}{\sqrt{2}} \\ 0 \end{pmatrix} \qquad Y\langle \Phi \rangle = \begin{pmatrix} 0 \\ \frac{\nu}{\sqrt{2}} \end{pmatrix} \tag{2.20}$$

We can relate the electric charge to weak isospin and weak hypercharge through the Gellman-Nishijima formula:

$$Q = I_3 + \frac{1}{2}Y \tag{2.21}$$

where we note the third component of isospin is generated by $\sigma_3/2$ and then see that Q as a generator has a zero eigenvalue - it is unbroken:

$$Q\langle\Phi\rangle = I_3\langle\Phi\rangle + \frac{1}{2}Y\langle\Phi\rangle \tag{2.22}$$

$$= \frac{1}{2}\phi_1\langle\Phi\rangle + \frac{1}{2}Y\langle\Phi\rangle \tag{2.23}$$

$$= \frac{1}{2}\begin{pmatrix} 0 \\ -\frac{\nu}{\sqrt{2}} \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 \\ \frac{\nu}{\sqrt{2}} \end{pmatrix} \tag{2.24}$$

$$= \mathbf{0} \tag{2.25}$$

This is the mechanism by which the local symmetry $SU(2)_L \times U(1)_Y$ breaks to $U(1)_Q$. Goldstone's theorem [4, 5] tells us that the number of Goldstone bosons generated is equal to the number of broken generators, which then in turn equals the number of massive gauge bosons - thus we have 3 unbroken generators corresponding after mixing to massive gauge bosons W^+, W^- and Z, and one unbroken generator corresponding to the massless photon.

Kinetic Term

The Kinetic term of the Higgs Lagrangian takes the form:

$$\mathcal{L}_{\text{kin}} = (D_\mu\Psi)^\dagger(D^\mu\Psi) \tag{2.26}$$

with:

$$D_\mu \equiv \partial_\mu - \frac{ig'}{2}YB_\mu - ig\frac{\sigma_i}{2}W_\mu^i \tag{2.27}$$

where W_μ^i (i=1,2,3) are the four-vector fields (gauge eigenstates), associated to the three generators σ_i described earlier. The other term B_μ is the four-vector field associated with the generator Y. The coefficients g and g' are the coupling constants of W_μ^i and B_μ respectively. We can re-write this in a more illuminating form as:

$$D_\mu = \partial_\mu - i\frac{g}{\sqrt{2}}\left(W_\mu^+T^+ + W_\mu^-T^-\right) - i\frac{g}{\cos\theta_w}Z_\mu\left(T^3 - \sin^2\theta_w Q\right) - ieQA_\mu \tag{2.28}$$

where, $T^{1,2,3}$ are the weak isospin components, $T^{\pm} = \frac{1}{2}(T^1 \pm iT^2)$ and most importantly:

$$W_{\mu}^{\pm} = \frac{1}{\sqrt{2}}\left[\left(W_{\mu}^1\right) \mp i\left(W_{\mu}^2\right)\right] \qquad Z_{\mu} = \frac{1}{\sqrt{g^2 + g'^2}}\left[gW_{\mu}^3 - g'B_{\mu}\right] \qquad (2.29)$$

$$A_{\mu} = \frac{1}{\sqrt{g^2 + g'^2}}\left[g'W_{\mu}^3 + gB_{\mu}\right] \qquad (2.30)$$

These three terms, $W_{\mu}^{\pm}$, Z_{μ} and A_{μ}, correspond to the W and Z bosons and the photon.

The masses of the electroweak bosons are given by:

$$M_W = \frac{g\nu}{\sqrt{2}} \qquad (2.31)$$

$$M_Z = \frac{g\nu}{\sqrt{2}\cos\theta_W} = \frac{M_W}{\cos\theta_W} \qquad (2.32)$$

$$M_A = 0 \qquad (2.33)$$

Potential Term

The Higgs potential of the standard model is given by:

$$V(\Phi^{\dagger}\Phi) = -\mu^2(\Phi^{\dagger}\Phi) + \lambda(\Phi^{\dagger}\Phi)^2 \qquad (2.34)$$

where μ^2 and λ are free parameters of the theory. Now we can constrain these parameters by ensuring the potential is bounded below, giving $\lambda > 0$, and then ensuring a non-unique minimum (for the sake of spontaneous symmetry breaking) $\mu^2 > 0$. We impose the same VEV's and symmetry breaking prescription seen above. This term of the Lagrangian gives the self-interaction of the Higgs.

Yukawa Lagrangian

To complete our Higgs sector we must introduce the part of the Lagrangian that gives interactions between the Higgs fields and fermions:

$$\mathcal{L}_Y = \eta_{ij}^U \bar{\Psi}_L \tilde{\Phi} U_R + \eta_{ij}^D \bar{\Psi}_L \Phi D_R + h.c. \qquad (2.35)$$

where $\tilde{\Phi} = i\sigma_2\Phi$, U_R and D_R are the right-handed fermion singlets from the up and down sectors (respectively) and $\bar{\Psi}_L$ are the left-handed fermion doublets. $\eta_{ij}^{U,D}$ are the free parameters that define the interaction vertices, where i, j are

the family indices. The Yukawa Lagrangian is the source of fermion masses when the Higgs doublet acquires its VEV.

2.1.2 Quantum Chromodynamics

Quantum Chromodynamics is the theory that describes the interaction of particles that carry 'colour' charge - quarks and gluons. In analogy to electromagnetic charge, which carries binary value of plus or minus, colour charge takes three values - denoted red, blue, green. As such mathematically it is a gauge theory of the $SU(3)$ Lie group of 3×3 unitary matrices of determinant one.

The quark fields, $[\psi(x)]^i$, are written as four-component spinor fields in the fundamental representation of $SU(3)$. The gluon fields, $[G_\mu(x)]^a$, are associated with the eight generators of the adjoint representation of $SU(3)$.

The QCD Lagrangian takes the form:

$$\mathcal{L}_{\text{QCD}} = \sum_q \bar{\psi}_q^i \left(i\gamma^\mu D_\mu^{ij} - \delta^{ij} m_q \right) \psi_q^j - \frac{1}{4} F_{\mu\nu}^a F^{a\mu\nu} \qquad (2.36)$$

where:

- $i, j = 1, 2, 3$ are the colour indices of the fundamental representation,

- $a = 1, 2, \ldots, 8$ are the colour indices of the adjoint representation,

- q are the flavour indices - up, down, charm, strange, bottom, top,

- The gauge covariant derivative takes the form $D_\mu^{ij} = \delta^{ij}\partial_\mu - ig_a^{ij}G_\mu^a$,

- The gluon field strength tensor takes the form $F_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + gf_{abc}G_\mu^b G_\nu^c$ with f_{abc} the structure constants of the $SU(3)$ Lie group,

- g is the QCD coupling constant.

Included in the above Lagrangian are the explicit mass terms for the fermions and pure gauge interactions. It is however missing a mass term for the gluon as it is in fact impossible to include one that is gauge invariant - thus the gluon is a massless boson. The Lagrangian is also invariant under the usual transformations - charge conjugation, parity, and time reversal. However it is not invariant under the combined charge-parity transformation.

There are two particularly unique properties of QCD that are worth discussing - asymptotic freedom and colour confinement.

Asymptotic Freedom

In analogy with electromagnetism, where isolated electric charges cause the generation of electron-positron pairs in their vicinity which leads to the apparent reduction of the effective electromagnetic coupling, isolated colour charges also leads to the generation of virtual coloured quark-anti quark pairs. However, in QCD the gluon carries colour charge (unlike the photon and electromagnetic charge) which allows an isolated coloured charge to generate gluons in its vicinity which leads to an apparent increase in effective strong coupling.

The QCD coupling constant is said to be a "running coupling" as it's apparent magnitude changes as a function of the energy scale being probed. The QCD β function encodes the details of how the QCD coupling strength runs and is defined as

$$\beta(g) = -\beta_0 \frac{g^3}{16\pi^2} \tag{2.37}$$

at leading order, where $\beta_0 = 11 - \frac{2}{3}N_f$ and N_f is the number of quark flavours. As long as $\beta(g) < 0$, i.e. $N_f < 17$ which is true in the SM, the effect of this is that the strong coupling asymptotes to zero at short distances or high energies.

Colour Confinement

The phenomenon of colour confinement is an open problem of mathematical physics. It states that in nature bare colour states can never be observed, only colour neutral states such as mesons and baryons can be. To date no individual quark or gluon has ever been detected, and yet no mathematical proof of colour confinement has ever been successfully developed.

Intuitively the concept of colour confinement can be understood that as the distance between a quark anti-quark pair increases, it becomes energetically more favourable for a quark anti-quark pair to be produced out of the vacuum between the original pair. Once this happens the result is two colour neutral objects.

Colour confinement leads to an extremely important experimental difficulty as it is the reason that coloured particles form objects known as "jets" in particle detectors. This will be discussed at length in chapter 4.

2.2 Short-comings and Failings of the Standard Model

The Standard Model has demonstrated absolutely stunning experimental success. Every predicted particle has been discovered and almost all measured model parameters do not break internal consistency of the model. However, there are still outstanding issues with the SM that broadly fall into two categories.

Short-comings

The first category contains those which can be considered short-comings, so to speak, where the model provides no explanation or an inadequate explanation of phenomena. Examples of this may be that the model does not explain dark matter, the mysterious gravitational source that permeates the universe, or that the model does not incorporate the fourth fundamental force gravity in any way. These problems are only really problems if one wishes the Standard Model to be an all encompassing theory of everything, but are not really problems in the sense that the model is wrong.

Other examples of short-comings may be areas where the SM appears to have *unnatural* or *finely-tuned* parameter values. The concept of naturalness and fine-tuning in the SM is a popular issue, with many physicists believing that the dimensionless parameters of the SM (or any model) should be of the same order. But in some places in the SM, such as the hierarchy problem, the parameters take wildly different values. Some physicists disagree that a lack of naturalness is an inherent problem such as in Ref [8]. The hierarchy problem will be discussed in detail in section 3.3.1.

Failings

The second category is far more concerning, but fortunately sparsely populated. These issues are those that are actually not self-consistent with the principles of the model itself - let us call them failings.

An example here is that the SM assumes that neutrinos should be massless - i.e. there is no mass term for them in the SM Lagrangian - and yet neutrino masses are a required property of the experimentally measured phenomenon of neutrino oscillations. It is possible that this issue can be fixed by simple addition of right handed neutrino fields to the SM Lagrangian, but until we pin down the mass ordering of the three flavours of neutrino we cannot say for sure and the way forward is uncertain [9].

Another major failing of the SM is the overall lack of charge-parity violation allowable in the theory. Charge-parity violation is one of the required conditions for the measured matter anti-matter asymmetry in the universe [10] and yet sources in the SM fall short of explaining this level by an enormous factor [11]. We will explore beyond the standard model theories that attempt to explain this missing charge-parity violation in the next chapter.

Conclusion

This chapter provided a review of the Standard Model, the current best theory of modern high energy particle physics. In the next chapter we will review a number of examples of extensions of the SM of varying levels of complexity, all of which attempt to provide solutions or at least improvements to the issues of the model that exist today.

Chapter 3

Beyond the Standard Model

The successes of the Standard Model have mostly pushed attempts to completely replace the theory to the fringe, meaning that beyond the Standard Model theories are extensions of the SM Lagrangian or symmetries of varying complexities. In this chapter we will cover three increasingly complex extensions of the SM that are all relevant to the published material in this thesis. The intention of this chapter is to help motivate and improve the digestibility of this thesis, especially chapters 6, 7 and 8.

3.1   Anomalous top-Higgs Coupling

As described in section 2.2 the lack of charge-parity violation present in the SM is vastly inadequate to explain the level of matter anti-matter asymmetry in the universe [10, 11]. Thus adding sources of charge-parity violation to the SM is an attempt to fix this inadequacy. One method of doing this is to look for certain couplings that are currently not pinned down to extremely high precision experimentally and add a charge-parity violating term to them in the SM Lagrangian. While generally this only allows for small (and insufficient) amounts of charge-parity violation it is still important to try and bridge this gap.

An example of a coupling that satisfies this criterion is the SM top-Higgs coupling. We can write the relevant Lagrangian term as follows:

$$\mathcal{L}_{tth} = -\frac{y_t}{\sqrt{2}}\bar{t}\left(\cos\xi_t + i\gamma_5\sin\xi_t\right)th \tag{3.1}$$

where, the mixing angle $\xi_t \in (-\pi, \pi]$ and the SM limit corresponds to when the mixing angle $\xi_t = 0$ and the Yukawa coupling adopts its SM value $y_t \to y_t^{SM} =$

$\sqrt{2}m_t/v$, where m_t is the mass of the top and $v \simeq 246$ GeV is the Standard Model Higgs vacuum expectation value.

This is an extremely simple extension of the SM but yields interesting top quark phenomenology which is explored in depth in chapter 7.

3.2 Two Higgs Doublet Model

A particularly simple extension of the SM Higgs Sector is to add a second Higgs doublet, called the Two Higgs Doublet Model (2HDM), which leads to a number of interesting phenemonological consequences.

Let us introduce a second doublet:

$$\Phi_1 = \begin{pmatrix} \phi_1^+ \\ \phi_1^0 \end{pmatrix} \quad \Phi_2 = \begin{pmatrix} \phi_2^+ \\ \phi_2^0 \end{pmatrix} \tag{3.2}$$

Now, both doublets acquire a VEV which can be defined generally as:

$$\langle \Phi_1 \rangle = \begin{pmatrix} 0 \\ \frac{\nu_1}{\sqrt{2}} \end{pmatrix} \quad \langle \Phi_2 \rangle = \begin{pmatrix} 0 \\ \frac{\nu_2}{\sqrt{2}} e^{i\theta} \end{pmatrix} \tag{3.3}$$

So we can write:

$$\Phi_1 = \begin{pmatrix} \phi_1^+ \\ \frac{h_1 + \nu_1 + ig_1}{\sqrt{2}} \end{pmatrix} \quad \Phi_2 = \begin{pmatrix} \phi_2^+ \\ \frac{h_2 + \nu_2 e^{i\theta} + ig_2}{\sqrt{2}} \end{pmatrix} \tag{3.4}$$

Where h_i and g_i are real scalar fields. The θ term is a CP-odd phase, and is a new source of charge-parity violation in the 2HDM.

3.2.1 Potential Term

As a result of the introduction of this second Higgs doublet we see many extra terms in the potential we did not see in the SM case. To ensure renormalisability we search for a potential of quartic order which we can write in a basis of gauge invariant Hermitian operators constructed from Φ_1 and Φ_2 [6], namely:

$$\hat{A} \equiv \Phi_1^\dagger \Phi_1 \tag{3.5}$$

$$\hat{B} \equiv \Phi_2^\dagger \Phi_2 \tag{3.6}$$

$$\hat{C} \equiv \frac{1}{2}\left[\Phi_1^\dagger \Phi_2 + \Phi_1 \Phi_2^\dagger\right] = \mathrm{Re}\left[\Phi_1^\dagger \Phi_2\right] \tag{3.7}$$

$$\hat{D} \equiv \frac{-i}{2}\left[\Phi_1^\dagger \Phi_2 - \Phi_1 \Phi_2^\dagger\right] = \mathrm{Im}\left[\Phi_1^\dagger \Phi_2\right] \tag{3.8}$$

Using these operators we can construct the potential by compiling all quadratic combinations of the above variables that still preserve gauge invariance. Doing this creates a potential of quartic order in Φ_1 and Φ_2 as intended:

$$
\begin{aligned}
V(\Phi_1, \Phi_2, \Phi_1^\dagger, \Phi_2^\dagger) = & - \mu_1^2 \hat{A} - \mu_2^2 \hat{B} - \mu_3^2 \hat{C} - \mu_4^2 \hat{D} + \lambda_1 \hat{A}^2 + \lambda_2 \hat{B}^2 + \lambda_3 \hat{C}^2 + \lambda_4 \hat{D}^2 \\
& + \lambda_5 \hat{A}\hat{B} + \lambda_6 \hat{A}\hat{C} + \lambda_7 \hat{A}\hat{D} + \lambda_8 \hat{B}\hat{C} + \lambda_9 \hat{B}\hat{D} + \lambda_{10} \hat{C}\hat{D}
\end{aligned}
$$

$$(3.9)$$

There are 14 free parameters here, compared to two in the SM Lagrangian. However, we can reduce this to 10 by enforcing charge symmetry, noting that the $\hat{D}$ is not invariant under this C-Symmetry:

$$
\begin{aligned}
D \to D' &= \frac{-i}{2} \left[(e^{i\alpha_1} \Phi_1^\star)^\dagger (e^{i\alpha_2} \Phi_2^\star) - (e^{i\alpha_1} \Phi_1^\star)(e^{i\alpha_2} \Phi_2^\star)^\dagger \right] \\
&= \frac{-i}{2} \left[e^{i(\alpha_2 - \alpha_1)} \Phi_2^\dagger \Phi_1 - e^{i(\alpha_1 - \alpha_2)} \Phi_1 \Phi_2^\dagger \right]
\end{aligned}
$$

$$(3.10)$$

Now in general, the relative phase of α_1 and α_2 is arbitrary but fixed. Choices of these phases lead to different conditions on the coefficients, one common one is $\alpha_1 - \alpha_2 = 0$ and in this case we see that $\hat{D}$ reverses sign. As a result, to ensure invariance of the Lagrangian μ_4, λ_9 and λ_{10} must all be zero. We may also have chosen $\alpha_1 - \alpha_2 = \pm\pi$ and we would of reversed the sign of $\hat{C}$ in which case the corresponding parameters would be set to zero. It is also important to note here that as we are dealing with a scalar theory we also have parity conserving terms, and thus C-symmetry is equivalent to CP-symmetry.

Now, this Lagrangian is CP-symmetric but it is possible for this symmetry to be broken when we choose our VEV. There are two ways to preserve the symmetry, the first is to demand Z_2 symmetry: $\Phi_1 \to \Phi_1$ and $\Phi_2 \to -\Phi_2$. Now under this condition clearly:

$$\hat{A} \to \hat{A} \tag{3.11}$$

$$\hat{B} \to \hat{B} \tag{3.12}$$

$$\hat{C} \to -\hat{C} \tag{3.13}$$

$$\hat{D} \to -\hat{D} \tag{3.14}$$

So obviously the coefficient of any term where this negative does not cancel must be zero, giving the Lagrangian:

$$V_{Z_2} = -\mu_1^2 \hat{A} - \mu_2^2 \hat{B} + \lambda_1 \hat{A}^2 + \lambda_2 \hat{B}^2 + \lambda_3 \hat{C}^2 + \lambda_4 \hat{D}^2 + \lambda_5 \hat{A}\hat{B} + \lambda_{10} \hat{C}\hat{D} \tag{3.15}$$

We could also impose global phase symmetry, $\Phi_2 \to e^{i\psi}\Phi_2$, which would induce the following shifts in our basis:

$$\hat{A} \to \hat{A} \tag{3.16}$$

$$\hat{B} \to \hat{B} \tag{3.17}$$

$$\hat{C} \to \hat{C}' = \frac{1}{2}\left[e^{i\psi}\Phi_1^\dagger\Phi_2 + e^{-i\psi}\Phi_1\Phi_2^\dagger\right] \tag{3.18}$$

$$\hat{D} \to \hat{D}' = \frac{-i}{2}\left[e^{i\psi}\Phi_1^\dagger\Phi_2 - e^{-i\psi}\Phi_1\Phi_2^\dagger\right] \tag{3.19}$$

Which effectively means that any term containing $\hat{C}$ or $\hat{D}$ must have a zero coefficient with one caveat:

$$\hat{C}'^2 = \frac{1}{4}\left[e^{2i\psi}\left(\Phi_1^\dagger\Phi_2\right)^2 + e^{-2i\psi}\left(\Phi_1\Phi_2^\dagger\right) - 2\right] = -\hat{D}'^2 \tag{3.20}$$

So under this transformation, if the $\hat{C}^2$ and $\hat{D}^2$ terms have the same coefficent they will cancel themselves. This leads to an extremely common form of the 2HDM Lagrangian that has (not spontaneously broken) CP-invariance:

$$V_\psi = -\mu_1^2\hat{A} - \mu_2^2\hat{B} + \lambda_1\hat{A}^2 + \lambda_2\hat{B}^2 + \lambda_3\left(\hat{C}^2 + \hat{D}^2\right) + \lambda_5\hat{A}\hat{B} \tag{3.21}$$

Our gauge eigenstates and mass eigenstates are related by [6]:

$$\begin{pmatrix} G^\pm \\ H^\pm \end{pmatrix} = \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix}\begin{pmatrix} \phi_1^+ \\ \phi_2^+ \end{pmatrix} \tag{3.22}$$

$$\begin{pmatrix} H^0 \\ h^0 \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \tag{3.23}$$

$$\begin{pmatrix} G^0 \\ A^0 \end{pmatrix} = \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix}\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} \tag{3.24}$$

Where,

$$\tan\beta \equiv \frac{\nu_2}{\nu_1} \tag{3.25}$$

$$\sin\beta = \frac{\nu_2}{\sqrt{\nu_1^2 + \nu_2^2}} \tag{3.26}$$

$$\cos\beta = \frac{\nu_1}{\sqrt{\nu_1^2 + \nu_2^2}} \tag{3.27}$$

and α is the CP-even Higgs mixing angle which is potential dependent.

In the above we see a particle spectrum consisting of two CP-even Higgs, h^0 and H^0, where the former denotes the lighter of the two (either could be the

Standard Model Higgs), a CP-odd pseudo-scalar Higgs, A^0, and two charged Higgs, $H^\pm$, which are the main focus of chapter 6 and chapter 8. We also have Goldstone bosons, G^0 and $G^\pm$, corresponding to the Z and $W^\pm$ vector bosons via the mechanism described in the previous chapter.

3.2.2 Kinetic Term

The kinetic sector extends simply by:

$$\mathcal{L}_{\text{kin}} = (D_\mu \Phi_1)^\dagger (D^\mu \Phi_1) + (D_\mu \Phi_2)^\dagger (D^\mu \Phi_2) \tag{3.28}$$

Where our covariant derivative is defined the same as for the Standard Model, as seen in section 2.1.1. By definition this Lagrangian is gauge invariant, but it is easily shown that it is also invariant under charge conjugation, Z_2 and global phase shift, meaning that it is unique (unlike the potential and the Yukawa term which will be seen in the next section). Now we can find the vector boson masses in the same way as for the Standard Model and this gives:

$$M_W^2 = \frac{1}{2} g^2 \left(\nu_1^2 + \nu_2^2 \right) \tag{3.29}$$

$$M_Z^2 = \frac{1}{2} \left(\nu_1^2 + \nu_2^2 \right) \left(g^2 + g'^2 \right) \tag{3.30}$$

$$M_\gamma^2 = 0 \tag{3.31}$$

3.2.3 Yukawa Lagrangian

As with the Standard Model Higgs sector the Yukawa Lagrangian provides the coupling of the Higgs doublets (and thus the Higgs bosons) to the fermions. It is this term in the Lagrangian in which a distinction is drawn between various types of Two Higgs Doublet Model, the most important of which are the Type-I and Type-II. Below is the most general gauge invariant Lagrangian that couples the Higgs fields to the fermions [6]:

$$\begin{aligned}
\mathcal{L}_Y = {}& \eta_{ij}^U \bar{Q}_{iL} \tilde{\Phi}_1 U_{jR} + \eta_{ij}^D \bar{Q}_{iL} \Phi_1 D_{jR} + \zeta_{ij}^U \bar{Q}_{iL} \tilde{\Phi}_2 U_{jR} \\
& + \zeta_{ij}^D \bar{Q}_{iL} \Phi_2 D_{jR} + \eta_{ij}^E \bar{\ell}_{iL} \Phi_1 E_{jR} + \zeta_{ij}^E \bar{\ell}_{iL} \Phi_2 E_{jR}
\end{aligned} \tag{3.32}$$

Where,

- $\tilde{\Phi} = i\sigma_2 \Phi$,

- η_{ij} and ζ_{ij} denote 3×3 matrices and i, j indicate generations,

- $U_R \equiv (u_R, c_R, t_R)^T$,

- $D_R \equiv (d_R, s_R, b_R)^T$,

- Q_L are weak isospin quark doublets,

- $E_{jR} \equiv (e_R, \mu_R, \tau_R)^T$,

- ℓ_L are weak isospin lepton doublets,

- All fields are gauge eigenstates (not mass eigenstates).

Type-I 2HDM

In this version of the model only one Higgs doublet couples to all of the fermions which is obtained by demanding the following Z_2 symmetries:

$$\Phi_1 \to \Phi_1 \tag{3.33}$$

$$\Phi_2 \to -\Phi_2 \tag{3.34}$$

$$U_{jR} \to -U_{jR} \tag{3.35}$$

$$D_{jR} \to -D_{jR} \tag{3.36}$$

$$E_{jR} \to -E_{jR} \tag{3.37}$$

In doing this we see that to preserve the Lagrangian the matrices η^U, η^D and η^E must all be zero. As a result the first Higgs doublet, Φ_1, no longer couples to fermionic matter and thus only Φ_2 contributes to the mass of the fermions.

$$\mathcal{L}_Y^I = \zeta_{ij}^U \bar{Q}_{iL} \tilde{\Phi}_2 U_{jR} + \zeta_{ij}^D \bar{Q}_{iL} \Phi_2 D_{jR} + \zeta_{ij}^E \bar{\ell}_{iL} \Phi_2 E_{jR} \tag{3.38}$$

Type-II 2HDM

The type-II 2HDM requires an extremely similar symmetry:

$$\Phi_1 \to \Phi_1 \tag{3.39}$$

$$\Phi_2 \to -\Phi_2 \tag{3.40}$$

$$U_{jR} \to -U_{jR} \tag{3.41}$$

$$D_{jR} \to D_{jR} \tag{3.42}$$

$$E_{jR} \to E_{jR} \tag{3.43}$$

In this scenario to preserve the Lagrangian the matrices η^U, ζ^U and ζ^E must be zero and thus the down type quarks and charged leptons couple to Φ_1, while only the up type quarks couple to Φ_2, each generating mass for their respective

fermions.

$$\mathcal{L}_Y^{II} = \eta_{ij}^D \bar{Q}_{iL} \Phi_1 D_{jR} + \zeta_{ij}^U \bar{Q}_{iL} \tilde{\Phi}_2 U_{jR} + \eta_{ij}^E \bar{\ell}_{iL} \Phi_1 E_{jR} \tag{3.44}$$

This is the type of 2HDM which is usually chosen as the Higgs sector in the MSSM and is used to tune the Yukawa couplings to provide a more natural quark mass hierarchy.

The most important points of this section are that the addition of a second Higgs doublet leads to a number of very well motivated phenomenological consequences and that the scalar particle spectrum of the 2HDM consists of 5 different Higgs bosons, including 4 that are not present in the Standard Model. We will see an investigation of the type-II 2HDM in chapter 6 where we explore the production of three top quarks at the LHC and the subsequent boost in cross section of this process that the existence of a charged Higgs boson would produce.

3.3 Supersymmetry

Supersymmetry is considered to be one of the most elegant solutions to many of the problems of the SM. Supersymmetric models are the result of extending the Poincaré algebra of the SM to include the commutation and anti-commutation relations:

$$[M^{\mu\nu}, Q_\alpha] = \frac{1}{2} (\sigma^{\mu\nu})_\alpha^\beta Q_\beta \tag{3.45}$$

$$[Q_\alpha, P^\mu] = 0 \tag{3.46}$$

$$\left\{ Q_\alpha, \bar{Q}_\beta \right\} = 2\sigma_{\alpha\beta}^\mu P_\mu \tag{3.47}$$

where Q_α and $\bar{Q}_\beta$ are Dirac spinors with α and $\beta = 1, 2$, $M_{\mu\nu}$ are the generators of the Lorentz transformations, $P_\mu = -i\partial_\mu$ are the generators of translation in the Poincaré algebra and σ^μ are the Pauli matrices.

There are many ways to handle the details of the supersymmetric extension which lead to vastly different phenomenology. In any case the consequences of this symmetry extension would lead to the existence of a relationship between fermions and bosons in the SM. That is to say that for every spin-$\frac{1}{2}$ fermion in the SM there would exist a spin-0 boson, and for every spin-1 boson (or spin-0 in the case of the SM Higgs) there would exist a spin-$\frac{1}{2}$ fermion.

3.3.1   Minimal Supersymmetric Standard Model

An extremely common supersymmetric model that we will discuss in detail here is the Minimal Supersymmetric Standard Model (MSSM), which is called this as it is the extension that introduces the minimum number of new particles and interactions that is still consistent with phenomenology [12]. This model arises from increasing the number of particles in the SM by replacing every SM fermion field by a "chiral superfield" and every SM vector field by a "vector superfield". A chiral superfield contains a Weyl spinor field (a fermion), a scalar field and an auxiliary field. A vector superfield contains a spin-1 boson, a spin-$\frac{1}{2}$ fermion and an auxillary field [13].

We can represent superfields of the quarks and leptons as follows:

$$Q_i \equiv \begin{pmatrix} u_{L_i} & \tilde{u}_{L_i} \\ d_{L_i} & \tilde{d}_{L_i} \end{pmatrix} \tag{3.48}$$

$$L_i \equiv \begin{pmatrix} \nu_{L_i} & \tilde{\nu}_{L_i} \\ e_{L_i} & \tilde{e}_{L_i} \end{pmatrix} \tag{3.49}$$

$$U_i \equiv \begin{pmatrix} u^c_{R_i} & \tilde{u}^c_{R_i} \end{pmatrix} \tag{3.50}$$

$$D_i \equiv \begin{pmatrix} d^c_{R_i} & \tilde{d}^c_{R_i} \end{pmatrix} \tag{3.51}$$

$$E_i \equiv \begin{pmatrix} e^c_{R_i} & \tilde{e}^c_{R_i} \end{pmatrix} \tag{3.52}$$

where the tilde indicates that the field is the superpartner to the SM field and i runs over the generations and the right-handed superfields are made up of the conjugate of the right-handed SM fields. The superpartner fields of the SM fermions are named by placing an 's' in front of the name of their SM relative, examples include the top and *stop*, electron and *selectron* or more generally quark and *squark* and lepton and *slepton*.

The superfields of the Higgs are slightly different. In supersymmetric models we are forced to introduce at minimum one extra Higgs doublet and couple different doublets to the up-type quarks and down-type quarks. In the MSSM we introduce exactly one extra doublet, exactly as seen in section 3.2 regarding the Type-II 2HDM. This means that the Higgs superfields are written as:

$$H_1 = \begin{pmatrix} H^0_1 & \tilde{H}^0_1 \\ H^-_1 & \tilde{H}^-_1 \end{pmatrix} \tag{3.53}$$

$$H_2 = \begin{pmatrix} H^0_2 & \tilde{H}^0_2 \\ H^+_2 & \tilde{H}^+_2 \end{pmatrix} \tag{3.54}$$

where the Higgs superpartner fields are called *Higgsinos*.

To complete the particle spectrum of the MSSM we must elevate our vector fields to vector superfields. These vector superfields, like their SM relatives, must transform under the adjoint of the gauge group. The superfields are written as:

$$V_S^A = \begin{pmatrix} G^{\mu A} & \tilde{G}^A \end{pmatrix} \tag{3.55}$$

$$V_W^I = \begin{pmatrix} W^{\mu I} & \tilde{W}^I \end{pmatrix} \tag{3.56}$$

$$V_Y = \begin{pmatrix} B^\mu & \tilde{B} \end{pmatrix} \tag{3.57}$$

where G and $\tilde{G}$ represent the gluon fields and their superpartner *gluino* fields, with $A = 1, 2, \ldots, 8$. W and $\tilde{W}$ are the electroweak fields (related to weak isospin) and their superpartner *wino* fields with $I = 1, 2, 3$ and finally B and $\tilde{B}$ are the remaining electroweak field (related to hypercharge) and its superpartner *bino* field.

R-Parity

In the SM there exists baryon and lepton number conservation. However in the MSSM, because the distinction between fermions and bosons disappears, many couplings arise that allow for baryon and lepton number to be violated. An important example of this can be seen in Fig 3.1 for the proton. This process shows a theoretically permissible process in the MSSM in which a proton turns into a lepton and meson. As the proton is extremely precisely experimentally measured to be stable for an extraordinarily long time we can make very strong statements about processes like these that violate baryon and lepton number in the MSSM.

As a result the following conservation law known as *R-parity* was put forth by Farrar and Weinberg [14]:

$$R_p = (-1)^{3(B-L)+2s} \tag{3.58}$$

where, B and L are the baryon and lepton number respectively and s is spin. Demanding that R_p be conserved in any interaction removes the above issue by ensuring that every interaction vertex has at least two supersymmetric partners. Remarkably, this also provides a fantastic candidate for dark matter as the lightest supersymmetric particle (LSP) cannot decay into any sparticles (as it is kinematically inaccessible) and cannot decay into SM particles as it would violate R-parity. Thus, the LSP provides a stable dark matter candidate.
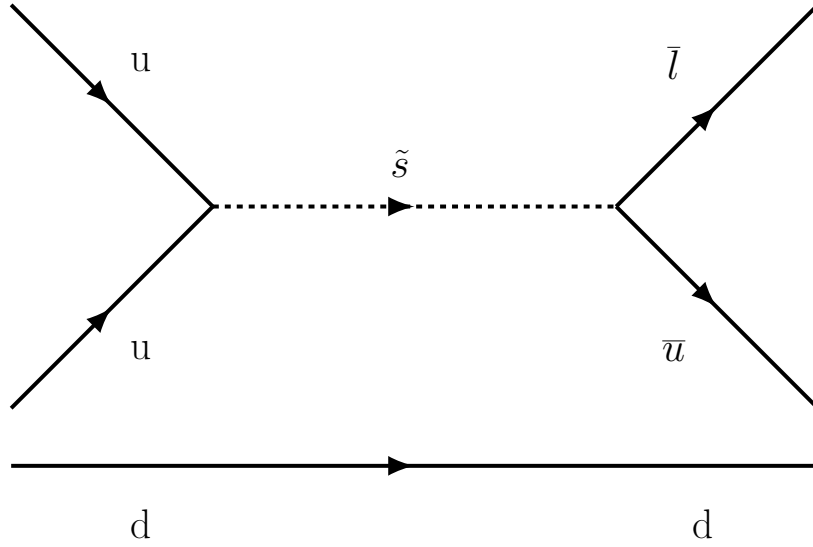
Fig. 3.1   An example of a proton spontaneously decaying to a lepton and a meson.

The Hierarchy Problem

In the SM the largest corrections to the Higgs mass come from higher order diagrams of the Higgs coupling to the most massive particles, predominantly the top quark. The Higgs mass correction can be written:

$$\Delta m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2} \left[\Lambda_{\text{UV}}^2 + \ldots\right] \tag{3.59}$$

where, Λ_{UV} is the assumed ultra-violet cut off of the SM. If we take Λ_{UV} to the Planck scale then the Higgs mass correction explodes. This is not proven to be an impossibility, but is considered by many physicists to be extremely unnatural and thus any high energy completion of the SM must solve this.

   If we consider a supersymmetric extension of the SM, this will in any case contain scalar superpartners of the fermions. The coupling of the Higgs to the new sparticle spectrum can cancel off the huge Higgs mass corrections - an example of this for the top and the stop can be seen in Fig 3.2.

   Mathematically, if we have Yukawa couplings of the Higgs to scalars in the new sparticle spectrum that are approximately equal to λ_f, say $\lambda_s \approx |\lambda_f|^2$, then the new Higgs mass correction will be:

$$\Delta m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2} \left[\Lambda_{\text{UV}}^2 + \ldots\right] + \frac{\lambda_s}{8\pi^2} \left[\Lambda_{\text{UV}}^2 + \ldots\right] \tag{3.60}$$

$$\approx 0 \tag{3.61}$$
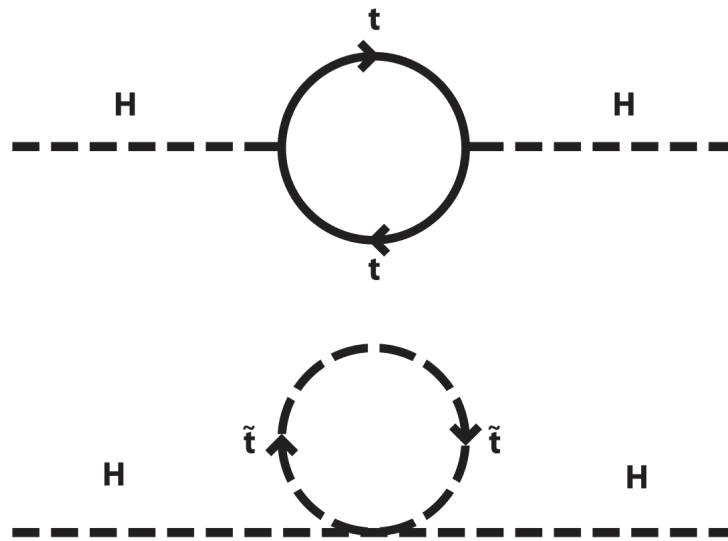
Fig. 3.2 Quantum corrections to the Higgs mass for the top quark and stop squark [15].

We can see that the extremely high mass correction has been cancelled off by the existence of the new sparticles. In the MSSM this problem is to a large extent fixed, however the loop corrections are reduced to about one order of magnitude higher than the Higgs mass - a huge improvement, but still not totally natural. Thus the MSSM still has its own *little hierarchy problem* [16].

In chapter 8 we see the MSSM used as a template to explore the often neglected experimental impact of quantum scattering interference between signals and backgrounds at the LHC. We explore how large the interference is in the production of a charged Higgs in association with a top quark and a bottom quark assuming two different benchmarks of the MSSM. As well as this in chapter 10 we see the MSSM used as a template for a study into the power of novel machine learning techniques at detecting anomalies in experimental data.

Conclusion

Now that we have explored a number of relevant beyond the Standard Model theories we will move onto an overview of the particle collider environment. In the next chapter the experimental hardware and terminology will be presented to aid in understanding of the manuscripts presented in this thesis.

Chapter 4

The Collider Environment

Studying modern particle physics theories experimentally requires the creation of high energy environments in which the high energy behaviour of physics models can manifest. To create these environments, particle colliders are used in which particles, usually electrons or protons (and their anti-particles), are accelerated to high energies typically on the order of GeV or TeV and then smashed together to stimulate interactions. The purpose of this chapter is to review the basic architecture of relevant particle colliders and detectors and the terminology and coordinate parametrisation used in particle physics studies.

4.1   The Large Hadron Collider

Residing on the border of Switzerland and France the Large Hadron Collider (LHC) is a 13 TeV centre of mass (CoM) energy proton-proton and heavy ion collider designed and built by the "European Organization for Nuclear Research" (CERN). Primary construction of the LHC began in 1998 and was completed in 2008 with first beam collisions occurring in 2010 at 7 and 8 TeV CoM energy and later upgrades to 14 TeV CoM energy occurring between 2013 and 2015 (although a decision was made to limit the machine to 13 TeV for normal operation).

The apparatus lies between 50 to 175 metres below ground in a 27 kilometre tunnel and has a main experimental ring approximately 8km in diameter. This is comprised of two parallel beam pipes which carry particle beams in opposite directions and cross over at four points where particle collisions occur. The beams are accelerated by 1,232 dipole magnets and focused by 392 quadrupole magnets with a number of higher multipole magnetic fields used to further improve the beam focus. The beams are comprised of "bunches" of protons allowing for collisions to occur in consistent discrete intervals.
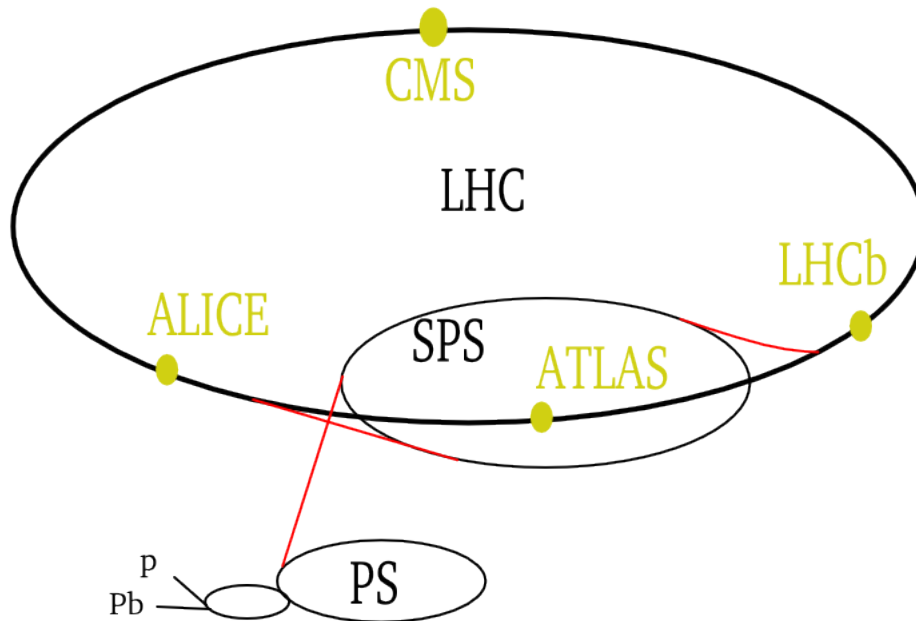
Fig. 4.1 A schematic of the Large Hadron Collider facility. [17]

At the four collision points of the main ring are four detectors - the ATLAS ("A Toroidal LHC ApparatuS") detector, CMS ("Compact Muon Solenoid"), ALICE ("A Large Ion Collider Experiment") and "LHCb". The first of these is important to the content of this thesis and will be expanded upon in section 4.2.

4.2   The ATLAS Detector

The ATLAS detector is a cylindrical barrel particle detector located 100 metres underground near the main CERN site in the town of Meyrin in Geneva, Switzerland. It is a general purpose detector designed and built to investigate the existence of the Higgs boson and many other BSM phenomena using the LHC proton-proton collisions.

The detector is comprised of a number of layers of various types of particle detector that are fit for different purposes such as the detection of different types of particles or particles of different orders of energy and angular positions. A cut away of the ATLAS detector can be found in Fig 4.2.

The innermost detector is comprised of an inner layer of high-resolution semiconductor pixel and strip detectors surrounded by a number of straw-tube tracking detectors to detect transition radiation. These two layers are immersed in a 2T magnetic field and together allow for momentum and vertex measurements and electron identification [18].

The next layer of the detector is a set of high granularity liquid-argon (LAr) electromagnetic sampling calorimeters with high energy and position resolution which cover the pseudorapidity range $|\eta| < 3.2$ (see section 4.3 for a description of pseudorapidity). In this layer is also a scintillator-tile calorimeter used for hadronic calorimetry covering the region $|\eta| < 1.7$. On either side ($|\eta| > 1.5$) are end-caps which contain LAr calorimeters used for the hadronic calorimetry extending the reach of the hadronic detection to $|\eta| \leq 4.9$ [18].

Finally the calorimeter is surrounded by a muon spectrometer. This is a large toroidal magnet with a long barrel and two end-cap magnets generating a strong magnetic field in a large volume in which the muon paths can curve through three layers of high precision tracking chambers to achieve high muon momentum resolution [18].

A technical overview of the ATLAS detector can be found in Ref [18] and in depth descriptions of each component can be found in Refs [19–32].
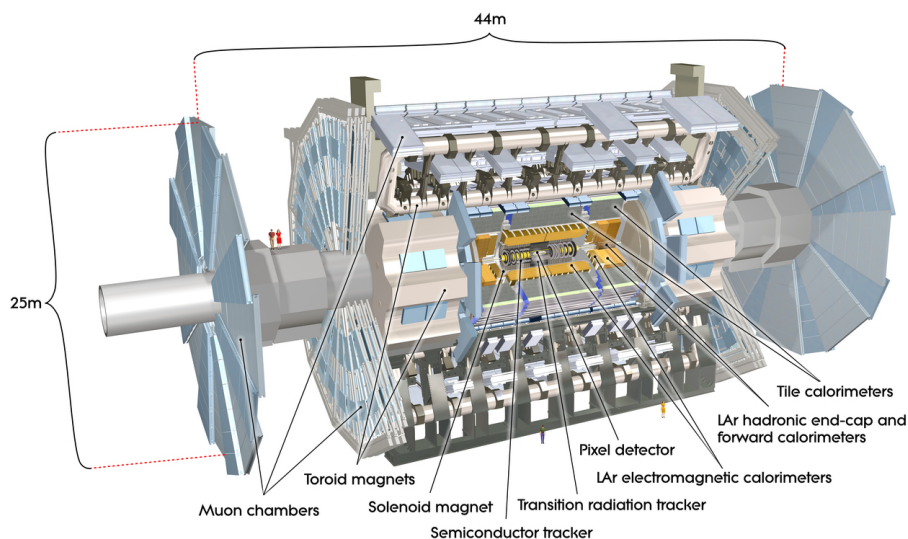
Fig. 4.2 A schematic of the ATLAS detector. [33]

4.3   Physics Objects

A description of subatomic particles and their behaviour has been provided in chapters 2 and 3 however this is not sufficient for a true understanding of particle phenomenology. The data taken by a detector such as ATLAS is extremely fine grained and complex (see Fig 4.3), with many levels of grooming undertaken to provide it in a useable form (not to mention the behemoth task of saving and storing the data which will not be covered in this work).

This resultant dataset does not necessarily contain the raw information about particles generated but rather signatures of them from the hardware in the detector.
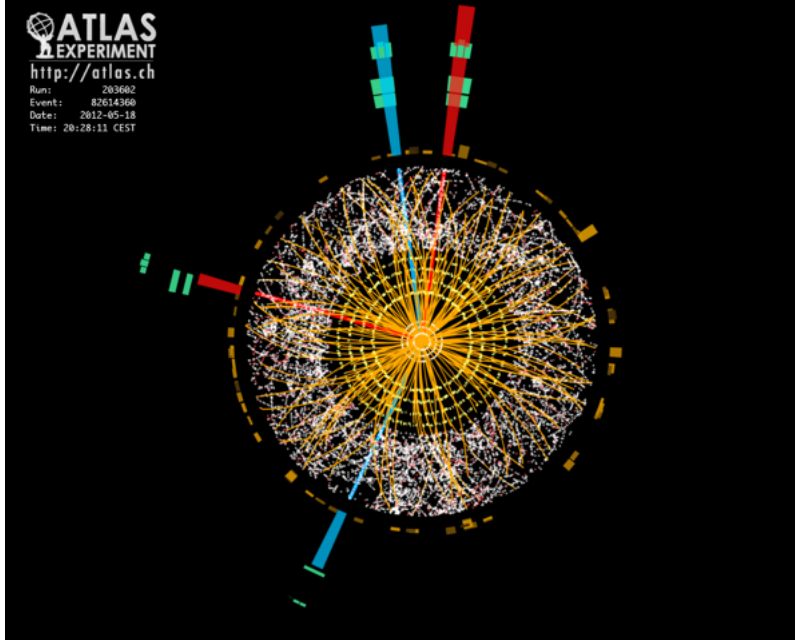
Fig. 4.3 An example of an event at ATLAS [34].

4.3.1 p_T-η-ϕ Description

The p_T-η-ϕ description is the coordinate system used in collider physics to describe the position of physics objects in space. Firstly, the beam-axis is defined as the z-axis. From this we then define $p_T = \sqrt{(p_x)^2 + (p_y)^2}$ as the transverse momentum of the physics object. The reason we employ the transverse momentum and not full momentum is that at hadron colliders such as the LHC, we only know the overall energy of each proton and not the specific energy of the constituents of protons (dubbed "partons") contained within them that have interacted. However, as both protons are moving along the beam axis we know that the total initial transverse momentum of the system is zero and thus the total transverse momentum of all produced particles in the event will sum to zero. As a note, when all detected particle transverse momenta are summed and the result is non-zero, this non-zero value is defined as "missing momentum" and its magnitude is called "missing energy" or $\not{E}_T$. It is generally identified as the presence of particles that have escaped the detector undetected - i.e. neutrinos, potential dark matter, etc.

ϕ is defined as the azimuthal angle around the beam axis. Finally η, also known as "pseudorapidity", is defined as:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{4.1}$$

where θ is the angle with the beam-axis. This value is useful as differences in pseudorapidity are Lorentz invariant under boosts along the beam axis - that is they transform additively like non-relativistic quantities. Fig 4.4 presents the behaviour of η for various values of θ between 0 and $\pi/2$. Using these three quantities we parametrise the collider environment.
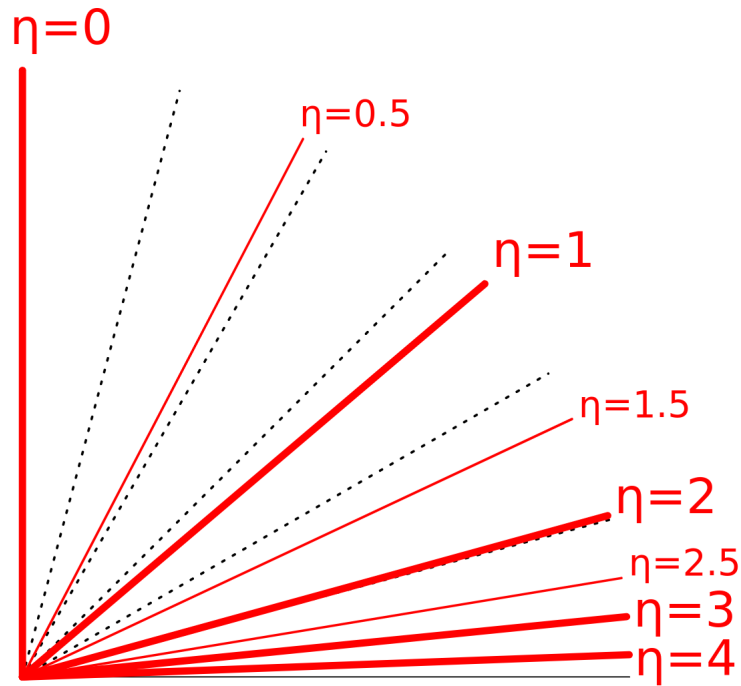
Fig. 4.4 A plot of pseudorapidity values for $0 \leq \theta \leq \frac{\pi}{2}$ [35].

4.3.2   Jets

In particle colliders colour charged objects such as quarks and gluons are extremely commonly produced. However, due to colour confinement, these bare colour objects cannot be observed. To avoid the bare colour these quarks and gluons will produce colour charged objects out of the vacuum to form colour neutral mesons and hadrons. These mesons and hadrons are not always stable and as such decay once again to bare colour charged objects. This process repeats, leading to a cascade of QCD particles in a conic shape onto

the detector. This process is known as "fragmentation" and "hadronisation" and the cluster of particles that hits the detector is called a "jet".

Jets are extremely important as they correlate with the quarks and gluons originally produced in a process and so are necessary for all physics analyses. The study of accurately and efficiently clustering fragmentation/hadronisation products into jets is a large field of study and many algorithms exist to do this. These algorithms broadly fall into two classes, "cone" algorithms [36] and "sequential recombination" algorithms. The former of these aim to define stable cones as circles of fixed radius in the $\eta - \phi$ plane in which the sum of all momenta of particles in that circle are in the same direction as the center of that circle. This seeks to define cones of energy flow of groups of particles and as such the conical shape of a jet. However, they are not commonly used in modern analyses and won't be discussed at length here.

The two most common sequential recombination algorithms are the "Anti-kt" [37] and Cambridge-Aachen [38]. These work as follows:

1. For each particle construct the following quantities:

   (a) $d_{ij} = \min(k_{T,i}^n, k_{T,j}^n)\frac{\Delta R_{i,j}^2}{R^2}$

   (b) $d_{iB} = k_{T,i}^n$

   where, n is an integer, $k_{T,i}^n$ is the transverse momentum of the ith particle, $\Delta R_{i,j}$ is the distance between the ith and j^{th} particle in the (η, ϕ) plane and R is a scaling parameter of order unity.

2. Then find $d_{min} = \min(d_{i,j}, d_{i,B})$ for all i, j.

3. If d_{min} is a $d_{i,j}$ then merge particles i and j into one by 4-vector summation, else if d_{min} is a $d_{i,B}$ then delete particle i from the list.

4. Repeat above iteratively until the list contains one particle.

The end result of this algorithm is one 4-vector that is a sum of all the particles which satisfied the above criterion. This 4-vector gives the momentum and energy of the jet the algorithm has found. The difference between the sequential recombination algorithms is in the scale of the energy contribution to the metric, n. For Anti-kt $n = -2$ is chosen and for Cambridge-Aachen $n = 0$ is chosen and also the variable $d_{i,B}$ is set to 1 for all i.

Typically one cannot tell which specific quark or gluon generated a jet, however there are exceptions to this. "Light jets" are dubbed as such because they were generated by a "light quark", that is an up, down or strange quark, or a gluon. It is possible, by assessing whether the vertex of the jet (the tip

of the cone shape) is displaced away from the interaction point of the overall process and by measuring how large this displacement is, to tag the jet as being generated by a charm quark or a bottom quark. This is known as "charm tagging" and "*b*-tagging" respectively. See Fig 4.5 for an example of this.
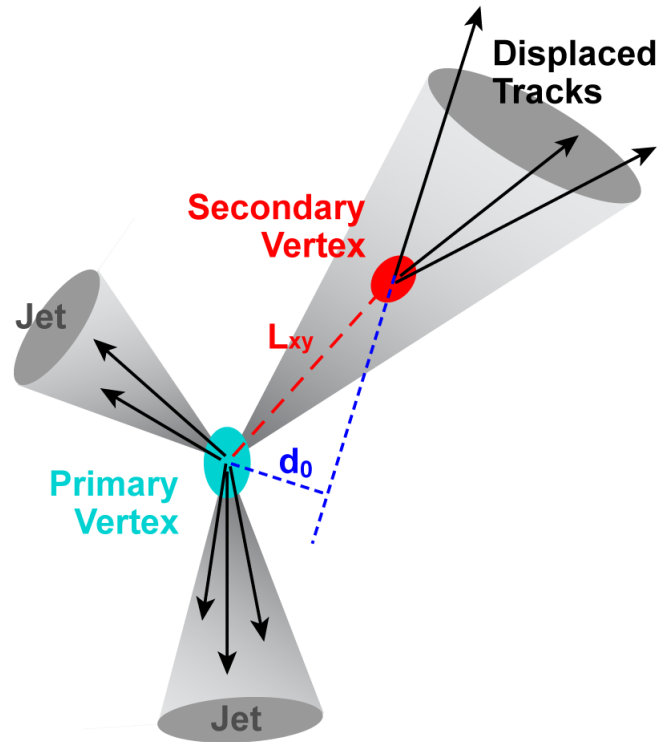
Fig. 4.5 An example of two non-displaced vertices indicative of light jets and a displaced vertex indicative of a *b*-jet [39].

Finally, a top quark decays via $t \to bW^+$ before it ever has a chance to go through the fragmentation/hadronization process because of its extremely high mass. As a result the top produces a *b*-jet and then either a lepton and a neutrino or a pair of jets depending on the W boson decay path. That being said, a highly boosted top quark may potentially generate such collimated decay products that the *b*-jet and the W boson decay products form a single conical shape which is clustered as what is called a "fat jet", an example of which is given in Fig 4.6.

The identification of fat jets is another large field of research and techniques range from the common "mass-drop tagger" approach [41] to using machine learning for image recognition [42]. It is important to note that the term fat jet is often also applied to any time that multiple distinct jet objects are boosted enough to be misidentified as a single jet object, but the principles of identifying them remain the same.
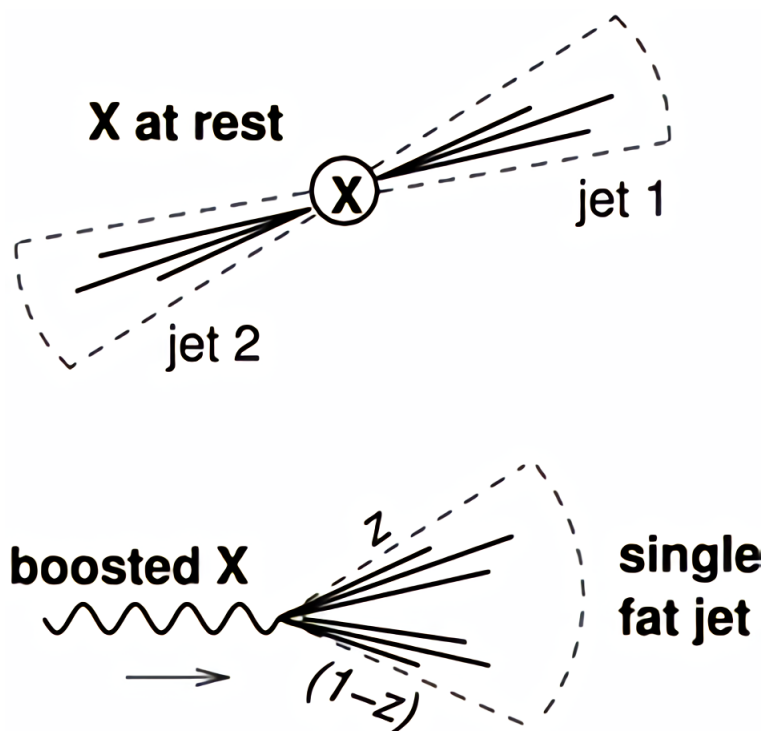
Fig. 4.6 An example of a boosted particle leading to the formation of a single fat jet. [40]

Conclusion

With a general overview of the nature of particle colliders, in particular the LHC and the ATLAS detector, and the collider environment complete, let us now move onto a discussion of analysis methods and tools which underpin the data science that is required to discover new physics.

Chapter 5

Analysis Methods and Tools

The purpose of this chapter is to provide an overview of the techniques and methodologies that are employed in particle physics analyses. Its structure is presented such that the reader will hopefully gain an understanding of the pipeline from theory to experimental analysis. To aid in understanding the workflow we will use a study of the production of a Higgs boson as an example.

5.1 Lagrangian and Feynman Rules

The first step in the workflow is the implementation of a Lagrangian. This Lagrangian includes the theoretical new physics in the form of new terms or symmetries in the SM Lagrangian that we wish to study.

In our Higgs example we will use an implementation of an effective Langragian to facilitate generation of higher order processes at tree level. Often Monte Carlo generators are limited to tree level generation, and yet some interesting processes have leading orders above tree level. The Higgs decaying to two photons is a good example of this - it occurs predominantly through a top loop, which can be parametrised as an effective vertex by integrating around the loop and sending the top mass to infinity. This is a very good approximation for a Higgs with mass less than 150 GeV. The effective Lagrangian takes the form of:

$$L_{h\gamma\gamma} = -\frac{1}{4}gF_{\mu\nu}F^{\mu\nu}H \tag{5.1}$$

There are a number of tools which facilitate the construction of Lagrangians computationally and the derivation of Feynman rules from the Lagrangian such as `FeynRules` [43], `FeynCalc` [44], `SARAH` [45] and `LanHEP` [46]. There are many overlapping functionalities of these packages with the most common usages being the calculation of cross sections, generation of Feynman diagrams and
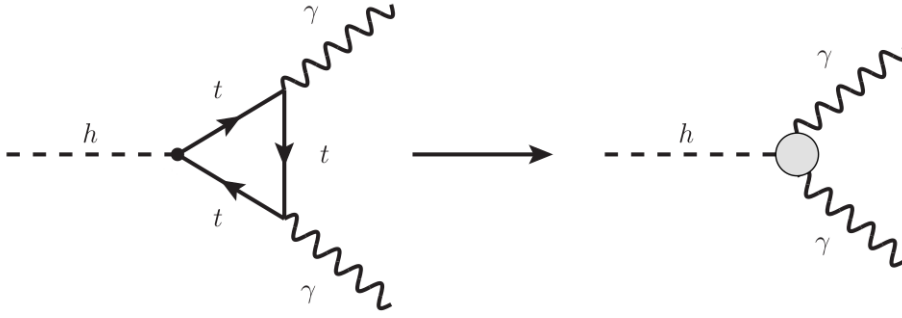
Fig. 5.1 Feynman diagram for Higgs decaying to two photons parametrised by
an effective vertex.

the generation of "model files" for implementation in Monte Carlo generators.
A model file is the Lagrangian implemented in an agreed upon way that is
understood by Monte Carlo generators to allow efficient and user friendly
generation of particle physics events which can then be analysed and compared
to real data.

So for this example we would choose a Feynman rules generator and add
on the above effective Lagrangian term. The software will generate the new
Feynman diagrams and Feynman rules leading to a replacement of diagrams
seen in Fig 5.1. This new diagram on the right is digestible by tree level Monte
Carlo generators and the Feynman rules would include the new coupling which
is a simple scalar equal to:

$$
\begin{aligned}
g = -\frac{\alpha}{\pi v}\frac{47}{18}\bigg(& 1 + \frac{66}{235}\tau_w + \frac{228}{1645}\tau_w^2 + \frac{696}{8225}\tau_w^3 + \frac{5248}{90475}\tau_w^4 \\
& + \frac{1280}{29939}\tau_w^5 + \frac{54528}{1646645}\tau_w^6 - \frac{56}{705}\tau_t - \frac{32}{987}\tau_t^2\bigg)
\end{aligned}
\tag{5.2}
$$

where, $\tau_t = m_h^2/(4m_t^2)$ and $\tau_w = m_h^2/(4m_W^2)$.

5.2   Signal and Background Definition

The definition of the signal is primarily motivated by what is phenomenologically
interesting, i.e. new and different to the SM, and what is experimentally
achievable, i.e. cross section is high enough and is non-degenerate with SM
backgrounds.

This is done by selection of a particular process and subsequent decay chain
that will include effects of the BSM parts of the Lagrangian introduced earlier.
Using the Higgs example, the choice of decay of the Higgs provides different

benefits and detriments. The Higgs decaying to two *b*-jets has the highest
branching ratio, however *b*-jets are experimentally messy and introduce larger
systematics and irreducible backgrounds. The Higgs decaying to two photons
provides an incredibly clean, low systematic signature at the expense of a
branching ratio that is roughly 1000 times lower.

Further, as can be seen in Fig 5.2, many of the leading production mecha-
nisms generate associated particles. Analyses can be constructed which target
these associated particles and their kinematics. For example, if one chooses to
investigate a top decaying to light jets (via a *W*-boson) the branching ratio is
roughly 3 times higher than if one chooses a top decaying to a lepton and neu-
trino. However, the lepton and neutrino scenario has much lower systematics
due to the very precise measurement of the lepton's momentum.
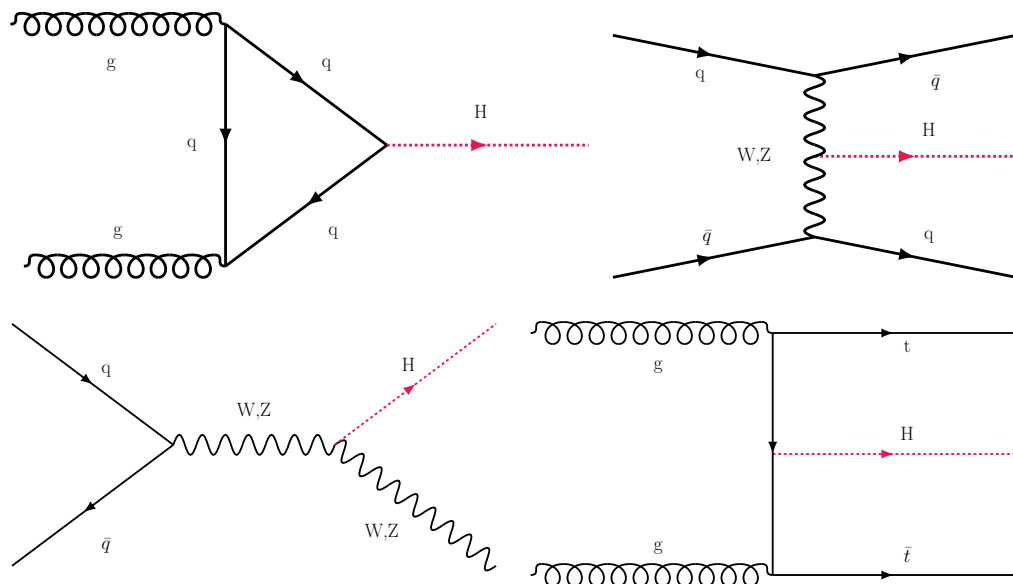
Fig. 5.2 Examples of Higgs production Feynman diagrams.

All of this is to say that the choice of signal is a delicate balance. Once
a signal, or set of signals, is chosen the backgrounds can be defined as those
with the same final state. It is important to note that this does not only mean
events with the exact same particles. It is possible that things like mis-tagging
of b-jets and light jets, or particles lost from the detector in its blind spots can
lead to experimental signatures that directly mimic the final state of the signal.
This must all be accounted for in the analysis.

Using the Higgs production, specifically the top left diagram of Fig 5.2,
followed by a decay to two photons as an example then we would define the
major background as any event with more than two energetic photons. Another
background could be any event that produces a photon and a light jet which is
then mis-tagged as a second photon. Of course we could go further and require

two or more mis-tagged light jets as a photons but with each mis-tag we are probing more and more unlikely occurrences which quickly suppresses the cross section to levels considered negligible.

5.3   Monte Carlo Generators

Monte Carlo generators such as `MadGraph` [47], `Pythia` [48], `SHERPA` [49], `POWHEG BOX` [50] and `HERWIG` [51] are absolutely indispensible to particle physics studies. They allow the intake of the aforementioned model files produced by the Feynman rule calculators and output a variety of information, most usefully, physically consistent particle collision events.

After utilizing one of these packages, and perhaps multiple if one would like to be extremely rigorous in cross checking the consistency of the results, the output is likely at production level - that is to say that the event will contain only the particles directly produced by the particle collision and not the decay products. However this is not what the detector sees and thus not what the data recorded will look like, so there are more steps to the process to simulate physically realistic events.

The produced particles must be realistically decayed and this will include subsequent hadronisation/fragmentation. Chapter 4 defined hadronisation and fragmentation - the generation of jets due to the production of bare colour particles in a collider - but this process is not well understood at all at a fundamental level. This is because this process is non-perturbative and the specific nature of quantum chromodynamic confinement itself is very poorly understood. Fortunately there exists an empirically tested phenomenological model that can simulate the hadronisation and fragmentation process - the "Lund String Model" [52]. This model is implemented in many of the aforementioned packages such as `Pythia`, `SHERPA` and `HERWIG` and those that do not such as `MadGraph` provide easy connectivity to other packages which do.

Finally, the events at this point contain only the values directly sampled during the Monte Carlo - it does not contain any effects that a realistic detector would create. As a result the final step in our simulation process is to introduce realistic detector effects which is often done by smearing the values of the MC data using functions of η, ϕ and transverse momentum that are empirically derived from real data. Many experiments have their own in house packages for very accurate effects of their specific experiment but general purpose packages such as `Delphes` [53] are also extremely commonly used.
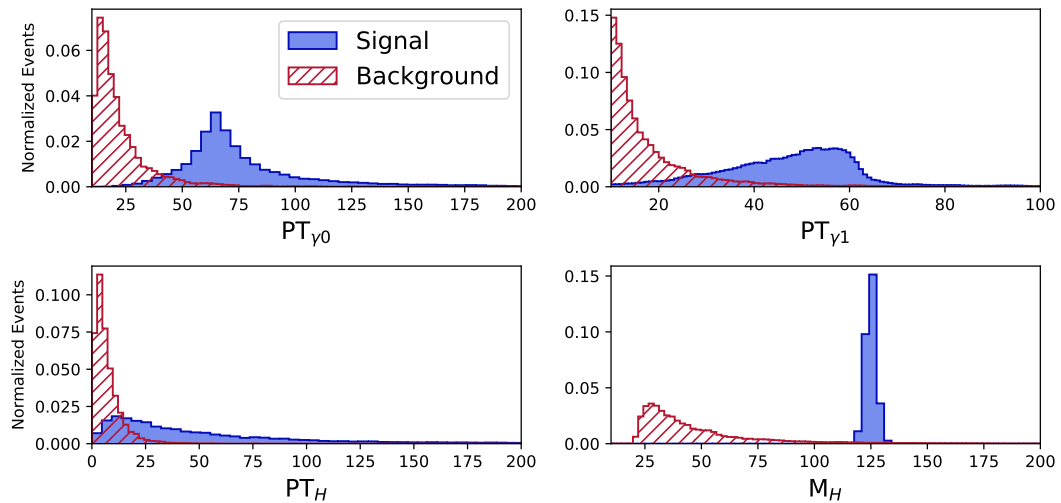
Fig. 5.3 Examples of variables for a Higgs produced then decaying to two photons.

To reiterate, the work flow of particle physics simulation is as follows: First generate a model file that includes Feynman rules from the new physics Lagrangian, define a signal and its corresponding backgrounds, then use this model file in MC generators that produce particle collision events. Next, decay the events and ensure that the relevant particles undergo hadronisation/fragmentation. Finally, smear the events to include the effects that a real detector would introduce. After the full dataset is produced, it is time to analyse this data. Fig 5.3 displays some examples of variables for the process $pp \rightarrow h \rightarrow \gamma\gamma$ and the relevant backgrounds of di-photon continuum events and light jet continuum events that have been mis-tagged as photons. In this figure γ_0 and γ_1 indicate the first and second most energetic photons in the event respectively, while PT_H and M_H are the transverse momentum and invariant mass, respectively, of the sum of 4-vectors of γ_0 and γ_1.

5.4   Analysis

The general purpose of an analysis is to quantify a statistical significance of the BSM effect when compared to the SM background. This is done by searching the data for an excess of events in the data over what would be expected from the SM background. Before we can get into specific methods of data analysis let us first take a look at a few common methods of quantifying signal significance in particle physics experiments.

5.4.1   Significance Tests

Poisson Signal to Noise Ratio

In any counting experiment the data is expected to be Poisson distributed. As a result the standard deviation in counting N events is expected to be $\sqrt{N}$. To quantify the significance of a count of S excess events over the expected count of B events, we count the number of standard deviations from the mean that our total count of $S + B$ events is. We do this in the typical fashion; measured value minus the mean divided by the standard deviation. That is:

$$\sigma = \frac{(S + B) - B}{\sqrt{S + B}} = \frac{S}{\sqrt{S + B}} \tag{5.3}$$

There are also many systematic errors present in any particle physics experiment and they are not incorporated into the above measure of significance. If all of these systematic errors are formed into a single value of expected absolute uncertainty called ΔB or some relative uncertainty of the background called β, then these can be included in the above test statistic as such:

$$\sigma = \frac{S}{\sqrt{S + B + (\Delta B)^2}} = \frac{S}{\sqrt{S + B + (\beta B)^2}} \tag{5.4}$$

This simple test is one of the most common tests of significance for simple searches for excess counts of events, especially in phenomenological papers. However, more robust statistical tests will often be used such as those seen below.

Chi-Squared Test

The Chi-Squared Goodness-of-Fit test [54] tests whether a sample of data was sampled from a population with a specific distribution. The null hypothesis H_0 is that the data <u>was</u> sampled from the chosen distribution and the alternate hypothesis H_a is that the data <u>was not</u> sampled from the chosen distribution. The test statistic is defined as:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{5.5}$$

where k is the number of bins of the histogrammed data, O_i is the number of events in the ith bin and E_i is the expected number of events in the ith bin. E_i

is calculated by:

$$E_i = N(F(Y_i^u) - F(Y_i^l)) \tag{5.6}$$

where F is the cumulative distribution function for the chosen distribution being tested, Y_i^u and Y_i^l are the upper and lower limits for the ith bin and N is the sample size.

The null hypothesis is rejected if the test statistic χ^2 satisfies:

$$\chi^2 > \chi^2_{1-\alpha,k-c} \tag{5.7}$$

where the right hand side of the equation is a scalar corresponding to the chosen distribution with α the chosen significance level, k the number of bins and c the number of parameters of the chosen distributions plus one, i.e. for a binomial distribution $c = 1$, a Poisson distribution $c = 2$ and a normal distribution $c = 3$.

For obvious reasons the Chi-Squared test is affected by the choice of binning of the data. There is no optimal binning, however a general rule of thumb is that for the test to work each bin should contain more than five events. Obviously, for any finite dataset and fine enough binning there will always either be a bin with less than five events in the tail of the distribution or some data will have to be excluded to remove this inevitability. To handle this, sometimes an "overflow" bin is used, where all of the bins to the right of a bin and with an unsatisfactory number of events are summed into a single bin.

CLs Technique

The CLs Technique [55] is a statistical test to set exclusion limits on model parameters that take non-zero values. Consider an experiment designed to measure N events resulting from a signal and a background process with expected events S and B respectively; thus $N \sim P(S + B)$. Now, if the experiment results in a count of N^* events we can construct the following quantity for all values of the parameter(s) of the signal model (and thus values of S) that satisfy:

$$\frac{P(n \leq n^*|s + b)}{P(n \leq n^*|b)} \leq \alpha \tag{5.8}$$

where α is chosen such that $1 - \alpha$ is our chosen confidence level. We would now have generated a limit on the parameter(s) at the confidence level of our choice.

Intuitively what is happening here is that we are finding all the values of the parameter(s) of the signal model that generate too many signal events such that the recorded data is more likely the result of statistical fluctuation in the background given no signal than the result of any real signal contribution.

The CLs Technique is used commonly in rigorous experimental studies, including famously in setting upper limits for the mass of the Higgs boson at LEP, ATLAS and CMS [56].

5.4.2 Cut Flows

Historically signal significance was optimised manually using "cuts" to generate a "cutflow"; this means that we discard events that do not meet certain criteria until the remaining events belong to a section of the phase space that, hopefully, represents a higher proportion of signal events relative to background events.

Looking at our Higgs example in Fig 5.3, it can be seen that the two photons produced by the decaying Higgs have a much higher transverse momentum than that of the background. So a reasonable cut may be that the leading photon in each event must have $PT \geq 40$ GeV and the sub-leading photon $PT \geq 30$ GeV. Another could be that one first reconstructs the Higgs by 4-vector summing the two photons and then demanding the invariant mass of the two photons be within 125 GeV plus or minus 20 GeV (note that this cut assumes a mass of the Higgs which is not known a priori). Either of these cuts would lead to the phase space being restricted specifically to an area where the signal is over-represented. An example of a cutflow on the Higgs boson can be seen in Tab 5.1, where S, $B1$ and $B2$ are the expected event counts for the signal, the di-photon continuum background and the photon and light jet background respectively. Note how the signal significance increases throughout the cutflow table as the cuts disproportionately remove the background relative to the signal.

Cut	S	$B1$	$B2$	σ_0	$\sigma_{0.01}$
No cuts:	4500	18853617	1888487	0.99	0.02
$N_\gamma \geq 2$:	2102	7146840	717036	0.75	0.03
$PT_{\gamma 0} \geq 40$ (GeV):	2025	624089	61292	2.44	0.29
$PT_{\gamma 1} \geq 30$ (GeV):	1830	546843	54201	2.36	0.30
$|M_{\gamma\gamma} - 125| \leq 15$ (GeV):	1828	126909	12186	4.87	1.27

Table 5.1 A simple cutflow table for a Higgs boson signal for 150 fb^{-1} of luminosity. Note: $B = B1 + B2$ and $\sigma_\alpha = S/\sqrt{S + B + (\alpha B)^2}$ is the significance with relative systematic uncertainty α.

It is much more common today to use the more advanced approach of supervised machine learning. These frameworks classify signal and background on an event by event basis leading to a far more surgical refinement of the phase space than the cutflow described above. The following subsections will describe some of these algorithms in detail.

5.4.3   Supervised Machine Learning

Supervised machine learning is the process of learning arbitrary functions that map inputs onto outputs via labelled training data. This is done by the minimisation of an error function defined by the difference between the network's output and the label of the input data; the error function is minimised by iteratively updating the parameters of the model.

The most common method of updating the parameters of the model to minimise error is called "gradient descent". Let us take a closer look at it before we move onto specific machine learning algorithms.

5.4.4   Gradient Descent

The idea behind "gradient descent" optimisation is that given some value of an arbitrary function the fastest path towards a minima of that function is in the opposite direction of the gradient of the function at that point. Intuitively one can picture this as standing on the side of a hill; the shortest path towards the bottom of the hill will be in the direction of the steepest slope.

Mathematically, if $F(\mathbf{x})$ is an arbitrary multi-variate function and $\mathbf{a_n}$ is a value of this function, then gradient descent can be written:

$$\mathbf{a_{n+1}} = \mathbf{a_n} - \gamma \nabla \mathbf{F(x)} \tag{5.9}$$

Where, $\mathbf{a_{n+1}}$ is a new point that should be closer to a maximum or minimum of $F(\mathbf{x})$ and γ is a scalar which acts to scale the step size of each step of the update process. This step size is not necessarily static and many algorithms exist to dynamically choose γ.

Two important examples of supervised algorithms that use gradient descent are Boosted Decision Trees and Artificial Neural Nets, let us take a closer look at these.

5.4.5 Boosted Decision Tree

To begin the overview of the Boosted Decision Tree algorithm we will first take a look at the concept of "decision trees". A decision tree is a set of cuts, similar to described earlier however in this case called decisions, but with the variables used and cut values optimised computationally [57, 58]. There are many methods used to choose the cut value of each decision in the tree, however the most common is to maximise "information gain" at each split [59]. An example of a decision tree for the Higgs signal can be found in Fig 5.4, note that it is honing in on the $PT_{\gamma0} > 40$ GeV and $PT_{\gamma1} > 30$ GeV region described earlier. The coloured rectangles are referred to as nodes, while the terminal nodes are referred to as "leaves". These leaves represent a purity of the given class after the cut, in this case $S/(S+B)$.
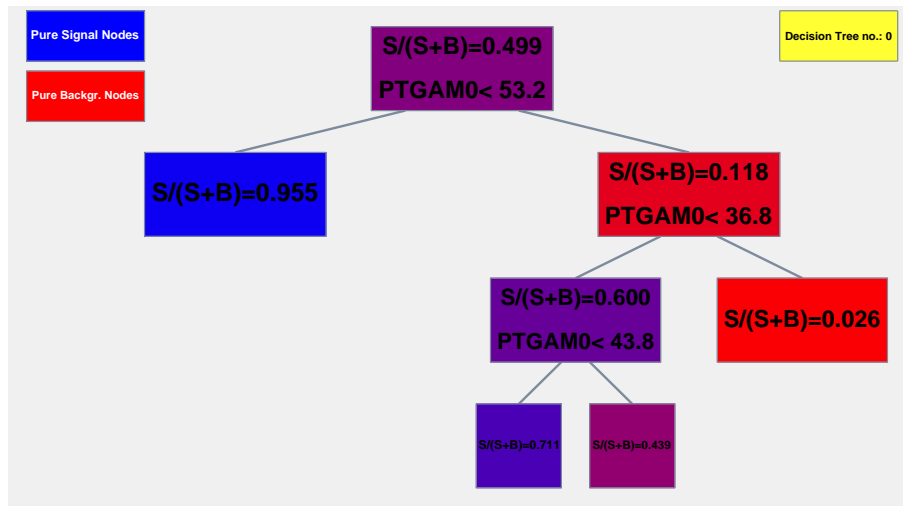
Fig. 5.4 An example of a decision tree.

However this individual tree suffers from very high bias and variability in its outputs, i.e. it is not able to catch a wide range of properties of the input sample. To improve this we replace the single tree with an ensemble of N trees, and the output is now a function, often a weighted mean, of the entire ensemble's outputs. Importantly, each tree is trained on a subset of the data sampled with replacement, this is called "bagging" or "bootstrap aggregation". Furthermore, the bagging methodology can be extended to allow for the random selection of input variables of the model, rather than cutting on all of the variables sequentially, this is known as "random forest".

Finally, to turn a decision tree into a boosted decision tree ensemble we implement "boosting". After the first tree in the series is trained, the second tree in the series is trained on a dataset with an increased weighting given to the events that the previous tree had the highest errors in classifying.

This is repeated along the series of trees leading to each subsequent tree being optimised to classify the most difficult to classify events of the previous sequence of trees. There are many methods of combining the classifiers outputs into one, some of the most common are "gradient boosting" [60–62] which includes gradient descent applied to a chosen error function and "AdaBoost" or "adaptive boosting" [63].

An example of the output of a boosted decision tree implemented in `ROOT` [64] using the Toolkit for MultiVariate Analysis (`TMVA`) [65] can be found in Fig 5.5. The top plot shows information regarding the classification output of the BDT for the training and testing sets. For a given event the BDT outputs a single value that represents whether that event belongs to the signal class or the background class called the "BDT response". This BDT response when histogrammed presents a very nice visualisation of the separation of the signal and background in one dimension. Further, the plot gives a sense of overtraining - when the model performs extremely well on the training set but does not perform well when presented with new data - by comparing the shape of the testing distribution, the hatched area, and the training distribution, the points crossed with error bars. In an overtrained model the two distributions will look very different for either the signal or background or both.

The bottom plot presents the statistical significance as a function of placing a cut on the BDT response distributions. If one were to throw away all events that have a BDT response less than 0.2438, the remaining set of events would have a much higher proportion of signal events surviving the cut than background events. In fact when signal significance is defined as $S/\sqrt{S+B}$ where S and B are the number of signal and background events remaining in the dataset respectively, the signal significance would be approximately 5.15σ. It is at this point that we should recall that this Higgs example is very much a toy model and is not taking into account all of the backgrounds and systematics that a full analysis would. It is also using a much higher luminosity than the original Higgs discovery. That being said, what is important is to note that the BDT analysis of the same data and same variables has led to a much higher signal significance than the manual cutflow.

5.4.6 Artificial Neural Nets

Artificial neural nets (ANN's) are arguably the most influential machine learning methodologies both in physics and in many facets of modern society. These techniques use a network of "nodes" connected by "weights" to approximate arbitrary non-linear functions [57].
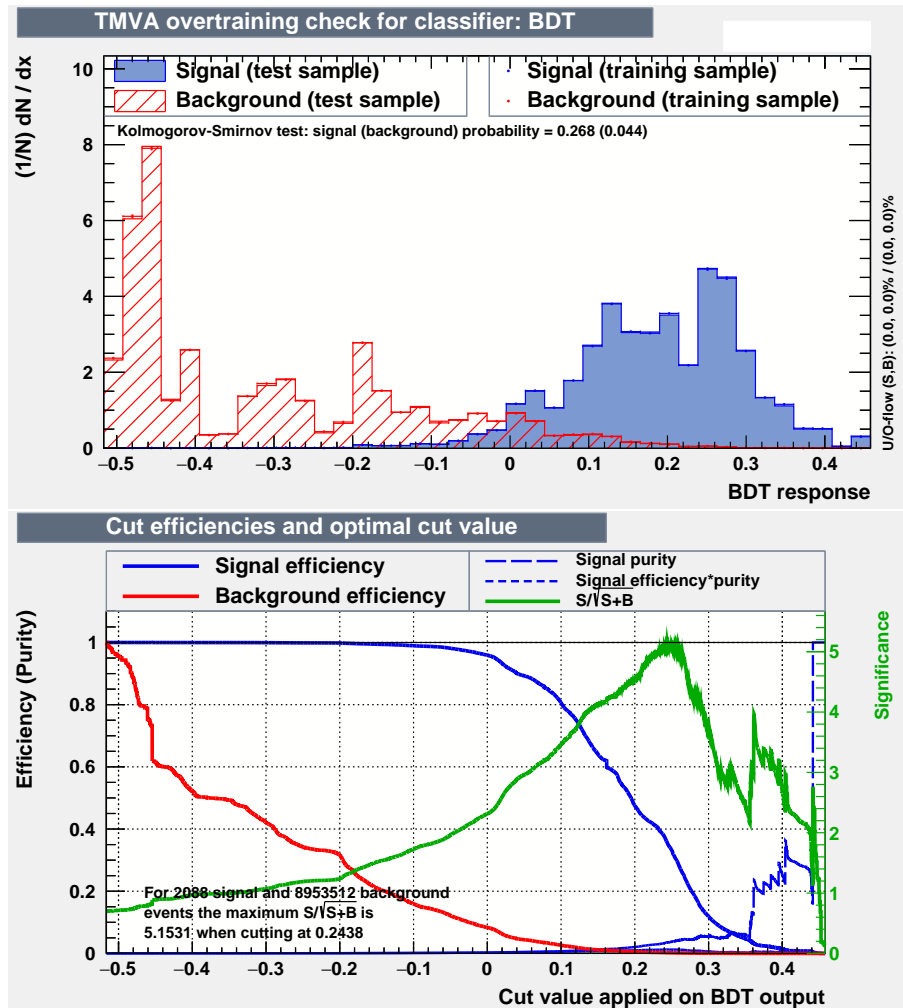
Fig. 5.5 Top: The BDT output for a sample of Higgs decaying to two photons signal and corresponding background events. Bottom: Signal significance as a function of where the cut is placed on the BDT output plot.
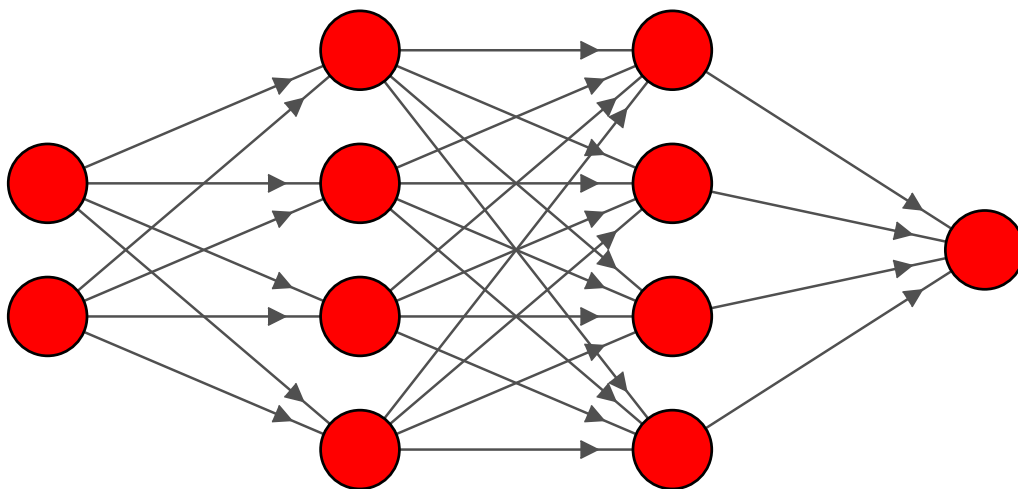
Fig. 5.6 An example of an artificial neural network.

Fig 5.6 presents a schematic of an ANN. This network takes two input values, has two layers in the middle called the "hidden layers" with four nodes (sometimes called neurons) and will generate one output value. The lines connecting each of the nodes represent weightings, values between zero and one, that will multiply the value stored in the node on the left and pass it to the node on the right. The value received by a hidden layer node is a sum of all nodes connected to it from the previous layer weighted by the weight of their connection plus a bias value (without a bias all zeros would always map to zeros). Once the node has received its input value from the preceding nodes it applies an "activation function" to it and passes it on to the nodes it is connected to in the next layer. Some examples of common activation functions are:

$$\text{Hyperbolic Tan:} \qquad \sigma(x) = \tanh(x)$$

$$\text{Leaky Rectified Linear Unit (Leaky ReLu):} \qquad \sigma(x) = \begin{cases} x & x > 0 \\ \alpha x & x \leq 0 \end{cases}$$

$$\text{Exponential Linear Unit (ELU):} \qquad \sigma(x) = \begin{cases} x & x > 0 \\ \alpha(e^x - 1) & x \leq 0 \end{cases}$$

$$\text{Logistic:} \qquad \sigma(x) = \frac{1}{1 + e^{-x}}$$

where, in the leaky rectified linear unit and exponential linear unit activation functions α is a parameter in $(0, 1)$. When $\alpha = 0$ we retrieve the standard rectified linear unit activation function which itself is a very common choice.

The above can be treated more generally, and rigorously, by writing the output of the jth node in the ith layer as follows:

$$a_j^i = \sigma \left(\sum_k (w_{jk}^i a_k^{i-1}) + b_j^i \right) \tag{5.10}$$

where,

- a_j^i is the output (also known as activation value) of the jth node in the ith layer,

- σ is the chosen activation function for that layer,

- w_{jk}^i is the weight connecting the kth node in the $(i-1)^{\text{th}}$ layer to the jth node in the ith layer,

- a_k^{i-1} is the output of the kth node in the $(i-1)^{\text{th}}$ layer,

- b_j^i is the bias value of the jth node in the ith layer.

The output of the network is just the repeated application of Eqn 5.10 from the input layer to the output layer.

Now, when the network returns its output it can be compared against the class label corresponding to the given input and from this the error, also called the cost or loss, can be calculated. There are many functions used to calculate error, each with their own benefits and implicit assumptions regarding the nature of the training data. A common example is Mean Absolute Error (MAE) defined as:

$$\mathcal{C}(y_i, \hat{y}_i) = \frac{1}{N} \sum_i^N |y_i - \hat{y}_i| \tag{5.11}$$

where $\mathcal{C}$ denotes the error, y_i is the true label of the ith input, $\hat{y}_i$ is the predicted label of the ith input and N is the number of samples used in the calculation (generally the batch size). This very simple function attempts to simply bring the absolute value of the predicted labels as close to their true values as possible. Note that it punishes "wrongness" linearly, i.e. being twice as wrong is twice as bad. Another common error function is Mean Squared Error (MSE) which takes the form:

$$\mathcal{C}(y_i, \hat{y}_i) = \frac{1}{2N} \sum_i^N (y_i - \hat{y}_i)^2 \tag{5.12}$$

This error function is useful when one would like to punish wrongness non-linearly, for example when being twice as wrong is much worse than being twice as bad. It assumes a Gaussian distribution of the underlying training data, an assumption that leads to decent results in a lot of cases.

The final example here is different in that it is useful for the very common classification task of binary classification, for example predicting whether an event is signal or background. While the above error functions can be used on binary classification tasks the performance will likely be poor for two reasons. Firstly, the underlying data is Bernoulli distributed (or in multi-class discrete classification at least discretely distributed) and not Gaussian distributed and secondly, MAE and MSE are non-convex for a binary classification problem as the range of the classification labels is $(0, 1)$. This means that there may not exist a minimum in the range $(0, 1)$ and using MAE or MSE may cause convergence to a false minimum or no convergence at all. In binary classification the common choice of error function is Binary Cross Entropy, also known as

Log Loss and it takes the form:

$$\mathcal{C}(y_i, \hat{y}_i) = -\frac{1}{N} \sum_i^N y_i \log(1 - \hat{y}_i) + \hat{y}_i \log(1 - y_i) \tag{5.13}$$

In binary classification problems y_i will either be 0 or 1 depending on its class, while $\hat{y}_i$ belongs to $(0, 1)$ and can be thought of as a probability of belonging to each class.

Once an error function is chosen and the network is applied to inputs the error can then be used to modify the weights of the network systematically. This is done via "backpropagation" of the errors, where the gradient of the error function is calculated with respect to each of the weights of the network one layer at a time backwards from the last layer to the first. This gradient can then be used in the gradient descent methodology to optimally update the weights of the network to ensure the error is smaller on the next iteration of training.

All of the above leads to a model that can create arbitrarily complex sets of combinations of the input values which can model any non-linear function that maps the input to the output.

To relate this to our Higgs example we could have the two input nodes take the transverse momentum of the two leading photons and the single output node return a value between zero and one where zero indicates a signal event and one indicates a background event. The exact architecture of the hidden layers is not entirely important as long as it is sufficiently complex (which in this example is not very complex at all). We would then be able to plot distributions of this output variable for the signal and background and choose a threshold in the range $(0, 1)$ at which we define events with output above that threshold as signal and events below that threshold as background, exactly like seen in the BDT example earlier.

Conclusion

The analysis methods and tools described in this chapter are absolutely indispensable to a particle physics phenomenologist undertaking studies of BSM physics. In chapters 6, 7 and 8 we will see many of the techniques described here applied in published works. Armed with the knowledge of this chapter hopefully these papers will be much more digestible and their results convincing.

Statement of Authorship

Title of Paper	Triple top signal as a probe of charged Higgs in a 2HDM

Publication Status	☒ Published	☐ Accepted for Publication
	☐ Submitted for Publication	☐ Unpublished and Unsubmitted work written in manuscript style

Publication Details	Published in Physics Letters B on 10 May 2018.

Principal Author

Name of Principal Author (Candidate)	Riley Patrick
Contribution to the Paper	• Simulation of signals • Cut Flow analysis • Preparation and editing of the paper, especially the analysis and results sections.
Overall percentage (%)	50%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a this thesis. I am the primary author of this
Signature	Date	09/04/2021

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i.   the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Dr. Pankaj Sharma
Contribution to the Paper	• Conceptualization of project. • Preparation and editing of paper especially the analysis and results sections.
Signature	ate	27/4/21

Name of Co-Author	Prof. Anthony Williams
Contribution to the Paper	• Conceptualization and direction of project. • Editing and proof-reading of manuscript.
Signature	Date	27/4/21

Please cut and paste additional co-author panels here as required.

Chapter 6

Publication 1: Triple top signal as a probe of charged Higgs in a 2HDM

Abstract

Within the framework of the type-II Two Higgs Doublet Model (2HDM-II) we study the production of three top quarks at the Large Hadron Collider (LHC). In the Standard Model the production cross section of three tops is low ($\approx$ 3fb), while it is expected to be significant in the 2HDM-II for reasonable choices of the parameters. We study the production of a charged Higgs in association with a top quark, followed by the decays $H^{\pm} \rightarrow W^{\pm}A$ and $A \rightarrow t\bar{t}$. We undertake a full detector simulation of the signal, and use simple conservative cuts, focussing on the final states that contain three or more leptons, and exactly one same sign di-lepton pair. Finally, we present the exclusion bounds dependent on charged Higgs and pseudoscalar Higgs masses expected in the near future at the 14TeV LHC.

6.1   Introduction

The last missing piece in the standard model (SM) puzzle, the Higgs boson, has been discovered at the Large Hadron Collider (LHC) in its first run. Thereafter various production and decay channels have been studied extensively in order to determine the couplings of the newly discovered boson to various SM particles and the measurements have been found to be close to the SM predictions. Despite this, there is a enough motivation to extend the Higgs sector of the SM in order to understand the underlying mechanism of electroweak symmetry

breaking (EWSB). Among several Beyond the SM (BSM) models which include an extended Higgs sector, the two Higgs doublet model (2HDM) is one of the simplest and extensively explored. The particle content of the model is enriched by additional scalars, i.e., two CP even Higgs (h, H), a CP odd Higgs (A) and a pair of singly charged Higgs ($H^\pm$). Reviews of the phenomenology of the 2HDM and charged Higgs can be found in ref. [66] and [67].

The dominant production of a heavy charged Higgs ($M_{H^\pm} > M_{\text{top}}$) at the LHC is in association with a single top quark occurring via $bg \to tH^-$ + c.c. fusion process [68]. Charged Higgs decay via bosonic mode i.e., $H^\pm \to W^\pm X$ (where $X \equiv h, H, A$) has received significant attention recently in refs. [69–74]. When a heavy neutral Higgs (H or A) decays to a top pair, the final state of the process contains a triple top signal and thus would be a unique and interesting probe of charged Higgs at the LHC.

In the SM, the dominant mode of top quark production at the LHC is pair-production, with cross section ~ 1000 pb at next-to-leading order (NLO), followed by the single-top quark production with total cross section of 250 pb with the t-channel mode having the largest cross section of 150 pb at the NLO.

In addition to single and pair production of top quarks at the LHC, there can also be multi-top quark production in the SM as well, such as three ($3t$) and four top quarks ($4t$). In the SM, the production of an even number of top quarks always occurs via the gg initial state with strong coupling. On the other hand, production of an odd number of top quarks always involves an EW $W^\pm tb$ coupling and often a b quark in the initial state. Thus the cross section for single and three top production is always suppressed with respect to the production of even number of top quarks in the SM. At the LHC with $\sqrt{s} = 14$ TeV, the leading order (LO) total cross section for $3t$ production is approximately 1.9 fb while for $4t$ [75] it is 11 fb, which is almost 6 times the former. In the SM, the $3t$ production occurs via three distinct channels at LO: (a) $pp \to 3t + W^\pm$ at $\mathcal{O}(\alpha_S^4)$; (b) $pp \to 3t + b$ at $\mathcal{O}(\alpha_S^2 \alpha_{\text{EW}}^2)$; and (c) $pp \to 3t +$ jets at $\mathcal{O}(\alpha_{\text{EW}}^4)$. Thus $3t + W^\pm$ has the largest cross section of all $3t$ production modes.

New physics effects may notably enhance the cross section for $3t$ production over the SM. Thus it could be a sensitive probe of BSM physics. There have been some attempts to investigate new physics in $3t$ and $4t$ production [75–79]. In ref. [75], the authors have studied two BSM models, namely the minimal supersymmetric standard model (MSSM) and the leptophobic Z' model. In the former, the pair production of gluinos and subsequent decays to stops lead to $3t$ production after stop decays via $\tilde{t} \to t\tilde{\chi}^0$. In the latter a t-channel exchange of a Z' boson leads to $3t$ production. At a 14 TeV LHC, the production cross

section in MSSM is found to be 41 fb, while for the leptophobic Z' model, it is 28 fb. These numbers are significantly larger than the SM cross section of ~ 2 fb. In the context of the 2HDM, multi-top production including both $3t$ and $4t$ production has been studied at the LHC and for the international linear collider (ILC) in ref. [77]. This study utilizes the subdominant charged Higgs production processes which are associated $H^{\pm}H$ and $H^{\pm}A$ production followed by $H^{\pm} \to tb$ and $A/H \to t\bar{t}$ decays.

The current limits on charged Higgs production can be found in Ref [80][81][82]. In Ref [80] charged Higgs production followed by the decay $H^{\pm} \to \tau +$jets is studied at the 13 TeV LHC, with 14.7fb^{-1} of luminosity. No signal is found in the mass range of 200-2000 GeV and upper limits between 2.0 to 0.008 pb is set. In Ref [81] the production of a charged Higgs in association with a top and bottom quark, followed by the decay $H^{\pm} \to tb$ is studied at the 13 TeV LHC with 13.2 fb^{-1} of luminosity. The mass range of 300 to 1000 GeV with multi-jet final states accompanied by one lepton (not tau). No significant excess is found and upper limits on the cross section times branching ratio is set between 1.09 (1.45) pb for $M_{H^{\pm}} = 300$ GeV to 0.18 (0.17) pb at $M_{H^{\pm}} = 1000$ GeV. Finally in Ref [82] the production of a charged Higgs in association with a single top quark and decaying via $H^{\pm} \to \tau\nu$ is studied at the 13 TeV LHC with 3.2 fb^{-1} of luminosity. No significant excess was found and upper limits on the cross section times branching ratio is set between 1.9 pb and 15 fb for charged Higgs masses between 200 and 2000 GeV.

In this letter, we will focus primarily on the triple top production $pp \to 3t + X$, facilitated by a charged Higgs and a pseudoscalar in a two Higgs doublet model (2HDM). We will demonstrate that $3t$ production can be an alternative probe of the charged Higgs at the LHC, especially for the scenario where both the charged Higgs and pseudoscalar are much heavier than top quarks. In our analysis, we make use of the dominant production mode of a heavy charged Higgs at the LHC, $pp \to tH^-$ followed by the decay of charged Higgs via the bosonic mode, $H^{\pm} \to W^{\pm}A$ and pseudoscalar Higgs via the $A \to t\bar{t}$ mode. Thus it leads to three top quarks in the final state. We perform a realistic simulation of the triple top signal, including detector effects, and apply a set of kinematical cuts to suppress the backgrounds. We present exclusion/discovery bounds after including the effects of all the irreducible and reducible backgrounds in the plane of charged Higgs mass and pseudoscalar mass for a 14 TeV LHC with 30 fb^{-1} of integrated luminosity.

The plan of the paper is following. The next section discusses the production cross section and decay branching ratio of the charged Higgs and triple top signal. Section 3 discusses the different signals and their corresponding backgrounds.

In section 4, we present our results after a full simulation and analysis of the events. Finally, we conclude and summarize in section 5.

6.2   Production and Decay

The process considered in this analysis is a charged Higgs production in association with a single top quark, the leading order Feynman diagrams can be seen in Fig. 6.1. Following the production we consider the decay $H^{\pm} \to W^{\pm} A$ and
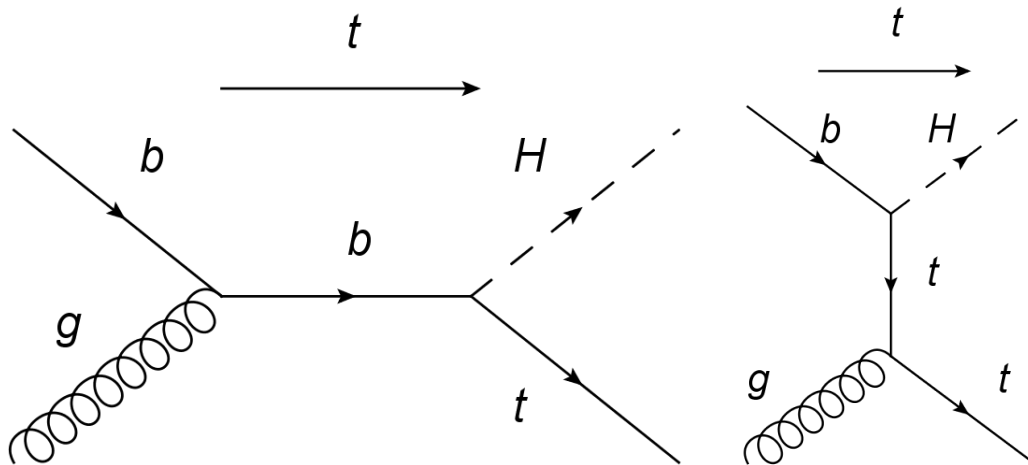
Fig. 6.1   The leading order production mechanisms for $pp \to tH^{\pm}$

$A \to t\bar{t}$ and thus leading to three top quarks in the final state. In Fig. 6.2 we present the production cross section of the triple top signal in type-II 2HDM obtained by multiplying the cross section of the process $pp \to tH^-$ with the branching ratios of the decays described above i.e., $H^{\pm} \to W^{\pm} A$ and $A \to t\bar{t}$ in the plane of the charged Higgs mass and pseudoscalar mass. We use the Two Higgs Doublet Model Calculator (2HDMC [83]) to obtain the corresponding branching ratios for each point in the parameter space. In order to evaluate the cross section and BRs, we consider $\tan \beta = 1$ and $\sin(\beta - \alpha) = 1$ throughout the analysis.

Of all 2HDM Yukawa types (see [66] for a review), we concentrate here on the type II 2HDM (2HDM-II). In the 2HDM Type II, constraints from $b \to s\gamma$ decays put a lower limit on the $H^{\pm}$ mass at about 580 GeV, rather independently of $\tan \beta$ [84, 85]. The additional neutral and charged scalars contribute to the gauge boson vacuum polarisation through their coupling to gauge bosons. As a result, the updated EW precision data provide important constraints on new physic models. In particular, the universal parameters S, T and U provides constraint on the mass splitting between the heavy states m_H, $m_{H^{\pm}}$ and m_A in the scenario in which h is identified with the SM-like Higgs
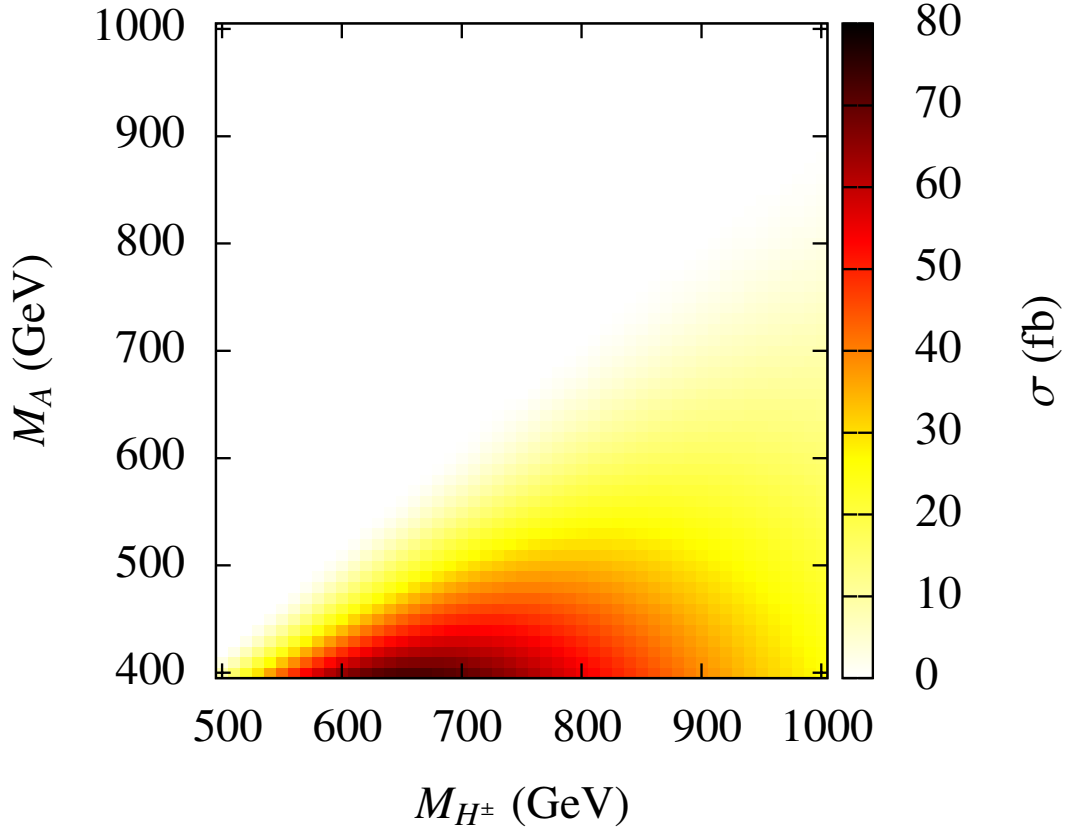
Fig. 6.2 Cross section for triple top signal ($t\bar{t}tW^{\pm}$) obtained by evaluating $\sigma(pp \to tH^{\pm}) \times \mathrm{Br}(H^{\pm} \to WA) \times \mathrm{Br}(A \to t\bar{t})$ in the plane of charged Higgs mass and pseudo scalar mass.

state. However, as pointed in ref. [86], it is clear that if either of neutral Higgs (CP even or CP odd Higgs) is degenerate with a charged Higgs in a 2HDM model, then the constraints on the mass splitting between the charged Higgs and other heavy neutral Higgs become quite relaxed. So far as the mass splitting between the $H^{\pm}$ and A is concerned for triple top production, it should be greater than 80 GeV so as to achieve a large BR to the $W^{\pm}A$ mode. Thus, the charged Higgs mass $M_{H^{\pm}}$ and pseudoscalar mass M_A are varied in the ranges (500 GeV - 1 TeV) and (400 GeV - 1 TeV) respectively. Also, in order to open the $A \to t\bar{t}$ decay channel, the pseudoscalar mass should be larger than $2M_{\mathrm{top}} \sim 350$ GeV. For such a pseudoscalar the dominant decay mode would be to a top pair as the coupling is proportional to mass of the top quark. The other decay $A \to Zh$ is suppressed by $\sin(\beta - \alpha)$ since the current LHC scenario prefers alignment scenario with $\sin(\beta - \alpha) = 1$. Therefore in the scenario where both $M_{H^{\pm}}$ and M_A are heavier than $2M_{\mathrm{top}}$, the triple-top production is the
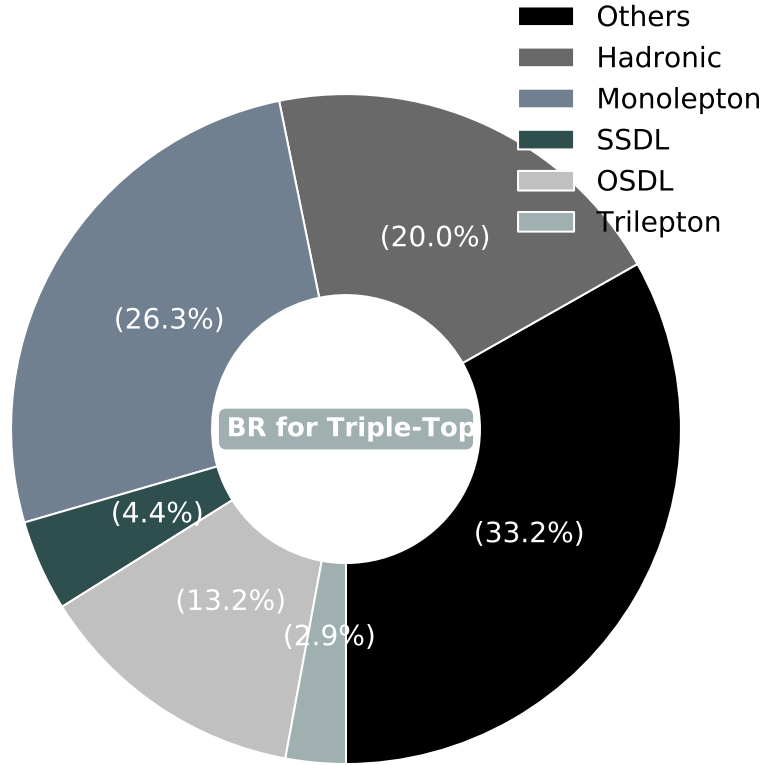
Fig. 6.3 The branching ratio associated with each possible final state for $t\bar{t}t + W^-$ and $t\bar{t}\bar{t} + W^+$ decays.

only possible signal to probe a charged Higgs at the LHC. Similar benchmark scenarios have been studied in refs. [87, 88].

We see from Fig. 6.2 that large cross section for triple top production in a 2HDM is obtained when $M_{H^\pm}$ is in the range (550 GeV - 750 GeV) and M_A in the range (400 GeV - 500 GeV), where the cross section is found to be in the range (50 fb - 80 fb). This is significantly larger than the SM cross section of 2 fb for triple top quarks at the 14 TeV LHC. Thus it is expected that the search for the triple top signal in the current and future runs at the LHC would significantly enhance the search prospects for a charged Higgs. If no such signal is found, it would enable a stringent bound to be set in the $M_{H^\pm} - M_A$ plane in 2HDM.

6.3   Signal and SM Backgrounds

6.3.1   Signal

The triple-top event $tt\bar{t}W^{\pm}$ decays into a large number of final state particles, $\sim \mathcal{O}(10)$. Various decay modes and their corresponding branching ratios have been presented in Fig.6.3. Fully hadronic decays of the triple top event which has 20 % of BR leads to a very high number of jets in final state, i.e., 8 light jets and 3 b jets. The monolepton signal includes one lepton (either $e^{\pm}$ or $\mu^{\pm}$, $\tau^{\pm}$ is not included here) associated with 9 jets and has fairly large BR of 26.3%. The dilepton signal can be classified into opposite sign dilepton (OSDL), which has 14.2% BR, and same-sign dilepton (SSDL), which has quite small 4.4% of BR. Finally the trilepton signal has only a 2.9% of BR.

In this analysis we focus on the multileptonic signals, namely SSDL and trilepton, which despite having low branching ratios are cleaner at the LHC and have backgrounds which are more efficiently manageable. The SSDL signal is accompanied by 7 additional jets while the trilepton signal arises along with 5 jets. In both cases, three of the jets are b jets.

6.3.2   Backgrounds

The irreducible background to the final state being considered is the SM production of three tops in association with a $W^{\pm}$ boson, which has a total production cross section of 1.37 fb. However, it is reasonable to expect any three top signal to behave as a background to the signal under the right circumstances (IS/FS radiation, jet mis-tagging etc.). However, the combined cross section of all three top production as the LHC is only 1.9 fb.

As well as this in the circumstance that a b-jet from the final state is lost from the detector, a four top process will successfully mimic our signal. This is concerning given the cross section of four top production at the LHC is much higher at 11 fb, though cuts on pseudo-rapidity and jet multiplicity should almost certainly remove most of these processes.

Other backgrounds come from various top pair production associated with one heavy SM particle and light jets of which at least one must be b jets, for instance, $t\bar{t}W^{\pm}$, $t\bar{t}h$ and $t\bar{t}Z$ processes. Processes such as $t\bar{t}\,b\,nj$ may contribute to both trilepton and SSDL signals when one of the light jets are faked as a lepton. $Q-$flip backgrounds occur when a real OSDL signal arising from some underlying process is misidentified as SSDL pair at the detector.

6.4 Signal and Background Efficiency

The accumulative efficiency associated with detector effects, jet finding/tagging and kinematic cuts is important to get a full understanding of signal significance. To achieve this we generate a million events using `Madgraph` [47], which are passed to `Pythia` [48] for parton showers and hadronization and finally to `Delphes` [53] for realistic detector effects. For detector acceptance, we apply the following cuts: (i) all leptons must have transverse momentum p_T larger than 20 GeV and be within pseudorapidity $|\eta| < 2.5$; (ii) all the jets must have $p_T > 25$ GeV and $|\eta| < 2.5$; (iii) all the objects must be well separated from with each other with $\Delta R_{ij} > 0.4$ where $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$. All the jets are clustered using an anti-k_T jet clustering algorithm with jet radius $\Delta R = 0.4$.

We then choose two final states to study, the trilepton and SSDL final states. In ref.[76], authors have studied four-top production and analysed the event in SSDL and trilepton signals at the 14 TeV LHC. The corresponding backgrounds to these signals are also modelled in great detail. We adopt the search strategy for selection of signal and background events presented in their analysis. In the tri-lepton case we demand that the event contains more than 5 light jets ($n_{\text{jet}} > 5$) and three or more b-tagged jets ($n_{\text{bjet}} \geq 3$). We also require that all same flavour opposite sign leptons invariant mass does not fall within 25 GeV of the Z-boson mass which serves to significantly cut into the SM background. In the SSDL case we use $n_{\text{jet}} > 6$ and $n_{\text{bjet}} \geq 3$.

The final accumulative signal efficiency for the trilepton and SSDL final states after the above selection requirements, and applying a 75% b-tagging efficiency and a 1% mis-tagging efficiency where necessary, are 49.9% and 77.2% respectively. The difference in these efficiencies can be attributed to the extra Z-boson mass window cut placed on the trilepton state.

As mentioned above, we adopt the same set of cuts employed in ref. [76] on the signal events. Thus in this analysis, we use the same efficiencies for the backgrounds presented in their study. The total background cross section for the trilepton signal is 60.67 fb, while for the SSDL signal it is 122.73 fb. As well as this the efficiencies associated with each background can be seen in Table 6.1 and Table 6.2.

Using the efficiencies for signals and background obtained above we now estimate the signal significance which is defined by the ratio $S/\sqrt{S+B}$ where S and B are the number of signal and background events respectively. We present in the Fig. 6.4 the signal significance using trilepton (top) and SSDL (bottom) signals in the plane of mass of the charged Higgs and mass of the

Backgrounds	Cross Section (fb)	Efficiency
ttW	31.34	6.46×10^{-3}
ttZ	48.47	2.21×10^{-3}
tth	7.25	2.53×10^{-3}
Fakes	16.57	3.48×10^{-3}
Q-flip	11.37	1.01×10^{-2}
Other	7.73	1.90×10^{-3}

Table 6.1 The background cross sections and corresponding efficiencies for the SSDL signal at the 14 TeV LHC.

Backgrounds	Cross Section (fb)	Efficiency
ttW	1.65	9.90×10^{-3}
ttZ	48.47	5.50×10^{-4}
tth	2.4	2.50×10^{-3}
Fakes	1.13	5.01×10^{-3}
Other	7.02	4.70×10^{-4}

Table 6.2 The background cross sections and corresponding efficiencies for the trilepton signal at the 14 TeV LHC.

pseudoscalar for the 14 TeV LHC with 30 fb^{-1} of integrated luminosity. As expected, we find the SSDL signals to be more constraining than the trilepton signal due to the larger cross section and smaller backgrounds as compared to the latter. We conclude that with the early data to be collected in the 14 TeV LHC, the SSDL and trilepton signal in the triple-top production can exclude the charged Higgs upto 1 TeV if the mass splitting between $H^{\pm}$ and the pseudoscalar is within the range (100 GeV - 300 GeV).
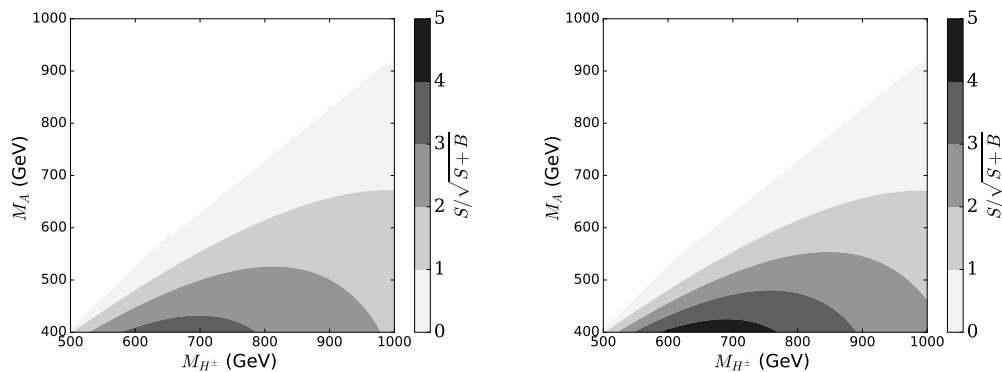
Fig. 6.4 Signal significance for charged Higgs in the trilepton (top) and SSDL signal (bottom) with 30 fb^{-1} of integrated luminosity at the 14 TeV LHC.

6.5 Conclusion

In this letter we have assessed the discovery/exclusion potential of a heavy charged Higgs boson in the type-II Two Higgs Doublet Model (2HDM) at the large hadron collider (LHC). We explored the case when a charged Higgs with mass ≥ 450 GeV is produced in association with a top quark, followed by the decay $H^{\pm} \to AW^{\pm}$ where A is the pseudo-scalar Higgs. This pseudoscalar Higgs was chosen to be ≥ 350 GeV so as to allow its decay to two top quarks, resulting in the state $t\bar{t}\bar{t}W^{+}$ or $tt\bar{t}W^{-}$. This is an exotic state that is extremely rare in the standard model (SM) (triple-top production in the SM has a cross section of less than 3 fb) and as such provides an interesting window into the search for charged Higgs in the 2HDM.

We further focused on the final state that included trilepton and also the final state with a same sign dilepton (SSDL) pair. The discovery/exclusion potential for these signals has been presented in Fig 6.4. It can be seen that with current LHC data, with approximately 30 fb of luminosity, we should be able to make strong statements regarding the existence of these states in almost all parts of the charged Higgs/pseudo-scalar mass plane for our chosen parameters.

Further directions for this work revolve predominantly around cut optimization. The above analysis used extremely conservative cuts and achieved great exclusion potential but the use of more advanced techniques such as Boosted Decision Tree or Artificial Neural Nets as seen in Refs. [71] [72] should significantly improve the exclusion potential but would require the full modelling of the SM backgrounds at the detector level. This was not undertaken in this study.

In addition, the remainder of the final states - opposite sign dilepton pair, monolepton and fully hadronic - could provide some level of exclusion potential but these are expected to be lower than the channels presented in this study. This is because while the branching ratios of these final states are higher, the SM backgrounds are much higher and may drown out the signal.

Finally, a study of the interference effects between the signal and irreducible SM backgrounds would be of great interest. It is expected that the interference terms of the full scattering amplitude would produce a peak dip structure in the distributions of the variables of the events. This may be exploited to further improve the discovery/exclusion potential or it may obscure the signal.

Acknowledgements

Statement of Authorship

Title of Paper	Top polarisation as a probe of CP-mixing top-Higgs coupling in tjh signals
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Published in Physical Review D on 20 May 2020.

Principal Author

Name of Principal Author (Candidate)	Riley Patrick
Contribution to the Paper	• Full simulation of signals and backgrounds. • Machine learning analysis including custom codes. • Preparation and editing of manuscript.
Overall percentage (%)	60%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a this thesis. I am the primary author of this
Signature	Date 09/04/2021

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i.   the candidate's stated contribution to the publication is accurate (as detailed above);

ii.   permission is granted for the candidate in include the publication in the thesis; and

iii.   the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Andre Scaffidi
Contribution to the Paper	• Production of custom FeynRules module for signal and background. • Preparation and editing of manuscript.
Signature	Date 09/04/2021

Name of Co-Author	Pankaj Sharma
Contribution to the Paper	• Conceptualization and direction of project. • Editing and proof-reading of manuscript.
Signature	Date 09/04/2021

Please cut and paste additional co-author panels here as required.

Chapter 7

Publication 2: Top polarisation as a probe of CP-mixing top-Higgs coupling in tjh signals

Abstract

In this letter we explore beyond the Standard Model top-Higgs Yukawa couplings as a function of a CP-mixing parameter ξ_t at the 14 TeV HL-LHC in the process $pp \to thj$. We observe that angular variables of the decay products of the top are non-trivially sensitive to ξ_t. This fact is exploited in a full detector level analysis that employs machine learning techniques to optimize signal sensitivity on a suite of variables, including lepton azimuthal angle. The key result of this study is an improved projected exclusion limit on ξ_t even when including the realistic effects of detector smearing and a conservative estimate of systematic error.

7.1   Introduction

Since the discovery of a scalar particle at 125 GeV as predicted by the Standard Model (SM) at the Large Hadron Collider (LHC) in 2012 [2, 1] attention has turned to narrowing down it's properties. It is of great interest as to whether this particle behaves exactly as the SM predicts or if it is perhaps a beyond the SM (BSM) scalar that exists within an expanded Higgs sector.

The measurement of the Higgs couplings to fermions is an on-going area of research as many of these couplings can be small and require large luminosities to probe, which motivates the application of intelligent analysis techniques to probe the signal as optimally as possible. The Higgs coupling to the top is of

particular interest as it is the largest, and so far the $t\bar{t}h$ coupling is probed via the loop-induced processes hgg and $h\gamma\gamma$ which rely on decay rate measurements of the Higgs. It is well known that the production cross section of both hgg and $h\gamma\gamma$ will be sensitive to a phase ξ_t of the top Yukawa coupling [89]. This phase mixes the CP-properties of the top-Higgs coupling and thus is CP violating. Assuming only the CP phase and keeping SM value for magnitude, strong constraints could be placed on the phase. However, those processes can only occur at one-loop level and beyond. Noting that new physics dynamics can contribute in the loop and thus may affect the accurate determination of top-Yukawa phase, the Higgs production in association with a top pair (dominant) and Higgs produced with a single top (sub-dominant) are the only means to directly probe this coupling. These processes have been established at above 5σ in Refs [90, 91]. It is found that increasing $|\xi_t|$ leads to a suppression of the $t\bar{t}h$ cross section and an enhancement of the tjh cross section [92, 93].

A study of this process has also been undertaken utilizing the Matrix-Element-Method (MEM) which shows that, with signal detection efficiencies on the order of a few percent, discovery could be made in the high luminosity phase of the LHC [94]. Also it has been shown in Ref [95] using matrix element techniques that 3σ seperation of the psuedo-scalar vs scalar tth coupling can be realized with 300fb^{-1} of data.

Further phenomenological studies of the top-Higgs coupling can be found in Refs [96–101].

The top quarks lifetime is so short that it decays before hadronization, meaning that its polarization information is preserved in the distribution of its decay products which can be measured directly by the detector, especially in its lepton angular distributions. In many new physics scenarios, it has been shown extensively in literature, Refs. [102–111], that charged-lepton azimuthal distribution is a powerful probe of top quark polarisation in the lab frame. There are two advantages of studying the charged lepton azimuthal distribution: first it does not require reconstruction of top-rest frame which would need full information of top-quark momentum and second, it is unaffected by any new physics in the top-quark decay and thus making it an uncontaminated probe of top quark polarization. This variable is constructed by taking the azimuthal angle of the lepton decaying from the top with respect to the x-z plane, where the top quarks x-component is positive. For other methodologies of studying top polarisation see Refs [112, 113, 99].

Top quark polarisation can be written in terms of ξ_t [114, 115] and thus the decay products differential distributions will be effected by ξ_t allowing for the improvement of analysis of this process. This has been exploited in a range of

studies [114–116], however these have not been undertaken at detector level to provide more accurate reflections of achievable sensitivity. Furthermore, these studies have employed traditional cutflow methods rather than more modern and advanced machine learning (ML) techniques to optimize signal sensitivity. In this letter we employ a full detector level analysis to calculate the angular variables of the decay products of the process $pp \rightarrow thj$ with the optimization of signal sensitivity through ML algorithms.

This paper is structured as follows: section 7.2 will cover the parametrization of the top-Higgs coupling and it's implementation, section 7.3 will outline signal and backgrounds, section 7.4 will outline event simulation and reconstruction, section 11.6 will present the results of this analysis and finally we will conclude in section 11.7.

7.2   CP-mixed Top-Higgs Coupling

In this study, a CP-mixing parameter ξ_t is introduced in the mass basis of the top-Higgs sector via the Lagrangian

$$\mathcal{L}_{tth} = -\frac{y_t}{\sqrt{2}} \bar{t} \left(\cos \xi_t + i\gamma_5 \sin \xi_t \right) th \; . \tag{7.1}$$

The SM limit corresponds to when the mixing angle $\xi_t = 0$ and the Yukawa coupling adopts its SM value $y_t \rightarrow y_t^{SM} = \sqrt{2}m_t/v$, where m_t is the mass of the top and $v \simeq 246$ GeV is the standard model Higgs vacuum expectation. We adopt a model independent approach in where the interaction Lagrangian in Eqn. 7.1 arises from an effective field theory (EFT) such as the dimention-6 opperators discussed in Refs. [117–121]. We assume that the new physics scale Λ of such an EFT is $\gtrsim 1$ TeV with couplings of $\mathcal{O}(1)$, such that the mixing angle $\xi_t \in (-\pi, \pi]$ [122].

Constraints on y_t and ξ_t from the hgg and $h\gamma\gamma$ loop processes can be found in Refs [114, 93, 123–127]. Further constraints including unitary violation in W and Z scattering with the top have been defined in Refs [128, 129]. The strongest constraints come from precision electron dipole-moment (EDM) measurements [130–132], however these are done under assumptions which when relaxed allow for much looser constraints. Assuming a standard model value for y_t, collider constraints have rendered $\xi_t \in [0, 2\pi/3]$ at 2σ [133]. For this study we also assume that y_t and WWh coupling adopt standard model values. Furthermore, for the sake of completeness, we perform the study on the entire region $\xi_t \in [0, \pi]$.

7.3   Signal and Background

The process studied is Higgs production with an associated top quark and jet, $pp \to tjh$, at the 14 TeV LHC. Fig 7.1 displays the dominant Feynman diagrams contributing to signal production.
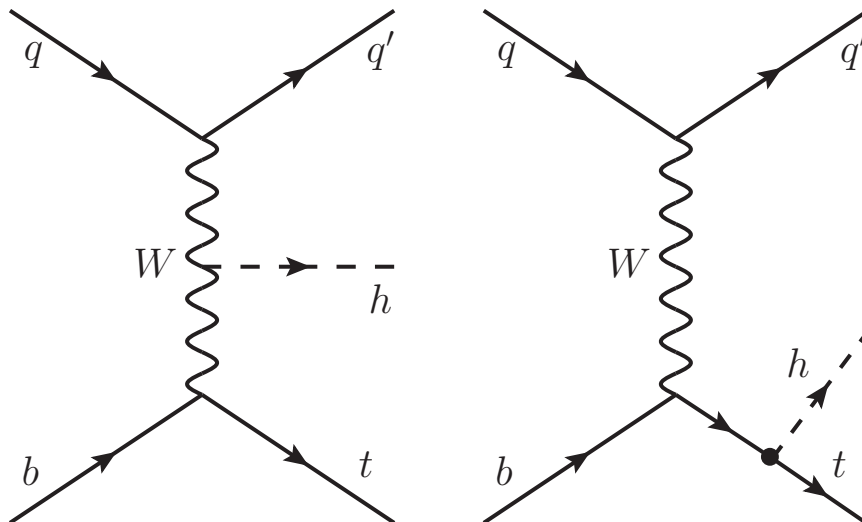
Fig. 7.1 Feynman diagrams for the dominant production process $bq \to thj$.

Due to the extremely clean signature it provides, the decay mode of $h \to \gamma\gamma$ has comparable signal sensitivity to the $h \to b\bar{b}$ decay despite a much smaller branching ratio. We hence choose this decay mode of the Higgs for our analysis. In Fig 7.2 one can see the effect of the CP-mixing parameter ξ_t on the production cross section of tjh. This effect is in agreement with previous results seen in Ref [116] and contains a maximum enhancement at $\xi_t = \pi$ of approximately 1200%.

We demand a final state containing 1 b-jet, exactly 1 lepton (but not τ) and at least 2 photons. The irreducible background for this process is $tj\gamma\gamma$ continuum. The sub-leading background is $t\bar{t}\gamma\gamma$ where a b-jet is lost, i.e. mistagged as a light jet or produced outside acceptance. The background $Wjj\gamma\gamma$ where a light jet is mistagged as a b-jet exists, however it is found to be at least an order of magnitude lower than the previous two backgrounds [134, 92] and is ignored in this study. Finally the backgrounds $t\bar{t}\gamma j$ and $t\gamma jj$ backgrounds are neglected in this study as they would require the mis-identification of a light jet as a photon which suppresses the already small cross section.
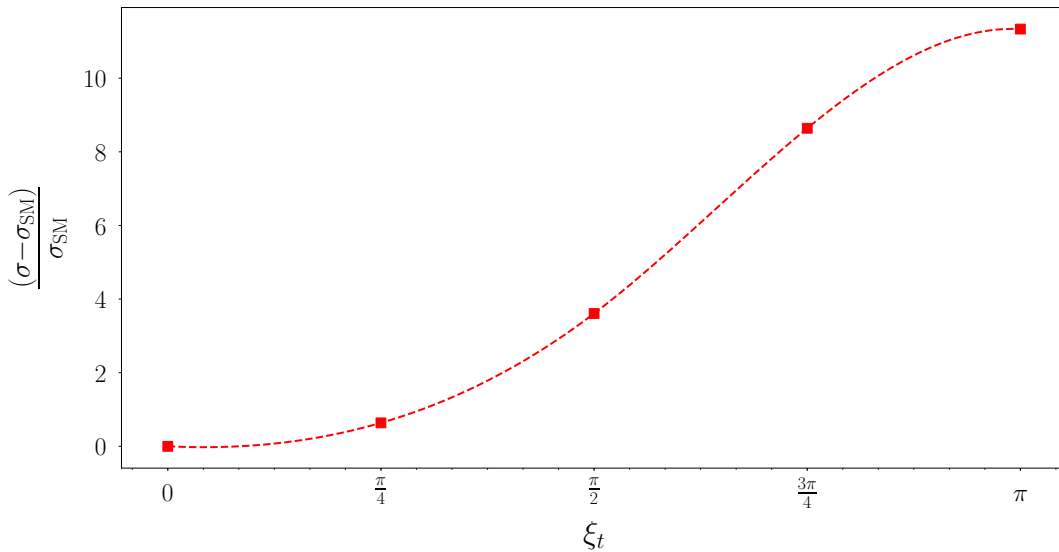
Fig. 7.2 The production cross section (red) and production cross section times $h \to \gamma\gamma$ branching ratio (blue) normalized with respect to standard model values for $\xi_t \in [0, \pi]$.

7.4 Event Simulation and Reconstruction

The parton level events are produced in `MG5_aMC_v2_6_0` [135] then passed to `PYTHIA8` [48] for hadronization/fragmentation and finally `Delphes` [53] for detector effects. Each sample consists of $100,000$ events, achieving an MC uncertainty on inclusive cross section values that is on the order of 0.1%.

We employ anti-kt jet clustering and take a b-tag efficiency of 77%, a mistagging efficiency of 1% and a lepton selection efficiency of 100%. We also employ the following detector acceptance cuts:

$$p_T^{b,\ell} > 20\text{GeV}, \ |\eta_{b,\ell}| < 2.5, \ p_T^j > 25\text{GeV}, \ |\eta_j| > 2.5 \qquad (7.2)$$

The cut on $|\eta_j|$ is employed to take advantage of the forwardness of the light jet which is characteristic of the tjh signal.

As we are selecting exactly 1 lepton we are able to calculate the longitudinal momentum of the neutrino decaying from the top. We do this using the following quadratic equation:

$$p_\nu^z = \frac{1}{2p_{\ell T}^2} \left(A_W p_\ell^z \pm E_\ell \sqrt{A_W^2 \pm 4p_{\ell T}^2 E_{\nu T}^2} \right), \qquad (7.3)$$

where, $A_W = M_{W^\pm}^2 + 2p_{\ell T} \cdot E_{\nu T}$. We choose the solution for p_ν^z that is real and that when combined with the remainder of the neutrino and lepton 4-vector components produces an invariant mass closest to the W boson mass. After

this is done the top quark is reconstructed from the neutrino, the lepton and the b-jet which best reproduces the top quark invariant mass.

In Fig 7.3 we present the variables selected for the numerical analysis to come. In the ℓ_0^ϕ plot (upper right of each subfigure) the lepton azimuthal distribution generated from the prescription above can be seen for the hardest lepton in each event. It is clear that the value of ξ_t is impacting this distribution significantly as the SM value of $\xi_t = 0$ presents a distribution identical to the background while the fully CP-odd value of $\xi_t = \frac{\pi}{2}$ presents a far more skewed distribution. In addition, the variables show that the reconstruction is faithfully producing invariant mass distributions for the top quark and Higgs.

(a)

(b)
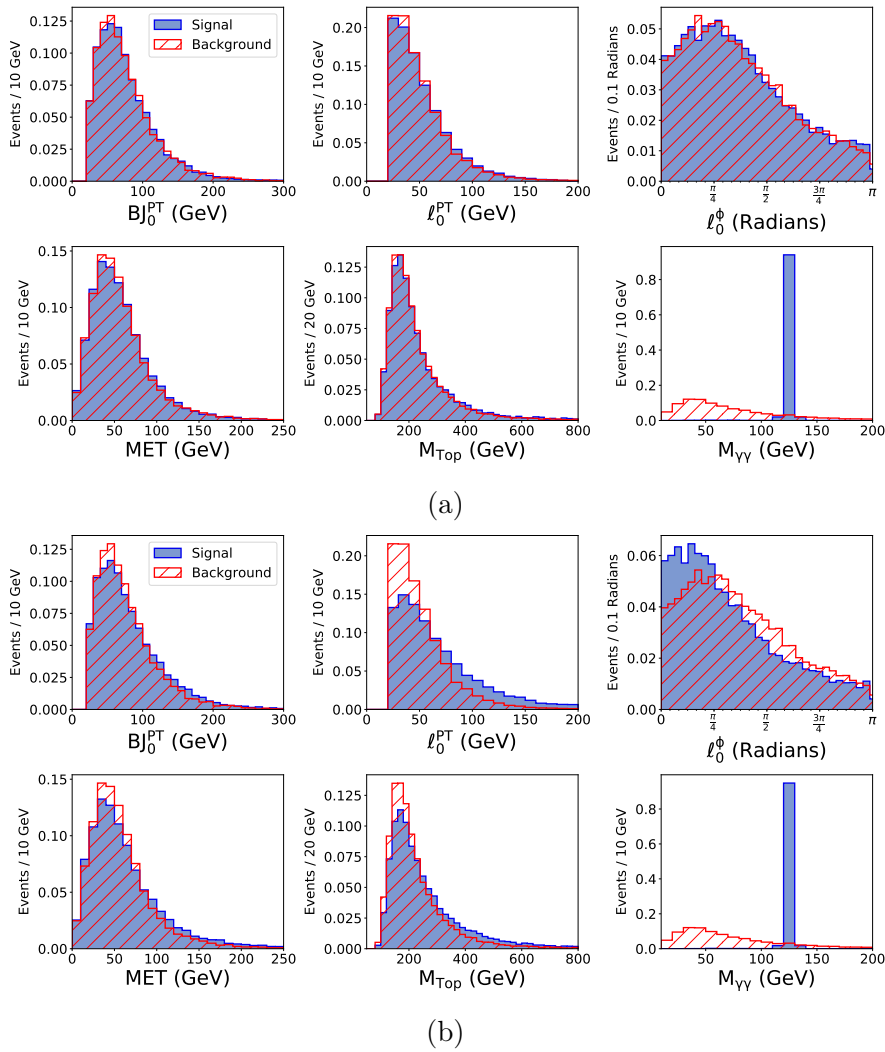
Fig. 7.3 Examples of variables employed in the analysis for the $\xi = 0$ (a) and $\xi = \frac{\pi}{2}$ (b) benchmarks.

7.5   Results

As seen in Ref [116], we can construct the lab frame left-right asymmetry of the charged lepton using:

$$A_\phi^\ell = \frac{\sigma\left(\cos\phi > 0\right) - \sigma\left(\cos\phi < 0\right)}{\sigma\left(\cos\phi > 0\right) + \sigma\left(\cos\phi < 0\right)} \tag{7.4}$$

Fig 7.4 displays this asymmetry as a function of ξ_t which takes a maximimum at $\xi_t = \frac{\pi}{2}$. However when compared to the parton level calculation of this asymmetry found in Ref [116] it can be seen that the detector effects lead to a flattening of this curve.
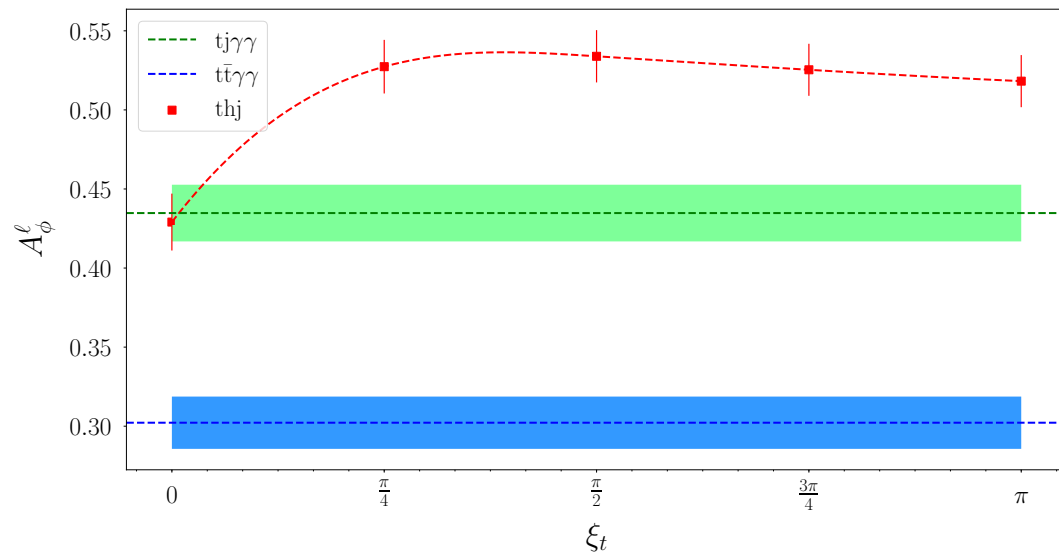
Fig. 7.4 The A_ϕ^ℓ asymmetry as a function of ξ_t for the signal and background.

Standard cut flows for this signal have been performed in the past such as in Refs [115][116] and thus we do not perform one in this work. Instead we employ a boosted decision tree analysis on the variables found in Fig 7.3 using the `Toolkit for Multivariate Data Analysis` (TMVA) [122].

Tab 7.1 presents the BDT results for each value of ξ. The column labels are as follows: N_s^b is the number of signal events before cuts while N_s^a and N_b^a are the number of signal and background events after cuts respectively. The number of background events before cuts was $N_b^b = 1076$. The column labelled "Cut" is the position of the optimal cut on the BDT classifier distribution. σ is the signal sensitivity considering no systematic error, while $\sigma_{0.2}$ is the signal sensitivity with a flat 20% systematic error. The systematic error of 20% was taken as an estimate of the overall level of systematic error in typical 1-lepton plus jets final state experiments [136]. The results presented in this table show,

ξ_t	N_s^b	Cut	N_s^a	N_b^a	Z	$Z_{0.2}$
0	8	0.1146	7	21	1.28	1.04
$\frac{\pi}{4}$	23	0.0910	20	24	3.05	2.44
$\frac{\pi}{2}$	89	0.0819	83	27	7.92	7.04
$\frac{3\pi}{4}$	198	0.0317	191	35	12.70	11.52
π	255	0.0503	244	34	14.64	13.55

Table 7.1

Table of signal sensitivity defined as: $Z = S/\sqrt{S + B + (\Delta B)^2}$ after applying the optimized cut generated by BDT analysis. A luminosity of 3000fb^{-1} is chosen and the number of background events before cuts is $N_b^b = 1076$.

as expected, high values of $\xi_t \geq \frac{\pi}{2}$ are strongly inconsistent with background only. The 95% C.L. exclusion of approximately $\xi_t = 0.54$ without systematic and $\xi_t = 0.68$ with systematic is expected for 3ab^{-1} of data at the HL-LHC, a significant improvement on the $\xi_t = 0.79$ exclusion placed in Ref [116]. We also see an improvement over the HL-LHC results found in Ref [99]. The parametrization in this work allows for complete freedom of the CP-even and CP-odd contributions to the top-Higgs coupling, called κ and $\tilde{\kappa}$ respectively. A detector level analysis using optimized angular variables is performed in the semi-leptonic final state and only the top pair background is considered. A significance value of $S/\sqrt{B} \approx 0.8$ for (it is assumed) the SM value of $\tilde{\kappa}$ utilizing a semi-leptonic final state is found and no meaningful bound on $\tilde{\kappa}$ can be extracted. The improvements are likely due to a combination of the BDT analysis and the additional top polarisation variable included. It is likely that had this analysis been undertaken at parton level like previous analyses of this process then improvements on the constraints would be even larger, that is to say that detector effects have likely reduced the overall positive impact of the BDT analysis and top polarisation variable.

7.6 Conclusion

The direct detection of top-Higgs coupling has now been achieved via the process $pp \rightarrow t\bar{t}h$, however the properties of this coupling still require further study. The process $pp \rightarrow tjh$ provides a good window into the charge-parity properties of the coupling as increasing values of the top-Higgs coupling phase ξ_t lead to increased cross sections, while the $t\bar{t}h$ process experiences decreases.

In this work we have introduced a CP-mixing parameter ξ_t to the SM top-Higgs coupling via an effective operator. We have explored the well studied effects of this variable on top and jet associated Higgs production. Previous

studies were then expanded on by performing a full detector level analysis of this process including the variable defined from the azimuthal distribution of the lepton decaying from the top which provides a powerful insight into top polarisation. Results were then further improved via the application of a ML algorithm, namely boosted decision tree analysis, to optimize signal sensitivity.

The key result of this study is a projected 95% median exclusion of $\xi_t \leq 0.54$ when not considering systematic errors and $\xi_t \leq 0.68$ when considering a conservative level of systematic error with 3ab^{-1} of luminosity, a significant improvement over previous analyses of this process. It is reasonable that the HL-LHC can provide very strong limits on ξ_t even in pessimistic scenarios.

Acknowledgements

Statement of Authorship

Title of Paper	Signal versus Background Interference in H+ → tb̄ Signals for MSSM Benchmark Scenarios
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Published in the Journal of High Energy Physics on 30 October 2020.

Author

Author (Candidate)	Riley Patrick
Contribution to the Paper	- Full simulation of signal and backgrounds. - Cut flow analysis - Preparation of the analysis and results sections. - Editing and proof-reading of full paper.
Overall percentage (%)	40%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis.
Signature	Date 09/04/2021

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i.   the candidate's stated contribution to the publication is accurate (as detailed above);

ii.   permission is granted for the candidate in include the publication in the thesis; and

iii.   the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Abdesslam Arhrib
Contribution to the Paper	- Discussion on the model and it's theoretical constraints. - Write up of the paper.
Signature	Date 19/03/2021
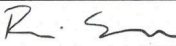
Name of Co-Author	Duarte Azevedo
Contribution to the Paper	My contributions were mainly related with bridging the gap between the MSSM benchmarks chosen and their implementation in a LHC-like toy analysis used as proof of concept. This was in parallel with Riley Patrick's own implementations in order to cross-check results.
Signature	Date 15/03/2021

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Rachid Benbrik
Contribution to the Paper	Permission is granted for the candidate to include the publication in the thesis.
Signature	_R. Benbrik_ Date 23/03/2021

Name of Co-Author	Hicham Harouiz
Contribution to the Paper	• Investigation of theoretical and experimental constraints of the model and production of plots relevant to this. • Assisted in preparation of the manuscript. (Filled in by Riley Patrick and confirmed by primary supervisor.)
Signature	Date 09/04/2021

Name of Co-Author	Stefano Moretti
Contribution to the Paper	I can confirm the above.
Signature	Date 16/03/2021

Name of Co-Author	Rui Santos
Contribution to the Paper	My contribution to the paper was mainly at the level of discussions and part of the writing and reviewing. I would say that I have contributed 5 to 10%.
Signature	_R. S._ Date 16/03/2021

Chapter 8

Publication 3: Signal versus Background Interference in $H^+ \to t\bar{b}$ Signal for MSSM Benchmark Scenarios

Abstract

In this paper, we investigate sizeable interference effects between a heavy charged Higgs boson signal produced dominantly via $gg \to t\bar{b}H^-$ (+ c.c.) followed by the decay $H^- \to b\bar{t}$ (+ c.c.) and the irreducible background given by $pp \to t\bar{t}b\bar{b}$ topologies at the Large Hadron Collider (LHC). We show that it may be possible that such effects could spoil current $H^\pm$ searches where signal and background are normally treated separately. The reason for this is that a heavy charged Higgs boson can have a large total width, in turn enabling such interferences, altogether leading to potentially very significant alterations, both at the inclusive and exclusive level, of the yield induced by the signal alone. This therefore implies that currently established LHC searches for such wide charged Higgs bosons might require modifications. We show such effects quantitatively using two different benchmark configurations of the minimal realisation of Supersymmetry, wherein such $H^\pm$ states naturally exist. However, on the basis of the limited computing resources available, we are unable to *always* bring the statistical error down to a level where all such interference effects are unequivocal, so that we advocate dedicated experimental analyses to confirm this with higher statistics data samples.

8.1 Introduction

After the discovery of a Higgs-like particle at Large Hadron Colliders (LHC) a few years ago [137, 138], a significant amount of both theoretical and experimental activities have taken place trying to identify the nature of this object. The mass of such a particle and its couplings to some Standard Model (SM) particles are now measured with a good precision [139, 140]. Their values indicate that such a Higgs-like particle is light and its properties (spin, CP quantum numbers and interactions) are consistent with those of the SM Higgs boson.

However, there are many theoretical and experimental indications that show that the SM can only be an effective theory of a more fundamental one that still needs to be discovered. Many Beyond the SM (BSM) scenarios have been put forward over the years and it is fair to say that one stems as the most appealing one - Supersymmetry (SUSY). This is because it solves the well-known hierarchy problem of the SM by protecting the Higgs mass from unstable higher order corrections thanks to the new symmetry between fermions and bosons that it predicts [141]. Furthermore, SUSY also has the capability to address the Dark Matter (DM) and gauge unification problems of the SM, indeed, without any proliferation of fundamental parameters if one assumes that SUSY can in turn be viewed as an effective realisation of some Grand Unified Theory (GUT), like Supergravity [142, 143]. The Minimal Supersymmetric Standard Model (MSSM) is the simplest realisation of SUSY that predicts a light Higgs boson h^0 that can be identified as the observed 125 GeV Higgs-like particle and can be as successful as the SM when confronted with experimental data, yet it can surpass it in all the above respects.

The Superpotential of the MSSM has to be holomorphic, thus one needs to introduce at least two Higgs doublets fields, one more than in the SM. One of these generates masses for up quarks and the other one generate masses for down quarks and charged leptons. From the 8 degrees of freedom present in such a 2-Higgs Doublet Model (2HDM), 3 are acquired by the longitudinal components of the gauge bosons $W^\pm$ and Z^0, so that the latter get a mass too, and the remaining 5 appears as new Higgs particles: 2 CP-even h and H (with $M_{h^0} < M_{H^0}$), one CP-odd A and a pair of charged ones $H^\pm$. Discovery of any such new states would be unmistakable evidence of BSM physics, yet, only charged Higgs states would be a clear hint towards a 2HDM structure, as required by the MSSM, as additional neutral Higgs states could be attributed to singlet structures entering an extended Higgs sector.

At a hadron collider, production of charged Higgs bosons proceeds through many channels. If the charged Higgs boson is light (i.e., $M_{H^\pm} < m_t + m_b$), it can be produced from top or anti-top decay: i.e., $gg, q\bar{q} \to t\bar{t}$ followed by, e.g., $\bar{t} \to \bar{b}H^-$ (+ c.c.). Given the fact that the $t\bar{t}$ cross section is very large, this mechanism would give an important source of light charged Higgs states. After the top-bottom threshold (i.e., $M_{H^\pm} > m_t + m_b$), a charged Higgs boson can be produced in association with top-bottom pairs, i.e., $bg \to tH^-$ [144]. In fact, these two channels are captured at once by the $gg \to t\bar{b}H^-$ (+ c.c.) 'complete' process, as explained in [145, 146]. There exist other production mechanisms too, such as $pp \to H^+H^-$, $pp \to A^0H^\pm$, $pp \to H^\pm W^\mp$, etc. which are however subleading compared to the previous ones[1].

At the LHC, a light charged Higgs boson, with $M_{H^\pm} < m_t + m_b$, can be detected from $t\bar{t}$ production followed by top or anti-top quark decay $t \to bH^+$ if the H^- state decays dominantly to $\tau\nu$. ATLAS and CMS have already set a limit on $\mathrm{BR}(t \to bH^+) \times \mathrm{BR}(H^+ \to \tau\nu)$ [147–150], which can be translated into a limit on the $(M_{H^\pm}, \tan\beta)$ plane, where $\tan\beta$ is the ratio of the two Higgs doublet vaccuum expectation values (VEVs). In the MSSM, for some specific benchmark scenarios, charged Higgs bosons with mass less than about 160 GeV are ruled out for almost any value of $\tan\beta$ [149, 150]. However, heavy charged Higgs states, with $M_{H^\pm} > m_t + m_b$, are generally allowed as they would decay dominantly into $t\bar{b}$, which is a rather difficult final state to extract due to large reducible and irreducible backgrounds associated with jets emerging from $H^- \to \bar{t}b$ decays. Even then, one could still get a moderate signal from such a channel for small $\tan\beta \le 1.5$ or large $\tan\beta \ge 40$ [151, 152]. Another possibility for detecting heavy charged Higgs states would be the search for $H^+ \to \bar{\tau}\nu_\tau$ (i.e., like the preferred one for a light state), which enjoys a smaller background in comparison. At the LHC Run 2, both channels have been searched for and no excess over the background only hypothesis have been reported. Therefore, limits are set on $\sigma(pp \to tH^-) \times \mathrm{BR}(H^- \to \bar{t}b/\tau\bar{\nu}_\tau)$ (+ c.c.) [82, 153–155]. In the MSSM, one can have additional SUSY channels that can contribute to $H^\pm$ production and/or decay, e.g., production from squark/gluino cascades [156, 157] and/or decays into chargino-neutralino states [158, 159], though these require special MSSM configuration assumptions, hence they are not currently pursued by ATLAS and CMS.

The current highest priority, in relation to charged Higgs boson searches at the LHC, is to further establish the $H^+ \to t\bar{b}$ decay channel in the heavy mass region. With this in mind, using the framework of a generic 2HDM [160], we

---

[1]For a recent review, see[67].

have investigated the possibility of having large interference effects between signals from a heavy charged Higgs boson via $bg \to tH^- \to tW^- A \to tW^- b\bar{b}$ (and similarly for h and H) and the irreducible background from $bg \to tW^- b\bar{b}$ processes. Therein, it was shown that such interference effects can modify any dedicated charged Higgs boson searches where signal and background are treated separately, which is the case for all aforementioned experimental analyses.

The purpose of this paper is to address similar issues for the MSSM, i.e., to quantify the impact of interference effects between the 'complete' signal $pp \to t\bar{b}H^-$ and the irreducible background in the $H^- \to b\bar{t}$ decay channel. We will show that such effects are indeed large for heavy $H^\pm$ masses for two MSSM benchmark scenarios, both for inclusive cross section calculations and after a full detector analysis. The plan of the paper is as follows. In the forthcoming section we describe the MSSM configurations used. Sect. 3 dwells on the MSSM spectra conducive to generate such large interference phenomena. Sect. 4 presents our numerical results. Finally, we conclude in Sect. 5.

8.2   Definition of the benchmark scenarios

At tree level, the MSSM Higgs sector is completely fixed by 2 parameters: $\tan\beta$ and a Higgs boson mass, e.g., the CP-odd one (M_{A^0}). One of the major predictions of SUSY is the presence of a light CP-even Higgs (lighter than Z boson at the lowest order) in the spectrum. However, high order corrections can shift such a mass in order to fit the observed Higgs-like particle mass [161–163]. It has been shown in [164] that high order corrections could raise the tree-level MSSM prediction for such a mass up to 135 GeV for large soft trilinear breaking terms and also that the theoretical uncertainties due to the unknown high order effects should be of the order of 3 GeV.

In the MSSM, the most important parameters relevant for the prediction of the masses, couplings and, hence, production cross sections and decay probabilities of the Higgs bosons are: $\tan\beta$, M_{A^0}, the soft SUSY-breaking masses for the stop and sbottom squarks (which, for simplicity, we assume all equal to a common mass parameter M_S), the soft SUSY-breaking gluino mass $m_{\tilde{g}}$, the Superpotential Higgs-mass parameter μ and the left-right mixing terms in the stop and sbottom mass matrices, i.e.,

$$X_t = A_t - \mu \cot\beta, \qquad X_b = A_b - \mu \tan\beta, \qquad (8.1)$$

respectively.

We use the PROSPINO public code [165] to compute the charged Higgs boson production cross section $\sigma(pp \to t(\bar{b})H^- + \text{c.c.})$, which includes Next-to-Leading Order (NLO) corrections to the $bg \to tH^- + \text{c.c.}$ (2-to-2) process. We use the inclusive cross section computed this way to test the viability of our proposed MSSM scenarios against data.

However, we adopt the tree-level $pp \to t\bar{b}H^- + \text{c.c.}$ (2-to-3) process for Monte Carlo (MC) event generation, because it produces a better description of the signal at the differential level in the detector region than the former channel (i.e., the additional b-(anti)quark is explicit in the phase space rather than integrated into the proton content) and because the corresponding irreducible background is only known at LO. This clearly implies that the normalisation used for the MC analysis is different from that used in the inclusive parameter scans, however, we note that we are primarily concerned here with the relative behaviour of signal, irreducible background and relative interference, rather than the overall normalisation. (Note that, hereafter, we always sum over both H^+ and H^-.) Both PROSPINO and FeynHiggs [166, 167] use the same (on-shell) renormalisation scheme, therefore, the input values of the MSSM parameters can be passed seamlessly from the Higgs spectrum generator to the cross section calculator. The MSSM parameter space is already highly constrained by asking that one of the CP-even neutral scalar states should mimic the properties of the SM-like Higgs boson observed at LHC while the additional Higgs bosons should satisfy the existing constraints obtained from ATLAS and CMS from different channels. For this purpose, the FeynHiggs code is linked to HiggsBounds-5.2.0beta [168–171] and HiggsSignals-2.2.0beta [172] allowing us to check the consistency of our parameter space against various LHC as well as Tevatron and LEP constraints. We list in Tab. 8.1 the specific charged Higgs boson searches that have been included in HiggsBounds. Additionally, a variety of lower energy constraints have been enforced, such as $B \to \tau\nu$, $B_{d,s} \to \mu^+\mu^-$, $B \to X_s\gamma$ and $\Delta m_{s,d}$ (see details in [180]).

All of the MSSM benchmark scenarios adopted in our analysis are characterised by relatively large values of the ratio X_t/M_S. This ensures that the mass of the SM-like Higgs state falls within the required range without the need for an extremely heavy stop. In addition, the gaugino mass parameters, M_2 and M_1, are usually assumed to be related via the GUT relation

$$M_1 = \frac{5}{3}\frac{\sin^2\theta_W}{\cos^2\theta_W}M_2. \tag{8.2}$$

We set the Higgs-sfermion interaction terms A_f to zero for the first and second generation fermions: $f = u, d, c, s, e, \mu$. Moreover, the masses of the gluino

Experiment	Luminosity [fb^{-1}]	Label	Channel
LEP	–	[173]	$e^+e^- \to H^+H^- \to qq'qq'$
LEP	–	[174] (DELPHI)	$e^+e^- \to H^+H^- \to qq'qq'$
LEP	–	[174] (DELPHI)	$e^+e^- \to H^+H^- \to \tau\nu\tau\nu$
D0	1.000	[175] (D0)	$t \to bH^+ \to bqq'$
CDF	2.200	[176] (CDF)	$t \to H^+b$
CDF	0.192	[177] (CDF)	$t \to bH^+ \to b\tau\nu$
CDF	0.335	[178] (CDF)	$t \to bH^+ \to b\tau\nu$
D0	1.000	[175] (D0)	$t \to bH^+ \to b\tau\nu$
ATLAS, 7 TeV	0.035	ATLAS-CONF-2011-094	$t \to H^+b \to c\bar{s}b$
ATLAS, 7 TeV	4.600	[179]	$t \to H^+b$
ATLAS, 8 TeV	19.500	ATLAS-CONF-2014-050	$t \to bH^+ \to b\tau\nu$
ATLAS, 13 TeV	36.100	[150]	$pp \to tbH^+ \to tb\tau\nu$
ATLAS, 13 TeV	36.100	[150]	$t \to bH^+ \to b\tau\nu$
CMS , 8 TeV	19.700	CMS-PAS-HIG-14-020	$t \to bH^+ \to b\tau\nu$
CMS, 8 TeV	19.700	CMS-PAS-HIG-13-035	$t \to H^+b \to c\bar{s}b$
CMS, 8 TeV	19.700	CMS-PAS-HIG-16-030	$t \to H^+b \to cbb$
CMS, 13 TeV	12.900	[153]	$t \to bH^+ \to b\tau\nu$
CMS, 13 TeV	35.900	CMS-PAS-HIG-18-014	$t \to H^+b \to cbb$
ATLAS, 13 TeV	14.700	ATLAS-CONF-2016-088	$pp \to tbH^+ \to tb\tau\nu$
ATLAS, 13 TeV	36.100	[155]	$pp \to tbH^+ \to ttbb$

Table 8.1 Constraints on charged Higgs boson processes implemented in Higgs-Bounds and used in our analyses.

MSSM Scenarios	hMSSM	$m_h^{\text{mod+}}$
$\tan\beta$	1–15	1–25
M_{A^0} (GeV)	150–1000	90–1000
$M_{Q_{1,2}} = M_{U_{1,2}} = M_{D_{1,2}}$ (TeV)	-	1.5
$M_{Q_3} = M_{U_3} = M_{D_3}$ (TeV)	-	1
$M_{L_{1,2}} = M_{E_{1,2}}$ (TeV)	-	0.5
$M_{L_3} = M_{E_3}$ (TeV)	-	1
μ (TeV)	-	0.2
X_t (TeV)	-	1.5
A_t (TeV)	-	$X_t + \mu/\tan\beta$
A_b (TeV)	-	A_t
A_τ (TeV)	-	A_t
M_1 (TeV)	-	GUT relation
M_2 (TeV)	-	0.2
M_3 (TeV)	-	1.5

Table 8.2 MSSM input parameters for our two MSSM benchmark scenarios.

and first two generation squarks are set to 1.5 TeV, large enough to evade the current ATLAS and CMS limits from SUSY searches. In Tab. 8.2 we list the MSSM parameters needed for the evaluation of the spectrum. We now move on to a detailed description of the MSSM benchmark scenarios to be used here, known as $m_h^{\text{mod+}}$ [181] and hMSSM [182].

8.2.1 The $m_h^{\mathrm{mod}+}$ scenario

The $m_h^{\mathrm{mod}+}$ scenario is a modification of the so-called maximal mixing scenario m_h^{max} [183] which was introduced to maximize M_{h^0} value by incorporating large high order effects and also to give conservative limit on $\tan\beta$ during Higgs boson searches at LEP. This scenario predicts a CP-even Higgs M_{h^0} slightly larger than the observed Higgs mass and that is why m_h^{max} scenario has been modified in order to accommodate the observed Higgs of 125 GeV. The modification is performed by reducing the amount of scalar top mixing such that the mass of the lightest Higgs state, M_{h^0}, is compatible with the mass of the observed Higgs boson within ± 3 GeV as a theoretical uncertainty. When confronting $m_h^{\mathrm{mod}+}$ with the LHC data [181], there is a substantial region in $(M_{A^0}, \tan\beta)$ plan with $\tan\beta > 7$ for which the light CP-even Higgs is in a good agreement with the measured Higgs mass at the LHC. The SUSY inputs for this scenario are given in the second column of Tab. 8.2 and the spectrum is computed by the use of FeynHiggs code.

8.2.2 The hMSSM scenario

In $m_h^{\mathrm{mod}+}$, one needs to input M_{A^0}, $\tan\beta$ and other SUSY parameters in order to make a prediction for M_{h^0} within the allowed range, [122, 128] GeV. However, plenty of points on the $(M_{A^0}, \tan\beta)$ plane would correspond to one value of M_{h^0}, the SM-like Higgs boson mass. In order to avoid such a situation, the hMSSM was introduced [182] in which M_{h^0} was enforced to be approximately 125 GeV as well as the SUSY breaking scale $M_{\mathrm{SUSY}} \approx M_S$ fixed to be rather high, $\gg 1$ TeV, in order to explain the non-observation of any SUSY particle at colliders. A key assumption of the hMSSM is to assume that radiative corrections to the diagonal mass of the heavy CP-even Higgs, ΔM_{22}, are much larger than the ones to the light CP-even Higgs, ΔM_{11}, and the mixing term between h and H, ΔM_{12} [182], that is: $\Delta M_{22} \gg \Delta M_{11}, \Delta M_{12}$. Therefore, ΔM_{22}, which parameterises the SUSY effects, is traded for the experimental value of M_{h^0}, $\tan\beta$ and M_Z. Therefore, the hMSSM setup describes the MSSM Higgs sector in terms of just M_{A^0} and $\tan\beta$, exactly like the tree-level predictions, given the experimental knowledge of M_Z and M_{h^0}. The SUSY input parameters in this scenario are given in the first column of Tab. 8.2 and the spectrum is computed by the HDECAY code [184].

8.3 Higgs boson masses and Branching Ratios (BR)

In our analysis, we include the measured signal rates from the ATLAS and CMS Run 2 results via HiggsSignals-2.2.0beta [172] which returns a χ^2 value for the consistency between the model predicted signal rates and the corresponding measurements. We then determine the minimal χ^2 value over the scanned parameter space, $\chi^2_{\min}$, and keep as allowed the portion of it that features a χ^2 value within $\Delta\chi^2 \equiv \chi^2 - \chi^2_{\min}$. For every benchmark scenario, we show the $\Delta\chi^2$ behavior, the best fit point, the charged Higgs total width, the typical BRs for charged Higgs decays into various final states and the charged Higgs production cross section.

8.3.1 The hMSSM case

In Fig. 8.1, we present $\Delta\chi^2$ (top-left) and the charged Higgs total width (top-right) in the $(M_{A^0}, \tan\beta)$ plane. The best fit point is located at $M_{A^0} \approx 1$ TeV and $\tan\beta \approx 2$. The green lines show the exclusion limits from HiggSignals at 1σ (solid) and 2σ (dashed) while the gray area is ruled out by the various LHC searches implemented in HiggsBounds. As one can see, the charged Higgs in the hMSSM scenario is rather heavy ≥ 550 GeV and the total width is large for small $\tan\beta$ and gets reduced for high $\tan\beta$ values. In the bottom panel we show the ratio $\Gamma_{H^\pm}/M_{H^\pm}$ as a function of the charged Higgs mass (left) as well as a function of the charged Higgs production cross section (right). The latter can be slightly above 1 pb. It is also visible from the lower panel that the charged Higgs total width can be about 4% of the charged Higgs mass at low $\tan\beta$.

In the hMSSM scenario, the charged Higgs decays mainly into top-bottom with more than 90% BR for $\tan\beta \leq 8$, see Fig. 8.2 (left), which decreases for larger $\tan\beta$ values. For small $\tan\beta$, the BR($H^+ \to t\bar{b}$) is very close to 100%. In this scenario, the $\tau\nu$ channel has a rather small BR, less than 10%, in most of the cases as depicted in Fig. 8.2 (right) and becomes negligible for low $\tan\beta$.

8.3.2 The $m_h^{\text{mod}+}$ case

In the $m_h^{\text{mod}+}$ scenario, the allowed parameter region is shown in Fig. 8.3 (top-left). The best fit point is located at $M_{H^\pm} \approx 1$ TeV and $\tan\beta = 20$. In order to have a low $\Delta\chi^2$ and simultaneously a light CP-even Higgs, close to 125 GeV, a value of $\tan\beta > 10$ is required. The latter requirement leads to a suppression
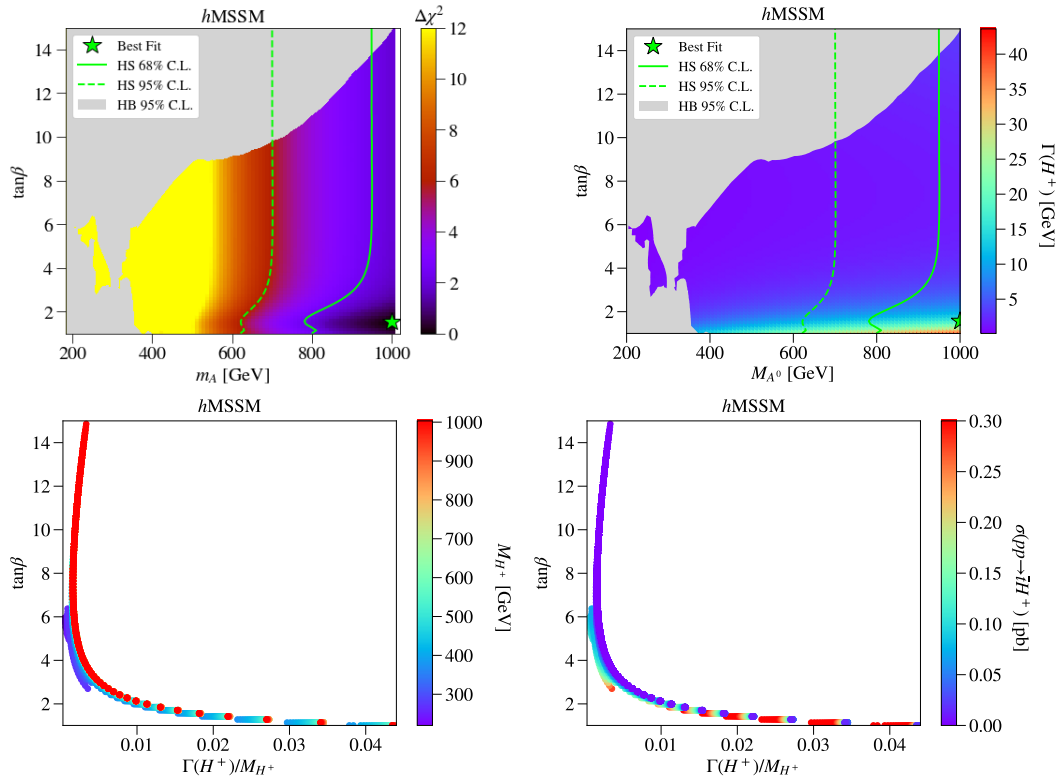
Fig. 8.1 $\Delta\chi^2$ (top-left) and the charged Higgs total width (top-right) in the $(m_A \equiv M_{A^0}, \tan\beta)$ plane. The best fit point is located at $M_{A^0} \approx 1$ TeV and $\tan\beta \approx 2$. The green lines show the exclusion limits from HiggSignals at 1σ (solid) and 2σ (dashed) while the gray area is ruled out by the various LHC searches implemented in HiggsBounds. The ratio $\Gamma_{H^\pm}/M_{H^\pm}$ as a function of the charged Higgs mass is shown in the bottom-left panel while in the bottom-right one it is presented as a function of the charged Higgs production cross section.
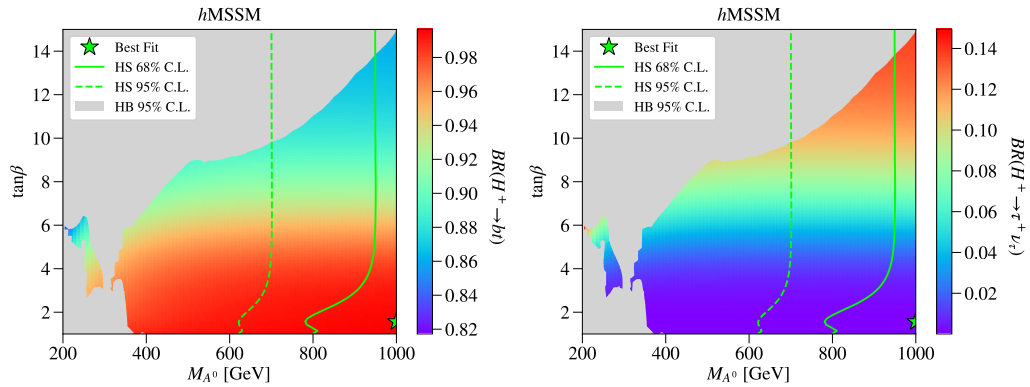
Fig. 8.2 Allowed regions from Fig. 8.1 in the $(M_{A^0}, \tan\beta)$ plane for hMSSM scenario, with the color representing the BRs $\mathrm{BR}(H^+ \to tb)$ (left) and $\mathrm{BR}(H^+ \to \tau\nu)$ (right).

of the total width of the charged Higgs because the BR to top-bottom is proportional to $m_t/\tan\beta$. The same argument holds for the charged Higgs

production cross section which becomes smaller than in the previous scenario. The total charged Higgs width is shown in the allowed parameter region in the top-right panel of Fig. 8.3. Again, we need a heavy charged Higgs boson to obtain sizeable total widths but, contrary to the previous scenario, we now need quite large values of $\tan\beta$. In the two bottom panels of Fig. 8.3, we present $\tan\beta$ as a function of $\Gamma_{H^\pm}/M_{H^\pm}$ with the colour showing the charged Higgs mass (left) and the charged Higgs production cross section (right). Clearly, a compromise has to be reached between the values chosen for the charged Higgs mass while having non-negligible values for the production cross section.

In Fig. 8.4, we again show the allowed region and the colour illustrates the charged Higgs BRs. In the top panels, we show $\mathrm{BR}(H^+ \to t\bar{b})$ and $\mathrm{BR}(H^+ \to \bar{\tau}\nu)$ where it can be seen that the tb BR is larger than the $\tau\nu$ one. In the bottom panels, we only illustrate the dominant chargino-neutralino channels, namely: $\mathrm{BR}(H^+ \to \chi_1^0\chi_1^+)$ and $\mathrm{BR}(H^+ \to \chi_2^0\chi_2^+)$.

8.3.3 Benchmark Points (BPs)

This section briefly outlines the BPs found via the methodology outlined in the previous section.

In Tab. 8.3 we present four BPs for the hMSSM scenario with a value of $\tan\beta = 1.01$ and 5 and with a range of charged Higgs masses between 275 and 633.91 GeV. BP4 in this table was chosen for the numerical analysis in the following section as it has a high charged Higgs width-to-mass ratio and thus is expected to have a high interference entering the signal cross section.

In Tab. 8.4 we present four BPs for the $m_h^{\mathrm{mod+}}$ scenario for a number of values of $\tan\beta$ between 3.42 and 20 and charged Higgs masses between 303.08 and 900 GeV. BP4 was chosen for the numerical analysis of the next section because it has a high cross section even though the charged Higgs width-to-mass ratio is very low. This BP provides insight into the behaviour of interference throughout the cutflow in the scenario where interference is small relative to the signal.

In both scenarios, BPs 1 to 3 are presented to motivate further research of interference in the future as they should provide interesting and, as yet un-excluded, points of the MSSM.

8.4 Results

As intimated, the process studied at MC level is $pp \to t\bar{b}H^- \to tb\bar{t}\bar{b}$ (+ c.c.), thus the signal is defined as all processes in the MSSM mediated by the charged
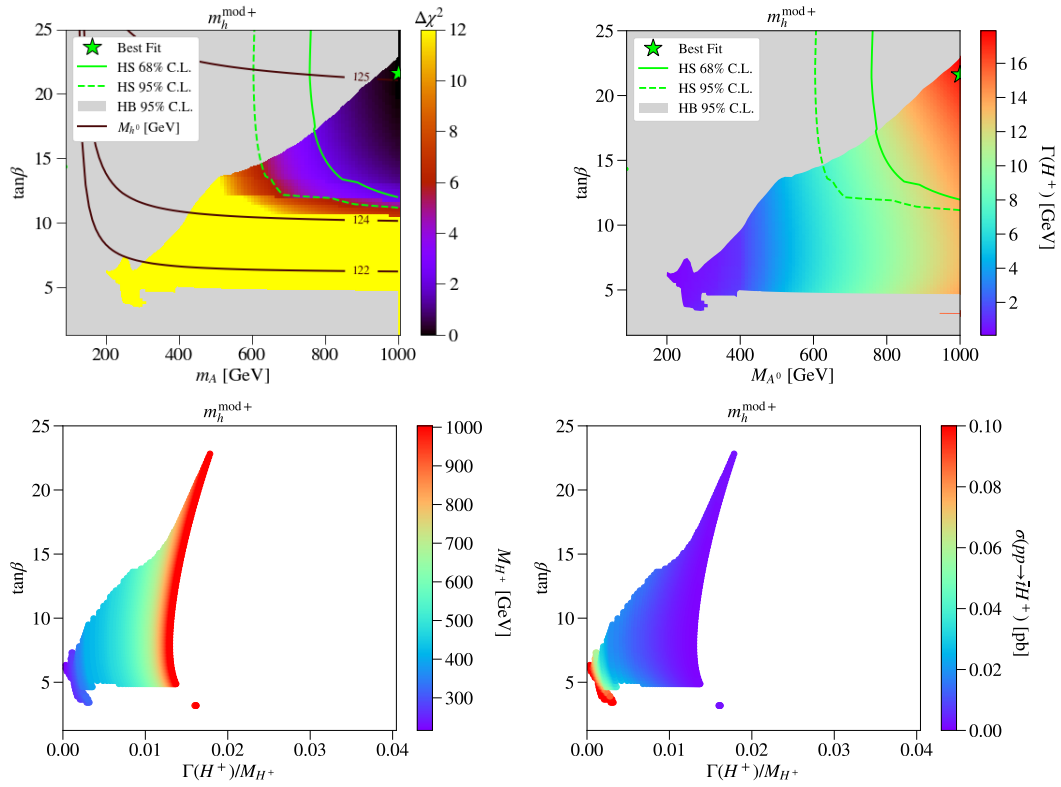
Fig. 8.3 Allowed parameter region in the $m_h^{\mathrm{mod}+}$ scenario over the $(m_A \equiv M_{A^0},$ $\tan\beta)$ plane with colour showing $\Delta\chi^2$ (top-left) and the charged Higgs boson mass (top-right). The LHC Higgs searches constraints are included. The light green contours are HiggsSignals exclusion limits at 1σ (solid) and 2σ (dashed). The light gray area is excluded by HiggsBounds at 2σ. The solid brown lines are contours for the lighter CP-even scalar h^0 mass. The best fit point is located at $M_{H^\pm} \approx 1$ TeV and $\tan\beta = 20$. In the two bottom panels of Fig. 8.3 we present $\tan\beta$ as a function of $\Gamma_{H^\pm}/M_{H^\pm}$ with the colour code showing the charged Higgs mass (left) and the charged Higgs production cross section (right).

Higgs with a $t\bar{t}b\bar{b}$ final state while the background is defined as all processes in the MSSM with the same final state which are not mediated by a charged Higgs state. Figs. 8.5 and 8.6 present some examples of signal and background diagrams.

Let us then define the scattering interference as I = T − S − B where 'T = Total' is the full scattering amplitude including all signal and background Feynman diagrams and the interference of these diagrams. 'S' is the signal scattering amplitude including only the signal diagrams and 'B' is the background scattering amplitude including only the background diagrams. As the same phase space is shared by all of these terms, we can perform the calculation of these terms independently and evaluate the interference via the equation presented above.
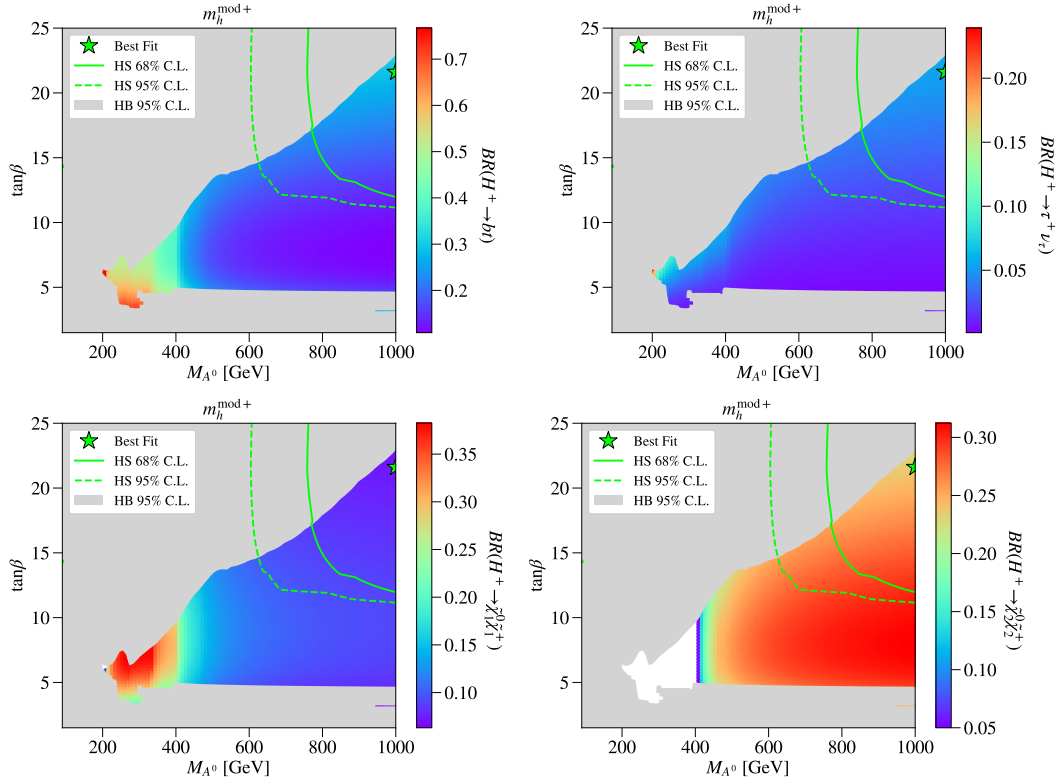
Fig. 8.4 Allowed regions, as shown in Fig. 8.3, in the $(M_{A^0}, \tan\beta)$ plane. We present the BR $H^+ \to t\bar{b}$ (top-left), $H^+ \to \bar{\tau}\nu$ (top-right), $H^+ \to \chi_1^0\chi_1^+$ (bottom-left) and $H^+ \to \chi_2^0\chi_2^+$ (bottom-right).

Parameters	BP1	BP2	BP3	BP4
MSSM inputs				
$\tan\beta$	5	5	1.01	1.01
Masses in GeV				
M_{h^0}	125	125	125	125
M_{H^0}	266.77	495.19	615.32	648.3
M_{A^0}	263	493.5	594.6	628.79
M_{H^+}	275.01	500	600.01	633.91
Total decay width in GeV				
$\Gamma(H^+)$	0.3499	1.0423	26.177	27.777
BR$(H^+ \to XY)$ in %				
BR$(H^+ \to bt)$	91.665	96.105	99.375	99.418
Ratios				
$\Gamma(H^+)/M_{H^+}$	0.0012723	0.0020846	0.043628	0.043819
Cross sections in pb				
$\sigma(pp \to \bar{t}H^+)$	0.0932	0.0177	0.2090	0.1431
$\sigma(pp \to \bar{t}H^+, H^+ \to bt)$	0.0854	0.0170	0.2077	0.1423

Table 8.3 BPs for the hMSSM scenario.

Parameters	BP1	BP2	BP3	BP4
MSSM inputs				
$\tan\beta$	6	10	20	3.42
Masses in GeV				
M_{h^0}	120.18	122.46	123.47	113.55
M_{H^0}	595.33	695.4	896.01	298.69
M_{A^0}	594.38	695.12	895.96	292.22
M_{H^+}	600	700	900	303.08
$M_{\tilde{\chi}_1^+}$	139.97	144.16	147.54	133.2
$M_{\tilde{\chi}_2^+}$	270.8	268.59	266.75	274.19
$M_{\tilde{\chi}_1^0}$	84.345	86.404	87.934	80.637
$M_{\tilde{\chi}_2^0}$	147.2	149.46	151.39	143.88
$M_{\tilde{\chi}_3^0}$	209.7	209.8	210.14	205.33
$M_{\tilde{\chi}_4^0}$	271.76	268.81	266.41	276.42
$M_{\tilde{b}_1}$	1000	999.82	996.08	998.97
$M_{\tilde{b}_2}$	1002	1002.2	1006	1002.8
$M_{\tilde{t}_1}$	876.49	876.45	876.43	876.61
$M_{\tilde{t}_2}$	1134.8	1134.8	1134.8	1134.9
Total decay width in GeV				
$\Gamma(H^+)$	5.8582	7.7229	14.311	0.9253
BR$(H^+ \to XY)$ in %				
BR$(H^+ \to \tilde{\chi}_1^0\tilde{\chi}_1^+)$	10.789	10.379	7.7896	20.73
BR$(H^+ \to \tilde{\chi}_2^0\tilde{\chi}_2^+)$	27.858	29.296	24.307	–
BR$(H^+ \to \tilde{\chi}_1^+\tilde{\chi}_3^0)$	13.003	12.161	9.1983	–
BR$(H^+ \to \tilde{\chi}_1^+\tilde{\chi}_4^0)$	18.454	18.648	15.061	–
BR$(H^+ \to \tilde{\chi}_3^0\tilde{\chi}_2^+)$	9.7934	11.002	9.5996	–
BR$(H^+ \to \tau^+\nu_\tau)$	0.73738	1.8127	5.031	0.7682
BR$(H^+ \to bt)$	15.728	13.718	26.989	72.036
Ratios				
$\Gamma(H^+)/M_{H^+}$	0.0097637	0.011033	0.015901	0.0031
Cross sections in pb				
$\sigma(pp \to \bar{t}H^+)$	0.007120	0.003170	0.002850	0.130750
$\sigma(pp \to \bar{t}H^+, H^+ \to \tilde{\chi}_1^0\tilde{\chi}_1^+)$	0.000768	0.000329	0.000222	0.027100
$\sigma(pp \to \bar{t}H^+, H^+ \to \tilde{\chi}_2^0\tilde{\chi}_2^+)$	0.001984	0.000929	0.000693	–
$\sigma(pp \to \bar{t}H^+, H^+ \to \tilde{\chi}_1^+\tilde{\chi}_3^0)$	0.000926	0.000386	0.000262	–
$\sigma(pp \to \bar{t}H^+, H^+ \to \tilde{\chi}_1^+\tilde{\chi}_4^0)$	0.001314	0.000591	0.000429	–
$\sigma(pp \to \bar{t}H^+, H^+ \to \tilde{\chi}_3^0\tilde{\chi}_2^+)$	0.000697	0.000349	0.000274	–
$\sigma(pp \to \bar{t}H^+, H^+ \to \tau^+\nu_\tau)$	0.000053	0.000057	0.000143	–
$\sigma(pp \to \bar{t}H^+, H^+ \to bt)$	0.001120	0.000435	0.000769	0.094200

Table 8.4 BPs for $m_h^{\mathrm{mod}+}$ scenario.

In order to explore the effects of interference on the search for a charged Higgs, we utilise BP4 found in Tab. 8.3 for the hMSSM case and BP4 found in Tab. 8.4 for the $m_h^{\mathrm{mod}+}$ case. These two points provide two kinematically distinct scenarios, one of which - the hMSSM one - has a high width-to-mass ratio for the charged Higgs boson, of 4.4%, while the other has a much lower ratio, of 0.31%.
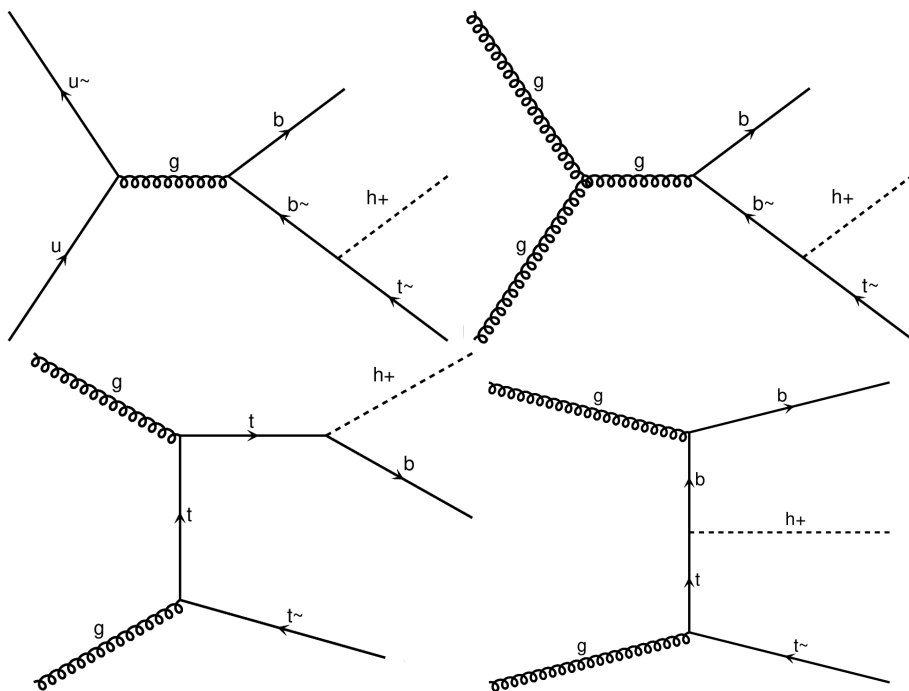
Fig. 8.5 A selection of signal Feynman diagrams.
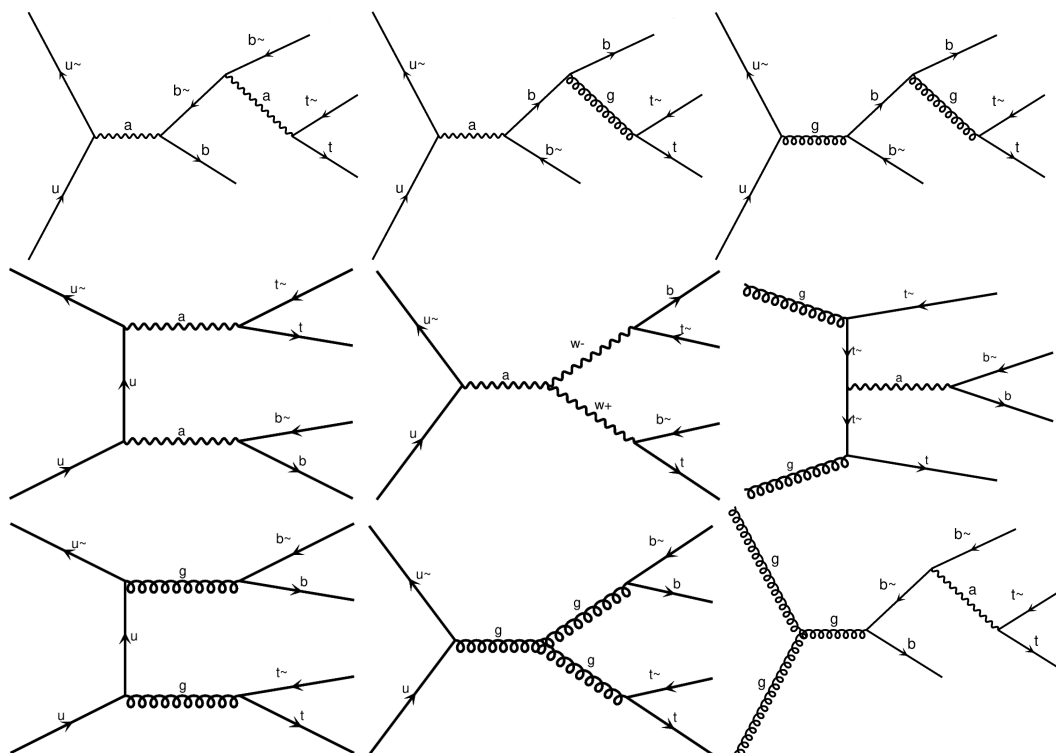
Fig. 8.6 A selection of background Feynman diagrams.

Signal cross sections are significantly smaller than background cross sections before cuts. Hence the simulation of the T and B terms outlined above must have low uncertainty, which requires very large MC samples. This mandates

a prudent use of computing resources and thus an extremely large sample
of events was generated for T, B and S at parton level to obtain a value
for the cross section of these processes with very low MC error. After this
was done, smaller detector level samples were generated for the purpose of
applying cuts and obtaining efficiencies. The parton level cross sections and
the detector level efficiencies are then used together to form the cutflow that
will be presented in this section. The parton level results for both BPs can be
found in Tab. 8.5. The error presented here is the MC error and is displayed
to show that a large enough parton level sample was generated to produce a
sufficiently low uncertainty on the interference cross section. There are of course
other sources of errors against which to tension the size of our interference
effects, such as systematic theoretical errors due to the finite order perturbative
modelling of the signal and background cross sections, which can be large,
especially for the latter (e.g., Ref. [155] finds that for the $t\bar{t}b\bar{b}$ background such
an error is of order 10%). Further, one ought to consider the systematic errors
coming with the choice of Parton Distribution Functions (PDFs) and of their
factorisation/renormalisation scale, which are expected to be similar in size.
Hence, one will truly need to worry about the systematic error due to the
presence of the interference effects studied here when they are beyond the 10%
or so level.

Model		S (pb)	B (pb)	T (pb)	I (pb)
hMSSM	σ	0.03240	13.078	13.139	0.028
	$\Delta\sigma$	1.4×10^{-5}	0.002	0.001	0.003
$m_h^{\mathrm{mod+}}$	σ	0.08854	13.095	13.197	0.014
	$\Delta\sigma$	3.3×10^{-5}	0.001	0.001	0.002

Table 8.5 Parton level results for the hMSSM and $m_{\mathrm{h}}^{\mathrm{mod+}}$ benchmarks.

The parton level sample for both scenarios contained 20,000,000 events
generated in MadGraph5 [47] at leading order with a Centre-of-Mass (CoM)
energy of 13 TeV, while the detector level samples for the hMSSM sample
contained 5,000,000 events comprised of 100 independent samples of 50,000
events and the $m_h^{\mathrm{mod+}}$ sample contained 10,000,000 events comprised of 200
independent samples of 50,000 events. Both were generated in MadGraph5 at
leading order and at 13 TeV CoM energy. The detector level samples were then
sent to Pythia8 [185] for hadronisation/fragmentation and finally passed to
Delphes [53] for detector smearing utilising the standard ATLAS card. Previous
sections of this work calculated cross sections at NLO, however, as previously
explained, this is not feasible for the background samples, so the MC analysis

was undertaken at LO only. All samples included the decay of the charged Higgs, $H^+ \to t\bar{b}$, to maximize statistics.

Typical detector acceptances were utilised, namely electrons and muons must have transverse momentum $p_T > 7$ GeV and pseudo-rapidity $|\eta| < 2.5$ with 100% lepton selection efficiency assumed. Jets must have $p_T > 20$ GeV and $|\eta| < 2.5$. Anti-k$_T$ jet clustering[37] was used and a b-tagging efficiency of 77% and mis-tagging efficiency of 1% employed. We demand exactly one lepton in the final state, so that the longitudinal momentum of the missing energy can be solved for via

$$p_\nu^z = \frac{1}{2p_{\ell T}^2} \left(A_W p_\ell^z \pm E_\ell \sqrt{A_W^2 \pm 4p_{\ell T}^2 E_{\nu T}^2} \right),\tag{8.3}$$

where, $A_W = M_{W^\pm}^2 + 2p_{\ell T} \cdot E_{\nu T}$.

Reconstruction was then undertaken via the simultaneous minimisation of the following equations by permuting through all combinations of jets in the process,

$$\chi_{\text{had}}^2 = \frac{(M_{\ell\nu} - M_W)^2}{\Gamma_W^2} + \frac{(M_{jj} - M_W)^2}{\Gamma_W^2} + \frac{(M_{\ell\nu j} - M_T)^2}{\Gamma_T^2}$$
$$+ \frac{(M_{jjj} - M_T)^2}{\Gamma_T^2} + \frac{(M_{jjjj} - M_{H^\pm})^2}{\Gamma_{H^\pm}^2}\tag{8.4}$$

and

$$\chi_{\text{lep}}^2 = \frac{(M_{\ell\nu} - M_W)^2}{\Gamma_W^2} + \frac{(M_{jj} - M_W)^2}{\Gamma_W^2} + \frac{(M_{\ell\nu j} - M_T)^2}{\Gamma_T^2}$$
$$+ \frac{(M_{jjj} - M_T)^2}{\Gamma_T^2} + \frac{(M_{\ell\nu jj} - M_{H^\pm})^2}{\Gamma_{H^\pm}^2}\tag{8.5}$$

The results of this reconstruction can be found in Figs. 8.7 and 8.9, normalised to unit area. This reconstruction requires one to use the width of the particles, which introduces a model dependence. Thus, in the aforementioned figures, we also present the same reconstruction methodology but without the use of particle widths to highlight how this affects the reconstruction. We refer to these methodologies as the model dependent and model independent reconstructions, respectively.

We apply a simple set of acceptance cuts to illustrate the sensitivity to the interference term, these cuts include a final state definition of 1 lepton, 5 or more jets, more than 2 or 3 b-jets, greater than 20 GeV missing transverse energy and, finally, the transverse mass of missing energy and the lepton must be higher

than 60 GeV. Specifically, $m_T^W = \sqrt{(\slashed{E}_x + \ell_x)^2 + (\slashed{E}_y + \ell_y)^2} > 60$ GeV. It is an interesting question to experimentalists as to how interference contributions change with respect to varying b-tagging. This motivates the usage of 2 and 3 b-tags regions even if the increased b-tagging does not necessarily lead to increased signal significance. This cutflow, applied to each of the BPs, can be found in Tabs. 8.6 and 8.7.

8.4.1   The hMSSM analysis

It can be seen in Fig. 8.7 that all particles appear to be reconstructed very well. The model-dependent reconstruction and the model-independent case perform equally well for the signal. However, for the background and total samples the reconstruction is quite different. The model-dependent assumption provides a much better separation from the signal, this is especially apparent in both the leptonic and hadronic charged Higgs invariant mass distributions.

The ratio of signal cross section to interference cross section before cuts is 86.7%. This is an alarmingly high level of interference that a traditional experimental study would not account for correctly. The ratio after cuts in both the ≥ 2 b-tag scenario and ≥ 3 b-tag scenarios is 103.3% and 85.5% respectively, both extremely large interferences showing that the cutflow has done little to mitigate the magnitude of the interference relative to the signal. It should be noted that the uncertainty on the values in the ≥ 3 b-tag region are approaching the same magnitude as the interference itself, thus strong conclusions in this region cannot be made.

It is important to note that the true effect of the interference is predicated on the overall shape of the interference distribution relative to the signal distribution. In general there are three cases [186–188]:

1. The interference takes the same shape as the signal and is positive, here we can expect a boosting of our new physics effects.

2. The interference takes the same shape as the signal and is negative, here we expect a cancellation of our new physics effects.

3. The interference takes a different shape and is either positive or negative, here we can expect a boosting and cancellation of new physics effects in different regions of phase space, manifesting as a "peak-dip" structure in the expected distributions.

In Fig. 8.8, an exploration of this shape at parton level before cuts can be seen in the $t\bar{t}b\bar{b}$ reconstructed invariant mass plane. This step was undertaken

at parton level to achieve the required per-bin statistics to discern the shape of the interference distribution. There appears to be a large interference impact in across the whole mass range, interestingly though the largest contributions are below the charged Higgs mass peak. The result of this would likely be a smearing of the charged Higgs mass bump towards lower values in actual data.
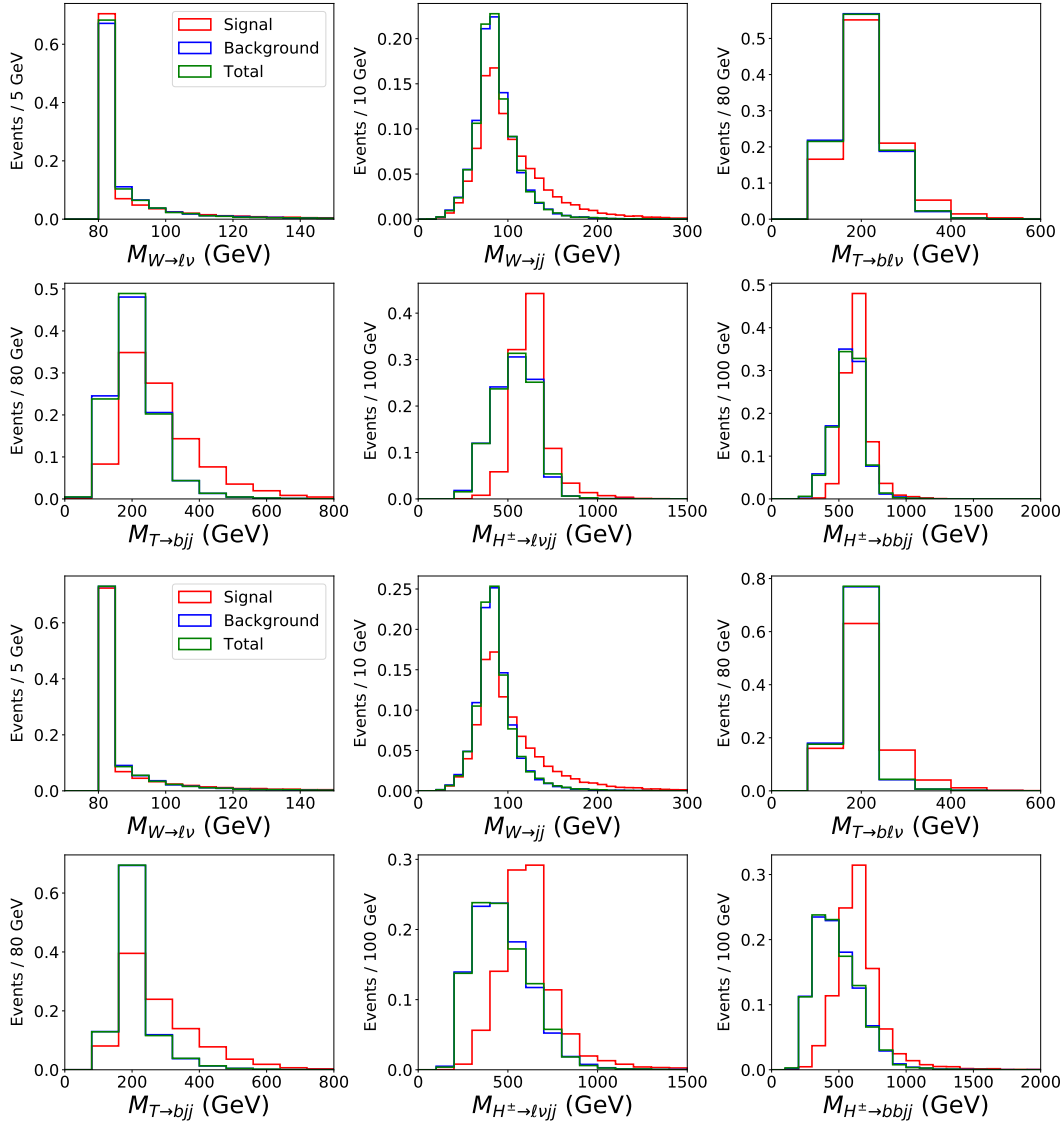
Fig. 8.7 Invariant mass distributions for reconstructed particles in the hMSSM benchmark. Top: Model independent. Bottom: Model dependent.

8.4.2 The $m_h^{\text{mod}+}$ analysis

The reconstruction in this scenario presents far less distinct signal distributions, which can be seen in the charged Higgs invariant mass plots of Fig. 8.9. As the mass difference between the charged Higgs and the sum of the t and b-quark

Cut	S	B	T	I	ΔI
No cuts:	9720	3923550	3941700	8429	2487
$N_\ell = 1$:	2160	904247	907925	1518	1193
$N_J \geq 5$:	1938	624001	627534	1594	992
$N_{BJ} \geq 2$:	1511	404919	408054	1623	799
$\not{E} > 20$ GeV:	1435	373648	376517	1433	768
$\not{E} + m_T^W > 60$ GeV:	1412	364026	366898	1458	758
Cut	**S**	**B**	**T**	**I**	**ΔI**
$N_{BJ} \geq 3$:	826	171918	173430	684	521
$\not{E} > 20$ GeV:	785	158921	160376	669	501
$\not{E} + m_T^W > 60$ GeV:	772	154880	156314	660	494

Table 8.6 Cut flow results presented in expected event yield with 300 fb^{-1} of luminosity for the hMSSM benchmark with 5,000,000 events for each sample.
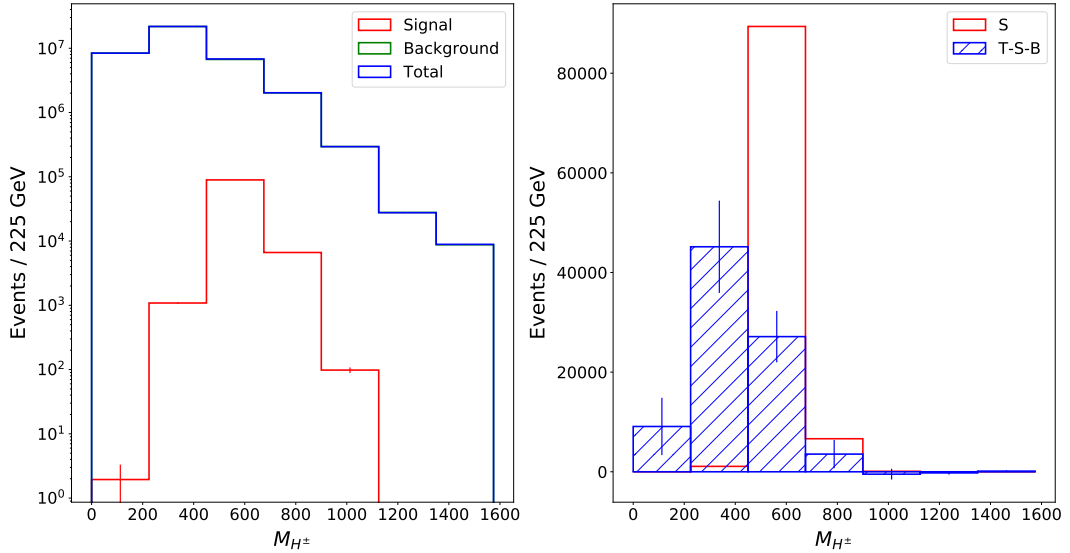
Fig. 8.8 The charged Higgs invariant mass distribution of the signal, background and total samples (left) and interference and signal (right) at parton level and without cuts in the hMSSM scenario.

masses is far smaller in this scenario, it appears that the reconstruction is performed very similarly for the signal and background. This can be further seen in the lack of difference between the model-dependent and model-independent reconstructions. Thus, extraction of the signal would be far more difficult in this case.

As the ratio of the charged Higgs mass to charged Higgs width is smaller in this benchmark than in the hMSSM one, we expect the interference effects to be smaller. However, the interference may become much larger relative to the signal after a cutflow. Tab. 8.7 displays the cutflow results for this BP and one

can see that the pre-cut ratio of interference cross section to signal cross section is 15.2%, while the ratio after cuts in the ≥ 2 b-tag region is 26.0%. In this scenario the more strict use of b-tagging decreased the ratio of the interference cross section to the signal cross section to 17.4%, though it should be noted the uncertainty on the values in the ≥ 3 b-tag region has grown beyond the magnitude of interference and conclusions in this region are tenuous.
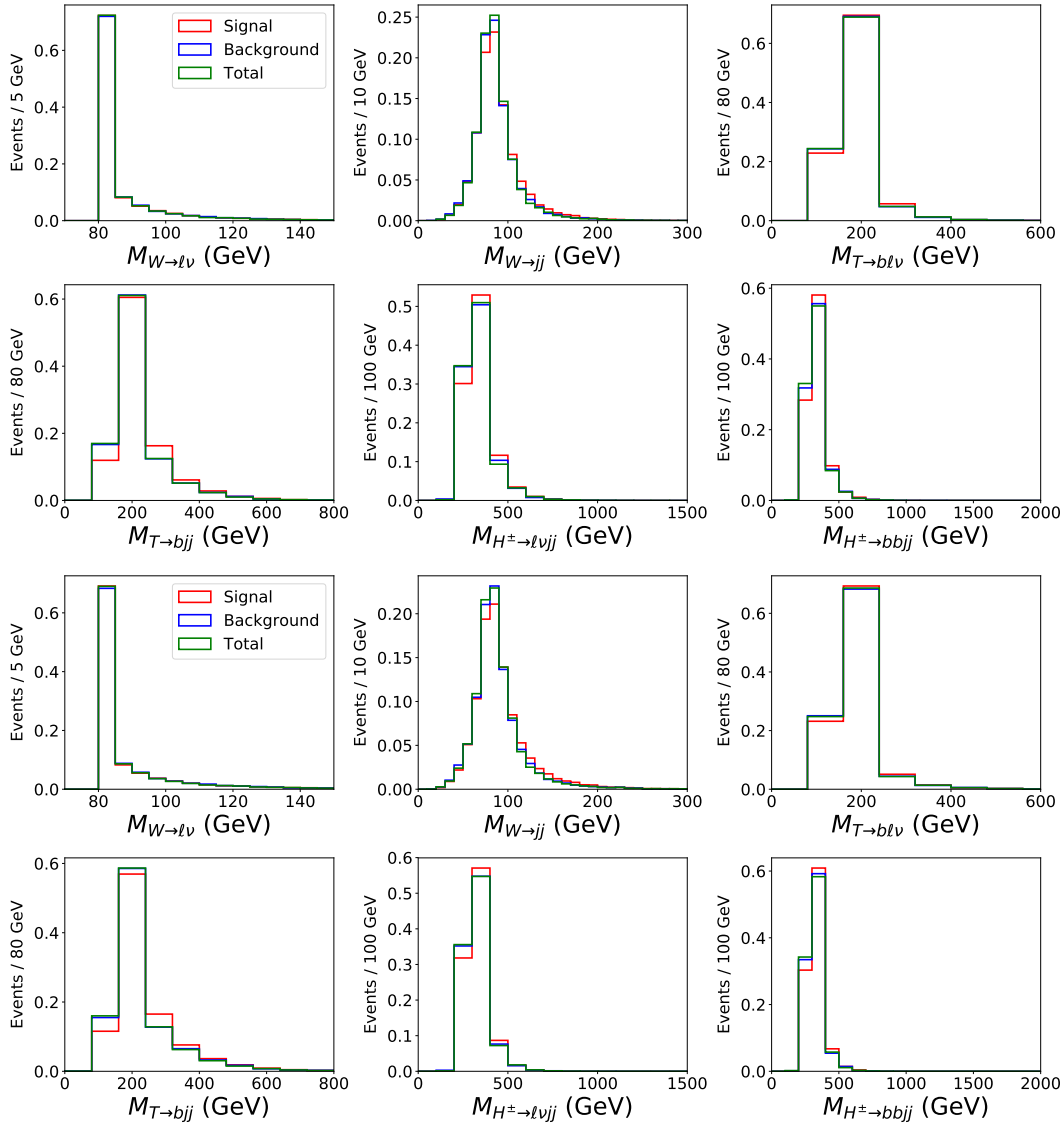
Fig. 8.9 Invariant mass distributions for reconstructed particles in the $m_h^{\mathrm{mod+}}$ benchmark. Top: Model independent. Bottom: Model dependent.

The interference shape of the reconstructed invariant mass of $t\bar{t}b\bar{b}$ at parton level and before cuts in this scenario can be found in Fig 8.10. In this scenario the signal invariant mass distribution peaks narrowly over the 300 GeV bin, however, interestingly the interference distribution appears to be a widely spread spectrum across the range $100 - 350$ GeV. These small, but non-negligible,

Cut	S	B	T	I	ΔI
No cuts:	26561	3928500	3959100	4038	1764
$N_\ell = 1$:	6017	903764	911292	1510	846
$N_J \geq 5$:	4964	623989	629532	578	703
$N_{BJ} \geq 2$:	3704	404776	409342	862	566
$\not{E} > 20$ GeV:	3432	373464	377885	989	544
$\not{E} + m_T^W > 60$ GeV:	3342	363876	368087	868	537
Cut	**S**	**B**	**T**	**I**	**ΔI**
$N_{BJ} \geq 3$:	1894	171654	173822	273	369
$\not{E} > 20$ GeV:	1757	158581	160686	347	354
$\not{E} + m_T^W > 60$ GeV:	1712	154576	156587	298	350

Table 8.7 Cut flow results presented in expected event yield with 300 fb^{-1} of luminosity for the $m_h^{\text{mod+}}$ benchmark with 10,000,000 events for each sample.
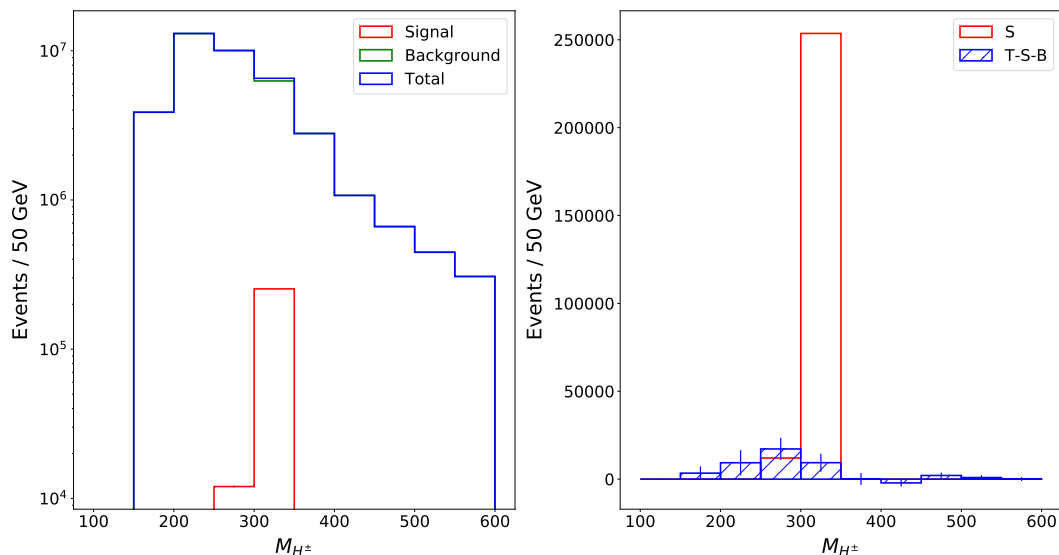
Fig. 8.10 The charged Higgs invariant mass distribution of the signal, background and total samples (left) and interference and signal (right) at parton level and without cuts in the $m_h^{\text{mod+}}$ scenario.

interference contributions would likely lead to a widening of an otherwise sharp signal bump in data, which further motivates the necessity of the signal shape analysis at detector level to discover whether this effect would be dominated by the smearing effect of the detector on the signal peak.

8.5 Conclusions

By borrowing the MSSM as a theoretical template that contains charged Higgs bosons, we have shown how experimental searches for these states cannot be

made immune from large interference effects between signal and background whenever they have a large mass and a width on the order of one percent of the mass and upwards. We have illustrated this for the case of the $H^+ \to t\bar{b}$ decay channel, which is onset dominantly by $gg \to b\bar{t}H^+$ production. In this case, the (irreducible) background intervening in such interference effects is $pp \to t\bar{t}b\bar{b}$, which can see both QCD and EW interactions. This study's goal was to show that signal and background are wrongly treated as separate in current LHC approaches.

In order to realistically assess the above phenomenon, we have decayed the $t\bar{t}$ pair semi-leptonically and carried out a full parton shower, hadronisation and detector analysis. In doing so, we have first prepared the MSSM parameter space regions amenable to phenomenological investigation by enforcing both theoretical (i.e., unitarity, perturbativity, vacuum stability, triviality) and experimental (i.e., from flavour physics, void and successful Higgs boson searches at the Tevatron and LHC, EW precisions observables from LEP and SLC) constraints, assuming two benchmark configurations of the MSSM, the so-called hMSSM and $m_h^{\mathrm{mod}+}$ scenarios.

After performing a sophisticated MC simulation, allowing for both model-independent and model-dependent selections, we have seen that such interference effects can be very large, even of $\mathcal{O}(100\%)$, both before and after $H^\pm$ detection cuts are enforced. This appears to be the case for the masses tested, approximately 300 and 630 GeV, in the MSSM scenarios adopted, though interference effects will manifest themselves at different LHC stages, depending on the overall cross sections, which vary significantly from one benchmark to another. Furthermore, the shapes of the signal and its interference (with the aforementioned irreducible background) appear to be different which would mean that it is not actually possible to proceed to a rescaling of the event yields due solely to the signal. In turn, in experimental analyses, one should account for such interference effects at the event generation level. We have proven this to be the case for a standard cut flow, while deferring the study of similar effects in the case of a machine learning framework to a future publication.

Acknowledgements

Chapter 9

Advanced Analysis Methods

The methodologies to be described in this chapter are very new to physics research, and in some cases very new altogether. The traditional techniques for extracting or excluding a possible physics signal have been explained and applied in detail in earlier chapters of this thesis - that is, a signal of interest and its relevant backgrounds are defined and then the phase space is cut away methodically to retain only areas where the signal is over-represented relative to its background. This is the way particle physics has been done throughout most of the collider era, and modern computers have only refined the process to be orders of magnitude more efficient. The issue is that unless that specific signal model exists in nature then we learn nothing about any signal model that may be present in the data. This chapter covers an entirely new paradigm which is surging in popularity.

The methodology is unsupervised machine learning for "signal agnostic searches". The key to these types of searches is in the name - they are performed without any (or at least minimal) signal defined. Rather than looking for something that looks like signal, we are instead looking for things which do *not* look like the Standard Model.

The concept of signal agnostic searches has existed for a long time, for example the model-independent and pseudo model-independent approaches outlined over a decade ago in Refs [189–195]. These studies employed general hunts for excesses in a variety of final states, but most did not employ modern machine learning techniques. Similarly there have been some more recent studies of this type such as Refs [196–200] at ATLAS and CMS.

As mentioned, unsupervised machine learning techniques have generated interest in recent years such as the use of autoencoders for fat jet and jet substructure analyses in Refs [201, 202] and more relevantly for anomaly detection in Ref [203]. Further anomaly detection methods using variational autoen-

coders and generative adversarial networks have been outlined in Refs [204–206]. Importantly we will see an exploration of Ref [206] in the coming chapter.

In this chapter we will give an overview of the general principles of signal agnostic searches and a few of the most powerful algorithms utilised to undertake them.

9.1   Signal Agnostic Searches

If a supremely powerful computer with sufficient storage capabilities existed one could undertake an enormous MC simulation of every single SM process with as many events as are expected for a given luminosity at the LHC. This dataset could then effectively be cross-referenced with the LHC dataset and any statistically significant differences may point in the direction of new physics (or at least a systematic error). Alas, this is not even close to being possible.

What we can do however is define a very simple set of acceptance cuts. Going back to our Higgs example in chapter 5, if we simply demanded that every ATLAS event contained two photons with transverse momentum over, for example, 50 GeV we would have narrowed down our potential backgrounds to all SM processes which produce a pair of high energy photons - a much smaller set of backgrounds. In the following, let us pretend that we live in world where the Higgs has not been discovered, or perhaps even theorised, we are simply looking for any physics particle that can decay to two high energy photons.

An anomaly detection algorithm, examples of which will be explained in the coming sections, can then be trained on MC samples of the di-photon backgrounds and each event given an "anomaly score". The distribution of these anomaly scores should be representative of the anomalousness of the background samples. An experiment such as ATLAS could then input their di-photon data into this same anomaly detection algorithm and if there is no new physics in this dataset then the resultant anomalousness distribution will be identical to the MC distribution up to statistics. However, if new physics does exist then these events should return a higher anomaly score than the MC backgrounds and will lead to an excess in the high anomaly area of the distribution.

Anomaly detection algorithms come in many forms, and the definition of anomaly score is based on the details of how each specific algorithm works. The coming sections cover three powerful algorithms.

9.2　Gaussian Mixture Models

Mixture models allow one to approximate a set of statistical distributions that a set of data was most likely sampled from. Specifically, Gaussian mixture models (GMMs) are an implementation of this methodology where the individual statistical distributions being fitted are Gaussian distributions [207].

Let us define a set of data points as $\boldsymbol{X} = \{\mathbf{x}_1, \ \ldots \ , \mathbf{x}_n, \ \ldots \ , \mathbf{x}_N\}$, where each $\mathbf{x}_n$ is a vector with d features. Let μ_k, $\boldsymbol{\Sigma}_k$ with $k = 1, \ldots, K$ be the mean vectors and covariance matrices of a chosen number (K) of d-dimensional Gaussian distributions, initialised arbitrarily. For each data point we introduce a vector of latent variables, $\mathbf{z}_n$ representing that it belongs to a particular Gaussian: if the nth data point belongs to the kth Gaussian we set $z_{nk} = 1$, otherwise it is zero.

We can write the probability of observing a given data point $\mathbf{x}_n$ from its Gaussian as

$$p(\mathbf{x}_n|\mathbf{z}_n) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{x}_n|\mu_k, \boldsymbol{\Sigma}_k)^{z_{nk}}, \tag{9.1}$$

where $\mathcal{N}$ denotes a Gaussian distribution. Note that this product occurs over all Gaussians but the way we have constructed the latent vector $\mathbf{z}_n$ suppresses all but the Gaussian $\mathbf{x}_n$ belongs to. Now by marginalisation over all $\mathbf{z}$ we get

$$p(\mathbf{x}_n) = \sum_{k=1}^{K} p(\mathbf{x}_n|\mathbf{z})p(\mathbf{z}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\mu_k, \boldsymbol{\Sigma}_k), \tag{9.2}$$

where we have defined a mixing parameter $\pi_k \equiv p(z_k = 1)$. These represent the probability that an arbitrary point belongs to the k-th mixture component (the k-th Gaussian), and hence the sum of π_k over all k is 1.

We aim to maximise $p(\mathbf{X})$, the overall probability of obtaining the data if it were sampled from the K Gaussians. The log-likelihood of this probability is given by

$$\log\left(p(\boldsymbol{X})\right) = \sum_{n=1}^{N} \log\left(p(\mathbf{x}_n)\right) = \sum_{n=1}^{N} \log\left[\sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\mu_k, \boldsymbol{\Sigma}_k)\right]. \tag{9.3}$$

The optimisation of this function can be performed using the Expectation-Maximisation (EM) algorithm. There are two steps to the EM algorithm: the expectation step (E-step) and maximisation step (M-step).

The E-step is performed by calculating the probability that each point was sampled from a particular Gaussian. This can be expressed in terms of the latent variables as $p(\mathbf{z}_k = 1|\mathbf{x}_n)$, which is often referred to as the *responsibility* of the distribution k for a given data point $\mathbf{x}_n$. Using Bayes law we can write

$$p(\mathbf{z}_k = 1|\mathbf{x}_n) = \frac{p(\mathbf{x}_n|\mathbf{z}_k = 1)p(\mathbf{z}_k = 1)}{\Sigma_{j=1}^K p(\mathbf{x}_n|\mathbf{z}_j = 1)p(\mathbf{z}_j = 1)} = \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\mu_k, \boldsymbol{\Sigma}_k)}{\Sigma_{j=1}^K \pi_j \mathcal{N}(\mathbf{x}_n|\mu_j, \boldsymbol{\Sigma}_j)} \equiv \gamma(z_{nk}).$$
(9.4)

Once we have calculated $\gamma(z_{nk})$ for all n and k we can undertake the M-step to estimate the updated parameters of each Gaussian. First, one calculates the number of points N_k for which Gaussian k is responsible

$$N_k = \sum_{n=1}^N \gamma(z_{nk}).$$
(9.5)

With this value, we update the mean of Gaussian k by calculating the mean of the data points that belongs to it, weighted by the responsibilities $\gamma(z_{nk})$

$$\mu'_k = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk})\mathbf{x}_n.$$
(9.6)

Similarly, the updated covariances for Gaussian k are given by the covariance of the points that belong to Gaussian k with the updated mean μ'_k, weighted by the responsibilities

$$\boldsymbol{\Sigma}'_k = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk})(\mathbf{x}_n - \mu'_k)(\mathbf{x}_n - \mu'_k)^T.$$
(9.7)

Finally, the mixing parameter π_k of Gaussian k is updated by calculating the percentage of the total dataset that belongs to it

$$\pi'_k = \frac{N_k}{N}.$$
(9.8)

The new log-likelihood may be computed directly using Eq. (9.3) with the new parameters for each Gaussian. The process is repeated iteratively, until we see convergence of the log-likelihood (within a set tolerance), or when the maximum number of epochs is reached.

An example of this convergence can be seen in Fig 9.1. In this figure, a two component Gaussian Mixture Model is fitting a two-dimensional dataset of points sampled from arbitrary Gaussians. The ellipses are defined by the

variance of each Gaussian in the x and y directions. Note the gradual improvement of the closeness of fit of the two Gaussian distributions to the two clear clusters that are found in the plot.
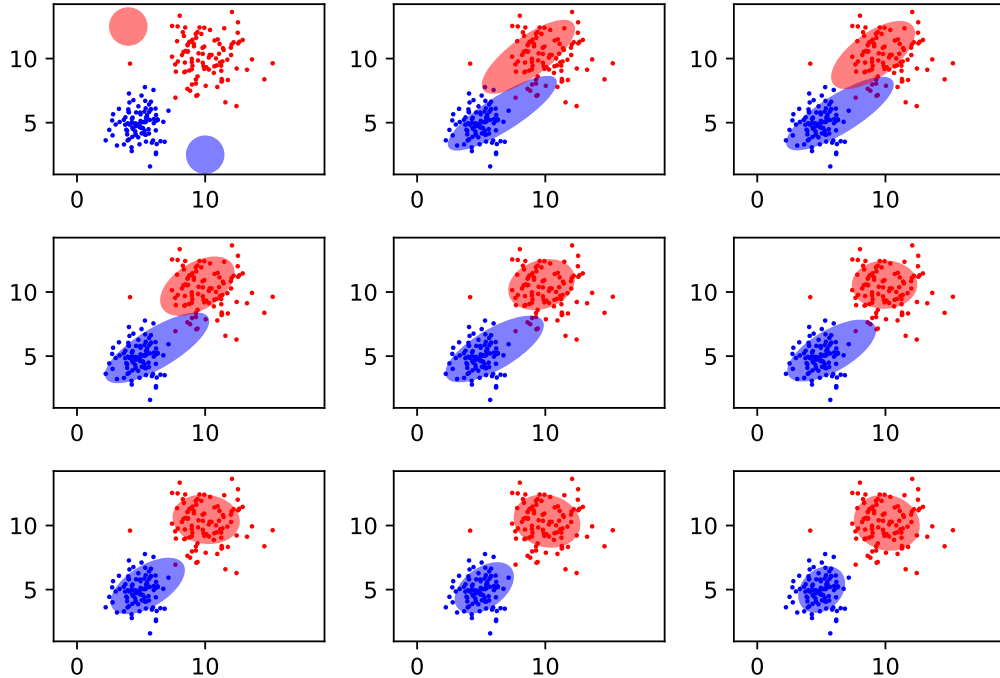
Fig. 9.1 An example of a Gaussian Mixture Model with two components undergoing parameter updates via the EM algorithm and converging to an optimum (to be read from left to right, top to bottom).

To apply this to physics, if we fit a K-component Gaussian Mixture Model to our background MC samples then the weighted probability of a point being sampled from the K-Gaussians is used as the anomaly score. But how do we select K? In the previous example shown in Fig 9.1 it is simple as the data was two dimensional and formed two clear clusters, but in a more complex case neither of these things will likely be true. The standard methodology for selecting the number of components of the mixture model is to choose the value of K that minimises the "information criterion", namely the "Akaike Information Criterion" (AIC) [208] or the "Bayesian Information Criterion" (BIC) [209]. If the number of data points in the dataset is n, the number of parameters of the mixture model is k and the maximum of the model's likelihood function given the data is $\hat{L}$, then we have:

$$\text{Akaike Information Criterion:} \quad \text{AIC} = 2k - 2\log(\hat{L}) \qquad (9.9)$$

$$\text{Bayesian Information Criterion:} \quad \text{BIC} = k\log(n) - \log(\hat{L}) \qquad (9.10)$$

As these information criteria contain both the maximum likelihood and also the number of parameters (and in BIC the number of data points) they measure not only the overall fit of the model, but also punish overfitting and over-complexity. Fig 9.2 displays the information criteria as functions of the number of components of our mixture model. It appears that approximately $K = 200$ minimises the BIC, though the AIC looks like it may continue to decrease. Due to the rapidly increasing computational complexity of high K we truncate our search here and use $K = 200$.
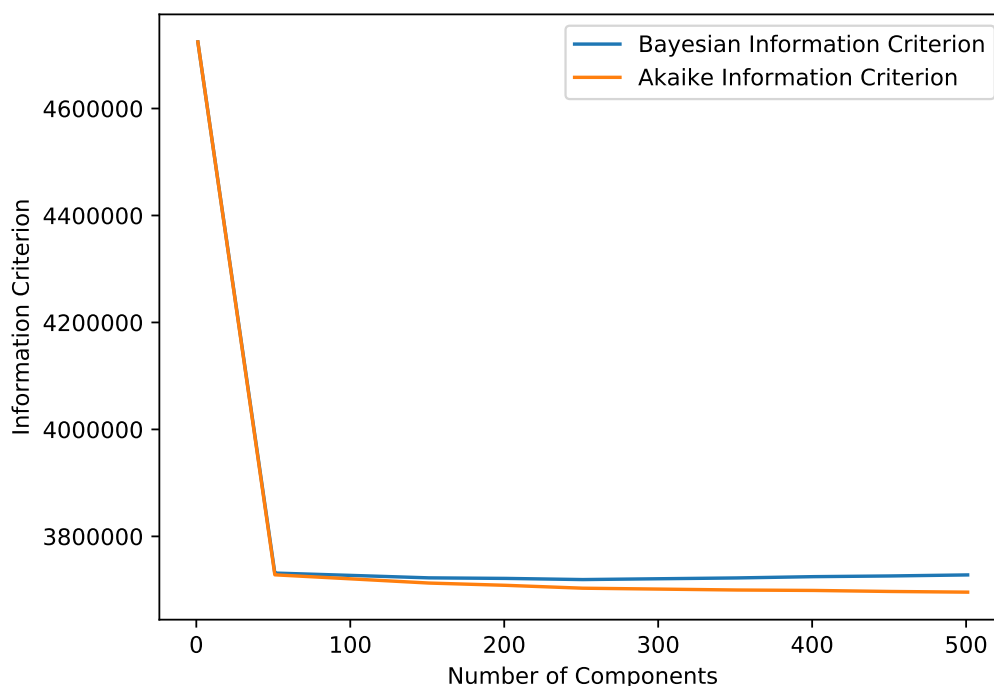
Fig. 9.2 The Akaike information criterion and Bayesian information criterion as a function of the number of components in the mixture model applied to the Higgs background dataset.

Fig 9.3 displays a scatter plot of the two leading photons in the background sample (left) and Higgs signal example (right) with a 200-component GMM trained entirely on the background sample overlaid on top. It is evident here that the signal events, by virtue of having a much higher density in the high transverse momentum area of the phase space would have a much lower likelihood of being sampled from the Gaussians present in the GMM.

Fig 9.4 shows two plots of anomaly score distributions for the Higgs example using a 200-component GMM. The top plot is illustrative and contains the anomaly score distribution of the background and the anomaly score distribution of a Higgs signal - it can be seen that the Higgs signal contains far more events of low probability and thus is much more anomalous. Note that in this case our
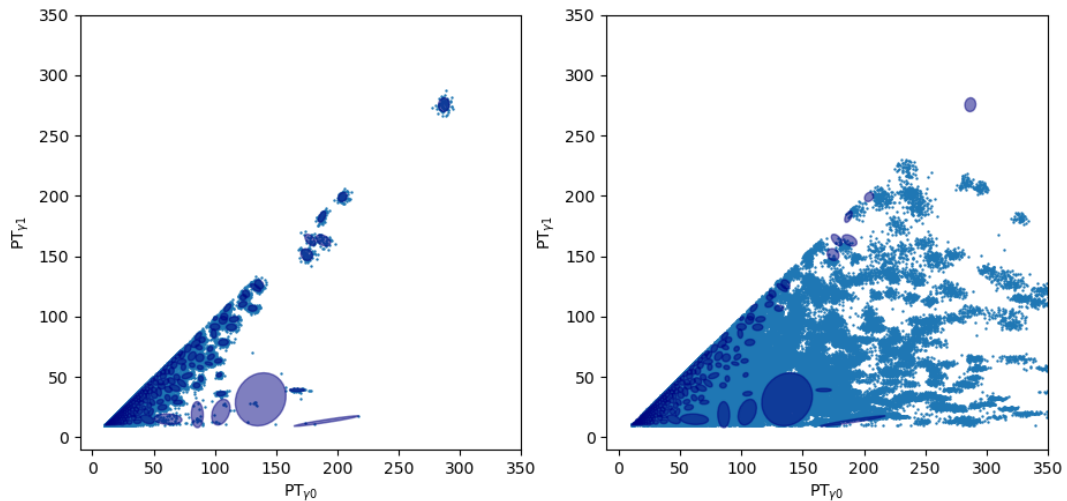
Fig. 9.3 A scatter plot of the two leading photons for the background (left) and signal (right) in the Higgs example. Overlaid on top is a 200-component GMM trained on the background sample.

definition of anomaly score is the probability of belonging to the background distributions and so a low anomaly score implies high anomalousness. The bottom plot contains the anomaly distribution of the background and the anomaly score distribution of the pseudo-data set constructed as described above, though with the injected signal cross section boosted to ensure it is visible for demonstration. These two distributions are almost identical, as expected, except for a excess in the 0%-0.05% bin of the distribution where the injected signal exists.

We will see Gaussian mixture models applied to a range of BSM models in chapter 10.

9.3 Isolation Forest

First outlined in Ref [210], "Isolation Forest" is an unsupervised learning algorithm that assigns each point in a dataset a value based on the ease with which it is isolated from the other points in the dataset. It is attractive due to its simple concept, linear time complexity and low memory requirement.

Given a set of data $X = \{\mathbf{x}_1, \mathbf{x}_2, \ \ldots \ , \mathbf{x}_n\}$ from a multivariate distribution, where each $\mathbf{x}_i$ is a vector with d dimensions, one first randomly chooses a feature $k \in \{1, ..., d\}$, and a "split value" p which lies between the maximum and minimum value of the feature k. Then all $\mathbf{x}_i$ of the dataset with $x_{ik} < p$ are placed in a set of points called X_l while if $x_{ik} \geq p$, it is placed in a set

Fig. 9.4 The anomaly score distributions generated by the Gaussian mixture model algorithm for (top) the signal and background and (bottom) the pseudo-data and background in a Higgs production decaying to two photons.

called X_r. This process is repeated recursively, selecting a new k each time, until *one* of the following stopping conditions is met:

- every data point $\mathbf{x}_i$ is isolated in its own set,

- all $\mathbf{x}_i$ in a given set are equal,

- a limit imposed on the number of splits is reached.

The sequence of splits generated are called *trees*, and the number of splits in them is called the *path length* of the tree. Each split is a *node* of the tree. Nodes that do not begin or end trees are *internal*, and those which do are *external*. By randomly selecting batches of size m from the dataset and constructing a

tree for the batch, we construct what is called a *forest*. The combination of many trees in this way improves stability and performance.

An anomaly is by definition an outlier, thus an anomaly should on average require a smaller number of splits to become isolated. The measure of anomalousness can then be defined via the average path length of the trees in the forest. This average path length is normalized using [210]:

$$c(m) = 2H(m-1) - \left(\frac{2(m-1)}{n}\right), \tag{9.11}$$

where n is the full dataset size, m is the size of a randomly sampled batch, and $H(m-1)$ indicates the harmonic number. The anomaly score of a point $\mathbf{x}_i$ is then defined as

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}}, \tag{9.12}$$

where $h(x)$ is the path length and $E(h(x))$ is the mean path length of all trees constructed for x. It can be seen from Eq. (9.12) that $s \approx 1$ implies a high level of anomalousness, whilst $s \approx 0$ indicates no anomaly at all. If the whole sample generates $s \approx 0.5$, we find that the entire sample is likely devoid of anomaly. An example of two trees, one for a non-anomalous point and one for an anomalous point, in two dimensions can be found in Fig 9.5. Note that the non-anomalous point required thirteen nodes (or splits) to isolate, while the anomalous point required only four, showing that their path lengths are vastly different.
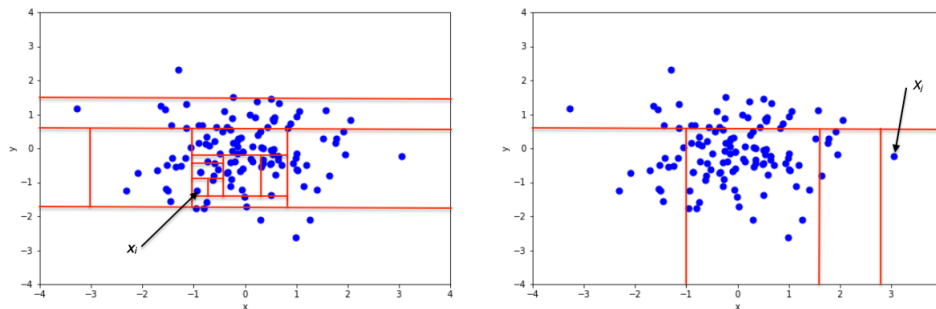
Fig. 9.5 An example of two trees formed in the isolation forest algorithm for an arbitrary 2D Gaussian distribution. Left: The isolation of a non-anomalous data point [211]. Right: The isolation of an anomalous data point [212].

In the top plot of Fig 9.6 we see the normalized anomaly score distributions generated by the isolation forest algorithm for the signal and background of the Higgs production decaying to two photons. The signal is clearly of higher anomaly score than the background. In the bottom plot we see the

background and pseudo-data anomaly score distributions normalized to cross section - though the injected signal model cross section was boosted to ensure it was visible. In the high anomaly areas of the plot we see excesses in the pseudo-data over the background corresponding to the same location as the signal distribution in the top plot.
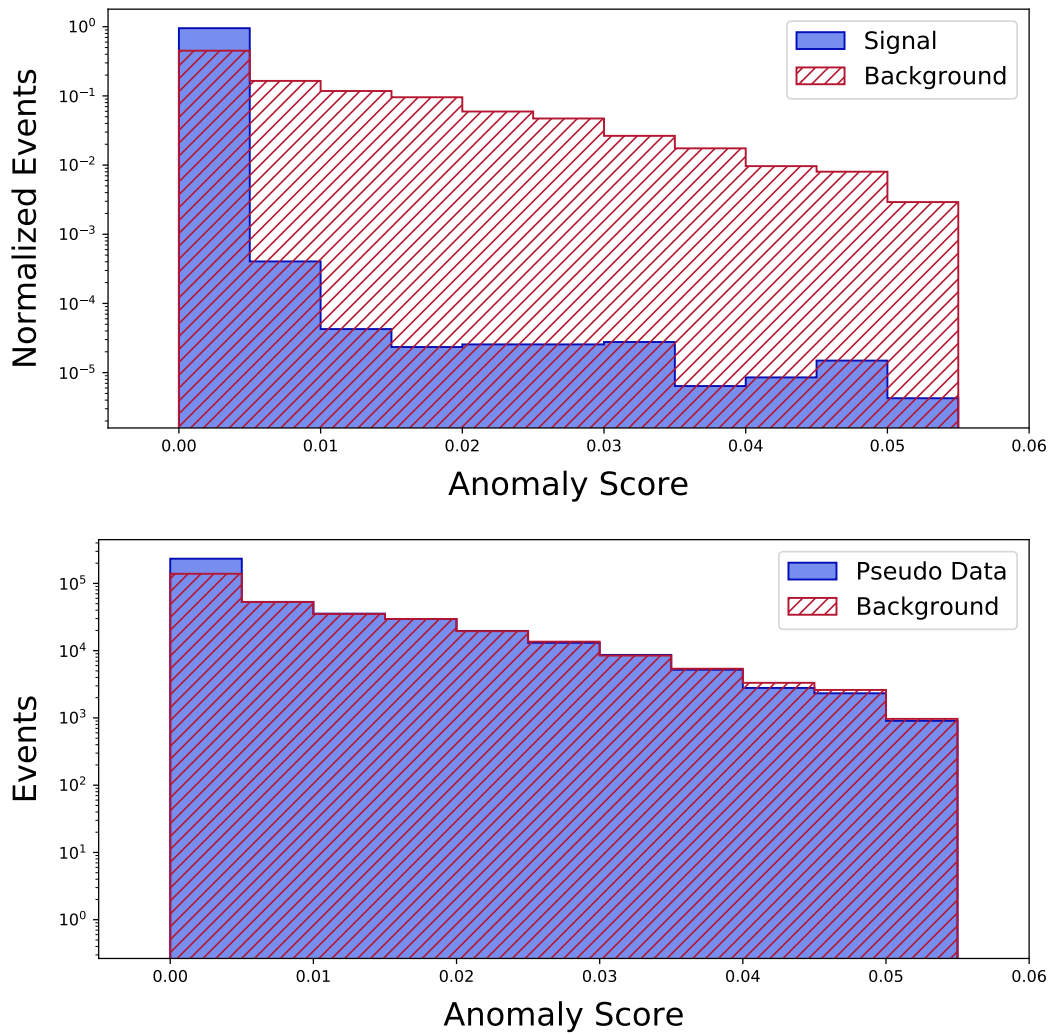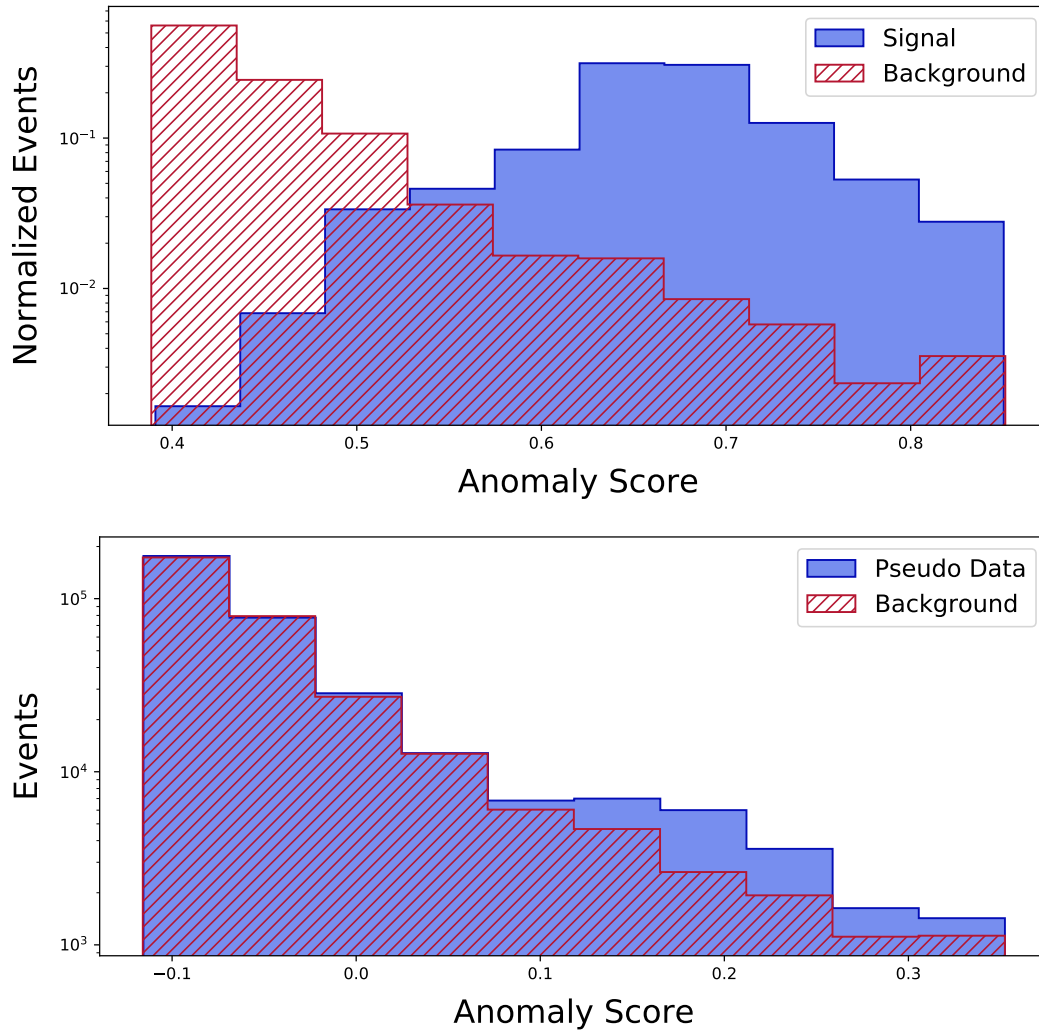

Fig. 9.6 The anomaly score distributions generated by the isolation forest algorithm for (top) the signal and background and (bottom) the pseudo-data and background in a Higgs production decaying to two photons.

9.4 Autoencoders

Autoencoders are a type of neural network (see chapter 5 for details) that attempt to return output that is as similar to their input as possible [213]. This is done to produce an encoding of the input data for the purposes of

dimensional reduction, feature learning and to produce generative models. In Fig 9.7 we see a simple example of an autoencoder.
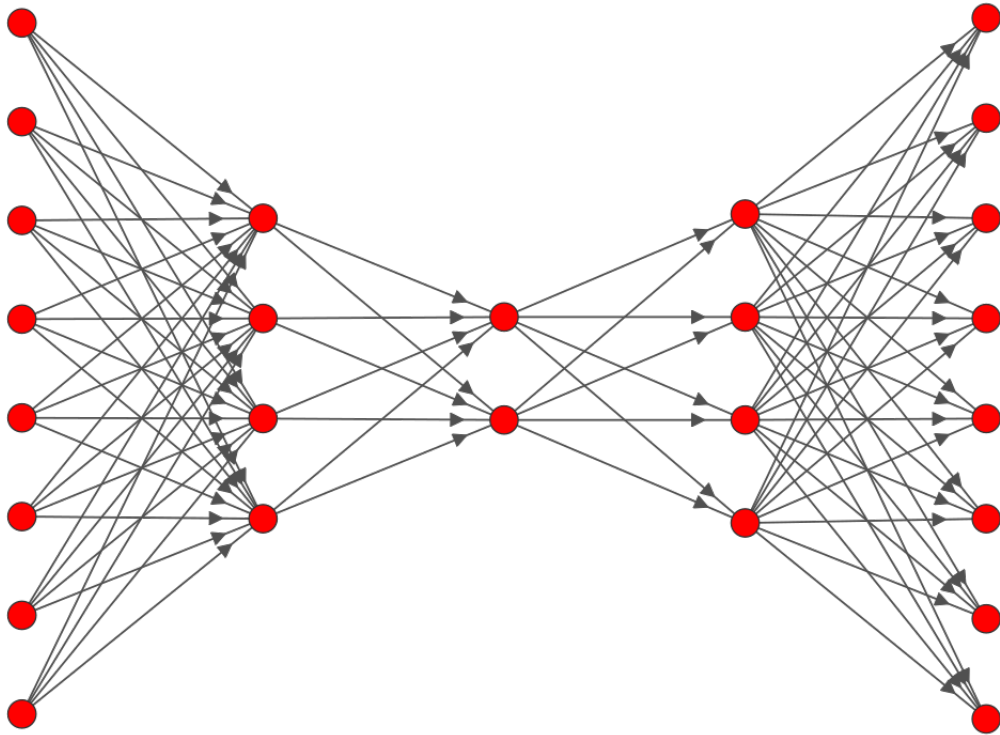
Fig. 9.7 An example of a simple autoencoder.

There are eight input nodes which condense to two nodes in the middle layer and which then expand back to eight output nodes - the exact architecture is in general case specific though of course there must be the same amount of input and output nodes and the middle layer almost always has the fewest nodes. The middle layer is known as the "latent space", and is the encoding of the input information to a lower dimensional representation. The error is called the "reconstruction loss" and is usually defined as the sum of mean squared errors between the input vectors and the output vectors in a training batch.

For our purposes we can use the reconstruction loss as a measure of the anomalousness of a physics event. For example, we could train our autoencoder on the background MC sample and then apply the trained model to the data. Any events which are very unlike the background MC sample will present with higher reconstruction loss as the autoencoder has never seen events like them.

The major downfall is that autoencoders are not in any way penalised for learning the identity function, or some function close to identity. This means that while reconstruction error will be very low, the model is trivial and unhelpful because it will be low on all inputs. There exist many methods to

regularise autoencoders to try and penalise this, however the most important is the promotion of autoencoders to "Variational Autoencoders" (VAEs).

9.4.1 Variational Autoencoders

Variational Autoencoders are very similar to autoencoders in the sense that they are neural networks with encoders and decoders, however mathematically they differ in that they can be considered to belong to a class of machine learning algorithms called generative models. That is, they attempt to understand the way in which data was generated or to approximate the distribution from which the data was sampled. In learning these relationships we can learn a great deal about representations of the data [214–216].

The major way in which VAEs differ from basic autoencoders is that before the latent space vector in an autoencoder we insert two vectors, a vector of means and a vector of variances of normal distributions. Rather than attempting to encode the input vector onto a lower dimensional vector of real numbers, we are now attempting to encode the input information onto a set of Gaussian distributions which when sampled from provide an output vector that is as similar to the input vector as possible. These distributions serve to "cover" the latent space with smooth distributions which offer a more faithful representation of smooth and continous data. They also inherently reduce the likelihood of the model approximating the identity function, and massively improve the performance of the architecture in general.

Theoretically, when a VAE is trained with a standard reconstruction loss it could simply learn Gaussians with extremely small variances (delta functions) or extremely disparate means (poor coverage of the latent space) and thus the VAE would effectively be a standard autoencoder and all benefits would be lost. Thus we regularise the reconstruction loss by penalising distributions that are different from standard normal distributions. This is done through the addition of the Kullback–Leibler divergence [217, 218] to the reconstruction loss (MSE used as an example) as seen below:

$$L = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 + \beta\sum_{j=1}^{K} D_{\mathrm{KL}}(N(\mu_j,\sigma_j)||N(0,1)) \qquad (9.13)$$

where, N is the number of input and output nodes and in general the Kullback-Leibler divergence for two distributions $P(x)$ and $Q(x)$ is defined as:

$$D_{\mathrm{KL}}(P||Q) = \int_x P(x)\log\left(\frac{P(x)}{Q(x)}\right)dx \qquad (9.14)$$

In our case $P(x)$ are the distributions in the latent space and $Q(x)$ is the standard normal distribution. This is a special case, and we can write the total Kullback-Leibler divergence loss, L_{KL}, for a latent space of K standard normal distributions as:

$$L_{\mathrm{KL}} = \sum_{i=i}^{K} \sigma_i^2 + \mu_i^2 - \log(\sigma_i) - 1 \qquad (9.15)$$

which is clearly minimised for $\sigma_i = 1$ and $\mu_i = 0$. The β factor is a parameter of $O(1)$ that scales the overall importance of the KL divergence to the reconstruction loss [219]. The default $\beta = 1$ is the definition for a standard VAE.

The Reparametrisation Trick

For a rigorous treatment of the reparametrisation trick see Refs [220] and [221]; here we will provide an intuitive explanation. When training a traditional autoencoder (or any ANN) via backpropagation, every node in the network is non-stochastic and there is no issue. However in the case of a VAE because we are sampling from a set of distributions in the latent space of the network we have introduced a set of nodes with stochasticity. We cannot compute a gradient of these nodes and backpropagation does not work. To fix this the reparametrisation trick was developed.

To produce the latent space vector $\mathbf{z}$ for the decoder to decode we calculate the following:

$$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \cdot \boldsymbol{\epsilon} \qquad (9.16)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are the mean and standard deviation of the Gaussian distributions in the latent space of the network and $\boldsymbol{\epsilon}$ is an introduced vector of stochastic variables sampled from standard normal distributions and independent of the network. Now backpropagation can be performed via $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ while stochasticity is maintained through $\boldsymbol{\epsilon}$ which does not need to be optimised. A schematic of this can be seen in Fig 9.8, where the yellow nodes represent input or independent nodes and the arrows represent the flow of information through the network. We can see that the values in the red bounding box on the left are the parts of the network being reparametrized using the reparametrization trick. The red bounding box on the right indicates that the Kullback-Leibler divergence is computed analytically via the known equation from the values in the latent space.
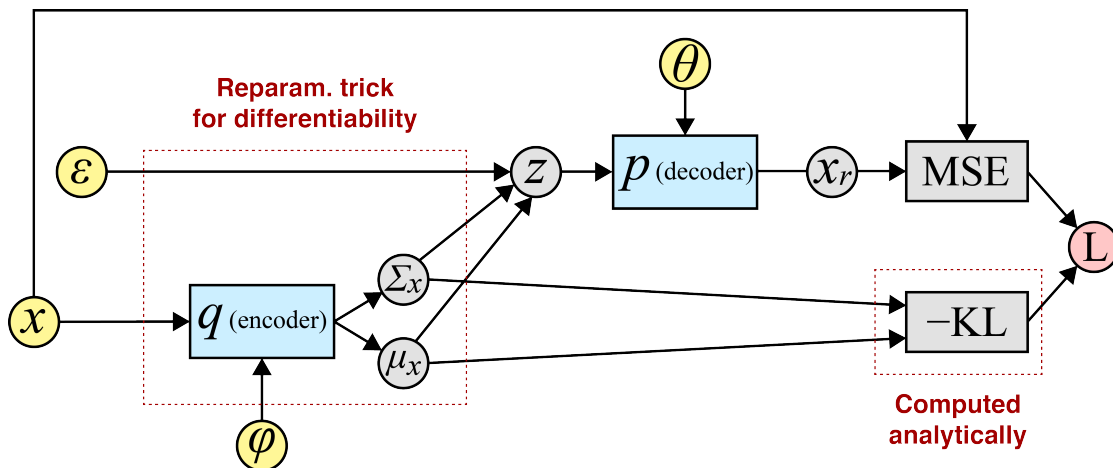
Fig. 9.8 A schematic of the reparametrisation trick [221].

9.4.2   Convolutional Variational Autoencoders

For chapter 11 of this thesis we employ a VAE to undertake anomaly detection for dark matter direct detection data. What is special about this study is that we do not use raw data for the machine learning inputs, we actually use images of the detector response.

An image is a set of pixels and computationally is represented in a number of ways - in our case it is an array of $(1,3)$ vectors with components in the range 0 to 255. These values represent the contribution of red, green and blue colour to the overall colour of the pixel on the screen and together form the image. The detector images we wish to train and test our VAE on are 64×64, which means in total we have 4096 vectors all of which have three components. This is an enormous number of inputs for our network that, at least with the computing resources available, is absolutely unfeasible to process. How do we reduce this input size to something that is acceptable?

Furthermore, if we simply flattened our array and inputted the result into our network we would be losing something fundamental about our image - adjacent pixels in an image are more related to each other than non-adjacent pixels. That is to say, if you imagine an image of a brown dog, standing on some green grass with some blue sky, the pixels are forming clusters of different colours and the boundaries between these clusters are incredibly important in detecting that that is an image of a dog. How do we ensure that this information is kept when inputting to our network?

Fortunately, both of these questions can be solved in the same way - "convolving" our input. This effectively boils down to sliding a window of height and width less than the size of the total image across and then down our image combining pixels into a single value. The size of the steps our window

takes across and then down the image is called the "stride size". There are a number of ways to do this, with two important ones being the use of a "kernel" and the other being "pooling" [222, 223].
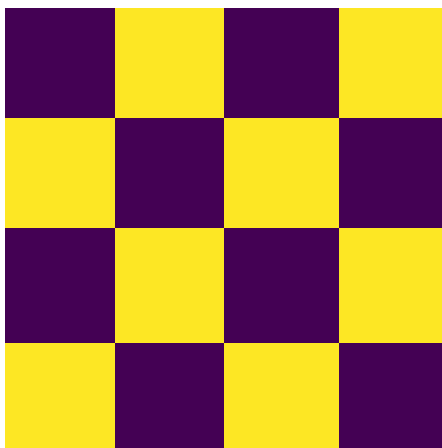
Fig. 9.9 A simple image.

Imagine a very simple 4×4 image such as that seen in Fig 9.9 which can be written as a matrix as:

$$I = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \tag{9.17}$$

Now, a kernel could be a matrix such as:

$$K = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{pmatrix} \tag{9.18}$$

where we call the center value of this kernel the "origin". We can then choose a stride length of 1 and construct the following matrix by aligning the origin of K with each value of I and summing the result of component-wise multiplying

the overlapping values. The resultant matrix is:

$$I' = \begin{pmatrix} -2 & 5 & -3 & 5 \\ 5 & -4 & 1 & -3 \\ -3 & 1 & -4 & 5 \\ 5 & -3 & 5 & -2 \end{pmatrix} \tag{9.19}$$

which looks like Fig 9.10. This is obviously a very trivial example, and is merely
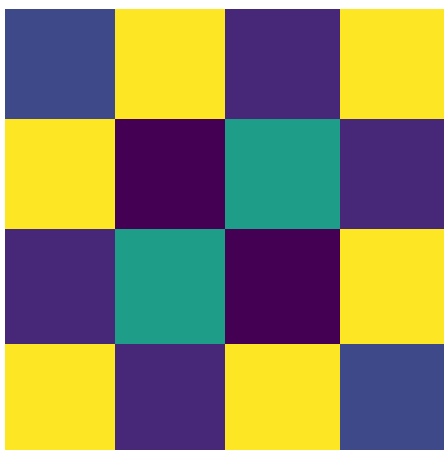
Fig. 9.10   The simple image seen in Fig 9.9 after applying a convolutional kernel.

done here to display the process. Now, if we wished to dimensionally reduce the original image we could simply choose a stride length greater than one, in which case the kernel would skip values of the original image leading to a smaller convolved image - for example if stride length is two this will reduce the dimension of the image by half.

Pooling is very similar, except instead of component-wise multiplying the kernel with values of the image we just take a simple function of the same elements - i.e. averaging them ("average pooling") or taking the maximum value of them ("max pooling"). This is useful for de-noising and extracting dominant features of an image, with max pooling being the superior choice.

In the context of convolutional neural networks we allow the values of the convolutional kernels to be parameters trained via backpropagation. By layering convolutional and pooling layers with different size kernels and stride lengths we can develop neural networks that provide extremely complex feature learning - we will see architectures like this in chapter 11.

9.5   ROC Curves

In the next chapter we will see Receiver Operating Characteristic (ROC) curves used to measure the performance of our anomaly detection algorithms. ROCs are a graphical way of displaying the performance of a classification model. The ROC curve is generated by plotting the "true-positive rate", the percentage of times the classifier correctly identifies a signal event as signal, vs the "false-positive rate", the percentage of times the classifier incorrectly identifies a background event as signal as a function of the discrimination threshold of the classifier [224]. The discrimination threshold is the value of the classifiers output, in our case this would be the anomaly score of a given algorithm, that we have chosen such that if the output is below the threshold we classify the event as background and if it is above we classify it as signal. Thus, a classifier that consists of purely guessing whether an event is signal or background will have a ROC curve that is a diagonal line from the origin to $(1, 1)$, while a classifier that is very effective will have a steep curve pushed to the upper left corner of the plot.
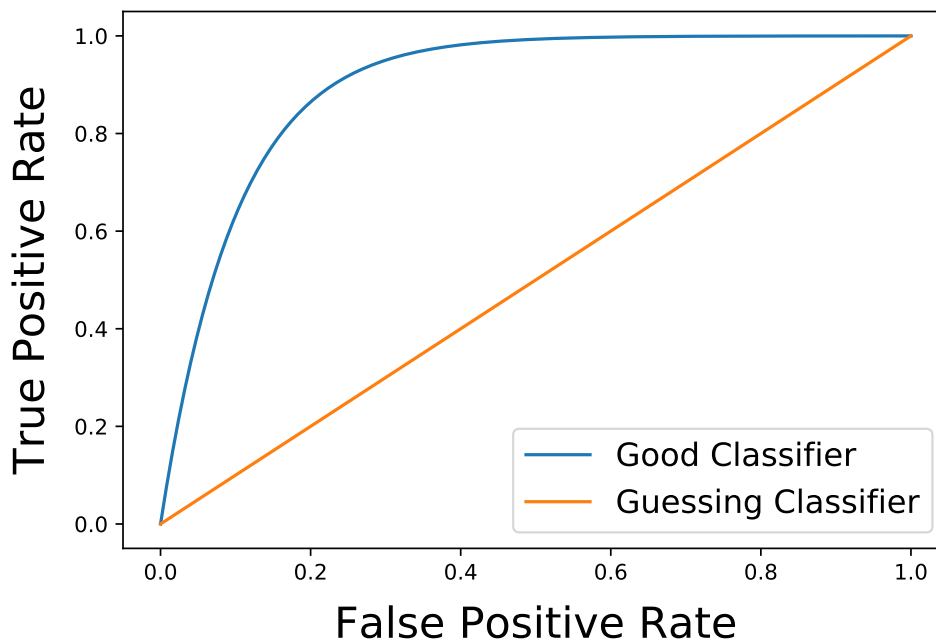
Fig. 9.11   An example of a ROC curve for a effective classifier and for a guessing classifier.

In Fig 9.11 one can see two simulated ROC curves, one for a effective classifier and one for a classifier that is guessing. Note that these are simply for illustration. In the next chapter we will see many real ROC curves for a

range of different machine learning models, so the features of the ROC curves for effective and ineffective machine learning models are important to recall.

Conclusion

In this chapter, we explored modern analysis techniques that attempt to detect anomalies in data without requiring a definition of specific signal topologies. Specifically, the isolation forest, Gaussian mixture model and variational autoencoder algorithms were detailed. In the coming chapters, we will see these algorithms applied to physically realistic scenarios in two different studies.

Chapter 10

Anomaly Detection at the LHC

In this chapter we will explore a study currently in peer review, see Ref [206], that seeks to apply the methodologies seen in chapter 9 to realistic BSM supersymmetric models. We will first discuss the unique aspects of the VAE that is employed then cover the benchmarks themselves and the details of our LHC simulations. Then finally, we will explore the effectiveness of the anomaly detection algorithms at extracting signals of varying cross sections and degeneracies with the SM background.

10.1   VAE Architecture

The architecture used here is defined as seen in Refs [225] and [206]. It consists of layers of 512, 256 and 128 neurons in the encoder part of the network, followed by a latent space of 26 nodes representing 13 means and 13 standard deviations, then 128, 256 and 512 nodes in the decoder part of the network. All layers are fully connected and the exponential linear unit activation function is used.

As each event does not necessarily contain the same number of particles each input does not necessarily have the same dimensionality. To handle this, all inputs are zero padded to match the largest input in the set. The continuous variables are then normalised by dividing by the largest value of each feature. The categorical input elements (particle type) are separated from the continuous input elements (pT, η and ϕ) and finally the number of particles in the event

is appended as its own input element. This creates an input as follows:

$$\mathbf{x} = \left(N, \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix}, \begin{bmatrix} (\mathrm{pT}, \eta, \phi)_0 \\ (\mathrm{pT}, \eta, \phi)_1 \\ \vdots \\ (\mathrm{pT}, \eta, \phi)_n \end{bmatrix} \right) \tag{10.1}$$

where N is the number of particles in the event and c_i is the categorical variable converted via a one-hot encoding.

Now, the reconstruction loss is defined in three parts: the mean squared error for the number of particles N and the continuous data, and the categorical cross-entropy is used for the categorical data. The total loss function is defined as:

$$L = 100\beta(x_n - y_n)^2 \tag{10.2}$$

$$- \frac{10\beta}{d_c} \sum_i^{d_c} (x_{c,i} \log(y_{c,i}) + (1 - x_{c,i}) \log(1 - y_{c,i})) \tag{10.3}$$

$$+ \frac{\beta}{d_r} \sum_i^{d_r} (x_{r,i} - y_{r,i})^2 \tag{10.4}$$

$$+ (1 - \beta) \sum_i^{d_z} D_{\mathrm{KL}}(\mathcal{N}(\hat{\mu}_i, \hat{\sigma}_i), \mathcal{N}(0, 1)) \tag{10.5}$$

where x_n is the true number of objects and y_n is the reconstructed number of objects, d_c is the dimensionality of the categorical data, $x_{c,i}$ is the i-th categorical label, $y_{c,i}$ is the i-th predicted categorical label, d_r is the dimensionality of the continuous elements, $x_{r,i}$ the i-th continuous data element and $y_{r,i}$ the i-th reconstructed continuous data element, d_z is the dimensionality of the latent space and μ_i and σ_i are the mean and standard deviation of the i-th latent space Gaussian respectively. β is a scaling factor that sets the overall impact of the Kullback-Leibler divergence as seen in chapter 9. In this case, we use $\beta = 10^{-3}$.

This loss function is much more advanced and far more successfully captures the properties of physics events than the basic VAE loss described in the previous chapter.

10.2   Benchmark 1: Gluinos

The first set of benchmarks contain pair-produced gluinos which then decay to a boosted top-quark pair and the LSP. The gluinos have mass in the range $1 - 2.2$ TeV in steps of 200 GeV and the LSP has mass 1 GeV. The branching ratio is assumed to be 100% $(\mathrm{BR}(\tilde{t} \to t\bar{t}\tilde{\chi}^0_1) = 100\%)$. More details can be found in Tab 10.2.

10.3   Benchmark 2: Stop Quarks

The second set of benchmarks contain pair-produced stop quarks followed by decay to top quarks and the LSP. There are four mass schemes in use here with masses (LSP, Stop) = $(20, 220)$, $(100, 300)$, $(100, 400)$, $(100, 800)$ GeV. From left to right we see a dropping cross section but also an increasing level of disparity with the background signature. That is to say, while the lightest stop mass point has by far the highest cross section it is almost degenerate in mass with the top quark and as such will have very similar final state kinematics. More details can be found in Tab 10.1.

Process	Process ID	σ (pb)	N_{tot} ($N_{\mathrm{10fb}^{-1}}$)
$pp \to \tilde{g}\tilde{g}$ (1 TeV)	Gluino 01	0.20	50000 (2013)
$pp \to \tilde{g}\tilde{g}$ (1.2 TeV)	Gluino 02	0.05	50000 (508)
$pp \to \tilde{g}\tilde{g}$ (1.4 TeV)	Gluino 03	0.014	50000 (144)
$pp \to \tilde{g}\tilde{g}$ (1.6 TeV)	Gluino 04	0.004	50000 (44)
$pp \to \tilde{g}\tilde{g}$ (1.8 TeV)	Gluino 05	0.001	50000 (14)
$pp \to \tilde{g}\tilde{g}$ (2 TeV)	Gluino 06	4.8×10^{-4}	50000 (5)
$pp \to \tilde{g}\tilde{g}$ (2.2 TeV)	Gluino 07	1.7×10^{-4}	50000 (2)
$pp \to \tilde{t}_1\tilde{t}_1$ (220 GeV), $m_{\tilde{\chi}^0_1} = 20$ GeV	Stop 01	26.7	500000 (267494)
$pp \to \tilde{t}_1\tilde{t}_1$ (300 GeV), $m_{\tilde{\chi}^0_1} = 100$ GeV	Stop 02	5.7	500000 (56977)
$pp \to \tilde{t}_1\tilde{t}_1$ (400 GeV), $m_{\tilde{\chi}^0_1} = 100$ GeV	Stop 03	1.25	250000 (12483)
$pp \to \tilde{t}_1\tilde{t}_1$ (800 GeV), $m_{\tilde{\chi}^0_1} = 100$ GeV	Stop 04	0.02	250000 (201)

Table 10.1 A list of our BSM benchmark points, with the cross section and number of events in both the generated count and also scaled to 10fb^{-1} of luminosity at 13 TeV CoM energy [226].

10.4   Backgrounds

For both sets of benchmarks, we utilise the background processes and MC described in table 2 of contribution 23 of Ref [226]. This is a huge dataset of many common SM backgrounds which serves as a fantastic background model

to train our unsupervised learning algorithms on. The full list of SM processes can be found in Ref [226] but is transcribed here for ease of use in Tab 10.2.

Process	Process ID	σ (pb)	N_{tot} ($N_{10\text{fb}^{-1}}$)
$pp \to jj$	njets	$19718_{H_T>600\text{GeV}}$	415331302 (197179140)
$pp \to W^{\pm}(+2j)$	w_jets	$10537_{H_T>600\text{GeV}}$	135692164 (105366237)
$pp \to \gamma(+2j)$	gam_jets	$7927_{H_T>600\text{GeV}}$	123709226 (79268824)
$pp \to Z(+2j)$	z_jets	$3753_{H_T>600\text{GeV}}$	60076409 (37529592)
$pp \to t\bar{t}(+2j)$	ttbar	541	13590811 (5412187)
$pp \to W^{\pm}t(+2j)$	wtop	318	5252172 (3176886)
$pp \to W^{\pm}\bar{t}(+2j)$	wtopbar	318	4723206 (3173834)
$pp \to W^+W^-(+2j)$	ww	244	17740278 (2441354)
$pp \to t + \text{jets}(+2j)$	single_top	130	7223883 (1297142)
$pp \to \bar{t} + \text{jets}(+2j)$	single_topbar	112	7179922 (1116396)
$pp \to \gamma\gamma(+2j)$	2gam	47.1	17464818 (470656)
$pp \to W^{\pm}\gamma(+2j)$	Wgam	45.1	18633683 (450672)
$pp \to ZW^{\pm}(+2j)$	zw	31.6	13847321 (315781)
$pp \to Z\gamma(+2j)$	Zgam	29.9	15909980 (299439)
$pp \to ZZ(+2j)$	zz	9.91	7118820 (99092)
$pp \to h(+2j)$	single_higgs	1.94	2596158 (19383)
$pp \to t\bar{t}\gamma(+2j)$	ttbarGam	1.55	95217 (15471)
$pp \to t\bar{t}Z$	ttbarZ	0.59	300000 (5874)
$pp \to t\bar{t}h(+1j)$	ttbarHiggs	0.46	200476 (4568)
$pp \to \gamma t(+2j)$	atop	0.39	2776166 (3947)
$pp \to t\bar{t}W^{\pm}$	ttbarW	0.35	279365 (3495)
$pp \to \gamma\bar{t}(+2j)$	atopbar	0.27	4770857 (2707)
$pp \to Zt(+2j)$	ztop	0.26	3213475 (2554)
$pp \to Z\bar{t}(+2j)$	ztopbar	0.15	2741276 (1524)
$pp \to t\bar{t}t\bar{t}$	4top	0.0097	399999 (96)
$pp \to t\bar{t}W^+W^-$	ttbarWW	0.0085	150000 (85)

Table 10.2 A list of SM processes in the background sample with cross section and number of events in both the raw generated count and also scaled to 10fb^{-1} of luminosity at 13 TeV CoM energy [226].

10.5 Details of the Monte Carlo Simulation

The events were generated at leading order using `Madgraph v6.3.2` with the NNPDF PDF set [227] and the 5 flavour proton scheme at 13 TeV CoM energy. The events were then fragmented/hadronised utilising `Pythia 8.2` with MLM matching. Detector effects were included using `Delphes 3` with a modified version of the ATLAS detector card.

10.6 Results

The important results of this study come in two parts, the first being the standard approach of applying various anomaly detection algorithms - isolation forest, Gaussian mixture model, autoencoder and VAE - to the raw data of each event. The second set of results are motivated by the fact that the latent space of the VAE is a lower dimensional encoding of the input data; theoretically it contains all the required information about the input data to still allow for decoding, but no excess information. It is believed that other anomaly detection algorithms, which often suffer from the curse of dimensionality, will perform better when applied to the latent space of the VAE. So we do exactly this and apply the isolation forest, Gaussian mixture model and autoencoder to the latent space of the VAE and compare the result to the results on raw data.

10.6.1 Raw Data

Fig 10.1 displays the anomaly score distributions for the background and signal using the four different algorithms on the 1 TeV gluino benchmark. Note that it has a logged y-axis and the background distributions while plotted individually are stacked and scaled correctly to cross section. This plot provides measure of how well the signal is identified as anomalous by observing how far to the right the signal curve is relative to the background. In this case the extreme kinematic differences of the particles produced by their very high mass gluino (relative to the SM) allow for easy identification of signal anomalies, and the performance is only restricted by the low signal cross section. In Fig 10.2, we see the same plot for the 404 GeV stop quark pair benchmark. In this case, the model is far less kinematically different due to the lower stop mass, but still sufficiently anomalous as the stop quark and top quark mass difference is still high.

Figs 10.3 and 10.4 display the ROC curve for the isolation forest, Gaussian mixture model, autoencoder and VAE applied to the gluino and stop benchmarks respectively. The $Z_1 00$ variable present in the legend is the z-score of cutting on the anomaly score distribution at the the optimal bin that still contains more than 100 background events. The inset plot is a magnification at the value of the ROC for this $Z_1 00$ cut on the VAE.

It is clear that the VAE is consistently the best anomaly detection algorithm in almost all benchmarks, though in the stop benchmarks Gaussian mixture models appear to outperform it. In the gluino benchmarks, the isolation forest performs quite poorly though it performs much better on the stop benchmarks.

It can be seen that the discrimination ability of every algorithm drops from gluino 01 to gluino 06 and stop 01 to stop 04, which can be attributed to the steeply falling cross section. However, the VAE still has no trouble correctly identifying the anomalies in the data.

This implies that while the VAE is a powerful tool at finding anomalies it may very well be constrained in its usefulness by low cross sections, though this is also true of many other techniques, both classical and machine learned.
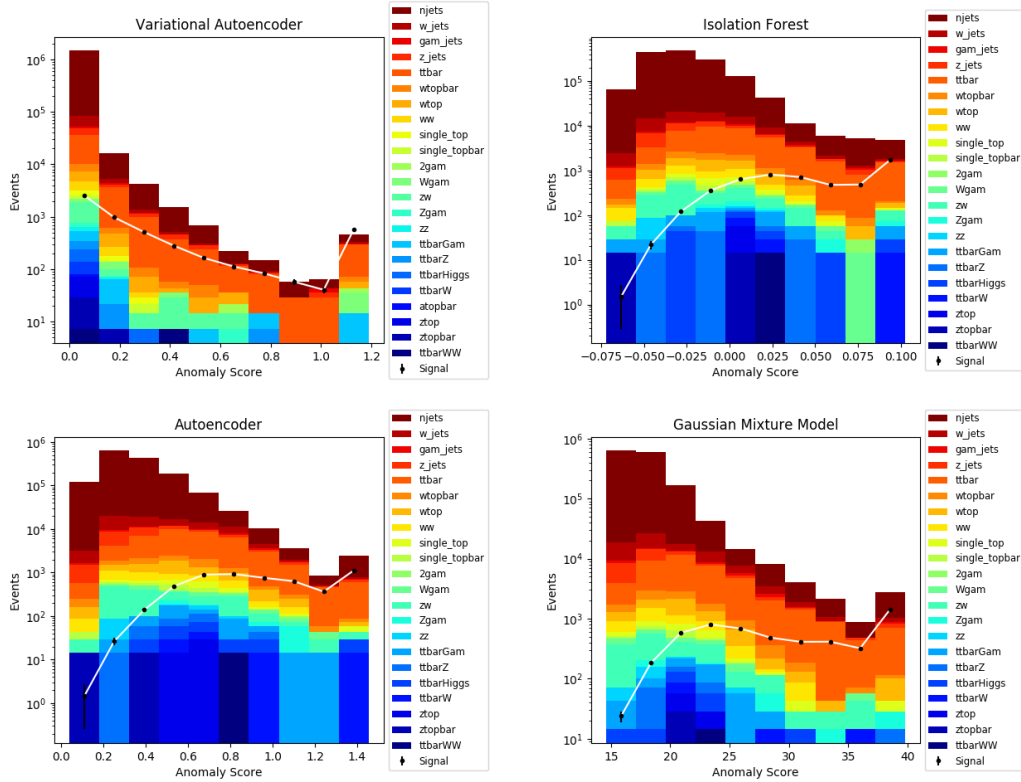
Fig. 10.1 Anomaly score histograms derived from various algorithms for the Gluino 01 benchmark - a 1 TeV Gluino signal - with 36 fb^{-1} of luminosity [206].

10.6.2 Latent Space

Figs 10.5 and 10.6 display the ROC curve for the isolation forest, Gaussian mixture model and basic autoencoder applied to the latent space of the VAE, as well as the performance of the VAE alone for the gluino and stop benchmarks respectively. The $Z_1 00$ variable present in the legend is the z-score of cutting on the anomaly score distribution at the the optimal bin that still contains more than 100 events. The inset plot is a magnification at the value of the ROC for this $Z_1 00$ cut on the VAE.
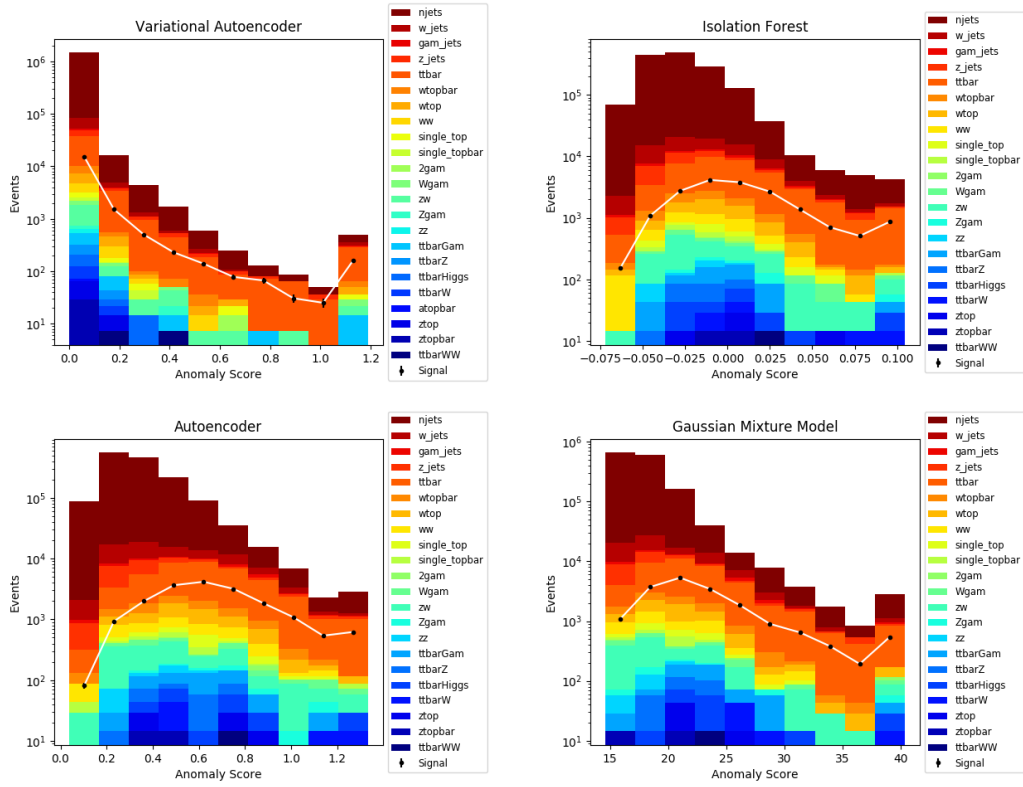
Fig. 10.2 Anomaly score histograms derived from various algorithms for the Stop 01 benchmark - a 400 GeV stop signal - with 36 fb^{-1} of luminosity [206].

It is clear that our hypothesis was in fact correct, using the VAE latent space as input to the other anomaly detection algorithms does greatly improve their performance, almost to the extent that they become equivalent to the VAE performance on the gluino benchmarks. In the Stop 01, Stop 02 and Stop 03 benchmarks the autoencoder and GMM actually outperform the VAE, though this is not true for Stop 04 perhaps because of the reduced cross section. This is both proof that the VAE is extremely effective at encoding background events to the latent space, and that the major stumbling block for the isolation forest, Gaussian mixture models and static autoencoder is the dimensionality of the input data.

Conclusion

The key result of this work is the power of the Variational Autoencoder in effectively encoding the SM backgrounds. This study utilised a huge swathe of background processes which cover the vast majority of non-negligible backgrounds and an enormous amount of events making it a very realistic representation of the performance of anomaly detection on ATLAS data. It is a powerful
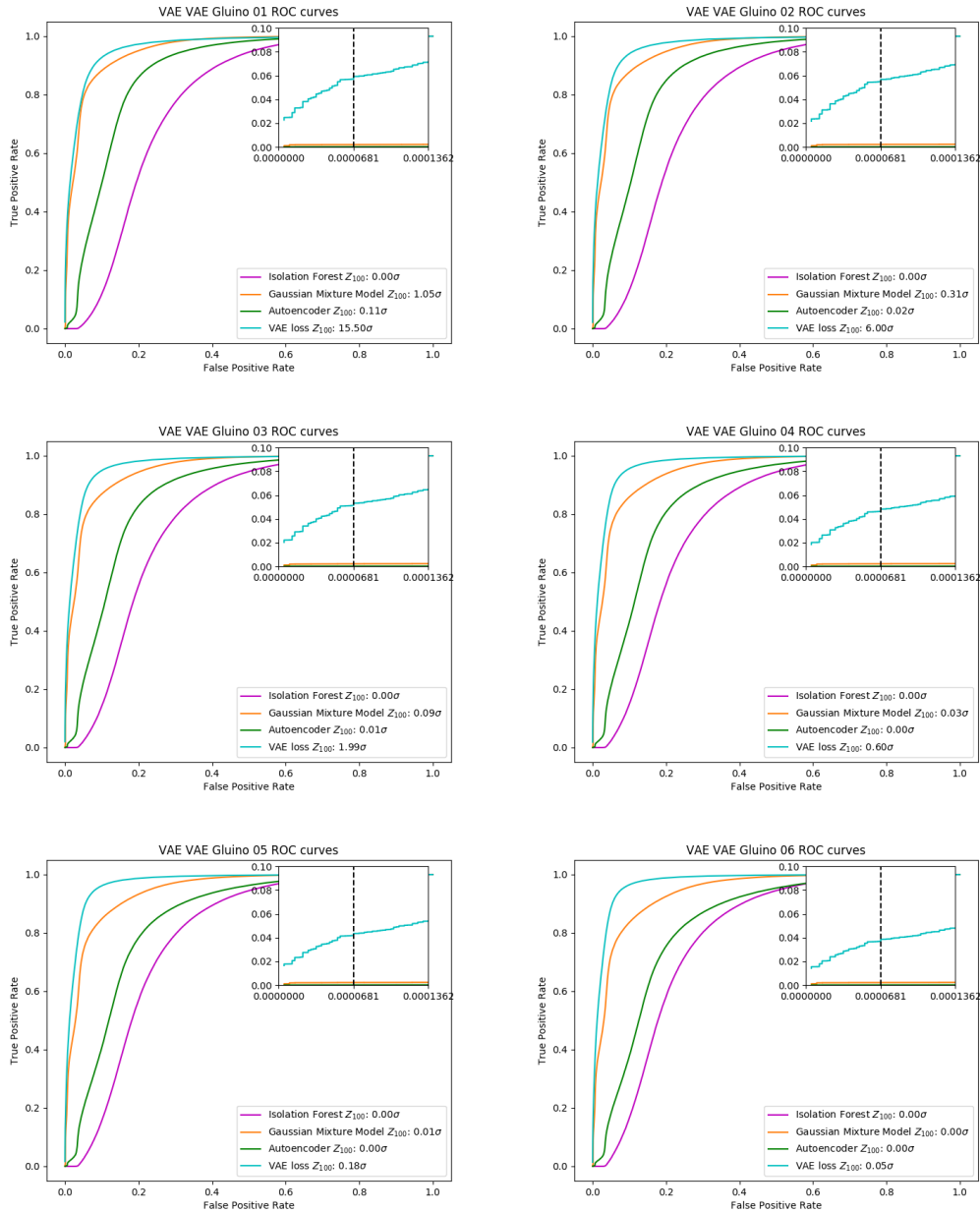
Fig. 10.3 ROC curves for the gluino benchmarks for various anomaly detection algorithms applied to original 4-vector data [206].

motivation for the adoption of signal agnostic searches by experimentalists using unsupervised machine learning.

Furthermore, while the isolation forest and Gaussian mixture model algorithms did not perform as well as the VAE, they are also much simpler models with lower training time and memory consumption making them useful prototyping and testing tools. The static autoencoder was overall not particularly useful as it sits in the middle ground of performance and high complexity, but provides an interesting contrast to the more advanced VAE architecture.
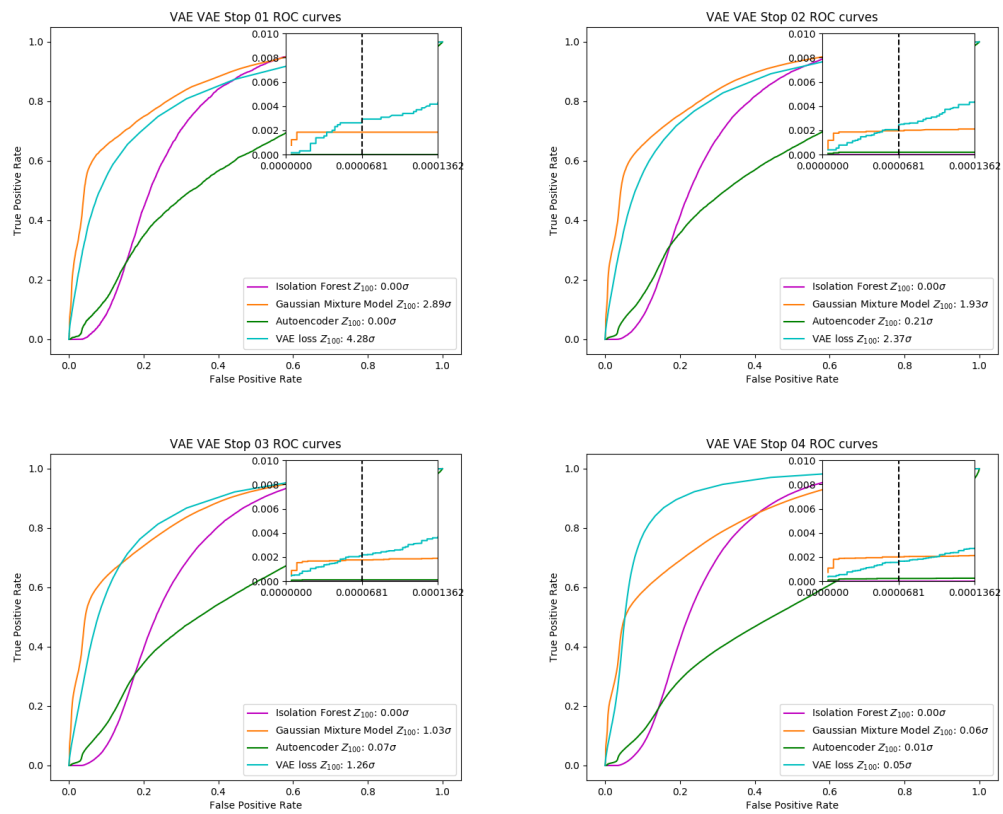
Fig. 10.4 ROC curves for the stop benchmarks for various anomaly detection algorithms applied to original 4-vector data [206].

In the next chapter we will see another study utilising a VAE, except this time it is a Convolutional VAE (CVAE) applied to images comprised of the detector response of the planned XENONnT dark matter direct detection experiment.
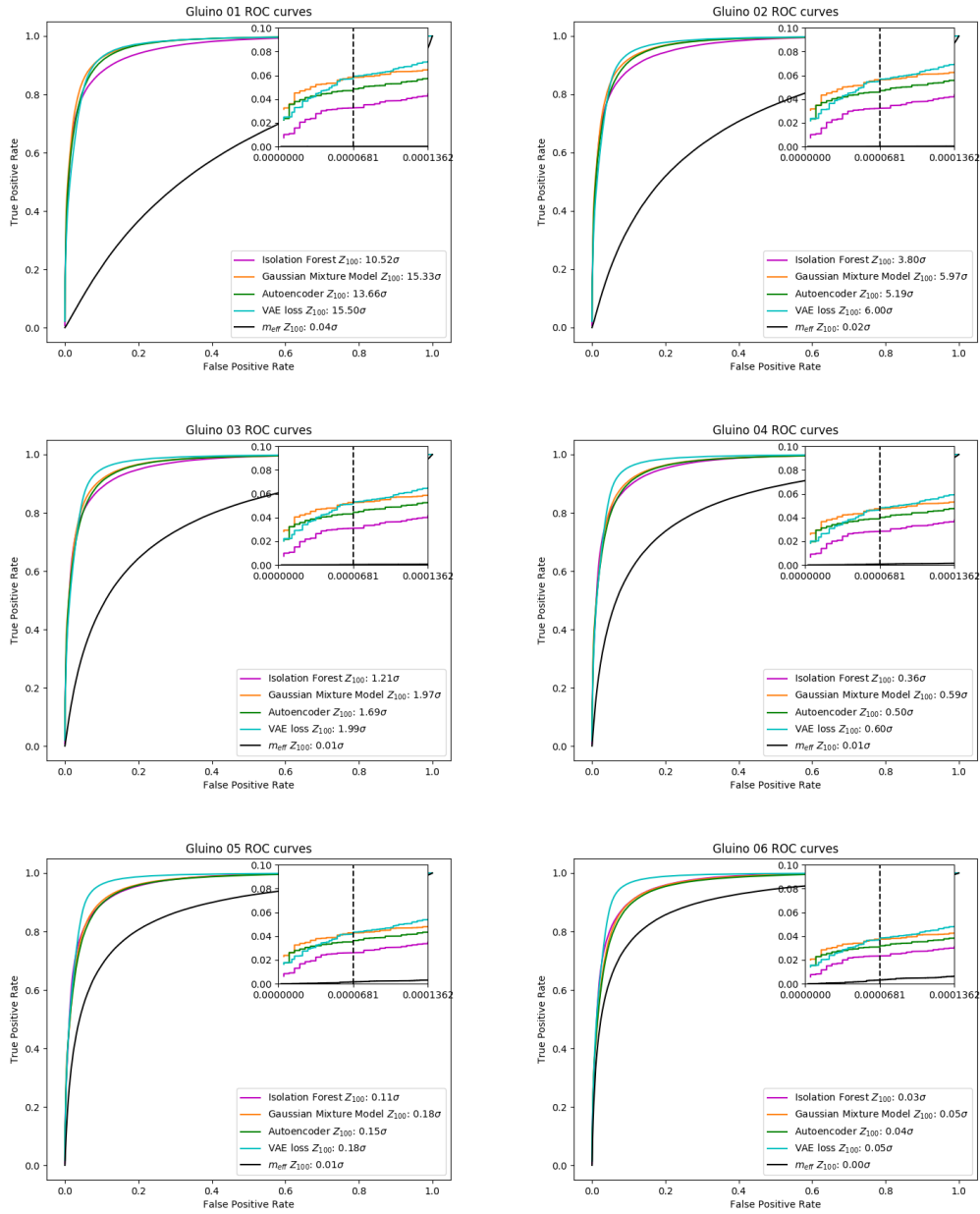
Fig. 10.5 ROC curves for the gluino benchmarks for various anomaly detection algorithms applied to the VAE latent space [206].
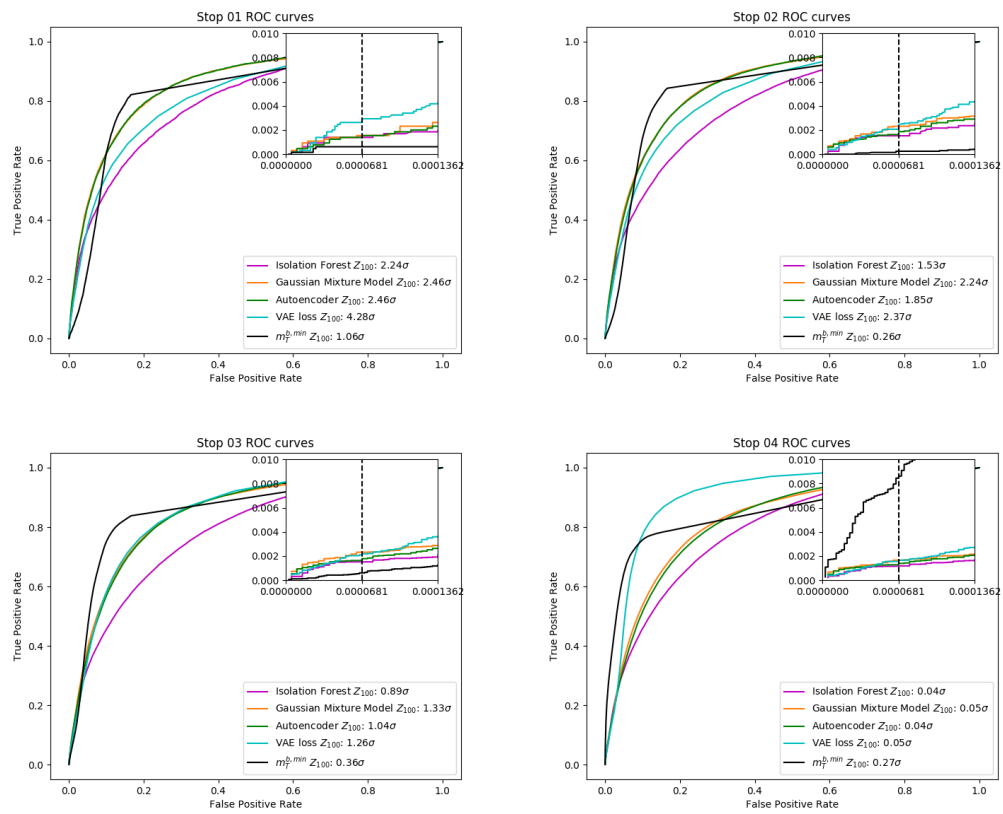
Fig. 10.6 ROC curves for the stop benchmarks for various anomaly detection algorithms applied to the VAE latent space [206].

Statement of Authorship

Title of Paper	Machine learning for dark matter direct detection I: Unsupervised analysis techniques
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☒ Unpublished and Unsubmitted work written in manuscript style
Publication Details	This paper is very close to being submitted for peer review.

Principal Author

Name of Principal Author (Candidate)	Riley Patrick
Contribution to the Paper	• Conceptualization and direction of project. • Design and implementation of the custom deep learning algorithm (convolutional variational autoencoder). • Preparation and editing of the manuscript, especially the deep learning architecture, analysis, results and conclusion sections.
Overall percentage (%)	60%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 09/04/2021

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i.   the candidate's stated contribution to the publication is accurate (as detailed above);

ii.   permission is granted for the candidate in include the publication in the thesis; and

iii.   the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Juan Herrero-Garcia
Contribution to the Paper	10%.
Signature	Date 03/12/2020

Name of Co-Author	Andre Scaffidi
Contribution to the Paper	• Conceptualization and direction of project. • Simulation of signal and background images. • Preparation and editing of the manuscript.
Signature	Date 09/04/2021

Please cut and paste additional co-au

Chapter 11

Publication 4: Convolutional Variational Autoencoders for Dark Matter Direct Detection

Abstract

The dark matter sector remains completely unknown. It is therefore crucial to keep an open mind regarding its nature and possible interactions. Focusing on the case of Weakly Interacting Massive Particles, in this work we make this general philosophy more concrete by applying unsupervised machine learning techniques to dark matter direct detection experiments. We do this by encoding and decoding the signal images in the photomultipliers of the XENONnT experiment with a convolution variational autoencoder. Any increase in the reconstruction loss with respect to just background is interpreted as an anomaly which deserves further investigation. We find that indeed our designed network is able to discriminate background with and without a dark matter signal on top at the 95% CL. Once a dataset is classified as anomalous, supervised analysis techniques become useful, which are intended to be analysed in a subsequent study.

11.1 Introduction

The nature of dark matter (DM) is one of the most puzzling problems of the Standard Model of particle physics. In recent decades a lot of work has been developed on the theory side to propose viable DM candidates. Among them, Weakly Interacting Massive Particles (WIMPs), although they have lost some momentum due to the absence of new physics signals at the LHC (e.g. no news

of low-energy Supersymmetry), remain a well-motivated option. Similarly, from the experimental side, impressive sensitivity has been reached, for example looking for the scattering of WIMPs in underground detectors, in what is known in the community as DM direct detection (DD). In the coming years an unprecedented experimental effort will be able to test large regions of the DM parameter space, especially with ton-scale noble gas experiments using xenon (PandaX [228], XENON1T [229], LZ [230] and ultimately DARWIN [231]) and argon (DEAP-3600 [232], DarkSide and Argo [233], ArDM [234]). In addition, the DAMA annual modulation signal will be tested in a model-independent way with several experiments using sodium iodine (COSINE [235], ANAIS [236], COSINUS [237] and SABRE [238, 239] in the southern hemisphere).

Model independent searches are an area of research that attempts to parametrise the concept of anomalous data. The key to this approach is to train a model to successfully capture the information contained within a background dataset such that data points that don't conform to the properties of the background can be identified. In this study we train a convolutional variational autoencoder (CVAE) on a data set of composite images containing the detector response and S1/S2 channels of simulated electron recoil events from the XENONnT dark matter detector. We then apply the trained CVAE to simulated WIMP events and show that the reconstruction loss for these events is on average higher, serving as a measure of anomalousness. While unsupervised methods such as this will likely never outperform supervised classification methods they are still powerfully motivated by their ability to identify events which are not background-like in a way that is agnostic to all signal models. In a context such as dark matter where the parameter space of possible models is extremely large the ability to find any anomalous events is helpful in motivating future studies, as well as identifying unknown systematic errors.

Machine learning approaches are a growing area of research in the physics community however their applications in dark matter are to some extent lagging behind. Supervised ML was applied in Ref [240] where a convolution neural network was applied to XENON1T detector response images of WIMP signal events and nuclear recoil background events achieving classification accuracy of above 90%. In Ref [241] a two-phase CNN architecture undertaking first classification and then regression was used to map 3D galaxy distributions to their underlying dark matter distribution. In Ref [242] one can see gradient boosted trees used to model dark matter halo formation. Similarly, in Ref [243] a deep CNN architecture is used to model halo formation.

In the collider physics context many studies have been published in recent years that attempt to parametrise the concept of anomalousness in data using unsupervised machine learning - see Refs [201–203, 244, 204, 205, 245–249, 225, 250, 206]. These studies employ a variety of different methods in a number of different ways, of particular relevance to this paper are Refs [204] and [206] which employ a variational autoencoder on simulated events of the CMS and ATLAS detectors respectively.

This paper will be structured as such: Section 11.2 will review the methodology by which time projection chamber (TPC) events are reconstructed at XENONnT and Section 11.3 will then cover the Monte Carlo methods by which events are simulated and combined into images for analysis. Section 11.4 will provide background on the concept of a variational autoencoder and convolution, followed by Section 11.5 where we present our specific convolutional variational autoencoder architecture. Finally in Section 11.6 we present the results of this study and conclude in Section 11.7.

11.2   TPC Event Reconstruction in the XENON Experiment

In this work we will simulate background and signal events for the XENONnT experiment located in the Gran Sasso National Laboratory. XENON1T has been running for some time and is to be superseded by XENONnT. The latter uses a dual-phase time projection chamber with 8 tons of natural xenon, of which around 6 tons constitute its fiducial volume. It consists of a liquid xenon target with a gaseous phase on top, with an applied electric field throughout the detector. The detector has two arrays of around 250 photo-multipliers in each of the top and bottom layers. It is embedded in a water tank, with muon and neutron vetos.

The main electronic background comes from gamma rays, being generated in the cryostat and in the PMTs. The detection technique makes use of both the prompt scintillation signal or S1, and the scintillation created from the drifted electrons due to the electric field (ionisation signal, or S2). Fiducialisation of the detector is possible thanks to a full 3D position reconstruction by using both the time delay between S1 and S2 signals, as well as the number of photons seen by each PMT (e.g. the hit-patterns). The key discrimination parameter to distinguish background (electronic recoils in this context[1]) from signal (nuclear

---

[1]Recently the XENON1T Collaboration reported an excess in electron recoils over known backgrounds at low energies that could be interpreted as dark matter in the form of solar

recoils) is the S2/S1 ratio, which is larger for electron recoils than for nuclear recoils.

For our purposes, it is important to emphasise that this constitutes the main discriminating power between nuclear and electronic recoils, and therefore the classification accuracy achievable by the neural network should be independent of the DM properties (i.e. the DM mass) as long as enough signal events in the relevant energy window are observed.

11.3   Data Generation

To generate the image event samples used for this signal/background classification we primarily follow the methodology of Ref. [240] and point readers to their `github` repository [252] for detailed information about event generation. In a similar fashion to this reference we utilize `wimprates` [253] to generate energy spectra for an ensemble of WIMP masses ranging from 10 GeV to 1 TeV all with a spin-independent interaction cross-section of $\sigma = 10^{-45} \text{cm}^{-2}$. Given that the classification analysis is done on an event-by-event basis, the coupling strength of the WIMP-nucleon interaction is irrelevant, and so the cross-section simply serves as an overall normalisation factor for the Monte Carlo (MC) event sampling. The ER background spectra are contained within `wimprates` and are empirically determined using the polynomial fit detailed in Ref. [254][2]. These spectra are then parsed to `laidbax` [255] which performs the MC generation of events, effectively sampling from the given energy spectrum producing detector observables for each event. The physically relevant variables taken for the analysis are recoil type (NR or ER), and corresponding S1, S2 signals as several auxiliary parameters. The model output from `laidbax` was then passed into `PaX` (Processor for analysing XENON) [256], which generates images of the largest S1 and S2 peaks as well as the hitpatterns for the top and bottom PMTs. Ref. [240] concluded that classification accuracy was maximised when the S1 and S2 hitpatterns and largest S1 and S2 peak graphs were displayed in the images. Generating images in this manner meant that only information unique to each WIMP or ER event were present, effectively increasing the robustness of classification. We assume that irrespective of the architecture used, the composite hitpattern + peak images yield the best classification accuracy. Examples of an ER background event and a WIMP event are shown

---

axions [251]. In this work we consider the standard case of a signal produced by WIMPs scattering on nuclei.

[2]It should be noted that the ER model used in `laidbax` is not the official model XENON collaboration, which includes several more systematic uncertainties.

in Fig. 11.1. In each of the four images the top-left component represents a view of the detector from below, the top-right component represents a view of the detector from above and the bottom-left and bottom-right components represent the S1 and S2 distributions respectively.

11.3.1 Image Pre-processing

For training we generate an ensemble of 10000 WIMP and 10000 ER images. WIMP images correspond to masses uniformly distributed in the range 10 GeV to 1 TeV. To make training manageable for the various architectures presented in this work, the images are scaled down from their original 800×800 resolution to 75×75 pixels. For the CNN analysis and validation we find that this resolution gives optimal classification accuracy as opposed to 64×64 which is often adopted as an industry standard. In Fig 11.1 we can see examples of an ER and WIMP event both before (left) and after (right) lowering the image resolution.

11.4 Convolutional Variational Autoencoders

Autoencoders are neural networks that attempt to return output that is identical to their input. This is done to produce an encoding of the input data for the purposes of dimensional reduction, feature learning and to produce generative models.

Fig 11.2 presents a simple example of an autoencoder. There are 8 input nodes followed by three hidden layers of 4, 2 and 4 nodes and then an output layer of 8 nodes. The middle layer is known as the *latent space*, and is the encoding of the input information to a lower dimensional representation. The error is defined as a reconstruction loss, commonly mean squared error:

$$L = \frac{1}{N} \sum_{i}^{N} (x_i - y_i)^2 \tag{11.1}$$

where x is the input vector and y is the output vector and N is the number of samples (generally the batch size).

Variational Autoencoders (VAEs) are more advanced autoencoders in that we add to the latent space two additional vectors - a vector of means and a vector of variances of normal distributions. Rather than encoding the input vector onto a lower dimensional vector of real numbers, we encode the input information onto a set of Gaussian distributions which are then sampled from
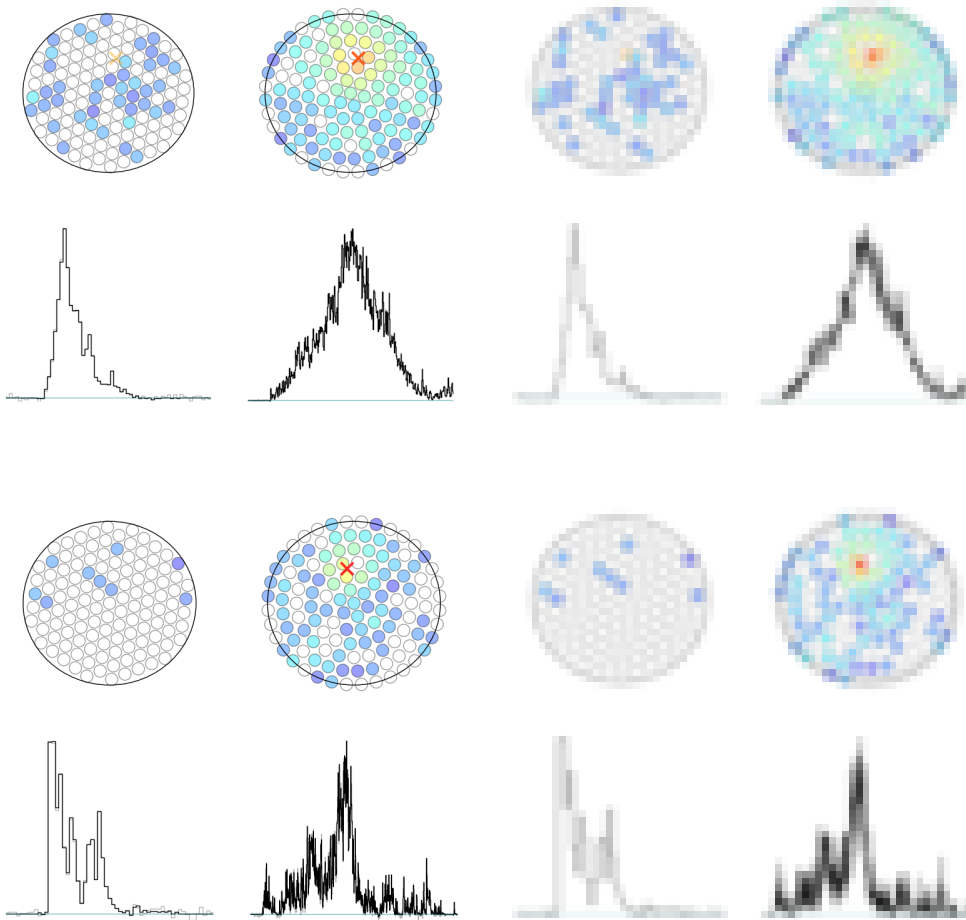
Fig. 11.1 Top: An example of an 800×800 electron recoil event image before (left) and after (right) reducing the resolution to 75×75. Bottom: Same but for a WIMP (653 GeV) event.

and this sample is decoded to produce an output vector. These distributions serve to "cover" the latent space with smooth distributions, which provides a much better representation of the input space.

It is found that when a VAE with a standard reconstruction loss such as mean squared error is trained it simply converges to Gaussians with extremely small variances (delta functions) or extremely disparate means (poor coverage of the latent space) and thus the VAE performs exactly like a standard autoencoder. The reconstruction loss can be regularised by penalising distributions that are different from standard normal distributions. This is done by addition of the
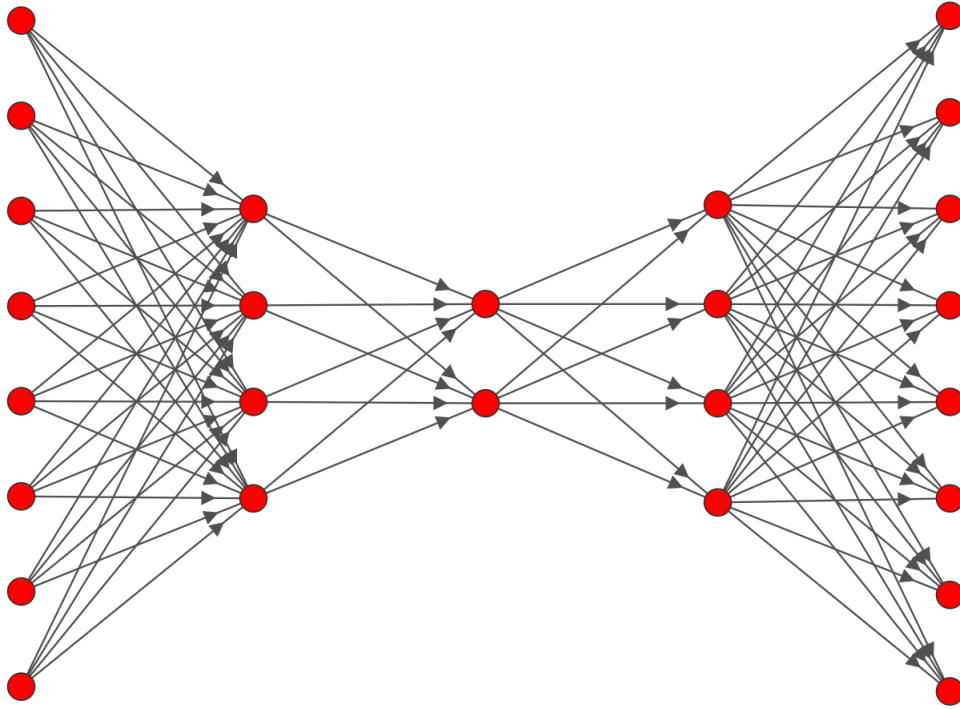
Fig. 11.2 An example of a simple autoencoder.

Kullback-Leibler divergence to the reconstruction loss as seen below:

$$L = \frac{1}{N} \sum_{i=0}^{N} (x_i - y_i)^2 + \beta L_{\text{KL}} \tag{11.2}$$

where, for a latent space of K standard normal distributions we have the special case of:

$$L_{\text{KL}} = \sum_{i=0}^{K} \sigma_i^2 + \mu_i^2 - \log(\sigma_i) - 1 \tag{11.3}$$

which is clearly minimised for $\sigma_i = 1$ and $\mu_i = 0$. The β factor is a parameter of $O(1)$ that scales the overall importance of the KL divergence relative to the reconstruction loss [219]. The default $\beta = 1$ is the definition for a standard VAE.

A Convolutional Variational Autoencoder (CVAE) is a VAE architecture which utilises convolutional layers to optimise feature extraction from images and reduce the dimensionality of inputs. For a full discussion of image convolution in the context of machine learning see [257].

11.5 Architecture

Our autoencoder architecture consists of a $75 \times 75 \times 3$ input layer, followed by three convolutional layers each with 128 filters, 3×3 kernels and stride lengths of 3. Each convolutional layer uses LeakyRelu activation with $\alpha = 0.05$, motivated by the hyper-parameter tuning undertaken in Ref [240]. The last convolutional layer is flattened and encoded onto latent space which consists of 512 nodes, corresponding to 256 means and 256 variances, before being decoded via a reflection of the encoder using convolution transpose with "same" padding.

11.6 Analysis and Results

We train the network for 200 epochs on 8000 electron recoil event images in mini-batches of 100 and test on 2000. The total loss and reconstruction loss per epoch can be seen in Fig 11.3. Both training and testing appears to converge with similar performance and overtraining does not appear to be an issue.
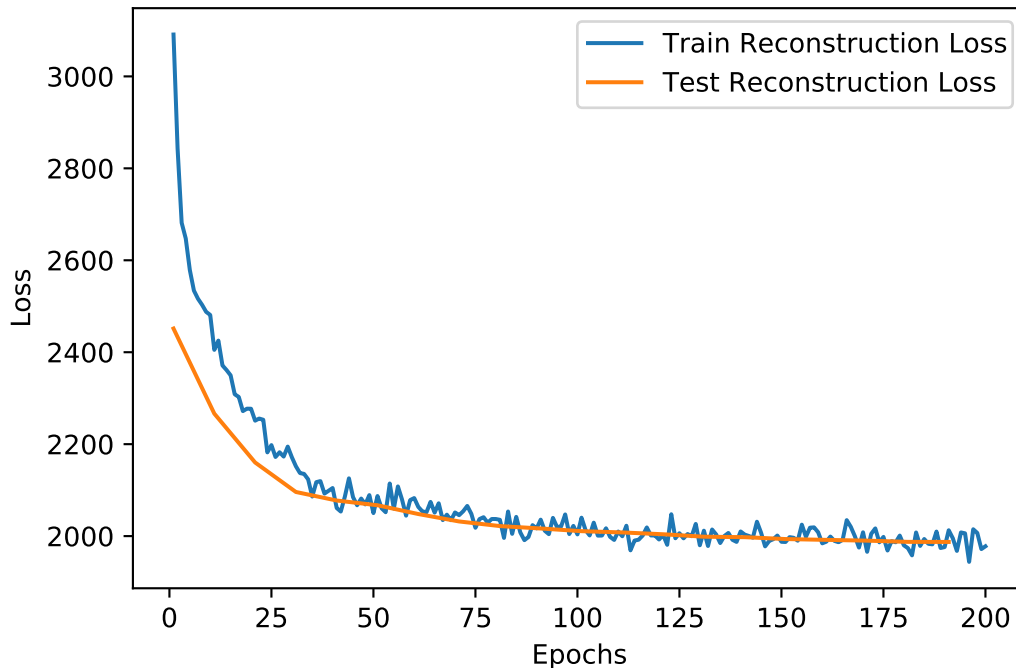
Fig. 11.3 The reconstruction loss per epoch for the training and testing sets for the CVAE.

We then apply the now trained CVAE model to a dataset of $10,000$ WIMP events with a mass of 500 GeV to assess the level of increased reconstruction loss, the results scaled to unit area can be seen in Fig 11.4. It can be seen that

the image reconstruction of the WIMP events is worse than the electron recoil events implying that the network is learning something fundamental about the electron recoil events that is not present in the WIMP events.
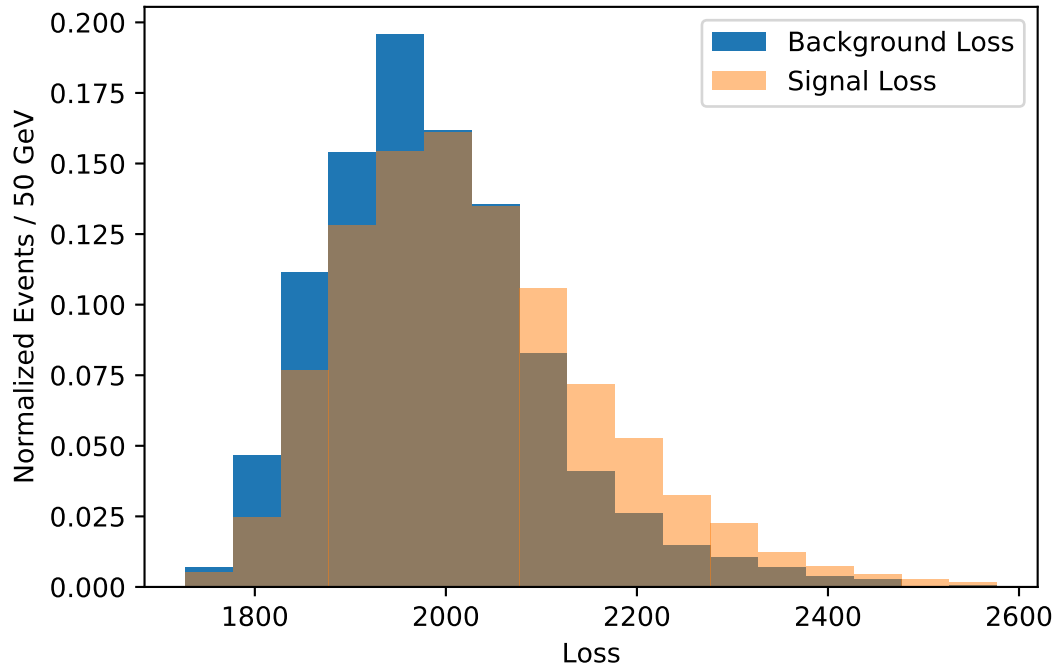
Fig. 11.4 Normalised reconstruction loss distributions for the electron recoil (background) sample and a 500 GeV WIMP particle (signal).

We also apply the network to a sample of "pseudo-data", a mock dataset created by injecting the WIMP signal events into a dataset of electron recoil events. The ER and 500 GeV WIMP events are correctly weighted to their realistic cross sections given an arbitrary exposure, in this case 5 years. We expect to see a pseudo-data distribution that matches the electron recoil distribution to a high degree in the low reconstruction loss area of the plots (the left) but with excesses seen in the high reconstruction loss areas of the plot (the right). This is presented in Fig 11.5 - where a binned χ^2-test results in a p-value of 0.008 implying we can reject our null hypothesis (that the distributions are the same) at 95% confidence.

11.7 Conclusion

In this study we presented an example of a model independent search for anomalous dark matter direct detection events using a convolutional variational
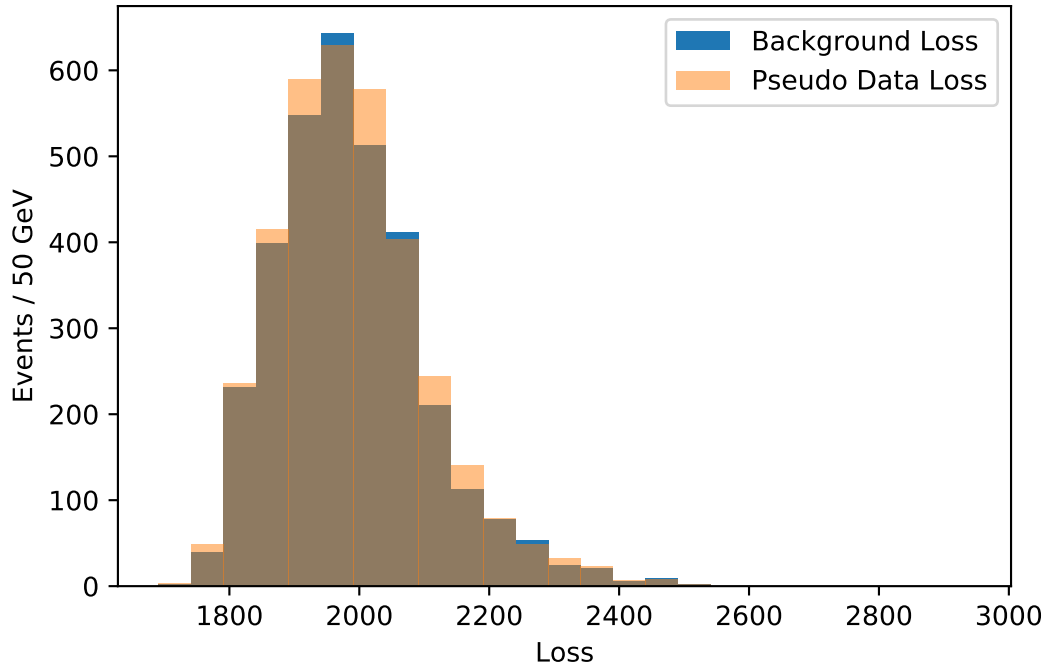
Fig. 11.5 The reconstruction loss distributions for the electron recoil (background) sample and the pseudo-data sample normalised to realistic expected event count for 5 years exposure.

autoencoder (CVAE). The planned XENONnT experiment was used as a testbed for this methodology wherein a set of composite images including the detector response and S1/S2 channels were used as input data. The CVAE was trained on a set of electron recoil background events, learning to encode the input images onto a smaller dimensional latent space, then decoding from this latent space back to the original image. The reconstruction loss is defined as the difference between the pixel information of the original image and the decoded image. The loss is thus a proxy for anomalousness as images that do not conform the the general behaviour and distribution of background events will inherently produce higher reconstruction losses as the network has not learned the properties of these events.

We demonstrate that the network would be able to show that a dataset containing a 500 GeV WIMP particle with a realistic cross section relative to the electron recoil background would have an excess identified at 95% significance. Note that this form of model independent search would provide us no information of what model was generating the excess of anomalous high reconstruction loss events but would indeed motivate further investigation into this area.

As a final note we draw attention to another well known use of generative models such as variational autoencoders - that of latent space sampling. As the latent space is a lower dimensional representation of the properties of the input dataset by sampling from the latent space randomly we can generate new events. This is an on-going and promising area of research, for example in Ref [225], as it is far more computationally efficient to sample from a latent space than to perform Monte Carlo. In the case of this study the input data are images, and the pixel information of these images presents an extremely large input space which makes latent space sampling quite difficult. However using generative models in place of Monte Carlo when the model utilizes raw detector data as inputs is an interesting avenue of research.

11.8   Acknowledgements

Chapter 12

Conclusion

In this thesis we have explored many different avenues for searches for new physics in the context of particle colliders, in particular the ATLAS detector at the LHC.

We began with an overview of the Standard Model and the various gauge theories that it is comprised of, as well as the particle content therein. We then explored a number of beyond the Standard Model theories, in particular charge-parity violating top-Higgs couplings, the two Higgs doublet model and the minimal supersymmetric model, all of which are pertinent to the contents of the published studies presented in this thesis. We then gained an understanding of the detector environment including the technology and terminology in the context of the ATLAS detector and the LHC. Next, we presented an end-to-end pipeline of an extremely common form of particle physics study from the calculation of Feynman rules, to full detector level Monte Carlo and finally analysis of data with both manual cut flow methods and more advanced supervised machine learning methods.

All of these chapters provided a firm footing to foster understanding of the next three chapters which are comprised of three published studies. The first was motivated by the incredibly high boosting of the cross section of processes that contain three top quarks that can be found with very reasonable choices of parameters in the two Higgs doublet model and as a result we set strong constraints on parameters of this model. The next study explored the phenomenology of a charge-parity violating coupling of the top quark to the Standard Model Higgs boson, the consequences of which lead to large modifications of the angular variables of the decay products of the top quark in the process $pp \rightarrow tjh$. In this study we also employed a boosted decision tree analysis which could lead to a large improvement on current constraints of the level of charge-parity violation in this coupling. The last of these studies

explored the size of quantum scattering interference in the process $pp \rightarrow t\bar{b}H^-$ in the context of the minimal supersymmetric model. For two different points in the parameter space of this model we can see large, non-negligible levels of interference which are not captured by standard analysis methodologies.

After this we change gears to a novel set of techniques in the field of particle physics - anomaly detection. These techniques stand in stark contrast to the preceding content of this thesis in that they do not seek to discover events that appear to come from a particular signal model but instead search for events that appear to be unlike background events, that is to say events that do not look like Standard Model. We present an overview of a number of algorithms that seek to do exactly this. In the following chapter we explore an in peer review study that applies these techniques at a very sophisticated level and with a wide variety of different Standard Model backgrounds and a huge number of events. This study is a very convincing example of the power of these methodologies - especially that of variational autoencoders. Finally another in progress study is presented that applies these same techniques in the vastly different environment of the planned XENONnT dark matter detector, wherein we use images of the detector response as inputs to a convolutional variational autoencoder and find great success in effectively modelling the electron recoil background of this detector.

The field of modern physics has not discovered a new particle since the 2012 discovery of the Higgs boson. This length of time without a major discovery in particle physics is small when compared to many of the particle droughts of physics history, however it is disappointing in that the current scale of experimental hardware that exists today is breathtaking compared to physics history. Supervised machine learning methods will always be powerful windows into particle physics data but it is certain that the development of novel anomaly detection techniques will be an integral part of the experimental and theoretical physics tapestry going forward. Perhaps, hopefully, we may even be able to re-visit some of the territory that we have already "ruled out" and find something that we missed. In either case the continued development of data science techniques will lay at the heart of the pursuit for new physics.

References

[1] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A.A. Abdelalim, O. Abdinov, R. Aben, B. Abi, and et al. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1 – 29, 2012.

[2] S. Chatrchyan, V. Khachatryan, A.M. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo, T. Bergauer, and et al. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30 – 61, 2012.

[3] Cush. *https://en.wikipedia.org/wiki/Standard_Model*, September 2019.

[4] Jeffrey Goldstone. Field theories with superconductor solutions. 1 1961.

[5] Jeffrey Goldstone, Abdus Salam, and Steven Weinberg. Broken symmetries. *Phys. Rev.*, 127:965–970, Aug 1962.

[6] Rodolfo Alexander Diaz. *Phenomenological analysis of the two Higgs doublet model.* PhD thesis, Colombia, U. Natl., 2002, hep-ph/0212237.

[7] Mardus. *https://en.wikipedia.org/wiki/File:Standard_Model_Of_Particle_Physics--Most_Complete_Diagram.png*, Dec 2015.

[8] Goran Senjanović. Natural Philosophy versus Philosophy of Naturalness. In *LHC Days in Split*, 1 2020, 2001.10988.

[9] S.F. King. Neutrino Mass Models: A Road map. *J. Phys. Conf. Ser.*, 136:022038, 2008, 0810.0492.

[10] A. D. Sakharov. Violation of cp in variance, c asymmetry, and baryon asymmetry of the universe. *Phys. Usp.*, 34(5):392–393, 1991.

[11] Laurent Canetti, Marco Drewes, and Mikhail Shaposhnikov. Matter and Antimatter in the Universe. *New J. Phys.*, 14:095012, 2012, 1204.4186.

[12] H. Baer and X. Tata. *Weak scale supersymmetry: From superfields to scattering events.* Cambridge University Press, 5 2006.

[13] Sudhir K. Vempati. Introduction to MSSM. 1 2012, 1201.0334.

[14] Glennys R. Farrar and Steven Weinberg. Supersymmetry at ordinary energies. ii. r invariance, goldstone bosons, and gauge-fermion masses. *Phys. Rev. D*, 27:2732–2746, Jun 1983.

[15] VermillionBird. *https://en.wikipedia.org/wiki/Minimal_ Supersymmetric_Standard_Model*, January 2008.

[16] Riccardo Barbieri and Alessandro Strumia. The 'LEP paradox'. In *4th Rencontres du Vietnam: Physics at Extreme Energies (Particle Physics and Astrophysics)*, 7 2000, hep-ph/0007265.

[17] Arpad Horvath. *https://en.wikipedia.org/wiki/Large_Hadron_Collider*, Apr 2006.

[18] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[19] ATLAS tile calorimeter: Technical design report. 1996.

[20] ATLAS liquid argon calorimeter: Technical design report. 1996.

[21] ATLAS central solenoid: Technical design report. 1997.

[22] ATLAS endcap toroids: Technical design report. 1997.

[23] ATLAS barrel toroid: Technical design report. 1997.

[24] ATLAS muon spectrometer: Technical design report. 1997.

[25] ATLAS magnet system: Technical design report. 1997.

[26] ATLAS inner detector: Technical design report. Vol. 2. 1997.

[27] ATLAS inner detector: Technical design report. Vol. 1. 1997.

[28] ATLAS first level trigger: Technical design report. 1998.

[29] M. S. Alam et al. ATLAS pixel detector: Technical design report. 1998.

[30] A. Airapetian et al. ATLAS: Detector and physics performance technical design report. Volume 1. 1999.

[31] A. Airapetian et al. ATLAS: Detector and physics performance technical design report. Volume 2. 1999.

[32] ATLAS high-level trigger, data acquisition and controls: Technical design report. 2003.

[33] Joao Pequenao. *http://cds.cern.ch/images/CERN-GE-0803012-01*, March 2008.

[34] Collaboration ATLAS. Event display of a H -gt; 4e candidate event. General Photo, Jul 2012.

[35] Meta501. *https://commons.wikimedia.org/wiki/File:Pseudorapidity_ plot.svg*, July 2012.

[36] Gavin P. Salam and Gregory Soyez. A Practical Seedless Infrared-Safe Cone jet algorithm. *JHEP*, 05:086, 2007, 0704.0292.

[37] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008, 0802.1189.

[38] A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging. 7 2009.

[39] D0 Collaboration. *https://www-d0.fnal.gov/Run2Physics/top/ singletop_observation/singletop_observation_updated.html*, March 2009.

[40] Sebastian Schätzel. *Boosted Top Quarks and Jet Structure*. PhD thesis, Heidelberg U., 2013, 1403.5176.

[41] Mrinal Dasgupta, Alessandro Fregoso, Simone Marzani, and Gavin P. Salam. Towards an understanding of jet substructure. *JHEP*, 09:029, 2013, 1307.0007.

[42] Katherine Fraser and Matthew D. Schwartz. Jet Charge and Machine Learning. *JHEP*, 10:093, 2018, 1803.08066.

[43] Adam Alloul, Neil D. Christensen, Céline Degrande, Claude Duhr, and Benjamin Fuks. FeynRules 2.0 - A complete toolbox for tree-level phenomenology. *Comput. Phys. Commun.*, 185:2250–2300, 2014, 1310.1921.

[44] Vladyslav Shtabovenko, Rolf Mertig, and Frederik Orellana. FeynCalc 9.3: New features and improvements. 1 2020, 2001.04407.

[45] F. Staub. SARAH. 6 2008, 0806.0538.

[46] A. Semenov. LanHEP: A Package for the automatic generation of Feynman rules in field theory. Version 3.0. *Comput. Phys. Commun.*, 180:431–454, 2009, 0805.0555.

[47] 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*, 07:079, 2014, 1405.0301.

[48] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008, 0710.3820.

[49] T. Gleisberg, Stefan. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter. Event generation with SHERPA 1.1. *JHEP*, 02:007, 2009, 0811.4622.

[50] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010, 1002.2581.

[51] Johannes Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, 76(4):196, 2016, 1512.01178.

[52] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand. Parton fragmentation and string dynamics. *Physics Reports*, 97(2):31 – 145, 1983.

[53] 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. *JHEP*, 02:057, 2014, 1307.6346.

[54] George W. Snedecor and William G. Cochran. *Statistical Methods*. Iowa State University Press, 1989.

[55] A L Read. Presentation of search results: the CLstechnique. *Journal of Physics G: Nuclear and Particle Physics*, 28(10):2693–2704, sep 2002.

[56] Alexander L. Read. Modified frequentist analysis of search results (The CL(s) method). In *Workshop on Confidence Limits*, pages 81–101, 8 2000.

[57] Kevin Gurney. *An Introduction to Neural Networks*. UCL Press, 1997.

[58] L. Breiman, J. Friedman, R. Olshen, and C. J. Stone. Classification and regression trees. 1983.

[59] Ian Witten, Eibe Frank, Mark Hall, and Christopher Pal. *Data Mining*. Elsevier, 2016.

[60] L Breiman. Arcing the edge. June 1997.

[61] Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, 2001.

[62] Jerome H. Friedman. Stochastic gradient boosting. *Computational Statistics Data Analysis*, 38(4):367 – 378, 2002. Nonlinear Methods and Data Mining.

[63] Yoav Freund and Robert E. Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In Paul Vitányi, editor, *Computational Learning Theory*, pages 23–37, Berlin, Heidelberg, 1995. Springer Berlin Heidelberg.

[64] I. Antcheva, M. Ballintijn, B. Bellenot, M. Biskup, R. Brun, N. Buncic, Ph. Canal, D. Casadei, O. Couet, V. Fine, L. Franco, G. Ganis, A. Gheata, D. Gonzalez Maline, M. Goto, J. Iwaszkiewicz, A. Kreshuk, D. Marcos Segura, R. Maunder, L. Moneta, A. Naumann, E. Offermann, V. Onuchin, S. Panacek, F. Rademakers, P. Russo, and M. Tadel. Root — a c++ framework for petabyte data storage, statistical analysis and visualization. *Computer Physics Communications*, 180(12):2499 – 2512, 2009. 40 YEARS OF CPC: A celebratory issue focused on quality software for high performance, grid and novel computing architectures.

[65] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov, D. Dannheim, K. Danielowski, S. Henrot-Versille, M. Jachowski, K. Kraszewski, A. Krasznahorkay Jr., M. Kruk, Y. Mahalalel, R. Ospanov, X. Prudent, A. Robert, D. Schouten, F. Tegenfeldt, A. Voigt, K. Voss, M. Wolter, and A. Zemla. Tmva - toolkit for multivariate data analysis, 2007, physics/0703039.

[66] G.C. Branco, P.M. Ferreira, L. Lavoura, M.N. Rebelo, Marc Sher, and Joao P. Silva. Theory and phenomenology of two-Higgs-doublet models. *Phys. Rept.*, 516:1–102, 2012, 1106.0034.

[67] A.G. Akeroyd et al. Prospects for charged Higgs searches at the LHC. *Eur. Phys. J. C*, 77(5):276, 2017, 1607.01320.

[68] J.F. Gunion, H.E. Haber, F.E. Paige, Wu-Ki Tung, and S.S.D. Willenbrock. Neutral and Charged Higgs Detection: Heavy Quark Fusion, Top Quark Mass Dependence and Rare Decays. *Nucl. Phys. B*, 294:621, 1987.

[69] Stefano Moretti, Rui Santos, and Pankaj Sharma. Optimising Charged Higgs Boson Searches at the Large Hadron Collider Across $bbW^{\pm}$ Final States. *Phys. Lett. B*, 760:697–705, 2016, 1604.04965.

[70] Abdesslam Arhrib, Rachid Benbrik, and Stefano Moretti. Bosonic Decays of Charged Higgs Bosons in a 2HDM Type-I. *Eur. Phys. J. C*, 77(9):621, 2017, 1607.02402.

[71] Jinmian Li, Riley Patrick, Pankaj Sharma, and Anthony G. Williams. Boosting the charged Higgs search prospects using jet substructure at the LHC. *JHEP*, 11:164, 2016, 1609.02645.

[72] Riley Patrick, Pankaj Sharma, and Anthony G. Williams. Exploring a heavy charged higgs using jet substructure in a fully hadronic channel. *Nuclear Physics B*, 917:19 – 30, 2017.

[73] Baradhwaj Coleppa, Felix Kling, and Shufang Su. Charged Higgs search via $AW^{\pm}/HW^{\pm}$ channel. *JHEP*, 12:148, 2014, 1408.4119.

[74] Felix Kling, Jose Miguel No, and Shufang Su. Anatomy of Exotic Higgs Decays in 2HDM. *JHEP*, 09:093, 2016, 1604.01406.

[75] Vernon Barger, Wai-Yee Keung, and Brian Yencho. Triple-Top Signal of New Physics at the LHC. *Phys. Lett. B*, 687:70–74, 2010, 1001.0221.

[76] Ezequiel Alvarez, Darius A. Faroughy, Jernej F. Kamenik, Roberto Morales, and Alejandro Szynkman. Four Tops for LHC. *Nucl. Phys. B*, 915:19–43, 2017, 1611.05032.

[77] Shinya Kanemura, Hiroshi Yokoya, and Ya-Juan Zheng. Searches for additional Higgs bosons in multi-top-quarks events at the LHC and the International Linear Collider. *Nucl. Phys. B*, 898:286–300, 2015, 1505.01089.

[78] Nathaniel Craig, Jan Hajer, Ying-Ying Li, Tao Liu, and Hao Zhang. Heavy Higgs bosons at low $\tan\beta$: from the LHC to 100 TeV. *JHEP*, 01:018, 2017, 1605.08744.

[79] Chuan-Ren Chen. Searching for new physics with triple-top signal at the lhc. *Physics Letters B*, 736:321 – 324, 2014.

[80] Search for charged Higgs bosons in the τ+jets final state using 14.7 fb-1 of pp collision data recorded at s$\sqrt{}$=13 TeV with the ATLAS experiment. 8 2016.

[81] Search for charged Higgs bosons in the $H^\pm \to tb$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector. Technical Report ATLAS-CONF-2016-089, CERN, Geneva, Aug 2016.

[82] Morad Aaboud et al. Search for charged Higgs bosons produced in association with a top quark and decaying via $H^\pm \to \tau\nu$ using pp collision data recorded at $\sqrt{s} = 13$ TeV by the ATLAS detector. *Phys. Lett. B*, 759:555–574, 2016, 1603.09203.

[83] David Eriksson, Johan Rathsman, and Oscar Stal. 2HDMC: Two-Higgs-Doublet Model Calculator Physics and Manual. *Comput. Phys. Commun.*, 181:189–205, 2010, 0902.0851.

[84] M. Misiak et al. Updated NNLO QCD predictions for the weak radiative B-meson decays. *Phys. Rev. Lett.*, 114(22):221801, 2015, 1503.01789.

[85] Mikolaj Misiak and Matthias Steinhauser. Weak radiative decays of the B meson and bounds on $M_{H^\pm}$ in the Two-Higgs-Doublet Model. *Eur. Phys. J. C*, 77(3):201, 2017, 1702.04571.

[86] Alessandro Broggio, Eung Jin Chun, Massimo Passera, Ketan M. Patel, and Sudhir K. Vempati. Limiting two-Higgs-doublet models. *JHEP*, 11:058, 2014, 1409.3199.

[87] G.C. Dorsch, S.J. Huber, K. Mimasu, and J.M. No. Hierarchical versus degenerate 2HDM: The LHC run 1 legacy at the onset of run 2. *Phys. Rev. D*, 93(11):115033, 2016, 1601.04545.

[88] G.C. Dorsch, S.J. Huber, K. Mimasu, and J.M. No. Echoes of the Electroweak Phase Transition: Discovering a second Higgs doublet through $A_0 \to ZH_0$. *Phys. Rev. Lett.*, 113(21):211802, 2014, 1405.5537.

[89] Jie Ren, Lei Wu, and Jin Min Yang. Unveiling CP property of top-Higgs coupling with graph neural networks at the LHC. 2019, 1901.05627.

[90] M. Aaboud et al. Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector. *Phys. Lett.*, B784:173–191, 2018, 1806.00425.

[91] Albert M Sirunyan et al. Observation of $t\bar{t}H$ production. *Phys. Rev. Lett.*, 120(23):231801, 2018, 1804.02610.

[92] Jung Chang, Kingman Cheung, Jae Sik Lee, and Chih-Ting Lu. Probing the Top-Yukawa Coupling in Associated Higgs production with a Single Top Quark. *JHEP*, 05:062, 2014, 1403.2053.

[93] Archil Kobakhidze, Lei Wu, and Jason Yue. Anomalous top-higgs couplings and top polarisation in single top and higgs associated production at the lhc. *Journal of High Energy Physics*, 2014(10):100, Oct 2014.

[94] Manfred Kraus, Till Martini, Sascha Peitzsch, and Peter Uwer. Exploring BSM Higgs couplings in single top-quark production. 2019, 1908.09100.

[95] Andrei V. Gritsan, Raoul Röntsch, Markus Schulze, and Meng Xiao. Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques. *Phys. Rev.*, D94(5):055023, 2016, 1606.03107.

[96] Sara Khatibi and Mojtaba Mohammadi Najafabadi. Exploring the Anomalous Higgs-top Couplings. *Phys. Rev.*, D90(7):074014, 2014, 1409.6553.

[97] Andreas Papaefstathiou and Kazuki Sakurai. Determining the Helicity Structure of Third Generation Resonances. *JHEP*, 06:069, 2012, 1112.3956.

[98] Pankaj Agrawal, Subhadip Mitra, and Ambresh Shivaji. Effect of Anomalous Couplings on the Associated Production of a Single Top Quark and a Higgs Boson at the LHC. *JHEP*, 12:077, 2013, 1211.4362.

[99] Darius A. Faroughy, Jernej F. Kamenik, Nejc Košnik, and Aleks Smolkovič. Probing the CP nature of the top quark Yukawa at hadron colliders. 2019, 1909.00007.

[100] Federico Demartin, Fabio Maltoni, Kentarou Mawatari, and Marco Zaro. Higgs production in association with a single top quark at the LHC. *Eur. Phys. J.*, C75(6):267, 2015, 1504.00611.

[101] Federico Demartin, Benedikt Maier, Fabio Maltoni, Kentarou Mawatari, and Marco Zaro. tWH associated production at the LHC. *Eur. Phys. J.*, C77(1):34, 2017, 1607.05862.

[102] Rohini M. Godbole, Kumar Rao, Saurabh D. Rindani, and Ritesh K. Singh. On measurement of top polarization as a probe of $t\bar{t}$ production mechanisms at the lhc. *Journal of High Energy Physics*, 2010(11):144, Nov 2010.

[103] Rohini M Godbole, Saurabh D Rindani, and Ritesh K Singh. Lepton distribution as a probe of new physics in production and decay of thetquark and its polarization. *Journal of High Energy Physics*, 2006(12):021–021, dec 2006.

[104] Rohini M. Godbole, Saurabh D. Rindani, Kumar Rao, and Ritesh Singh. Top polarization as a probe of new physics. *AIP Conference Proceedings*, 1200(1):682–685, 2010, https://aip.scitation.org/doi/pdf/10.1063/1.3327703.

[105] Sudhansu S. Biswal, Saurabh D. Rindani, and Pankaj Sharma. Probing chromomagnetic and chromoelectric couplings of the top quark using its polarization in pair production at hadron colliders. *Phys. Rev. D*, 88:074018, Oct 2013.

[106] Katri Huitu, Santosh Kumar Rai, Kumar Rao, Saurabh D. Rindani, and Pankaj Sharma. Probing top charged-higgs production using top polarization at the large hadron collider. *Journal of High Energy Physics*, 2011(4):26, Apr 2011.

[107] R. M. Godbole, L. Hartgring, I. Niessen, and C. D. White. Top polarisation studies in h-t and w t production. *Journal of High Energy Physics*, 2012(1):11, Jan 2012.

[108] Saurabh D. Rindani and Pankaj Sharma. Probing anomalous tbw couplings in single-top production using top polarization at the large hadron collider. *Journal of High Energy Physics*, 2011(11):82, Nov 2011.

[109] Saurabh D. Rindani and Pankaj Sharma. Cp violation in tbw couplings at the lhc. *Physics Letters B*, 712(4):413 – 418, 2012.

[110] Saurabh D. Rindani, Rui Santos, and Pankaj Sharma. Measuring the charged higgs mass and distinguishing between models with top-quark observables. *Journal of High Energy Physics*, 2013(11):188, Nov 2013.

[111] Saurabh D. Rindani, Pankaj Sharma, and Anthony W. Thomas. Polarization of top quark as a probe of its chromomagnetic and chromoelectric couplings in tw production at the large hadron collider. *Journal of High Energy Physics*, 2015(10):180, Oct 2015.

[112] Abdesslam Arhrib, Adil Jueid, and Stefano Moretti. Top quark polarization as a probe of charged Higgs bosons. *Phys. Rev.*, D98(11):115006, 2018, 1807.11306.

[113] Adil Jueid. Probing anomalous Wtb couplings at the LHC in single t-channel top quark production. *Phys. Rev.*, D98(5):053006, 2018, 1805.07763.

[114] John Ellis, Dae Sung Hwang, Kazuki Sakurai, and Michihisa Takeuchi. Disentangling higgs-top couplings in associated production. *Journal of High Energy Physics*, 2014(4):4, Apr 2014.

[115] Jason Yue. Enhanced thj signal at the LHC with $h \to \gamma\gamma$ decay and $\mathcal{CP}$-violating top-Higgs coupling. *Phys. Lett.*, B744:131–136, 2015, 1410.2701.

[116] Saurabh D. Rindani, Pankaj Sharma, and Ambresh Shivaji. Unraveling the CP phase of top-Higgs coupling in associated production at the LHC. *Phys. Lett.*, B761:25–30, 2016, 1605.03806.

[117] J. A. Aguilar-Saavedra. A Minimal set of top-Higgs anomalous couplings. *Nucl. Phys.*, B821:215–227, 2009, 0904.2387.

[118] J. A. Aguilar-Saavedra. A Minimal set of top anomalous couplings. *Nucl. Phys.*, B812:181–204, 2009, 0811.3842.

[119] Cen Zhang, Nicolas Greiner, and Scott Willenbrock. Constraints on Non-standard Top Quark Couplings. *Phys. Rev.*, D86:014024, 2012, 1201.6670.

[120] Hermes Belusca-Maito. Higgs Couplings in an Effective Theory Framework. 2015, 1507.05657.

[121] Xiao-Gang He, Yong Tang, and German Valencia. Interplay between new physics in one-loop Higgs couplings and the top-quark Yukawa coupling. *Phys. Rev.*, D88:033005, 2013, 1305.5420.

[122] Roni Harnik, Adam Martin, Takemichi Okui, Reinard Primulando, and Felix Yu. Measuring CP Violation in $h \to \tau^+\tau^-$ at Colliders. *Phys. Rev.*, D88(7):076009, 2013, 1308.1094.

[123] Nicolas Mileo, Ken Kiers, Alejandro Szynkman, Daniel Crane, and Ethan Gegner. Pseudoscalar top-higgs coupling: exploration of cp-odd observables to resolve the sign ambiguity. *Journal of High Energy Physics*, 2016(7):56, Jul 2016.

[124] Kenji Nishiwaki, Saurabh Niyogi, and Ambresh Shivaji. tth anomalous coupling in double higgs production. *Journal of High Energy Physics*, 2014(4):11, Apr 2014.

[125] Fawzi Boudjema, Diego Guadagnoli, Rohini M. Godbole, and Kirtimaan A. Mohan. Laboratory-frame observables for probing the top-higgs boson interaction. *Phys. Rev. D*, 92:015019, Jul 2015.

[126] Kingman Cheung, Jae Sik Lee, and Po-Yan Tseng. Higgs precision (higgcision) era begins. *Journal of High Energy Physics*, 2013(5):134, May 2013.

[127] Kingman Cheung, Jae Sik Lee, and Po-Yan Tseng. Higgs precision analysis updates 2014. *Phys. Rev. D*, 90:095009, Nov 2014.

[128] Debajyoti Choudhury, Rashidul Islam, and Anirban Kundu. Anomalous higgs couplings as a window to new physics. *Phys. Rev. D*, 88:013014, Jul 2013.

[129] Gautam Bhattacharyya, Dipankar Das, and Palash B. Pal. Modified higgs couplings and unitarity violation. *Phys. Rev. D*, 87:011702, Jan 2013.

[130] Joachim Brod, Ulrich Haisch, and Jure Zupan. Constraints on cp-violating higgs couplings to the third generation. *Journal of High Energy Physics*, 2013(11):180, Nov 2013.

[131] V. Cirigliano, W. Dekens, J. de Vries, and E. Mereghetti. Is there room for CP violation in the top-Higgs sector? *Phys. Rev.*, D94(1):016002, 2016, 1603.03049.

[132] Y. T. Chien, V. Cirigliano, W. Dekens, J. de Vries, and E. Mereghetti. Direct and indirect constraints on cp-violating higgs-quark and higgs-gluon interactions. *Journal of High Energy Physics*, 2016(2):11, Feb 2016.

[133] Hannes Mildner. *Measurements of the Top-Higgs Coupling with the CMS Experiment*. PhD thesis, Karlsruher Institut für Technologie (KIT), 2016.

[134] Lei Wu. Enhancing *thj* Production from Top-Higgs FCNC Couplings. *JHEP*, 02:061, 2015, 1407.6113.

[135] R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro. The automation of next-to-leading order electroweak calculations. *JHEP*, 07:185, 2018, 1804.10017.

[136] Morad Aaboud et al. Search for top-squark pair production in final states with one lepton, jets, and missing transverse momentum using 36 fb^{-1} of $\sqrt{s} = 13$ TeV pp collision data with the ATLAS detector. *JHEP*, 06:108, 2018, 1711.11520.

[137] Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1 – 29, 2012.

[138] Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30 – 61, 2012.

[139] Georges Aad et al. Combined Measurement of the Higgs Boson Mass in *pp* Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, 2015, 1503.07589.

[140] Georges Aad et al. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 08:045, 2016, 1606.02266.

[141] Stefano Moretti and Shaaban Khalil. *Supersymmetry Beyond Minimality: From Theory to Experiment*. December 2017.

[142] H.P. Nilles. Supersymmetry, supergravity and particle physics. *Physics Reports*, 110(1):1 – 162, 1984.

[143] H.E. Haber and G.L. Kane. The search for supersymmetry: Probing physics beyond the standard model. *Physics Reports*, 117(2):75 – 263, 1985.

[144] Vernon D. Barger, R.J.N. Phillips, and D.P. Roy. Heavy charged Higgs signals at the LHC. *Phys. Lett. B*, 324:236–240, 1994, hep-ph/9311372.

[145] K.A. Assamagan, M. Guchait, and S. Moretti. Charged Higgs bosons in the transition region M(H+-) ~ m(t) at the LHC. In *3rd Les Houches Workshop on Physics at TeV Colliders*, 2 2004, hep-ph/0402057.

[146] Monoranjan Guchait and Stefano Moretti. Improving the discovery potential of charged higgs bosons at tevatron run 2. *Journal of High Energy Physics*, 2002(01):001–001, jan 2002.

[147] Georges Aad et al. Search for charged Higgs bosons decaying via $H^{\pm} \to \tau^{\pm}\nu$ in fully hadronic final states using *pp* collision data at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 03:088, 2015, 1412.6663.

[148] CMS Collaboration. Search for a charged higgs boson in pp collisions at $\sqrt{s} = 8$ tev. *Journal of High Energy Physics*.

[149] Albert M Sirunyan et al. Search for charged Higgs bosons in the H$^{\pm} \to \tau^{\pm}\nu_{\tau}$ decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 07:142, 2019, 1903.04560.

[150] Morad Aaboud et al. Search for charged Higgs bosons decaying via $H^\pm \to \tau^\pm \nu_\tau$ in the τ+jets and τ+lepton final states with 36 fb^{-1} of pp collision data recorded at $\sqrt{s} = 13$ TeV with the ATLAS experiment. *JHEP*, 09:139, 2018, 1807.07915.

[151] David Miller, Stefano Moretti, D. Roy, and W. Stirling. Detecting heavy charged higgs bosons at the cern lhc with four b-quark tags. *Phys. Rev. D*, 61, 02 2000.

[152] S. Moretti and D.P. Roy. Detecting heavy charged Higgs bosons at the LHC with triple b tagging. *Phys. Lett. B*, 470:209–214, 1999, hep-ph/9909435.

[153] Search for charged Higgs bosons with the $H^\pm \to \tau^\pm \nu_\tau$ decay channel in the fully hadronic final state at $\sqrt{s} = 13$ TeV. 10 2016.

[154] Search for charged Higgs bosons in the $H^\pm \to tb$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector. 8 2016.

[155] Morad Aaboud et al. Search for charged Higgs bosons decaying into top and bottom quarks at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 11:085, 2018, 1808.03599.

[156] Filip Moortgat, Salavat Abdullin, and Daniel Denegri. Observability of MSSM Higgs bosons via sparticle decay modes in CMS. 12 2001, hep-ph/0112046.

[157] M. Bisset, N. Kersting, J. Li, F. Moortgat, S. Moretti, and Q.L. Xie. Pair-produced heavy particle topologies: MSSM neutralino properties at the LHC from gluino/squark cascade decays. *Eur. Phys. J. C*, 45:477–492, 2006, hep-ph/0501157.

[158] Mike Bisset, Monoranjan Guchait, and Stefano Moretti. Signatures of MSSM charged Higgs bosons via chargino neutralino decay channels at the LHC. *Eur. Phys. J. C*, 19:143–154, 2001, hep-ph/0010253.

[159] Mike Bisset, Filip Moortgat, and Stefano Moretti. Trilepton + top signal from chargino neutralino decays of MSSM charged Higgs bosons at the LHC. *Eur. Phys. J. C*, 30:419–434, 2003, hep-ph/0303093.

[160] Abdesslam Arhrib, Rachid Benbrik, Stefano Moretti, Rui Santos, and Pankaj Sharma. Signal to background interference in $pp \to tH^- \to tW^- b\bar{b}$ at the LHC Run-II. *Phys. Rev. D*, 97(7):075037, 2018, 1712.05018.

[161] S. Heinemeyer, W. Hollik, and G. Weiglein. The Masses of the neutral CP - even Higgs bosons in the MSSM: Accurate analysis at the two loop level. *Eur. Phys. J. C*, 9:343–366, 1999, hep-ph/9812472.

[162] S. Heinemeyer, W. Hollik, and G. Weiglein. QCD corrections to the masses of the neutral CP - even Higgs bosons in the MSSM. *Phys. Rev. D*, 58:091701, 1998, hep-ph/9803277.

[163] S. Heinemeyer, W. Hollik, and G. Weiglein. Precise prediction for the mass of the lightest Higgs boson in the MSSM. *Phys. Lett. B*, 440:296–304, 1998, hep-ph/9807423.

[164] G. Degrassi, S. Heinemeyer, W. Hollik, P. Slavich, and G. Weiglein. Towards high precision predictions for the MSSM Higgs sector. *Eur. Phys. J. C*, 28:133–143, 2003, hep-ph/0212020.

[165] W. Beenakker, R. Hopker, and M. Spira. PROSPINO: A Program for the production of supersymmetric particles in next-to-leading order QCD. 11 1996, hep-ph/9611232.

[166] S. Heinemeyer, W. Hollik, and G. Weiglein. FeynHiggs: A Program for the calculation of the masses of the neutral CP even Higgs bosons in the MSSM. *Comput. Phys. Commun.*, 124:76–89, 2000, hep-ph/9812320.

[167] T. Hahn, S. Heinemeyer, W. Hollik, H. Rzehak, and G. Weiglein. Feyn-higgs: A program for the calculation of mssm higgs-boson observables – version 2.6.5. *Computer Physics Communications*, 180(8):1426 – 1427, 2009.

[168] Philip Bechtle, Oliver Brein, Sven Heinemeyer, Georg Weiglein, and Karina E. Williams. HiggsBounds: Confronting Arbitrary Higgs Sectors with Exclusion Bounds from LEP and the Tevatron. *Comput. Phys. Commun.*, 181:138–167, 2010, 0811.4169.

[169] Philip Bechtle, Oliver Brein, Sven Heinemeyer, Georg Weiglein, and Karina E. Williams. HiggsBounds 2.0.0: Confronting Neutral and Charged Higgs Sector Predictions with Exclusion Bounds from LEP and the Tevatron. *Comput. Phys. Commun.*, 182:2605–2631, 2011, 1102.1898.

[170] Philip Bechtle, Oliver Brein, Sven Heinemeyer, Oscar Stål, Tim Stefaniak, Georg Weiglein, and Karina E. Williams. HiggsBounds − 4: Improved Tests of Extended Higgs Sectors against Exclusion Bounds from LEP, the Tevatron and the LHC. *Eur. Phys. J. C*, 74(3):2693, 2014, 1311.0055.

[171] Philip Bechtle, Sven Heinemeyer, Oscar Stal, Tim Stefaniak, and Georg Weiglein. Applying Exclusion Likelihoods from LHC Searches to Extended Higgs Sectors. *Eur. Phys. J. C*, 75(9):421, 2015, 1507.06706.

[172] Philip Bechtle, Sven Heinemeyer, Oscar Stål, Tim Stefaniak, and Georg Weiglein. *HiggsSignals*: Confronting arbitrary Higgs sectors with measurements at the Tevatron and the LHC. *Eur. Phys. J. C*, 74(2):2711, 2014, 1305.1933.

[173] Search for charged Higgs bosons: Preliminary combined results using LEP data collected at energies up to 209-GeV. In *2001 Europhysics Conference on High Energy Physics*, 7 2001, hep-ex/0107031.

[174] J. Abdallah et al. Search for charged Higgs bosons at LEP in general two Higgs doublet models. *Eur. Phys. J. C*, 34:399–418, 2004, hep-ex/0404012.

[175] V.M. Abazov et al. Search for Charged Higgs Bosons in Top Quark Decays. *Phys. Lett. B*, 682:278–286, 2009, 0908.1811.

[176] T. Aaltonen et al. Search for charged Higgs bosons in decays of top quarks in p anti-p collisions at s**(1/2) = 1.96 TeV. *Phys. Rev. Lett.*, 103:101803, 2009, 0907.1269.

[177] CDF Collaboration. Cdf note 7712.

[178] CDF Collaboration. Cdf note 8353.

[179] Georges Aad et al. Search for charged Higgs bosons decaying via $H^+ \to \tau\nu$ in top quark pair events using pp collision data at $\sqrt{s} = 7$ TeV with the ATLAS detector. *JHEP*, 06:039, 2012, 1204.2760.

[180] Abdesslam Arhrib, Rachid Benbrik, Hicham Harouiz, Stefano Moretti, and Abdessamad Rouchad. A Guidebook to Hunting Charged Higgs Bosons at the LHC. 10 2018, 1810.09106.

[181] M. Carena, S. Heinemeyer, O. Stål, C.E.M. Wagner, and G. Weiglein. MSSM Higgs Boson Searches at the LHC: Benchmark Scenarios after the Discovery of a Higgs-like Particle. *Eur. Phys. J. C*, 73(9):2552, 2013, 1302.7033.

[182] A. Djouadi, L. Maiani, G. Moreau, A. Polosa, J. Quevillon, and V. Riquer. The post-Higgs MSSM scenario: Habemus MSSM? *Eur. Phys. J. C*, 73:2650, 2013, 1307.5205.

[183] S. Heinemeyer, W. Hollik, and G. Weiglein. Constraints on tan Beta in the MSSM from the upper bound on the mass of the lightest Higgs boson. *JHEP*, 06:009, 2000, hep-ph/9909540.

[184] A. Djouadi, J. Kalinowski, and M. Spira. HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension. *Comput. Phys. Commun.*, 108:56–74, 1998, hep-ph/9704448.

[185] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015, 1410.3012.

[186] K.J.F. Gaemers and F. Hoogeveen. Higgs Production and Decay Into Heavy Flavors With the Gluon Fusion Mechanism. *Phys. Lett. B*, 146:347–349, 1984.

[187] D. Dicus, A. Stange, and S. Willenbrock. Higgs decay to top quarks at hadron colliders. *Phys. Lett. B*, 333:126–131, 1994, hep-ph/9404359.

[188] Jeonghyeon Song, Yeo Woong Yoon, and Sunghoon Jung. Interference effect on a heavy higgs resonance signal in the $\gamma\gamma$ and zz channels. *Phys. Rev. D*, 93:055035, Mar 2016.

[189] B. Abbott et al. Search for new physics in $e\mu X$ data at DØ using Sherlock: A quasi model independent search strategy for new physics. *Phys. Rev.*, D62:092004, 2000, hep-ex/0006011.

[190] D0 Collaboration. Quasi-model-independent search for new physics at large transverse momentum. *Phys. Rev. D*, 64:012004, 2001, hep-ex/0011067.

[191] D0 Collaboration. Quasi-Model-Independent Search for New High p_T Physics at DO. *Phys. Rev. Lett.*, 86:3712–3717, 2001, hep-ex/0011071.

[192] D0 Collaboration. Model independent search for new phenomena in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev. D*, 85:092015, 2012, 1108.5362.

[193] H1 Collaboration. A general search for new phenomena in *ep* scattering at HERA. *Phys. Lett. B*, 602:14–30, 2004, hep-ex/0408044.

[194] H1 Collaboration. A general search for new phenomena at HERA. *Phys. Lett. B*, 674:257–268, 2009, 0901.0507.

[195] Georgios Choudalakis. On hypothesis testing, trials factor, hypertests and the BumpHunter. In *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN,Geneva, Switzerland 17-20 January 2011*, 2011, 1101.0390.

[196] Morad Aaboud et al. A strategy for a general search for new phenomena using data-derived signal regions and its application within the ATLAS experiment. *Submitted to: Eur. Phys. J.*, 2018, 1807.07447.

[197] MUSIC – An Automated Scan for Deviations between Data and Monte Carlo Simulation. 2008.

[198] The ATLAS collaboration. A model independent general search for new phenomena with the ATLAS detector at $\sqrt{s} = 13$ TeV. 2017.

[199] Pouya Asadi, Matthew R. Buckley, Anthony DiFranzo, Angelo Monteux, and David Shih. Digging Deeper for New Physics in the LHC Data. *JHEP*, 11:194, 2017, 1707.05783.

[200] Model Unspecific Search for New Physics in pp Collisions at sqrt(s) = 7 TeV. 2011.

[201] Marco Farina, Yuichiro Nakai, and David Shih. Searching for New Physics with Deep Autoencoders. 2018, 1808.08992.

[202] Theo Heimel, Gregor Kasieczka, Tilman Plehn, and Jennifer M. Thompson. QCD or What? *SciPost Phys.*, 6(3):030, 2019, 1808.08979.

[203] Jan Hajer, Ying-Ying Li, Tao Liu, and He Wang. Novelty Detection Meets Collider Physics. 2018, 1807.10261.

[204] Olmo Cerri, Thong Q Nguyen, Maurizio Pierini, Maria Spiropulu, and Jean-Roch Vlimant. Variational autoencoders for new physics mining at the large hadron collider. *Journal of High Energy Physics*, 2019(5):36, 2019.

[205] Oliver Knapp, Guenther Dissertori, Olmo Cerri, Thong Q Nguyen, Jean-Roch Vlimant, and Maurizio Pierini. Adversarially learned anomaly detection on cms open data: re-discovering the top quark. *arXiv preprint arXiv:2005.01598*, 2020.

[206] Melissa van Beekveld, Sascha Caron, Luc Hendriks, Paul Jackson, Adam Leinweber, Sydney Otten, Riley Patrick, Roberto Ruiz De Austri, Marco Santoni, and Martin White. Combining outlier analysis algorithms to identify new physics at the LHC. *JHEP*, 09:024, 2021, 2010.07940. Thank you to Adam Leinweber for providing these plots.

[207] Ivo D Dinov. Expectation maximization and mixture modeling tutorial. ucla: Statistics online computational resource, 12 2008.

[208] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[209] Ernst Wit, Edwin van den Heuvel, and Jan-Willem Romeijn. 'all models are wrong...': an introduction to model uncertainty. *Statistica Neerlandica*, 66(3):217–236, 2012, https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9574.2012.00530.x.

[210] F. T. Liu, K. M. Ting, and Z. Zhou. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422, 2008.

[211] Sal Borelli. *https://en.wikipedia.org/wiki/Isolation_forest*, September 2019.

[212] Sal Borelli. *https://en.wikipedia.org/wiki/Isolation_forest*, September 2019.

[213] Mark A. Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37(2):233–243, 1991, https://aiche.onlinelibrary.wiley.com/doi/pdf/10.1002/aic.690370209.

[214] I. Goodfellow, Yoshua Bengio, and Aaron C. Courville. Deep learning. *Nature*, 521:436–444, 2015.

[215] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[216] Diederik P. Kingma and Max Welling. An introduction to variational autoencoders. *CoRR*, abs/1906.02691, 2019, 1906.02691.

[217] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 03 1951.

[218] Solomon Kullback. *Information Theory and Statistics*. Wiley, New York, 1959.

[219] Irina Higgins, Loïc Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew M Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. In *ICLR*, 2017.

[220] Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2013, 1312.6114.

[221] Gregory Gundersen. The reparametrization trick. April 2018.

[222] M.V. Valueva, N.N. Nagornov, P.A. Lyakhov, G.V. Valuev, and N.I. Chervyakov. Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. *Mathematics and Computers in Simulation*, 177:232 – 243, 2020.

[223] Wei Zhang, Kazuyoshi Itoh, Jun Tanida, and Yoshiki Ichioka. Parallel distributed processing model with local space-invariant interconnections and its optical architecture. *Appl. Opt.*, 29(32):4790–4797, Nov 1990.

[224] Christopher D. Brown and Herbert T. Davis. Receiver operating characteristics curves and related decision measures: A tutorial. *Chemometrics and Intelligent Laboratory Systems*, 80(1):24 – 38, 2006.

[225] Sydney Otten, Sascha Caron, Wieske de Swart, Melissa van Beekveld, Luc Hendriks, Caspar van Leeuwen, Damian Podareanu, Roberto Ruiz de Austri, and Rob Verheyen. Event Generation and Statistical Sampling for Physics with Deep Generative Models and a Density Information Buffer. 1 2019, 1901.00875.

[226] G. Brooijmans et al. Les Houches 2019 Physics at TeV Colliders: New Physics Working Group Report. In *11th Les Houches Workshop on Physics at TeV Colliders: PhysTeV Les Houches*, 2 2020, 2002.12220.

[227] Andy Buckley, James Ferrando, Stephen Lloyd, Karl Nordström, Ben Page, Martin Rüfenacht, Marek Schönherr, and Graeme Watt. LHAPDF6: parton density access in the LHC precision era. *Eur. Phys. J. C*, 75:132, 2015, 1412.7420.

[228] Xiangyi Cui et al. Dark Matter Results From 54-Ton-Day Exposure of PandaX-II Experiment. *Phys. Rev. Lett.*, 119(18):181302, 2017, 1708.06917.

[229] E. Aprile et al. Conceptual design and simulation of a water Cherenkov muon veto for the XENON1T experiment. *JINST*, 9:P11006, 2014, 1406.2374.

[230] D.S. Akerib et al. LUX-ZEPLIN (LZ) Conceptual Design Report. 9 2015, 1509.02910.

[231] J. Aalbers et al. DARWIN: towards the ultimate dark matter detector. *JCAP*, 11:017, 2016, 1606.07001.

[232] Nasim Fatemighomi. DEAP-3600 dark matter experiment. In *35th International Symposium on Physics in Collision*, 9 2016, 1609.07990.

[233] C.E. Aalseth et al. DarkSide-20k: A 20 tonne two-phase LAr TPC for direct dark matter detection at LNGS. *Eur. Phys. J. Plus*, 133:131, 2018, 1707.08145.

[234] J. Calvo et al. Commissioning of the ArDM experiment at the Canfranc underground laboratory: first steps towards a tonne-scale liquid argon time projection chamber for Dark Matter searches. *JCAP*, 03:003, 2017, 1612.06375.

[235] G. Adhikari et al. Search for a Dark Matter-Induced Annual Modulation Signal in NaI(Tl) with the COSINE-100 Experiment. *Phys. Rev. Lett.*, 123(3):031302, 2019, 1903.10098.

[236] J. Amaré et al. First Results on Dark Matter Annual Modulation from the ANAIS-112 Experiment. *Phys. Rev. Lett.*, 123(3):031301, 2019, 1903.03973.

[237] G. Angloher et al. The COSINUS project - perspectives of a NaI scintillating calorimeter for dark matter search. *Eur. Phys. J. C*, 76(8):441, 2016, 1603.02214.

[238] Emily Shields, Jingke Xu, and Frank Calaprice. SABRE: A New NaI(T1) Dark Matter Direct Detection Experiment. *Phys. Procedia*, 61:169–178, 2015.

[239] Francis Froborg. SABRE: WIMP modulation detection in the northern and southern hemisphere. *J. Phys. Conf. Ser.*, 718(4):042021, 2016, 1601.05307.

[240] Charanjit K Khosa, Lucy Mars, Joel Richards, and Veronica Sanz. Convolutional neural networks for direct detection of dark matter. *Journal of Physics G: Nuclear and Particle Physics*, 47(9):095201, Jul 2020.

[241] Xinyue Zhang, Yanfang Wang, Wei Zhang, Yueqiu Sun, Siyu He, Gabriella Contardo, Francisco Villaescusa-Navarro, and Shirley Ho. From Dark Matter to Galaxies with Convolutional Networks. 2 2019, 1902.05965.

[242] Luisa Lucie-Smith, Hiranya V. Peiris, and Andrew Pontzen. An interpretable machine learning framework for dark matter halo formation. *Mon. Not. Roy. Astron. Soc.*, 490(1):331–342, 2019, 1906.06339.

[243] Mauro Bernardini, Lucio Mayer, Darren Reed, and Robert Feldmann. Predicting dark matter halo formation in N-body simulations with deep regression networks. *Mon. Not. Roy. Astron. Soc.*, 496(4):5116–5125, 2020, 1912.04299.

[244] Mikael Kuusela, Tommi Vatanen, Eric Malmi, Tapani Raiko, Timo Aaltonen, and Yoshikazu Nagai. Semi-Supervised Anomaly Detection - Towards Model-Independent Searches of New Physics. *J. Phys. Conf. Ser.*, 368:012032, 2012, 1112.3329.

[245] Anders Andreassen, Benjamin Nachman, and David Shih. Simulation Assisted Likelihood-free Anomaly Detection. *Phys. Rev. D*, 101(9):095004, 2020, 2001.05001.

[246] Benjamin Nachman and David Shih. Anomaly Detection with Density Estimation. *Phys. Rev. D*, 101:075042, 2020, 2001.04990.

[247] Jack H. Collins, Kiel Howe, and Benjamin Nachman. Extending the search for new resonances with machine learning. *Phys. Rev. D*, 99(1):014038, 2019, 1902.02634.

[248] Lucio Mwinmaarong Dery, Benjamin Nachman, Francesco Rubbo, and Ariel Schwartzman. Weakly Supervised Classification For High Energy Physics. *J. Phys. Conf. Ser.*, 1085(4):042006, 2018.

[249] Jack H. Collins, Kiel Howe, and Benjamin Nachman. Anomaly Detection for Resonant New Physics with Machine Learning. *Phys. Rev. Lett.*, 121(24):241803, 2018, 1805.02664.

[250] Charanjit K. Khosa and Veronica Sanz. Anomaly Awareness. 7 2020, 2007.14462.

[251] E. Aprile et al. Excess electronic recoil events in XENON1T. *Phys. Rev. D*, 102(7):072004, 2020, 2006.09721.

[252] https://github.com/LucyMars/SearchForDarkMatter .

[253] https://github.com/JelleAalbers/wimprates.

[254] Jelle Aalbers. *Dark matter search with XENON1T*. PhD thesis, Amsterdam U., 2018.

[255] https://github.com/XENON1T/laidbax.

[256] https://github.com/XENON1T/pax.

[257] Vincent Dumoulin and Francesco Visin. A guide to convolution arithmetic for deep learning, 2018, 1603.07285.